# Agentic AI for Earth Observation

Open concepts and architectures for the next generation of EO workflows

**Karol Bot Gonçalves**
Researcher · IT4Innovations National Supercomputing Center, VSB–TUO
22 September 2026 · 11:00–12:00 CEST

WELCOME

# Today, in 60 minutes

1. **Why now** — the EO data deluge meets agentic AI — 10 min
2. **What "agentic" actually means in EO** — planning, tools, memory, feedback — 10 min
3. **The open-source building blocks** — data → APIs → tools → models → agents → HPC — 15 min
4. **Three walkthroughs, step by step** — forest change · solar farms · flood impact — 13 min
5. **Trust, governance & discussion** — benchmarks, failure modes, EU AI Act, Q&A — 12 min

## You'll leave with

- A working mental model of agentic EO systems
- A reference architecture and a map of open tools
- A critical checklist: reliability, evaluation, governance

**Format notes**

No live demos or exercises — static, reproducible walkthroughs instead. Slides + an annotated link list may follow by e-mail.

**Questions: you can email me and I am happy to discuss specifics! karol.bot.goncalves@vsb.cz**

01 · WHY NOW

# Earth observation has a scale problem

**22 TB+**

of new Copernicus data, every single day

*EU Space Programme*

**50+ PB**

online at the Copernicus Data Space Ecosystem — 500,000+ registered users

*Nature Sci. Data, 2026*

**245 PB**

NASA Earth-science archive projected by 2025; NISAR alone adds ~80 TB/day

*NASA JPL*

**50+ yrs**

of Landsat — more than 10 million scenes and counting

*USGS*

**Data-rich, insight-poor.** The bottleneck is no longer pixels — it is expert time.

Every question still needs someone who knows sensors, formats, indices and GIS tooling. So most monitoring questions are never asked — or answered once a year instead of continuously.

# Why answering one question takes weeks

**Find data** → catalog jargon, tiles, orbits, revisit dates

**Download** → big archives, quotas, formats, storage

**Preprocess** → clouds, atmosphere, co-registration

**Analyze** → sensor-specific code, indices, thresholds

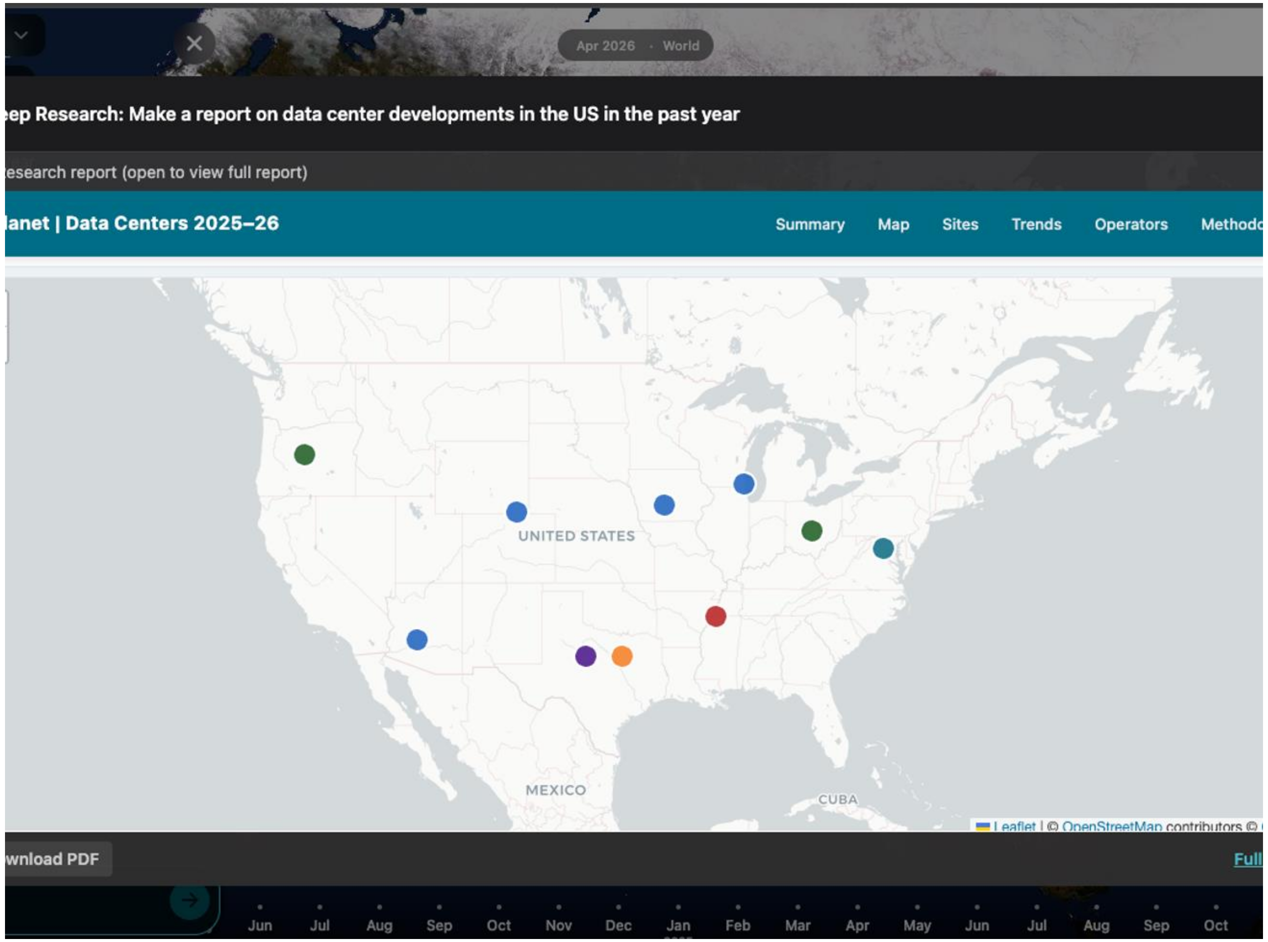
**Validate** → expert judgement, reference data

