

GeoScout: A Conversational Agentic GeoAI Framework for Spatial Decision Support in Methane Leak Detection and Repair

Jinya Wang
University of Calgary
Calgary, Canada
jinya.wang@ucalgary.ca

Mingke Erin Li
University of Calgary
Calgary, Canada
mingke.li@ucalgary.ca

Steve H.L. Liang
University of Calgary
Calgary, Canada
steve.liang@ucalgary.ca

ABSTRACT

Field-inspection programs are inherently spatial because site locations, spatial arrangement, and crew movement shape what can be inspected and how quickly problems can be addressed. Recent studies increasingly combine large language models (LLMs) with simulation systems to lower the technical barrier to configuring and running complex models through natural-language interaction. However, existing work has largely emphasized scenario generation and execution, while result interpretation is rarely grounded in explicit spatial computation, limiting its use for spatial field operations such as methane Leak Detection and Repair (LDAR). We present GeoScout, a conversational agentic GeoAI framework that extends LLM-assisted simulation into spatial decision support. GeoScout configures two simulation engines with different geographic representations from a single plain-language program specification and compares their outcomes. Holding the program fixed while varying only facility layout increases modeled per-transition crew travel roughly fourfold in the distance-aware engine while leaving the geography-agnostic engine unchanged. GeoScout also grounds post-simulation interpretation by routing spatial questions to deterministic spatial (GIS) tools and retaining tool traces so that spatial claims remain traceable to their computations. We apply GeoScout to methane LDAR equivalency under Alberta's Directive 060. These capabilities provide an assumption-transparent and auditable framework for agentic geospatial simulation and spatial decision support.

CCS CONCEPTS

• Information systems → Geographic information systems • Computing methodologies → Modeling and simulation • Applied computing → Environmental sciences

KEYWORDS

spatial simulation, geospatial AI, conversational agent, spatial analysis, leak detection and repair, methane emissions, large language models, regulatory compliance

ACM Reference Format:

Jinya Wang, Mingke Erin Li, and Steve H.L. Liang. 2026. GeoScout: A Conversational Agentic GeoAI Framework for Spatial Decision Support in Methane Leak Detection and Repair. In *The 9th ACM SIGSPATIAL International Workshop on Geospatial Simulation (GeoSim '26)*, November 03–06, 2026, Riverside, CA, USA. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3849737.3849835>

1 Introduction

Field-inspection programs are inherently spatial. How often sites are surveyed, which detection technologies are used, and how crews move between sites all depend on the geography. Site spacing, clustering, and travel distance can shape how much can be surveyed and how quickly detected problems can be addressed. We study methane Leak Detection and Repair (LDAR), in which operators routinely survey oil-and-gas facilities to detect and repair leaks. An operator proposing an alternative LDAR program must demonstrate outcomes equivalent to a prescribed regulatory standard. The stakes are both environmental and economic. Methane is a potent, short-lived greenhouse gas and a major contributor to near-term warming [1], while survey costs vary with inspection frequency and detection technology [2].

Evaluating a candidate LDAR program through simulation is challenging for two reasons. First, simulators differ in how they represent geography. Some derive crew travel from inter-site distance, while others assume a fixed travel time per site, and this modeling choice can affect the resulting compliance assessment. Second, operating simulation tools and interpreting their spatial outputs require expertise that many operators lack, so analyses often rely on specialist consultants and proceed through slow, offline iterations. Recent work uses large language models (LLMs) to lower this barrier in other simulation domains but does not address comparing simulators with different geographic representations or grounding the spatial interpretation of results (Section 2).

We present GeoScout, a conversational agentic geospatial AI (GeoAI) framework that addresses these gaps through a staged workflow. A user describes a program in plain language, and GeoScout interprets the request, plans the comparison, prepares the required inputs, and runs the simulators as tools, then returns a recommendation and a spatial analysis of the completed run.



This work is licensed under a Creative Commons Attribution 4.0 International License. GeoSim '26, Riverside, CA, USA
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-3059-7/2026/11
<https://doi.org/10.1145/3849737.3849835>

GeoScout integrates two simulation engines with contrasting treatments of geography. The spatially explicit GeoLLAMA (Geospatial Leak Lifecycle Analysis and Mitigation Agent) [3] derives crew travel time from inter-site distances, whereas the geography-agnostic LDAR-Sim v4 [4] assumes a fixed travel time per site. Running the same program on both allows users to test whether this difference in geographic representation changes the assessment (Section 4.1).

