A Graph-based and LLM-agnostic Agentic Approach for Natural Language Interaction with PostGIS Databases

Giovanni Piccolo[1,2] Cinzia Sada[1]

[1]Università degli Studi di Padova — Dept. of Physics & Astronomy “Galileo Galilei”, Padova, Italy
[2]Engineering Ingegneria Informatica S.p.A., Rome, Italy

1st International Conference on Geospatial Artificial Intelligence
GeoAI 2026 • Ghent, Belgium • 3–6 June 2026

UNIVERSITÀ DI PADOVA

eng

The GIS Complexity Barrier

PostGIS: Powerful yet Complex

High **technical barrier** for **non-GIS** experts

Hundreds of specialized functions: `ST_DWithin`, `ST_Intersects`, `ST_Buffer`, `ST_Area`…

Spatial Reference Systems (**SRIDs**) are **error-prone**

Spider & BIRD [1] [2]: general Text-to-SQL, no geospatial coverage

GeoSQL-Eval [3]: LLMs struggle with spatial predicates and SRID transformations

Geo-Gap in NLP Benchmarks

UNIVERSITÀ DI PADOVA
Dipartimento
di Fisica e Astronomia
"Galileo Galilei"

1. Yu, Tao, et al. 2018. "Spider: A Large-Scale Human-Labeled Dataset for Complex and Cross-Domain Semantic Parsing and Text-to-SQL Task." In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing.
2. Li, Jinyang, et al. 2023. "Can LLM Already Serve as a Database Interface? A BIG Bench for Large-Scale Database Grounded Text-to-SQLs."
3. Hou, S., et al. 2025. "GeoSQL-Eval: First Evaluation of LLMs on PostGIS-Based NL2GeoSQL Queries." arXiv preprint arXiv:2509.25264.

Two Critical Failure Modes in Spatial LLMs

#1 — Linguistic Ambiguity

Vague spatial prepositions are context-dependent and lack direct SQL equivalents.

"near a city" ≈ 15km
"near a lake" ≈ 1km
⇒ no single `ST_Dwithin` threshold!

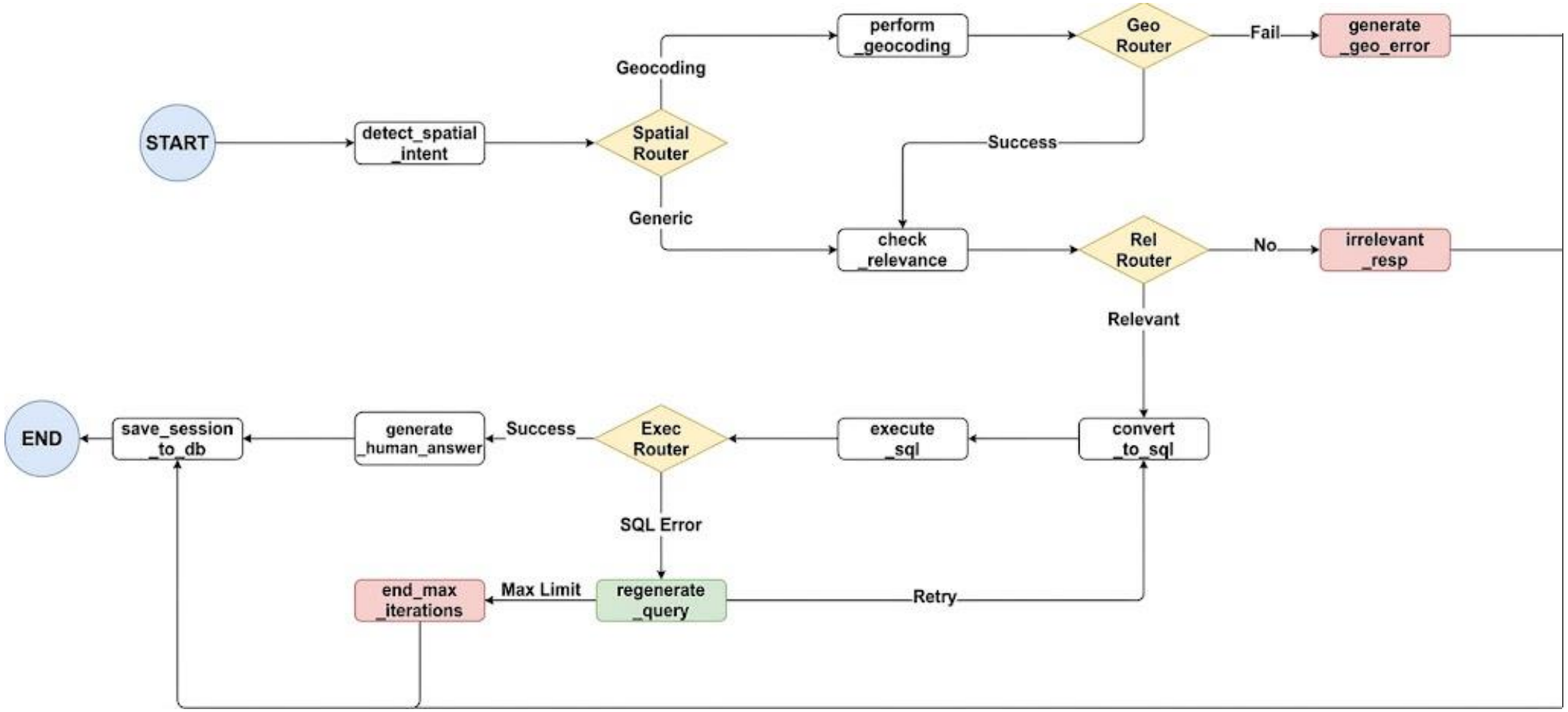
#2 — Function Hallucination

LLMs invent non-existent functions or misuse signatures due to lack of domain grounding.

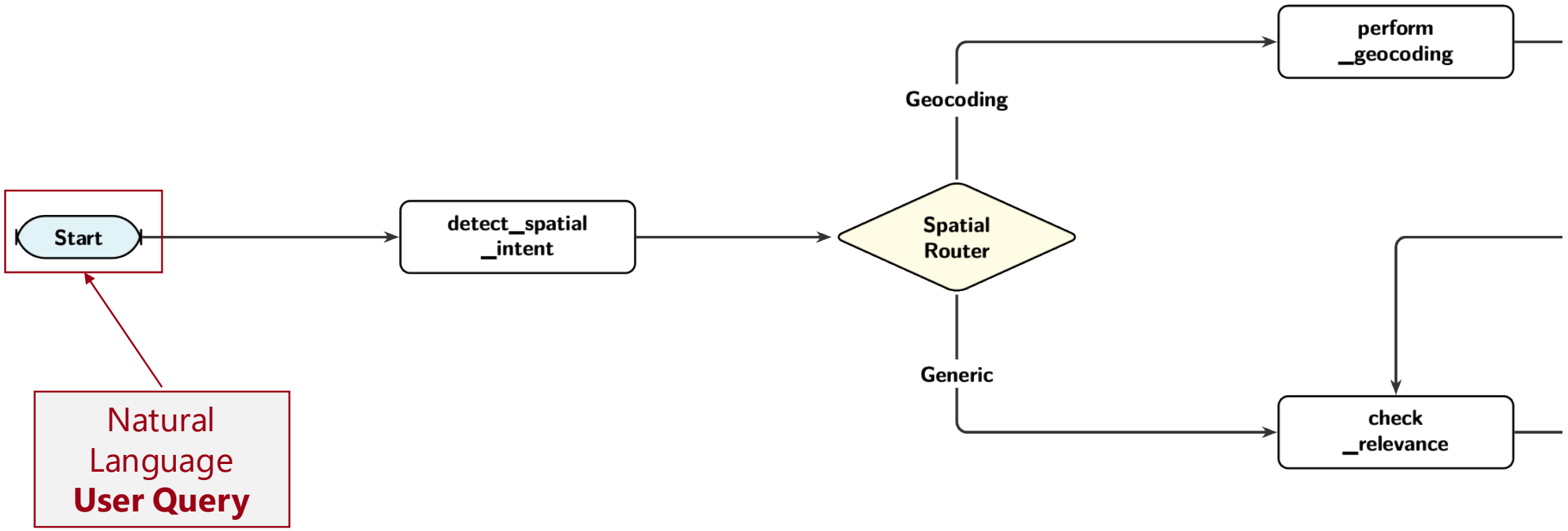
Hallucinated:
`ST_NearBy(geom, 100)`
⇒`ProgrammingError`: **function does not exist**

Key Insight The solution lies in a superior control architecture — shifting from linear chains to stateful, cyclic graph-based workflows

