American Journal of Information Science and Technology
2026, Vol. 10, No. 1, pp. 25–34
https://doi.org/10.11648/j.ajist.20261001.14

Research Article

The "Airlock" and the "Fortress": A Strategic Blueprint for Secure Large Language Model Adoption in National Defence Contexts

Partha Majumdar*

Department of Computer Science, Swiss School of Business and Management (SSBM), Geneva, Switzerland

Abstract

This report outlines a strategic blueprint for the secure adoption of Large Language Models (LLMs) within national defence contexts, addressing the trilemma of needing state-of-the-art AI, prohibiting the exposure of sensitive data, and the prohibitive cost of building a sovereign foundation model. It rejects a monolithic "one-size-fits-all" approach as strategically flawed, proposing instead a Tiered Hybrid AI Architecture that aligns deployment models with existing military data classification hierarchies. This framework is built upon three concurrent solutions. First, the "Secure Enclave" leverages government-grade cloud platforms like Azure OpenAI for Government, enabling the use of powerful proprietary models over private, isolated networks with contractual guarantees that data is never exposed or used for training, making it suitable for confidential and secret information. Second, for top-secret data requiring absolute sovereignty, the "Private Fortress" model involves deploying high-performance, pre-trained open-source models (e.g., Llama 3) on-premise in fully air-gapped environments. This provides maximum security while being significantly more feasible than building a model from scratch. Finally, the "Intelligent Airlock," an application-layer proxy, filters, redacts, and sanitises prompts and responses to prevent data leakage and malicious inputs. It serves as a primary control for low-risk data and as a crucial defence-in-depth component for the other two tiers. By integrating these solutions, this tiered strategy offers a pragmatic, secure, and financially viable roadmap for defence organisations to harness the transformative power of LLMs while upholding the non-negotiable mandate of data secrecy.

Keywords

Tiered Hybrid AI Architecture, Secure Enclave, Private Fortress, Intelligent Airlock, Data Classification

1. Executive Briefing

This report addresses the strategic trilemma facing organisations, such as the Indian Army, that operate with highly sensitive information: the perceived necessity of using state-of-the-art (SOTA) Large Language Models (LLMs) like OpenAI's GPT, the non-negotiable prohibition on sharing sensitive information with any public-facing service, and the prohibitive financial and technological cost of building a "sovereign" foundation model from scratch.

The central finding of this analysis is that the "public versus private" dichotomy is a false choice. A "one-size-fits-all" AI strategy is strategically flawed, operationally restrictive, and financially irresponsible. The optimal solution is a

*Corresponding author: partha.majumdar@hotmail.com (Partha Majumdar)

Received: 18 December 2025; **Accepted:** 30 December 2025; **Published:** 20 January 2026

Tiered Hybrid AI Architecture that precisely aligns secure deployment models with the military's existing data classification hierarchy.

This architecture is built on three distinct, concurrent solution models:

1) *The "Secure Enclave":* This model utilises government-grade cloud platforms, such as Azure OpenAI for Government or AWS GovCloud. It allows the use of SOTA models (including OpenAI's) over completely private, isolated networks, with contractual and architectural guarantees that data is never exposed to the public internet, shared with other customers, or used for model training.
2) *The "Private Fortress":* This model deploys high-performance, pre-trained *open-source* models (e.g., Llama 3) on-premise within fully air-gapped environments. This provides absolute data sovereignty for the most sensitive "Top Secret" data and is magnitudes cheaper and more feasible than attempting to build a foundation model from scratch.
3) *The "Intelligent Airlock":* This is an application-layer proxy that sits within the organisation's network to filter, sanitise, and redact sensitive information from prompts and responses. It serves as a primary security control for low-risk tasks and as a defence-in-depth component for the other two models.

This tiered strategy is technologically and financially viable. It provides a pragmatic and secure roadmap for the Indian Army to leverage the transformative power of LLMs, balancing immediate operational necessity with the non-negotiable mandate of data secrecy.

2. The Strategic Imperative: Data Classification and AI in Military Operations

Military adoption of artificial intelligence is not merely a technological upgrade but a strategic decision shaped by risk, control, and accountability. In defence environments, the value of AI is inseparable from the sensitivity of the information it touches. Strategic advantage depends on aligning AI capabilities with long-established security doctrines rather than disrupting them. Any serious integration effort must therefore be grounded in institutional realities rather than abstract innovation narratives.

2.1. Defining the Asset: The Centrality of Data Classification

Any discussion of AI adoption within a defence organisation must begin not with the technology, but with the data it will process. The foundation of military information security is a rigid, hierarchical data classification system. This system provides the essential framework for tiering AI solutions.

