TENEIGHT CYBER | run:ai

SECURE AI

Executive Guide to Secure LLM Deployment

–

Achieving strong business results with secure and compliant AI solutions

The Executive's Guide to Secure LLM Deployment

Introduction: The Call to Action

In today's rapidly evolving AI landscape, businesses must leverage large language models (LLMs) to enhance productivity, drive innovation, and secure a competitive edge. Yet, deploying these powerful tools comes with significant risks. As AI models proliferate, so do threats—sensitive data disclosure, data poisoning, model theft, infrastructure vulnerabilities, and unauthorized access are just a few challenges organizations must navigate.

In this executive guide, we will explore how organizations can embark on their AI journey by securing LLM deployments. With the guidance of expert partners at 10-8 Cyber and Run:ai, executive teams can emerge as heroes, driving secure, business-driven applications while safeguarding their infrastructure. This guide will take executives through the journey of confronting AI threats, navigating challenges, and ultimately prevailing with robust governance and security frameworks.

Authored by Gregory Crabb, Founder and Principal Cybersecurity Consultant, 10-8 Cyber

As Founder and Principal Consultant at 10-8 LLC and CISO in Residence at Ballistic Ventures, Gregory Crabb brings over 30 years of elite cybersecurity expertise across government and private sectors. Leveraging his deep understanding of threat actor behaviors and tactics, he delivers secure solutions that enable organizations to defend against evolving threats. Crabb collaborates with groundbreaking AI security innovators like Pangea and Noma, specializing in cyber risk management, threat modeling, secure software development, and workforce development for defensible architectures and secure AI solutions.

The Journey Begins: Understanding LLM Governance

Business Risks of Poor Governance, Security, and Compliance

Poor governance, inadequate security, and non-compliance can have severe business repercussions, including:

- **Regulatory Fines and Legal Liability:** Non-compliance with regulations such as GDPR, HIPAA, and the EU AI Act can lead to significant fines and legal consequences. These compliance requirements are designed to protect data privacy and ensure transparency, and violations could cost organizations millions in penalties.
- **Reputational Damage:** A data breach or compliance failure can severely damage an organization's reputation, eroding customer trust and investor confidence. Negative publicity often results in lost business opportunities and a decline in market value.
- **Operational Disruption:** Poor governance can lead to gaps in AI model reliability and integrity, resulting in operational disruptions. Inaccurate or biased AI outputs can affect decision-making, productivity, and ultimately the organization's bottom line.
- **Intellectual Property Loss:** Inadequate security measures can result in proprietary information being leaked or stolen, allowing competitors to replicate innovations and gain a market advantage.
- **Financial Losses:** Security incidents, regulatory fines, and reputational damage can lead to direct financial losses, decreased revenues, and increased costs due to remediation efforts and incident management.
- **Employee and Customer Trust Erosion:** If sensitive employee or customer data is mishandled or exposed, trust will be severely impacted, potentially resulting in higher turnover rates and customer attrition.

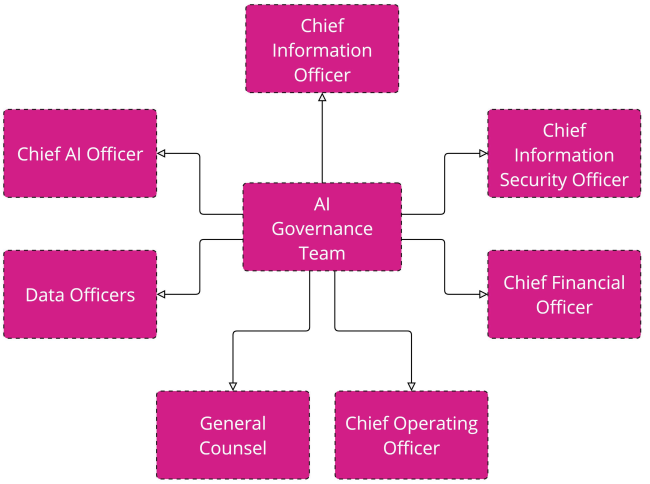
Addressing these risks through proper AI governance, security, and compliance frameworks is essential for minimizing business impact and ensuring that AI deployments contribute positively to business objectives.