**Report** → manual GIS, cartography, writing

Each stage assumes a different specialism — remote sensing, data engineering, statistics, cartography. **Multiply by every region, hazard and reporting cycle → weeks of specialist effort per question.**

**Agentic systems attack the orchestration, not the science:** they automate the gluing between stages, while the measurements stay in verifiable tools.

# Planet turns a question into a map and a research report




### What is actually visible in the beta

| Surface | Status | Input | Process | Output |
|---|---|---|---|---|
| Map + chat | Open beta | Plain-language question | Multi-step research | Interactive map |
| Planet archive | May 2026 | AOI + topic | Change detection | Site evidence |
| Public sources | Product demo | Time window | Synthesis | Shareable report |
| Importa nt | MCP ≠ analytics | MCP: search/order | Beta: analysis | Check provenance |

### A deployed-looking interface is not proof of scientific validity

- Use it as motivation: the interaction model is real. Then ask what scene IDs, model versions, checks and uncertainty are exposed behind the polished report.

# Google's research agent grounds damage assessment in specialist models

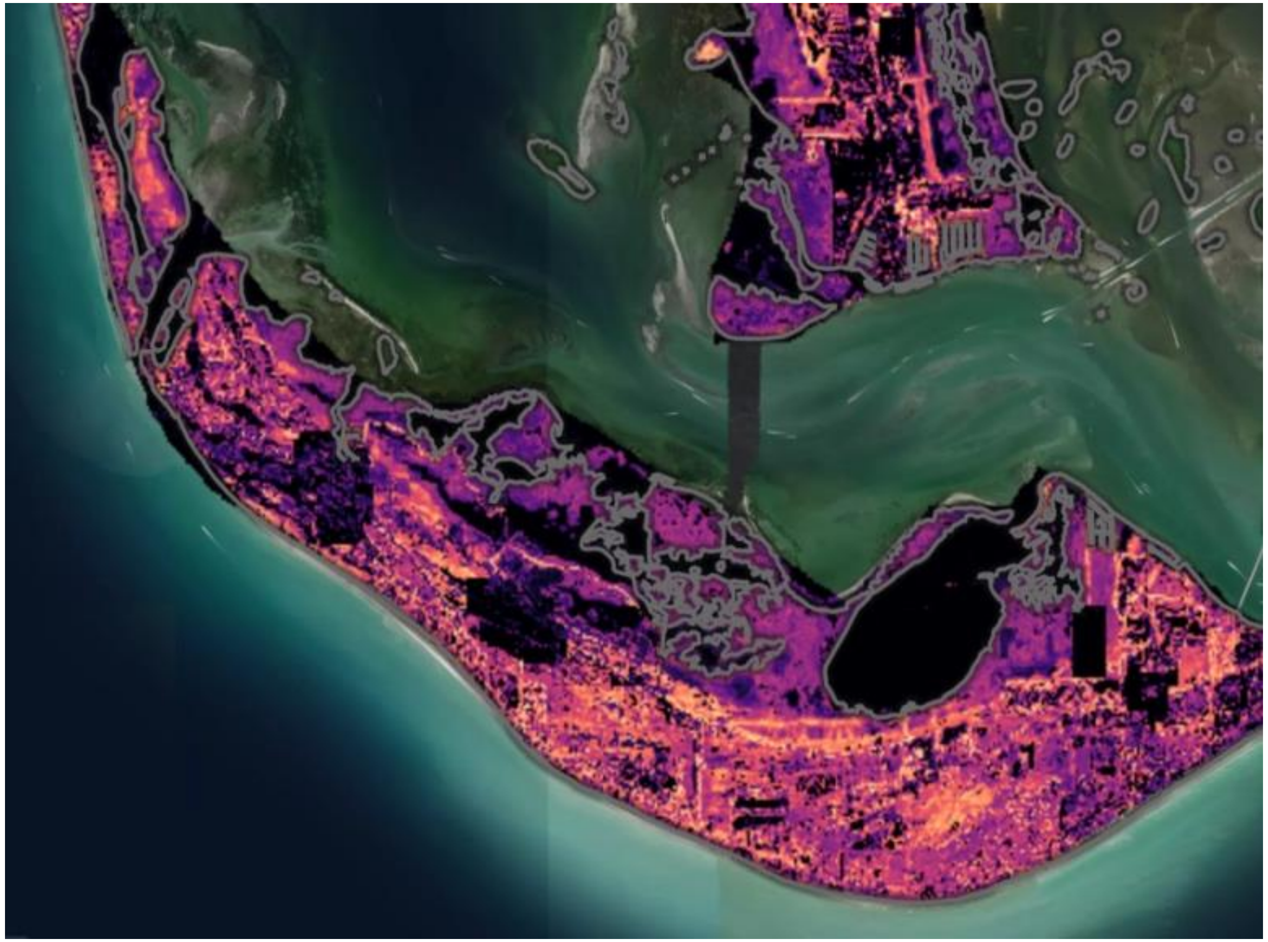

### What the research demonstration chains

| Stage | System | Evidence | Role | Maturity |
|---|---|---|---|---|
| Retrieve | Cloud Storage | Aerial image | Input | Demo |
| Detect | OWL-ViT RS | Bounding boxes | Tool | Research |
| Assess | Gemini / captioner | Image patches | Reason | Research |
| Summarise | Agent workflow | Damage narrative | Output | Edited demo |

**The visible boxes are tool outputs—not the language model "looking harder"**

- Google states that its demonstration video was edited for brevity. This is evidence of a workable architecture, not an independent accuracy claim.

