ORAL PRESENTATION

1st International Conference on Geospatial Artificial Intelligence (GeoAI 2026)
3-6 June 2026, Ghent, Belgium

Constraint-Aware AoV Planning for Orchestrating Autonomous Geospatial Analysis with LLMs

Annisa Puspa Kirana[1], Dr. Mahdi Farnaghi [2], Prof. Dr. Raul Zurita-Milla[3]

puspa.kirana@utwente.nl [1] m.farnaghi@utwente.nl[2] r.zurita-milla@utwente.nl [3]

Faculty of Geo-information Science and Earth Observation (ITC)
Department of Geo-information Processing

Supported by the Indonesian Endowment Fund for Education (LPDP) under Grant Number (0008346/REM/D/ASN-2024)

What If You Could Analyze Geospatial Data Just by Asking?

The vision: natural language → geospatial workflow

"Find the blind spots in Amsterdam's fire emergency response map."

↓ **LLM translates to executable steps**

- **Load** neighborhoods & fire stations
- **Reproject** to local coordinate system
- **Create** 500m around each fire station
- **Difference overlay** → identify blind spots
- **Visualize** result on map

Why this matters

GIS analysis is powerful but complex

Requires strict spatial rules: CRS, geometry types, data alignment

LLMs can plan multi-step workflows

LLM excel at decomposing complex tasks into sequences of operations

But LLMs lack of spatial rules knowledge

Without domain knowledge, workflows look correct but might geometrically invalid

Why Can't We Just Use Any LLM for Geospatial Workflows?

Root cause: LLMs lack geospatial domain knowledge

which leads to three types of spatial errors:

CRS Incompatibility

Buffering in degrees instead of metres → wrong distance calculations

Geometry Type Mismatch

Area calculation on LineString → mathematically impossible

Mismatched Spatial Extents

Overlaying non overlapping datasets → empty result

How existing systems fail to address this

Ignore it entirely

MetaGPT · AutoGen

Task planning without geospatial constraint awareness [Hong 2023, Li 2023, Wu 2023]

Catch error too late

GeoGPT · AutonomousGIS

Connect GIS tools to LLM reasoning | Spatial errors detected at runtime not pre-validated [Zhang 2024]

