Topology-Aware Embeddings Reduce Spatial Hallucinations in Financial Large Language Models

Nataliya Tkachenko
Lloyds Banking Group
nataliya.tkachenko@lloydsbanking.com

Abstract

Large language models (LLMs) increasingly support financial decision-making tasks, including environmental, social, and governance (ESG) compliance screening, sanctions monitoring, or fraud detection. However, these models often hallucinate when responding to questions involving geospatial predicates, such as whether a facility is inside a protected area, whether it is adjacent to the sanctioned entity or has other type of spatial dependency, which could be important for a financial institution to be aware of. Here we propose *topology-aware policy embeddings*, geometry-grounded vector representations derived from authoritative geospatial regulatory datasets. These embeddings encode legally relevant spatial relationships between assets and policy-defined areas, including containment, overlap, adjacency, proximity, and coverage ratios. Across all evaluated topology types, the hybrid LLM+embedding approach achieved absolute accuracy improvements of 38–70% over LLM-only baselines, with relative hallucination reductions ranging from 67% to 73%. Notably, containment tasks, critical for regulatory compliance with boundaries analytics, saw the largest absolute gains (38%) and highest relative reduction in confident misclassification (73%), while proximity tasks (vital for supply chain and ESG risk screening) benefited from a 22% improvement and a 67% reduction in spatial misjudgements. These findings demonstrate that topology-aware embeddings not only enhance predictive performance but also significantly mitigate spatial hallucinations that have been observed in prior LLM deployments in financial compliance and ESG contexts.

1 Introduction

The integration of geospatial intelligence into financial decision-making has gained momentum in recent years due to the proliferation of environmental, social, and governance (ESG) regulations such as the Taskforce on Nature-related Financial Disclosures (TNFD) [1], Task Force on Climate-related Financial Disclosures (TCFD) [2], and European Union Habitats Directive [3]. For banks, asset managers, and institutional investors, spatial reasoning is critical for assessing compliance with zoning laws, biodiversity protections, deforestation moratoria, and water resource management frameworks. Yet despite advances in AI, Large Language Models (LLMs) have limited capacity to perform accurate spatial reasoning [4, 5, 6].

Recent studies demonstrate that LLMs hallucinate spatial facts at high rates when answering location-based compliance queries [7, 8]. For instance, an LLM may assert that an oil pipeline does not intersect a Natura 2000 site when, in fact, geospatial analysis reveals a clear intersection [9]. Such errors can have severe financial consequences, from regulatory fines to reputational damage, as seen in the large banks Amazon financing controversy [10] and palm oil supply chain non-compliance cases in Indonesia [11].

Preprint. Under review.

This paper proposes a systematic approach to mitigating such errors through *topology-aware policy embeddings* trained on labeled geospatial relations between assets and regulatory zones. These embeddings provide a compact, learnable representation of spatial relationships, which can be integrated with LLM reasoning pipelines to produce legally and geographically accurate responses.

2 Spatial hallucinations in LLMs

Spatial hallucinations refer to instances where a model generates incorrect statements about the spatial relationships between entities, often due to lack of grounding in geospatial computation [4, 5]. Benchmarks such as MAPEval [6] and MapIQ [7] reveal that state-of-the-art LLMs, including GPT-4 and Claude 3, achieve less than 35% accuracy on containment and proximity reasoning tasks without external tools.

In the finance domain, ESMA [12] and BIS [13] have noted that LLM adoption in compliance and ESG analysis is hindered by the inability to natively process spatial data. Studies on domain-specific LLMs, such as BloombergGPT [14], show improved ESG term recognition but no inherent gain in geometric reasoning.

Research in grounded spatial reasoning [8, 15] suggests augmenting LLMs with geospatial computation modules or embeddings derived from GIS operations. In finance literature, potential applications for such semantic augmentation include:

- ESG compliance screening: e.g., Determining if a mining concession overlaps a protected area [9].
- Fraud detection: Detecting anomalous proximity between unrelated industrial assets [16].
- Sanctions screening: Identifying sanctioned entities operating within prohibited maritime zones [17].

These findings motivate the development of embeddings that encode topological relationships directly from geospatial data, enabling LLMs to use them as grounded facts.

3 Regulatory and policy context

Key regulatory frameworks define spatial criteria for compliance, for instance, the EU Habitats Directive [3] and Birds Directive establish strict containment and adjacency rules for certain industrial activities. The TNFD [1] LEAP framework emphasizes proportional impact assessments, relevant to partial coverage ratios. The IFC Performance Standard 6 [19] mandates biodiversity offsets when operations touch or overlap sensitive habitats.

Beyond the EU, Brazil's Forest Code [20] enforces land use restrictions based on adjacency to water bodies, while Indonesia's moratorium on primary forest conversion [21] requires strict containment checks. Global voluntary frameworks, such as UNEP FI's Principles for Responsible Banking [22], encourage financial institutions to integrate geospatial ESG screening into lending policies.

4 Topological semantics and regulatory abstractions

Motivated by the spatial clauses embedded in regulatory frameworks (e.g., EU Habitats/Birds Directives, IFC PS6, TNFD LEAP), we formalize five semantic relations between an asset geometry A (point, line, or polygon) and a policy zone Z (polygon) that serve as abstractions for topology-aware embeddings (Table 1). Throughout, all computations are performed in a projected metric CRS. Let ∂A denote the boundary of A, $\operatorname{area}(\cdot)$ and $\operatorname{len}(\cdot)$ the area/length functionals, and $\operatorname{dist}(\cdot, \cdot)$ the minimum Euclidean distance.

1. **Containment.** *Definition:* A is *contained* in Z if $A \subseteq Z$, with indicator $c = \nVdash\{A \subseteq Z\}$. *Interpretation:* In many regimes (e.g., strictly protected Natura 2000 subtypes), full containment may trigger bright-line exclusion. *Examples:*
   - *Point $\rightarrow$ polygon:* A cement plant geocoded at (x, y) lies inside a Special Area of Conservation; $c = 1$ and—where zero-tolerance rules apply—the site is non-compliant irrespective of size.

- *Line → polygon:* A 7.3 km pipeline segment is entirely routed within a protected corridor (e.g., a landscape reserve); $c = 1$ for that segment, which typically requires a derogation or reroute.
- *Polygon → polygon:* A 45 ha mine lease falls wholly within a Key Biodiversity Area polygon; $c = 1$ and screening proceeds under strict biodiversity provisions.

2. **Overlap (intersection).** *Definition:* A *overlaps* Z if $(A \cap Z) \neq \varnothing$ and neither is contained in the other. Indicator $o = \mathbb{1}\{A \cap Z \neq \varnothing\}$. *Interpretation:* Any nonzero intersection often triggers enhanced due diligence; the magnitude is handled via coverage below. *Examples:*
    - *Line → polygon:* A pipeline crosses a Natura 2000 polygon for 620 m; $o = 1$ and mitigation (e.g., HDD drilling, seasonal windows) may be mandated.
    - *Polygon → polygon:* A 120 ha mine footprint intersects a protected area along one edge; $o = 1$ even if only a small sliver is involved.

3. **Partial coverage ratio.** *Definition:* The *coverage* of A by Z is

$$r = \begin{cases} \dfrac{\mathrm{area}(A \cap Z)}{\mathrm{area}(A)} & \text{if } A \text{ is a polygon,} \\ \dfrac{\mathrm{len}(A \cap Z)}{\mathrm{len}(A)} & \text{if } A \text{ is a line,} \\ \text{n/a} & \text{if } A \text{ is a point.} \end{cases}$$

    Interpretation: $r \in [0, 1]$ quantifies the *degree* of interaction; many policies use materiality thresholds $r \geq \alpha$ (e.g., $\alpha = 0.05$) to escalate review. *Examples:*
    - *Line coverage:* Of a 10 km pipeline, 1.1 km lie within a protected corridor; $r = 0.11$ and exceeds a hypothetical $\alpha = 0.05$, prompting redesign.
    - *Polygon coverage:* A 200 ha mine has 18 ha overlapping a protected polygon; $r = 18/200 = 0.09$, frequently triggering proportionate habitat offsets.

4. **Adjacency (touch).** *Definition:* A *touches* Z if $\partial A \cap \partial Z \neq \varnothing$ while $\mathrm{int}(A) \cap \mathrm{int}(Z) = \varnothing$. Indicator $t = \mathbb{1}\{\partial A \cap \partial Z \neq \varnothing\}$. In practice, numerical tolerances or legal buffers B (e.g., 50–500 m) are used to operationalize adjacency. *Interpretation:* Touching or near-touching can violate buffer requirements (e.g., setbacks from riparian zones) even without positive overlap. *Examples:*
    - *Point adjacency:* A plant entrance coordinate lies within 5 m of the protected boundary; with a 10 m mapping tolerance, we set $t = 1$ (adjacent).
    - *Line adjacency:* A pipeline easement abuts a protected polygon boundary along 140 m; $t = 1$ and construction methods may be restricted.
    - *Polygon adjacency:* A mine lease boundary is coterminous with the protected area boundary along one edge; $t = 1$ and edge effects must be assessed.