GeoAgent: Full Directed Cyclic Graph Architecture

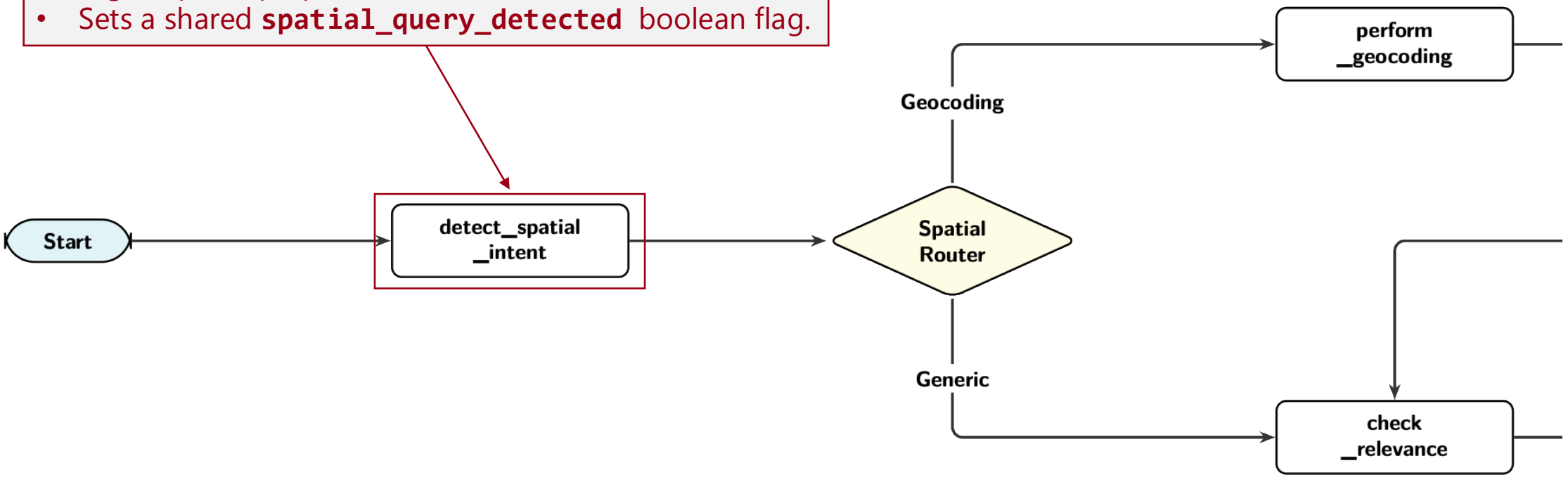


GeoAgent: Full Directed Cyclic Graph Architecture

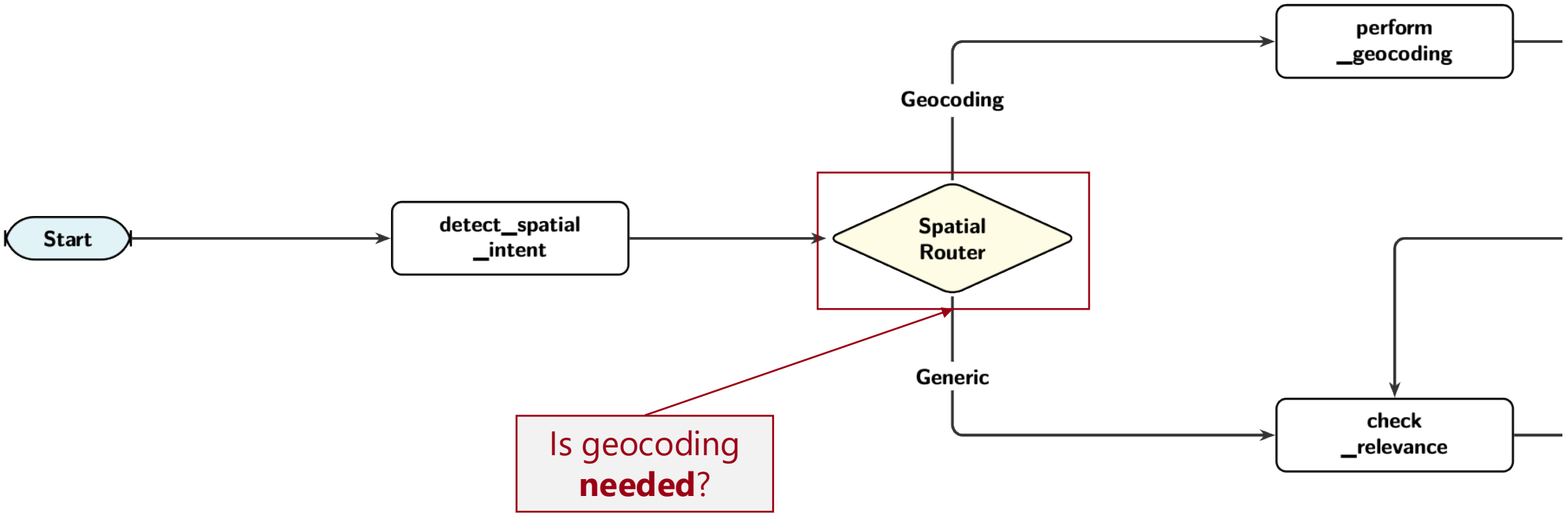


GeoAgent: Full Directed Cyclic Graph Architecture

- Domain-tuned **few-shot prompts** to disambiguate vague spatial prepositions
- Sets a shared **`spatial_query_detected`** boolean flag.