Most government and military classification systems are organised into hierarchical levels of sensitivity, such as Restricted, Confidential, Secret, and Top Secret. [14] These are not just labels; they are legal and procedural mandates that govern data handling, access, and transfer. International agreements and national legislation, for example, outline formal "Measures of Protection" for classified information. [14] These measures stipulate that access shall be granted "only to authorised persons in the course of their official duties" and that such information cannot be transferred to a third party "without the prior written consent of the Party under whose order it has been classified". [14]

This existing, multi-level classification system is the key organising principle for a secure AI strategy. The fact that the military already segregates data by risk proves that a monolithic, "one-size-fits-all" AI solution is strategically incorrect. A single solution would be either (a) too insecure for sensitive data, violating formal data protection rules [14], or (b) too expensive and restrictive for unclassified data, stifling innovation. A Top Secret battlefield operation plan and an unclassified public affairs press release summary cannot and should not transit the same AI system. This reality mandates a hybrid architecture from the outset, where the security posture of the AI solution is matched precisely to the data's classification.

2.2. Defining the Requirement: High-Impact Military Use Cases for LLMs

The push for LLM adoption is not theoretical; urgent, high-impact operational requirements across all domains of warfare drive it. The technology is a "must-have" capability. Key use cases include:

1) *Logistics and Supply Chain Management:* LLMs can analyse historical and real-time data to "identify potential bottlenecks, improve inventory management, and ensure timely delivery of critical supplies." [3] This extends to predictive maintenance, where AI models can forecast mechanical failures in military assets, reducing downtime and enhancing readiness. [3]
2) *Intelligence Analysis:* This is a primary driver. LLMs are required for "secure natural language processing for intelligence analysis" and "document analysis on classified sources." [19] This involves processing and summarising petabytes of raw signal intelligence (SIGINT) and human intelligence (HUMINT) to accelerate the "OODA loop" (Observe, Orient, Decide, Act).
3) *Operational Planning:* The technology can assist commanders and staff in strategy development and complex planning, such as for air operations. [20]
4) *Cyber Defence and Autonomous Systems:* LLMs are crucial components in modern cyber defence, helping to "accelerate threat detection. They are also fundamental to processing the vast data streams from autonomous systems like "drones or unmanned subma-

rines." This also includes the vital task of ensuring robust cyber defence *for* the AI systems themselves. [20]

These use cases span the entire data classification spectrum. For example, a predictive maintenance analysis for a non-critical transport vehicle [3] might use "Restricted" or "Confidential" data. In contrast, the real-time "document analysis on classified sources" [19] or planning for autonomous vehicle operations in a hostile environment [16] would involve "Top Secret" data. Both are valid and necessary LLM applications, but the same system cannot serve them. This reinforces the core conclusion: the specific *use case* dictates the *data classification*, which in turn must dictate the *security architecture*.

3. Deconstructing the Sovereign AI Fallacy: The True Cost of Foundation Models

The pursuit of a fully sovereign, frontier-scale AI model is often framed as a question of national autonomy and strategic independence. In practice, this ambition collides with economic, infrastructural, and organisational realities that are frequently underestimated. Separating symbolic aspiration from operational feasibility is essential when evaluating long-term AI strategy. A rigorous cost-based examination exposes where sovereignty genuinely adds value and where it becomes an unsustainable burden.

3.1. Validating the "Impossible Cost" Premise

The premise that building a "sovereign GPT-4" from scratch is "financially impossible" for most organisations, including a national army, is correct. The costs are staggering and extend far beyond a single training run.

Direct Financial Costs (Training Runs):

1) *GPT-4:* While exact figures are secret, estimates for the training cost of GPT-4 in 2023 were around $63 million [21], with OpenAI's CEO confirming the cost was "more than" $100 million.
2) *Llama 3:* One public estimate for Meta's Llama 3 training exceeded $720 million.
3) *Claude 3.5 Sonnet:* Anthropic's CEO stated its training cost was in the "tens of millions" of dollars.

Systemic Costs (Total Development): The cost of a single training run is just one part of the total development cost. The trend of growing training costs is precipitous, increasing at a rate of 2.4x per year since 2016. [1] At this rate, the largest training runs are projected to cost over $1 billion by 2027. [1] These figures make building a competing, frontier model a financial non-starter for most government entities.

3.2. The Deeper Costs: Hardware and Human Capital

The "technologically impossible" aspect of the query is validated by the deeper costs associated with hardware and human capital.

