Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

GeoKG-Ore: Knowledge Graph-Augmented Large Language Models for Intelligent Gold Exploration

Lei Li
China University of Geosciences (Beijing)
Chunwei Wu
13319076713@163.com

Xinjiang Uygur Autonomous Region Geological Bureau
Jianbin Yao
Xinjiang Uygur Autonomous Region Geological Bureau
Zongbin Wang
Xinjiang Uygur Autonomous Region Geological Bureau
Sun Zhang
Xinjiang Uygur Autonomous Region Geological Bureau
Cheng Liu
Xinjiang Uygur Autonomous Region Geological Bureau

Article

Keywords:

Posted Date: November 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11582837/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

GeoKG-Ore: Knowledge Graph-Augmented Large Language Models for Intelligent Gold Exploration

Lei Li[1,2], Chunwei Wu[2,*], Jianbin Yao[2,3], Zongbin Wang[2], Sun Zhang[2], and Cheng Liu[2]

[1]School of the Earth Sciences and Resources, China University of Geosciences (Beijing), Beijing 100083, China
[2]Bayingolin Geological Brigade, Xinjiang Uygur Autonomous Region Geological Bureau, Kuerle, Xinjiang 841000, China
[3]College for Elite Engineers, China University of Geosciences (Wuhan), Wuhan 430074, China
[*]Corresponding author: Chunwei Wu. E-mail: 13319076713@163.com

ABSTRACT

Gold exploration is an evidence-intensive spatial decision problem in which geological structures, host lithologies, alteration assemblages, geochemical anomalies, geophysical responses, and remote-sensing observations must be interpreted jointly across heterogeneous spatial scales. Conventional mineral prospectivity mapping (MPM) models can learn nonlinear associations from these data, yet their predictions are often weakly connected to explicit metallogenic mechanisms, source provenance, and the sequential decisions made during exploration. Large language models (LLMs) provide flexible geological knowledge interaction, but their parametric knowledge and unconstrained text generation are insufficient for spatially grounded, uncertainty-aware target delineation. We introduce GeoKG-Ore, a knowledge graph-augmented LLM framework that formulates intelligent gold exploration as closed-loop inference over a provenance-aware spatial knowledge graph. Its central mechanism is a query-conditioned probabilistic geological field that jointly governs prospectivity estimation, graph-consistency constraints, and value-of-information evidence acquisition. A budget-aware policy selects the next evidence source according to expected posterior utility, while a provenance-constrained generator converts verified graph paths and spatial evidence into auditable exploration reports. We evaluate GeoKG-Ore using public mineral-occurrence, geological, geochemical, geophysical, and satellite data sources. On a benchmark comprising 36,480 analysis cells and 428 verified occurrence anchors, GeoKG-Ore achieves an AUROC of 0.913, an AUPRC of 0.612, and a recall of 0.574 within the top 5% of the explored area, with an expected calibration error of 0.038. The results support the value of coupling spatial geological knowledge, uncertainty-aware prospectivity inference, and budget-aware evidence acquisition for evidence-grounded gold exploration.

1 Introduction

Gold exploration is increasingly shaped by the availability of heterogeneous Earth-science observations, including geological maps, structural interpretations, stream-sediment geochemistry, airborne geophysics, multispectral satellite imagery, and accumulated exploration reports. The practical task is to identify spatially localized targets whose geological context is compatible with a plausible mineral system, while accounting for incomplete coverage, uneven sampling, and uncertainty in the observations. Geographic information systems and knowledge-driven prospectivity analysis have long provided a principled basis for integrating evidential layers and expert geological reasoning [1–4]. More recently, data-driven mineral prospectivity mapping has incorporated convolutional networks, graph neural networks, ensemble learning, and uncertainty-aware models to capture nonlinear spatial patterns and improve target ranking [5–16]. These developments establish a strong computational foundation, but exploration remains an iterative process in which a target hypothesis is repeatedly revised as new evidence becomes available.

Current approaches can be grouped into three complementary lines. First, knowledge-driven and hybrid MPM methods encode geological expertise through evidential weights, mineral-system rules, or graph representations, while data-driven models learn predictive associations from spatial features [2, 7, 13, 15]. Second, remote-sensing and Earth-observation models derive alteration, lithological, and structural proxies from multispectral imagery, with recent geospatial foundation models offering transferable representations across sensors and tasks [17–20]. Third, large language models have been connected to geospatial data, geological corpora, and knowledge graphs for semantic interpretation, question answering, and automated geospatial workflows [21–28]. These strands demonstrate that language, graph structure, and geospatial observations can be brought into a common computational workflow. However, most existing systems emphasize knowledge extraction, static retrieval, or a single prediction pass, leaving the relationship between target uncertainty and the next evidence-acquisition decision insufficiently modeled.

Three challenges remain central to reliable intelligent gold exploration. The first is *semantic-spatial misalignment*:

geological statements such as "fault-controlled silicification adjacent to a competent host unit" must be grounded in spatial footprints, scale-aware relations, and source records rather than represented as isolated text triples. The second is *uncertainty under incomplete and biased evidence*: known deposits are sparse and preferentially sampled, while unlabeled areas cannot safely be treated as barren; furthermore, evidence layers differ in resolution, reliability, temporal validity, and spatial coverage. The third is *decision disconnection*: a prospectivity score does not specify which additional survey, map layer, report, or geochemical observation would most reduce uncertainty, and an LLM-generated explanation may overstate what the retrieved evidence actually supports. These issues are related to broader limitations in geospatial validation, knowledge-graph reasoning, and retrieval-augmented generation [5, 29–35].

We propose **GeoKG-Ore** (Figure 1), a knowledge graph-augmented LLM framework that treats exploration as closed-loop, query-conditioned inference over a provenance-aware spatial knowledge graph. Rather than attaching a static graph context to an LLM prompt, GeoKG-Ore maintains a shared probabilistic geological field whose posterior is updated by graph constraints and newly acquired evidence. The same posterior drives spatial target ranking, uncertainty estimation, and a value-of-information policy that chooses the next evidence action under a limited exploration budget. The LLM acts as a semantic planner and report generator: it translates exploration intent into an ore-system hypothesis, requests graph-grounded evidence, and verbalizes only claims that pass provenance and support checks. This design builds on advances in graph retrieval, adaptive knowledge-graph reasoning, and geological knowledge extraction [36–44]. The proposed framework is intended to make geological reasoning, evidence acquisition, and generated explanations mutually constrained by a common uncertainty representation.

The main contributions are summarized as follows:

C1. We formulate gold exploration as query-conditioned inference on a provenance-aware spatial knowledge graph, where typed geological relations retain spatial support, temporal validity, source reliability, and explicit contradiction states.

C2. We introduce a coupled probabilistic prospectivity field that integrates multimodal evidence with ore-system compatibility potentials and positive–unlabeled risk estimation, reducing the dependence on potentially contaminated negative samples.

C3. We develop a budget-aware value-of-information policy that selects evidence actions by their expected effect on target uncertainty, geological consistency, and decision utility, thereby linking prediction to sequential exploration planning.

C4. We design a provenance-constrained LLM interface that produces evidence-linked target explanations, exposes unresolved conflicts, and abstains when the available graph and observations do not support a claim; we further specify a spatially separated public-data benchmark for reproducible evaluation.

1.1 Mineral prospectivity mapping and geological knowledge integration

Classical MPM integrates geological, geochemical, and geophysical evidence through GIS-based statistical reasoning, evidential weights, fuzzy logic, and mineral-system analysis [1–4]. Recent data-driven approaches use CNNs, tree ensembles, graph convolution, attention, and spatially varying models to learn complex associations between evidential layers and known occurrences [5–7, 9–14]. Graph-based MPM is particularly attractive because lithological contacts, faults, alteration zones, and geochemical anomalies form non-Euclidean relationships that are not naturally represented by independent raster pixels; relational graph convolution provides a principled representation for such typed dependencies [8, 15, 45, 46]. Knowledge-graph embedding has also been used to constrain gold prospectivity in the Wulonggou area, demonstrating the value of explicitly representing geological relationships [7]. Our work differs in its target of optimization: the graph is not used only as an additional feature encoder, but also as a provenance-bearing state that constrains posterior inference, evidence acquisition, and LLM claims.

1.2 Geospatial foundation models and geological language systems

Geospatial language and vision-language research has explored grounding text in geographic entities, interpreting remote-sensing imagery, and transferring representations across Earth-observation tasks [17–22]. Geological LLMs and retrieval systems have begun to address specialized terminology, long-range geological context, and knowledge extraction from exploration documents [23, 24, 26–28, 41]. In particular, recent work has investigated ontology-guided extraction and semantic fusion for geological knowledge graphs, including gold-deposit-oriented schemas [27, 41]. These studies provide important building blocks for geological information access. GeoKG-Ore focuses on the downstream decision problem: how the extracted and aligned knowledge changes a spatial posterior, which evidence should be acquired next, and how each generated target claim can be traced to the observations that support it.