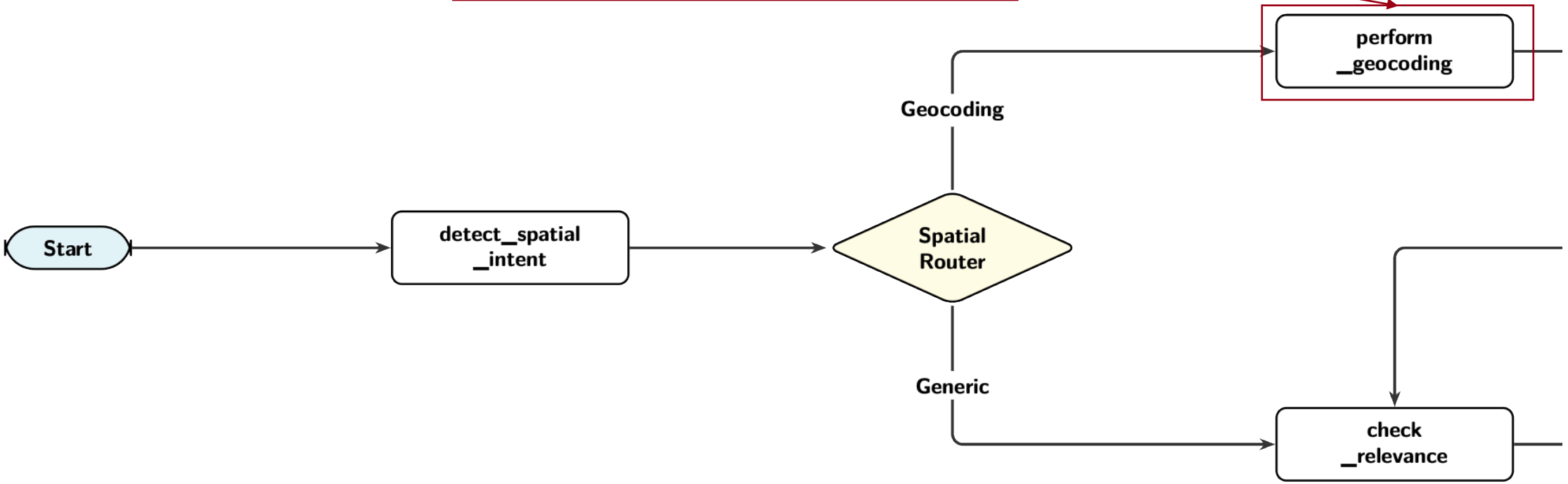


GeoAgent: Full Directed Cyclic Graph Architecture

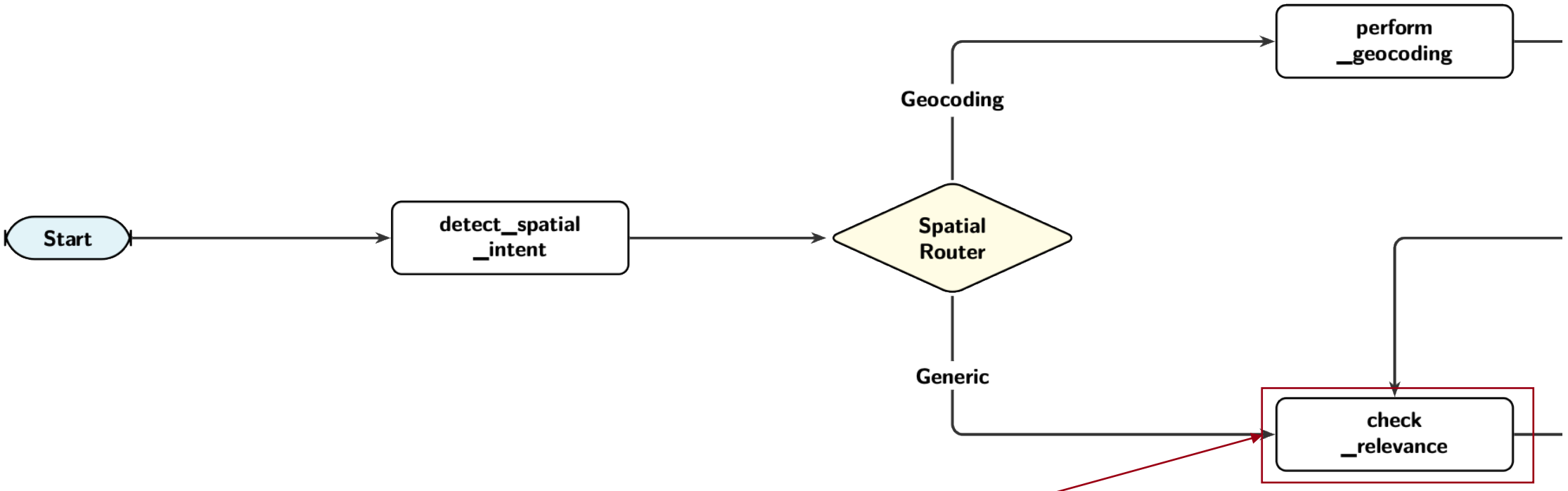


GeoAgent: Full Directed Cyclic Graph Architecture

Transforms toponyms (e.g., “Milan”) into precise coordinates via **OSM Nominatim APIs.**

