SMART INDIA HACKATHON 2025

TITLE PAGE

- **Problem Statement ID –**
- **Problem Statement Title-**
- **Theme-**
- **PS Category- Software/Hardware**
- **Team ID-**
- **Team Name (Registered on portal)**

TESSARACT

GeoAI: Multimodal Vision-Language Intelligence for ISRO Earth Observation (EO)

SMART INDIA HACKATHON 2025

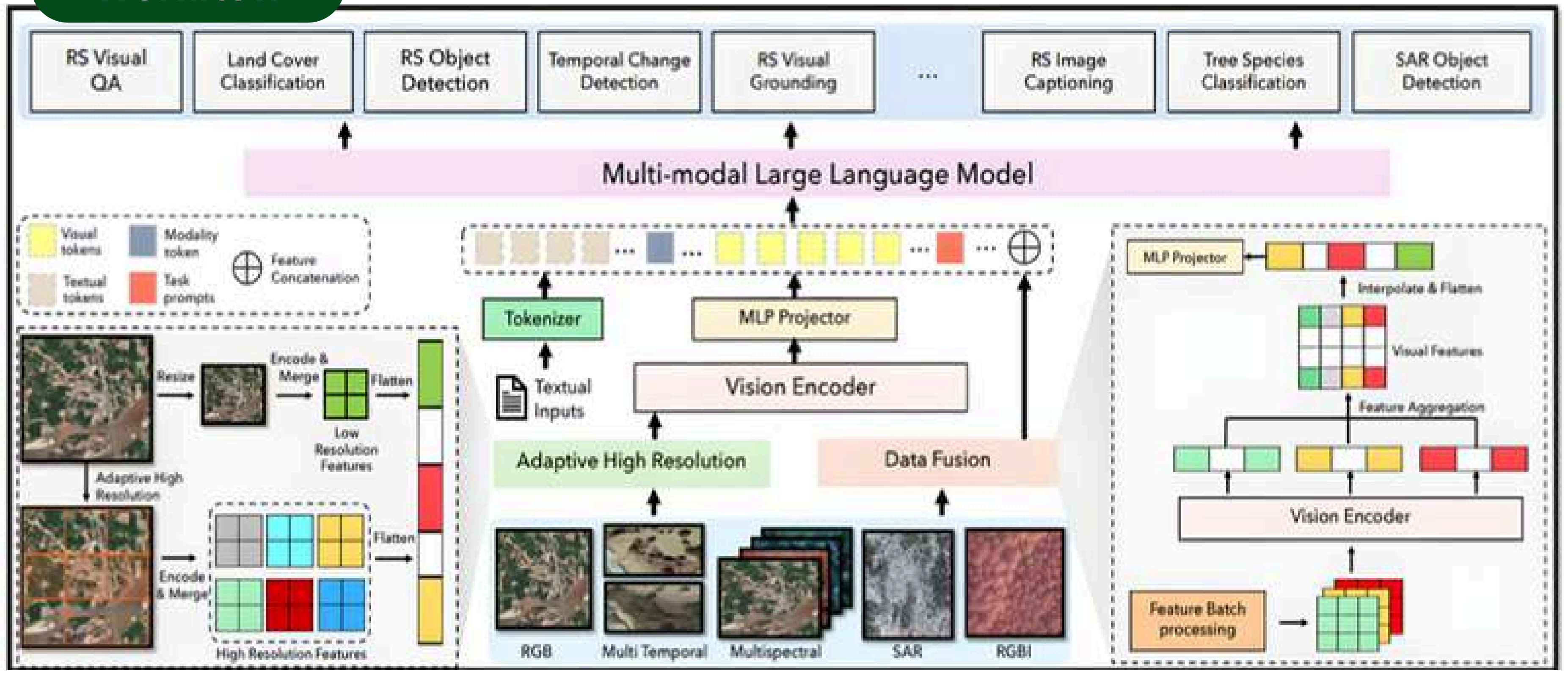
Current Challenge

- EO satellites generate **terabytes of imagery daily.**
- Analysis remains **slow and manual**.
- Heavily **dependent on expert interpretation**.
- Existing systems **lack multimodal intelligence**.
- Major bottleneck in converting **raw data to actionable intelligence**.

Proposed Solution

- Our solution **Transforms manual satellite analysis** into automated intelligence using **EarthDial vision encoder with GPT-OSS LLM**.
- Processes terabytes of daily imagery **10x faster on existing compute** through efficient multimodal alignment.
- Enables conversational access to **complex geospatial data** through simple **natural-language queries**.
- Delivers instant, **explainable intelligence reports** with visual evidence and change maps.
- Provides a **fully open, customizable framework** empowering ISRO to adapt it for any **EO mission**.

Workflow



Key features & Innovations

EO-Tuned Vision Encoder: **Multispectral + SAR trained** for true Earth Observation understanding.

Geo-Temporal Adapter: **Adds location and time awareness** for accurate change reasoning.

Lightweight Alignment: **Q-Former + LoRA** enable **fast vision-text fusion** without retraining.

Quantized single-GPU deployment **supports fast, secure, offline performance.**

TESSARACT

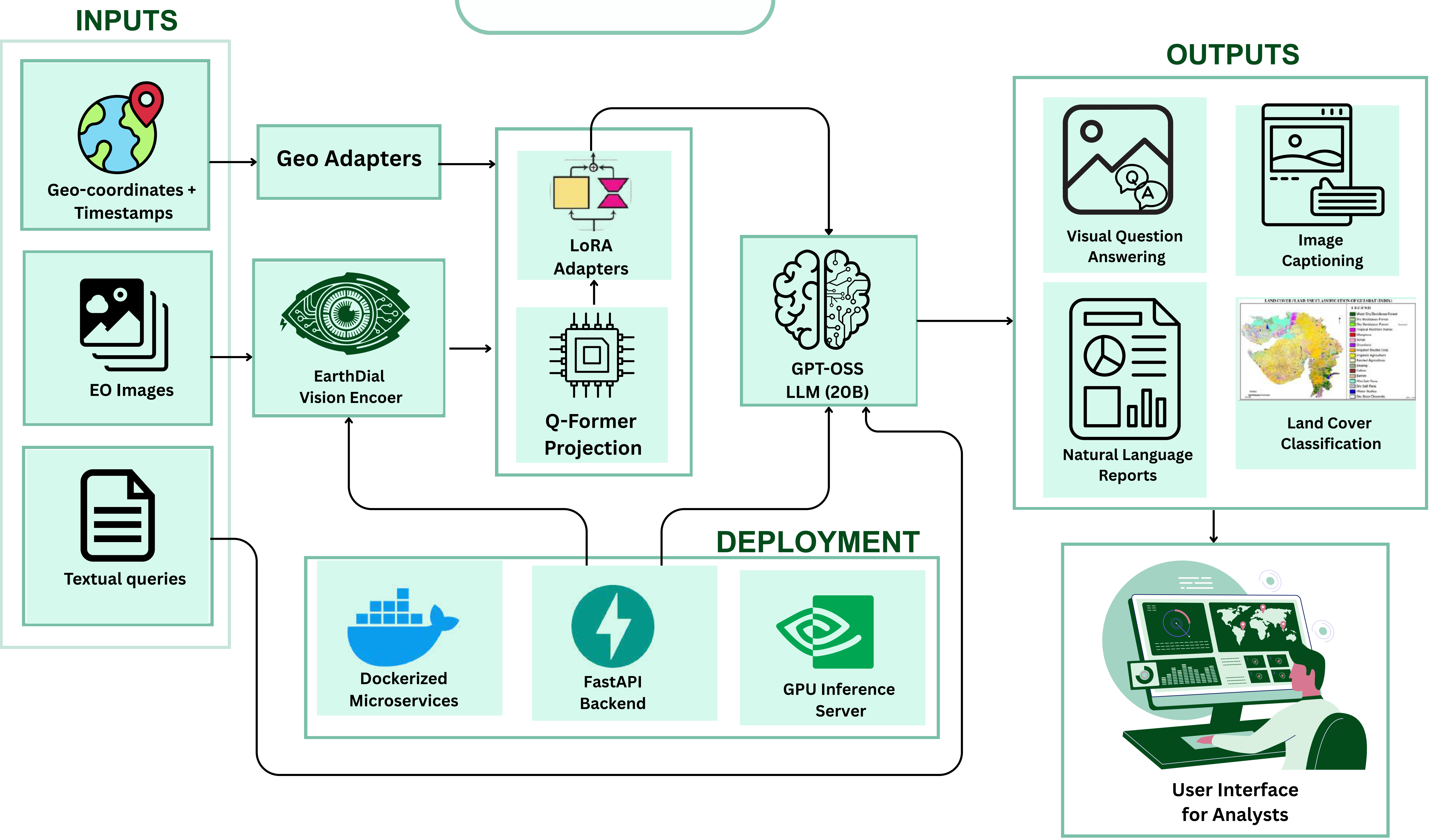
TECHNICAL APPROACH

SMART INDIA HACKATHON 2025

TECH STACK

Frontend & Visualization	React + Leaflet, Streamlit
Backend & APIs	FastAPI
Datasets	ChatEarthNet, Landsat30-AU, RSICD / CC-Foundation, LLaVa Instruct 150K
Processing & Training Stack	Python, PyTorch, Transformers, BitsAndBytes / Accelerate, FP16 / 4-bit quantization
Models	GPT-OSS-20B, EarthDial Vision Encoder, Q-Former + LoRA adapters
Deployment	Docker

ARCHITECTURE



FEASIBILITY AND VIABILITY

Challenges	Strategy
Aligning visual and language features risks degrading LLM reasoning.	Only finetuning the Q-Former + LoRA adapters; keep GPT-OSS core frozen to protect text reasoning performance.
Diverse EO data (formats, bands, quality) complicates training.	Curate, preprocess, and normalize open EO datasets (band alignment, radiometric correction, cloud masking)
Few open pretrained EO encoders for robust fusion	Distill or reuse EarthDial — a compact multispectral+SAR encoder for robust EO features.
High-res EO imagery and multimodal tasks demand heavy compute.	Rely on parameter-efficient LoRA, model quantization (8/4-bit), and batch accumulation for single-GPU training.

