PROTOPIA + Meta

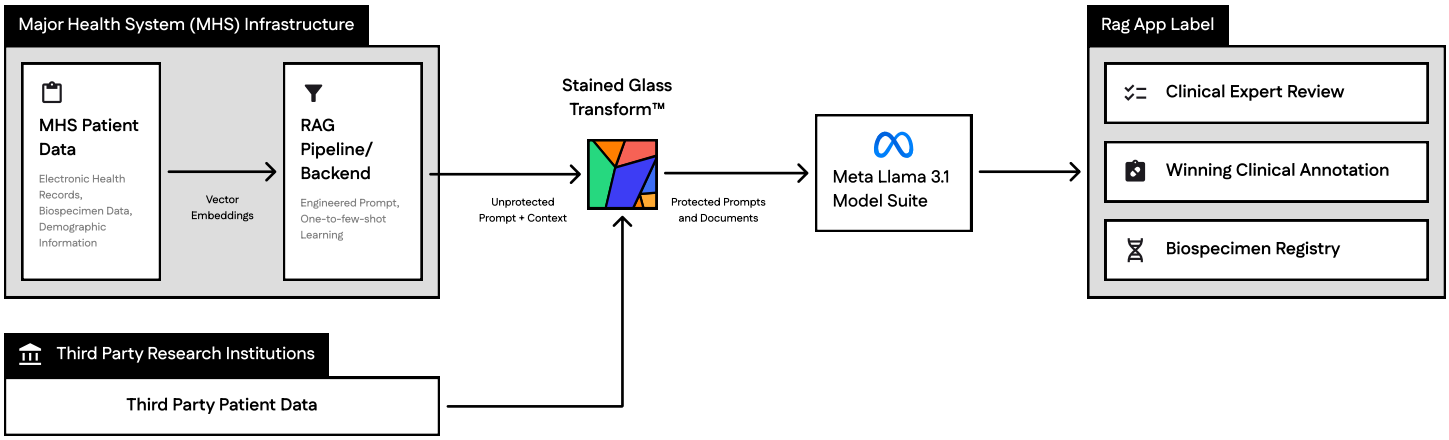
Accelerating AI-Driven Medical Research While Preserving Data Privacy

Addressing Data Privacy Barriers For AI-Powered Oncology Research

Medical research must process sensitive patient records without exposing confidential data. Traditional security and data privacy alternatives like homomorphic encryption and federated learning introduced significant computational overhead, high costs, and performance degradation. These blockers stalled AI medical use cases and prevented third-party collaboration across research teams.

Secure RAG Pipelines Powered By Protopia SGT & Meta Llama 3.1B Model Suite

Protopia partnered with a leading Major Health System (MHS) and Meta to securely implement a RAG application for critical oncology research and AI-assisted clinical annotation. By deploying Protopia AI's Stained Glass Transform (SGT) the MHS overcame strict privacy PHI data privacy constraints, enabling secure AI-powered collaboration with third-party researchers.



Scaling AI Research While Protecting Private Medical Data

By using Protopia SGT and Meta Llama models for data protection and third-party enrichment, the MHS productionized a new RAG application that could be used across multiple research teams while ensuring patient data remained protected. Meta's open-weight Llama models enabled CPT (Continued Pre-Training) for domain-specific adaptation without black-box risks, while Protopia's SGT ensured secure, privacy-preserving data transformation, allowing MHR to share research data without compromising confidentiality.

Increased Data Utilization for Medical AI Use Cases

Overcame strict privacy constraints, enabling AI-powered research on sensitive medical data without exposure risks.

Bioresearch Cost Savings via AI-Powered RAG

By automating clinical annotations, the MHR reduced reliance on manual workloads, saving $200K per 1K patient cases.

Secure Collaboration Across Third-Party Teams

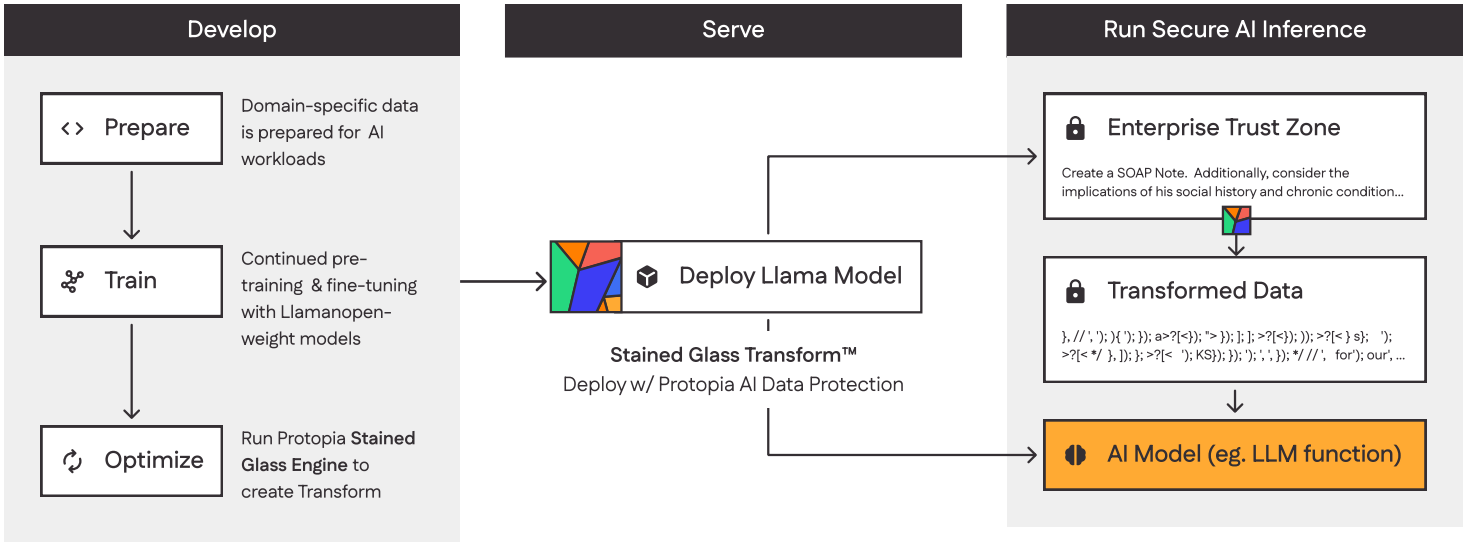
Facilitated secure data sharing and AI-powered collaboration with research institutions while ensuring patient confidentiality.

Expanded Research & Medical Services

SGTs add minimal latency (typically ms) and can even run on CPUs or embedded devices, including sensors.

How Protopia Secures AI Workloads Using Open-Weight Llama Models

Protopia AI makes Meta's Llama models enterprise-ready by solving the biggest barrier to adoption: data security.



Maintain High Performance While Eliminating Plain-Text Data Exposure

Model	Average Tokens Transformed	Sentence Completion (HellaSwag)	Language Understanding (MMLU)	Model Truthfulness	Abstraction Reasoning (ARC)
Llama 3.1 70B W/ SGT	98.44%	77.97%	77.88%	62.33%	51.02%
Llama 3.1 70B without SGT	0% (ie., plain-text)	77.61%	80.52%	66.9%	50.94%

Secure AI Inference Without Performance Trade-Offs

<1%
Added to inference time

~25 milliseconds
Latency for Llama 70B SGT for ~200 token

Data Protection Without Disruption
Model weights, accuracy, and AI workload performance rates remain intact

Seamless Integration: Secure AI with Standard ML Tooling

No Deviation From Existing AI/ML Deployment Pipelines:

- No change to the weights of trained AI model being targeted.
- Stained Glass Engine introduces minimal latency compared to target model training.
- Stained Glass Transform introduces next to no latency in inference pipeline.

No Dependency On Non-Proven Technology (HW or SW):

Built on widely used standard SW packages:

- Python 3.9, 3.10
- Pytorch >= 2.1.0
- Hugging Face Transformers
- Hugging Face Tokenizers

**Hugging Face Libraries are not required for non LLM use cases*

Data Protection Across Inference Inputs, Outputs, and Embeddings

Protopia SGT ensures that sensitive information never exists in plain-text outside of the enterprise zone of trust, securing data through every layer of the AI pipeline—from user inputs to model outputs. By protecting data used in AI, Protopia is ideal to help maximize the ROI on AI infrastructure.