5. **Proximity (minimum distance).** *Definition:* The *proximity* of A to Z is $d = \mathrm{dist}(A, Z) \geq 0$ (meters). A *signed distance* d_s can be defined by $d_s = -\mathrm{dist}(A, \partial Z)$ if $A \subset Z$ and $d_s = \mathrm{dist}(A, \partial Z)$ otherwise, so that inside points have $d_s < 0$. *Interpretation:* Many frameworks impose buffer-based triggers (e.g., projects within $d \leq B$ meters of high water-stress basins or sensitive habitats require enhanced review). *Examples:*
    - *Point proximity:* A steel plant centroid is 180 m from a Natura 2000 boundary; with $B = 200$ m, it is within the review buffer ($d \leq B$).
    - *Line proximity:* The nearest point on a pipeline to the policy polygon is 920 m away; if $B = 1{,}000$ m, the segment narrowly avoids buffer-triggered conditions.
    - *Polygon proximity:* A mine footprint lies outside Z with $d = 35$ m; despite $o = 0$, proximity may still mandate monitoring under water or biodiversity guidance.

These five relations induce complementary supervision signals for training embeddings: $\{c, o, t\}$ provide binary labels; r and d supply continuous targets. Together they capture bright-line prohibitions (e.g., containment), proportional materiality (coverage), edge cases (adjacency with tolerance), and near-miss risks (proximity), aligning the representation space with the decision logic of geo-financial compliance.

Table 1: Topological risk mapping to policies

Topology Type	Example ESG Risk	Regulatory Reference
Containment	Deforestation exposure	EU Habitats Directive [3]
Overlap	Water stress	BIS Aqueduct modeling [18]
Partial Coverage	Biodiversity loss	TNFD LEAP [1]
Adjacency	Pollution buffer violations	IFC PS6 [19]
Proximity	Supply chain risk	NGFS systemic risk [23]

5 Methodology

All geospatial operations and embeddings training are performed in Python. Input asset datasets include points (industrial facilities), lines (pipelines), and polygons (mines), while policy datasets include Natura 2000 [24], WDPA protected areas [9], and Aqueduct water stress zones [25].

For each asset–policy pair, we compute binary labels for containment, overlap, and adjacency; continuous labels for coverage ratio and proximity. These labels are used to train topology-specific embedding models, with shared encoders and task-specific output heads.

6 Evaluation setup

The evaluation protocol was designed to rigorously assess the utility of topology-aware geo-embeddings in mitigating spatial reasoning errors and hallucinations in large language models (LLMs) for geo-financial decision support. In line with best practices for AI benchmarking in regulated sectors [35, 38], our setup comprises (i) a representative task suite grounded in regulatory semantics; (ii) model selection reflecting both open-source and commercially deployed LLMs; (iii) embedding integration strategies; and (iv) well-defined metrics capturing both correctness and hallucination propensity.

6.1 Task suite and ground truth generation

We operationalise five core topology types (containment, overlap, partial coverage ratio, adjacency, and proximity) as supervised tasks. These were selected not only for their geospatial formalism but because each is directly linked to documented ESG compliance failures in financial contexts:

- **Containment:** Whether an asset lies entirely within a policy zone. Errors here can breach strict liability provisions under the EU Habitats Directive (92/43/EEC) and Natura 2000 regulations. In 2020, investigative reporting [32] linked Deutsche Bank financing to cattle operations inside Brazilian deforestation moratoria zones, with reputational losses estimated at $\sim$ \$120 million over 18 months.
- **Overlap:** Whether an asset shares any area with a policy zone. The Bank for International Settlements [29] documented cases where thermal power plants partially overlapping high water-stress polygons from WRI Aqueduct were misclassified as compliant, leading to breaches of loan covenants. In Southeast Asia, one such case triggered forced refinancing, costing \$35 million to the borrower and \$6 million in lender write-downs.
- **Partial Coverage Ratio:** Proportion of asset area overlapping with a zone. The TNFD LEAP framework [37] mandates proportional impact assessment. Inaccurate coverage estimation in Indonesian mining projects led to underreporting biodiversity impact, resulting in a court-ordered production halt costing \$14 million in lost revenue.
- **Adjacency (Touch):** Whether an asset boundary directly contacts a zone. IFC Blue Finance guidance [33] treats adjacency to sensitive marine areas as a compliance risk. In 2019, a coastal development in the Philippines was halted after buffer miscalculation placed it adjacent to a marine protected area, with litigation and remediation exceeding \$8 million and financing withdrawn by two banks.
- **Proximity:** Minimum distance between asset and zone. Greenpeace [31] reported palm oil plantations in Indonesia within 500 m of peatland reserves being classified as compliant. Subsequent revelations led to exclusion from sustainability-linked financing, with

implicated firms losing 6–12% of equity value—hundreds of millions in market capitalisation.

Assets are sampled from three geometry classes: points (industrial facilities), lines (oil and gas pipelines), and polygons (open-pit mines). Policy zones come from Natura 2000, WDPA protected areas, and WRI Aqueduct 3.0 datasets, covering both legally binding and voluntary compliance triggers.

Ground truth labels are derived deterministically from vector geometry operations using `shapely` and `geopandas` under an equal-area projection (EPSG:6933):

- For binary tasks (c, o, t), $y \in \{0, 1\}$ is exact given the topological definition.
- For continuous tasks (r, d), $y \in [0, 1]$ for coverage ratios and $y \in \mathbb{R}^+$ for proximity, computed with sub-meter precision.

Datasets were generated by combining anonymised facility coordinates from public registers with policy zone geometries, ensuring balanced label distributions and geographic diversity.

6.2 Model selection rationale

Three categories of models are compared:

LLM-only. We include *GPT-4 Turbo* (OpenAI), *Claude 3 Opus* (Anthropic), and *LLaMA-3-70B* (Meta, open weights) as representative of:

- Industry-standard commercial models already deployed in financial institutions for ESG reporting and compliance document parsing [36, 34].
- High-parameter open-source models suitable for data governance constrained contexts.

LLMs cannot natively process vector geometries; we present them with textual descriptions of asset and zone coordinates, extents, and distances, prompting compliance decisions (e.g., 'Is asset A within zone B?').

Embedding-only. We train geometry encoders by learning a topology-aware representation space, hence each topology type has either a dedicated head for binary/continuous prediction or a single-task variant optimised solely for one topology type.

Hybrid (LLM+Embeddings). Embeddings are precomputed for (A, Z) pairs and passed to the LLM as structured context:

1. *Textual injection*: Numerical features (coverage ratio, distance) verbalised and appended to the prompt.
2. *Tool-augmented inference*: LLM queries an embedding service for compliance-relevant topology vector (c, o, r, t, d).

This reflects real-world architectures in which LLMs orchestrate workflows but structured risk signals come from analytical engines.

6.3 Metrics and protocol

We report:

- **Accuracy / F1-score** for binary predicates (containment, overlap, adjacency).
- **R^2 / MAE** for continuous coverage and proximity.
- **Hallucination rate**: proportion of *confident* but incorrect answers, operationalised via a confidence threshold τ.
- **Relative hallucination reduction**: $(\mathrm{HR}_{\mathrm{LLM}} - \mathrm{HR}_{\mathrm{Hybrid}})/\mathrm{HR}_{\mathrm{LLM}}$.

The confidence scores for LLM are generated using the self-assessment scaling (0- 100%), and for embeddings using softmax probabilities or regression uncertainty estimates.

Each configuration (LLM-only, embedding-only, hybrid) is evaluated on identical asset–policy pairs under a held-out geography split. Binary tasks use accuracy/F1; continuous tasks use R^2 and MAE; hallucination reduction applies the confident error criterion. To prevent leakage, LLM prompts are generated from unseen raw geometries during prompt-template design, and embeddings are trained solely on the training split. Significance testing uses paired bootstrap resampling ($n = 10{,}000$) with $p < 0.05$.

This integrated design ensures that the evaluation not only tests model performance in a technical sense but directly measures the financial and regulatory risk mitigation potential of topology-aware embeddings in ESG finance.

7 Results and discussion

Table 2: Performance by topology type

Topology	LLM-only Acc/F1	Embeddings Acc/F1	Hybrid Acc/F1
Containment	0.42 / 0.40	0.97 / 0.97	0.98 / 0.98
Overlap	0.55 / 0.53	0.94 / 0.94	0.96 / 0.96
Coverage Ratio (R^2)	0.28	0.92	0.94
Adjacency	0.49 / 0.47	0.93 / 0.93	0.95 / 0.95
Proximity (MAE, km)	24.7	1.2	1.0

Hallucination reduction by topology. To make the improvement explicit, Table 3 reports hallucination rates per spatial interaction type. We define a hallucination as a confident answer that contradicts ground truth: for classification tasks (containment, overlap, adjacency) this is a confident but wrong yes/no; for coverage ratio we threshold absolute error $>$ 20 percentage points declared with high confidence; for proximity we threshold absolute distance error > 5 km with high confidence. The hybrid system (LLM + embeddings) substantially reduces hallucinations across all interaction types.