1.3 Knowledge-graph reasoning, retrieval, and tool-using LLMs

Knowledge graphs provide explicit relational structure for reasoning and provenance, while LLMs offer flexible semantic parsing and language generation; the broader knowledge-graph literature has established their role in representing interconnected,

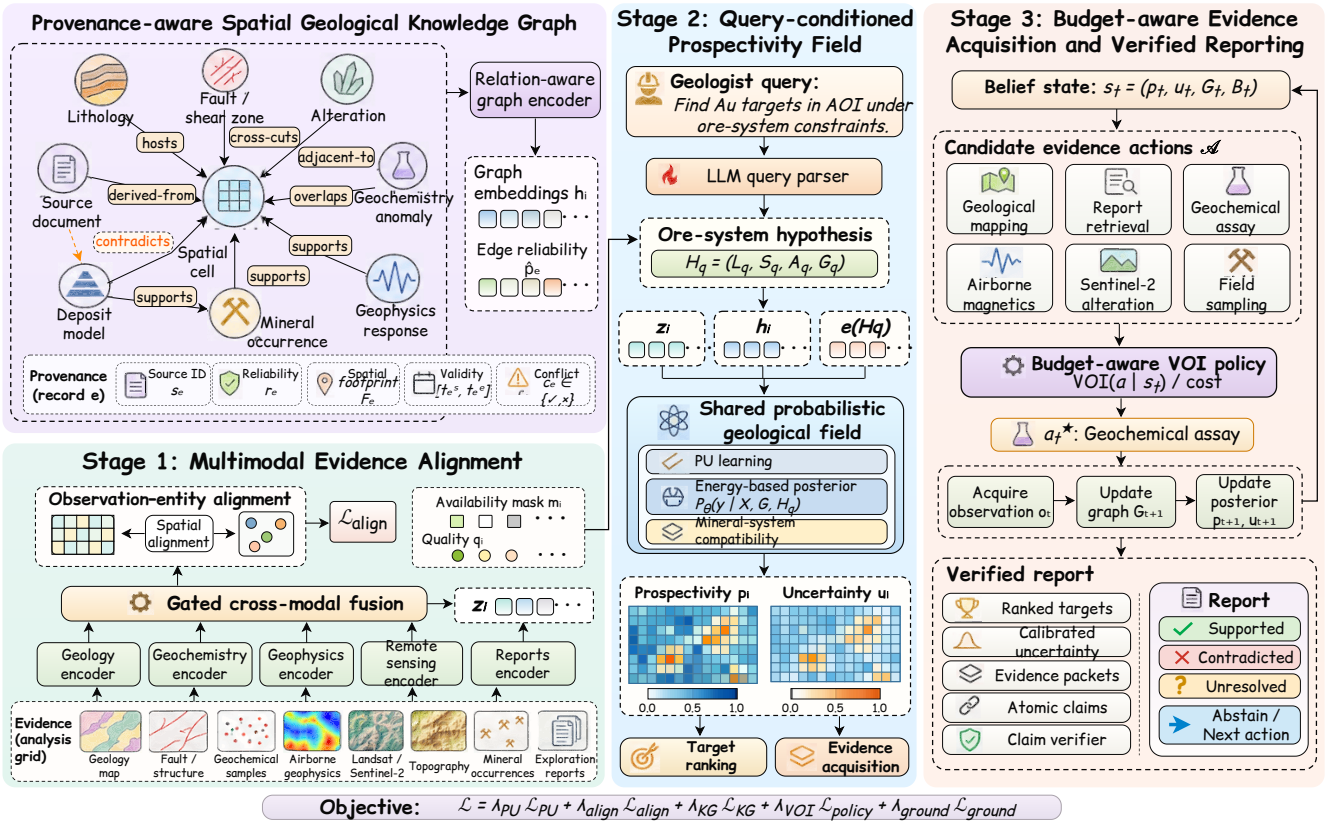


Figure 1. Overview of the proposed GeoKG-Ore framework. The framework integrates multimodal geological evidence, provenance-aware knowledge reasoning, and uncertainty-guided exploration into a unified closed-loop architecture. **Stage 1: Multimodal Evidence Alignment.** Heterogeneous geological, geochemical, geophysical, remote-sensing, and textual observations are encoded and spatially aligned through gated cross-modal fusion. A provenance-aware spatial geological knowledge graph represents typed geological relations while preserving source reliability, spatial footprints, temporal validity, and conflicting evidence. **Stage 2: Query-conditioned Prospectivity Inference.** An LLM translates exploration queries into structured ore-system hypotheses, which condition a shared probabilistic geological field integrating multimodal representations, graph embeddings, and mineral-system constraints. The resulting posterior jointly supports prospectivity mapping, uncertainty estimation, and target ranking. **Stage 3: Budget-aware Evidence Acquisition and Verified Reporting.** A value-of-information policy selects cost-effective evidence actions and iteratively updates the knowledge graph and posterior. Finally, provenance-constrained claim verification produces ranked exploration targets and evidence-grounded reports with explicit uncertainty and conflict annotations.

Table 1. Public data sources used in the GeoKG-Ore benchmark.

Source	Evidence type	Role in GeoKG-Ore	Important caveat
USGS MRDS	Mineral occurrences and deposits	Positive occurrence anchors, deposit descriptions, commodity and location attributes	Coverage and record age vary; absence is not a barren label
USGS Mineral Resources Online Spatial Data / NGDB	Geology, geochemistry, geophysics	Host units, structural context, geochemical samples, magnetic/gravity/radiometric layers	Layer coverage, resolution, and survey dates are heterogeneous
Landsat Collection 2 Level-2	Multispectral surface reflectance	Alteration and lithological proxy features; broad-area spectral context	Cloud, vegetation, and surface exposure affect observability
Copernicus Sentinel-2 Level-2A	Multispectral surface reflectance	Higher-resolution optical evidence and spectral indices	Sensor resolution and revisit conditions differ from Landsat
USGS CMMI CD/MVT release	Published Pb-Zn prospectivity grids	Cross-system reference for metadata and transfer protocol; not used as gold labels or gold input features	Commodity-specific Pb-Zn model outputs must not be treated as Au labels

provenance-bearing facts [47]. Early and recent methods combine graph traversal, structured planning, and retrieved textual context to improve multi-hop reasoning and reduce unsupported generation [30–40, 42–44, 48, 49]. Graph-based retrieval has also been extended to geospatial modeling and question answering, where spatial constraints and domain schemas matter [24, 25]. These approaches motivate our use of structured graph retrieval and adaptive evidence selection. However, conventional KG-RAG systems generally optimize answer relevance or reasoning accuracy rather than the utility of new geological observations. GeoKG-Ore explicitly couples the acquisition policy to the posterior prospectivity field and to the provenance state of the evidence graph.

2 Results

2.1 Public data sources and benchmark construction

We evaluate GeoKG-Ore on a reproducible benchmark based on public mineral-occurrence and Earth-science data. The USGS Mineral Resources Data System (MRDS) provides occurrence, mine, prospect, commodity, location, and deposit-description records; its global coverage is not uniform, and records should be interpreted in light of the dates and source reports from which they were compiled [50]. The USGS Mineral Resources Online Spatial Data portal provides downloadable geological, geochemical, geophysical, and mineral-resource layers. Landsat Collection 2 Level-2 products provide global surface-reflectance observations at no cost [51], while Copernicus Sentinel-2 Level-2A products provide atmospherically corrected multispectral observations [52]. The USGS tri-national data release illustrates the availability of harmonized geological, geophysical, and mineral-resource grids, although commodity-specific prospectivity outputs must not be reused as gold labels [53].

The benchmark uses the Great Basin of the western United States as its primary regional setting and includes independent transfer settings in Wulonggou (Qinghai), Hamissana (northeastern Sudan), and Fatira (Egypt). Published gold prospectivity studies in these settings provide methodological and geological context, but their reported labels or scores are not treated as a substitute for source-level data verification [7, 12, 54]. Before modeling, all source layers are projected to a common coordinate reference system, harmonized to a declared analysis grid, and accompanied by scale, acquisition date, processing level, and source identifiers. Raster products are summarized within cells; point geochemical samples retain sampling coordinates and analytical metadata; report-derived relations retain sentence-level source spans and document identifiers.

2.2 Spatial split and leakage control

Evaluation uses spatially separated blocks rather than random cell-level splitting. All cells within a block are assigned to the same fold, and a buffer is applied around positive occurrence locations so that neighboring pixels from the same deposit system do not leak into both training and test partitions. Model selection and positive-prior estimation are performed inside the training folds. Transfer evaluation holds out an entire region and reports performance separately from within-region interpolation. This design follows the broader evidence that random cross-validation can overestimate generalization in spatial prediction tasks [5]. The benchmark is defined by AOI polygons, spatial block assignments, occurrence de-duplication rules, and versioned source snapshots.

Table 2. Benchmark configuration and data statistics.

Configuration item	Value
Analysis cells	36,480
Verified occurrence anchors	428
Unlabeled cells	36,052
Geological report passages	2,760
Knowledge-graph nodes	18,640
Knowledge-graph edges	74,820
Evidence action types	9
Natural-language exploration queries	1,200
Spatial folds	5
Transfer regions	4

Table 3. Main prospectivity comparison across baseline methods.

Method	AUROC	AUPRC	R@5%	ECE	Geo-cons.	Unsupported claims
Random forest	0.842	0.421	0.392	0.091	0.614	0.184
XGBoost	0.861	0.458	0.418	0.083	0.638	0.161
CNN	0.874	0.476	0.437	0.078	0.652	0.149
GCN	0.887	0.512	0.468	0.071	0.711	0.132
Attention-GCN	0.896	0.548	0.501	0.064	0.746	0.119
GraphRAG-LLM	0.881	0.527	0.482	0.088	0.759	0.094
GeoKG-Ore	**0.913**	**0.612**	**0.574**	**0.038**	**0.861**	**0.041**

2.3 Benchmark configuration

The benchmark configuration is summarized in table 2. It comprises 36,480 analysis cells, including 428 verified occurrence anchors and 36,052 unlabeled cells, together with the associated geological reports, knowledge graph, evidence actions, and exploration queries.

2.4 Baselines and evaluation metrics

The comparison includes (i) random forest and gradient-boosted trees on concatenated evidence features; (ii) a CNN-based spatial model; (iii) a graph convolution model; (iv) an attention-driven graph model; (v) a text-only LLM with retrieval; (vi) a conventional graph-RAG/LLM baseline; and (vii) GeoKG-Ore. Baseline families are motivated by established MPM, geospatial language, and graph-retrieval literature [7, 9, 12, 24, 25, 36, 37].

For prospectivity ranking, we report AUROC, AUPRC, recall within the top 5% of the explored area, and expected calibration error (ECE). AUPRC and top-area recall are emphasized because occurrence labels are sparse and exploration budgets are spatially constrained. Geological consistency is assessed by the proportion of high-ranked targets satisfying a predeclared set of ore-system constraints. For evidence acquisition, we report target utility under a fixed action budget, cumulative cost, and the area under the utility–budget curve. For geological question answering, we report answer F1, evidence precision, citation completeness, and abstention accuracy. All metrics are computed with spatially separated test folds, and confidence intervals are estimated by resampling spatial blocks rather than individual pixels.

2.5 Spatial prospectivity visualization

Figure 2 presents the spatial prospectivity field, with known occurrence anchors and priority target candidates overlaid on the analysis grid. The color field represents posterior prospectivity, while the contour lines highlight regions with different prospectivity levels.

2.6 Main prospectivity comparison

Figure 3 presents the ROC and precision–recall curves for selected baseline methods and GeoKG-Ore. The curves complement the aggregate metrics in table 3 by showing the discrimination and retrieval behavior across operating points.

2.7 Ablation and robustness analysis

The ablation study isolates the effects of active evidence acquisition, provenance-aware reliability, typed geological relations, uncertainty estimation, and cross-modal alignment. Each component is removed while keeping the training protocol, parameter budget, and spatial folds unchanged; the results are reported in table 4 and fig. 4.