# NASA Earth Copilot explores natural-language access to open Earth data



### Prototype facts—not a production claim

| Owner | Status | Interface | Backend | Access |
|---|---|---|---|---|
| NASA IMPACT | Prototype | Natural language | Azure OpenAI | Internal test |
| NASA VEDA | Exploration | Map + chat | Earth datasets | Researchers |
| Goal | Discovery | Complex questions | Cloud services | Broader access |
| Caveat | Evaluation | Misuse checks | Before release | No GA claim |

### Why this matters at IT4I: open data still needs a trustworthy access layer

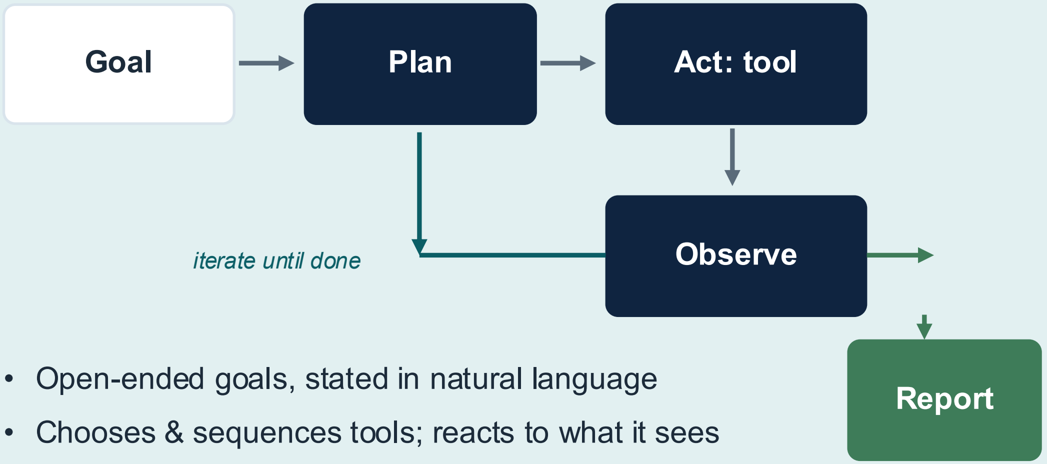
- Public systems cannot stop at fluent answers. They need reproducible queries, transparent data selection, access controls and reviewable evidence.

# From one-shot models to agentic systems

## A one-shot ML model


- One fixed task: cloud mask, crop map, ship detector…
- No context, no follow-up questions, no goal
- You still build the whole pipeline around it

## An agentic system




- Open-ended goals, stated in natural language
- Chooses & sequences tools; reacts to what it sees
- Stops when the goal is met — or asks for help

**Agentic AI** = an LLM-driven system that plans multi-step work, calls tools, keeps state, and checks its own progress toward a goal.

# Four capabilities — and one mental shift

### Planning

Decomposes “monitor deforestation” into concrete, ordered, checkable steps.

### Tool use

Calls catalogues, code and models — instead of guessing what pixels say.

### Memory

Keeps the AOI, intermediate products and decisions across steps and sessions.

### Feedback

Checks outputs, retries, re-plans — or escalates to a human.

**The mental shift: the LLM is the orchestrator — the science happens in the tools it calls.**

In EO terms: the model that writes the plan should never be the instrument that measures the world. Language models reason about the workflow; rasters are measured by deterministic, testable code.

# How an agent decomposes a real EO question

*"Has deforestation increased in region X since 2023 — and by how much?"*

### 1 Resolve the AOI
geocoder · OSM boundaries

### 2 Find usable imagery
STAC search: S2 L2A, cloud < 20%

### 3 Build pre/post composites
openEO / xarray median

### 4 Detect change
dNDVI + threshold + min-area

### 5 Quantify
zonal statistics, areas per year

### 6 Report
map + numbers + caveats + provenance

*Same skeleton for floods, urban growth, solar farms — only steps 4–5 change. The agent's job is choosing and wiring these steps for each new question.*

# An agent is not a script — and not just a model

| | Scripted pipeline | One-shot model | Agentic system |
|---|---|---|---|
| **Interface** | code & parameters | an API call | natural language |
| **Scope** | one workflow, fixed | one task, fixed | open-ended goals |
| **On failure** | crashes; needs a developer | fails silently | retries, re-plans, asks |
| **Determinism** | high | high | **stochastic — runs differ** |
| **Auditability** | read the code | model card | **must be engineered: logs** |

**Agents trade determinism for flexibility.** That trade is precisely what the governance section at the end is about.

# Division of labour: what the LLM must never do

### Let the LLM...

- Interpret the user's intent and constraints
- Plan and order the tool calls
- Write small snippets of analysis code
- Draft the report — from tool outputs only

### Never let the LLM...

- Read pixel values “by eye” from an image
- Count objects in a scene
- Produce numbers from memory
- Decide alone what is “true” on the ground

