Enabling Agentic capabilities in Earth Observation using EVE – applications in the EO Dashboard and drought monitoring

[1]**Riccardo D'Ercole**[*], [1]Eva Gmelich Meijling[*], [3]Kaushik Pillalamarri[*], [2]Anca Anghelea, [1]Chiara Maria Cocchiara, [1]Nicolas Longépé

[1]**European Space Agency Φ-lab, Frascati, Italy;** [2]**European Space Agency, Frascati, Italy;** [3]**North Carolina State University, North Carolina, United States**

This work investigates how agentic AI can build a connected EO ecosystem, using **EVE (Earth Virtual Expert)**, a **domain-specialized LLM** funded by **ESA's Φ-lab**, to autonomously interact with EO tools, APIs, and databases via the Model Context Protocol, enabling dynamic reasoning and cross-source information retrieval.

From Q&A to Agentic EO

LLMs are moving **beyond isolated Q&A toward autonomous agents** that plan, reason, and chain tools. In EO, scientific analysis of satellite data demands **multi-step reasoning grounded in domain knowledge**; retrieving imagery, computing indices, and interpreting results in scientific context.

General-purpose LLMs struggle here, suffering from geospatial naivety and weak handling of domain-specific terminology. **EVE** [1] is a 24B-parameter LLM built on Mistral Small 3.2, specialized for Earth Observation and Earth Sciences. We use EVE as the **reasoning core** of a multi-agent pipeline, where it orchestrates EO tools through the Model Context Protocol (MCP).

MCP Servers

MCP servers act as a standardised "USB port" for connecting LLM agents to real tools, APIs, and data sources.

EO Dashboard

Searches narratives, themes and indicators from the joint ESA / NASA / JAXA EO Dashboard and resolves each result to its underlying STAC catalogue items for direct data access.

Google Earth Engine [2]

Downloads satellite imagery (Sentinel-2 and other GEE collections) as GeoTIFFs for a given AOI, data range, and band selection, with cloud-cover filtering and automatic tiling.

arXiv [3]

Searches, downloads and reads full arXiv papers, allowing agents to retrieve and reason over recent scientific publications.

GIS [4]

Provides 90+ spatial tools spanning geometry operations, vector/raster processing, spatial statistics, map visualization and geospatial data downloads.

EVE's Knowledge Base

RAG*

Contains curated domain-specific sources (open-access, proprietary, ESA documents, and private collections) totalling ~365k documents, supporting hybrid and metadata retrieval.

ESA Science Strategy

Graph-RAG*

Combines dense vector retrieval with a knowledge graph of ESA's EO Science Strategy (the roadmap for ESA's Earth Observation priorities) recovering both passages and structured relations.

* RAG stands for Retrieval-Augmented Generation

Benchmarking Against Other Models

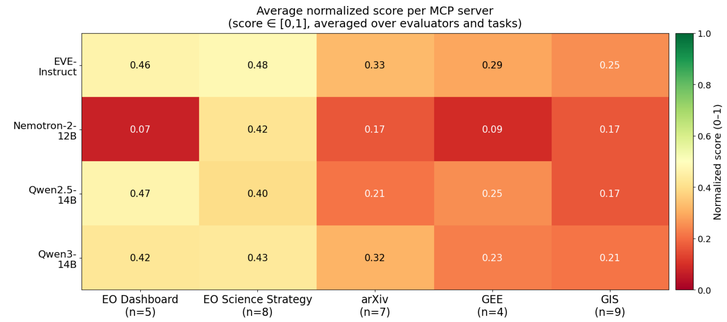
Building on **GeoBenchX** [5], we introduce a benchmark of **103 executable tasks** that evaluates the full tool-call trajectory of the multi-agent system across five of the framework's MCP servers.

Category	# Tasks	What is tested
Tool Grounding	37	Single, correctly parametrised tool call
Sequential Reasoning	18	Multi-step chain within one server
Cross-Agent Synthesis	26	Propagating intermediate results across servers
Scope Awareness	22	Refusing out-of-domain queries (no tool call)

Preliminary Findings

Each generated trace is scored 0/1/2 against reference solutions by **an LLM-as-judge ensemble** (Gemma3-12B, Nemotron-2-12B, Qwen3-14B); scores are normalised to [0, 1] and averaged over judges and tasks.

Retrieval-oriented servers (EO Dashboard, EO Science Strategy, arXiv) are the most reliably handled across models, with **EVE-Instruct leading** on the EO Science Strategy and arXiv. Action- and computation-oriented servers (GEE, GIS) remain the hardest.

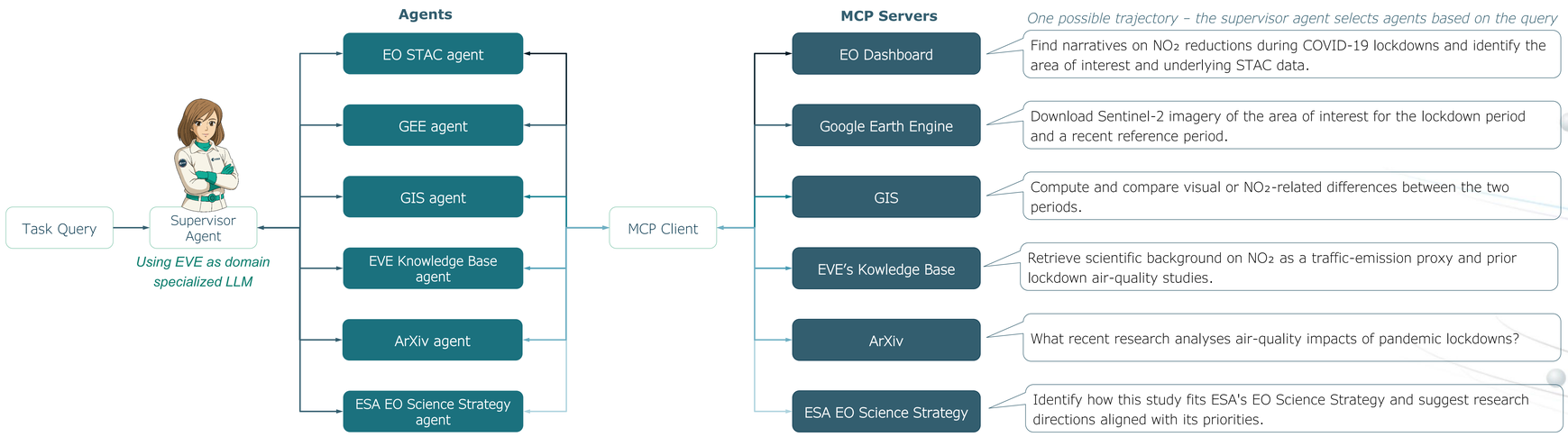


Multi-Agent Framework

We adopt a **hierarchical supervisor / sub-agent design** implemented using LangChain and LangGraph. A central supervisor receives the user query, maintains conversational history for **multi-turn interaction**, and dynamically routes work to **specialized sub-agents**, each pairing EVE with its own MCP-backed toolkit.

Sample user story

"Building on the EO Dashboard story on NO_2 reductions during COVID-19 lockdowns, extend the analysis to a recent period, ground it in the literature, and suggest research directions aligned with ESA's EO Science Strategy."



So, What's Next?

- **Expand the benchmark** – more tasks, more models, broader EO tool coverage
- **Grow the MCP ecosystem** – add servers and experiment with agent-tool distribution
- **Mixed-LLM configurations** – explore smaller specialized models per sub-agent to cut compute, with EVE reserved for reasoning-heavy steps
- **Smarter multi-turn context** – move beyond naïve history concatenation (summarization, selective retrieval, memory)
- **Spatial analytics & RS tools** – integrate Geospatial Foundation Models for learned multi-modal tasks such as segmentation and change detection
- **Agents for drought monitoring & insurance**– integrate MCP server to compute insurances based on satellite imagery

Satellite imagery → Vegetation Index (NDVI) → Anomaly (VCI) → Threshold trigger → Insurance payout

[1] Atrio, À. R., Lopez, A., Rohit, J., Ouahidi, Y. E., Politi, M., Iyer, V., Jamil, U., Bratières, S., and Longépé, N.: EVE: A Domain-Specific LLM Framework for Earth Intelligence, arXiv preprint arXiv:2604.13071, 2026.
[2] https://github.com/kuzhang/mcp-gee-satellite-download
[3] https://github.com/blazickjp/arxiv-mcp-server
[4] https://github.com/mahdin75/gis-mcp
[5] Krechetova, V. and Kochedykov, D.: GeoBenchX: Benchmarking LLMs in agent solving multistep geospatial tasks, in: Proceedings of the 1st ACM SIGSPATIAL International Workshop on Generative and Agentic AI for Multi-Modality Space-Time Intelligence, pp. 27–35, 2025.