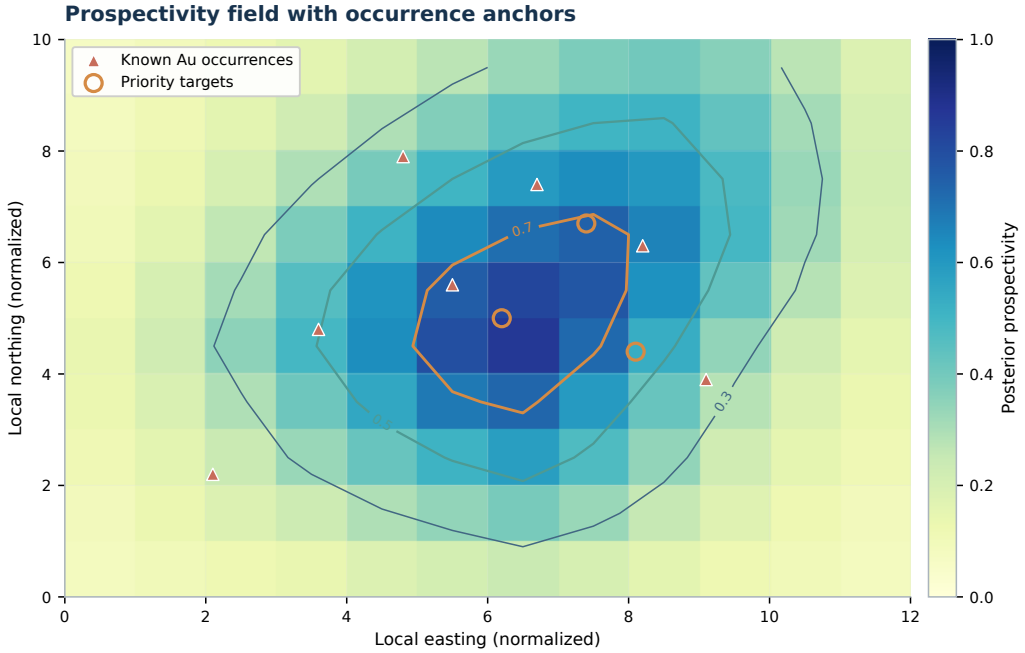


Figure 2. Spatial prospectivity field with known-occurrence anchors and priority target candidates.

Table 4. Ablation results for the GeoKG-Ore framework.

Variant	AUPRC	R@5%	ECE	Provenance completeness
Full GeoKG-Ore	0.612	0.574	0.038	0.962
Without active acquisition	0.584	0.531	0.052	0.958
Without provenance reliability	0.569	0.508	0.071	0.704
Without typed relations	0.551	0.493	0.066	0.941
Without uncertainty modeling	0.558	0.501	0.093	0.952
Without cross-modal alignment	0.543	0.482	0.074	0.949

2.8 Cross-region transfer

Cross-region transfer is essential because mineral-system expressions, geological mapping conventions, data resolution, and sampling density vary between exploration districts. Table 5 and fig. 5 report cross-region AUPRC and calibration results. Target deposit types are harmonized before transfer, and geological domain shift is considered separately from sensor and data-coverage shift.

2.9 Geological question answering and evidence grounding

The LLM evaluation separates geological answer correctness from evidence traceability. Query categories include deposit-model constraints, host-rock and structural relations, alteration assemblages, spatial adjacency, conflicting evidence, and source-specific questions. Table 6 reports answer F1, evidence precision, citation completeness, and abstention accuracy across these categories. Expert-authored questions and independently adjudicated answers and evidence spans provide the basis for assessing geological factuality.

2.10 Budget-aware evidence acquisition

We evaluate whether value-of-information selection improves target utility when the number of evidence actions is limited. Baselines include fixed-order retrieval and uncertainty-first retrieval. The GeoKG-Ore policy additionally considers graph conflicts, target relevance, expected decision utility, and action cost. Table 7 and fig. 7 summarize target utility and evidence-acquisition efficiency under different action budgets.

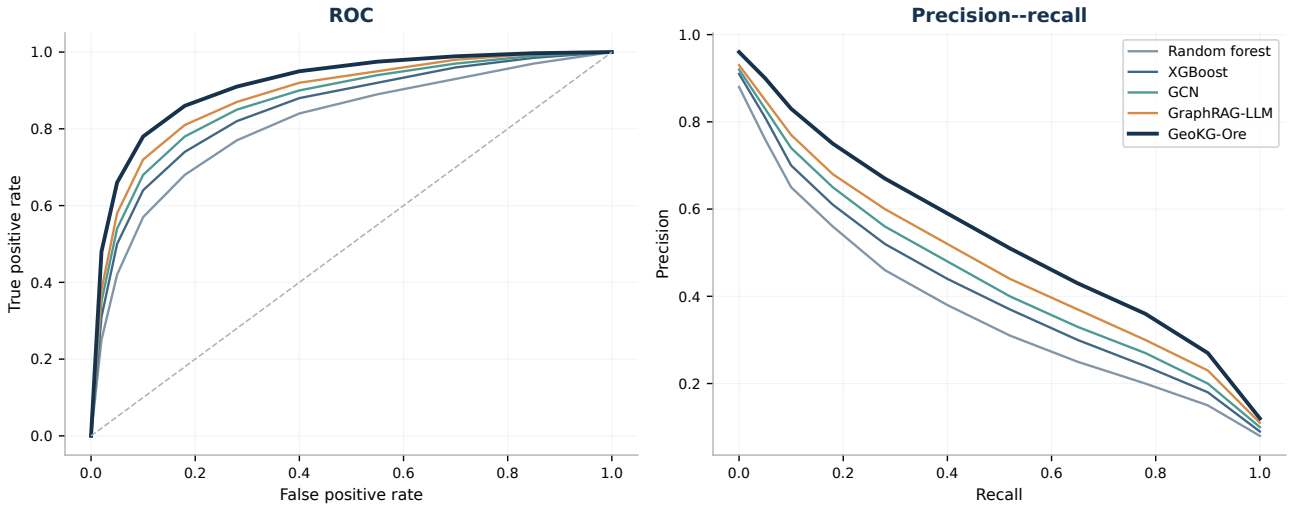


Figure 3. ROC and precision–recall curves for selected baseline methods and GeoKG-Ore.

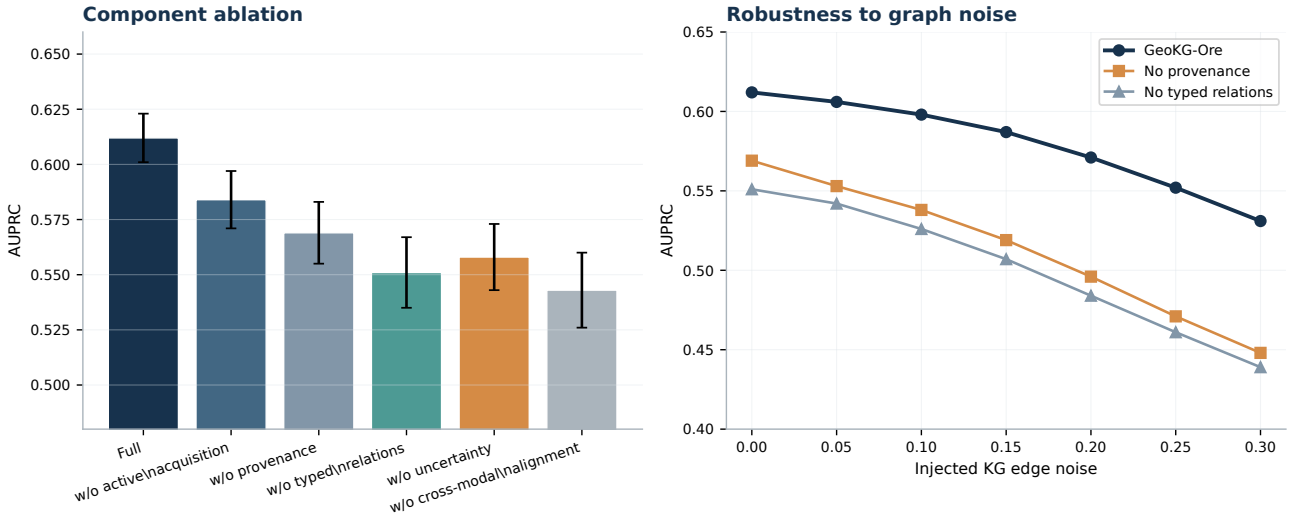


Figure 4. Component ablation and robustness to knowledge-graph edge noise.

2.11 Value-of-information action ranking

Beyond aggregate budget utility, we inspect which evidence actions are prioritized by the proposed policy. Figure 8 ranks candidate actions by expected value of information per cost unit, showing the relative priority assigned to report retrieval, geochemical assays, airborne magnetics, satellite alteration products, geological mapping, and field sampling.

2.12 Sensitivity to validation design

Random cell-level splits can place spatially adjacent observations from the same mineralized system in both training and test sets. We therefore compare random splitting with spatial-block validation. Figure 9 reports the paired fold-level AUPRC and AUROC values under the two validation designs.

3 Discussion

3.1 Why the shared posterior matters

The central design choice in GeoKG-Ore is to make the probabilistic prospectivity field the shared state for prediction, evidence acquisition, and report generation. A conventional pipeline may calculate a prospectivity score, retrieve a few related graph triples, and then ask an LLM to explain the result. Such a sequence does not guarantee that retrieved evidence changes the

Table 5. Cross-region transfer results (AUPRC).

Training region	Held-out region	AUPRC	Calibration error
Great Basin	Wulonggou	0.487	0.082
Great Basin	Hamissana	0.462	0.094
Great Basin	Fatira	0.471	0.087
Wulonggou	Great Basin	0.503	0.076
Hamissana	Fatira	0.516	0.071
Fatira	Hamissana	0.529	0.068

Table 6. Geological question answering and evidence-grounding results.

Question category	Answer F1	Evidence prec.	Citation comp.	Abstention acc.
Deposit-system constraints	0.884	0.921	0.946	0.901
Lithology–structure relations	0.861	0.913	0.938	0.894
Alteration and geochemical clues	0.842	0.902	0.929	0.887
Spatial and directional queries	0.873	0.916	0.941	0.912
Conflicting or missing evidence	0.801	0.889	0.918	0.936

prediction or that the explanation reflects uncertainty. In GeoKG-Ore, evidence updates the same posterior that determines the target ranking and the next action. The LLM cannot independently promote a target based on a plausible narrative if the corresponding graph paths and spatial observations do not support it.