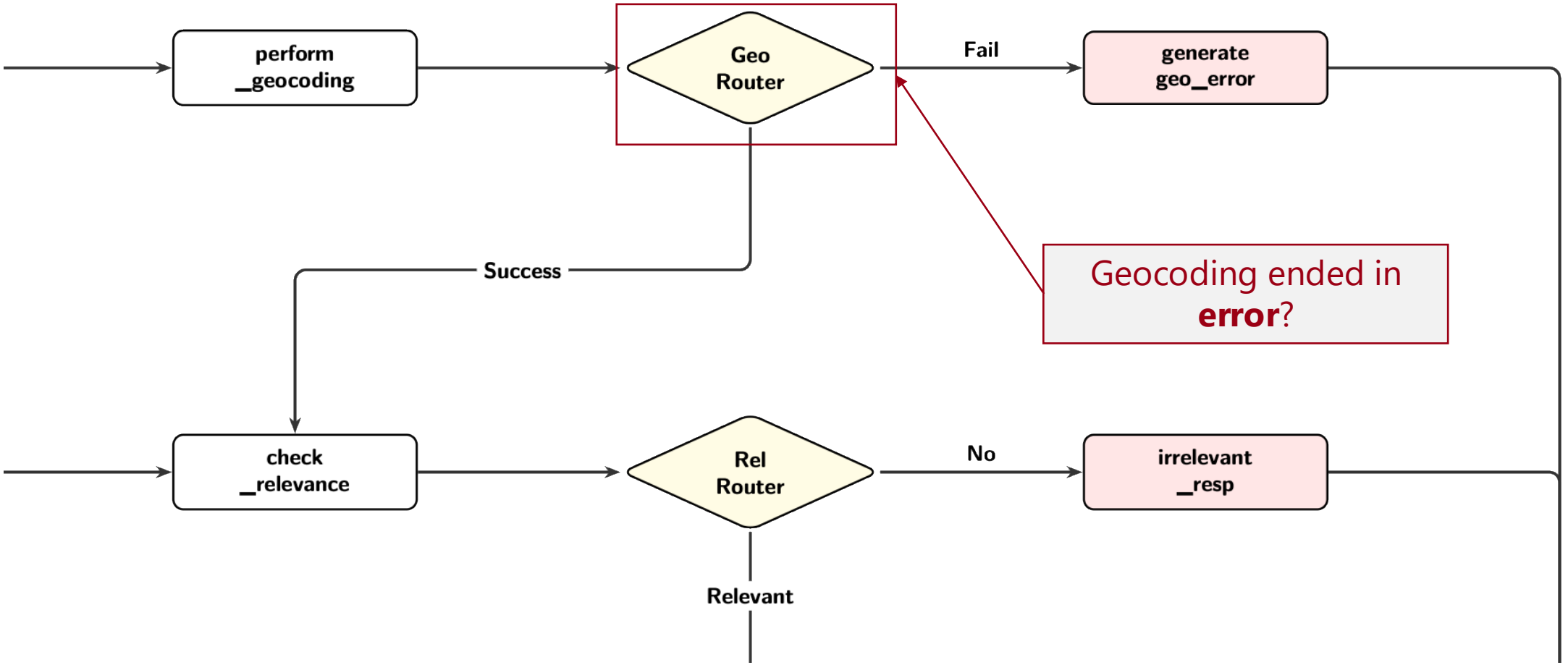


GeoAgent: Full Directed Cyclic Graph Architecture

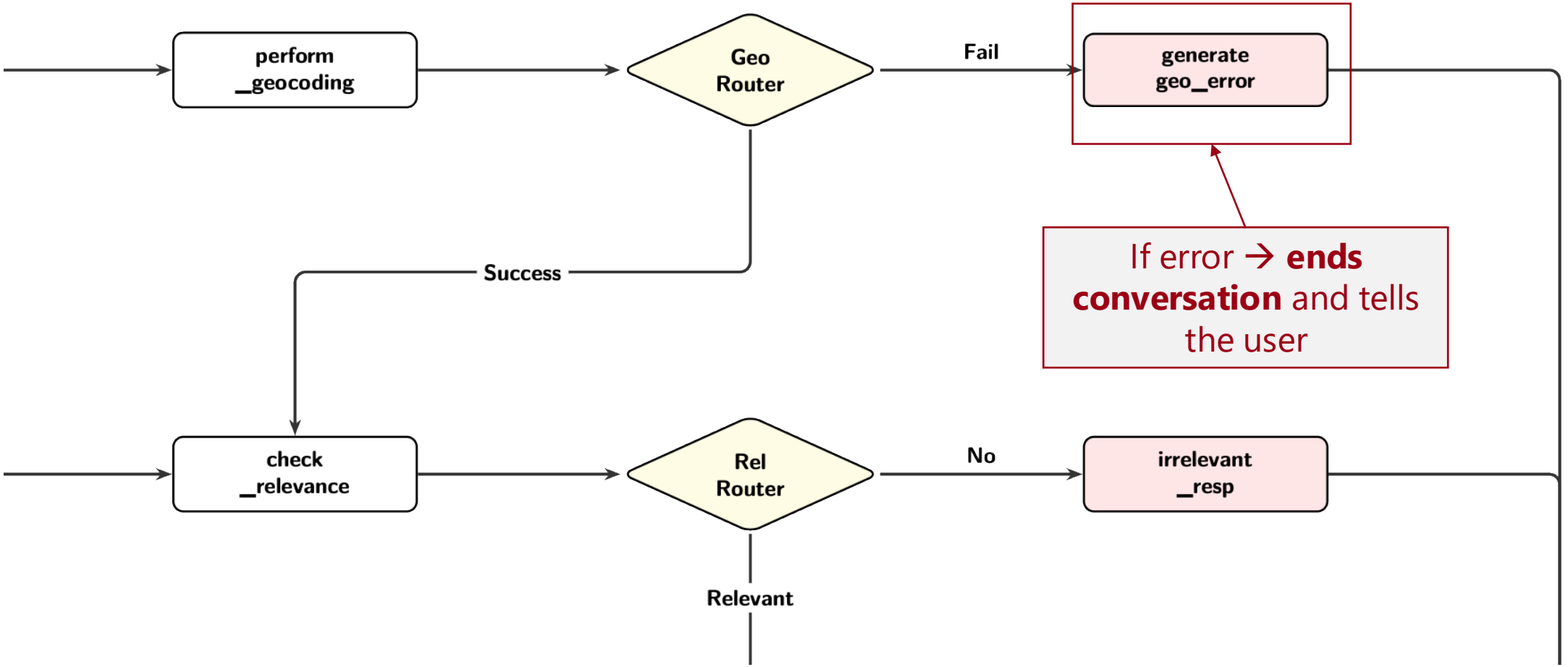


If geocoding is not needed we **check if** the question is **geo-related**

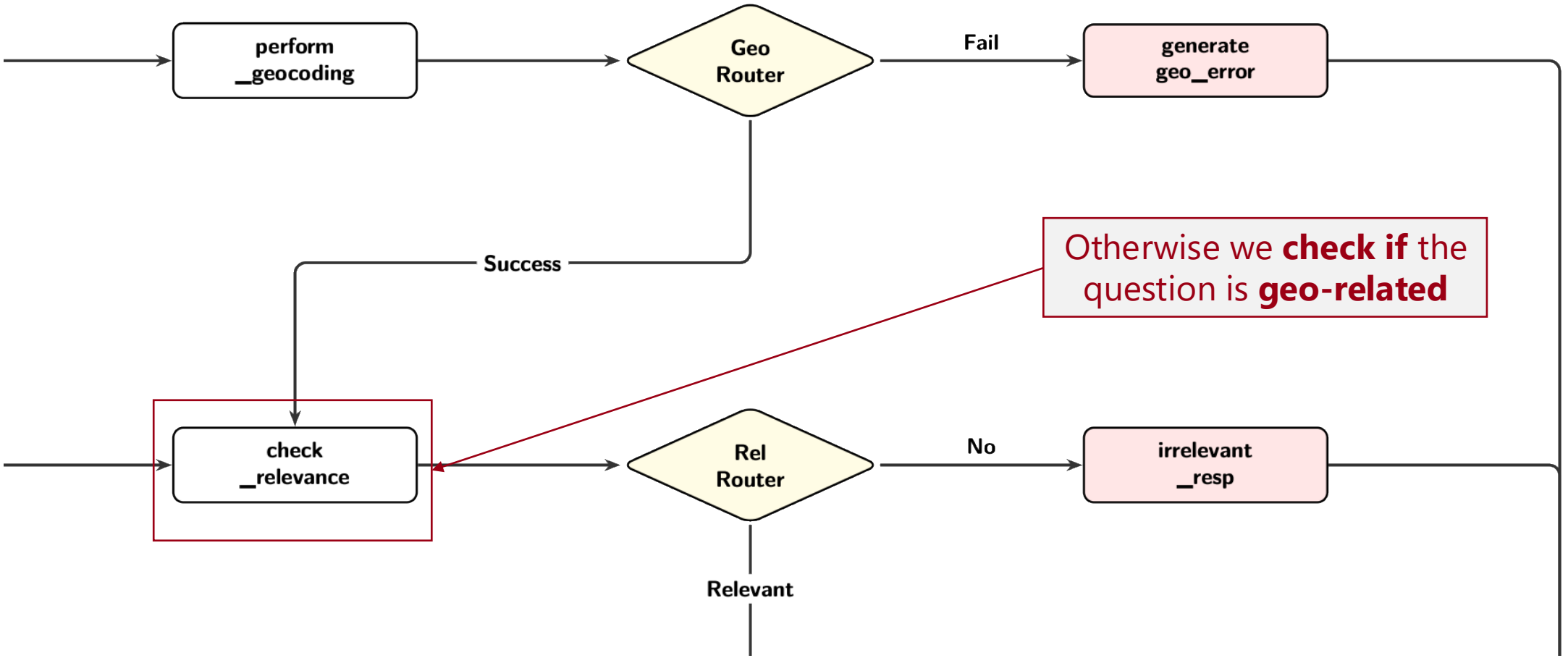
GeoAgent: Full Directed Cyclic Graph Architecture



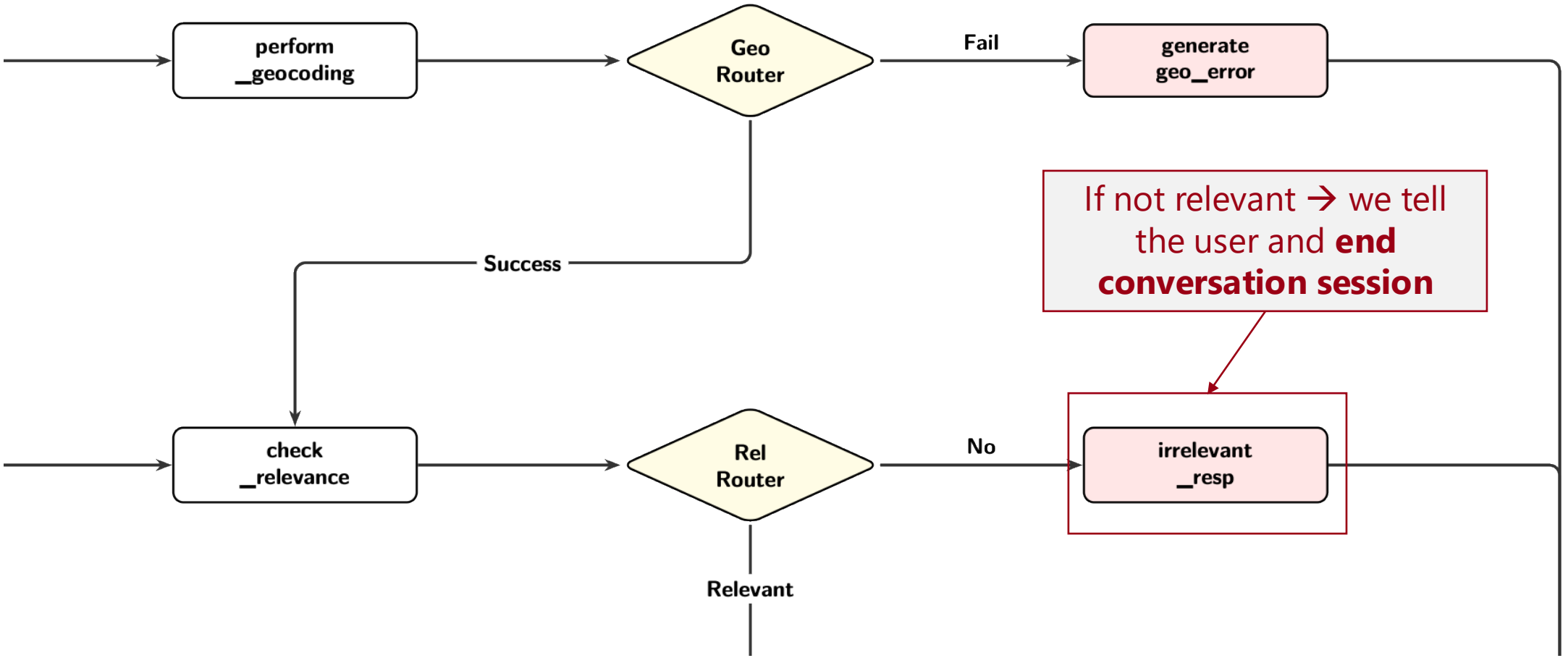
GeoAgent: Full Directed Cyclic Graph Architecture