GeoScout also extends the interaction into spatially grounded post-simulation analysis. Existing conversational simulation interfaces can analyze completed runs through natural language [5], but errors in spatial interpretation may be difficult to identify without explicit geographic validation. GeoScout addresses this limitation through a dedicated result-analysis agent that answers free-form spatial questions by calling deterministic spatial tools rather than inferring spatial relationships directly from coordinate tables (Section 3.3). Each spatial result therefore remains traceable to the computation that produced it.

We make three contributions. First, we present an agentic geospatial simulation framework that runs a single plain-language program across simulation engines with different geographic representations and compares their outcomes. Second, we integrate a tool-grounded post-simulation analysis agent that routes spatial questions to deterministic spatial tools rather than inferring spatial relationships from coordinates, and retains the corresponding tool traces, keeping spatial results auditable. Third, we apply GeoScout to methane LDAR equivalency under Alberta's Directive 060 [6]; to our knowledge, it is the first framework to couple cross-engine geographic comparison with tool-grounded, auditable spatial analysis for this regulated equivalency assessment.

2 Related Work

Recent systems use LLMs to configure and operate simulations through natural-language interaction. AgentSUMO plans and executes traffic simulations in SUMO and supports interactive analysis of simulation results [5], while related systems generate simulation scenarios from user prompts [7]. GeoScout differs by coordinating simulation engines with different geographic representations and grounding spatial analysis of both the study area and the completed simulation results in deterministic spatial tools, rather than relying on LLM inference. This approach builds on spatial decision support systems that integrate geographic data, analytical methods, and simulation to inform spatially distributed decision-making [8, 9], as well as autonomous GIS agents that compose deterministic spatial operations from natural-language requests [10, 11]. GeoScout adopts this separation between language-based interpretation and tool-based computation within an end-to-end simulation workflow.

For methane Leak Detection and Repair (LDAR), established simulators include LDAR-Sim [4] and FEAST [12], while GAMA supports spatially explicit agent-based simulation across application domains [13, 14]. Our earlier GeoLLAMA model, a spatially explicit LDAR simulator, was developed on GAMA (GIS Agent-based Modeling Architecture) [3]. GeoScout integrates

GeoLLAMA and LDAR-Sim as simulation tools, compares their different treatments of geography, and carries their outputs into spatially grounded decision support. GeoScout's contribution lies in integrating established conversational-simulation and deterministic spatial tool-grounding methods into an end-to-end, auditable workflow for spatially explicit LDAR decision support.

3 Methodology

GeoScout supports the full workflow from an initial recommendation request to post-simulation result analysis (Figure 1). The framework follows a staged workflow in which a planner interprets the request and defines the analysis, a simulation executor runs the simulators as tools, a synthesizer evaluates the outputs against the regulatory profile and generates a recommendation, and a dedicated tool-calling agent answers follow-up questions about the completed simulation. Throughout the workflow, deterministic tools compute all spatial quantities, while the language model interprets their outputs, keeping the resulting spatial information verifiable. GeoScout is implemented as a local Python and Streamlit application. Its workflow is orchestrated with LangGraph, and the LLM components use GPT-4o at temperature 0 through OpenAI function calling.

Users interact with GeoScout through a plain-language request or structured form describing the program to be evaluated, for example, quarterly optical gas imaging (OGI) surveys of all sites by two crews. GeoScout requests any missing parameters or files, then presents a complete plan for user confirmation before simulation. After the run, the system returns a spatial characterization of the fleet and compliance metrics (emission reduction, average repair delay, and cost-effectiveness), together with an exportable report and tool-grounded answers to free-form questions about the completed simulation.

3.1 Simulation and regulatory assessment

The executor runs both engines using matched program parameters, so that each engine evaluates the same LDAR program under its own modeling assumptions. The synthesizer then scores these outputs against Directive 060 [6]: an alternative program must achieve at least the same emission reduction as the standard program, and every detected leak must be repaired within 30 days. The assessment is compiled into an equivalency report that includes the recommendation and supporting metrics presented as tables and figures. GeoScout verifies each reported value against the computed results before finalizing and exporting the report to support regulatory review.

