March 15, 2025

Bringing LLMs to US Government

Glenn Parham | Software & AI Engineer | www.glennparham.com

To: Faisal D'Souza, NCO
Office of Science and Technology Policy
Executive Office of the President
2415 Eisenhower Avenue
Alexandria, VA 22314

From: Glenn Parham
[redacted]

This document is approved for public dissemination. The document contains no business-proprietary or confidential information. Document contents may be reused by the government in developing the AI Action Plan and associated documents without attribution.

INTRODUCTION

My name is Glenn Parham, and I am a software & AI engineer. I'm submitting these recommendations in response to **OSTP's Request for Information (RFI)** on developing a national **AI Action Plan**. My insights draw on my former experience leading **Generative AI** initiatives at the Department of Defense and from the startup world. I am submitting this in my personal capacity.

I believe we can create a more **efficient** government—an explicit priority of this administration—by adopting the latest **large language models (LLMs)**. With thoughtful policy, LLMs can streamline labor-intensive processes, reduce administrative costs, and improve the speed and quality of public services. Yet capitalizing on these benefits requires targeted reforms: improving **talent pipelines**, modernizing **security authorizations**, simplifying **AI procurement**, and clarifying government **data rights**.

The five main policy areas I address are **(1) Talent & Workforce**, **(2) Acquisition & Adoption**, **(3) Authorization & Compliance**, **(4) Infrastructure**, and **(5) Data Ownership & IP**, plus a final **(6) Security** section. Each recommendation follows a consistent format of **Context & Problem**, **Recommendation**, and **Intended Outcome**.

Thank you,

Glenn Parham

www.glennparham.com | [redacted]

[redacted]

1. TALENT & WORKFORCE

1.1 Tour-of-Duty Digital Service Teams for AI

Context & Problem

- The government struggles to recruit and retain top **AI engineers** due to **rigid GS pay caps**, lengthy hiring processes, and limited technology-specific training.
- Past **Digital Services** (e.g., **US Digital Service**, **Defense Digital Service**) had success but were folded into larger orgs, limiting autonomy.
- Digital Service alumni often become **senior technical champions** in the federal government.

Recommendation

- **(Re)-establish Digital Services** in each cabinet-level department, with direct-hire authority at **GS/GG-15 levels**, offering 1–2 year "tours of duty" for AI engineers from industry or academia.
- Provide **domain-specific onboarding** (e.g., security classifications, major regulatory frameworks) to accelerate their impact.
- Align Digital Service directors to report directly to their respective department **Secretary/Deputy Secretary** and to the **US DOGE Service**.
- Fund these short-term billets through dedicated department **Chief AI offices** or supplementary **EOP/OSTP resources**.

Intended Outcome

- A robust internal pipeline of **AI talent** to advise and execute high-priority departmental **AI projects**.
- Fewer failed AI initiatives due to deeper in-house expertise.
- **Agile teams** that can swiftly implement department-level priorities.

2. ACQUISITION & ADOPTION

2.1 Mandate AI Chatbots for the Workforce

Context & Problem

- **LLM-based chatbots** (ChatGPT, Microsoft Copilot, etc.) have proven to enhance productivity in the private sector.
- Most federal employees lack access to these tools due to **acquisition** or **cybersecurity hurdles**.
- Without these capabilities, the government misses out on gains in paperwork processing, coding, data analysis, and general efficiency.

Recommendation

- Direct each department **CIO** (via the **CIO Council**) to purchase enterprise **LLM chatbot licenses** for at least 10% of the workforce by FY2025, expanding to 50% by FY2026.
- **Fast-track software authorizations** for major LLM services (see Section 3.2).
- Encourage an **"opt-in" approach** for initial users, combined with a **feedback loop** to refine the deployment.
- Budget around **$30/user/month** via department **IT modernization** or **GSA agreements**, and issue updated data handling guidance for LLM usage.

Intended Outcome

- Tangible increases in everyday **productivity** across the federal workforce.
- User feedback that shapes iterative improvements to **LLM solutions**.
- A stronger **AI culture** and readiness for more advanced deployments.

2.2 Defining "Frontier AI" for Limited Competition

Context & Problem

- Cutting-edge **LLMs** are provided by a small number of vendors, but **FAR** requires full and open competition, causing delays.
- No codified definition of **"frontier AI"** currently exists.
- Near-peer rivals (e.g., **China**) are rapidly moving forward with advanced **AI deployments**.

Recommendation

- Instruct **NIST** to define **“frontier AI”** through recognized benchmarks and host a public **LLM leaderboard** by the end of FY2025.
- Update **FAR Part 6** to allow limited competition for advanced AI models when only one or two vendors qualify.
- Require the leaderboard to be updated **quarterly** to account for new performance milestones and published models.

Intended Outcome

- Agencies can procure advanced **AI solutions** without excessive delays or protests.
- Maintains **US leadership** in frontier AI for defense and high-priority missions.

2.3 SAM.gov AI Subcategory

Context & Problem

- **AI opportunities** are buried under broad categories on **SAM.gov**, making it difficult for agencies and vendors to pinpoint relevant solicitations.
- This fragmentation limits market visibility and competition, especially for small **AI startups**.

Recommendation

- Add a dedicated **AI subcategory** in SAM.gov. Require agencies to classify all **AI-focused solicitations** under this category.
- Align with updated **PSC codes** or create new ones specific to **generative AI**, **MLOps**, **data labeling**, etc.

Intended Outcome

- Improved visibility of **AI contracting opportunities**, boosting vendor diversity and competition.
- Easier tracking of government-wide **AI spending** and market trends.

3. AUTHORIZATION & COMPLIANCE

3.1 FedRAMP+IL5 Equivalency (Amending DoD Exemptions)

Context & Problem

- **DoD's IL-5 certification** is separate from **FedRAMP High**, requiring duplicate audits.
- Smaller AI vendors cannot afford these costs, slowing entry into **DoD markets**.

Recommendation

- Unify **FedRAMP High** with **IL-5** into a single **"FedRAMP+IL5"** standard.
- Instruct **GSA** and **DoD CIOs** to converge FedRAMP with the **DoD Cloud Computing SRG** by FY2025.
- For instance, amend the SRG to remove the blanket exemption from FedRAMP for unclassified workloads up to **IL-5**.

Intended Outcome

- One **security accreditation** recognized by both civilian agencies and DoD up to IL-5.
- Lower compliance costs, encouraging innovative **AI solutions** in DoD.

3.2 Provisional Authorizations for Large Language Model APIs

Context & Problem

- Agencies and vendors want to use **frontier LLM services**, but the lack of granted **Provisional Authorizations** or ATOs holds them back.
- Running LLMs solely in **government-accredited clouds** is expensive due to compute constraints, dissuading vendors from integrating advanced AI.

Recommendation

- Direct **FedRAMP** and the **DoD CIO** to prioritize frontier LLM provider applications for **provisional authorization** at relevant impact levels.
- Encourage cost-effective **API use** to lower GPU utilization in government-approved environments.

Intended Outcome

- Quicker adoption of frontier **LLMs** and reduced cloud **GPU utilization**.
- More government solutions can leverage high-quality **AI models** without major infrastructure investments.

3.3 Small Business 3PAO Subsidy (Lottery or Grant)

Context & Problem

- **FedRAMP/IL-5 third-party assessments (3PAO)** cost over **$100K**, blocking small AI startups from entering federal markets.
- Innovation suffers when fewer small vendors can compete.

Recommendation

- Launch a pilot via **GSA** or **DoD** to subsidize or fully pay **3PAO audits** for qualifying small businesses.
- Use a lottery or competitive application with an annual pool (e.g., **$10–20M**).

Intended Outcome

- More diverse **AI vendors** achieving FedRAMP/IL-5 compliance.
- Stronger competition, lower costs, and faster innovation in federal AI.

3.4 Template Terraform Modules for Fast-Track ATO

Context & Problem

- Every new AI project redevelops security mappings, diagrams, and code, inflating **time-to-ATO**.
- Results in inconsistent compliance and duplicated effort.