1) *Hardware (CapEx):* The acquisition of AI accelerator chips (GPUs/TPUs) is the single largest expense, accounting for 47%–67% of the total development cost. [1] This includes not just the GPUs—which can have procurement lead times of 12 weeks or more [18]—but also servers and high-performance cluster-level interconnects. [1]
2) *Personnel (OpEx):* R&D staff costs are the second-largest component, representing a substantial 29%-49% of the total development budget. [1] This is the "hidden anchor" in any sovereign AI project. [8] The technological impossibility lies not in the *science*, which is often public, but in the *logistics* of "hiring for the breadth and depth of required skills." [15] This includes scarce, high-cost talent like MLOps engineers (avg. salary ~$134,000/year), DevOps engineers (~$145,000/year), and specialised data and software engineers. [8] This ongoing personnel cost makes a sovereign model a permanent, billion-dollar-plus strategic drag, not a one-time purchase.

3.3. The Strategic Pivot: "Build" vs. "Deploy"

The analysis in sections 3.1 and 3.2 confirms the premise. Therefore, the strategic goal for the Indian Army must shift from "building" a foundation model from scratch to "deploying" existing, SOTA pre-trained models. This approach leverages the multi-billion-dollar R&D investments of private industry while focusing the Army's resources on the achievable and more critical tasks of *secure deployment, validation, and fine-tuning*.

The following sections analyse the two viable "deploy" pathways: using a third-party SOTA model in a secure enclave and deploying an open-source SOTA model on-premise.

Table 1. *Foundation Model Training Cost Analysis (2024-2025).*

Model	Estimated Training Cost	Total Development Cost Components
GPT-4	$63M - $100M+	Hardware: 47%–67% Personnel: 29%–49%

Model	Estimated Training Cost	Total Development Cost Components
		Energy: 2%–6%
Claude 3.5 Sonnet	"Tens of millions"	(Component breakdown similar to GPT-4)
Llama 3 (Meta)	~$720M+ (public estimate)	(Component breakdown similar to GPT-4)
Frontier Models (2027 Proj.)	$1B+	(Trend of 2.4x growth per year)

4. Solution I: The "Secure Enclave" (PaaS) - Using Public Models Without Public Data Exposure

This solution directly answers the primary question: "How can an organisation like the Indian Army use a public LLM like GPT... and yet maintain secrecy?" The answer lies in the critical distinction between consumer-grade products and contractually-binding enterprise-grade platforms, combined with a secure network architecture.

4.1. Deconstructing the "Public API" Myth

The fear of data exposure is justified when using free, consumer-facing tools like ChatGPT. However, enterprise-grade API platforms operate under entirely different, legally-binding terms that explicitly prevent data leakage and misuse.

1) *OpenAI API Platform:* By default, OpenAI does not train its models on business data (inputs and outputs) submitted via its API. API data submitted after March 1, 2023, is not used for training. Data is retained for a maximum of 30 days for abuse monitoring and then removed. Customers with qualifying use cases can also request Zero Data Retention (ZDR). The platform is SOC 2 Type 2 compliant, encrypts all data at rest (AES-256) and in transit (TLS 1.2+), and will execute a Data Processing Addendum (DPA) for GDPR compliance or a Business Associate Agreement (BAA) for HIPAA compliance.
2) *Industry-Standard Guarantees:* These policies are the industry standard for enterprise customers.
   a) Anthropic (Claude): For commercial API and Enterprise users, data is never used for training. [2] Standard retention is 7 days, and ZDR is available. [2]
   b) Google (Gemini Enterprise): Google's commitment to Gemini Enterprise states, "Google doesn't use your data to train our models without your permission." Your content is "not used for any other customers" and "stays within your organisation."

These *policy* guarantees are the first line of defence. They contractually forbid the exact behaviour that defence organisations fear.

4.2. Architectural Blueprint: The Secure Cloud Enclave

Policy alone is insufficient for national security. The true solution is *architectural*, using government-focused cloud platforms that isolate data at the network level. This is how an organisation can use the OpenAI model *without* sending data over the *public* internet.

1) *Azure OpenAI for Government:* This is the primary solution. The Azure platform allows an OpenAI resource to be created and secured within a private Azure Virtual Network (VNet). By creating a private endpoint for this resource, it receives a private IP address within the VNet. All traffic to the model is then "sent privately instead of over the internet."
2) *The Amplified Guarantee:* The data privacy terms for Azure Direct Models (which includes Azure OpenAI) are even stronger: "Your prompts... are NOT available to other customers... [and] are NOT available to OpenAI". The model is hosted by Microsoft in the secure Azure environment and does *not* interact with OpenAI's public-facing services.
3) *AWS Bedrock for GovCloud (The Precedent):* This model is not just theoretical; it is already approved and in use by the U.S. government for high-stakes data.
4) *DoD & FedRAMP Authorisation:* AWS is the first cloud provider to achieve FedRAMP High and Department of Defence (DoD) Cloud Computing Security Requirements Guide Impact Level 4 and 5 (IL4/5) authorisations for Anthropic's Claude and Meta's Llama models within AWS GovCloud (US). This is a pivotal precedent, proving SOTA models can meet high-bar defence compliance standards.
5) *Data Isolation:* Like Azure, AWS uses AWS PrivateLink to "establish private connectivity from your... (VPC) to Amazon Bedrock, without having to expose your VPC to internet traffic."
6) *Private Model Copies:* When an organisation fine-tunes a model on Bedrock, it is a "private copy." This means "your data is not shared with model providers, and is not used to improve the base models."