3.2 Deterministic spatial toolset

GeoScout uses deterministic spatial tools to compute geographic quantities for fleet characterization and post-simulation analysis. Table 1 summarizes the tools, their inputs, and outputs. Clustering uses DBSCAN, allowing the number of clusters to emerge from the site geography rather than being fixed in advance.

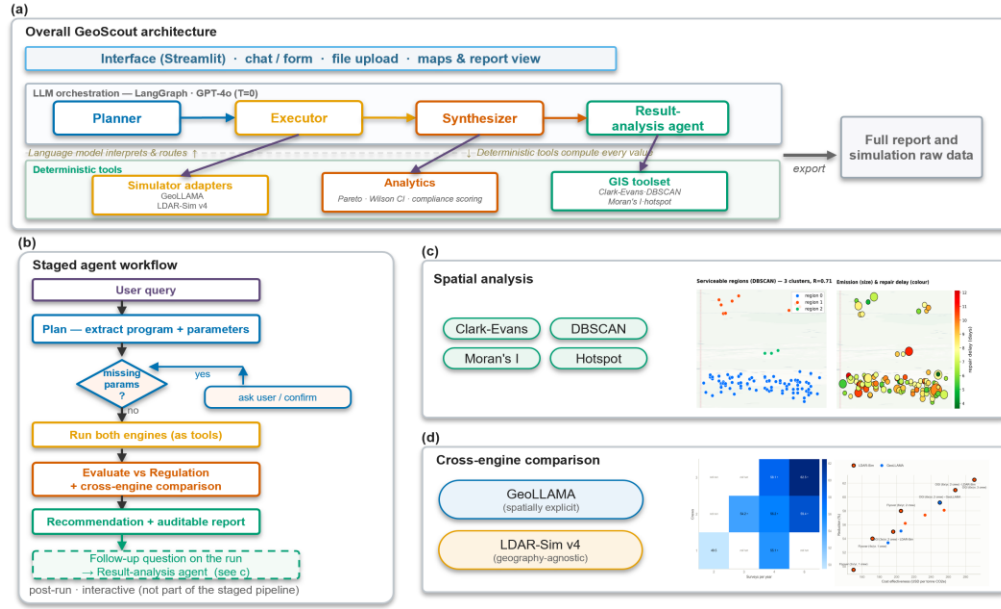


Figure 1: Overview of GeoScout. (a) overall architecture, (b) staged agent workflow, (c) spatial analysis, and (d) cross-engine comparison.

The default adjacency radius is 25 km, corresponding to the straight-line distance represented by LDAR-Sim's 30-minute inter-site travel assumption at rural driving speed, and can be adjusted to local conditions. Moran's I uses eight-nearest-neighbour row-standardized weights. These settings remain fixed across questions rather than being tuned for individual queries.

Table 1: Deterministic spatial tools used by GeoScout.

Tool	Input	Output
Point-pattern index	Coordinates	Clark-Evans index and pattern class
Clustering	Coordinates	DBSCAN clusters, per-cluster spread flagged
Emission profile	Sites, emission history	Emission by site and cluster, coverage, top-site share
Repair-window risk	Sites, delays	Per-cluster exposure to the 30-day window
Spatial autocorrelation	Sites, an attribute	Moran's I (8-NN), whether high values cluster

3.3 Post-simulation result-analysis agent

After a simulation run, users can ask follow-up questions about the results, such as which cluster emits the most, whether slow-to-repair sites are spatially concentrated, or how the fleet's sites are spatially clustered. GeoScout handles these queries through a dedicated result-analysis agent that operates on the completed run's results table and site geography, separate from the interface used to initiate a new simulation.

The result-analysis agent uses two tool families, one for retrieving compliance metrics from the results table and one for spatial analysis using the tools in Table 1. The language model selects the appropriate tool and can chain multiple tools when required, but it does not infer spatial relationships directly from

coordinates. Each response follows four rules. (1) Numerical values are obtained from tools. (2) Data coverage is disclosed. (3) Dominant single-site effects are identified. (4) Questions beyond the available tools, such as crew routing or weather, are identified as unsupported. Every response also retains the trace of the tools invoked (Figure 2).

Ask about these results

- Where do emissions concentrate across these sites
- Emissions are concentrated primarily in Cluster 0, which contains 87 sites and accounts for a total emission of 32,906.36 units (mean 378.23 per site). They are spread across several sites, with no single site dominating: the top-emitting site, ABRAND00095, contributes 2,044.04 units, about 6.2% of the cluster's total. Cluster 1 (10 sites) totals 4,993.7 units (mean 499.37) and Cluster 2 (3 sites) totals 1,963.97 units (highest mean, 654.66). Overall the site distribution is clustered, with a Clark-Evans R index of 0.71.

