Qualitative Reasoning and LLMs: Experiments about integrated spatial reasoning using a relation algebra and Graph-based RAG

Reinhard Moratz[1] , **Niklas Daute**[1] , **James Ondiecki**[1] and **Markus Kattenbeck**[2]
Ioannis Giannopoulos[2]
[1]Geoinformatik University of Münster, Germany
[2]Geoinformation, TU Wien, Austria
{firstname.lastname}@uni-muenster.de, {firstname.lastname}@tu-wien.ac.at

Abstract

This paper deals with improving the capabilities of Large Language Models to provide route instructions for pedestrian wayfinders by means of qualitative spatial relations. To this end we use Retrieval-Augmented Generation (RAG) to support a large language model with context information based on the specific query. We assess the impact the added information has on model performance for generating pedestrian route instructions. Our findings encourage further integration of qualitative spatial data into large language model applications, potentially benefiting areas such as digital navigation aids, smart city tools, and accessibility technologies.

1 Introduction

There is growing interest (see [Roberts *et al.*, 2023]) in the geospatial community in the spatial reasoning capabilities of Large Language Models (LLMs). While LLMs perform well across many geospatial discourse domains [Roberts *et al.*, 2023], their performance is much weaker in the domain of navigation assistance (see also [Claramunt *et al.*, 2026] for a very recent overview of the limitations of LLMs). This limitation is particularly evident (see our companion website at https://cloud.geo.tuwien.ac.at/s/faXWE45pPPDfDgY, documenting the weak performance for route instructions for pedestrians) in tasks that require precise spatial relations, such as turn-by-turn route instructions or the interpretation of local street configurations.

Consider the following example: In January 2026, we have prompted ChatGPT-4o [OpenAI, 2026] to provide us with route instructions for pedestrians from Münster central station to nearby street Hafenweg. We received the following answer(text bold-faced by the model):

Distance: ∼2.5–3 km
Estimated walking time: 30–35 minutes
Directions:

1. **Exit Münster Hauptbahnhof** through the main entrance onto **Willy-Brandt-Allee**.
2. **Turn right** and walk southeast on Willy-Brandt-Allee toward the city center.
3. After about **500 meters**, you'll reach **Albersloher Weg**. **Turn left** onto Albersloher Weg.
4. Continue along Albersloher Weg for about **1 km**, passing the **Aasee park area** on your right.
5. When you reach the intersection with **Hammer Straße**, **turn right**.
6. Walk along Hammer Straße for around **700 meters**. You'll pass a few residential and commercial buildings.
7. Look for **Hafenweg** on your left. **Turn left** onto Hafenweg.

You have arrived at **Hafenweg, Münster**.

The model responded quickly with a seemingly convincing set of directions for the given problem. However, upon closer inspection, it becomes evident that LLM navigation is heavily prone to hallucinations and other errors. In the first step, the model suggests leaving the central station onto "Willy-Brandt-Allee", which is a street that does not exist in Münster. Subsequently, the model suggests walking toward the city center, while the Hafenweg is actually located away from the city center. Next, turning onto the "Albersloher Weg" is suggested. While this road exists in Münster and is close to the actual destination, it does not make sense to take the road for the given navigation task. The same can be said for the following two steps, where turning onto "Aaseepark" and "Hammer Straße" is suggested, with neither of them being sensible choices for the problem, while also being completely disconnected, making navigation along the suggested route impossible. Finally, the model suggests turning onto the destination street "Hafenweg", which is the correct destination, but impossible given the previous incorrect steps. When visually comparing the route proposed by ChatGPT-4o to a route provided by Google Maps[Google, 2026], the problem becomes even more evident.

Figure 1 shows the correct route from "Start" to "Finish" in green: The route is short and requires only a few turns to reach the destination. The generated LLM route displayed in red on the map is however far from correct: Multiple disconnected segments show a complete lack of "understanding' the actual navigation task as only one of the segments is anywhere close to the actual route. Observations like these highlight the current limitations of LLMs in navigation performance, and motivated us to find approaches to boost their capabilities in this domain.

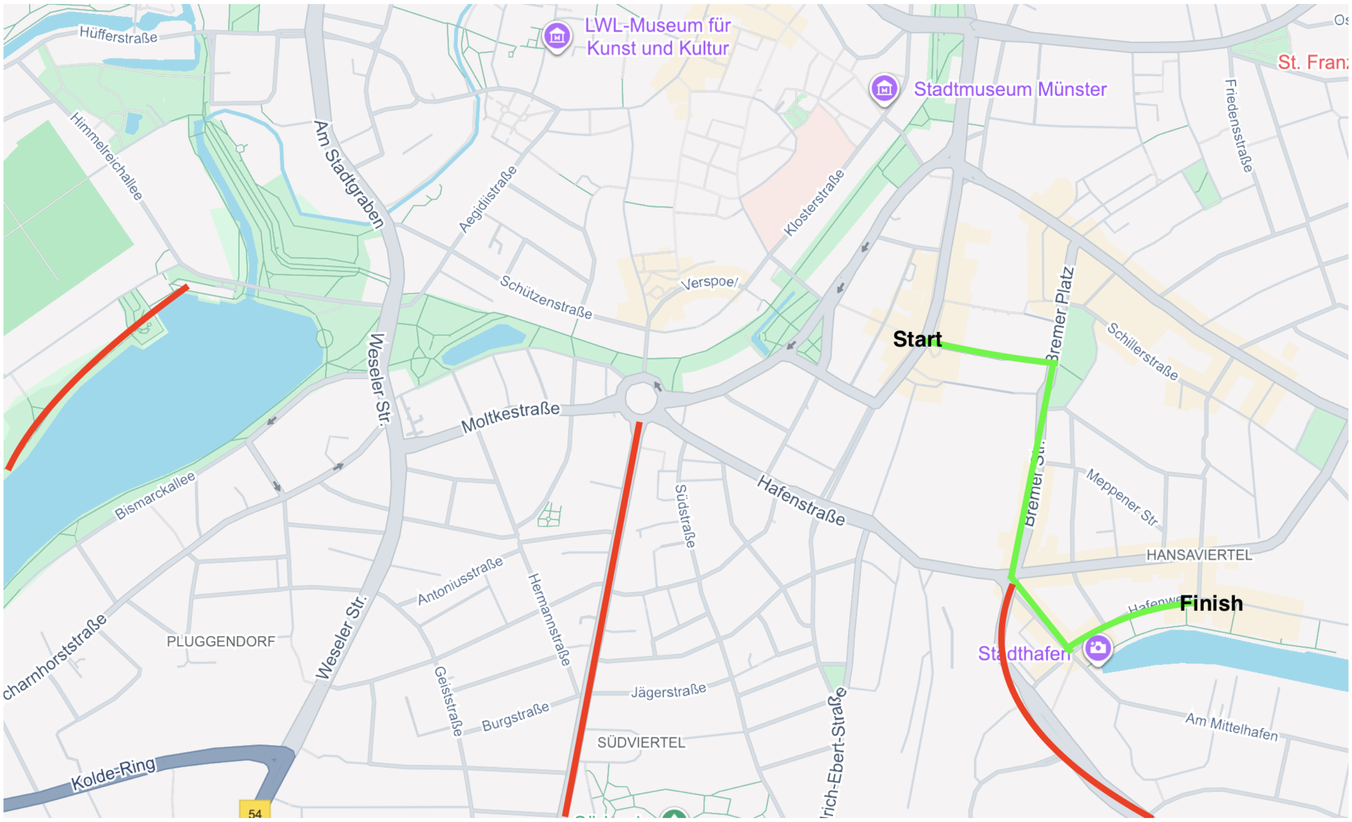

Figure 1: Map detail showing erroneous route instructions in red and the correct route in green. Basemap taken from Google Maps [Google, 2026]

A better performance is highly desirable because LLMs can tailor the actual phrasing of route instructions to an individual's needs (see, for example, [Nuhn and Timpf, 2017] for personalized landmark selection for route instructions) beyond turn-by-turn-instructions (see, e.g., [Krukar *et al.*, 2020] for the concept of orientation instructions), while providing a natural language user interface (see, for example, [Richter *et al.*, 2008] for an early attempt of a dialogue between humans and systems for route instructions and [Krieger *et al.*, 2020] for a natural language corpus providing evidence in how humans reach common ground in such situations).

To address this issue, we propose a graph-based Retrieval-Augmented Generation (Graph-RAG, see [Peng *et al.*, 2026] for a recent review of this technique) approach that integrates qualitative spatial and temporal reasoning (QSTR, [Ligozat, 2013]).

In contrast to purely textual augmentation, our approach uses structured graph representations of street networks consisting of qualitative spatial relations (see [Van de Weghe *et al.*, 2025] who provide an argument how GIScience and LLMs can be mutually beneficial). These qualitative spatial relations provide a finite and compositional representation of spatial configurations, enabling consistent reasoning over local environments. By embedding such representations into a graph-based retrieval pipeline, we enable LLMs to access and utilize structured spatial knowledge during generation.

Roberts et al. [Roberts *et al.*, 2023] demonstrate that LLMs perform substantially better with qualitative spatial descriptions than with quantitative specifications, while struggling with coordinate transformations, distance calculations, and metric spatial reasoning. This performance gap reveals why qualitative spatial representations can serve as a representation suited for extending LLM-based reasoning. Rather than expecting language models to interpret raw elevation matrices or coordinate arrays, qualitative spatial reasoning translates continuous spatial data into discrete relational vocabularies that correspond to linguistic expression. Specifically, this means that qualitative spatial reasoning can serve as an effective bridge between computational geometry and language-based reasoning systems. Complex topographic relationships can be encoded in qualitative terms that language models can process meaningfully. This emphasizes the strengths of QSTR based representation in the context of LLMs.

We, therefore, use a qualitative spatial representation framework which is based on oriented line segments (dipoles) as basic entities [Moratz *et al.*, 2011a]. In our context a *qualitative* representation provides mechanisms which characterize central essential properties of objects or configurations. Qualitative spatial spatial representations usually deal with elementary objects (e.g., positions, directions, regions) and qualitative relations between them (e.g., ”adjacent”, ”on the left of”, ”included in”). A *quantitative* representation in contrast establishes a measure in relation to a unit of measurement which has to be generally available.

A strength of qualitative spatial relations is, consequently, that they are able to express the typical relations between object classes (e.g. human concepts) in addition to the relations of specific instances of these classes. So we can say that the human heart is at the *left* side of the human body and this spatial relation typically would hold for all instances of hearts even if the metrical details differ from body to body. So often we can make relevant spatial expressions using qualitative relations where a precise metric spatial statement would impossible. There are comprehensive results from the research field of (QSTR) that offer many spatial representations which are near linguistic concepts [Ligozat, 2013; Cohn, 1997].

2 Graph-Based Retrieval Augmented Generation (Graph-RAG)

Graph-based Retrieval Augmented Generation represents an advanced evolution of traditional RAG systems that leverages knowledge graphs to enhance information retrieval and generation quality [Edge *et al.*, 2024]. A Graph-RAG enhanced language model receives not only natural language text chunks, but structured information including entity mentions and their properties, explicit relationships between entities, graph paths showing logical connections, and multi-level abstractions (detailed facts and high-level summaries).

Traditional RAG retrieves relevant text chunks from a vector database using semantic similarity, then feeds these chunks to a language model for generation. Graph-RAG enhances this by organizing information in a knowledge graph structure, where entities are nodes and relationships are edges, enabling more sophisticated retrieval strategies.

The system first constructs a knowledge graph from source documents by extracting entities (people, places, concepts, events), identifying relationships between entities, creating nodes for entities and edges for relationships, and storing both structural information and associated text content. Retrieved information is assembled into context by combining node content (entity descriptions), including edge information (relationships), incorporating graph structure (how entities connect), and adding hierarchical summaries when relevant.

A central design question in Graph-RAG is the choice of an appropriate representation for retrieval structures. Throughout this paper, we provide evidence that qualitative spatial and temporal representations (QSTR) are particularly well-suited for Graph-RAG in spatial domains. There are three different reasons for this rationale [Ligozat, 2013]:

1. Qualitative spatial calculi such as dipole calculi provide a finite set of relations, which enables efficient indexing and retrieval in graph structures.
2. In contrast to continuous geometric representations, qualitative relations reduce the search space while preserving essential spatial distinctions.
3. QSTR representations are invariant under scaling, translation, and small perturbations.

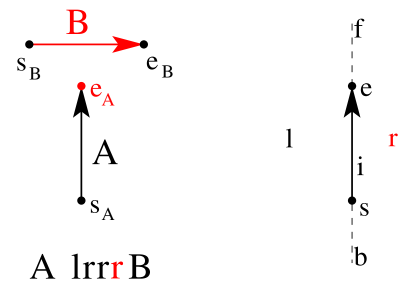

Figure 2: Orientation between two dipoles based on four dipole-point relations

Taken together, qualitative relations ensure that structurally equivalent configurations are treated uniformly. This robustness is particularly beneficial for RAG pipelines, where retrieved contexts may originate from heterogeneous or noisy data sources.

Route instructions and spatial descriptions are typically expressed in qualitative terms such as “left of”, “ahead”, or “next to”. QSTR provides a formalization of these concepts, enabling a direct mapping between retrieved graph structures and natural language generation. By constraining the retrieval step with qualitative relations, the generation process is grounded in a consistent relational structure. This reduces ambiguity and improves the reliability of generated outputs. Qualitative spatial representations combine discreteness, robustness, and compositionality. These properties make them a natural fit for Graph-RAG architectures in spatial reasoning tasks.

3 Dipole Calculi

Human natural language spatial propositions often express relative spatial positions based on reference directions derived from the shape (and function) of one of the objects involved [Levinson, 1996] (e.g., *The hill is to the left of the train*). This leads to binary relations between objects in which at least one of the objects has the feature of orientedness.

For that reason, in our conception, orientedness is an important feature of natural objects. In a corresponding qualitative calculus it is necessary to use more complex basic entities than points. One option for building more complex basic entities is to use oriented line segments (see Fig. 2) as basic entities. In this abstraction we loose the specific shape of the object, but preserve the feature of orientedness. With this approach we can design relative position calculi in which directions are expressed as binary relations [Moratz *et al.*, 2011b]. Oriented straight line segments (which were called *dipoles* by Moratz et al. [Moratz *et al.*, 2000]) may be specified by their start and end points. Then a dipole is a pair of different points in the Euclidean plane.

Using dipoles as basic building blocks, more complex objects can be constructed (e.g., polylines or polygons) in a straightforward manner. Therefore, dipoles can be used as basic units in numerous applications. To give an example, line segments are central to edge-based image segmentation and grouping in computer vision. In addition, GIS systems often have line segments as basic entities [Hoel and Samet, 1991]. Polylines are particularly interesting for representing paths in cognitive robotics [Musto *et al.*, 1999] and can serve as the geometric basis of a mobile robot when autonomously mapping its working environment [Wolter and Latecki, 2004]. Dipole calculi are qualitative calculi that abstract from metric information. They focus on directional relations, but can also be used to express certain topological relations [van Delden and Moratz, 2014].

To formally derive dipole-dipole relations we derive those relations from dipole-point relations. A dipole-point relation distinguishes between whether a point lies to the left, to the right, or at one of five qualitatively different locations on the straight line that passes through the corresponding dipole. The corresponding regions are shown on the right side of Fig. 2, see [Isli and Moratz, 1999; Ligozat, 1993; Scivos and Nebel, 2005]. Using these seven possible relations between a dipole and a point, the relations between two dipoles may be specified according to the following four relationships:

$$A \ \mathrm{R}_1 \ \mathbf{s}_B \wedge A \ \mathrm{R}_2 \ \mathbf{e}_B \wedge B \ \mathrm{R}_3 \ \mathbf{s}_A \wedge B \ \mathrm{R}_4 \ \mathbf{e}_A,$$

where $\mathrm{R_i} \in \{\mathrm{l, r, b, s, i, e, f}\}$ with $1 \leq i \leq 4$. Theoretically, this gives us 2401 relations, out of which 72 relations are geometrically possible: { rrrr, rrrl, rrlr, rrll, rlrr, rllr, rlll, lrrr, lrrl, lrll, llrr, llrl, lllr, llll, ells, errs, lere, rele, slsr, srsl, lsel, rser, sese, eses, lllb, llfl, llbr, llrf, lirl, lfrr, lril, lrri, blrr, irrl, frrr, rbrr, lbll, flll, brll, rfll, rlli, rrlf, illr, rilr, rrbl, rlir, rrfr, rrrb, ffbb, efbs, ifbi, bfii, sfsi, beie, ffbb, bsef, biif, iibf, sisf, iebe, ffff, fefe, fifi, fbii, fsei, ebis, iifb, eifs, iseb, bbbb, sbsb, ibib }.

These constitute the $\mathcal{DRA}_f$ calculus, see [Moratz *et al.*, 2000; Moratz *et al.*, 2011b] for details. For example, in Fig. 2, the relation A lrrr B holds.

3.1 Representation of street networks in the dipole calculus

There are several ways to represent street networks with qualitative spatial calculi [Chipofya *et al.*, 2016]. In our application to street networks we relax the condition for the dipoles to be straight lines. The streets have unique names (e.g. A, ... , Z in figure 3). In a pairwise manner all street segments forming an intersection have a spatial relation which can by qualitatively expressed by a $\mathcal{DRA}_f$ base relation. The relations between the streets in figure 3 are the following: A rlir B, A rllr C, A rlli D, E lril F, E lrrl G, E lrri H, ...

In addition to the orientation information at crossings we also represent the cardinal direction between the dipoles representing the streets. We use the foot points of the shortest segment connecting the two dipoles. Then the cardinal direction between these foot points is represented using Andrew Frank's cardinal direction calculus [Frank, 1991]. Based on the definitions of foot points and acceptance sectors in the cardinal direction calculus, it follows that streets which intersect are assigned the cardinal direction *same* (see Figure 4).

3.2 Implementation specifics for current use case

In our LLM-based application we do not use the standard relation names for the Dipole relations, as they appear exclusively in research papers and are, hence, insufficiently

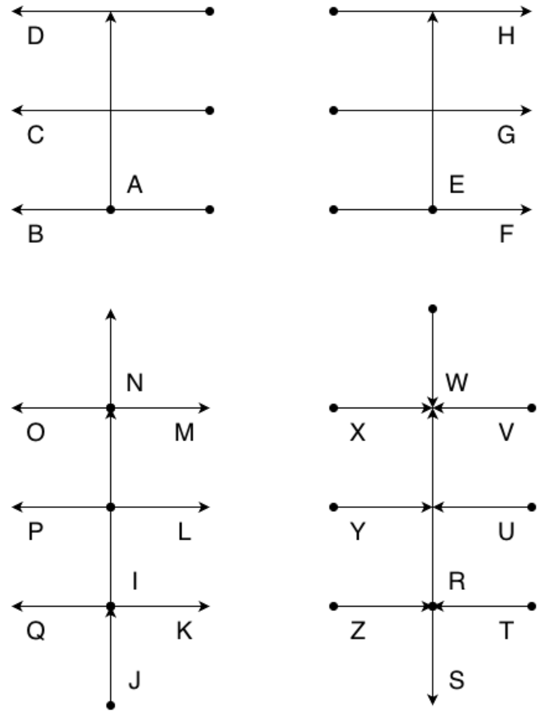

Figure 3: All qualitatively different street relations

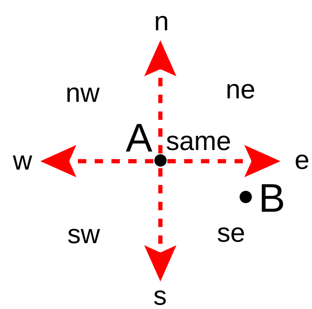

Figure 4: Cardinal Direction Calculus

represented in the training data of common LLMs. As a consequence, we designed linguistic patterns that the LLMs can directly relate to common meanings for linguistic spatial expressions. For example the dipole relation rlir is mapped on the linguistic pattern "starts at". The relation rllr is mapped on "is crossed by". We provide the full list of dipole relations and their corresponding linguistic pattern on our companion website https://cloud.geo.tuwien.ac.at/s/faXWE45pPPDfDgY.

4 Initial, small-scale evaluation in Münster and Hamburg

We test our Graph-RAG approach by example of a small area of HAMBURG, GERMANY and a slightly larger MÜNSTER, GERMANY. We base our dipole relation knowledge graph on OpenStreetMap [OpenStreetMap contributors, 2017] data using a two step process:

Figure 5: Detail of the dipole relation part of our knowledge graph for the HAMBURG, GERMANY test area

1. We semi-automatically extracted the dipole relations and the cardinal direction relations for both areas. This means, any street with a name in OpenStreetMap is mapped on an individual dipole.
2. We stored the relations in a knowledge graph using the Neo4j software.

On our companion website https://cloud.geo.tuwien.ac.at/s/faXWE45pPPDfDgYwe provide the verbalization (e.g. list of linguistic patterns/human natural language sentences corresponding to a list of triples of two street names and the linguistic pattern of their spatial relation) of the qualitative relation knowledge graph that corresponds to our small test area of HAMBURG, GERMANYand MÜNSTER, GERMANY.

4.1 Results

In total, 240 individual trials were conducted across various LLMs, namely Gemini Pro 2.5[Google DeepMind, 2025], Claude Sonnett 4.5[Anthropic, 2025], ChatGPT 4o[OpenAI, 2026] , to find out whether the inclusion of qualitative geographic context helps improve the navigation performance of LLMs. We, first, provide an overview of the obtained results obtained and, second, break the results down by different factors such as test city or LLM tested.

In total, 240 navigation tasks were answered by the LLMs and manually labeled as either correct or incorrect. We compared two groups: We refer to the *control group* to the subset of trials which we performed without Graph-RAG, whereas the *test group* comprises those cases in which we provide context to the LLM using our dipole approach. In each group, the exact same 120 navigation tasks were executed. Table 1 summarizes the overall results of our experiments.

There is an obvious difference w.r.t. to the factual correctness of rout instructions between the control and test group.

No trial in the control group could be labeled as successful, i.e., there was no case in which the resulting route instruction returned was factually correct. Consequently, all trials in this group were labeled as failures, leading to a task success rate of 0%. In the test group, however, 75 out of 120 trials were assessed as successful, resulting in a task success rate of 62.5%.

Comparing the test group performance across the two test cities, yields several key observations. First, the task success rate was increased in both cities compared to the control group which had a success rate of 0% in both cities. In the smaller HAMBURG, GERMANY test area, the task success rate increased to 86.6% whereas in the larger Münster test area it reached just 38.3%. We assume that the reason for this difference is the difference in size of both data sets. The MÜNSTER, GERMANY data set contains 128 individual streets and the average distance for the routes is 1272 meters. The HAMBURG, GERMANY data set contains 38 individual streets and the average distance for the routes is 899 meters. Apparently, we see a decline in success rate as route length increases in which case the configuration of more connected streets is required. It seems like our Graph-RAG approach gets less successful with longer routes which need the configuration of more connecting streets. We will investigate this in more detail in the future. Second, comparing the test group performance across the three tested LLMs, there are noteworthy differences in their performance increase. While all models performed better compared to their control group performance of 0%, the model which performed best across all trials in the test group was Google's Gemini 2.5 Pro. It achieved a task success rate of 70%, answering 28 out of 40 navigation tasks correctly. The next best performing model was Claude 4.5, which achieved a slightly lower task success rate of 65% (26 out of 40 navigation tasks). In total, the worst performing model in the test group was GPT-4o, with 21 correct answers out of 40 possible, amounting to a task success rate of 52.5%.

Group	# Experiments	# Successful	# Failed	Success Rate (%)
Control Group	120	0	120	0%
Test Group	120	75	45	62.5%

Table 1: Overview of experiment results in control and test groups with success and failure counts and complimentary success rates.

4.2 Limitations

So far in our first experiments we only worked with smaller data for which we could manually verify the consistency. We, therefore, can currently not judge how well our suggested approach scales. A second, related, limitation is related to the network similarity of the environments tested. Both of them are ”organically grown”. Thirdly, we have not tested current state-of-the-art reasoning models.

5 Conclusion and Future Work

In summary, our experiments revealed that LLM street navigation performance with additional qualitative geographic context was substantially higher than without it. This effect was not limited to a specific city or model, but could be observed across all combinations tested in this study. In this sense our approach helps reducing the discursive and conceptual risks of generative AI [Claramunt *et al.*, 2026, p. 5] by following their suggestion of incorporating spatial logic into LLMs. Having said this, our results suggest at least four-different strands of future work: 1) We will investigate the correlation between factual correctness of the instruction and the straight-line distance between start and destination of the route. 2) As we have found substantial differences w.r.t. the performance of different LLMs, we will look for ways to use our scenario as benchmark for the spatial reasoning capability of LLMs, including more fine-grained assessment of different error categories and severity classes. 3) We enhance the scalability of our approach. To this end, will be to improve our qualitative abstraction algorithm to extract larger data from OpenStreetMap automatically. This will allow us to assess the reasons why the scale of the dipole data seems to matter. In addition to that, we will develop an automatic verification of suggested routes. 4) We will investigate whether prompt-based training enables LLMs to adopt formal Dipole notation, comparing its directional generation performance against linguistic mappings, and evaluating the semantic alignment between linguistic terms and formal relations through point-line configuration tasks and characterisation queries.

References

[Anthropic, 2025] Anthropic. Claude sonnet 4.5. https://www.anthropic.com/claude, September 2025. Large language model.

[Chipofya *et al.*, 2016] Malumbo Chaka Chipofya, Carl Schultz, and Angela Schwering. A metaheuristic approach for efficient and effective sketch-to-metric map alignment. *International Journal of Geographical Information Science*, 30(2):405–425, 2016.

[Claramunt *et al.*, 2026] Christophe Claramunt, Lars De Sloover, and Nico Van de Weghe. Computational and urban science amid generative ai: the enduring relevance of mcdermott’s critique (opinion paper). *Computational Urban Science*, 6:29, 4 2026.

[Cohn, 1997] A.G. Cohn. Qualitative spatial representation and reasoning techniques. In G. Brewka, C. Habel, and B. Nebel, editors, *KI-97: Advances in Artificial Intelligence*, Lecture Notes in Artificial Intelligence, pages 1–30. Springer-Verlag, Berlin, 1997.

[Edge *et al.*, 2024] Darren Edge, Ha Trinh, Newman Cheng, Joshua Bradley, Alex Chao, Apurva Mody, Steven Truitt, Dasha Metropolitansky, Robert Osazuwa Ness, and Jonathan Larson. From local to global: A graph rag approach to query-focused summarization. *arXiv preprint arXiv:2404.16130*, 2024.

[Frank, 1991] A. Frank. Qualitative spatial reasoning with cardinal directions. In *Proceedings of 7th Österreichische Artificial-Intelligence-Tagung*, pages 157–167, Berlin, 1991. Springer.

[Google DeepMind, 2025] Google DeepMind. Gemini 2.5 Pro, 2025. Large Language Model.

[Google, 2026] Google. Google maps, 2026. Accessed: May 20, 2026.

[Hoel and Samet, 1991] E. G. Hoel and H. Samet. Efficient processing of spatial queries in line segment databases. In *Proceedings of the 2nd Symp. on Large Spatial Databases (SSD’91)*, pages 237–255, Zürich, 1991.

[Isli and Moratz, 1999] A. Isli and R. Moratz. *Qualitative Spatial Representation and Reasoning: Algebraic Models for Relative Position*. Universität Hamburg, FB Informatik, Technical Report FBI-HH-M-284/99, Hamburg, 1999.

[Krieger *et al.*, 2020] Pauline Krieger, Markus Kattenbeck, Bernd Ludwig, Johannes Helmbrecht, and Ioannis Giannopoulos. Hey you! let’s talk. dialogue-initiatives revisited for wayfinding instructions. *AGILE: GIScience Series*, 1:1–19, 2020.

[Krukar *et al.*, 2020] Jakub Krukar, Vanessa Joy Anacta, and Angela Schwering. The effect of orientation instructions on the recall and reuse of route and survey elements in wayfinding descriptions. *Journal of Environmental Psychology*, 68, 2020.

[Levinson, 1996] S. C. Levinson. Frames of Reference and Molyneux’s Question: Crosslinguistic Evidence . In P. Bloom, M. Peterson, L. Nadel, and M. Garrett, editors, *Language and Space*, pages 109–169. MIT Press, Cambridge, MA, 1996.

[Ligozat, 1993] G. Ligozat. Qualitative triangulation for spatial reasoning. In *COSIT 1993*, Berlin, 1993. Springer.

[Ligozat, 2013] Gérard Ligozat. *Qualitative spatial and temporal reasoning*. John Wiley & Sons, 2013.

[Moratz *et al.*, 2000] R. Moratz, J. Renz, and D. Wolter. Qualitative spatial reasoning about line segments. In Horn. W., editor, *ECAI 2000. Proceedings of the 14th European Conference on Artifical Intelligence*, Amsterdam, 2000. IOS Press.

[Moratz *et al.*, 2011a] Reinhard Moratz, Dominik Lücke, and Till Mossakowski. A condensed semantics for qualitative spatial reasoning about oriented straight line segments. *Artificial Intelligence*, 175(16-17):2099–2127, 2011.

[Moratz *et al.*, 2011b] Reinhard Moratz, Dominik Lücke, and Till Mossakowski. A condensed semantics for qualitative spatial reasoning about oriented straight line segments. *Artificial Intelligence*, 175:2099–2127, 10 2011.

[Musto *et al.*, 1999] Alexandra Musto, Klaus Stein, Andreas Eisenkolb, and Thomas Röfer. Qualitative and quantitative representations of locomotion and their application in robot navigation. In *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*, IJCAI '99, page 1067–1073, San Francisco, CA, USA, 1999. Morgan Kaufmann Publishers Inc.

[Nuhn and Timpf, 2017] Eva Nuhn and Sabine Timpf. A multidimensional model for selecting personalised landmarks. *Journal of Location Based Services*, 11:153–180, 2017.

[OpenAI, 2026] OpenAI. ChatGPT (gpt-4o). https://chat.openai.com/, 2026. Large language model.

[OpenStreetMap contributors, 2017] OpenStreetMap contributors. Planet dump retrieved from https://planet.osm.org . https://www.openstreetmap.org, 2017.

[Peng *et al.*, 2026] Boci Peng, Yun Zhu, Yongchao Liu, Xiaohe Bo, Haizhou Shi, Chuntao Hong, Yan Zhang, and Siliang Tang. Graph retrieval-augmented generation: A survey. *ACM Transactions on Information Systems*, 44:1–52, 2 2026.

[Richter *et al.*, 2008] Kai-Florian Richter, Martin Tomko, and Stephan Winter. A dialog-driven process of generating route directions. *Computers, Environment and Urban Systems*, 32:233–245, 2008. Discrete Global Grids.

[Roberts *et al.*, 2023] Jonathan Roberts, