Agentic Frameworks for Space Data Management: Semantics, Memory, and Scaling

Gregory Tsagkatakis

Assist. Professor, Computer Science Department, UoC
Affiliated Researcher, Institute of Computer Science, FORTH

Signal Processing Lab (SPL) Team



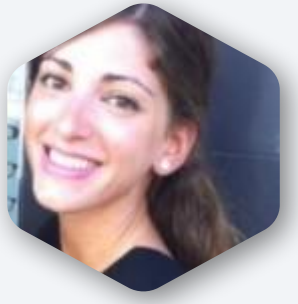
P. Tsakalides
SPL HEAD
PROFESSOR



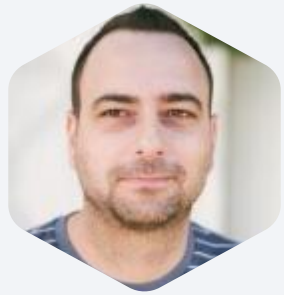
G. Tzagkarakis
PRINCIPAL
RESEARCHER



G. Tsagkatakis
ASSISTANT
PROFESSOR



E. Doutsi
AFFL.
RESEARCHER



N. Stefanakis
UNIVERSITY
PROFESSOR

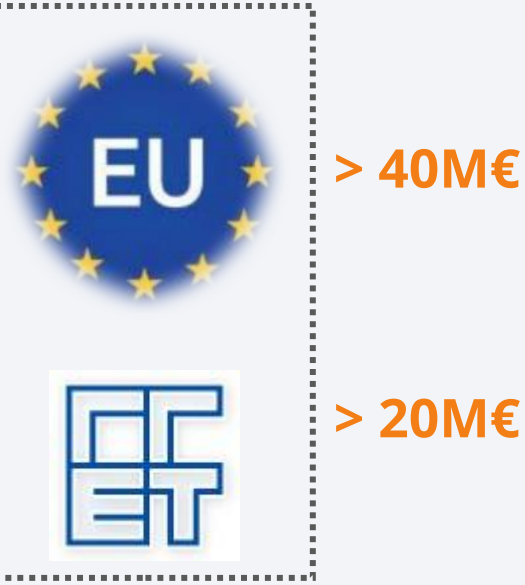


2006



- **07** Perm./Affil. Researchers
- **05** Post-doc Researchers
- **06** Post-grad Students
- **03** Research Engineers





> 40M€

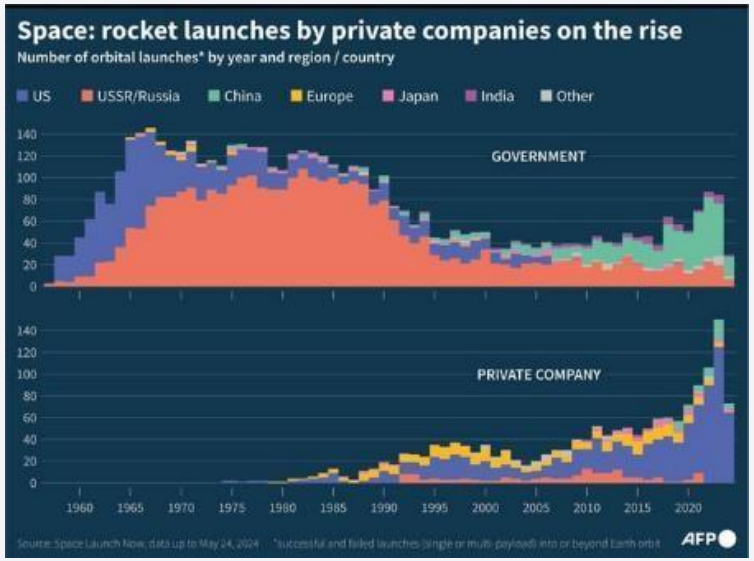
> 20M€

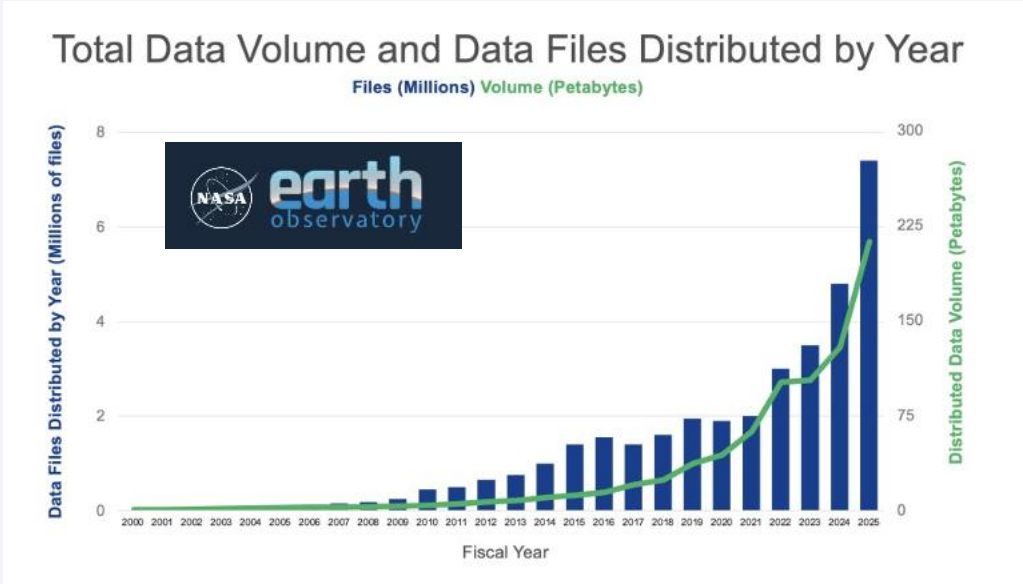
FORTH
INSTITUTE OF COMPUTER SCIENCE

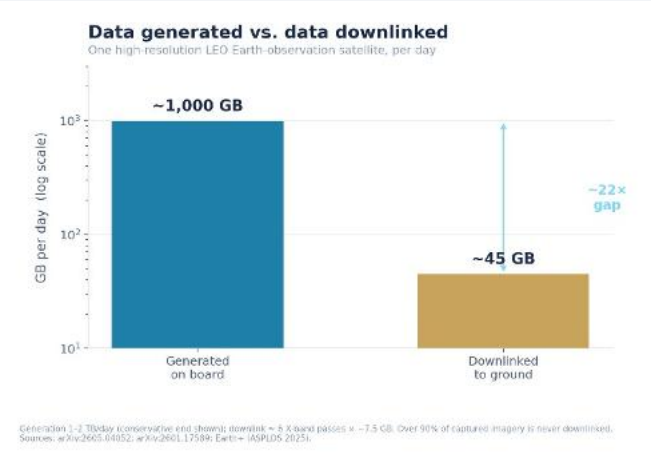
SPL

Big Data

#Earth Observation (EO)

