ZER0TRUSTED.AI

ZeroTrusted.ai’s Alignment with the Blueprint for an AI Bill of Rights

Waylon Krush
CEO of ZeroTrusted.ai

INTRODUCTION

Artificial Intelligence (AI) and automated systems have brought significant advancements, revolutionizing industries from healthcare to finance. However, with the rise of AI, concerns regarding privacy, fairness, and security have also emerged. To address these issues, the **Blueprint for an AI Bill of Rights**, introduced by the White House Office of Science and Technology Policy, outlines five key principles aimed at protecting civil liberties and ensuring that AI is used responsibly and ethically.

ZeroTrusted.ai is a cutting-edge cybersecurity company dedicated to ensuring the safe, secure, and trustworthy use of AI systems. Our solutions, such as AI Health Check and AI Governance System (AGS), align closely with the core principles of the AI Bill of Rights, including safety, algorithmic discrimination protection, data privacy, transparency, and human alternatives.

Safe and Effective Systems

Blueprint Principle:
AI systems should be designed with input from diverse communities and stakeholders, undergoing thorough pre-deployment testing, risk identification, and ongoing monitoring to ensure they are safe and effective.

How ZeroTrusted.ai Aligns:
ZeroTrusted.ai's AI Health Check continuously monitors AI systems, including LLMs (Large Language Models), SLMs (Small Language Models), Vector Databases, and AI Agents, to ensure their security, reliability, and privacy. Through real-time tracking of potential risks and vulnerabilities, our solution provides organizations with an ongoing assessment of their AI systems' health, ensuring that these systems remain effective and safe for their intended use.

Our STIGed container with an optional Version of Llama, a STIG-compliant (Security Technical Implementation Guide) version of AI models, further guarantees that AI is deployed securely, in compliance with industry standards, and free from potential vulnerabilities. We also work with your existing AI assets in both cloud and on-premise and even closed networks.

Algorithmic Discrimination Protections

Blueprint Principle:
Automated systems must be designed to prevent discrimination based on race, color, ethnicity, sex, gender identity, disability, and other legally protected characteristics.

How ZeroTrusted.ai Aligns:

At ZeroTrusted.ai, we understand the importance of equitable AI deployment. Our AI Governance System (AGS) ensures that AI models are designed, deployed, and maintained in a way that prevents algorithmic discrimination. Through continuous evaluation, disparity testing, and proactive assessments, ZeroTrusted.ai helps organizations detect and mitigate algorithmic biases, ensuring fair outcomes for all individuals and communities.

The AGS offers transparent reporting and independent evaluations to confirm that these protections are upheld. We also monitor for any data drift or model input or output that changes overtime that may indicate updates or training issues that could result in algorithmic discrimination. Lastly, we have an optional and organizationally tailored AI-Ethics based dirty word list to prevent jail breaking of the models.

Data Privacy

Blueprint Principle:
Individuals should have control over how their data is collected, used, and shared, and AI systems should incorporate privacy protections by design.

How ZeroTrusted.ai Aligns:
ZeroTrusted.ai prioritizes data sovereignty and privacy as core elements of our approach. We employ Encryption & Anonymization techniques that protect AI data in transit, at rest, and during processing. AI models are trained and fine-tuned using only the minimum amount of data necessary, adhering to privacy by design principles.

Our system gives organizations control over how data is shared and used within AI models, preventing unauthorized data collection or reuse. We also let the organization set data privacy requirements in accordance with their privacy policy or compliance requirements around PII, PHI, NIST 800-53, GDPR, PCI, and several international privacy requirements. Moreover, ZeroTrusted.ai's API Integration allows organizations to retain full ownership of their fine-tuning data, ensuring compliance with data privacy regulations and giving them the ability to port their AI configurations when switching providers. This provides a full chain of custody of each token (input and output), response, organizational specific fine tuning data, and any possible deviations against organizational data privacy requirements to include international requirements such as GDPR and

Notice and Explanation

Blueprint Principle:

Users should be informed when AI systems are being used and provided with clear explanations of how these systems operate and how outcomes are determined.

How ZeroTrusted.ai Aligns:

ZeroTrusted.ai enables full transparency by providing plain language documentation and explanations of how AI models operate. Through real-time logging and reporting, organizations can access clear, understandable information about the decisions made by AI systems, how inputs are processed, and the logic behind any given outcome. This transparency helps organizations comply with regulatory requirements and offers users the ability to understand the implications of automated decision-making.

Additionally, our AI Health Check system continuously documents and explains how the AI operates, flagging any anomalies or changes in performance that could impact decision-making.

Human Alternatives, Consideration, and Fallback

Blueprint Principle:
Users should have access to human alternatives when interacting with automated systems and be able to contest or appeal decisions made by AI.

How ZeroTrusted.ai Aligns:
ZeroTrusted.ai ensures that human consideration is always an option. Our Zero Trust Architecture empowers organizations to provide human oversight for AI-driven decisions, especially in sensitive contexts such as healthcare, employment, and finance. Users have the ability to opt out of automated decisions in favor of human alternatives, and the system supports fallback mechanisms that allow users to escalate and challenge AI outcomes.

ZeroTrusted.ai also offers tailored oversight and intervention capabilities, enabling timely human review in situations where AI errors or high-risk decisions are detected. This ensures that no critical decisions are made without adequate human consideration. We also allow organizational users to tailor responses based on human-reinforced feedback.

Conclusion

ZeroTrusted.Ai is committed to meeting the standards set forth in the Blueprint for an AI Bill of Rights. By focusing on Safe and Effective Systems, Algorithmic Discrimination Protections, Data Privacy, Notice and Explanation, and Human

Alternatives, ZeroTrusted.Ai ensures that AI is used in a way that protects individuals' rights and enhances trust in automated systems.
As AI continues to shape the future, ZeroTrusted.ai stands at the forefront of AI governance, empowering organizations to deploy secure, reliable, and compliant AI systems that work for the American people.

For more information on how ZeroTrusted.ai can help your organization meet the Blueprint for an AI Bill of Rights, please contact us at waylon@zerotrusted.ai or visit our website at www.zerotrusted.ai.

Also note our AGS and AI Health Check can be tailored to your specific AI Cyber and Privacy requirements, but we also currently support:

ZeroTrusted.ai supports the following security and privacy compliance requirements:

1. Payment Card Industry Data Security Standard (PCI DSS)
   - Ensures secure handling of credit card information.
2. Open Web Application Security Project (OWASP) Top 10
   - Focuses on addressing the most critical security risks to web applications.
3. USA AI Bill of Rights
   - Provides a framework for the ethical and responsible use of AI, focusing on safety, fairness, and transparency.
4. NIST AI 100-01
   - National Institute of Standards and Technology's framework for managing AI risks and ensuring AI system trustworthiness.
5. NIST AI 600-1
   - An additional NIST framework focusing on risk management for AI systems, ensuring responsible development and deployment.
6. NIST SP 800-53
   - A set of security and privacy controls for federal information systems and organizations.
7. OMB M-24-10
   - Office of Management and Budget's directive for improving federal cybersecurity and AI system resilience.
8. CWE/SANS Top 25
   - Identifies the most dangerous software errors that can lead to security vulnerabilities.
9. Personally Identifiable Information (PII) Data Compliance
   - Ensures proper protection and handling of sensitive personal data.
10. Protected Health Information (PHI) Data Compliance
    - Protects patient health information, especially in healthcare systems.
11. General Data Protection Regulation (GDPR)

- European Union regulation for data privacy and protection of individuals' personal data.

12. California Consumer Privacy Act (CCPA)
    - California law providing data privacy rights for California residents.
13. Health Insurance Portability and Accountability Act (HIPAA)
    - U.S. law focused on protecting the privacy and security of health information.
14. Health Information Technology for Economic and Clinical Health Act (HITECH)
    - Promotes the adoption of health information technology and strengthens HIPAA enforcement.
15. HITRUST
    - A certifiable framework that addresses security, privacy, and regulatory compliance in the healthcare industry.
16. Gramm-Leach-Bliley Act (GLBA)
    - U.S. law that requires financial institutions to explain their information-sharing practices and protect customer data.
17. Lei Geral de Proteção de Dados (LGPD)
    - Brazil's data protection law focused on the protection of personal data and digital privacy rights.
18. DHS Common Vulnerabilities and Exposures (CVE)
    - Identifies and catalogs publicly known cybersecurity vulnerabilities.
19. Automated Indicator Sharing (AIS)
    - A DHS program enabling the real-time exchange of cyber threat indicators between the private sector and the government.
20. MIT AI Risk Management Framework (https://airisk.mit.edu/)
    - A framework from MIT focused on AI risk management, providing guidance on identifying and mitigating AI risks in various domains.

ZeroTrusted.ai integrates these compliance frameworks to ensure that AI systems meet the highest standards of security, privacy, and transparency while adhering to critical regulatory requirements across industries and sectors.