4.3. The Compliance Precedent: HIPAA

While not as stringent as "Top Secret," the Health Insurance Portability and Accountability Act (HIPAA) in the U.S. provides a robust, legally-binding model for handling highly sensitive Protected Health Information (PHI). All major providers, including OpenAI, AWS, and their compliant partners [22], will sign Business Associate Agreements (BAAs). This is a legal contract that mandates end-to-end encryption [12], strong access controls, and complete audit logs. [12]

The "Secure Enclave" is the definitive answer to the question. It combines the *policy* guarantees of the enterprise API (no training) with the *architectural* guarantees of a private cloud (VNet/PrivateLink) and the *compliance* precedent of DoD IL4/5 authorisation. This is a solved, productized, and defence-approved solution. It allows the use of OpenAI's SOTA model (via Azure) with *zero* data exposure to the public internet *or even to OpenAI itself*.

Table 2. Comparative Analysis of Enterprise AI Platform Security Guarantees.

Platform	Default Training on User Data?	Data Retention Policy (Default)	Zero-Data Retention (ZDR) Available?	Private Connectivity (VNet/PrivateLink)?	Key Compliance
OpenAI API (Standard)	No (since 3/1/2023)	30 days (abuse monitoring)	Yes (for eligible endpoints)	No	SOC 2, HIPAA (BAA)
Anthropic API (Commercial)	No (never)	30 days (7 days from 9/2025)	Yes	No	SOC 2, HIPAA (BAA)
Google Gemini (Enterprise)	No (w/o permission)	Per-organization policy	Yes	Yes (Vertex AI)	SOC 2, HIPAA (BAA)
Azure OpenAI Service	No (Data is NOT available to OpenAI)	30 days (abuse monitoring)	Yes	Yes (VNet, Private Endpoints)	SOC 2, HIPAA (BAA)
AWS Bedrock GovCloud	No (Data NOT shared with model providers)	Per-organization policy	Yes	Yes (VPC, PrivateLink)	SOC 2, HIPAA (BAA), FedRAMP High, DoD IL4/5

5. Solution II: The "Private Fortress" (On-Premise) - Sovereign Capability with Open-Source Models

This solution addresses the highest-security use cases, where data is classified "Top Secret". It *cannot* leave a physically-controlled environment under any circumstances, not even to a trusted GovCloud.

5.1. The Viable Alternative: Pre-Trained, On-Premise, Open-Source

This model is *not* "building a private LLM." It is deploying a pre-trained, SOTA open-source model on infrastructure that is wholly-owned and controlled by the Indian Army. This approach is financially and technologically feasible.

1) The Architecture: These solutions are explicitly "designed to run in air-gapped environments with no internet access required." The entire system, including vector databases and administrative tools, is "fully self-contained." This ensures "complete data sovereignty," as all prompts and data remain within the organisation's physical perimeter.
2) The Use Case: This architecture is essential for "government and defence organisations" handling "classified information." [19] It offers an "unparalleled level of protection" by "mitigating the risks of unauthorised access, data leakage, and cyber-attacks." [19]

5.2. The Air-Gapped Deployment Playbook & Security Requirements

Deploying an on-premise LLM is a major infrastructure project, not a simple software installation. A typical deployment playbook runs 9–12 weeks and involves [18]:

1) Phase 1 (Weeks 1-4): Infrastructure procurement (noting long GPU lead times) and setup of the air-gapped network environment.
2) Phase 2 (Weeks 5-8): Model weight transfer via secure physical media and setup of the model serving infrastructure.
3) Phase 3 (Weeks 9-12): Security and compliance hardening, including penetration testing and audit logging.

A comprehensive 9-point security plan is required for any such air-gapped deployment [5]:

1) *Identity and Access Control:* Enforce strict Role-Based Access Control (RBAC).
2) *Data Protection (End-to-End):* Encrypt all data at rest (AES-256) using Customer-Managed Keys (CMKs) stored in Hardware Security Modules (HSMs). Encrypt all data in transit (TLS 1.2+).
3) *System Hardening:* Follow CIS/NIST hardening benchmarks and disable all unnecessary ports and services.
4) *Network Security:* Block all outbound traffic by default. There must be no "phone home" traffic or hidden internet dependencies.
5) *Monitoring and Logging:* Implement offline dashboards and ensure logs are consumable by the organisation's local SIEM.
6) *Data Loss and Recovery:* Implement encrypted offline backups and documented offline recovery procedures.
7) *Third-Party Risk Management:* The vendor must provide a complete Software Bill of Materials (SBOM).
8) *Security Validation:* Conduct regular penetration tests before major releases.
9) *Offline Patching:* This is the most critical and difficult challenge of an air-gapped system. The vendor *must* provide "signed offline update bundles" and "secure update workflows that work without Internet connectivity." [5]