GeoAgent: Full Directed Cyclic Graph Architecture

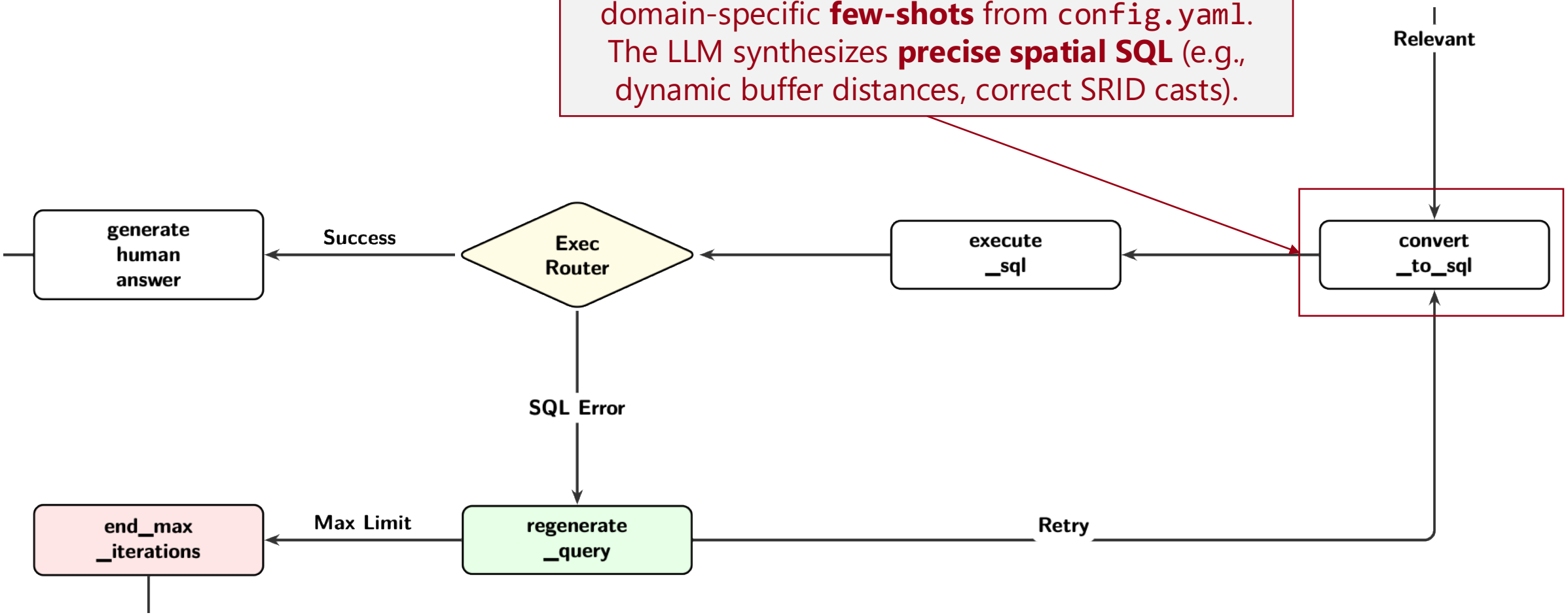


GeoAgent: Full Directed Cyclic Graph Architecture



GeoAgent: Full Directed Cyclic Graph Architecture

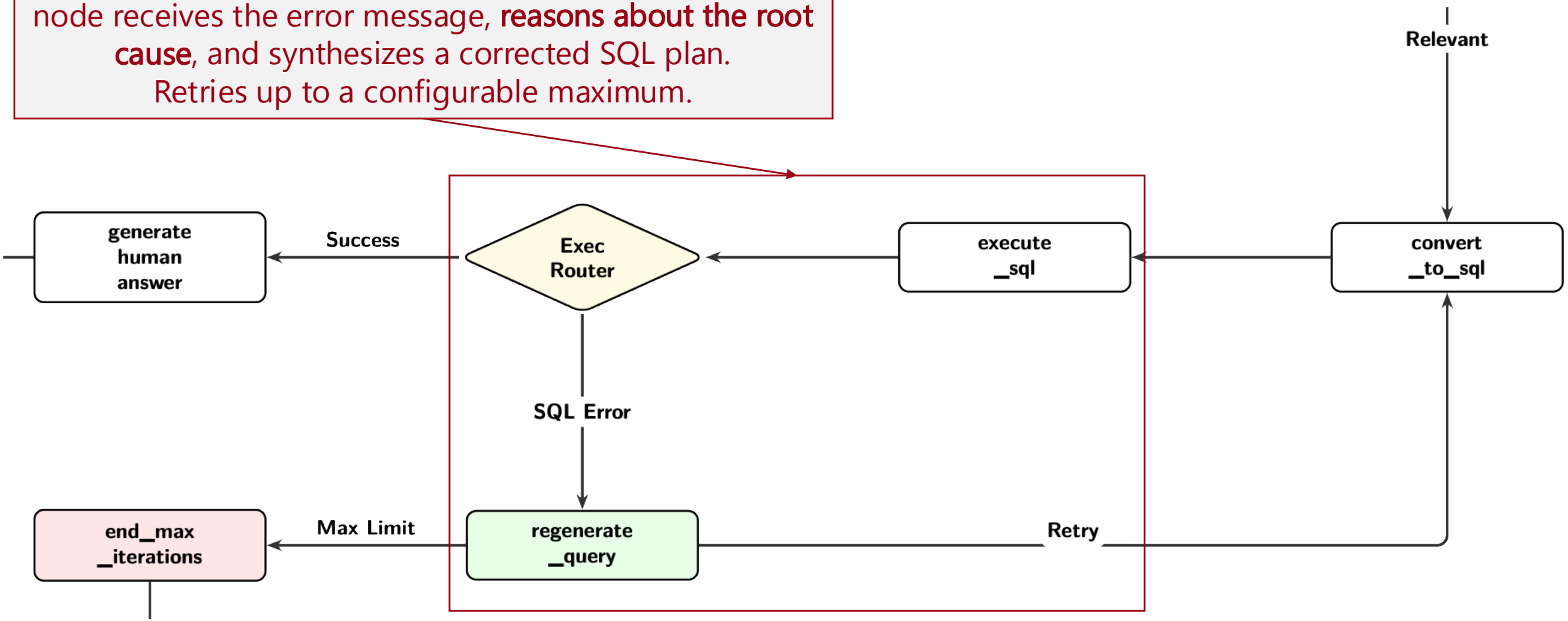
Dynamically injects the **PostGIS schema** and domain-specific **few-shots** from `config.yaml`. The LLM synthesizes **precise spatial SQL** (e.g., dynamic buffer distances, correct SRID casts).



UNIVERSITÀ DI PADOVA
Dipartimento
di Fisica e Astronomia
"Galileo Galilei"

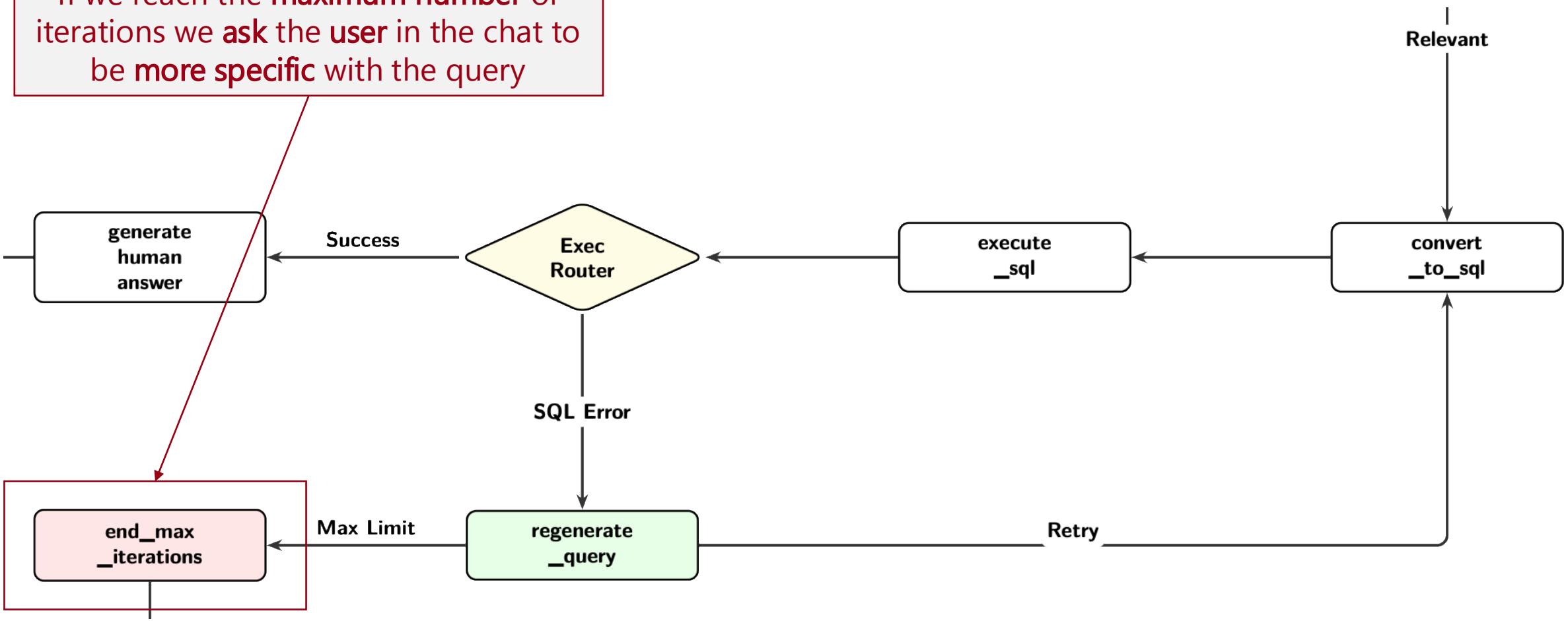
GeoAgent: Full Directed Cyclic Graph Architecture

Self-correcting loop. On any DB exception, the debugger node receives the error message, **reasons about the root cause**, and synthesizes a corrected SQL plan. Retries up to a configurable maximum.