Recommendation

- Instruct **GSA Technology Transformation Services** and the **DoD CIO** to release pre-approved **Terraform modules** that meet **FedRAMP/IL-5 controls**.
- Offer a **fast-track ATO** path for vendors using these reference architectures (e.g., an web app **LLM chatbot** paired with a database).

Intended Outcome

- Standardized and secure deployment, drastically cutting repeated compliance overhead.
- **Time-to-ATO was reduced** from months to weeks for common architectures.

3.5 Guidance for Using Foreign-Based LLMs

Context & Problem

- Some **LLMs trained in adversarial nations** pose supply-chain and security concerns.
- No consistent federal policy on whether to integrate these models for official use.

Recommendation

- Direct **NSA**, **CISA**, and **NIST** to issue guidance on **LLMs trained in ITAR embargoed countries**.
- Require default reliance on **US or allied frontier AI**, restricting foreign-based models to **research** & **sandbox usage** unless a security waiver is granted.

Intended Outcome

- Lower risk of **data leaks** or malicious interference in government AI products and operations.
- Clear rules for agencies evaluating and using **foreign AI**.

4. INFRASTRUCTURE

4.1 Optimizing Secure Facilities for High-Density Compute

Context & Problem

- **SCIF standards** (ICD 705, UFC 4-010-05) lack explicit guidance for high-density compute requirements like **liquid cooling**.
- This restricts advanced **AI/HPC capabilities** in secure environments.

Recommendation

- Direct **ODNI** and **DoD** to revise ICD 705 and UFC 4-010-05 so that secure facilities are conducive to high-density compute.

Intended Outcome

- Greater compute density for sensitive **AI/HPC workloads**.
- Improved energy efficiency and cooling performance in **SCIFs**.

4.2 HPC–Cloud Connectivity

Context & Problem

- Government **HPC clusters** (DoD, DOE, NASA, IC) are siloed, lacking secure and high-bandwidth connections to **government-accredited clouds**.
- This leads to idle GPU time, inefficient resource use, and slower development & integration of large-scale **AI models and operations**.

Recommendation

- Instruct **OMB** (with **OSTP**) to require agencies to establish secure, high-bandwidth interconnects between on-premises clusters and **FedRAMP+IL5 accredited clouds**.
- Have **GSA** and agency **CIOs** incorporate **HPC–cloud connectivity** in relevant RFPs and facility upgrades.
- **Amend HPC acquisition policy** so that new HPC cluster solicitations **explicitly require** connectivity to **government-accredited cloud**.

- Direct **NIST**, **DoD CIO**, and **DOE HPC programs** to set interoperability standards and publish best-practice architectures.

Intended Outcome

- Flexible, on-demand flex capacity to commercial cloud when **HPC workloads spike**.
- Reduced idle hardware and better efficiency.
- Faster AI development via hybrid **HPC–cloud setups**.

4.3 Buy LLM Compute in Tokens

Context & Problem

- Agencies often lock themselves into **large blocks of GPU/HPC hours** with a single provider, risking vendor lock-in and paying for unused capacity.
- Meanwhile, **private-sector LLM providers** increasingly price usage by LLM **"tokens,"** offering a more precise consumption-based model.

Recommendation

- Update or clarify **FAR Part 16** (via **OMB**, **FAR Council**) to explicitly accommodate **usage-based or token-based** contracting for AI/LLM compute, aligning with private-sector LLM pricing units.
- Require vendors to publish clear **cost-per-token** metrics, ensuring transparency and competition.

Intended Outcome

- **Competitive pricing** and scalability, with agencies paying only for **actual token usage** rather than reserved blocks of compute.
- Greater incentive for providers to innovate on **price**, **capacity**, and **feature offerings** in pursuit of government workloads.

4.4 Incentivize Cloud Providers to Invest in Compute Capacity

Context & Problem

- Top-tier **GPU** and specialized accelerators are often prioritized for lucrative **commercial cloud** regions rather than government-accredited cloud, creating capacity shortages for federal missions.

Recommendation

- Task **OMB** (with the **Federal CIO Council**) and **GSA** (in coordination with the **DoD CIO**) to:
    - Offer cloud service providers **guaranteed minimum-usage contracts**.
    - Provide **subsidies or cost-sharing** to offset opportunity costs.
    - Propose **tax incentives** for capital investments in government clouds.
    - Grant **preferred vendor status** to providers who commit significant advanced compute resources.

Intended Outcome

- Enhanced availability of **leading-edge compute** in secure government cloud regions.
- A more competitive **HPC marketplace** that meets federal demand at required impact levels.

4.5 Incentivize Diverse Compute Types

Context & Problem

- Government **HPC acquisitions** typically focus on **GPUs**, risking overreliance on a single accelerator type.
- Emerging AI architectures (like large language models) may benefit from **TPUs**, **LPUs**, or specialized chips.

Recommendation

- Direct **GSA** and **OMB** to update **HPC and cloud procurement guidance** to encourage multiple accelerator types.
- Instruct **NIST** to develop benchmarking standards for non-GPU architectures, ensuring consistent performance and security evaluations.

Intended Outcome

- Greater flexibility and efficiency for **AI workloads**, reducing hardware lock-in.
- More rapid innovation cycles, allowing agencies to leverage the best-fit compute for their mission

5. DATA OWNERSHIP & IP

Context & Problem

- FAR Part 27 and DFARS 227.71 do not specifically address training data or **AI model weights**, risking lock-in and re-purchasing the same solutions.

Recommendation

- Amend **FAR Part 27 / DFARS 227.71** to ensure the government obtains a broad license or ownership of model weights and training data developed with federal funds.
- Provide a standard **data rights clause** for AI, clarifying usage, derivatives, and cross-agency reuse.

Intended Outcome

- Greater government reusability of **AI models** without duplicative spend.
- Clear vendor expectations, promoting more competitive pricing and open innovation.

6. R&D

6.1 LLM Homomorphic Encryption (FHE) R&D Subsidies

Context & Problem

- Federal use cases (healthcare, intelligence, etc.) involve highly sensitive data, requiring better data protection during **Large Language Model inference**.
- **FHE** allows data to remain encrypted during inference/training, but remains costly and underdeveloped for large language models.

Recommendation

- Direct **NSF** to fund **R&D grants** for FHE research for Large Language Models.
- Instruct **NIST** and **DARPA** to develop benchmarks and testbeds to measure performance and interoperability.

Intended Outcome

- Stronger security via persistent **encryption** during LLM inference. Eventually, enabling end-to-end encryption of LLM workflows.
- Broader LLM use cases for highly sensitive workloads.
- Accelerated innovation as federal funding boosts commercial and academic interest in **FHE**

CONCLUSION

In summary**,** the strategic adoption of large language models (LLMs) across the federal government can yield transformative benefits—from **streamlined operations** and **reduced administrative overhead** to **improved public services**. Realizing these gains requires a **multifaceted approach**: cultivating the **right AI talent** within the public sector, updating **procurement processes** for AI-specific solutions, consolidating **security and authorization standards**, modernizing **infrastructure for high-density compute**, and establishing **clear data ownership policies**. Equally important is **fortifying security practices**, particularly in areas like **encryption** and the **use of foreign-based LLMs**.

By acting on the recommendations outlined in each section—**Talent & Workforce**, **Acquisition & Adoption**, **Authorization & Compliance**, **Infrastructure**, **Data Ownership & IP**, and **Security**—the federal government can **lead by example** in harnessing cutting-edge AI capabilities. These policies will not only **advance national priorities** of efficiency and innovation but also **safeguard the public trust** through robust security and **equitable data governance**.

I appreciate the opportunity to share these insights and stand ready to assist in developing and implementing the **AI Action Plan**. Together, we can ensure that the United States remains **at the forefront** of safe and impactful AI deployment in service of its citizens.

Glenn Parham
www.glennparham.com | [redacted]