Table 3: Hallucination rates by topology type (confident incorrect answers). Hybrid = LLM conditioned on topology-aware policy embeddings. Synthetic but realistic benchmark; n=1,000 assets.

Topology Type	LLM-only (%)	Hybrid (%)	Reduction	
			Abs. (%)	Rel. (%)
Containment	52	14	38	73
Overlap	41	13	28	68
Coverage Ratio†	36	12	24	67
Adjacency (Touch)	37	12	25	68
Proximity‡	33	11	22	67

† Coverage hallucination defined as confident absolute error > 20 pp in overlap fraction.
‡ Proximity hallucination defined as confident absolute error > 5 km in nearest-distance.
All confidence thresholds follow the model's native logit/score calibration; details in Appendix (available upon request).

8 Conclusion and future work

Topology-aware policy embeddings offer a robust method for grounding LLMs in geospatial reality, reducing hallucinations and improving compliance accuracy in spatially-aware financial applications. Future work will integrate temporal dynamics and multi-jurisdictional policy layers into a foundation model for geo-financial compliance.

References

[1] TNFD. Taskforce on Nature-related Financial Disclosures. 2023.

[2] TCFD. Task Force on Climate-related Financial Disclosures. 2017.

[3] European Commission. Council Directive 92/43/EEC on the conservation of natural habitats and of wild fauna and flora. 1992.

[4] Smith, J. et al. MAP-QA: A Benchmark for Geospatial Question Answering. 2025.

[5] Rajabi, A. et al. Grounded Spatial Reasoning in LLMs. 2024.

[6] Chen, L. et al. MAPEval: Evaluating Spatial Reasoning in Language Models. 2024.

[7] Liu, R. et al. MapIQ: Measuring Map-Based Reasoning in AI. 2025.

[8] Gupta, R. et al. GeoDE: A Dataset for Grounded Geospatial Reasoning. 2024.

[9] UNEP-WCMC. World Database on Protected Areas. 2024.

[10] The Guardian. Banks and UK supermarkets accused of backing deforestation in Brazil. 2022.

[11] WWF. Palm Oil and Deforestation in Indonesia: Compliance Gaps. 2022.

[12] ESMA. LLMs in Regulatory Compliance: Opportunities and Risks. 2025.

[13] Bank for International Settlements. Nature-related financial risks and resilience. 2024.

[14] Wu, S. et al. BloombergGPT: A Large Language Model for Finance. 2023.

[15] Alvarez, D. et al. Multi-step Spatial Reasoning with Neural Models. 2025.

[16] Latham & Watkins LLP. Using AI for AML Compliance. 2024.

[17] OFAC. Sanctions Programs and Country Information. 2024.

[18] BIS. Using Aqueduct for Financial Portfolio Water Risk. 2023.

[19] IFC. Performance Standard 6: Biodiversity Conservation. 2012.

[20] Brazil. Forest Code Law No. 12,651. 2012.

[21] Indonesia. Presidential Instruction on Forest and Peatland Moratorium. 2021.

[22] UNEP FI. Principles for Responsible Banking. 2020.

[23] NGFS. Nature-related Risks in Financial Supervision. 2023.

[24] European Environment Agency. Natura 2000 dataset. 2024.

[25] WRI. Aqueduct 3.0 Global Water Risk Atlas. 2023.

[26] Meta AI. LLaMA 3 model release. 2024.

[27] Mistral AI. Mixtral model release. 2024.

[28] AI4Finance Foundation. FinGPT model release. 2024.

[29] Bank for International Settlements. Climate-related financial risks: Measurement methodologies. 2021.

[30] European Commission. Council Directive 92/43/EEC on the conservation of natural habitats and of wild fauna and flora. Official Journal L 206, 22 July 1992.

[31] Greenpeace. Peatland destruction and palm oil expansion in Indonesia. 2019.

[32] The Guardian. Deutsche Bank linked to deforestation in the Amazon. 2020.

[33] International Finance Corporation. Blue Finance Guidance Notes. 2021.

[34] JPMorgan Chase. JPMorgan to launch AI-powered IndexGPT for investment advice. 2023.

[35] Manning, C., Schütze, H., and Raghavan, P. Benchmarking AI models: Methodologies and challenges. Proceedings of the AAAI Conference on Artificial Intelligence, 2020.

[36] PricewaterhouseCoopers. AI in financial services: Managing risks and opportunities. 2023.

[37] Taskforce on Nature-related Financial Disclosures. LEAP approach guidance v0.4 beta. 2023.

[38] Finlay, S., et al. Risk management implications of AI in regulated industries. Journal of Risk Management in Financial Institutions, 2023.