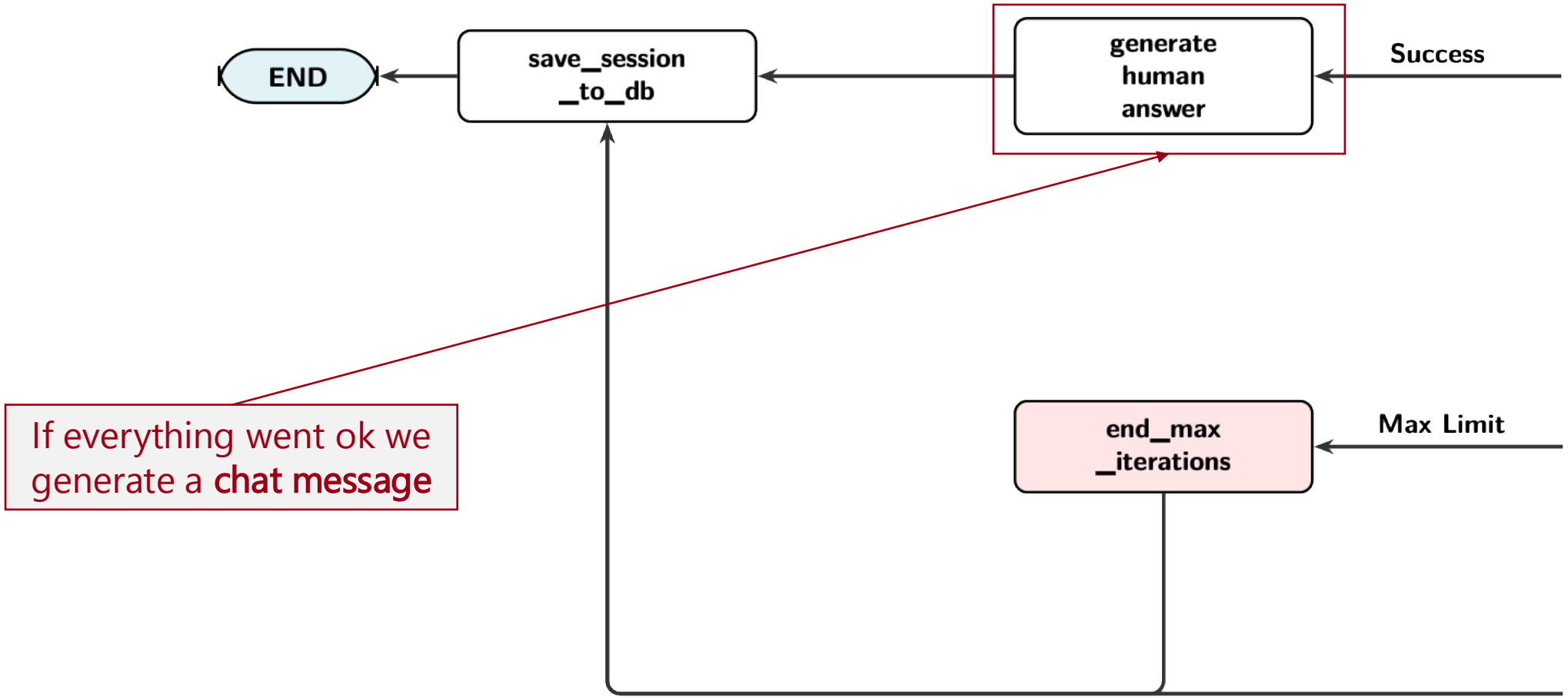


GeoAgent: Full Directed Cyclic Graph Architecture

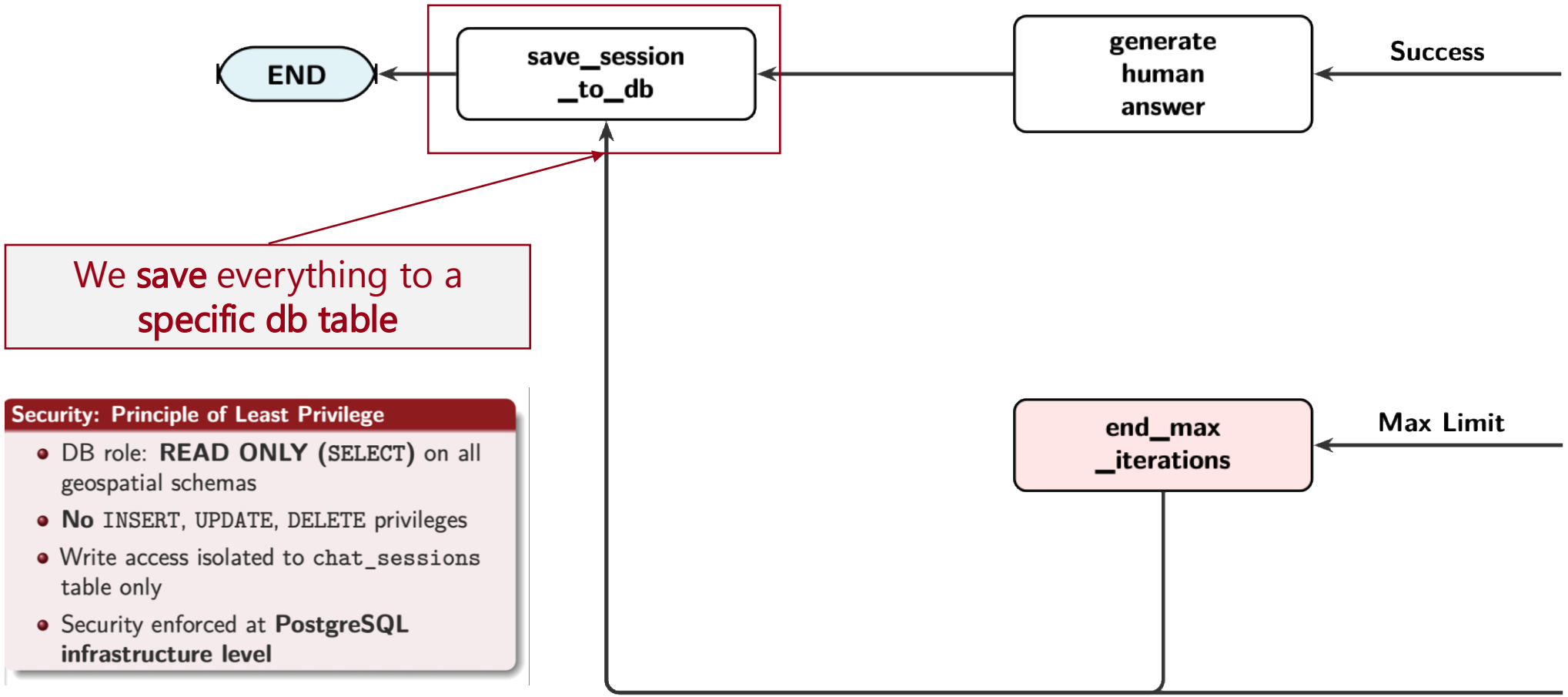
If we reach the **maximum number** of iterations we **ask** the **user** in the chat to be **more specific** with the query