Evidence: GPT-4V on EO imagery — “good at captioning, bad at counting” (VLEO-Bench, CVPR Workshops 2024)

**Rule № 1 of agentic EO: every number in the report must be traceable to a tool output.**

# A reference architecture — every layer has open options

**L5 Interface & outputs** chat · notebook · QGIS → maps, statistics, reports

**L4 Agent runtime** open-weight LLM · agent framework · MCP tool bus

**L3 Tools & models** Python geostack · indices & change detection · EO foundation models

**L2 Discovery & access** STAC · openEO · S3 / COG / Zarr

**L1 Open data** Sentinel-1/2 · Landsat · DEMs · land cover · embeddings

### Runs through every layer

- Logging
- Provenance
- Human oversight

*Keep this picture in mind — the three walkthroughs will light it up, layer by layer.*

# Layer 1 — the open data is already there

### Sentinel-2 (optical)

10 m, 5-day revisit, 13 bands. Vegetation, land cover, water, change — the workhorse.

### Sentinel-1 (radar)

Sees through clouds and darkness. Floods, ice, ships, ground deformation — weather-proof monitoring.

### Landsat

30 m since 1972 — 10M+ scenes. The long time machine for trends and baselines.

### EO as embeddings — new

AlphaEarth Satellite Embeddings (annual, 10 m, open in GEE/AWS) · Major TOM (ESA Φ-lab, on Hugging Face): search the planet by similarity.

**All of it free and global. Copernicus data is full, free and open by regulation — a strategic asset for building agentic systems in Europe.**

# Layer 2 — discovery & access have standardised

**STAC — a common JSON language for "what imagery exists, where, when".** One search client works across providers — exactly what a tool-calling agent needs.

STAC v1.1 | OGC Community Standard 2025 | openEO: OGC 2026

**ONE SEARCH, ANY CATALOGUE — PYSTAC-CLIENT**

```
from pystac_client import Client
cat = Client.open("https://stac.dataspace.copernicus.eu/v1")
items = cat.search(
    collections=["sentinel-2-l2a"],
    intersects=aoi_geojson,
    datetime="2026-06-01/2026-08-31",
    query={"eo:cloud_cover": {"lt": 20}},
).item_collection()    # → scenes, as JSON
```

### Copernicus Data Space (CDSE)

free Sentinel archive: STAC + openEO + Sentinel Hub APIs + S3; cloud-native Zarr (EOPF) rolling out now

### Microsoft Planetary Computer

free STAC API over 100+ open collections (the classic catalogue — distinct from the paid "Pro")

### Earth Search (AWS / Element 84)

no-auth STAC for Sentinel-2/-1 COGs and Landsat on AWS open data

### NASA Earthdata CMR-STAC

all of NASA's Earth-science holdings, exposed as STAC catalogues

# Layer 3a — the analysis toolbox agents call

## The Python geostack

- pystac-client
- odc-stac
- xarray
- rioxarray
- Dask
- GeoPandas
- GDAL 3.12
- samgeo
- spyndex

*…all mature, maintained, scriptable — i.e. callable as agent tools.*

**The classic recipe (80% of monitoring):**
*search → composite → index → difference → threshold → vectorise → summarise*

## Spectral indices — simple, robust, explainable

**NDVI**
(NIR−Red) / (NIR+Red)
*vegetation vigour*

**dNBR**
NBRpre − NBRpost
*burn severity*

**NDWI**
(Green−NIR) / (Green+NIR)
*open water*

CHANGE DETECTION IN FIVE LINES — XARRAY

```
cube = odc.stac.load(items, bands=["red","nir"])
ndvi = (cube.nir - cube.red) / (cube.nir + cube.red)
change = ndvi.sel(time="2026-07").median("time")
       - ndvi.sel(time="2023-07").median("time")
loss = (change < -0.25) & (baseline > 0.6)

# deterministic, testable, unit-checkable —
# this is what the agent calls, not what it is
```

# Layer 3b — open geospatial foundation models

### Prithvi-EO-2.0

**NASA × IBM · 300M/600M params · Apache-2.0**

Multi-temporal ViT on NASA HLS data. Floods, burn scars, crop mapping. A variant has even flown in orbit (2026).

### TerraMind 1.0

**ESA × IBM · Apache-2.0**

Any-to-any multimodal: S-1, S-2, DEM, land cover, text. ESA-reported best-in-class in evaluations at up to 10× less compute.

### Clay v1.5

**Nonprofit · Apache-2.0**

General-purpose Earth embeddings for search, similarity and light-weight classification heads.

### SAM 2 + samgeo

**Meta · Apache-2.0 (SAM 3: custom licence)**

Promptable segmentation for imagery: fields, roofs, solar panels — click or text-prompt, get polygons.

**Fine-tune with TerraTorch / TorchGeo (community-run).** LoRA-scale budgets, not training from scratch. And remember: these models are tools the agent calls — not the agent.