Before venturing into the world of LLM deployment, organizations must establish a solid foundation: AI governance. Many companies rush to implement AI without fully understanding the strategic importance of governance. This leaves gaps in compliance, operational integrity, and ethical oversight. As businesses integrate LLMs, the role of governance becomes even more critical.

Why LLM Governance Matters

Without proper governance, LLM deployments can result in unintended consequences, from biased outputs to security breaches. Establishing role clarity for critical stakeholders ensures that AI initiatives are aligned with business goals and sufficiently protected. Key stakeholders in an AI governance framework include:

- **Chief Information Officer (CIO):** Oversees technology strategy and ensures AI initiatives align with overall business objectives.
- **Chief Information Security Officer (CISO):** Manages cybersecurity risks and ensures that AI deployments adhere to security best practices.
- **Chief AI Officer (CAIO):** Drives AI strategy and implementation while ensuring ethical AI practices.
- **Chief Operating Officer (COO):** Focuses on operational efficiency and ensures that AI initiatives contribute to the organization's productivity and business outcomes.



- **Chief Financial Officer (CFO):** Oversees financial investments in AI technologies, ensuring they provide a strong return on investment and align with business growth objectives.
- **Data Officers:** Responsible for data governance, ensuring data quality, privacy, and compliance.
- **General Counsel:** Ensures that AI deployments are compliant with regulatory and legal standards, mitigating liability risks.

Addressing these risks through proper AI governance, security, and compliance frameworks is essential for minimizing business impact and ensuring that AI deployments contribute positively to business objectives.

The Call to Arms: Executive Leadership

Leadership teams must take charge, forming cross-functional AI governance committees to oversee the development and deployment of LLMs. These committees should be tasked with addressing governance challenges, such as ensuring compliance, managing risk, driving business value, and maintaining ethical standards throughout the AI lifecycle. Including key stakeholders such as the CIO, CISO, CAIO, COO, CFO, Data Officers, and General Counsel ensures that all perspectives are considered—from technology and security to legal, ethical, and operational implications. Through successful cross-functional collaboration, executives can establish a governance framework that mitigates risks and aligns AI initiatives with the organization's strategic business objectives.

To set AI governance on the right path, leveraging highly effective collaboration tools is essential. From our work, 10-8 Cyber has identified several critical tools that have proven to maximize organizational outcomes in AI governance:

- **Situational Awareness Map:** This tool helps establish a baseline mindset by providing a clear picture of the current situation, helping teams to identify gaps and develop informed strategies.
- **Goal Hierarchy:** By identifying how individual, team, and multi-team goals fit into the organization's larger objectives, this tool ensures alignment and helps drive business outcomes.
- **Boundary Spanning:** This tool is designed to identify who communicates between teams and under what circumstances, ensuring cross-team collaboration and preventing silos from hindering governance effectiveness.
- **Multi-Team System Mapping:** Identifying when and how teams interact can help streamline collaboration, reduce miscommunications, and ensure that governance processes involve the right people at the right times.
- **Conflict Resolution Protocol:** Ensuring healthy conflict resolution is critical in cross-functional AI governance. This tool helps articulate intentions during disagreements and fosters productive conversations, which is essential for maintaining strong collaborative relationships.

These collaboration tools collectively establish a strong foundation for effective AI governance, driving proactive decision-making, alignment, and robust governance outcomes. With partners like 10-8 Cyber providing a structured governance framework and Run:ai optimizing AI workloads, businesses can build an ethical, secure foundation for LLM success.

Facing the Adversary: Security Challenges in LLM Deployments

The next phase of the journey involves confronting the adversary: the security threats are real and increasing in frequency. As more businesses adopt LLMs, adversaries are quick to exploit vulnerabilities like information leakage, data poisoning, and model theft.

Information Leakage, Data Poisoning, and Model Theft

Many organizations are increasingly concerned about proprietary information being inappropriately injected into an LLM during training, leading to the exposure of sensitive business details. Furthermore, there is a risk of sensitive information inadvertently getting into the hands of employees who lack proper authorization, highlighting the importance of strict access control measures and data segregation protocols. Attackers also embed malicious data during the training phase (data poisoning) to corrupt the model's behavior, resulting in skewed/biased outputs or unauthorized access to sensitive data. Additionally, model theft allows competitors to replicate proprietary AI systems through techniques like reverse engineering.