GeoAgent: Full Directed Cyclic Graph Architecture



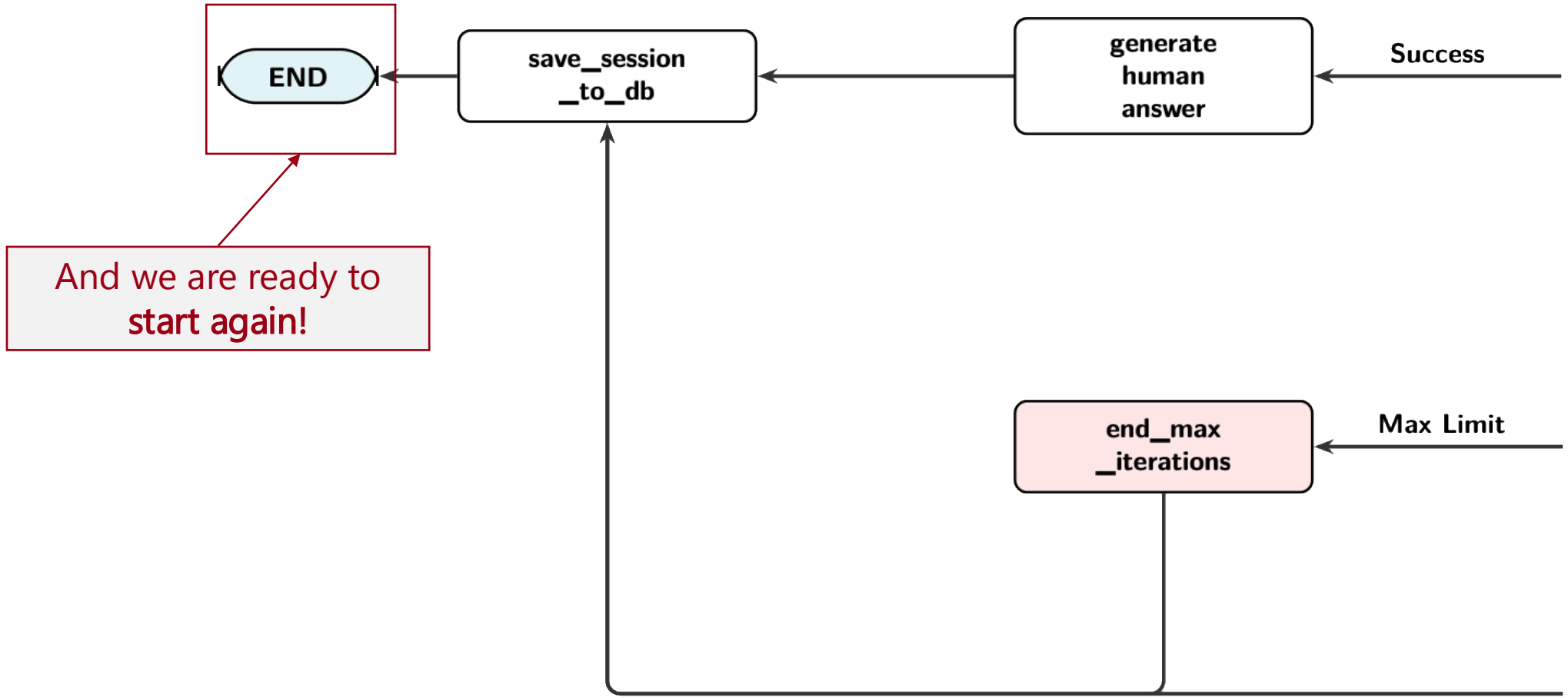
GeoAgent: Full Directed Cyclic Graph Architecture



Security: Principle of Least Privilege

- DB role: **READ ONLY** (`SELECT`) on all geospatial schemas
- **No** `INSERT`, `UPDATE`, `DELETE` privileges
- Write access isolated to `chat_sessions` table only
- Security enforced at **PostgreSQL infrastructure level**

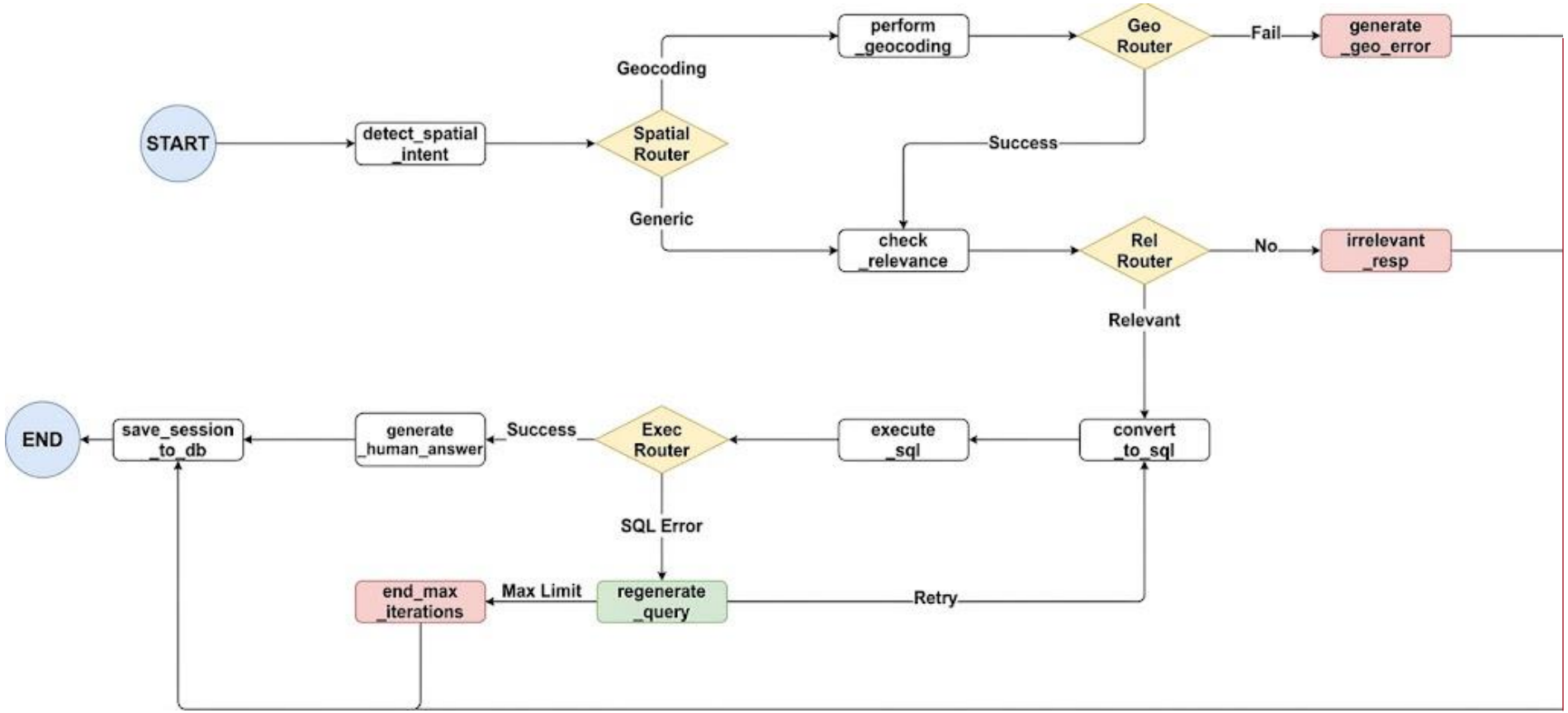
GeoAgent: Full Directed Cyclic Graph Architecture



GeoAgent: Full Directed Cyclic Graph Architecture

Zero-shot schema adaptation

The agent autonomously queries `information_schema` at initialisation ⇒ no fine-tuning required for new schemas.



UNIVERSITÀ DI PADOVA
Dipartimento
di Fisica e Astronomia
"Galileo Galilei"

Experimental Setup

Dataset 1 — EO GeoSQL Query Set

Proprietary PostGIS database

13 questions x 4 complexity tiers on Italian state

1. Spatial Analysis & Validation
2. Spatial-Temporal Filtering
3. Spatial Aggregation
4. KNN Optimization

Dataset 2 — Public Benchmark

`geosql-llm-eval`

Hugging Face [4] 🤗

- 108 queries on Florida state
- US Census TIGER/Line 2023 (7 layers, EPSG:4269)
- Strict zero-shot domain transfer
- Per-query timeout: 60 seconds

UNIVERSITÀ DI PADOVA
Dipartimento
di Fisica e Astronomia
"Galileo Galilei"

4. Naamche. 2024. geosql-llm-eval: A Geospatial Text-to-SQL Evaluation Dataset. https://huggingface.co/datasets/naamche/geosql-llm-eval. Dataset accessed: March 2026.