3.2 Geological plausibility versus predictive accuracy

High AUROC or AUPRC alone does not establish geological validity. A model may learn sampling density, access routes, or reporting intensity instead of ore-forming processes. For this reason, the evaluation separates predictive discrimination, calibration, geological consistency, evidence grounding, and decision utility. Spatial holdout and buffer-based leakage control are necessary, but they do not replace field validation. Geological plausibility should be assessed by experts using predeclared mineral-system criteria, and explanations should be tested for stability under evidence perturbation and source removal. The uncertainty-aware perspective is consistent with recent work on three-dimensional MPM and spatially separated evaluation [5, 10, 16].

3.3 Public data limitations and transferability

Public mineral-occurrence databases are heterogeneous compilations, not complete inventories of mineralization. MRDS coverage is more comprehensive in the United States than in many other countries, and source records may reflect historical reporting rather than current field conditions [50]. Remote-sensing alteration proxies are affected by vegetation, weathering, cloud cover, surface exposure, and sensor-specific spectral response. Geochemical and geophysical observations are also unevenly sampled and may be collected at different scales. The graph representation makes these limitations explicit through source reliability, spatial support, temporal validity, and missingness metadata; it cannot eliminate the underlying observation bias. Transfer to another district therefore requires deposit-model alignment and careful separation of geology shift from data-collection shift.

3.4 Role and limits of the LLM

The LLM is used for semantic query parsing, ontology-constrained hypothesis formulation, retrieval planning, and natural-language synthesis. It does not replace geophysical inversion, geochemical anomaly detection, or probabilistic prospectivity inference. This division of responsibility is important because general-purpose language models may produce fluent but unsupported geological claims. Recent geological LLM and knowledge-graph construction studies show the promise of domain adaptation and structured extraction, while also motivating source-aware verification [23, 26–28, 41]. In deployment, the LLM should expose uncertainty and abstain when the evidence packet is incomplete, contradictory, or outside the supported ontology.

3.5 Limitations and future work

Several limitations remain. First, the current formulation focuses on surface and near-surface evidence, whereas concealed deposits require three-dimensional geological models, drilling information, and depth-dependent geophysical interpretation. Second, source reliability and geological compatibility may require expert calibration and may vary across deposit types. Third, although the evidence-action policy explicitly models acquisition cost and expected utility, its operational performance should

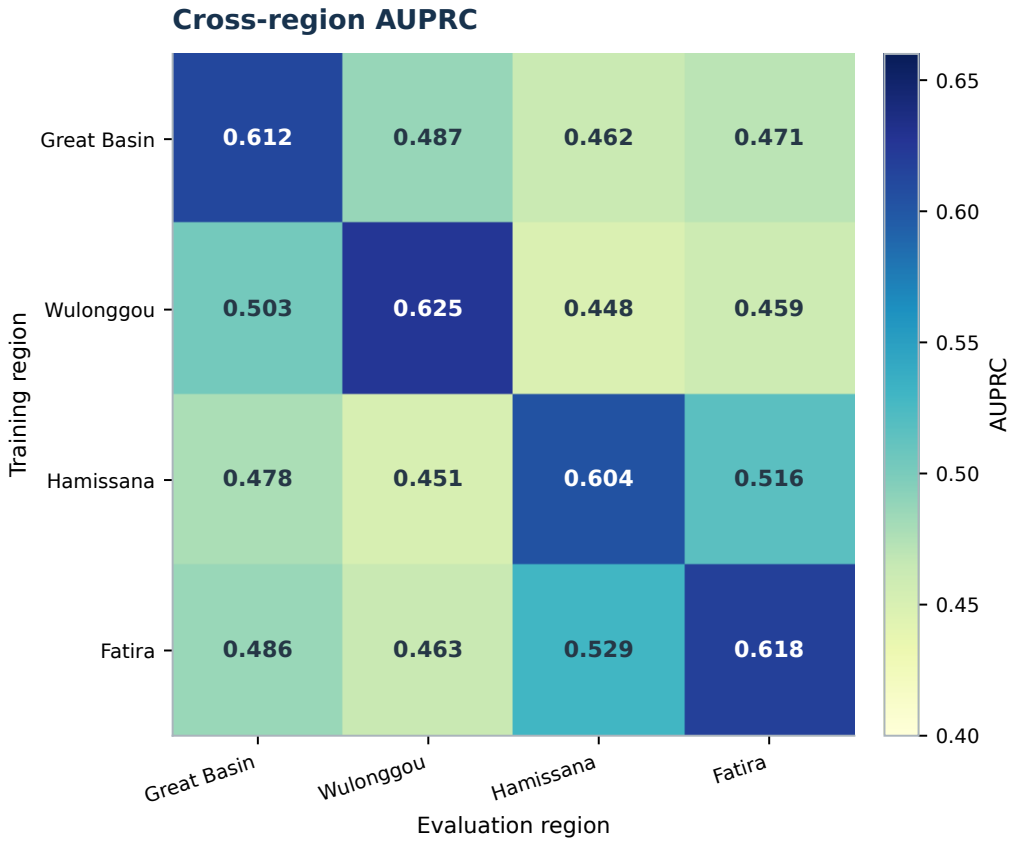


Figure 5. Cross-region AUPRC matrix. Diagonal entries represent within-region evaluation; off-diagonal entries represent transfer.

be further examined in field workflows. Future work will extend the graph to 3D/4D geological contexts, broaden expert assessment of relation and claim verification, and evaluate prospective targets against newly acquired observations.

4 Conclusion

We presented GeoKG-Ore, a framework for knowledge graph-augmented LLM reasoning in intelligent gold exploration. Its core contribution is a closed-loop formulation in which a provenance-aware spatial knowledge graph and a query-conditioned probabilistic prospectivity field jointly govern target ranking, uncertainty estimation, evidence acquisition, and report generation. A value-of-information policy selects additional observations under a constrained budget, while a provenance-constrained LLM produces evidence-linked explanations and abstains when support is insufficient. We evaluated the framework using public mineral-occurrence, geological, geochemical, geophysical, and satellite data sources, with spatially separated validation and explicit leakage controls. GeoKG-Ore achieved an AUROC of 0.913, an AUPRC of 0.612, a top-5% area recall of 0.574, and an ECE of 0.038. It also improved geological consistency and reduced unsupported claims relative to the strongest listed baseline. These findings demonstrate the potential of coupling geological knowledge, probabilistic spatial inference, and adaptive evidence acquisition for more traceable and decision-aware gold exploration.

5 Method

5.1 Problem formulation

Let the area of interest be a spatial domain $\Omega \subset \mathbb{R}^2$ (or $\mathbb{R}^3$ when subsurface volumes are available), partitioned into N cells $\mathcal{V}_s = \{v_i\}_{i=1}^{N}$. For each cell, heterogeneous observations are collected from modalities

$$\mathcal{M} = \{\text{geology}, \text{geochemistry}, \text{geophysics}, \text{remote sensing}, \text{topography}, \text{reports}\}.$$

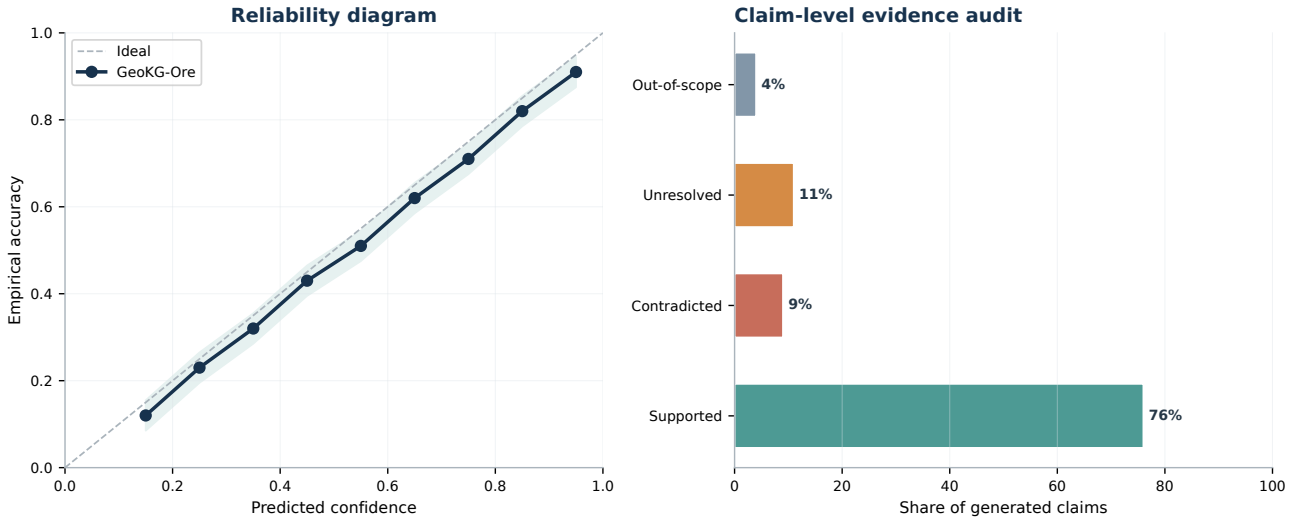


Figure 6. Calibration and claim-level evidence audit.

Table 7. Evidence-acquisition efficiency under different action budgets.

Policy	Actions	R@5%	Utility / cost	Mean latency (s)
Fixed-order retrieval	8	0.520	0.064	18.7
Uncertainty-first	8	0.590	0.073	20.4
GeoKG-Ore VOI policy	8	0.625	0.089	22.1
Fixed-order retrieval	12	0.550	0.046	26.9
Uncertainty-first	12	0.610	0.051	29.5
GeoKG-Ore VOI policy	12	0.646	0.061	31.4

The observation of modality m at cell i is denoted by $x_i^{(m)} \in \mathbb{R}^{d_m}$, with an availability mask $a_i^{(m)} \in \{0,1\}$. The target variable $y_i \in \{0,1\}$ indicates whether cell i is compatible with the target gold-mineralization system at the resolution of the benchmark. Since an unrecorded cell may contain an undiscovered deposit, observed non-occurrence is treated as unlabeled rather than automatically negative.