Grounded in: spatial tools

How I got this — 3 tool call(s)

```
1. cluster_sites ()
  "method": "dbscan", "n_clusters": 3, "n_noise": 0
  "eps_km": 25, "min_samples": 3, "most_dispersed_cluster_id": 0
  "how_decided": "DBSCAN at eps=25 km; a cluster is a contiguous serviceable region, not one crew-day; the count emerged from the geography ..."
  "cross_check_silhouette_best_k": 2 (consistent with 3)
2. describe_distribution ()
  "n_sites": 180, "mean_nearest_neighbour_km": 3.62
  "clark_evans_R": 0.71, "pattern": "clustered"
  "extent_bbox_diagonal_km": 381.0, "convex_hull_area_km2": 10322.6
3. emission_profile ("by_cluster": True, "top_n": 5)
  "unit": "leak volume (model units, summed per site)"
  "total": 39864.03, "hotspot_cluster_id": 0
  "top_sites": [...], "coverage": {...}, "by_cluster": {...}
  "hotspot_concentration": {
    "cluster_total": 32906.36, "n_sites_in_cluster": 87
    "dominant_site": "ABRAND00095", "dominant_site_share": 0.062
    "single_site_dominated": False
  }
```

Figure 2: The result-analysis agent answering a spatial question from deterministic tools, with the tool trace shown.

4 Case Study and Results

The case study uses real facility locations from an Alberta operating region, while emissions and repair outcomes are simulated from empirically derived inputs rather than replayed from a specific LDAR campaign; the results therefore represent a simulation-based case study, not a one-to-one reconstruction of observed operations.

Validation against campaign-level operational records remains future work. For the 100-site Alberta fleet, GeoScout identifies a clustered spatial pattern with a Clark-Evans index of 0.71 and groups the sites into three clusters at the 25-km serviceable radius, with no isolated outliers (Figure 3). One cluster contains 87 of the 100 sites and accounts for more than 80% of the fleet's emissions. This concentration provides operationally relevant context for prioritizing limited inspection capacity, translating spatial characterization into an operationally relevant prioritization.

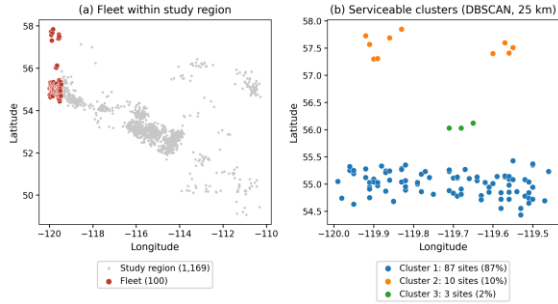


Figure 3: Spatial characterization of the Alberta fleet. (a) The 100-site Alberta facility fleet (red) within the 1,169-facility study region. (b) The three serviceable clusters identified by GeoScout with DBSCAN at a 25-km adjacency radius.

4.1 Cross-engine spatial comparison

The two engines represent crew travel differently. Holding the program and leak characteristics fixed while varying only facility layout, we increased the fleet's mean nearest-neighbour spacing from the observed 3.6 km to about 43 km, a roughly twelvefold increase, with ten runs at each spacing level. In the distance-aware engine, mean per-transition crew travel increases from about 10 to 40 minutes, while LDAR-Sim v4 remains unchanged because it assumes a fixed per-site travel time. Because the engines differ in implementation beyond geography, we do not attribute differences in their absolute outputs solely to geographic representation; the sweep instead tests how each engine responds to the same change in facility spacing.

4.2 Result-analysis agent evaluation

Direct interpretation of coordinate tables by the language model is unreliable for spatial questions. When asked to classify the fleet's spatial pattern and rank its clusters using coordinates alone, GPT-4o answered correctly in only 5 of 45 cases (11%).

Table 2: Result-analysis agent routing and refusal (three site sets, seven questions each, GPT-4o at temperature 0).