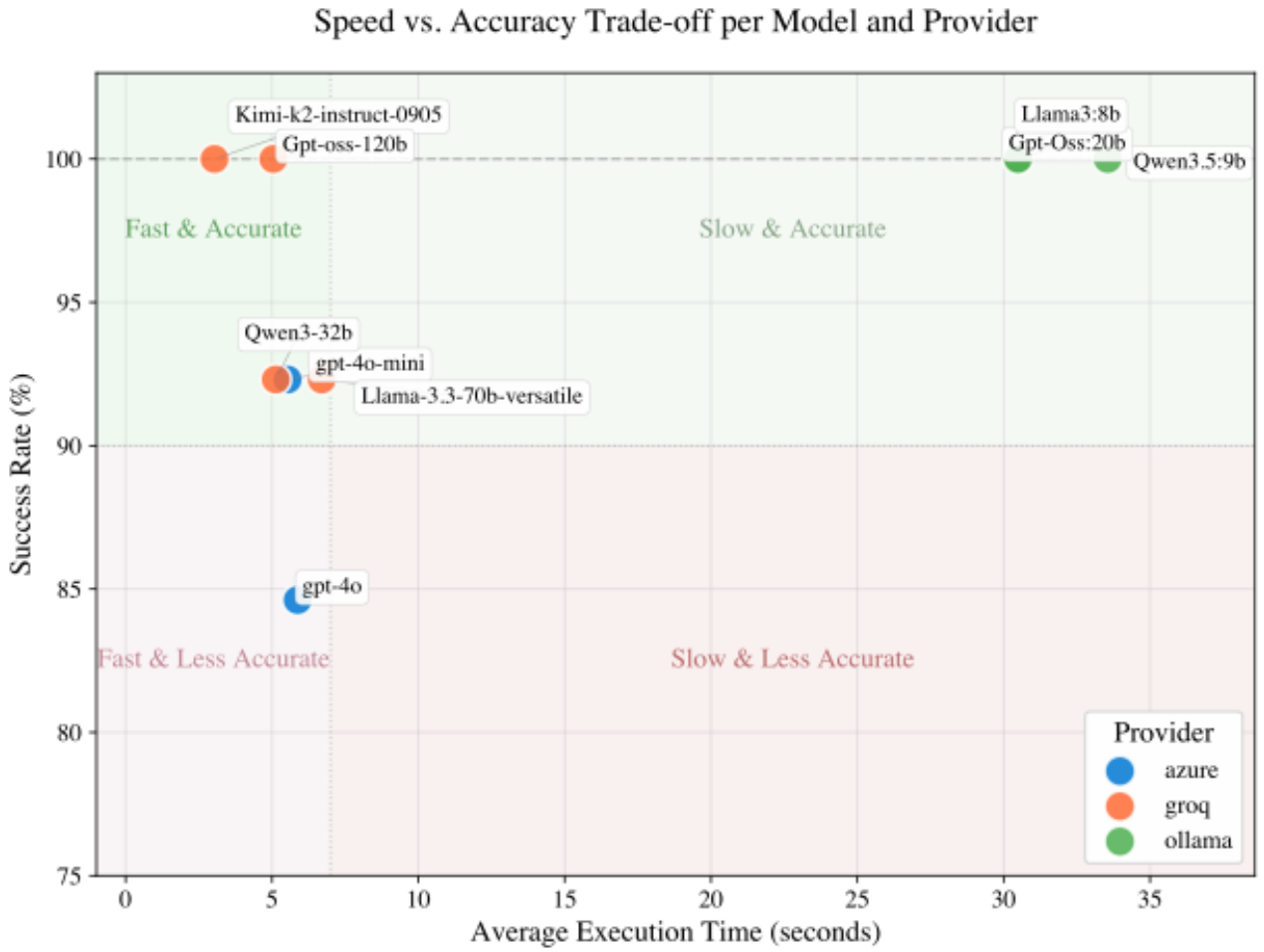
Cross-Model Evaluation: Architectural Robustness on Dataset 1

Model	Success	Avg. Time (s)
Kimi-k2-instruct (Groq)	**100.0%**	3.03
Gpt-oss-120b (Groq)	**100.0%**	5.05
Llama3:8b (Ollama)	**100.0%**	30.48
Gpt-Oss:20b (Ollama)	**100.0%**	30.50
Qwen3.5:9b (Ollama)	**100.0%**	33.55
Qwen3-32b (Groq)	92.3%	5.13
Gpt-4o-mini (Azure)	92.3%	5.53
Llama-3.3-70b (Groq)	92.3%	6.70
Gpt-4o (Azure)	84.6%	5.88

Error Taxonomy (5 failures / 117 total calls)

Error Type	N	%
SRID Mismatch	3	60.0%
Schema Drift	1	20.0%
Semantic Refusal	1	20.0%
Topology Error	0	0.0%

SRID Failures
Models fail to align SRID 0 geometries with SRID 4326 envelopes.



Baseline Comparison: GeoAgent vs. State of the Art on Dataset 1

Overall Execution Accuracy (EA)

System	EA (%)
B1 – LangChain ReAct Agent [5]	0.0
B2 – DIN-SQL (4 steps) [6]	45.5
GeoAgent (ours)	**100.0**

EA by Query Complexity Tier

System	B1 (%)	B2 (%)	GeoAgent (%)
Simple	0	100	**100**
Medium	0	60	**100**
High	0	25	**100**
Very High	0	0	**100**

B1 Structural Failure (0% EA)

LangChain ReAct **cannot perceive PostGIS geometry types or spatial indices** — blind to the spatial domain entirely.

B2 Partial Failure (45.5% EA)

DIN-SQL fails on metric-distance queries: unable to inject **`::geography` casts** required by PostGIS for accurate spatial measurements.

GeoAgent

Perfect EA across all tiers.
Architecture-driven spatial grounding and iterative self-repair make the difference.

UNIVERSITÀ DI PADOVA
Dipartimento
di Fisica e Astronomia
"Galileo Galilei"

5. Yao, Shunyu and Zhao, et al. 2023, ReAct: Synergizing Reasoning and Acting in Language Models, arXiv
6. Pourreza, Mohammadreza, and Davood Rafiei. 2023. "Din-sql: Decomposed in-context learning of text-to-sql with self-correction." Advances in neural information processing systems 36:36339–36348.

Zero-shot Generalization on Public GeoSQL Benchmark (Dataset 2)

- **Original domain**: Italian Change-Detection DB
- **Target domain**: US Census TIGER/Line 2023
- **No fine-tuning**; no dataset-specific prompts
- Relevance classifier **disabled** for this evaluation
- **LLM**: `llama3:8b`

Metric	Value
Agent Success Rate	**93.5% (101/108)**
SQL Execution Rate	**89.8% (97/108)**
Queries with retry (≥ 2)	12.0% (13/108)

Insights

1. zero unhandled exceptions; **robust under domain shift**
2. valid, **executable SQL in~9/10 queries** on an **unseen** schema
3. **self-correction** loop actively **contributes** to final success rate

🌍 GeoAgent - Change Detection Assistant

Ask questions about the PostGIS change detection database

Your question:

E.g.: What change detection categories are available?

Send

Conclusions & Future Work

Main Contributions

1. **GeoAgent**: graph-based, LLM-agnostic **agentic NL2GeoSQL** system
2. **100% EA on EO dataset** across all complexity tiers
3. **93.5% zero-shot success** on public GeoSQL benchmark (**unseen schema & domain**)
4. **Cross-model robustness** : 9 LLMs, 84.6% - 100%

Key Paradigm Shift

- **Architecture dominates** over model scale
- Small, open-weights **local models** ≈ **flagship cloud models**
- **Parameter** count **is not** the **bottleneck**

Future Work

- **RAG integration**: Retrieval-Augmented Generation to handle massive schema scaling beyond context window limits
- **SLM fine-tuning**: Use curated interaction logs to train Small Language Models explicitly optimised for the PostGIS dialect

UNIVERSITÀ DI PADOVA
Dipartimento
di Fisica e Astronomia
"Galileo Galilei"

Conclusions & Future Work

Main Contributions

1. **GeoAgent**: graph-based, LLM-agnostic **agentic NL2GeoSQL** system
2. **100% EA on EO dataset** across all complexity tiers
3. **93.5% zero-shot success** on public GeoSQL benchmark (**unseen schema & domain**)
4. **Cross-model robustness** : 9 LLMs, 84.6% - 100%

Key Paradigm Shift

- **Architecture dominates** over model scale
- Small, open-weights **local models** ≈ **flagship cloud models**
- **Parameter** count **is not** the **bottleneck**

Future Work

- **RAG integration**: Retrieval-Augmented Generation to handle massive schema scaling beyond context window limits
- **SLM fine-tuning**: Use curated interaction logs to train Small Language Models explicitly optimised for the PostGIS dialect

UNIVERSITÀ DI PADOVA
Dipartimento
di Fisica e Astronomia
"Galileo Galilei"

Conclusions & Future Work

Main Contributions

1. **GeoAgent**: graph-based, LLM-agnostic **agentic NL2GeoSQL** system
2. **100% EA on EO dataset** across all complexity tiers
3. **93.5% zero-shot success** on public GeoSQL benchmark (**unseen schema & domain**)
4. **Cross-model robustness** : 9 LLMs, 84.6% - 100%

Key Paradigm Shift

- **Architecture dominates** over model scale
- Small, open-weights **local models** ≈ **flagship cloud models**
- **Parameter** count **is not** the **bottleneck**

Future Work

- **RAG integration**: Retrieval-Augmented Generation to handle massive schema scaling beyond context window limits
- **SLM fine-tuning**: Use curated interaction logs to train Small Language Models explicitly optimised for the PostGIS dialect

UNIVERSITÀ DI PADOVA
Dipartimento
di Fisica e Astronomia
"Galileo Galilei"

Thank you for your attention!

Giovanni Piccolo
giovanni.piccolo.3@phd.unipd.it

UNIVERSITÀ DI PADOVA
Dipartimento
di Fisica e Astronomia
“Galileo Galilei”