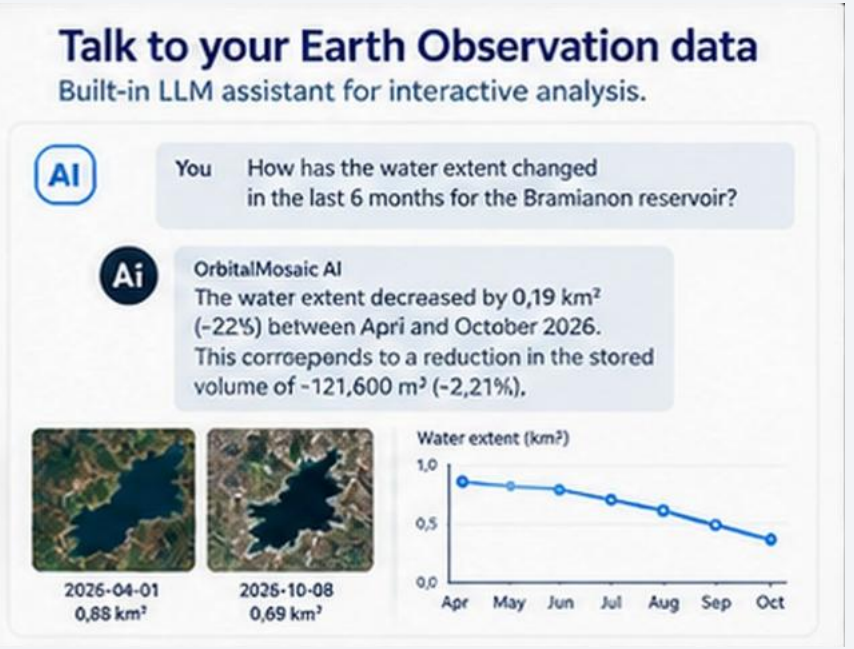






Video



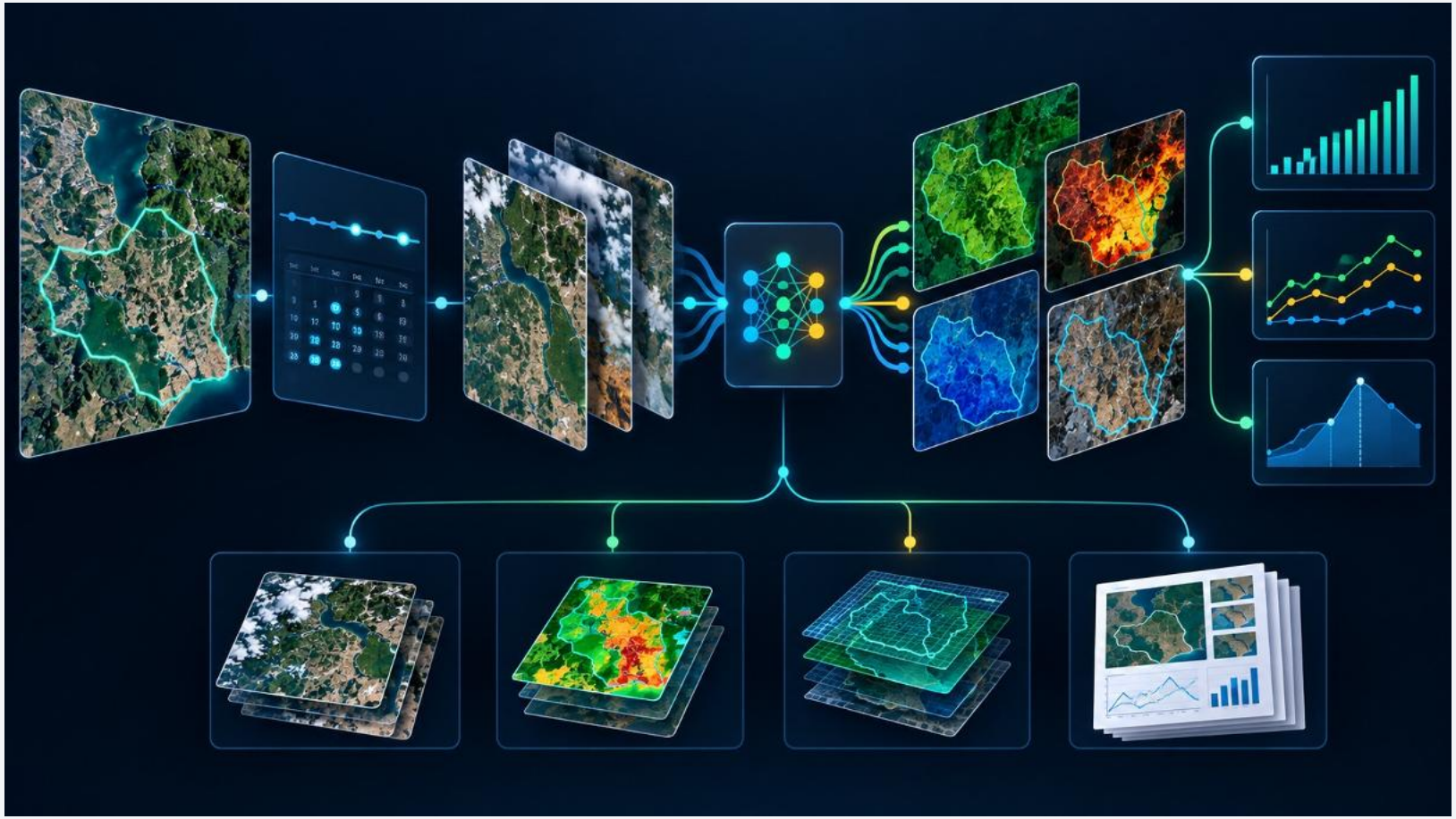


WEBSITE

FORTH
INSTITUTE OF COMPUTER SCIENCE

SPL

OrbitalMosaic platform



01 **Define area and period**

02 **Align EO and context**

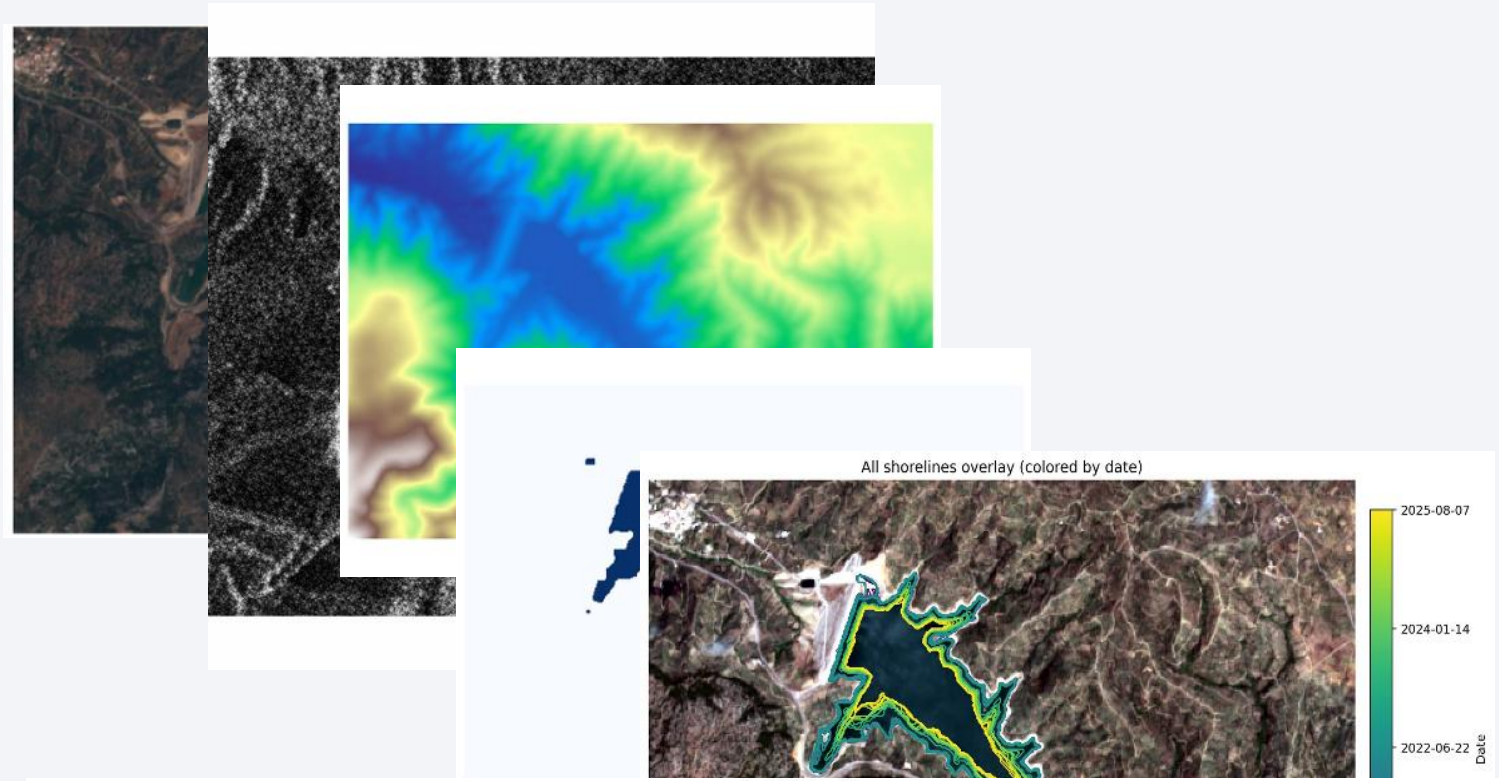
03 **Run an AI workflow**

04 **Review and export**

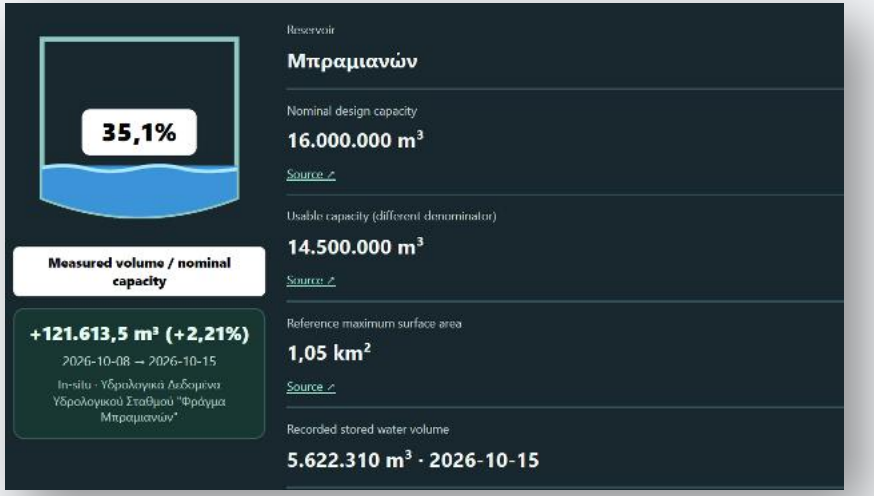
FORTH