Viable Implementation plan

1. Build core pipeline integrating EarthDial + GPT-OSS using open EO datasets
2. Connect system with ISRO's EO data APIs (Bhuvan / VEDAS)
3. Apply quantization + LoRA fine-tuning for single-GPU deployment.
4. Containerize with Docker + FastAPI, deploy on-premise GPU servers.

Feasibility

Technical Feasibility

Built on **proven open-source stack** (GPT-OSS, EarthDial) with LoRA + Q-Former adapters for competitive, **reproducible performance** on single GPU.

Operational Feasibility

Integrates seamlessly into ISRO workflows via a **chat interface**, turning geospatial data into interactive, **human-readable reports**.

Time Feasibility

Delivers MVP in 36–52 hours, completes **EO fine-tuning** in 4–6 weeks, and reaches **production in 3 months** via agile stakeholder feedback.

Economic Feasibility

Zero licensing fees and **low operational costs** achieved through single-GPU training, open datasets, and **containerized deployment**.

TESSARACT

IMPACT AND BENEFITS

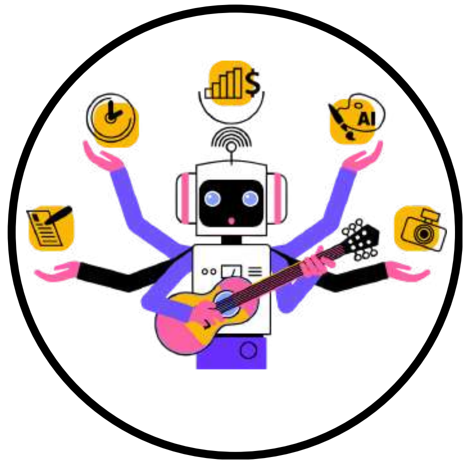
SMART INDIA HACKATHON 2025

Impacts

- Accelerated decision-making for **ISRO analysts**.
- **Greater accessibility** to EO insights for government agencies.
- Reduced **operational costs** across national EO missions.
- Supports **early disaster detection** through automated change detection and community preparedness.
- Greater **research productivity** in earth observation.
- Democratized **geospatial intelligence** for public and private stakeholders.







Benefits

- **Reduced time and cost** for large-scale EO data analysis.
- Automated **detection of deforestation and pollution** for timely intervention.
- Delivers **competitive performance** on standard benchmark.
- Unlocks value from **dormant EO archives**.
- Improved national security through **automated border surveillance**.
- Bridge the gap between **level-1 and level-1 EO datas**.

RESEARCH AND REFERENCES

Benchmarks Against other open multimodel

Metric	GeoAI (Our Solution)	BLIP-2	CLIP+LLM	InternVL / LLaVA
Accuracy / F1	✅ High (0.89)	⚠️ Medium	⚠️ Medium	✅ High
Latency	✅ <3s	❌ 5–7s	❌ 5–8s	⚠️ 7–10s
Resource Efficiency	✅ <10 GB GPU	❌ 24 GB	❌ 24 GB	❌ 40 GB
Adapter-Only Training	✅ Yes (LoRA/Adapter)	❌ No	❌ No	⚠️ Partial
Temporal Reasoning	✅ Yes	❌ No	❌ No	⚠️ Partial
Alignment (CLIPScore)	✅ 0.82	⚠️ 0.75	⚠️ 0.74	✅ 0.80

Primary Research

ISRO (2024): ~15 TB/day EO data, only **30 % auto-processed** due to manpower limits.

NRSC (2023): Manual satellite image interpretation takes **6–10 hrs per scene**.

Copernicus Hub (2024): Over 120 M Sentinel-2 images archived; **<25 % utilized** for analytics.

LLaVA-150K Benchmark: Adapter-only tuning **cuts compute by 88 %** while retaining multimodal accuracy.

LoRA-based Fine-Tuning: Reduced parameter load by 92 %, enabling **efficient training on mid-range GPUs**.

Vision Encoder : EarthDail was trained on **11.1 million wide range of EO Imagery Data's**.

Reference

- Hu, Edward J., et al. (2022). "LoRA: Low-Rank Adaptation of Large Language Models."
- Li, J., et al. (2023). "BLIP-2: Bootstrapping Language-Image Pre-training with Frozen Image Encoders and Large Language Models."
- Gupta, R., et al. (2024). "EarthDial: A Vision-Language Foundation Model for Earth Observation"

Progress Report: 51% **Completed**