The "Private Fortress" is a highly secure and viable solution, but its primary challenge is not the initial deployment, but the long-term "Day 2" operational cost of maintenance. The "Offline Patching" requirement is a complex, physical-security-meets-cyber-security process that is slow and operationally intensive. This must be a central consideration in any on-premise decision.

5.3. Technical Trade-Offs: Selecting the Right On-Premise Model

The Army does not need to build a model; it can choose from a variety of SOTA open-source models, such as Llama, Mixtral, Falcon, or Phi. The primary choice today is often between two leading architectures: Meta's Llama 3 (a dense model) and Mixtral (a Mixture-of-Experts model).

Meta Llama 3 70B (Dense Model):

1) Pros: Superior accuracy and reasoning. It achieves "strong results on tasks requiring multi-step logical reasoning." [17] It demonstrates better "faithfulness and factual grounding," making it "suited for high-stakes use cases" where accuracy is critical, such as intelligence analysis. [17]
2) Cons: Higher hardware cost. As a dense model, it "demands more memory and compute" because all 70 billion parameters are active during inference, requiring "specialised infrastructure. [17]

Mixtral 8x7B (Mixture-of-Experts, or MoE, Model):

1) Pros: Superior speed and efficiency. It is "optimised for speed and resource efficiency" [17] and can achieve inference speeds "six times faster" than older dense models. Its MoE design provides "lower compute consumption" and a "reduction in latency and operational cost" because only a subset of its parameters are used for any given token. [17]
2) Cons: Lower reasoning fidelity. It "doesn't quite match LLaMA's comprehension depth when dealing with multi-document reasoning or layered logic." [17]

The choice is a direct trade-off: Accuracy (Llama 3) vs. Efficiency (Mixtral). For deep, high-stakes intelligence analysis, Llama 3 is the superior choice. For real-time, high-throughput tasks like field-data summarisation, Mixtral may be more appropriate.

Table 3. *On-Premise SOTA Model Comparison (Llama 3 70B vs. Mixtral 8x7B).*

Model	Architecture	Key Strength	Best Use Case	Hardware/ VRAM Requirement	Key Weakness
Llama 3 70B	Dense	Accuracy & Reasoning	High-stakes intelligence analysis, multi-step logical reasoning, RAG pipelines	High (All 70B parameters active)	Slower inference, higher compute cost
Mixtral 8x7B	Mixture-of-Experts (MoE)	Speed & Efficiency	Real-time summarization, high-throughput queries, latency-sensitive tasks	Lower (Only a subset of parameters active)	Lower reasoning fidelity, less adept at multi-document reasoning

6. Solution III: The "Intelligent Airlock" (Guardrail Proxy) - A Hybrid Filtering Approach

This third solution is an application-layer proxy that can be deployed as a standalone security measure for low-risk data or as an essential "defence-in-depth" component for the "Secure Enclave" and "Private Fortress" models.

6.1. Architectural Blueprint: The Application-Layer Proxy

This architecture involves a "guardrail" proxy system [9] that sits *inside* the Army's secure network. Every prompt from a user (input) and every response from the LLM (output) is intercepted and inspected by this proxy *before* it proceeds.

6.2. Core Functions of the Airlock

This system is the primary mitigation for the application-layer risks outlined in the OWASP Top 10 for LLM Applications. [6]

1) Input Filtering (Prompt Protection):
   a) Preventing Prompt Injection (LLM01): The proxy validates all user input to block "adversarial or off-topic user prompts." [9] This is a critical defence against "jailbreaking" attacks, where a user attempts to bypass the model's safety restrictions. [13]
   b) Data Sanitisation: The proxy can be configured to "redact sensitive PII" or, more importantly for a military context, specific keywords (e.g., base locations, operation names, personnel IDs, equipment specifications) *before* the prompt is ever sent to the LLM.
2) Output Filtering (Response Protection):
   a) Preventing Sensitive Information Disclosure (LLM02): The guardrail scans the LLM's response to "prevent the model from fielding requests outside the application's domain boundary." [9] This stops the model from accidentally revealing sensitive data it may have been trained on or has access to. [9]
   b) Preventing Insecure Output Handling (LLM05): The proxy can filter outputs to "eliminate vulnerabilities like XSS and SQL injection in LLM-generated outputs" [6], protecting downstream applications.

The "Intelligent Airlock" is not just a third, competing architecture; it is a *modular security component* that enhances the other two solutions. The "Secure Enclave" (Solution I) provides *network-layer* security (a private VNet). The "Private Fortress" (Solution II) provides *physical-layer* security (an air-gap). Neither of these inherently protects against *application-layer* attacks, such as a malicious or compromised insider using a valid, authenticated prompt to exfiltrate data. [7]

The "Airlock" [9] provides this missing application-layer defence. A truly robust "Secure Enclave" implementation would therefore *also* include an "Airlock" proxy *inside* the VNet, sanitising prompts before they are sent to the private endpoint. This combination of network, physical, and application-layer security creates a best-practice, Zero-Trust architecture.

7. Synthesis: A Tiered Hybrid AI Architecture for National Defence

A defensible AI strategy for national defence must reconcile operational ambition with legal constraint, security doctrine, and economic realism. Effectiveness emerges not from maximising model power everywhere, but from aligning capability, risk, and control with precision. Architectural coherence becomes the decisive factor when multiple AI deployment modes must coexist within a single institution. Strategic integration, rather than isolated optimisation, defines sustainable military AI adoption.

7.1. The "Hub-and-Spoke" Model: A Unified Strategy

The final, synthesised strategy rejects a "one-size-fits-all" approach and integrates all three solutions into a single, cohesive framework. This is the Tiered Hybrid AI Architecture, built on a Zero-Trust, "Hub-and-Spoke" network model.

In this model, a central "AI Trust Layer" or "AI Gateway" acts as the "hub." This hub is responsible for authentication and authorisation. Based on the user's credentials (enforcing RBAC) [5] and the classification of the data they are attempting to access or input (from the classification framework) [14], this hub intelligently and automatically routes the user's query to the correct "spoke"—the appropriate, risk-adjusted LLM backend.

7.2. Mapping Solutions to Data Classification Tiers

This framework allows the Army to match the operational need to the required security posture.

1) Tier 1: UNCLASSIFIED / RESTRICTED Data
   a) Solution: The "Intelligent Airlock" (Solution III) connected to a standard, commercial-grade API (e.g., standard OpenAI or Anthropic API).
   b) Use Cases: Summarising public news reports, drafting public affairs announcements, generating code for non-critical administrative tools, and initial review of non-sensitive recruiting materials.
   c) Security Rationale: The data is low-sensitivity. The "Airlock" [9] provides a sufficient security layer by sanitising any inadvertent PII and logging all queries for a full audit trail, without the cost of a full GovCloud deployment.

2) Tier 2: CONFIDENTIAL / SECRET Data
   a) Solution: The "Secure Enclave" (Solution I).
   b) Use Cases: Logistics planning [3], predictive maintenance schedules [3], non-classified intelligence summaries, supply chain management [4], summarising sensitive but not Top Secret field reports.
   c) Security Rationale: This is the "sweet spot" for this architecture. The data is too sensitive for the public internet but requires the most powerful SOTA models. The Azure VNet / AWS PrivateLink provides full network isolation, and the DoD IL4/5 / FedRAMP High compliance precedent provides the necessary, independently-audited security guarantee. The "Airlock" (Solution III) should be added as a defence-in-depth layer *inside* the VNet.
3) Tier 3: TOP SECRET / Mission-Critical Data
   a) Solution: The "Private Fortress" (Solution II).
   b) Use Cases: Real-time battlefield intelligence analysis, cyber defence kill-chain automation, analysis of *highly* classified signal/human intelligence [19], operational planning for autonomous weapons systems or drones. [16]
   c) Security Rationale: This data, by law and policy [14], *cannot* leave a physically-controlled environment or be seen by any third party (including a cloud vendor). The air-gapped, on-premise model [19] is the *only* acceptable solution. The trade-off of using a slightly less powerful (but still SOTA) open-source model like Llama 3 [17] is more than justified by the 100% guarantee of data sovereignty.

Table 4. *Tiered Hybrid AI Architecture - Data Classification vs. Solution.*

Data Classification Tier	Primary Solution	Key Technologies	Example Use Cases	Key Security Control
UNCLASSIFIED / RESTRICTED	Intelligent Airlock (Solution III)	Application-layer proxy, commercial API	Public news summarization, public affairs drafts, non-critical code generation	Application-layer data sanitization & logging
CONFIDENTIAL / SECRET	Secure Enclave (Solution I)	Azure Private Endpoints, AWS PrivateLink, GovCloud, SOTA Models (GPT-4, Claude)	Logistics planning, supply chain management, predictive maintenance	Network-layer isolation; DoD IL4/5 & FedRAMP compliance
TOP SECRET	Private Fortress (Solution II)	On-premise, air-gapped server, Open-Source Models (Llama 3 70B)	Battlefield intelligence analysis, autonomous systems C2, cyber defence	Physical air-gap; 100% data sovereignty

8. Strategic Recommendations and Implementation Roadmap

Strategic clarity must translate into concrete, sequenced action to avoid fragmentation and risk accumulation. The recommendations that follow are designed to convert architectural intent into operational capability while preserving security, control, and institutional accountability.

1) Recommendation 1: Adopt the Tiered Hybrid AI Architecture. Formally reject a "one-size-fits-all" model. Mandate the "Tiered Hybrid" framework as the central policy for all AI procurement and deployment.
2) Recommendation 2: Prioritise "Secure Enclave" (Tier 2) Deployment. Begin immediate engagement with "GovCloud" providers (e.g., Microsoft Azure, AWS) that have a domestic presence and can provide the necessary private endpoint / private link infrastructure. This Tier 2 solution unlocks the *majority* of high-value, medium-sensitivity use cases (like logistics) [3] *immediately* and with the *best* SOTA models.
3) Recommendation 3: Initiate a "Private Fortress" (Tier 3) Pilot Program. Do not attempt a force-wide rollout. Select a specific, high-impact use case (e.g., an intelligence analysis unit) for a pilot program. The primary focus of this pilot must be solving the "Offline Patching" problem [5], as this is a *process* and *logistics* challenge, not just a technical one. Procure hardware [18] for an on-premise Llama 3 70B deployment, given its superiority in high-stakes reasoning. [17]
4) Recommendation 4: Develop a Central "AI Trust Layer" (Airlock). Invest in building or procuring a central "Intelligent Airlock" (Solution III) proxy. This proxy is a *force multiplier*: it is the standalone solution for Tier 1, the defence-in-depth component for Tier 2, and the internal access control/audit layer for Tier 3.
5) Recommendation 5: Update Personnel Training and Security Protocols. Begin focused training and recruitment for MLOps and DevOps engineers [8] required to manage the on-premise systems. Update all information

security protocols to include the OWASP Top 10 for LLMs [10] and train all personnel on new application-layer risks like prompt injection. [11]

Abbreviations

AI	Artificial Intelligence
API	Application Processing Interface
AWS	Amazon Web Services
B	Billion
BAA	Business Associate Agreement
CapEx	Capital Expenditure
CEO	Chief Executive Officer
CIS	Centre for Internet Security
CMK	Customer Managed Key
DoD	Department of Defence
DPA	Data Processing Addendum
GPT	General Purpose Transformers
GPU	Graphics Processing Unit
HIPAA	Health Insurance Portability and Accountability Act
HSM	Hardware Security Module
HUMINT	Human Intelligence
LLM	Large Language Model
M	Million
MLOps	Machine Learning Operations
MoE	Mixture of Experts
NIST	National Institute of Standards and Technology
OODA	Observe, Orient, Decide, Act
OpEx	Operational Expenditure
OWASP	Open Web Application Security Project
PII	Personal Identification Information
R&D	Research and Development
RBAC	Role-Based Access Control
SBOM	Software Bill of Material
SIGINT	Signal Intelligence
SOC	Security Operations Centre
SOTA	State-of-the-art
SQL	Structured Query Language
TPU	Tensor Processing Unit
VNet	Virtual Network
VPC	Virtual Private Cloud
XSS	Cross Site Scripting
ZDR	Zero Data Retention

Author Contributions

Partha Majumdar is the sole author. The author read and approved the final manuscript.

Conflicts of Interest

The author declares no conflicts of interest.

References

[1] Cottier, B., Rahman, R., Fattorini, L., Maslej, N., Besiroglu, T., & Owen, D. (2025, February 7). *THE RISING COSTS OF TRAINING FRONTIER AI MODELS*. Juma. Retrieved December 18, 2025, from https://arxiv.org/pdf/2405.21015

[2] DATA STUDIOS (n.d.). *Claude: Data retention policies, storage rules, and compliance overview*. Retrieved December 18, 2025, from https://www.datastudios.org/post/claude-data-retention-policies-storage-rules-and-compliance-overview

[3] Department of Defense Management, Naval Postgraduate School (2025, May 5). *Simplifying the Complex: A Conversational Approach to Configuring Military Simulators*. Retrieved December 18, 2025, from https://dair.nps.edu/bitstream/123456789/5411/1/SYM-AM-25-359.pdf

[4] DLA Public Affairs (2025, September 8). *Machine learning has potential to revolutionize agency's planning processes*. Retrieved December 18, 2025, from https://www.dla.mil/About-DLA/News/News-Article-View/Article/4294549/machine-learning-has-potential-to-revolutionize-agencys-planning-processes/

[5] Freitas, T. (2025, September 17). *Securing On-Prem LLM Platforms: Key Requirements for Air-Gapped Deployments*. Medium. Retrieved December 18, 2025, from https://medium.com/@tatielefreitas/securing-on-prem-llm-platforms-key-requirements-for-air-gapped-deployments-8eb9f280448b

[6] GenAI Security Project (n.d.). *LLM01: 2025: Prompt Injection*. Retrieved December 18, 2025, from https://genai.owasp.org/llmrisk/llm01-prompt-injection/

[7] GenAI Security Project (n.d.). *OWASP Top 10 for Large Language Model Applications*. Retrieved December 18, 2025, from https://owasp.org/www-project-top-10-for-large-language-model-applications/

[8] Gubanova, O. (2025, August 1). *LLM Total Cost of Ownership*. Ptolemay. Retrieved December 18, 2025, from https://www.ptolemay.com/post/llm-total-cost-of-ownership

[9] Ip, J. (2025, August 8). *LLM Guardrails for Data Leakage, Prompt Injection, and More*. Confident AI. Retrieved December 18, 2025, from https://www.confident-ai.com/blog/llm-guardrails-the-ultimate-guide-to-safeguard-llm-systems

[10] Johnson, S. K. (n. d.). *Securing Government AI: Why Federal Agencies Need a Trust Layer for Accountable, Compliant Deployment*. Carahsoft. Retrieved December 18, 2025, from https://www.carahsoft.com/blog/nuggets-securing-government-ai-why-federal-agencies-need-a-trust-layer-blog-2025

[11] Klingen, M. (n.d.). *LLM Security & Guardrails*. Github. Retrieved December 18, 2025, from https://langfuse.com/docs/security-and-guardrails

[12] Kuzmych, A., & Teres, K. (2025, June 17). *HIPAA-Compliant LLMs: Guide to Using AI in Healthcare Without Compromising Patient Privacy*. Retrieved December 18, 2025, from https://www.techmagic.co/blog/hipaa-compliant-llms

[13] Lumelsky, A. (2025, January 6). *OWASP Top 10 LLM, Updated 2025: Examples and Mitigation Strategies*. Oligo. Retrieved December 18, 2025, from https://www.oligo.security/academy/owasp-top-10-llm-updated-2025-examples-and-mitigation-strategies

[14] Ministry of Home Affairs, Government of India (2003, August 12). *Agreement between the Republic of India and Ukraine on the Mutual Protection of Classified Documents*. Retrieved December 18, 2025, from https://www.mea.gov.in/Portal/LegalTreatiesDoc/UK03B4291.pdf

[15] Neptune (2025, May 6). *State of Foundation Model Training Report 2025*. Neptune.ai. Retrieved December 18, 2025, from https://neptune.ai/state-of-foundation-model-training-report

[16] Onsu, M. A., Lohan, P., & Kantarci, B. (2024, December 28). *Leveraging Edge Intelligence and LLMs to Advance 6G-Enabled Internet of Automated Defense Vehicles*. Retrieved December 18, 2025, from https://arxiv.org/html/2501.06205v1

[17] Sobolik, T., & George, V. (2025, October 22). *LLM guardrails: Best practices for deploying LLM apps securely*. Datadog. Retrieved December 18, 2025, from https://www.datadoghq.com/blog/llm-guardrails-best-practices/

[18] SOO Group Engineering (2025, June 10). *Sandboxed AI: Deploying LLMs in Air-Gapped Environments*. SOO. Retrieved December 18, 2025, from https://thesoogroup.com/blog/sandboxed-ai-deploying-llms-airgapped

[19] Tzanev, M. E. (n.d.). *Mastering LLM Security: An Air-gapped Solution for High Security Deployments*. Dynamiq. Retrieved December 18, 2025, from https://www.getdynamiq.ai/post/mastering-llm-security-an-air-gapped-solution-for-high-security-deployments

[20] US Air Force (2025, April 8). *AIR FORCE DOCTRINE NOTE 25-1: ARTIFICIAL INTELLIGENCE (AI)*. Retrieved December 18, 2025, from https://www.doctrine.af.mil/Portals/61/documents/AFDN_25-1/AFDN%2025-1%20Artificial%20Intelligence.pdf

[21] Valchanov, I. (2024, July 12). *How Much Did It Cost to Train GPT-4? Let's Break It Down*. Juma. Retrieved December 18, 2025, from https://juma.ai/blog/how-much-did-it-cost-to-train-gpt-4

[22] Vidals, G. (2025, October 3). *Is ChatGPT or Google Gemini HIPAA Compliant? A Complete Guide to HIPAA-Safe LLMs*. HIPAA Vault. Retrieved December 18, 2025, from https://www.hipaavault.com/resources/hipaa-compliant-hosting-insights/hipaa-compliant-llm-chatgpt-gemini/