The OWASP LLM Top 10

To combat these evolving threats, the OWASP LLM Top 10 security framework provides a comprehensive overview of critical risks, including:

- **Prompt Injection Attacks:** Malicious inputs designed to manipulate LLM outputs.
- **Adversarial Inputs:** Crafted inputs that trick models into producing undesirable results.
- **Sensitive Data Disclosure:** The risk of exposing proprietary or sensitive information through inadequate access controls or data leaks.

To address these vulnerabilities, 10-8 Cyber will leverage partnerships to bring composable security services to Run:ai customers, ensuring a robust, secure environment for AI applications.

These services include:

- **Access Control:** Ensuring authentication and authorization are enforced at each point where sensitive information or models are accessed, thus preventing unauthorized access and data leakage.
- **Secure Audit Trails:** Tracking all interactions within the LLM pipeline to provide visibility into access and changes, which is essential for maintaining accountability and compliance.
- **Protecting Sensitive Information:** This service will help ensure that sensitive information such as Personally Identifiable Information (PII) is detected, sanitized, or blocked before entering LLM pipelines, reducing risks of unintended data exposure.
- **Preventing Prompt Injection:** A tool designed to detect and block prompt injections, thereby preventing adversaries from manipulating the model's behavior or bypassing safeguards.
- **Reputation Services:** File, URL, IP, and domain reputation services will be integrated to prevent malicious data or commands from reaching the model. These services will automatically scan and validate inputs, ensuring the integrity of information used by the AI.

By providing these services, 10-8 Cyber will support Run:ai customers in meeting the OWASP LLM Top 10 challenges effectively and efficiently. Security leaders must also implement a range of AI DevSecOps considerations to ensure the integrity and security of AI applications throughout their lifecycle.

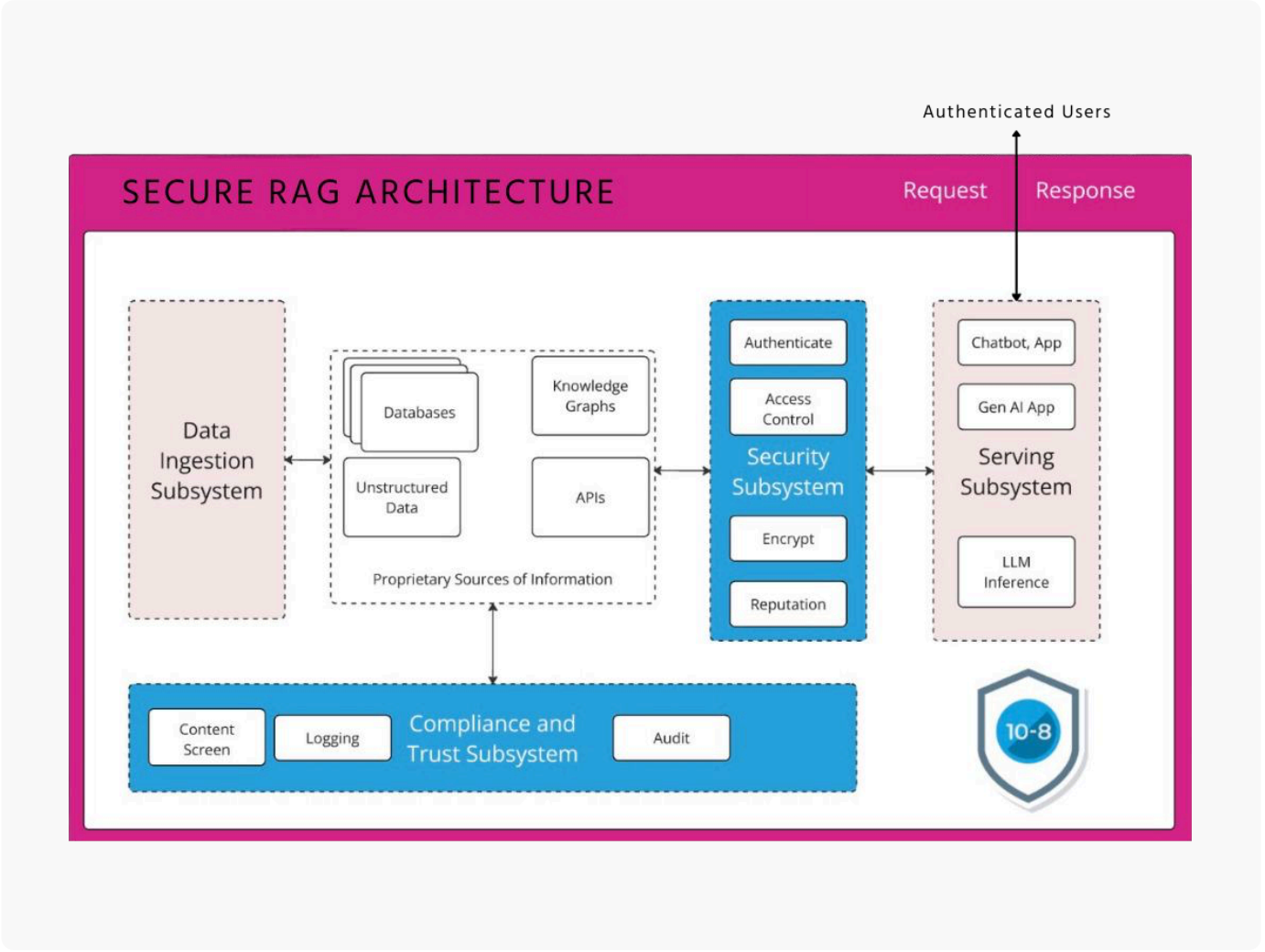
Additionally, it is critical to have effective incident response plans and ongoing monitoring to identify and mitigate risks as they arise. These capabilities collectively help organizations maintain compliance, protect proprietary information, and reduce vulnerabilities in AI deployments.