# Layer 4 — an agent brain you can self-host

### Open-weight LLMs with tool calling (Aug 2026)

- **Qwen 3.x** · Apache-2.0 — strong agentic focus, all sizes
- **gpt-oss-120b** · Apache-2.0 — sized for a single 80-GB GPU
- **Mistral 3** · Apache-2.0 — EU-based provider
- **DeepSeek V3/V4** · MIT — frontier-scale open weights
- **Llama 4** · community licence — open weights, some terms
- **Gemma 4** · Apache-2.0 (new in 2026) — small & efficient

### Orchestration (open source)

**LangGraph 1.0 · smolagents · Pydantic AI · CrewAI**

*(AutoGen lives on as MS Agent Framework / AG2 — don't cite it by its old name)*

### Serving on your hardware

vLLM — the de-facto standard server
SGLang — big MoE models, multi-node
Ollama / llama.cpp — laptop-scale prototyping
*(HF TGI is in maintenance mode — avoid for new work)*

**One GPU node ≈ a private, per-token-free agent brain — your prompts and your data never leave your infrastructure.**

# MCP — the “USB-C port” between agents and EO tools


## Model Context Protocol

- An open standard for connecting agents to tools & data (born at Anthropic, Nov 2024)
- Since Dec 2025 governed by the Linux Foundation's Agentic AI Foundation
- Write a tool definition once — any MCP-capable model can use it
- **77+ geospatial MCP servers already catalogued (Sparkgeo, Aug 2026)**

### A TOOL, AS THE AGENT SEES IT

```
{ "name": "search_imagery",
  "description": "Find scenes in a STAC catalogue",
  "inputSchema": { "properties": {
    "bbox":     {"type": "array"},
    "datetime": {"type": "string"},
    "max_cloud": {"type": "number"}}}}
```

# Shortcuts — high-level tooling that exists today

### geoai (opengeos) *MIT*

Detection/segmentation pipelines, ~20 RS foundation models, agent-powered interactive mapping, QGIS plugin.

### samgeo *MIT*

SAM 1/2/3 on georeferenced imagery — text-promptable segmentation straight to GeoPackage.

### leafmap *MIT*

Interactive maps, STAC/COG browsing and results display in notebooks.

### QGIS agents *research-grade*

IntelliGeo (U. Twente) · GIS Copilot (Penn State): natural language → desktop GIS workflows.

### Curated maps of the field *open*

“Agentic AI in RS” survey + reading list (Jan 2026) · Sparkgeo geo-MCP index.

### Maturity, honestly *caveat*

Research-grade and moving fast: pin versions, expect rough edges, read the issues.

**Net effect: a capable agentic-EO prototype is days of work now — not months.**

# Where HPC fits — and how you actually get it

## Three jobs for a supercomputer

**Serve open-weight LLMs privately**
vLLM on GPU nodes: data governance + no per-token costs

**Fine-tune EO foundation models**
TerraTorch across multi-GPU nodes, at LoRA-scale budgets

**Scale the analysis itself**
Dask over national or continental AOIs, petabyte archives

## The access ladder — free for research

1. **IT4I Open Access** — 3 calls/year for Czech research: Karolina GPUs, Barbora NG — and the Czech share of LUMI
2. **EuroHPC JU calls** — EU-level, peer-reviewed — incl. a dedicated “AI & Data-Intensive Applications” track
3. **AI Factories — LAIF** — LUMI AI Factory (Czechia is a partner via IT4I/VSB-TUO): Playground access in ~2 days, Fast Lane ≤ 50k GPU-h, free for academia & SMEs. Plus CZAI — the Czech AI Factory (launched 2026, “KarolAIna” system incoming)

**The LAIF Hub Ostrava is open at IT4I**

# Three walkthroughs, step by step

Static, reproducible narratives — no live calls, no magic. What an open agentic EO system would actually do.

### A · Forest change

optical time series, change detection, trend

### B · New solar farms

object detection + verification against OSM

### C · Flood impact

SAR through the clouds + land-cover overlay

**How to read the next slides:** every step names a real open tool · every figure is an illustrative schematic · every number is invented for teaching (and marked so). *The pattern is real; the pixels are not.*

# The request, the plan, the calls

*“How much forest was cleared in NW Rondônia between July 2023 and July 2026? Map it, give me area per year — and tell me how sure you are.”*

## THE AGENT'S PLAN

1. **Resolve AOI to a polygon** *OSM / gazetteer*
2. **Find scenes: S2 L2A, July windows, cloud < 20%** *pystac-client → CDSE*
3. **Median composites for 2023 & 2026** *odc-stac + xarray*
4. **dNDVI < −0.25 on forest baseline → patches ≥ 1 ha** *numpy + GDAL*
5. **Areas per year, map, report with caveats** *GeoPandas + matplotlib*

## AGENT TRACE (ABRIDGED — VALUES ILLUSTRATIVE)

```
PLAN   5 steps → approved (auto: read-only tools)
CALL   stac.search(s2-l2a, bbox=[-64.3,-9.9,-63.1,-9.0],
       2023-06/2023-08, cloud<20)
  ↳    96 items · 19 usable
CALL   stac.search(…, 2026-06/2026-08, cloud<20)
  ↳    118 items · 22 usable
CALL   composite(median) × 2  ↳  2 cloud-free mosaics
CALL   ndvi_diff(thr=-0.25, min_patch=1 ha)
  ↳    214 patches · 186 km² → loss_polygons.gpkg
CHECK  cloud-gap coverage = 4.2% of AOI
  ↳    recorded as uncertainty in the report
CALL   zonal_stats(by=year) → drafting report…
```

*Note the CHECK step: a good agent budgets for its own uncertainty — it was planned, not improvised.*

# Real Landsat evidence: repeat observation makes change measurable



**2005 · Machadinho d'Oeste, Rondônia**

*The expanding 'fishbone' road-and-clearing pattern is visible in the real Landsat record. This image is evidence of land-cover change—not the pseudo-demo's invented 2023–26 result.*

### 1984 → 2005: what the pixels establish

- Repeat observations expose spatial change
- The agent still needs a declared AOI, comparable dates and cloud handling
- Area estimates require a reproducible classification—not visual counting

# Detection needs a verification design

*"Which solar farms have been built within 30 km of Ostrava since 2022? Give me a list I can actually verify."*

## THE AGENT'S PLAN

1. **Build a 30-km ring around the city** *GeoPandas buffer*
2. **Composites for 2022 and 2026** *STAC + odc-stac*
3. **Screen candidates: spectral + geometric signature** *raster ops*
4. **Segment & measure each candidate** *samgeo (SAM)*
5. **Cross-check OSM; rank by confidence; queue reviews** *Overpass API*

## AGENT TRACE (ABRIDGED — VALUES ILLUSTRATIVE)

```
CALL  buffer(city_boundary, 30 km) ↳ ring AOI
CALL  stac.search + composite for 2022, 2026
CALL  screen(low_NDVI + blue-grey + regular grid)
  ↳   31 raw candidates
CALL  sam_segment(31 candidates)
  ↳   17 panel-like · 14 rejected by shape
CALL  osm_check(power=plant, source=solar)
  ↳   7 matched · 5 high-score unmatched
  ↳   5 rejected (greenhouses, gravel pads)
PLAN  output = ranked list + human review queue

# values illustrative
```

*Teaching point: the agent plans its own cross-checks — detection without a verification design is just guessing at scale.*

# A real open workflow already maps photovoltaic farms from Sentinel-2

### ESA APEx example · openEO process

| Data | Execution | Training | Output | Licence |
|---|---|---|---|---|
| Sentinel -2 L2A | openEO | OSM polygons | PV mask | CC BY 4.0 |
| Time series | Cloud masking | Open data | GeoTIFF | Public |
| Feature extraction | Process graph | Pixel models | netCDF | Reproducible |
| Known limit | Validated ≤20×20 km | Proof of concept | Artifacts remain | Read docs |

**Open data + open process ≠ automatic generalisation**

- This is the kind of callable tool an agent can orchestrate. Its documented scale limit and known artifacts must travel into the final report.

# When the agent must know physics

*“Estimate the flooded area along the Odra during the September 2024 event — and how much farmland was affected.”*

**Real context: Storm Boris, Sep 2024 — rapid-mapping teams did this manually**

## THE AGENT'S PLAN

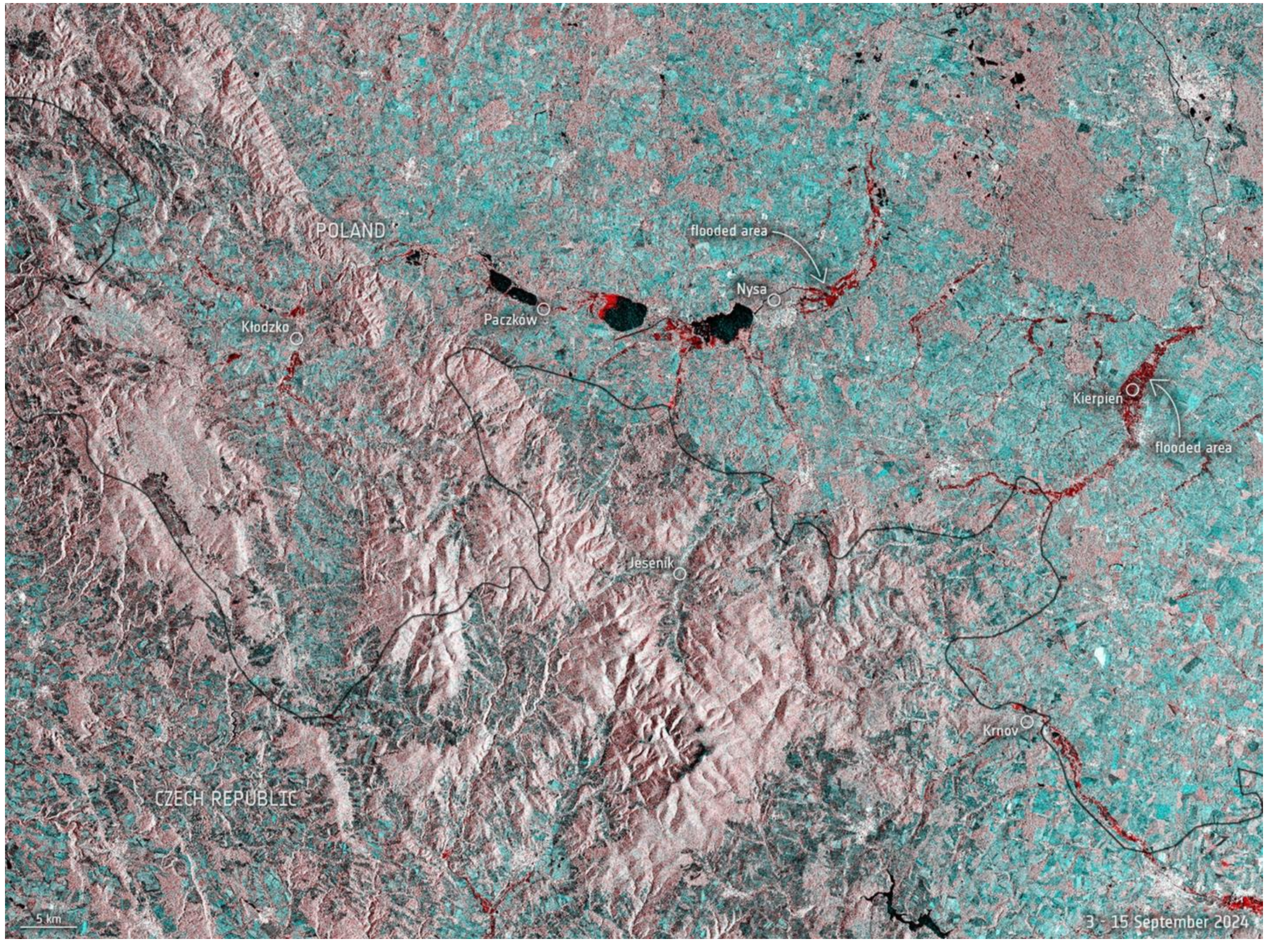
1. **Define event window + river-reach AOI** *gazetteer + hydro data*
2. **Sentinel-1 GRD pairs, pre & post event** *pystac-client → CDSE*
3. **Backscatter drop → water mask** *thresholding (Otsu)*
4. **Subtract permanent water** *JRC Global Surface Water*
5. **Overlay land cover & settlements → stats + report** *ESA WorldCover, GHSL*

## AGENT TRACE (ABRIDGED — VALUES ILLUSTRATIVE)

```
PLAN   optical unusable (storm clouds) → use SAR
CALL   stac.search(sentinel-1-grd, pre: 2024-09-08,
       post: 2024-09-16..20, same orbit)
  ↳    3 pre + 4 post scenes, VV+VH
CALL   water_mask(backscatter_drop, otsu)
CALL   subtract(jrc_permanent_water)
  ↳    flood_extent.gpkg
CALL   overlay(worldcover, ghsl_settlements)
CHECK  urban SAR caveat: flooded streets under-
       detected between buildings → flag in report
```

**The agent must know SAR beats optical in a storm.** That knowledge lives in tool descriptions and system prompts — engineered, not hoped for.

# Sentinel-1 saw Storm Boris through cloud and rain




**3 Sep**
pre-event Sentinel-1

**15 Sep**
post-event Sentinel-1

**RED**
areas affected by flooding

## WHAT THE REAL IMAGE PROVES

**sensor choice is part of the reasoning**

- Sentinel-1 radar observes through cloud and rain
- Multi-temporal comparison reveals change
- Flood extent still needs thresholds, permanent-water masks and validation

**So — what could go wrong?** That's the last section. →

# Reality check: agents still fail EO benchmarks

**33%**

of 140 real Earth-science questions answered correctly by state-of-the-art agents writing Earth-Engine code — and >58% of code runs failed.

*arXiv 2504.12110 · 2025*

**>84%**

of numerical answers fell outside acceptable tolerance — across ALL frontier models — on 403 executable Earth-system tasks.

*TerraBench · 2026*

**60.8%**

best model on a production-style environmental-API benchmark. Notably: open-weight models led the cost–accuracy frontier.

*GeoNatureAgent · 2026*

**Demos ≠ deployments.** Step-level audits (ThinkGeo, CVPR 2026) show the bottlenecks are tool selection and numeric grounding — not eloquence.

**Open, domain-specific evaluation is the research frontier — and it is wide open.**

# Failure modes → engineering countermeasures

| Failure mode | Countermeasure |
|---|---|
| **Hallucinated numbers** | numbers may only be copied from structured tool outputs — never generated |
| **Wrong tool, wrong parameters** | typed schemas, validated arguments, constrained tool choices |
| **Silent data gaps (clouds, missing scenes)** | mandatory coverage & uncertainty checks in every plan |
| **Overconfident prose** | report templates with caveats + links to intermediate products |
| **Fake or altered imagery** | provenance chains: STAC metadata · W3C PROV · C2PA (emerging) |

### Cautionary tale — July 2026

Google added generative imagery to Google Earth. Within hours, researchers produced fake scenes of sensitive sites. The feature was pulled in a day, pending “stronger guardrails”.

***In agentic EO, provenance is not optional.***

# EU AI Act: what applies now — and what to do meanwhile

TODAY

**Aug 2024**
in force

**Aug 2025**
GPAI duties + Code of Practice

**Aug 2026**
Art. 50 transparency · GPAI enforcement

**Dec 2027 / Aug 2028**
high-risk regime (deferred by the Digital Omnibus, Jul 2026)

## No agent-specific rules (yet)

- Agents inherit duties indirectly: via the GPAI model that powers them, via transparency (Art. 50) — and later via the high-risk regime
- Public-service decision support (floods, compliance…) will likely be high-risk when Annex III applies (Dec 2027)
- Neither the GPAI Code of Practice nor the guidelines mention agents — soft law and standards will move first

## Design to the spirit, today

- Human oversight (Art. 14): review queues before publication
- Automatic logging (Art. 12): keep the full agent trace
- Disclosure (Art. 50): say when users talk to a machine
- Provenance: every product traceable to scenes + code

***For public-sector EO decision support: build high-risk-grade discipline before it binds.***

Further reading: OECD “Agentic AI Landscape” (2026) · IAPS “AI Agent Governance: A Field Guide” · International AI Safety Report 2026

# An open research agenda — where you come in

### Evaluation

EO-agent benchmarks, error taxonomies, uncertainty scoring — the field's biggest gap and fastest publication route.

### Reliability engineering

Constrained tool use, self-verification steps, uncertainty propagation through multi-step pipelines.

### Domain agents

Fine-tuned open models for EO tool use: an 8B model matched much larger ones after fine-tuning (2025).

### Trustworthy deployment

Human-in-the-loop patterns, provenance chains, AI-Act-ready logging — with real public-sector pilots.

**Support exists:** compute & consulting via LAIF, CZAI and IT4I Open Access · community at the ESA × BIFOLD “Agentic AI for EO” workshop (Berlin, 19–21 Oct 2026) · open reading lists to start from

WRAP-UP

# Four takeaways

1. **Agentic = a loop, not a model** — planning + tools + memory + verification around an LLM orchestrator. The LLM never does the science alone.

2. **The open stack is complete** — Copernicus data → STAC/openEO → Python tools + open foundation models → MCP → open-weight LLMs on EuroHPC.

3. **Reliability is the frontier** — impressive demos, sobering benchmarks. Verification and evaluation are where research wins.

4. **Govern by design** — human oversight, automatic logs, provenance — from day one. Regulation will catch up to you.

DISCUSSION & Q&A

# Let's discuss

**Q1** Which recurring EO question in your work would you hand to an agent first?

**Q2** What evidence would it take for you to trust an agent-made map in an official report?

**Q3** Which decisions must stay human — even if the agent is usually right?

### After the session

Slides + an annotated link list by e-mail. The session was recorded.

### Contact

Karol Bot Gonçalves
IT4Innovations National Supercomputing Center

## Thank you.

# Selected sources — systems, data & tools

### Planet — agentic EO
planet.com/pulse: “How Agentic AI Transforms EO…” (21 May 2026) · Planet × Anthropic (6 Mar 2025) · Planet MCP server beta (12 Jan 2026)

### Google
Geospatial Reasoning (research.google, Apr 2025) · AlphaEarth Foundations / Satellite Embeddings (DeepMind, Jul 2025) · Earth AI (Oct 2025)

### NASA × Microsoft
Earth Copilot (Nov 2024) · Earth Copilot 2.0 coverage (Nov 2025) · Planetary Computer (free STAC catalogue)

### ESA
Φ-lab DA4DTE assistant (Mar 2025) · TerraMind with IBM (Apr 2025, Apache-2.0) · EarthCODE · ESA × BIFOLD Agentic-EO workshop, Berlin (Oct 2026)

### Standards
STAC 1.1 — OGC Community Standard (Oct 2025) · openEO — OGC Community Standard (May 2026) · MCP — Linux Foundation / AAIF (Dec 2025)

### Open data & APIs
dataspace.copernicus.eu (STAC · openEO · S3; EOPF Zarr) · planetarycomputer.microsoft.com · earth-search.aws.element84.com · cmr.earthdata.nasa.gov/stac

### Foundation models
Prithvi-EO-2.0 (huggingface.co/ibm-nasa-geospatial) · TerraMind (ibm-esa-geospatial) · Clay (clay-foundation) · SAM/samgeo

### GeoAI tooling
opengeos: geoai · samgeo · leafmap (opengeoai.org) · TerraTorch & TorchGeo (github.com/torchgeo) · IntelliGeo · GIS Copilot (QGIS plugins)

### Agent stack
LangGraph · smolagents · Pydantic AI · vLLM · SGLang · Ollama · geospatial MCP index: github.com/sparkgeo/geo-mcp-servers

### Data-volume stats
eu-space.europa.eu (22 TB/day) · Nature Sci. Data (CDSE, Feb 2026) · NASA JPL (245 PB, NISAR) · USGS EROS (Landsat archive)

*Full clickable link list ships with the slides.*

# Selected sources — research & governance

## Surveys & lineage

"Agentic AI in Remote Sensing" survey — arXiv 2601.01891 (Jan 2026, WACV-W) · LLM-Geo / Autonomous GIS — arXiv 2305.06453 · GIS Copilot — arXiv 2411.03205

## Benchmarks

Towards LLM Agents for EO — arXiv 2504.12110 (33% / >58%) · TerraBench — arXiv 2606.13148 (>84% numeric misses) · GeoNatureAgent — arXiv 2606.12821 (60.8%; open-weight Pareto) · ThinkGeo — arXiv 2505.23752 (CVPR 2026) · Earth-Agent/Earth-Bench — arXiv 2509.23141 (ICLR 2026)

## VLM limits

"Good at captioning, bad at counting" (VLEO-Bench) — arXiv 2401.17600 (CVPR-W 2024)

## EU AI Act

Regulation (EU) 2024/1689 · GPAI Code of Practice (10 Jul 2025) · Digital Omnibus on AI — Reg. (EU) 2026/1744 (OJ 24 Jul 2026): Annex III → 2 Dec 2027, Annex I → 2 Aug 2028 · artificialintelligenceact.eu (Art. 12, 14, 50)

## Agent governance

OECD "The Agentic AI Landscape" (Feb 2026) · IAPS "AI Agent Governance: A Field Guide" (arXiv 2505.21808) · International AI Safety Report 2026 · Chan et al., "Visibility into AI Agents" (FAccT 2024)

## EO quality & provenance

QA4EO (CEOS) uncertainty framework · W3C PROV-O · C2PA Content Credentials (emerging for imagery) · conformal prediction for EO — Sci. Rep. 2024

## HPC access

it4i.cz/en/for-users/open-access-competition · eurohpc-ju.europa.eu (AI & Data-Intensive Access; AI Factories access modes) · lumi-ai-factory.eu (LAIF · Hub Ostrava)

*Status of facts: verified August 2026. Re-check dates and versions before reuse.*