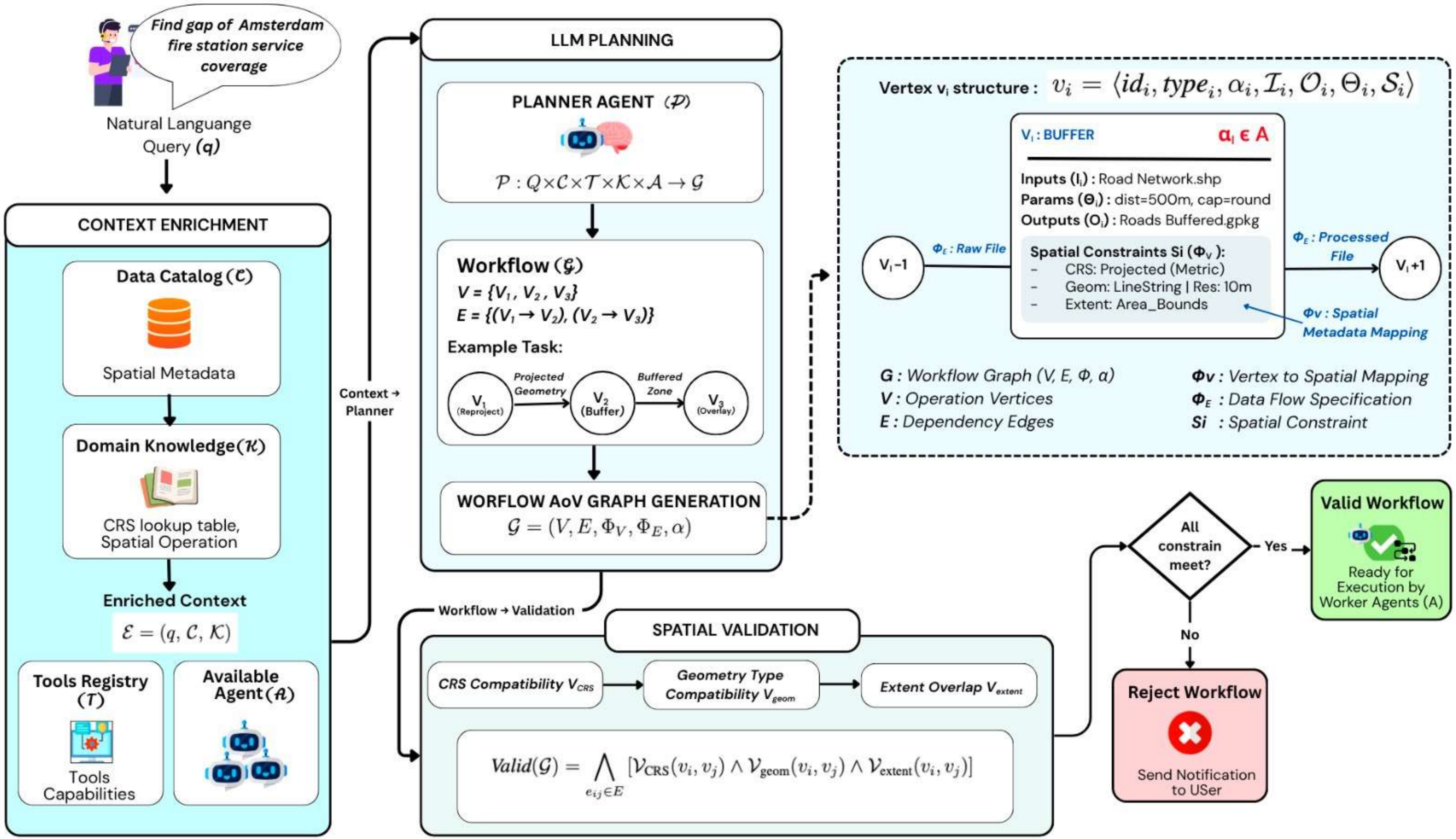
Solve parallelism, not spatial workflow

Flow · GeoFlow

Parallel AoV (Activity on Vertex) (**Flow**) , API/Tools optimization (**GeoFlow**) | Neither enforces spatial constraints on edges [Niu 2025, Bhattaram 2025]

Our contribution: Validate CRS · Geometry · Spatial Extent on every workflow edge BEFORE execution

Three-Stage Pipeline: Context Enrichment → LLM Planning → Spatial Validation



Building Spatial Awareness Before Planning

q	+	C	+	K	= Enriched Context E
User Query		Data Catalog		Domain Knowledge	

Available at planning time (from Data Catalog C)

Data Catalog C :

- **File paths and data formats**
- **Geometry roles** described in **metadata** ("polygons", "points")
- **Attributes**

Domain Knowledge K:
CRS lookup rule, operation constraints

Requires inference or file inspection

Actual Coordinate Reference System (CRS) : may differ from defaults (resolved from place names or file headers)

Precise bounding box : require reading the actual geometries

Cross-dataset spatial compatibility : whether two datasets overlap and share a common CRS

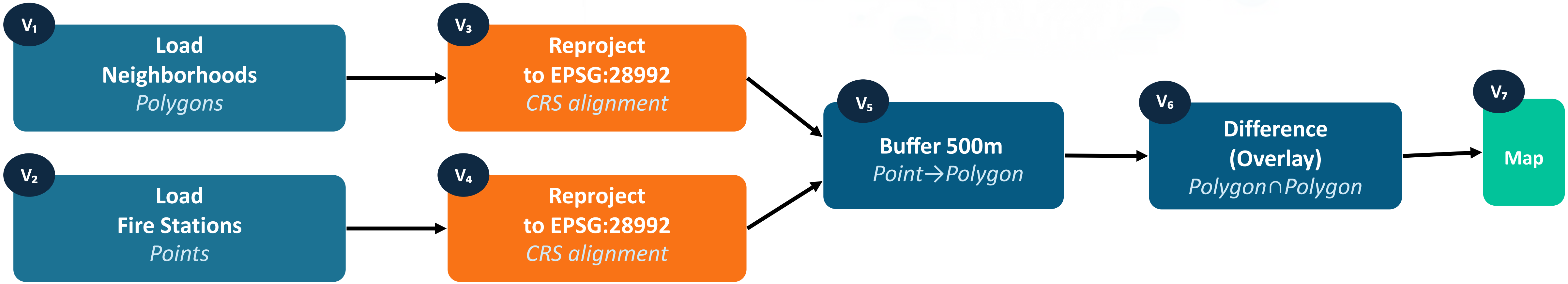
The enrichment bridges the gap between what users describe in natural language and what GIS operations require as structured spatial input

The Planner Agent Translates Queries into Workflows

$$P : Q \times C \times T \times K \times A \rightarrow G$$

- **P** — **Planner Agent** : LLM function orchestrating the workflow
- **Q** — **User queries** : natural language geospatial requests
- **C** — **Data Catalog** : Dataset metadata *(file paths, geometry types, attributes)*
- **T** — **Tools Registry** : available GIS operations *(buffer, overlay, reproject, etc.)*
- **K** — **Domain Knowledge** : spatial rules and constrains
- **A** — **Available Agents** : specialized workers (GeoprocessingAgent, VisualizationAgent, etc.)
- **G** — **Output Workflow** : the executable AoV graph: G = (V, E, Φ_V, Φ_E, α)

Example DAG — "Find blind spots in Amsterdam's fire emergency response"



STEP3: SPATIAL VALIDATION

LLM output JSON workflow → Parser extracts spatial fields → Check each edge (vi→vj) → All pass → ✓ Execute / Any fail → ✗ Reject)

V_CRS

CRS Compatibility

① Rearead output CRS of source

② Compare to required CRS of target

③ Match → pass

④ Differ → insert reproject

⑤ Infeasible reprojection → reject workflow

V_geom

Geometry Type

① Read output geometry type of vi

② Look up what vj's operation requires:

- buffer → accepts any geometry
- calculate_area → Polygon only
- overlay → Polygon ↔ Polygon

③ Incompatible → reject edge

V_extent

Extent Overlap (AABB)

① Read bounding box of vi and vj

② Apply Axis-Aligned Bounding Boxes (AABB) intersection test:

$(Xmin_1 \leq Xmax_2) \wedge (Xmin_2 \leq Xmax_1)$

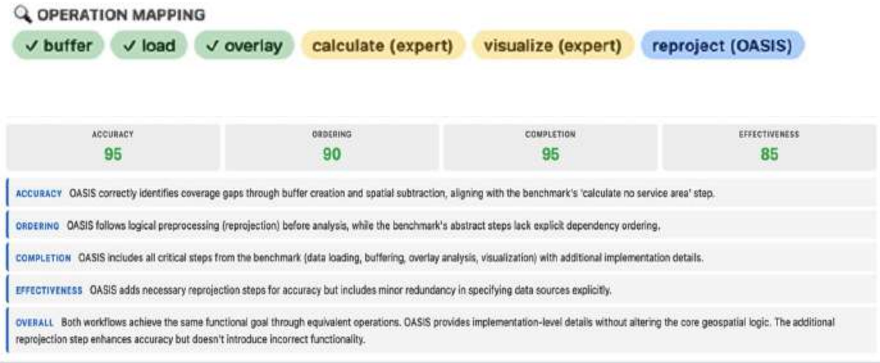
$(Ymin_1 \leq Ymax_2) \wedge (Ymin_2 \leq Ymax_1)$

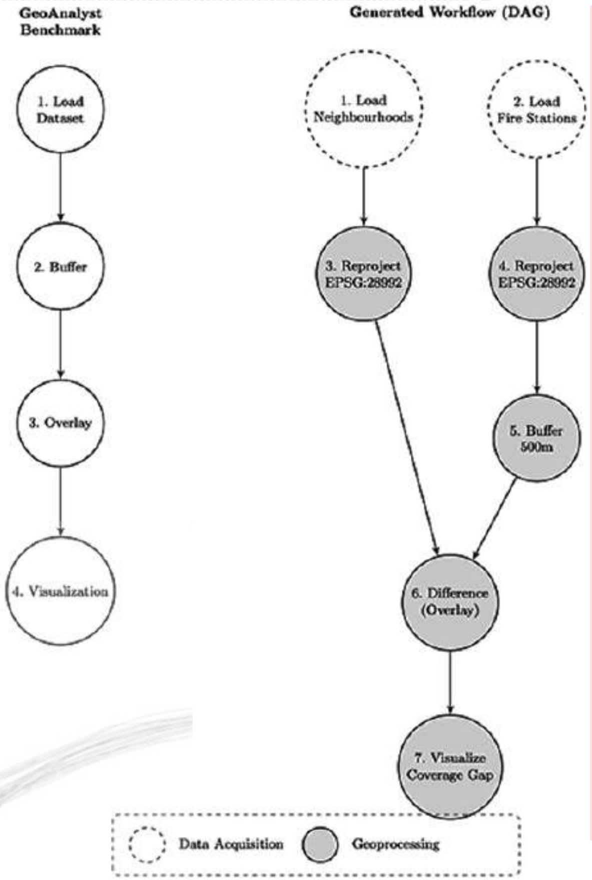
③ No overlap → reject workflow + notify user

$$\text{Valid}(G) = \wedge\ [\ V\_CRS(vi,vj) \wedge V\_geom(vi,vj) \wedge V\_extent(vi,vj)\ ] \quad \text{for all edges} \in E$$

EXAMPLE

"Find gap for Amsterdam fire station service coverage" - GeoAnalystBench





Expert Reference (4 steps, GeoAnalyst Benchmark)

1. Load Dataset
2. Buffer
3. Calculate no service area (Overlay)
4. Visualization

No CRS specified. No reprojection step.
No explicit file names or geometry types.

Our Agentic System (OASIS) - (7 steps, explicit)

1	load_geodata	Load neighborhoods (polygons)
2	load_geodata	Load fire stations (points)
3	reproject	Reproject neighborhoods
4	reproject	Reproject fire stations
5	buffer	500m service coverage
6	difference	Neighborhoods minus buffer
7	create_map	Display coverage gaps

Jaccard (string match): 53.57
LLM judge Overall Similarity: 93 / 100

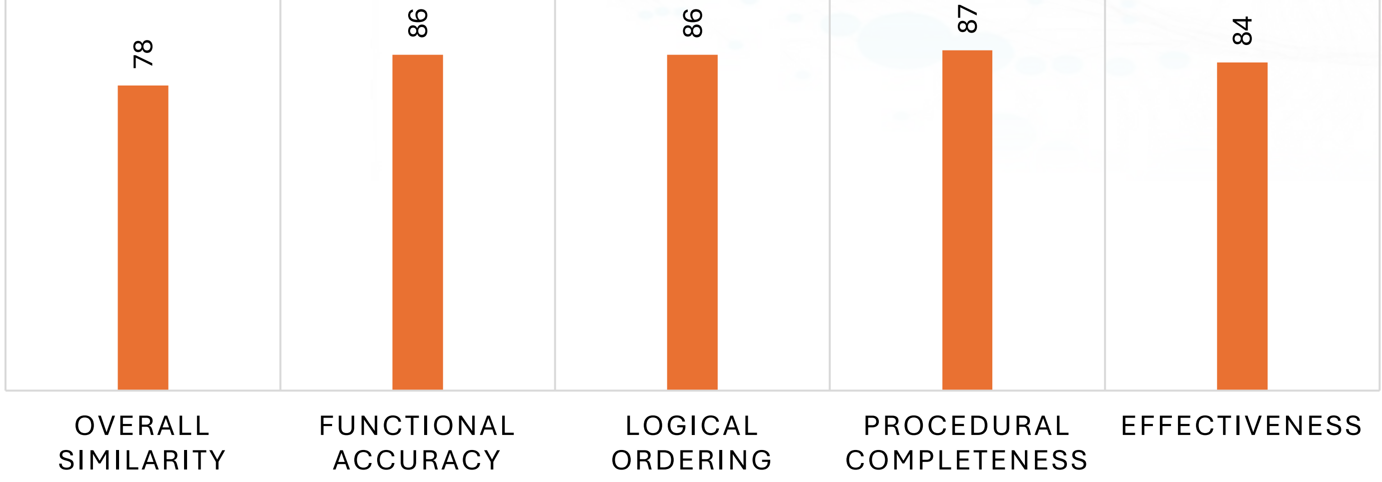
50 Expert Tasks | 3 Models | 150 Workflows | GPT-4o-mini Judge

50
Expert-validated geospatial tasks

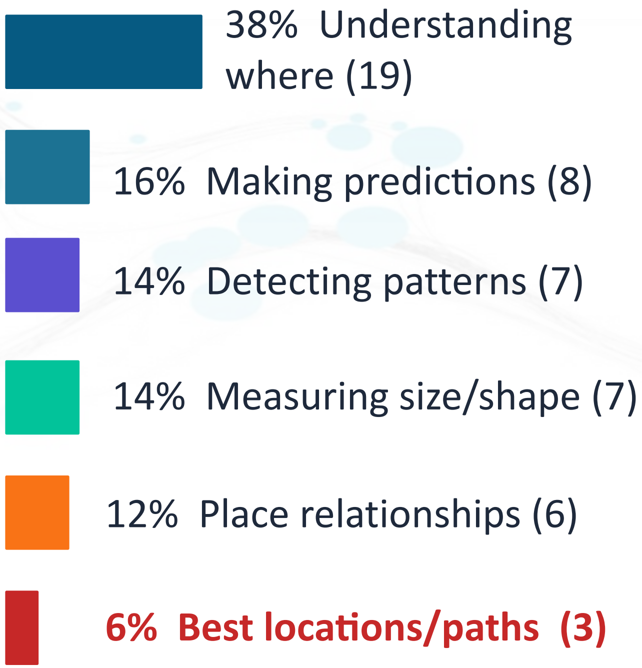
3
Open-weight models (local)

150
Total workflows evaluated

5 Scoring Metrics : 0-100 scale:



Task Category Distribution (GeoAnalystBench)



The Verbosity Gap: 4 abstract steps → 7 explicit steps

Metric	Qwen3-8B	Phi-4-14B	Gemma3-27B	Mean
Overall Similarity	**78.5**	78.0	77.5	**78**
Functional Accuracy	86.7	86.5	85.3	**86**
Logical Ordering	**88.2**	84.3	85.3	**86**
Procedural Completeness	**89.3**	87.7	85.0	**87**
Effectiveness ↓ (verbosity)	87.1	83.5	80.8	84

OASIS produces more steps than reference by design:

Benchmark (abstract)	OASIS (explicit)
"Load Dataset"	*load_geodata(neighborhoods) + load_geodata(fire_stations) : one per file*
"Buffer"	*reproject(EPSG:28992) + buffer(500m) : CRS alignment required first*
"Overlay"	*difference(neighborhoods, buffer_union) : explicit operation + inputs*

Limitations

Structured output schema + spatially enriched prompts appear to normalize performance similarity accuracy across model scales. Qwen3-8B ≈ Gemma3-27B at ~78%. But this is a hypothesis ablation study needed to confirm.

1

Single Judge Bias

GPT-4o-mini may reward workflows matching its training patterns. Two valid but differently-ordered workflows could score differently. [Gu 2024]

2

No Ablation Study

Individual contribution of V_CRS, V_geom, V_extent unquantified. We will do it in the future publication

3

No Execution Testing

90% scoring workflow planning could still fail when actually run due to data or tools incompatibilities.

Conclusion & Future Work

KEY CONTRIBUTIONS

- ✓ Spatial constraint injection + rule-based pre-validation → geometrically valid geospatial workflows from LLMs
- ✓ 78% semantic alignment with expert workflows : consistent across 8B to 27B models (Qwen3, Phi-4, Gemma3)
- ✓ Fully local deployment via Ollama reproducible, no proprietary API, no cloud dependency

FUTURE WORK

→ Second judge and manual evaluation using sequence-based metrics (e.g., LCS : Longest Common Subsequence) to validate functional equivalence between generated and reference workflows step-by-step

→ Ablation to quantify V_CRS · V_geom · V_extent individually

→ Execution-based metrics : actually run the workflows

→ Evaluation using additional benchmarks

Thank You

Questions?

Annisa Puspa Kirana
PhD Candidate
ITC-TECH
puspa.kirana@utwente.nl

Dr. Mahdi Farnaghi
Assistant Professor
ITC-TECH
m.farnaghi@utwente.nl

Prof. Dr. Raul Zurita-Milla
Professor
ITC-TECH
r.zurita-milla@utwente.nl

LinkedIn

Faculty of Geo-information Science and Earth Observation (ITC)

UNIVERSITY OF TWENTE. | ITC lpdp