Secure Reference Architecture for RAG

To further enhance the deployment of LLMs securely, the following is a **reference architecture** for a secure Retrieval-Augmented Generation (RAG) implementation. This architecture is built to integrate stringent security measures while maintaining high-performance retrieval and generation capabilities.

Reference Architecture for Secure RAG Implementation: High-Level Overview

A secure RAG architecture integrates three main components—**ingestion, security and serving layer**—into a streamlined system. The architecture leverages secure communication, authentication, and access control measures to safeguard data throughout the retrieval and generation processes. This architecture provides end-to-end security, from user query handling to response generation and delivery, maintaining regulatory compliance and protecting sensitive information.



Core Components and Flow

The **Secure RAG Architecture** involves a series of interconnected subsystems that work together to securely process user queries, retrieve relevant data, and generate appropriate responses. This architecture ensures that each component not only performs its function efficiently but also adheres to stringent security standards to protect the data being processed.

- The process begins with the **Data Ingestion Subsystem**, where proprietary data sources are integrated into the system. This subsystem processes information from structured databases, unstructured data repositories, knowledge graphs, and external APIs. The ingestion phase includes initial data validation, ensuring that only authorized data sources are incorporated into the RAG pipeline. This data ingestion is the foundation upon which the retrieval processes depend, providing access to well-defined and secure sources of knowledge.

- Once the data has been ingested, the **Security Subsystem** comes into play to secure all interactions within the system. This subsystem is responsible for authenticating users, applying access control, encrypting data at various stages of its lifecycle, and assessing the reputation of both users and data sources. Through multi-layered security measures, the Security Subsystem ensures that data access and handling meet the highest standards of privacy and regulatory compliance. Each interaction is governed by strict protocols to prevent unauthorized access and data leakage, adopting a zero-trust approach where each action must be validated before proceeding.

- The **Serving Subsystem** is the engine that generates responses for the user. This subsystem includes applications like chatbots or generalized AI applications, which work alongside the **LLM Inference** component to deliver relevant, contextually appropriate answers. The Serving Subsystem processes data retrieved from external knowledge repositories and transforms it into actionable responses, considering the user query while adhering to security protocols. Each response is filtered for accuracy and compliance to avoid unintended exposure of sensitive information.

- Throughout the architecture, a critical role is played by the **Compliance and Trust Subsystem**. This subsystem is responsible for monitoring data flows, logging activities, and maintaining audit trails. By continuously logging interactions, the system ensures transparency, accountability, and adherence to regulatory requirements like GDPR and HIPAA. It also screens the content for compliance issues, identifying and mitigating risks related to sensitive information being used or exposed inappropriately.

Secure Management of Proprietary Data Sources

The **secure management of proprietary data sources** is central to the Secure RAG Architecture. Only those data sources that meet the strict security profile of the AI application are authorized for input and retrieval within the system. This involves stringent validation and categorization of data repositories, such as databases, unstructured data sources, knowledge graphs, and external APIs. Before data can be integrated into the RAG pipeline, each source must pass a security assessment to ensure it complies with privacy regulations, data encryption standards, and internal governance policies.

Access to these data sources is carefully managed through role-based access control (RBAC) and robust authentication mechanisms, such as OAuth2 tokens. These security controls prevent unauthorized personnel from interacting with sensitive information and ensure that data input and retrieval within the system are aligned with both organizational requirements and regulatory standards. By adopting a zero-trust model, the system validates each interaction before it is allowed, thereby ensuring that only secure and verified data contributes to the AI application's processing and output. This meticulous management of proprietary data sources not only enhances the overall reliability and trustworthiness of the system but also protects valuable intellectual property and sensitive business information from exposure or misuse.

Crossing the Threshold: Securing AI Infrastructure

Even with proper governance and secure LLMs, AI systems are only as safe as the infrastructure supporting them. This is where many organizations encounter vulnerabilities, particularly in securing their Kubernetes clusters, which are critical for managing and scaling AI workloads.

Kubernetes and RBAC: Protecting AI Workloads

Kubernetes provides flexibility for AI deployments but can expose businesses to significant risks if not properly secured. Misconfigurations, vulnerable container images, and weak network security create openings for attackers. Role-Based Access Control (RBAC) is essential for ensuring that only authorized personnel have access to sensitive AI resources.

Best Practices for Securing AI Infrastructure

- **Encryption and Data Security:** Ensure data is encrypted both in transit and at rest.
- **Pod Security Policies:** Limit the capabilities of containers to reduce the attack surface.
- **Automated Vulnerability Scanning:** Integrate vulnerability scanning into the CI/CD pipeline to catch and patch risks early.

By partnering with Run:ai, organizations can ensure that their AI infrastructure is optimized and secure, leveraging advanced orchestration tools to manage AI workloads without compromising security.

The Hero's Triumph: Building a Secure AI Future

The path to secure LLM deployment is fraught with challenges, from governance gaps to evolving security threats and infrastructure vulnerabilities. Yet, with the right tools and partners, executive teams can emerge as heroes—leading their organizations through the complexities of AI with confidence and security.
10-8 Cyber brings the expertise in AI governance and security needed to navigate the treacherous landscape of LLM deployment, while Run:ai provides the optimized infrastructure and AI orchestration capabilities that make it possible to scale AI workloads safely and efficiently. Together, they empower businesses to deploy AI securely, enabling innovation while protecting key assets and maintaining compliance.

Taking the Next Step

Executives embarking on the journey of AI deployment need a clear strategy, expert partners, and a secure framework to succeed. By prioritizing LLM governance, addressing security challenges, and ensuring infrastructure integrity, you can drive AI innovation while safeguarding your organization.

As you prepare to take the next steps, remember that success in the AI space is not just about deploying models—it's about doing so with security, ethics, and business alignment in mind. For a more detailed consultation on how 10-8 Cyber and Run:ai can help guide your organization on this journey, contact us today.

About 10-8 Cyber

10-8 Cyber, LLC is renowned for its expertise in tackling complex cyber challenges, particularly nation-state cyber aggression. The firm offers a broad array of services, including post-incident cybersecurity roadmaps, cybersecurity assessments, technical validation, incident preparedness, workforce planning, and other strategic advisory services.

About Run:ai

Run:ai is revolutionizing the AI infrastructure landscape with its platform, designed to optimize the efficiency, scalability, and accessibility of AI and machine learning operations. By addressing the challenges of AI infrastructure, Run:ai empowers enterprises to accelerate their AI initiatives and foster innovation.

To learn more about secure LLM deployment, visit:

www.teneightcyber.com
www.run.ai