Question category	Expected behavior	Correct
Spatial (where questions)	route to a spatial tool	9 / 9
Data (compliance numbers)	route to results table	6 / 6
Out-of-scope (route, weather)	decline	6 / 6
All		21 / 21

We evaluated the result-analysis agent using seven questions for each of three site sets, covering spatial, data, and out-of-scope queries (Table 2). The agent routed all supported questions to the

appropriate tools and correctly identified all out-of-scope queries. Each question was evaluated once using GPT-4o at temperature 0 to reduce sampling variability. This validates tool routing and refusal behavior rather than the accuracy of the tools themselves or general LLM spatial reasoning; robustness across models, larger question sets, and multi-turn interactions remains future work.

5 Conclusion and Future Work

We presented GeoScout, a conversational agentic GeoAI framework that carries an LDAR program from plain-language specification through simulation, cross-engine comparison, and interactive analysis of completed results. GeoScout grounds spatial claims in deterministic spatial tools and compares engines with different geographic representations, making both the reported results and the influence of modeling assumptions traceable. Applied to methane Leak Detection and Repair equivalency under Alberta's Directive 060, the framework provides an auditable basis for simulation-supported spatial decision-making. Future work will extend this approach and develop memory that learns transferable spatial patterns across campaigns. A separate prototype that uses an LLM to order inspections and prioritize repairs based on spatial context provides an initial step in this direction. Although demonstrated for methane LDAR, GeoScout represents a broader agentic geosimulation approach for spatial field operations whose outcomes must be interpreted and acted upon geographically.

REFERENCES

- [1] International Energy Agency. Global Methane Tracker 2025. IEA, Paris, 2025.
- [2] C.E. Kemp and A.P. Ravikumar. New technologies can cost-effectively reduce oil and gas methane emissions. *Environmental Science & Technology*, 55(11):7660–7668, 2021.
- [3] J. Wang and S.H.L. Liang. LLM-Driven ABM Framework for Cost-Effective LDAR Program Design. In *Spatial Knowledge and Information Canada (SKI Canada 2026)*, Banff, AB, Canada, 2026.
- [4] T.A. Fox, T.E. Barchyn, D. Risk, A.P. Ravikumar, and C.H. Hugenholtz. An agent-based model for estimating emissions reductions equivalent from leak detection and repair programs. *Journal of Cleaner Production*, 282:125237, 2021.
- [5] M. Jeong, J. Chang, and Y. Yoon. Speak to Simulate: An LLM-Guided Agentic Framework for Traffic Simulation in SUMO. In *Proc. 8th ACM SIGSPATIAL Int'l Workshop on Geospatial Simulation (GeoSim '25)*. DOI: 10.1145/3764921.3770151.
- [6] Alberta Energy Regulator. Directive 060: Upstream Petroleum Industry Flaring, Incinerating, and Venting. 2023.
- [7] S. Li, T. Azfar, and R. Ke. ChatSUMO: Large Language Model for Automating Traffic Scenario Generation in Simulation of Urban Mobility. *arXiv:2409.09040*, 2024.
- [8] P.J. Densham. Spatial decision support systems. In *Geographical Information Systems: Principles and Applications*, pages 403–412. Longman, London, 1991.
- [9] R. Sugumaran and J. DeGroot. *Spatial Decision Support Systems: Principles and Practices*. CRC Press, Boca Raton, FL, 2010.
- [10] Z. Li and H. Ning. Autonomous GIS: the next-generation AI-powered GIS. *International Journal of Digital Earth*, 16(2):4668–4686, 2023.
- [11] T. Akinboyewa, Z. Li, H. Ning, and M.N. Lessani. GIS Copilot: towards an autonomous GIS agent for spatial analysis. *International Journal of Digital Earth*, 18(1):2497489, 2025.
- [12] C.E. Kemp, A.P. Ravikumar, and A.R. Brandt. Comparing Natural Gas Leakage Detection Technologies Using an Open-Source Virtual Gas Field Simulator. *Environmental Science & Technology*, 50(8):4546–4553, 2016.
- [13] P. Taillandier, B. Gaudou, A. Grignard, Q.-N. Huynh, N. Marilleau, P. Caillou, D. Philippon, and A. Drogoul. Building, composing and experimenting complex spatial models with the GAMA platform. *GeoInformatica*, 23(2):299–322, 2019.
- [14] T.-D. Vu, B. Gaudou, and K.S. Oberoi. Modeling realistic human behavior using generative agents in a multimodal transport system. *arXiv:2510.19497*, 2025.