INSTITUTE OF COMPUTER SCIENCE

SPL

Data





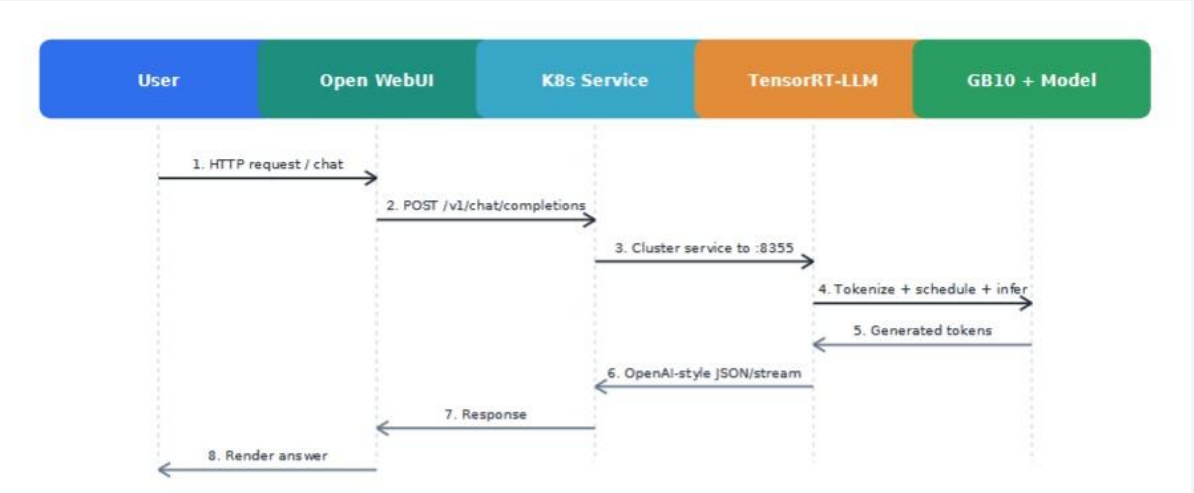


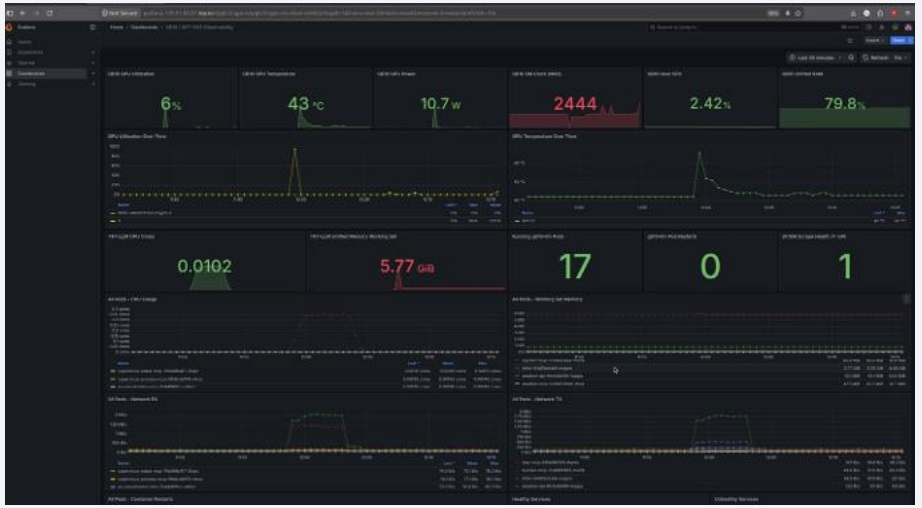
FORTH
INSTITUTE OF COMPUTER SCIENCE

SPL

Platform

UI	Open WebUI
LLM SERVING	NVIDIA TensorRT-LLM
MODELS	OpenAI, OmbriaNet (CNN)
TOOLS	Python MCP servers
CONTAINERS	Docker · Kubernetes
GPU SHARING	Kubernetes time-slicing
HARDWARE	NVIDIA GB10

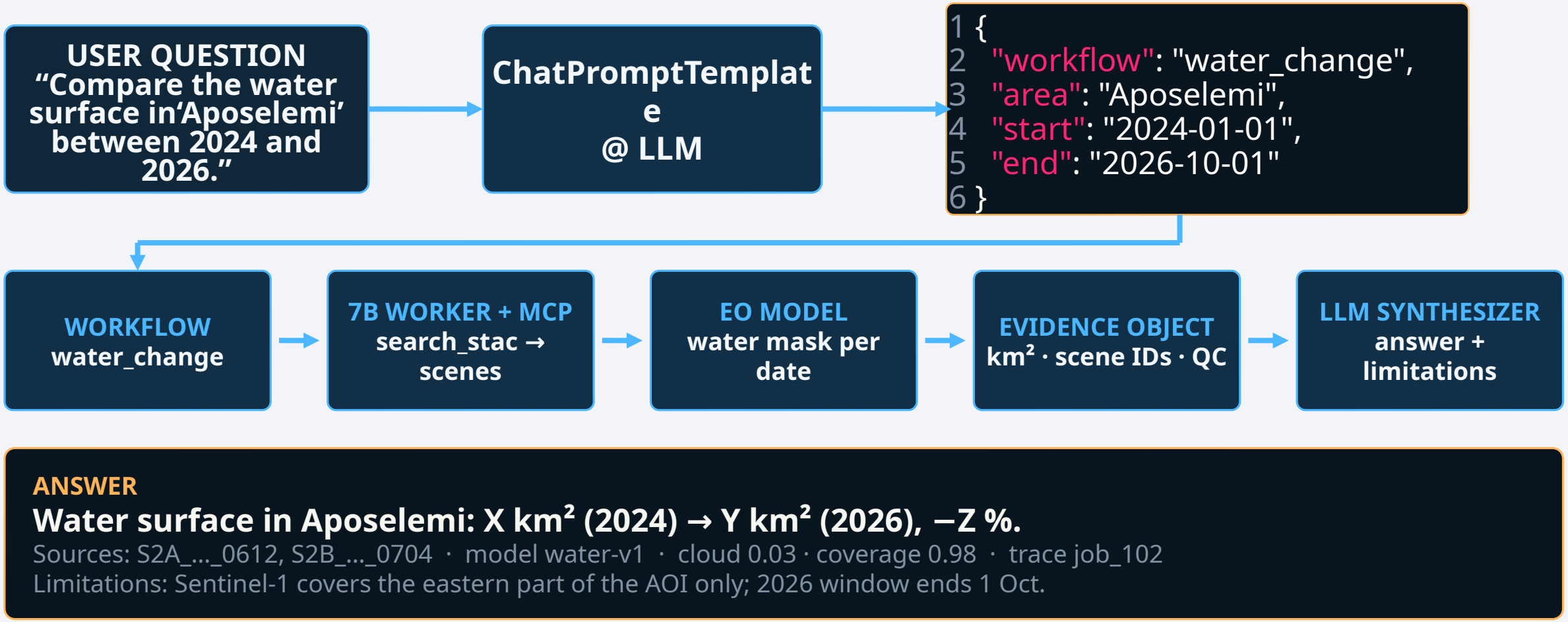




FORTH
INSTITUTE OF COMPUTER SCIENCE

SPL

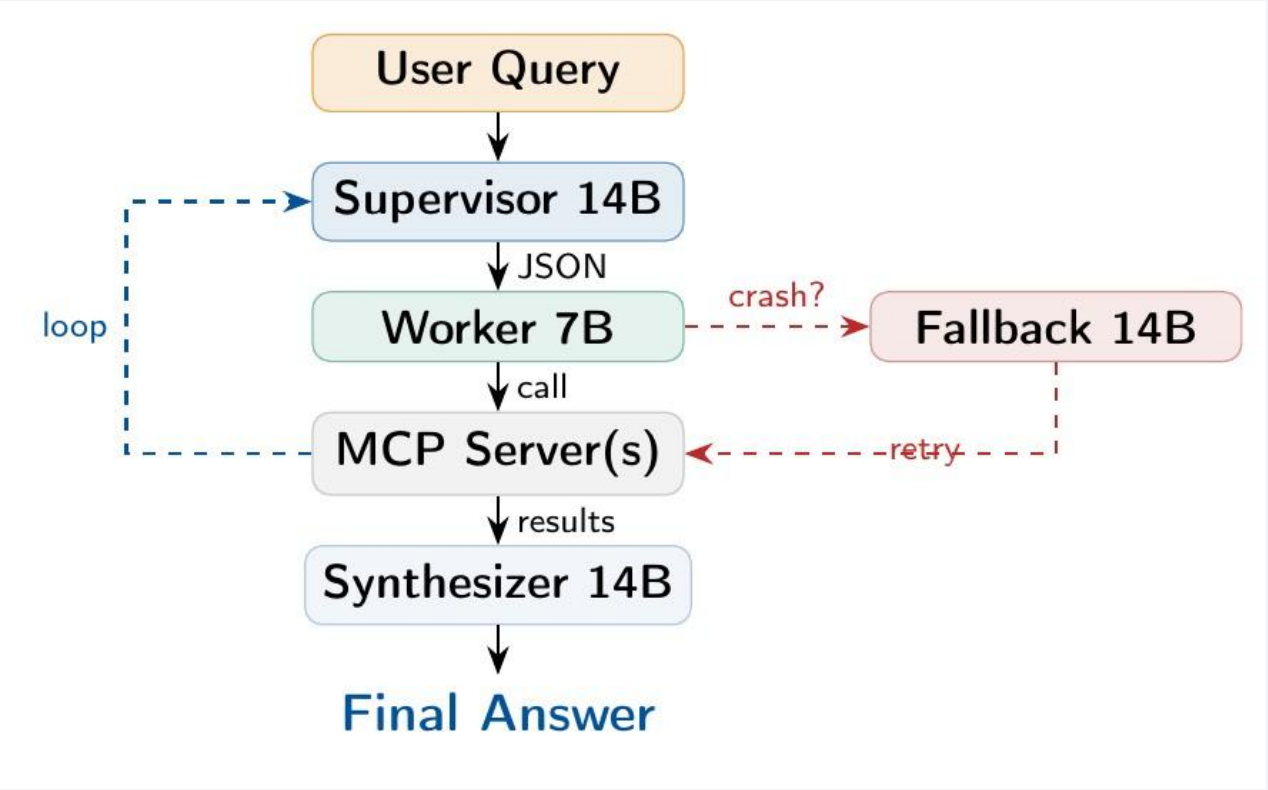
The Agentic Workflow



FORTH INSTITUTE OF COMPUTER SCIENCE

SPL

Agentic OrbitalMosaic architecture



LLM — one shared service
Supervisor · Synthesizer (14B testbench / 120B production)

SLM-7B — one shared service
Worker · Specific Role

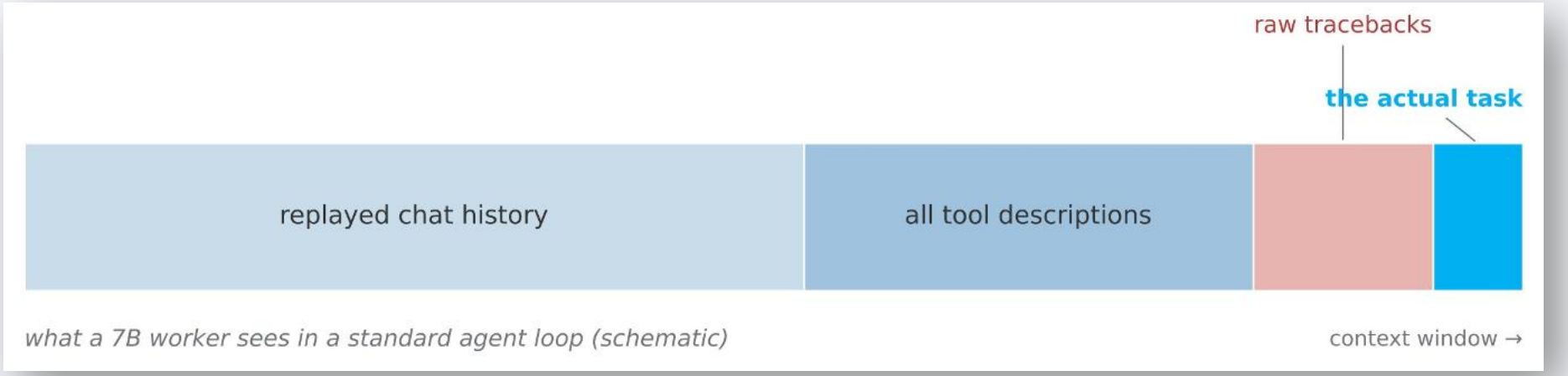
MCP servers
strict schemas · one capability per tool

STAC · S3 / COG · PostGIS · TSDB · EO models
Semantic layer & Databases

FORTH INSTITUTE OF COMPUTER SCIENCE

SPL

Limitations of typical Agentic frameworks



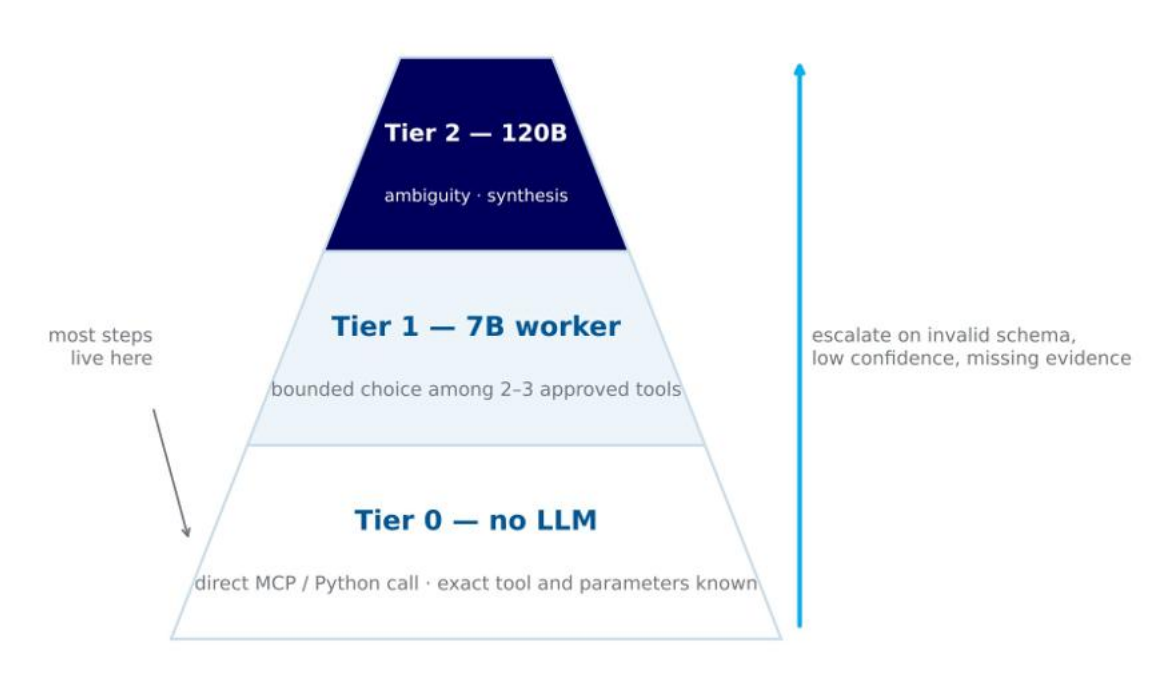
Failure modes
- chat noise
- tool flood
- loop fragility
- planning–execution gap

ReAct: think → act → observe → think → act… (loops)

↓

Ours: receive payload → call tool → done

Hierarchical Agentic Capabilities



LLM-120B MODEL — SHARED

LLM-120B INFERENCE SERVICE

SUPERVISOR | **SYNTHESIZER**

SLM-7B MODEL — SHARED

SLM-7B INFERENCE SERVICE

DATA DISCOVERY
STAC · S3 · PostGIS · TSDB

GEOSPATIAL CONTEXT
OSM · LULC · DEM

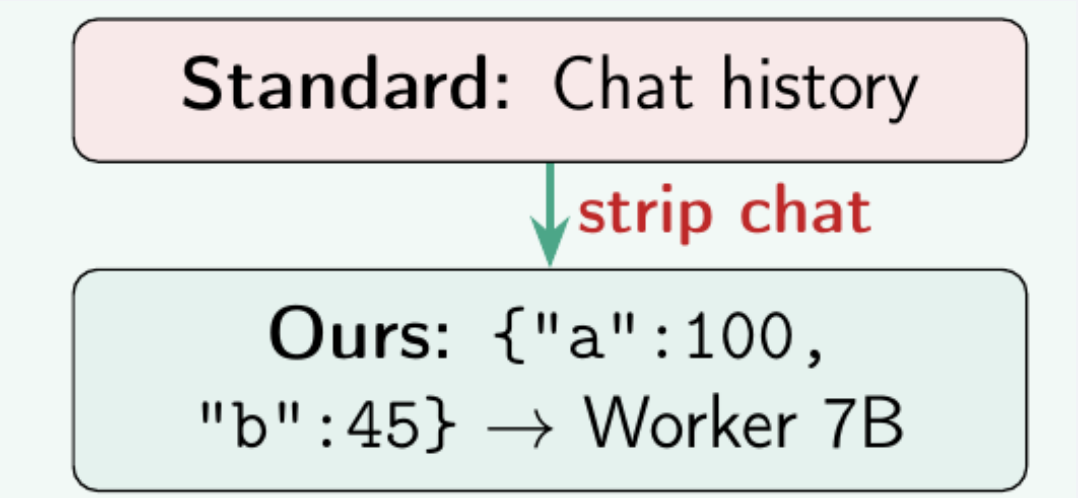
EO INFERENCE
Water · Burn · Change Detection

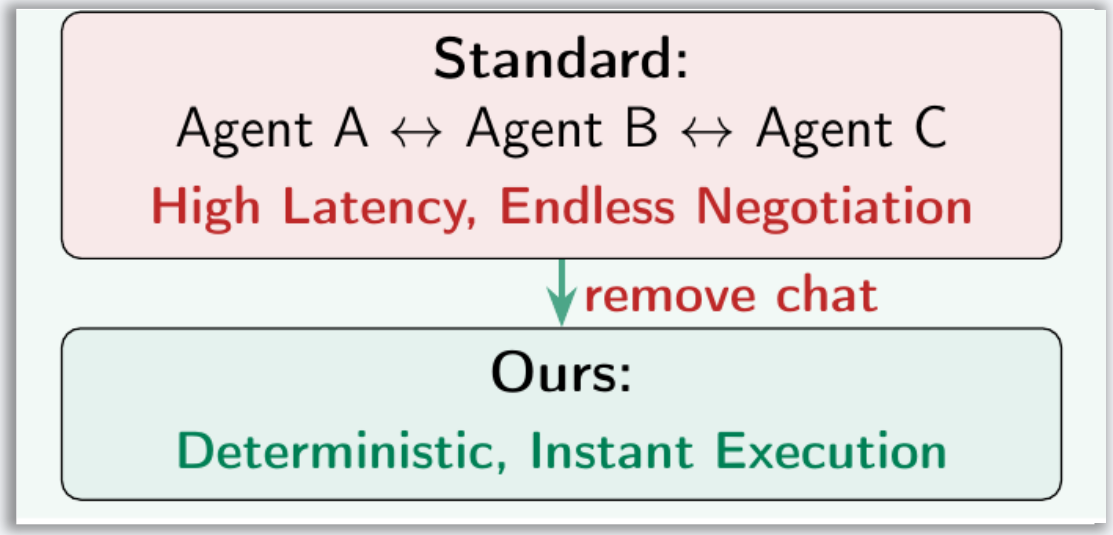
CLIMATE / WEATHER
ERA5 · Meteostat · APIs

FORTH INSTITUTE OF COMPUTER SCIENCE

SPL

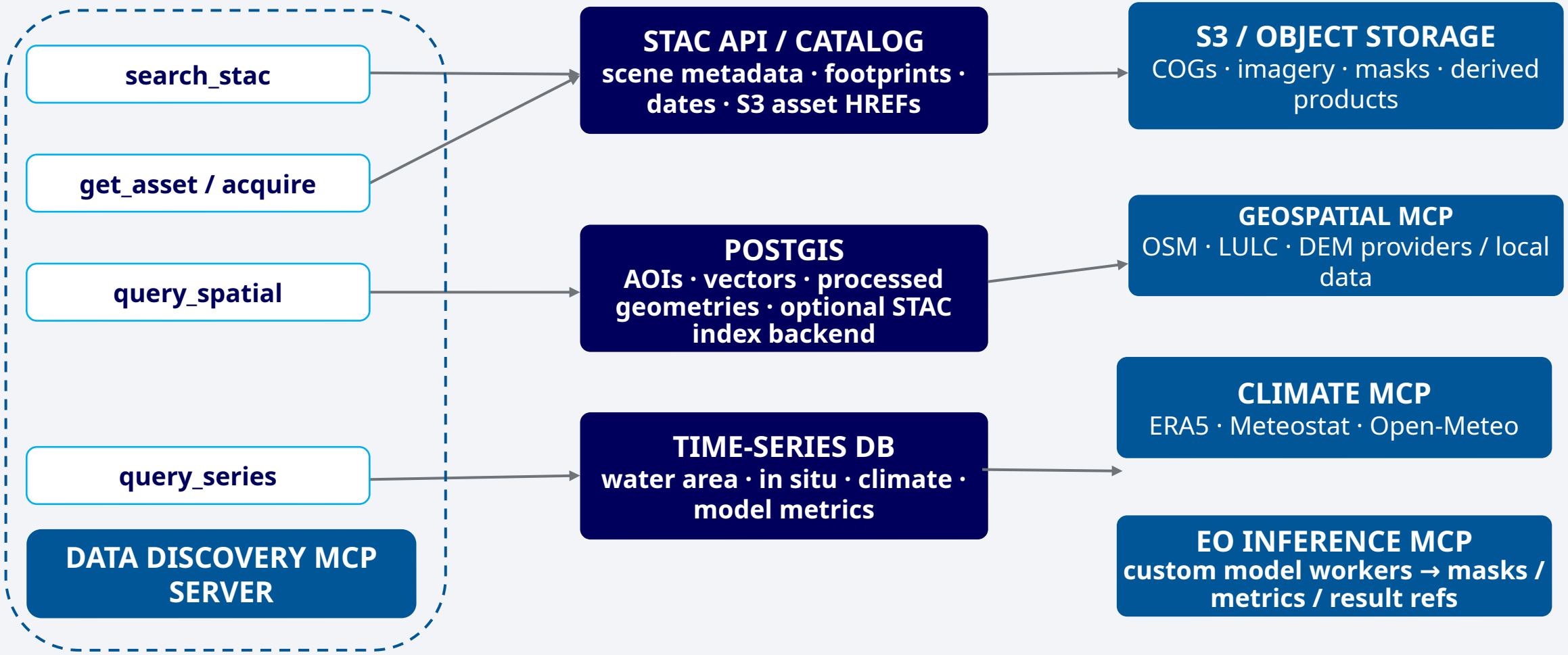
Design principle: Limit the chat history






```
{ "job_id": "job_102", "step": "select_stac_scenes",
  "aoi_ref": "postgis://aoi/42", "dates": ["2024-01-01","2024-12-31"],
  "tools": ["search_stac", "get_metadata"] }
```

MCP: Workers ↔ Independent data services



Agentic Interaction & Memory

PERSISENT STATE MODEL

Chat log
full transcript · recent turns

Long-term memory
saved AOIs · preferences · prior jobs

Job state
step · retries · asset refs · QC

EO data stores
STAC · S3 · PostGIS · TSDB

Tool registry
approved MCP schemas · permissions

WORKING MODEL MEMORY

1 Retrieve
tenant / project-scoped records only

2 Filter
drop unrelated turns, stale facts, excess tools

3 Compress
summarize under a fixed token budget

4 Assemble
question + relevant context + evidence refs

5 Call
120B prompt — or a 7B sterile payload

MEASURED — what reached each model (EMSR671 run, 16k wir

162
tokens per worker call
≈ 1 % of the window · 111 calls · 33 out

1,218
tokens per synthesizer call
≈ 7 % of the window · 73 calls · 68 out

ONE STERILE PAYLOAD

```
{ "job_id": "job_102", "step": "select_stac_scenes",
  "aoi_ref": "postgis://aoi/42",
  "dates": ["2024-01-01", "2024-12-31"],
  "tools": ["search_stac", "get_metadata"] }
```

≈ 160 tokens · 2 schemas

FORTH INSTITUTE OF COMPUTER SCIENCE

SPL

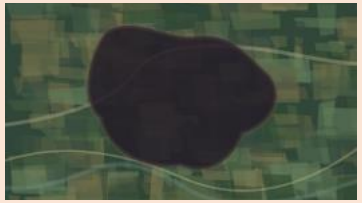
Consistency and reuse

1 Result cache
validated metrics + provenance

key = hash(tenant / project, AOI geometry, date range, source scene IDs, parameters, model version, quality scope)

2 Intermediate products
masks · mosaics · aligned assets


PRE-FIRE


POST-FIRE


MODEL MASK

3 Source scenes
reuse before acquiring again

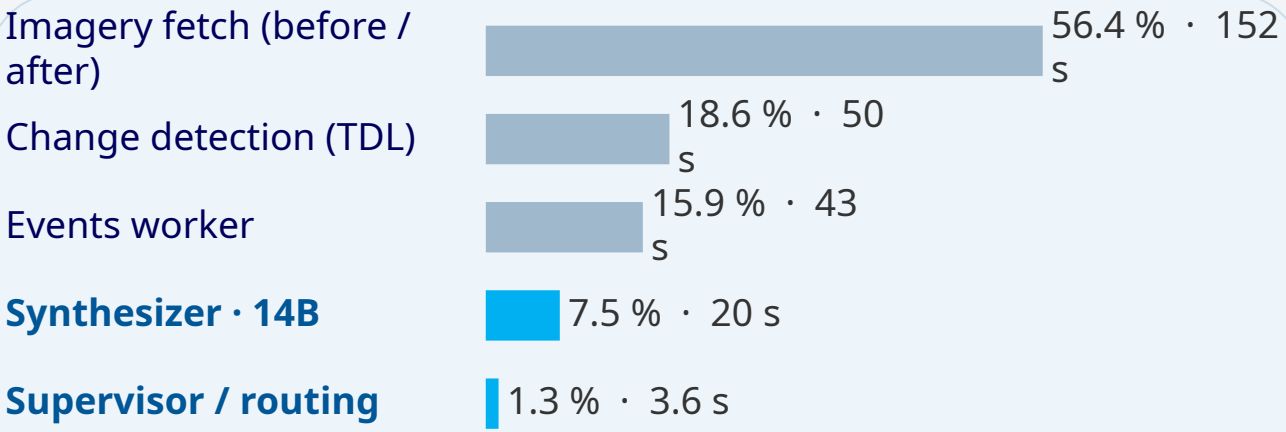


FORTH INSTITUTE OF COMPUTER SCIENCE

SPL

Time & token allocation

WALL-CLOCK PER QUESTION



Language models: < 9 % of wall-clock.

TOKENS — producing vs. judging answers (EMSR671)

115,586
application tokens
184 calls · 1,583 / question

≥ 501,714
judge tokens
85 attempts · lower bound

4.3× more tokens to grade the answers than to produce them

- Evaluation (not inference) is the dominant token cost of a grounded EO agent.

Takeaways

01 **Memory & Intelligence**
- Hierarchical Capabilities
- Databases & Persistent Storage

02 **Semantics**
- EO ontologies (silos)
- User Profile

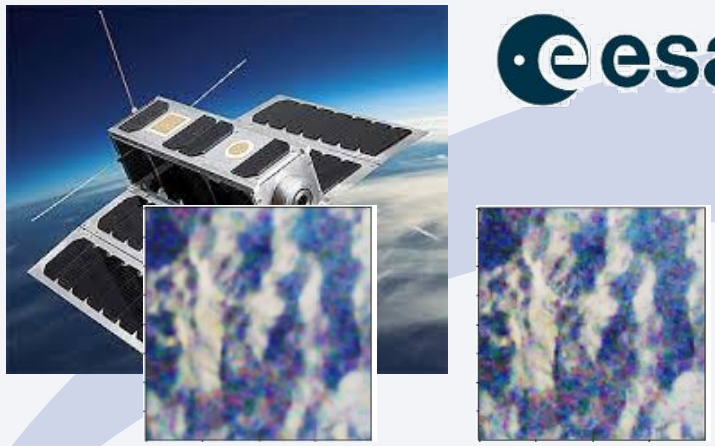
03 **Agent-User Coordination**
- Protocols
- Reusability

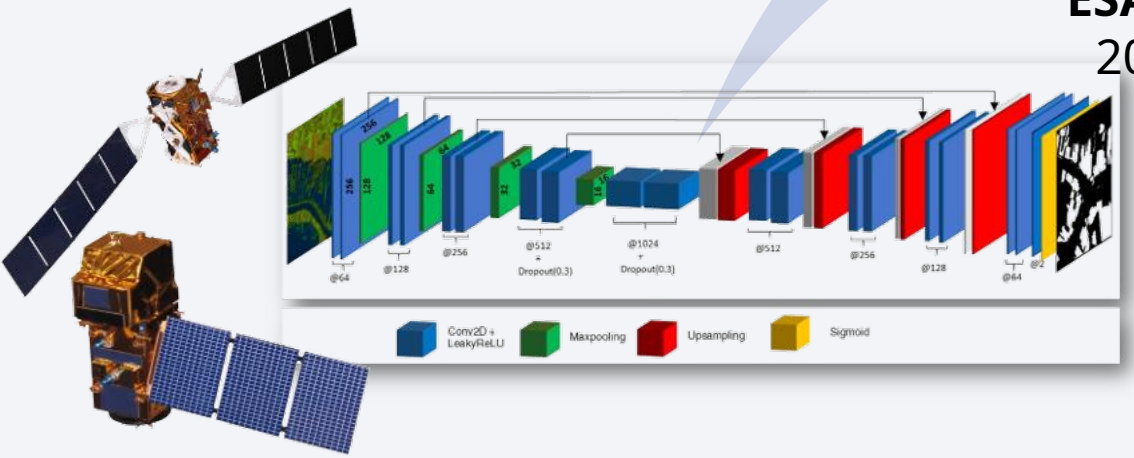
04 **ML analysis**
- Multiple models with different specs
- Platform data synchronization

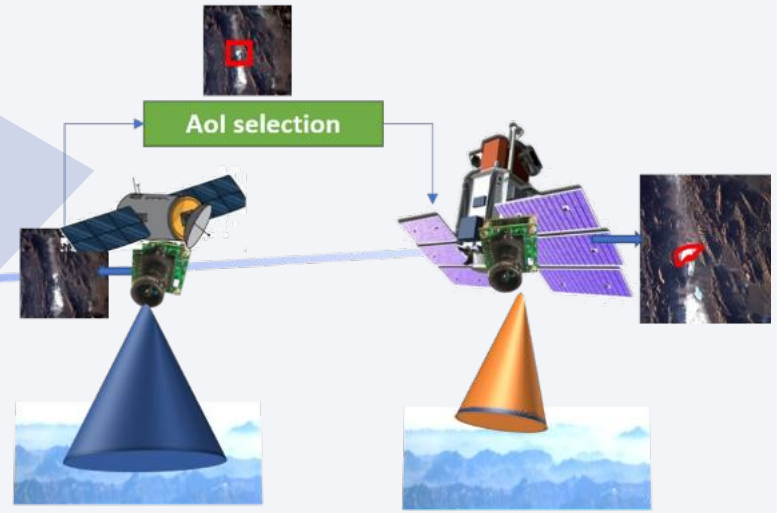
05 **Extending ML**
- Online & Incremental Learning
- Uncertainty Quantification

06 **Edge Computing**
- Model compression
- Data compression (embedding)

AI and *New Space*







ESA OPS-SAT
2022-2024

Thank you



https://spl.ics.forth.gr

FORTH
INSTITUTE OF COMPUTER SCIENCE

SIGNAL PROCESSING LAB