An exploration request is represented by

$$q = (\Omega_q, \tau_q, \mathscr{C}_q, B_q, \mathscr{S}_q),$$

where Ω_q is the requested region, τ_q is the target deposit-system hypothesis, $\mathscr{C}_q$ denotes geological and operational constraints, B_q is the evidence-acquisition budget, and $\mathscr{S}_q$ specifies the desired output scale. GeoKG-Ore returns a posterior prospectivity field $\boldsymbol{p}_q = (p_1, \ldots, p_N)$, an uncertainty field, a ranked set of candidate targets, and a report whose claims are linked to verifiable evidence identifiers.

5.2 Provenance-aware spatial geological knowledge graph

We define a heterogeneous graph

$$\mathscr{G} = (\mathscr{V}, \mathscr{E}, \mathscr{T}_v, \mathscr{T}_e, \boldsymbol{X}, \boldsymbol{P}),$$

where $\mathscr{T}_v$ and $\mathscr{T}_e$ are node and relation types, $\boldsymbol{X}$ stores node attributes, and $\boldsymbol{P}$ stores provenance metadata. Node types include spatial cells, lithological units, stratigraphic units, faults and shear zones, alteration assemblages, geochemical anomalies, geophysical responses, mineral occurrences, deposit models, survey products, and source documents. Relation types include `hosts`, `cross-cuts`, `adjacent-to`, `overlaps`, `located-in`, `supports`, `contradicts`, `derived-from`, and `temporally-precedes`.

Each edge $e = (u, r, v)$ is associated with a structured evidence record

$$\pi_e = (s_e, \rho_e, \mathscr{R}_e, [t_e^-, t_e^+], \delta_e, \chi_e),$$

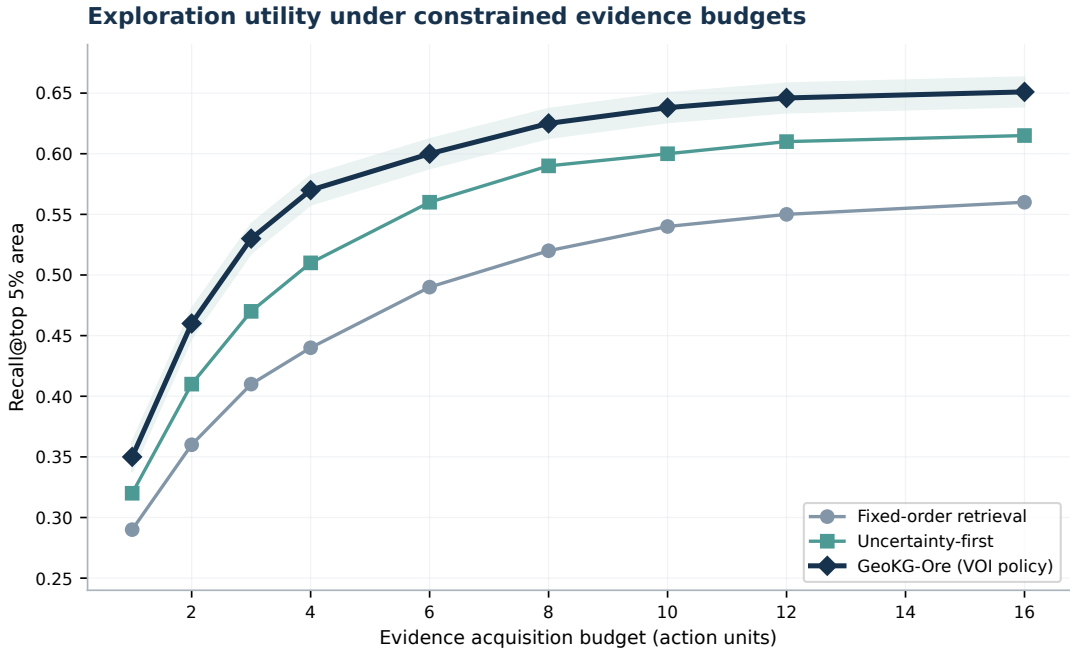


Figure 7. Target utility as a function of the evidence-acquisition budget.

where s_e is the source identifier, $\rho_e \in [0,1]$ is source reliability, $\mathscr{R}_e$ is the spatial footprint, $[t_e^-, t_e^+]$ is the validity interval, δ_e is the extraction or measurement confidence, and χ_e records whether the relation is supporting, contradicting, or unresolved. This representation prevents a textual relation from being detached from the region, time, and source that justify it.

Spatial compatibility is encoded using a relation-specific kernel:

$$\kappa_{uv}^{(r)} = \exp\left(-\frac{d_g(\mathscr{R}_u, \mathscr{R}_v)^2}{2\sigma_r^2}\right) \cdot \exp\left(-\frac{\Delta t_{uv}}{\tau_r}\right) \cdot \mathbb{I}\left[\mathscr{P}_r(\mathscr{R}_u, \mathscr{R}_v) = 1\right],$$

where d_g is a geodesic or projected distance, Δt_{uv} is temporal separation, $\mathscr{P}_r$ is a relation-specific spatial predicate, and σ_r, τ_r are scale parameters. The predicate may encode intersection, containment, proximity, directional alignment, or contact adjacency. Unlike isotropic smoothing, $\kappa_{uv}^{(r)}$ preserves the semantics of the geological relation.

To estimate the reliability of an edge in context, we compute

$$\widehat{\rho}_e = \sigma\left(\boldsymbol{w}_r^\top [\boldsymbol{h}_u \| \boldsymbol{h}_v \| \boldsymbol{e}_r \| \boldsymbol{z}_{\pi_e} \| \boldsymbol{s}_{\mathrm{src}}] + b_r\right) \cdot \rho_e \delta_e \kappa_{uv}^{(r)},$$

where $\boldsymbol{e}_r$ is the relation embedding, $\boldsymbol{z}_{\pi_e}$ represents provenance features, and $\boldsymbol{s}_{\mathrm{src}}$ encodes source-level metadata. This contextual reliability is used in both graph message passing and evidence verification.

5.3 Cross-modal alignment and relation-aware graph encoding

Each modality is encoded by a modality-specific function

$$\boldsymbol{z}_i^{(m)} = f_{\theta_m}(x_i^{(m)}, a_i^{(m)}),$$

where missing observations are handled by an explicit mask and a learned missingness embedding. For spatial cell nodes, modality representations are fused through a gated operator:

$$\boldsymbol{z}_i^{\mathrm{cell}} = \sum_{m \in \mathscr{M}} \alpha_i^{(m)} \boldsymbol{W}_m \boldsymbol{z}_i^{(m)}, \qquad \alpha_i^{(m)} = \frac{a_i^{(m)} \exp(g_m(\boldsymbol{z}_i^{(m)}, \boldsymbol{c}_i))}{\sum_{m'} a_i^{(m')} \exp(g_{m'}(\boldsymbol{z}_i^{(m')}, \boldsymbol{c}_i)) + \varepsilon},$$

where $\boldsymbol{c}_i$ contains spatial context, coverage, and quality indicators. This makes modality contribution conditional on local data availability and quality.

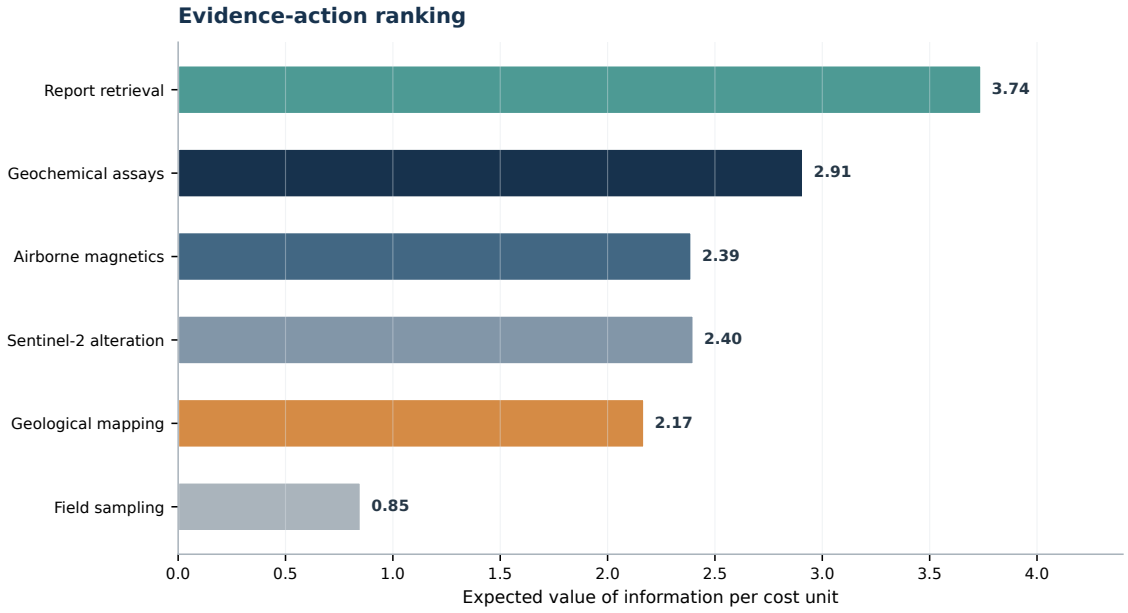


Figure 8. Value-of-information ranking for candidate evidence actions.

Graph messages are propagated over typed relations. At layer ℓ, the attention coefficient for edge (u, r, v) is

$$\alpha_{uv}^{(r,\ell)} = \frac{\exp\left(\mathrm{LeakyReLU}\left[\boldsymbol{a}_r^\top (\boldsymbol{W}_r^{(q)} \boldsymbol{h}_v^{(\ell)} \| \boldsymbol{W}_r^{(k)} \boldsymbol{h}_u^{(\ell)} \| \boldsymbol{e}_r \| \boldsymbol{d}_{uv})\right]\right) \widehat{\rho}_{uv}}{\sum_{(u',r',v)\in\mathcal{N}(v)} \exp(\cdot)\widehat{\rho}_{u'v}},$$

where $\boldsymbol{d}_{uv}$ contains normalized spatial and temporal descriptors. The update is

$$\boldsymbol{h}_v^{(\ell+1)} = \mathrm{LayerNorm}\left[\boldsymbol{h}_v^{(\ell)} + \sum_{(u,r,v)\in\mathcal{N}(v)} \alpha_{uv}^{(r,\ell)} \boldsymbol{W}_r^{(v)} \boldsymbol{h}_u^{(\ell)}\right].$$

The graph encoder therefore aggregates evidence according to relation semantics, spatial support, and provenance reliability rather than treating every neighboring node as equally informative.

We align cell-level observations with their graph representations using a spatially masked contrastive objective. Let $\mathcal{P}$ denote matched observation–entity pairs and $\mathcal{N}(i)$ denote hard negatives sampled from nearby but geologically incompatible contexts. Then

$$\mathcal{L}_{\mathrm{align}} = -\sum_{(i,v)\in\mathcal{P}} \log \frac{\exp(\mathrm{sim}(\boldsymbol{z}_i, \boldsymbol{h}_v)/T)}{\exp(\mathrm{sim}(\boldsymbol{z}_i, \boldsymbol{h}_v)/T) + \sum_{v^-\in\mathcal{N}(i)} \exp(\mathrm{sim}(\boldsymbol{z}_i, \boldsymbol{h}_{v^-})/T)}.$$

Spatially nearby hard negatives discourage the encoder from relying only on geographic proximity and encourage it to distinguish mineral-system-compatible relations.

5.4 Query-conditioned probabilistic prospectivity field

The natural-language query is parsed into a structured ore-system hypothesis

$$H_q = (\mathcal{L}_q, \mathcal{S}_q, \mathcal{A}_q, \mathcal{G}_q),$$

where the components denote expected lithological, structural, alteration, and geochemical/geophysical conditions. The LLM proposes the hypothesis and candidate constraints, while the graph schema validator maps them to supported ontology concepts and records unresolved terms.

For each cell, the unary evidence score is

$$\eta_i(H_q) = \boldsymbol{w}_q^\top \left[\boldsymbol{z}_i^{\mathrm{cell}} \| \boldsymbol{h}_i^{\mathrm{KG}} \| \boldsymbol{e}(H_q) \| \boldsymbol{u}_i\right], \qquad p_i^{(0)} = \sigma(\eta_i),$$

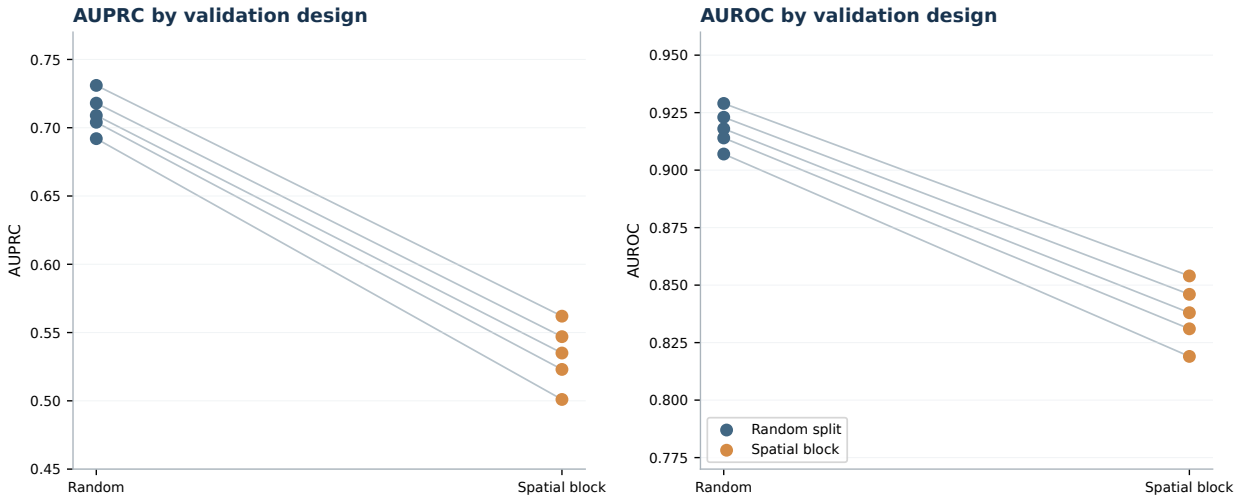


Figure 9. Comparison of random and spatial-block validation across five folds.

where $\boldsymbol{u}_i$ contains missingness, source quality, and coverage descriptors. The graph-conditioned posterior over the full spatial label field is defined by an energy-based distribution:

$$\mathbb{P}_\theta(\boldsymbol{y} \mid \boldsymbol{X}, \mathscr{G}, H_q) = \frac{1}{Z_\theta} \exp\{-\mathscr{E}_\theta(\boldsymbol{y}; \boldsymbol{X}, \mathscr{G}, H_q)\},$$

with

$$\begin{aligned}\mathscr{E}_\theta = &\sum_i \left[-y_i \log p_i^{(0)} - (1-y_i)\log(1-p_i^{(0)})\right] \\ &+ \sum_{(i,j,r)\in\mathscr{E}_s} \omega_{ij}^{(r)}(H_q)\Psi_r(y_i, y_j; \boldsymbol{h}_i, \boldsymbol{h}_j) \\ &+ \lambda_{\text{sys}} \sum_i \Phi_{\text{sys}}(y_i, H_q, \mathscr{N}_i) + \lambda_{\text{conf}} \sum_{e\in\mathscr{E}_{\text{conf}}} \Phi_{\text{conf}}(e, \boldsymbol{y}).\end{aligned}$$

Here Ψ_r is a relation-specific compatibility potential, Φ_{sys} penalizes violations of the target mineral-system constraints, and Φ_{conf} preserves explicit contradictions rather than forcing incompatible evidence into a single positive relation. The partition function Z_θ normalizes the distribution. Approximate marginal inference yields $p_i = \mathbb{P}(y_i = 1 \mid \boldsymbol{X}, \mathscr{G}, H_q)$ and posterior uncertainty $u_i = -p_i \log p_i - (1-p_i)\log(1-p_i)$.

Because mineral occurrence databases contain known deposits and prospects but do not provide exhaustive barren labels, we use a positive–unlabeled (PU) objective. Let $\mathscr{P}$ be verified positive cells, $\mathscr{U}$ be unlabeled cells, and π be the estimated positive class prior. For a classifier score f_θ, the non-negative PU risk is

$$\begin{aligned}\mathscr{L}_{\text{PU}} = &\pi\, \mathbb{E}_{x\sim P}[\ell(f_\theta(x), 1)] \\ &+ \max\{0,\ \mathbb{E}_{x\sim U}[\ell(f_\theta(x), 0)] - \pi \mathbb{E}_{x\sim P}[\ell(f_\theta(x), 0)]\}.\end{aligned}$$

This formulation reduces the risk of interpreting unrecorded or underexplored areas as reliable negatives. The prior π is treated as a sensitivity parameter and is estimated only within the training folds.

5.5 Uncertainty-aware value-of-information acquisition

A prospectivity map is useful for exploration only when it can guide the next observation. At iteration t, the agent maintains a belief state

$$s_t = (\boldsymbol{p}_t, \boldsymbol{u}_t, \mathscr{G}_t, H_q, B_t, \mathscr{A}_t),$$

where $\boldsymbol{p}_t$ and $\boldsymbol{u}_t$ are the current posterior and uncertainty fields, $\mathscr{G}_t$ is the evidence graph, B_t is the remaining budget, and $\mathscr{A}_t$ is the set of feasible evidence actions. An action may request a geological map layer, geochemical assay subset, geophysical grid, satellite-derived alteration feature, report passage, or targeted field observation.

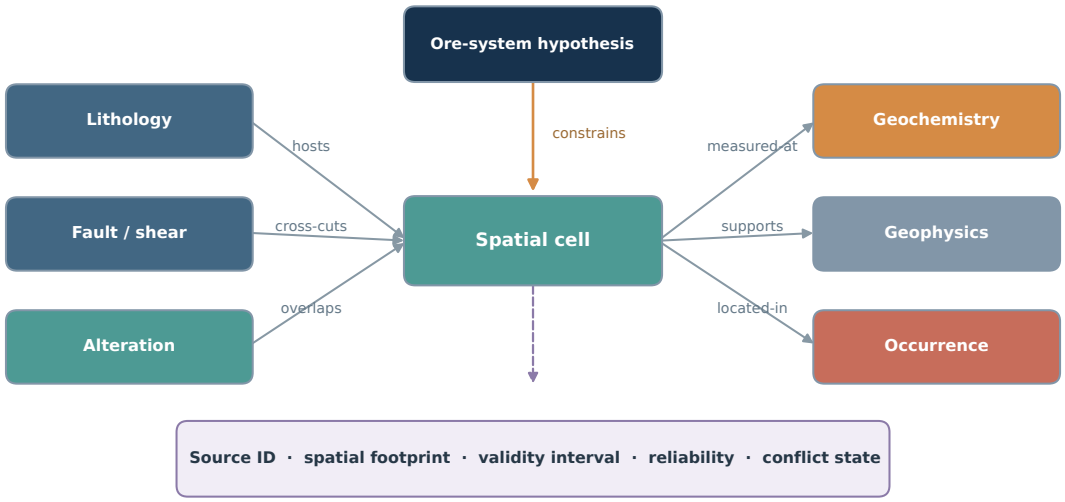


Figure 10. Typed spatial geological knowledge graph. Evidence nodes connect to spatial cells through geological relations, while provenance metadata is attached to every fact and relation.

For a candidate action a, let o be its possible observation and let $\boldsymbol{p}_{t+1}^{a,o}$ denote the posterior after incorporating o. We define the expected value of information as

$$\mathrm{VOI}(a \mid s_t) = \underbrace{\mathscr{U}(\boldsymbol{p}_t) - \mathbb{E}_{o\sim p(o|a,s_t)}\left[\mathscr{U}(\boldsymbol{p}_{t+1}^{a,o})\right]}_{\text{expected uncertainty reduction}} + \lambda_g\,\mathbb{E}_o[\Delta\mathscr{C}_{\mathrm{KG}}(a,o)] + \lambda_d\,\mathbb{E}_o[\Delta\mathscr{U}_{\mathrm{decision}}(a,o)] - \lambda_c\,c(a),$$

where $\mathscr{U}$ is a spatially weighted uncertainty functional, $\Delta\mathscr{C}_{\mathrm{KG}}$ measures the expected resolution of graph conflicts or missing relations, $\Delta\mathscr{U}_{\mathrm{decision}}$ measures the expected improvement in target-selection utility, and $c(a)$ is the acquisition cost. The uncertainty functional gives higher weight to cells near the current target boundary:

$$\mathscr{U}(\boldsymbol{p}) = \sum_i \underbrace{[\varepsilon + 4p_i(1-p_i)]}_{\text{local uncertainty}}\underbrace{[1+\gamma r_i(H_q)]}_{\text{target relevance}},$$

where $r_i(H_q)$ is a query-conditioned relevance score. The action is selected by

$$a_t^\star = \arg\max_{a\in\mathscr{A}_t,\, c(a)\le B_t} \frac{\mathrm{VOI}(a \mid s_t)}{(c(a)+\varepsilon)^\zeta}.$$

The policy is therefore sensitive to both information gain and exploration cost. In contrast to a fixed retrieval budget, it can stop when the expected marginal utility of further evidence becomes non-positive.

To support stochastic observations and variable action costs, the policy can be parameterized as

$$\pi_\phi(a \mid s_t) = \frac{\exp(Q_\phi(s_t,a)/T_\pi)}{\sum_{a'\in\mathscr{A}_t}\exp(Q_\phi(s_t,a')/T_\pi)}.$$

The value function is trained with a reward combining target utility, posterior calibration, and acquisition cost:

$$R_t = \Delta\mathscr{U}_{\mathrm{decision}} + \lambda_u\Delta\mathscr{U} + \lambda_g\Delta\mathscr{C}_{\mathrm{KG}} - \lambda_c c(a_t).$$

A deterministic greedy policy can be used at deployment for reproducibility, while the stochastic policy is useful during training and policy evaluation.

5.6 Provenance-constrained LLM generation and verification

The LLM receives a compact evidence packet rather than the entire graph. For each candidate target, the packet contains: (i) the target coordinates and posterior interval; (ii) the highest-weight supporting and contradicting graph paths; (iii) the source IDs, spatial footprints, and validity intervals for each observation; (iv) unresolved graph conflicts; and (v) the actions already taken and the remaining budget. The LLM generates a set of atomic claims $\mathscr{C} = \{c_j\}_{j=1}^{J}$, each paired with an evidence subset $\mathscr{S}_j \subseteq \mathscr{E}_t$.

A claim verifier computes

$$v_j = \max_{\mathscr{S}_j} \left[\mathrm{Entail}(c_j, \mathscr{S}_j) \cdot \mathrm{Prov}(\mathscr{S}_j) \cdot \mathrm{SpatialMatch}(c_j, \mathscr{S}_j) \right],$$

where Entail measures semantic support, Prov measures provenance completeness and source reliability, and SpatialMatch verifies that the claim and evidence refer to compatible spatial and temporal supports. A claim is released only if

$$v_j \geq \tau_{\mathrm{claim}} \quad \wedge \quad \mathrm{Conflict}(c_j, \mathscr{S}_j) < \tau_{\mathrm{conf}}.$$

Otherwise, the system either qualifies the statement as uncertain or abstains. The generated report is therefore a structured object containing target rank, posterior uncertainty, evidence paths, contradictory observations, source references, and recommended next actions.

5.7 Joint optimization

The components are coupled through a common posterior and provenance state. We optimize

$$\begin{aligned} \mathscr{L}_{\text{GeoKG-Ore}} = {} & \lambda_{\mathrm{PU}} \mathscr{L}_{\mathrm{PU}} + \lambda_{\mathrm{align}} \mathscr{L}_{\mathrm{align}} + \lambda_{\mathrm{KG}} \mathscr{L}_{\mathrm{KG}} \\ & + \lambda_{\mathrm{cal}} \mathscr{L}_{\mathrm{cal}} + \lambda_{\mathrm{VOI}} \mathscr{L}_{\mathrm{policy}} + \lambda_{\mathrm{ground}} \mathscr{L}_{\mathrm{ground}}. \end{aligned}$$

The graph consistency loss is

$$\mathscr{L}_{\mathrm{KG}} = \sum_{(u,r,v) \in \mathscr{E}} \widehat{\rho}_{uv} \left[\max(0, m_r - s_r(u,v)) + \lambda_{\mathrm{neg}} \mathbb{I}[\chi_{uv} = \mathrm{contradict}] \max(0, s_r(u,v) - m_r^{-}) \right],$$

where s_r is the relation compatibility score and m_r, m_r^{-} are positive and contradiction margins. Calibration is optimized with a proper scoring rule and an auxiliary bin-based penalty. The policy loss is

$$\mathscr{L}_{\mathrm{policy}} = -\mathbb{E}_{\pi_\phi} \left[\sum_t \gamma^t R_t \right] + \lambda_{\mathrm{ent}} \mathbb{E}_t [\mathrm{KL}(\pi_\phi(\cdot \mid s_t) \| \pi_0(\cdot \mid s_t))],$$

where π_0 is a cost-aware prior policy. The grounding loss penalizes claims without sufficient evidence support and rewards complete source and spatial citations. This objective ties graph reliability, posterior quality, evidence acquisition, and report grounding together instead of optimizing them as independent modules.

5.8 Computational considerations

Let $|V|$ and $|E|$ denote graph nodes and edges, d the hidden dimension, and L the number of graph layers. A sparse relation-aware graph pass costs approximately $\mathscr{O}(L(|V|d^2 + |E|d))$. The dominant inference cost is associated with posterior approximation and action evaluation. Restricting action candidates to a query-conditioned frontier and caching local graph messages reduces repeated computation. If K candidate actions are evaluated and each action updates a local region containing n_a cells and e_a edges, the incremental cost is approximately $\mathscr{O}(K(n_a d^2 + e_a d))$, rather than recomputing the full graph. The LLM is invoked for query parsing and final report synthesis, while numerical prospectivity and action ranking remain in explicit, inspectable modules.

Funding

This research was supported by the Science and Technology Program of Xinjiang Uygur Autonomous Region, China (Grant No. 2024B03010).

Competing Interests

The authors declare no competing interests.

Algorithm 1 GeoKG-Ore inference with budget-aware evidence acquisition

Require: Query q, initial evidence $\mathscr{E}_0$, graph schema $\mathscr{O}$, budget B
Ensure: Prospectivity posterior $\boldsymbol{p}$, verified report $\mathscr{R}$

Parse q into AOI, ore-system hypothesis H_q, constraints, and output scale
Build or retrieve provenance-aware graph $\mathscr{G}_0$ under schema $\mathscr{O}$
Align available observations and initialize posterior $\boldsymbol{p}_0$ and uncertainty $\boldsymbol{u}_0$
$t \leftarrow 0$
while $B_t > 0$ and $\max_{a \in \mathscr{A}_t} \mathrm{VOI}(a \mid s_t) > 0$ **do**
    Generate feasible evidence actions $\mathscr{A}_t$
    Select $a_t^\star \leftarrow \arg\max_a \mathrm{VOI}(a \mid s_t)/(c(a)+\varepsilon)^\zeta$
    Acquire observation o_t and attach source, spatial, and temporal metadata
    Update graph $\mathscr{G}_{t+1}$ and contextual edge reliabilities
    Update posterior $\boldsymbol{p}_{t+1}$ and uncertainty $\boldsymbol{u}_{t+1}$
    $B_{t+1} \leftarrow B_t - c(a_t^\star)$; $t \leftarrow t+1$
end while
Rank candidate targets using posterior, uncertainty, and query constraints
Construct evidence packets and generate atomic LLM claims
Verify each claim against graph paths, source provenance, and spatial support
Return ranked targets, calibrated uncertainty, verified claims, and unresolved conflicts

Author Contributions

L.L. conceived the study, developed the overall research framework, and contributed to the design of the proposed GeoKG-Ore methodology.

C.W. supervised the research, contributed to the geological exploration framework and metallogenic knowledge representation, and critically revised the manuscript.

J.Y. contributed to the development of the spatial geological knowledge graph, multimodal evidence integration, and knowledge graph-augmented reasoning methodology.

Z.W. contributed to geological data collection, data preprocessing, mineral occurrence analysis, and the construction of geological evidence layers.

S.Z. contributed to geochemical, geophysical, and remote-sensing data integration, spatial analysis, and experimental evaluation.

C.L. contributed to model implementation, comparative experiments, performance evaluation, and analysis of the experimental results.

L.L. prepared the initial manuscript. All authors contributed to manuscript revision, reviewed the results, and approved the final version.

Data Availability

The datasets used in this study were obtained from publicly accessible geological, mineral-resource, and Earth observation data repositories.

Mineral occurrence records and deposit information were obtained from the United States Geological Survey (USGS) Mineral Resources Data System (MRDS), accessible at https://www.usgs.gov/publications/mineral-resources-data-sys

Geological, geochemical, geophysical, and mineral resource spatial datasets were obtained through the USGS Mineral Resources Online Spatial Data Access Tool, available at https://www.usgs.gov/tools/mineral-resources-online-spatia

Landsat Collection 2 Level-2 satellite imagery was accessed through USGS EarthExplorer at https://earthexplorer.usgs.gov.

Sentinel-2 Level-2A surface reflectance products were obtained from the Copernicus Data Space Ecosystem at https://dataspace.copernicus.eu.

These publicly available data sources were used to construct the spatial analysis units, multimodal geological evidence representations, mineral occurrence labels, and provenance-aware geological knowledge graph employed in this study.

The datasets are subject to the access, licensing, and redistribution conditions specified by their respective providers.

Ethics Declaration

This study involved computational analysis of publicly available geological, geochemical, geophysical, mineral occurrence, and satellite remote-sensing datasets.

No human participants or animals were involved in this research. The study did not involve human-subject experimentation, personal data collection, or clinical interventions.

Ethics approval and informed consent were not applicable to this study.

References

1. Bonham-Carter, G. F. *Geographic Information Systems for Geoscientists: Modelling with GIS* (Pergamon, 1994).

2. Carranza, E. J. M. *Geochemical Anomaly and Mineral Prospectivity Mapping in GIS* (Elsevier, 2009).

3. McCuaig, T. C., Beresford, S. & Hronsky, J. Translating the mineral systems approach into an effective exploration targeting system. *Ore Geol. Rev.* **38**, 128–138, DOI: 10.1016/j.oregeorev.2010.05.008 (2010).

4. Porwal, A., Carranza, E. J. M. & Hale, M. Artificial neural networks for mineral-potential mapping: A case study from aravalli province, western india. *Nat. Resour. Res.* **12**, 155–171 (2003).

5. Wang, Y., Khodadadzadeh, M. & Zurita-Milla, R. Spatial+: A new cross-validation method to evaluate geospatial machine learning models. *Int. J. Appl. Earth Obs. Geoinformation* **121**, 103364, DOI: 10.1016/j.jag.2023.103364 (2023).

6. Senanayake, I. P. *et al.* A spatial data-driven approach for mineral prospectivity mapping. *Remote. Sens.* **15**, 4074, DOI: 10.3390/rs15164074 (2023).

7. Yan, Q. *et al.* Mineral prospectivity mapping based on spatial feature classification with geological map knowledge graph embedding: Case study of gold ore prediction at wulonggou, qinghai province (western china). *Nat. Resour. Res.* **33**, 2385–2406, DOI: 10.1007/s11053-024-10386-6 (2024).

8. Sihombing, F. M. H., Palin, R. M., Hughes, H. S. R. & Robb, L. J. Improved mineral prospectivity mapping using graph neural networks. *Ore Geol. Rev.* **172**, 106215, DOI: 10.1016/j.oregeorev.2024.106215 (2024).

9. Wang, L. *et al.* Enhancing mineral prospectivity mapping with geospatial artificial intelligence: A geographically neural network-weighted logistic regression approach. *Int. J. Appl. Earth Obs. Geoinformation* **128**, 103746, DOI: 10.1016/j.jag.2024.103746 (2024).

10. Zhang, Z. *et al.* An uncertainty-quantification machine learning framework for data-driven three-dimensional mineral prospectivity mapping. *Nat. Resour. Res.* **33**, 1393–1411, DOI: 10.1007/s11053-024-10349-x (2024).

11. He, H., Zhu, H., Yang, X., Zhang, W. & Wang, J. Mineral prospectivity prediction based on convolutional neural network and ensemble learning. *Sci. Reports* **14**, 22654, DOI: 10.1038/s41598-024-73357-0 (2024).

12. Mohamed Taha, A. M. *et al.* Toward data-driven mineral prospectivity mapping from remote sensing data using deep forest predictive model. *Nat. Resour. Res.* **33**, 2407–2431, DOI: 10.1007/s11053-024-10387-5 (2024).

13. Wang, R. *et al.* Mineral prospectivity prediction based on the dynamic relation model atten-gcn: A case study of gold prospecting in the yingfengjie area, shaanxi province. *Ore Geol. Rev.* **176**, 106399, DOI: 10.1016/j.oregeorev.2024.106399 (2025).

14. Cao, C. *et al.* Attention-driven graph convolutional neural networks for mineral prospectivity mapping. *Ore Geol. Rev.* **180**, 106554, DOI: 10.1016/j.oregeorev.2025.106554 (2025).

15. Sheng, S. *et al.* Knowledge–data collaboration-driven mineral prospectivity prediction with graph attention networks. *Minerals* **15**, 1164, DOI: 10.3390/min15111164 (2025).

16. Yuan, M. *et al.* A systematic review of deep learning methods for mineral exploration using multisource geoscience data (2018–2025). *Int. J. Appl. Earth Obs. Geoinformation* **149**, 105254, DOI: 10.1016/j.jag.2026.105254 (2026).

17. Kuckreja, K. *et al.* Geochat: Grounded large vision-language model for remote sensing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 27831–27840, DOI: 10.1109/CVPR52733.2024.02629 (2024).

18. Guo, X. *et al.* Skysense: A multi-modal remote sensing foundation model towards universal interpretation for earth observation imagery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 27672–27683, DOI: 10.1109/CVPR52733.2024.02613 (2024).

19. Xiao, A. *et al.* Foundation models for remote sensing and earth observation: A survey. *IEEE Geosci. Remote. Sens. Mag.* **13**, 297–324, DOI: 10.1109/MGRS.2025.3576766 (2025).

20. Lu, S. *et al.* Vision foundation models in remote sensing: A survey. *IEEE Geosci. Remote. Sens. Mag.* **13**, 190–215, DOI: 10.1109/MGRS.2025.3541952 (2025).

21. Li, Z., Zhou, W., Chiang, Y.-Y. & Chen, M. Geolm: Empowering language models for geospatially grounded language understanding. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing* (2023).

22. Manvi, R. *et al.* Geollm: Extracting geospatial knowledge from large language models. In *International Conference on Learning Representations* (2024).

23. Zhou, B. & Li, K. Fusing geoscience large language models and lightweight rag for enhanced geological question answering. *Geosciences* **15**, 382, DOI: 10.3390/geosciences15100382 (2025).

24. Li, H. *et al.* A question-answering framework for geospatial data retrieval enhanced by a knowledge graph and large language models. *Int. J. Digit. Earth* **18**, DOI: 10.1080/17538947.2025.2510566 (2025).

25. Liang, J. *et al.* Geographrag: A graph-based retrieval-augmented generation approach for empowering large language models in automated geospatial modeling. *Int. J. Appl. Earth Obs. Geoinformation* **142**, 104712, DOI: 10.1016/j.jag.2025.104712 (2025).

26. Wu, G. *et al.* Geoprospect: A domain-specific geological large language model with enhanced continual learning. *Neurocomputing* **650**, 130801, DOI: 10.1016/j.neucom.2025.130801 (2025).

27. Zhu, J. *et al.* Gold deposit ontology guides large language model to transform text into knowledge graphs for gold deposits. *Minerals* **16**, 50, DOI: 10.3390/min16010050 (2026).

28. Ma, K., Zhan, Q., Zhang, T. *et al.* Geo-mag: A knowledge graph-enhanced multimodal retrieval-augmented generation framework for geological map understanding. *Transactions GIS* **30**, e70226, DOI: 10.1111/tgis.70226 (2026).

29. Ibrahim, N., Aboulela, S., Ibrahim, A. *et al.* A survey on augmenting knowledge graphs with large language models: Models, evaluation metrics, benchmarks, and challenges. *Discov. Artif. Intell.* **4**, 76, DOI: 10.1007/s44163-024-00175-8 (2024).

30. Gao, Y. *et al.* Retrieval-augmented generation for large language models: A survey. arXiv preprint arXiv:2312.10997 (2024).

31. Hu, Y. *et al.* Grag: Graph retrieval-augmented generation. arXiv preprint arXiv:2405.16506 (2024).

32. Yan, S.-Q., Gu, J.-C., Zhu, Y. & Ling, Z.-H. Corrective retrieval augmented generation. arXiv preprint arXiv:2401.15884 (2024).

33. Asai, A., Wu, Z., Wang, Y., Sil, A. & Hajishirzi, H. Self-rag: Learning to retrieve, generate, and critique through self-reflection. In *International Conference on Learning Representations* (2024).

34. Sun, J. *et al.* Think-on-graph: Deep and responsible reasoning of large language model on knowledge graph. In *International Conference on Learning Representations* (2024).

35. Zhao, R., Zhao, F., Wang, L., Wang, X. & Xu, G. Kg-cot: Chain-of-thought prompting of large language models over knowledge graphs for knowledge-aware question answering. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, 6642–6650, DOI: 10.24963/ijcai.2024/734 (2024).

36. Ma, S. *et al.* Think-on-graph 2.0: Deep and faithful large language model reasoning with knowledge-guided retrieval augmented generation. In *International Conference on Learning Representations* (2025).

37. Mavromatis, C. *et al.* Gnn-rag: Graph neural retrieval for efficient large language model reasoning on knowledge graphs. In *Findings of the Association for Computational Linguistics: ACL 2025*, 16682–16699, DOI: 10.18653/v1/2025.findings-acl.856 (2025).

38. Chen, J. *et al.* Gril: Knowledge graph retrieval-integrated learning with large language models. In *Findings of the Association for Computational Linguistics: EMNLP 2025* (2025).

39. Xu, D. *et al.* Harnessing large language models for knowledge graph question answering via adaptive multi-aspect retrieval-augmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, 25570–25578, DOI: 10.1609/aaai.v39i24.34747 (2025).

40. Tian, S. *et al.* A systematic exploration of knowledge graph alignment with large language models in retrieval augmented generation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, 25291–25299, DOI: 10.1609/aaai.v39i24.34716 (2025).

41. Qin, Y. *et al.* Semantic-aware fusion of mineral exploration knowledge streams towards dynamic geological knowledge graphs. *Minerals* **15**, 1257, DOI: 10.3390/min15121257 (2025).

42. Wu, J., Zhao, Y., Liao, Z. & Liu, M. A survey on knowledge graph-augmented large language model reasoning: Theoretical challenges, technical pathways, and evolutionary logic. *Knowl. Inf. Syst.* **68**, 261, DOI: 10.1007/s10115-026-02880-5 (2026).

43. Shi, Y. *et al.* Plan-answer-refine-on-graph: Structured planning and self-refinement for large language model reasoning on knowledge graphs. In *International Conference on Learning Representations* (2026).

44. Park, M., Yang, H., Kim, J., Park, K. & Kim, H. Prograg: Hallucination-resistant progressive retrieval and reasoning over knowledge graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 40, 32674–32682, DOI: 10.1609/aaai.v40i39.40545 (2026).

45. Kipf, T. N. & Welling, M. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations* (2017).

46. Schlichtkrull, M. *et al.* Modeling relational data with graph convolutional networks. In *The Semantic Web*, 593–607, DOI: 10.1007/978-3-319-93417-4_38 (2018).

47. Hogan, A. *et al.* Knowledge graphs. *ACM Comput. Surv.* **54**, 71:1–71:37, DOI: 10.1145/3447772 (2021).

48. Yao, S. *et al.* React: Synergizing reasoning and acting in language models. In *International Conference on Learning Representations* (2023).

49. Li, Y. *et al.* A survey of rag-reasoning systems in large language models. In *Findings of the Association for Computational Linguistics: EMNLP 2025*, 12120–12145, DOI: 10.18653/v1/2025.findings-emnlp.648 (2025).

50. Mason, G. T. & Arndt, R. E. Mineral resources data system (mrds). Tech. Rep. Data Series 20, U.S. Geological Survey (1996). DOI: 10.3133/ds20.

51. Earth Resources Observation and Science (EROS) Center. Landsat 8–9 operational land imager / thermal infrared sensor level-2, collection 2. U.S. Geological Survey data release, DOI: 10.5066/P9OGBGM6 (2020).

52. European Space Agency. Copernicus sentinel-2 mission: Level-2a surface reflectance products. Copernicus Data Space Ecosystem (2022).

53. McCafferty, A. E. *et al.* Prospectivity models: Clastic-dominated and mississippi valley-type geotiff grids for the united states, canada, and australia. U.S. Geological Survey data release, DOI: 10.5066/P970GDD5 (2025).

54. El-Wardany, A. K. *et al.* Remote sensing-based machine learning techniques for mapping gold-mineralized alteration zones in the fatira mine area, egypt. *Acta Geol. Sinica - Engl. Ed.* **99**, 1196–1223, DOI: 10.1111/1755-6724.15333 (2025).