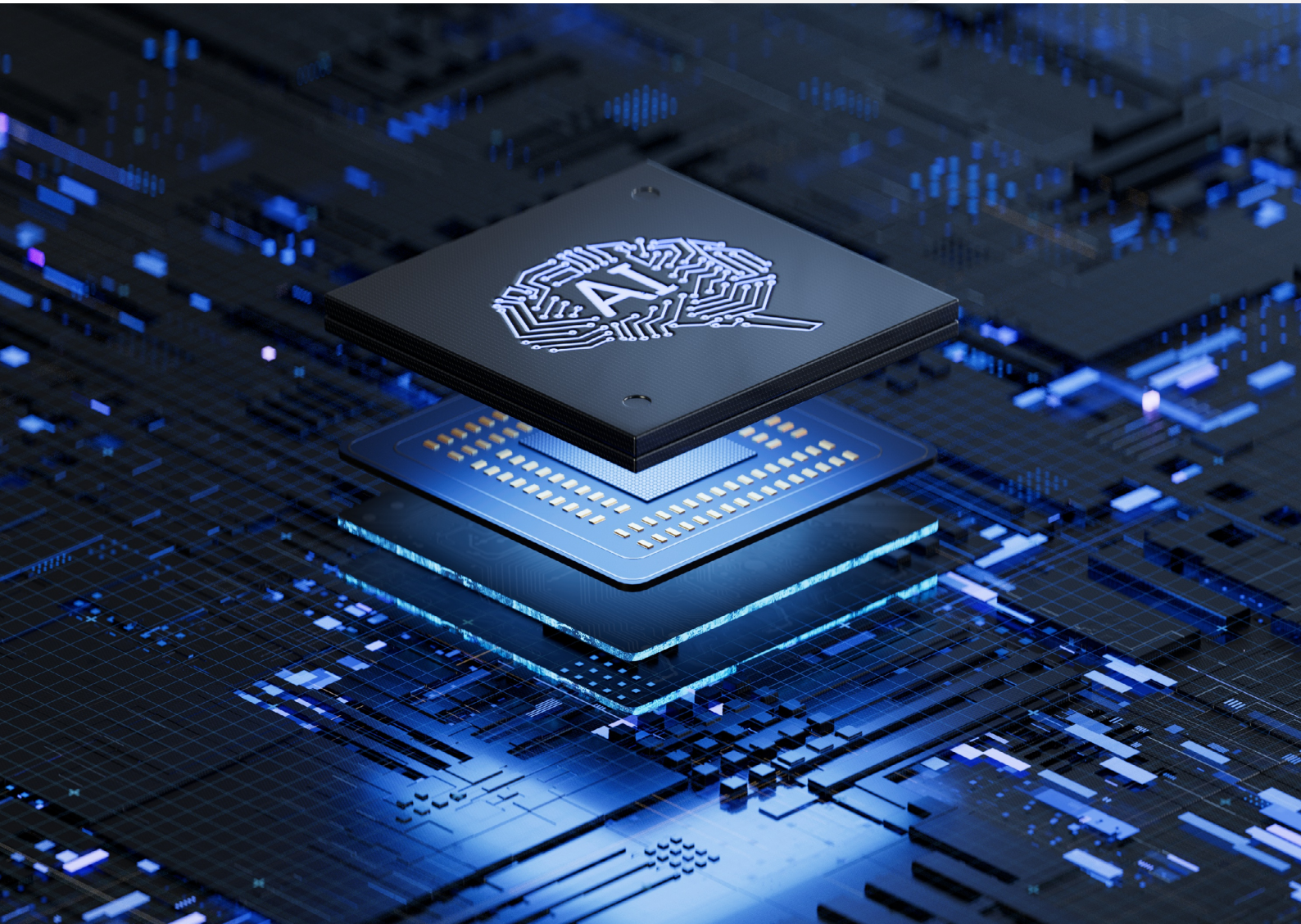
opentext™

WHITE PAPER

Future proofing with trusted AI



Contents

Introduction 3

The role of governance in trusted AI 4

Exploring trust-critical AI use cases 6

Enhancing trust through security measures 7

Conclusion 11

Introduction

Artificial intelligence (AI) offers immense opportunities and is reshaping today's business landscape. AI tools are highly actionable and deployable, driving compelling business outcomes and efficiencies. A significant 64% of businesses note that AI helps increase their overall productivity, and 53% use AI to improve production processes.[1] Given the impact today's AI investments will have, embracing AI has become a strategic imperative for businesses to remain competitive. To reap the benefits, AI systems must be reliable, transparent, and ethical, providing better insights for more informed decision-making.

Despite the promise AI holds, its adoption also brings significant risks to consider, such as data poisoning, privacy breaches, and security threats like remote code execution. For example, consider a global manufacturing company that integrates AI into its production processes. The AI system is trained to optimize supply chain management, predict maintenance needs, improve quality control, delivering on the promised outcomes of cost savings and increased productivity. However, if the company experiences a data breach that compromises the AI system, it would likely lead to faulty predictions and production delays.

Given the impact today's AI investments will have, embracing AI has become a strategic imperative for businesses to remain competitive.

As businesses increasingly rely on AI, these potential risks cannot be overlooked. Indeed, 54% of companies list adapting to AI as a top business challenge they'll need to navigate in the next year.[2] The key is charting a course for strategic AI implementation with a focus on adopting a framework for trusted AI. The trustworthiness of AI systems hinges on securing the digital landscape.

In this whitepaper, we'll explore how trusted AI practices fortify enterprises for AI adoption, ensuring resilience, data integrity, and competitive edge in evolving digital landscapes.

The role of governance in trusted AI

AI governance is invaluable for achieving compliance, trust, and efficiency in the development and application of AI technologies.

Without proper oversight, AI can cause significant harm, underscoring the importance of governance in managing the risks associated with advanced AI. AI governance is invaluable for achieving compliance, trust, and efficiency in the development and application of AI technologies.

AI governance refers to the framework, policies, and regulations that guide the development, deployment, and management of AI technologies to ensure ethical, responsible, and secure practices. It helps to ensure that AI is developed and deployed responsibly, ethically, and securely, addressing risks while balancing AI innovation with safety.

How governance accelerates and optimizes AI adoption

AI governance is crucial for managing the rapid advancements in AI technology, particularly with the emergence of generative AI. Generative AI, which includes technologies capable of creating new content and solutions such as text, images, and code, has vast potential across many use cases. Overall, AI governance defines the extent to which algorithms can influence daily life and establishes who is responsible for overseeing their operation. Key principles of responsible AI governance that can help organizations optimize their adoption and use include: Through information management, businesses can consolidate and integrate information so it can be managed transparently throughout its entire lifecycle.

Ethical AI practices

Establishing ethical guidelines ensures that AI systems operate transparently and fairly, mitigating biases and promoting inclusivity.

Establishing ethical guidelines ensures that AI systems operate transparently and fairly, mitigating biases and promoting inclusivity. For example, the FTC issued a warning in February 2024 about companies covertly changing their AI guidelines to mislead stakeholders or bypass ethical standards.[3] This emphasizes the need for clear, consistent, and publicly available ethical practices to build trust and accountability in AI applications.

Data security and privacy

Implementing robust data security measures protects sensitive information and ensures compliance with privacy regulations, fostering trust among users and stakeholders. By prioritizing cybersecurity and data privacy, organizations can reassure users that their personal information is safe and handled with care, which is essential for maintaining trust in AI systems.

Accountability and transparency

Clear accountability frameworks and transparent processes enable organizations to trace AI decisions, ensuring responsibility and fostering trust. This involves documenting decision-making processes, maintaining detailed logs, and providing explanations for AI-driven outcomes.

Transparency in AI operations allows stakeholders to understand how decisions are made, who is responsible for them, and how to address potential issues. This practice enhances the credibility and reliability of an organization's AI systems.

Continuous monitoring and evaluation

Regular monitoring and evaluation of AI systems help identify and mitigate risks, ensuring ongoing compliance with governance standards. This includes monitoring the quality of the data used for training and detecting security vulnerabilities like data poisoning, which can significantly reduce the accuracy and reliability of AI engines.

Continuous oversight allows for the detection of biases, performance issues, and security threats, ensuring AI systems remain effective, fair, and secure. By routinely assessing AI systems, organizations can adapt to evolving standards, address emerging risks, and ensure their AI technologies remain ethical, reliable, and aligned with governance policies.management fundamentally integrates the parts of your business that matter, delivering the right information to the right people at the right time.

Cybersecurity in trusted AI

A common metaphor that "data is the new oil" highlights how valuable data is as a resource that fuels economies, countries, and daily lives. And there is perhaps no other system with the capacity and appetite for data quite like AI.

However, the sources of that data, the methods of its processing, and the resulting outputs all require robust identity and security measures. Understandably, many people are worried about this in practice. Indeed, 81% are concerned about the security risks associated with ChatGPT and generative AI.[4]

AI data is susceptible to various threats and vulnerabilities. Malicious actors are skilled at attacking AI models and finding ways to corrupt the training data to skew outcomes. Data quality issues can be inadvertently introduced during training, which can significantly impair model performance. Likewise, malicious and naturally occurring data corruption can produce erroneous outputs during inference and potential leaks of sensitive data, which poses a substantial security and privacy risk for organizations.

Thus, strong cybersecurity measures are indispensable in AI technologies. Without a secure foundation, AI projects are vulnerable to failure, as evidenced by the fact that only 54% of AI initiatives progress from pilot stages to full production.[5] By integrating robust security practices into their AI governance frameworks, organizations can strategically establish trusted AI.

Before we delve into the cybersecurity framework for trusted AI, let's first explore some key industry use cases where AI delivers benefits while also demanding trust.

Exploring trust-critical AI use cases

AI encompasses machine learning (ML), deep learning, and generative AI, collectively revolutionizing industries by enhancing efficiency and scalability.

Instances of AI pervade society, encompassing a wide range of applications such as chatbots, financial fraud detection systems, and navigation software. AI encompasses machine learning (ML), deep learning, and generative AI, collectively revolutionizing industries by enhancing efficiency and scalability. These technologies continue to redefine operational standards across diverse sectors.

As with any technology, it was not long before bad actors began leveraging AI for nefarious purposes. This misuse underscores the urgency for organizations to adopt trusted AI ecosystems that prioritize security, transparency, and ethical standards. By establishing robust frameworks for trusted AI, organizations can harness the benefits of AI while mitigating associated risks.

Let's explore applications of trusted AI in high-consequence environments.

Healthcare

In healthcare, AI plays a transformative role by enabling personalized treatment plans and predictive diagnostics. AI algorithms can analyze vast amounts of patient data to identify patterns and predict outcomes, helping healthcare providers deliver more precise and effective care. In fact, 20% of healthcare organizations have already adopted AI models for their healthcare solutions, underscoring the growing reliance on AI to enhance medical outcomes and operational efficiency.[6]

However, the stakes are incredibly high. A flawed AI diagnosis could lead to incorrect treatment, putting patient lives at risk. Additionally, breaches in patient data privacy could erode trust in the healthcare system and lead to significant legal and HIPAA compliance repercussions. The potential for AI to perpetuate biases also exists, potentially resulting in unequal treatment across different patient demographics.

Trusted AI in this context requires confidence in the accuracy and reliability of AI-driven medical recommendations. Ensuring trust involves rigorous security for AI models and continuous monitoring to mitigate data corruption and ensure patient safety. Robust cybersecurity measures must be in place to protect sensitive health information and maintain patient confidentiality.

Financial services

Within financial services, AI is instrumental in enhancing fraud detection and risk modeling capabilities. In fact, AI technologies have helped reduce fraud losses by 20% in financial institutions.[7] AI-powered systems analyze transactional data in real-time to detect anomalies and potential fraud, improving operational efficiency and safeguarding against financial crimes. The benefits are so vast that 60% of banks and financial institutions have already adopted AI and machine learning technologies.[8]

However, AI-related risks can significantly impact financial operations. Incorrectly flagged transactions can disrupt legitimate business activities while undetected fraudulent activities can result in substantial financial losses. Additionally, breaches in data security can expose sensitive customer information, leading to identity theft and financial fraud. The opaque nature of some AI models can also make it difficult to understand and explain decisions, which can erode trust among customers and regulators.

Given these challenges, the implementation of trusted AI becomes a cornerstone for effectively managing the risks. Trust is crucial in the AI's ability to protect sensitive financial data and provide accurate assessments of risk.

Critical infrastructure

For critical infrastructure sectors like communications, transportation, and energy, the AI use cases are vast. In the energy sector, AI-driven technologies optimize grid management and predictive maintenance strategies. Trust in AI's predictions is essential to ensure the efficiency, reliability, and sustainability of energy operations. AI algorithms analyze data from sensors and historical patterns to forecast demand, optimize energy distribution, and predict equipment failures before they occur.

Yet, an inaccurate prediction can lead to grid failures, causing widespread power outages and affecting millions of people. Cyber attacks on AI systems can disrupt energy supply, endangering public safety and national security. Furthermore, incorrect maintenance predictions can lead to premature equipment failures or unnecessary maintenance costs, affecting operational efficiency and profitability.

The value of trusted AI in the energy sector lies in its ability to ensure the safety, integrity, and reliability of energy systems. By leveraging AI that stakeholders trust, the energy sector can maintain stable operations, respond swiftly to changing conditions, and uphold the highest standards of service and security. This trust is crucial for the continued adoption and integration of AI technologies, enabling the energy sector to meet future demands and challenges with confidence.

These examples illustrate how AI applications in high-consequence environments are reshaping industries by improving decision-making, operational efficiency, and risk management. At the core of these advancements lies the critical importance of cybersecurity. Adopting robust security measures is foundational to trusted AI implementations, safeguarding sensitive data, protecting against cyber threats, and preserving the integrity and reliability of AI systems.

The ability to trust AI outputs will depend on factors such as the level of data quality and data protection.

Enhancing trust through security measures

Building trust through robust security measures is foundational to ensuring the integrity, reliability, and ethical use of AI technologies across industries. ISACA aptly notes that "the ability to trust AI outputs will depend on factors such as the level of data quality and data protection. The AI model must be developed based on trusted code and transparency to include operation decision making.

The end user or consumer of AI-enabled services trusts the AI algorithms driving decisions are proven to be accurate, protects the privacy of the end user, is secure, and is free from bias."

To turn this into reality, the pragmatic framework to trusted AI is comprised of three pillars: trust in data, trust in access, and trust in applications—each essential for safeguarding sensitive information, securely managing access, and ensuring the secure deployment of AI applications.

Trust in data

Trusted data forms the bedrock of AI applications, directly impacting the accuracy and reliability of AI-driven insights and decisions. Poor quality or tampered training data can lead to flawed outputs, compromising the effectiveness of AI tools and eroding trust in their outcomes. And poor data quality destroys business value. Organizations report that poor data quality is responsible for an average of $15 million per year in losses.[9]

Securing trust in AI data involves implementing rigorous data security practices that go beyond mere protection to encompass proactive management and governance. It starts with comprehensive data discovery processes that identify all sources of data that go into the organization's AI systems. This step is crucial for gaining visibility into data flows and pinpointing potential vulnerabilities that could compromise AI data privacy or integrity.

Data sensitivity-centric security approaches... ensure that data protection measures align closely with organizational risk management strategies, elevating trust in AI applications.

Once identified, data should be classified based on sensitivity levels, enabling organizations to apply tailored security measures. This categorization ensures that appropriate encryption techniques are applied to protect data both at rest and in transit, safeguarding it from unauthorized access and potential breaches. Secure storage practices also play a crucial role in maintaining data integrity and confidentiality by ensuring that data remains protected against unauthorized access, data breaches, and cyber threats throughout its lifecycle.

Robust access controls and auditing mechanisms are then employed throughout the data lifecycle, ensuring that only authorized personnel can access and manage sensitive information. This approach not only enhances security but also builds confidence in the reliability and integrity of AI-driven processes.

Data sensitivity-centric security approaches, such as enhance data security by focusing on securing data at rest, in use and in transit based on its sensitivity and criticality. This targeted approach ensures that data protection measures align closely with organizational risk management strategies, elevating trust in AI applications and maintaining compliance with regulatory standards.

By implementing these proactive measures, organizations can mitigate risks associated with data quality and security, laying a solid foundation for trustworthy AI applications. This systematic approach not only safeguards sensitive information but also reinforces organizational resilience in the face of evolving cyber threats.

Trust in access

Ensuring trust in access control is paramount for AI systems as the proliferation of identities complicates governance and increases the risk of unauthorized access. Non-human identities (machine identities, service accounts, API Security) which are the primary driver of identity growth and are considered the riskiest type of identity, significantly increase the vulnerability of AI models. Notably, 90% of organizations experienced two or more identity-related breaches in the past year alone, highlighting the pervasive threat landscape.[10]

These risks underscore the importance of trust in AI access as a foundational pillar. Unauthorized access not only compromises data integrity but also undermines the reliability of AI-driven decisions, impacting organizational performance and trustworthiness. To mitigate these unauthorized access attempts and identity-related breaches, organizations need to establish robust identity and access management (IAM) capabilities.

Effective Identity and Access Management (IAM) is essential for maintaining trust in AI systems.

Ensuring trust through IAM

Effective Identity and Access Management (IAM) is essential for maintaining trust in AI systems. IAM frameworks encompass role-based and risk-based adaptive access controls, multi-factor authentication (MFA), identity governance and administration (IGA), and privileged access management (PAM). These mechanisms ensure that access to AI systems is granted based on the principle of least privilege (PoLP), which dictates that individuals should have only the minimum level of access necessary to perform their job functions. This approach minimizes exposure to potential threats and unauthorized use.

IAM's role-based access controls (RBAC) assign permissions based on users' roles within the organization, limiting access to AI data and functionalities strictly necessary for their job functions. Risk-based adaptive access controls employ behavioral analytics to assess user behavior patterns and detect anomalies, triggering additional security measures when suspicious activities are identified.

Strengthening security with multi-factor authentication (MFA)

MFA enhances security for AI systems by requiring users to authenticate their identity through multiple credentials, such as passwords, biometrics, or one-time passcodes. This layered authentication approach mitigates the risk of unauthorized access, particularly crucial for sensitive AI applications that handle confidential or proprietary data.

Manage application and data access permissions with IGA

Identity Governance and Administration (IGA) helps protect data with simplified compliance and user access review processes. By continuously verifying identities, managing user lifecycles, and applying granular access policies, this approach minimizes exposure to potential threats and unauthorized use, thereby enhancing the overall security and integrity of AI systems.

Managing privileged access

Privileged access management (PAM) governs access to critical AI systems and data through continuous monitoring and control of privileged accounts. This rigorous oversight ensures that only authorized administrators can execute administrative tasks. By enforcing stringent access policies and conducting regular audits of privileged activities, organizations can thwart insider threats and diminish the risk of outsider malicious activities.

Collectively, practices for trust in AI access are indispensable for safeguarding data integrity, maintaining operational continuity, and upholding regulatory compliance. By implementing comprehensive IAM strategies and leveraging advanced authentication methods, organizations can mitigate identity-related risks and fortify the foundation of trustworthy AI deployments.

Trust in applications

Businesses are adopting AI applications at a breakneck pace driven by vast opportunities to meet market demand, enhance customer engagement, and gain a competitive edge. Organizations report that lead business drivers for adopting AI applications include increased tech resilience and uptime (52%), improving management at scale (51%), and reducing staff workload (50%).[11]

However, alongside the rapid adoption of AI applications comes an increase in cyber threats. Last year, 20% of businesses experienced attacks on their AI applications, highlighting the growing importance of implementing robust security measures.[12]

These risks underscore the critical need for secure coding practices to ensure trust in AI applications. This strategy not only prevents security risks in the applications but also enhances the overall reliability and security of AI systems. Techniques such as static application security testing (SAST) play a crucial role in identifying and mitigating security flaws in AI-generated code before the application is deployed. Use of AI to ensure an organization is adhering to OSS licensing agreement and preventing supply chain attacks. Additionally, by conducting thorough scans for vulnerabilities specific to large language models (LLMs), organizations can proactively address and resolve weaknesses that could otherwise be exploited by malicious actors.

Additionally, AI is increasingly being employed to advance application security with greater speed and accuracy. AI-driven security solutions can autonomously detect and respond to threats in real-time, offering a proactive defense against evolving cyber threats. This integration of AI in security operations not only enhances response times but also augments the effectiveness of the organization's security measures.

Adopting these application security measures empowers organizations to instill trust in their AI applications. By doing so, businesses not only safeguard their operations but also enhance the reliability and security of AI applications. This strategy fosters confidence among stakeholders, ensuring that AI-driven innovations continue to deliver value while mitigating potential risks.

Explore further how trusted AI practices can secure your AI investments.

To learn more, go to:

IGA Buyers Guide ›

Data Security ›

Conclusion

AI's rapid integration into industries promises transformative benefits, from enhancing operational efficiency to revolutionizing customer engagement. However, realizing these benefits hinges on taking pragmatic steps to manage and secure AI systems.

At the heart of AI adoption lies the imperative for ethical guidelines and governance. Ensuring transparency in AI algorithms, mitigating biases, and promoting inclusivity are crucial steps towards a mature governance framework. Likewise, trusted AI practices, centered around cybersecurity measures, are indispensable for organizations navigating the complexities of AI adoption. By prioritizing trust in data integrity, access control, and application security, businesses can improve resilience, safeguard sensitive information, and maintain a competitive edge. These practices not only mitigate risks associated with data breaches and unauthorized access but also instill confidence among stakeholders in the reliability and ethical use of AI technologies.

Looking forward, the adoption of robust cybersecurity frameworks remains crucial for organizations to harness the full potential of AI while mitigating the cyber risks. By embedding trust into every facet of AI deployment—from data handling to access management and application security—businesses can ensure that their AI initiatives deliver value ethically, securely, and sustainably. This commitment not only enhances operational efficiency but also positions organizations at the forefront of innovation, driving growth and differentiation in an increasingly digital world.

1 Forbes. How Businesses Are Using Artificial Intelligence In 2024.

2 Certinia. Global Service Dynamics 2024 Report.

3 FTC. AI (and other) Companies: Quietly Changing Your Terms of Service Could Be Unfair or Deceptive. 2024.

4 CSO. Survey reveals mass concern over generative AI security risks.

5 Forbes. Why AI Methods Need Heightened Security. 2024.

6 Radix. The Statistical Landscape of AI Adoption in Healthcare. 2024.

7 WifiTalents. Ai In Finance Industry Statistics. 2024.

8 WifiTalents. Ai In Finance Industry Statistics. 2024.

9 Forbes. Flying Blind: How Bad Data Undermines Business. 2022.

10 SecurityToday. Study: 90 Percent of Organizations Experienced an Identity-Related Incident. 2024.

11 Statistica. Key artificial intelligence (AI) application and infrastructure upgrade drivers in businesses worldwide as of 2023.

12 Forbes. AI Models Under Attack: Protecting Your Business From AI Cyberthreats. 2023.

Copyright © 2026 Open Text • 06.26 | 262-000116-002

opentext™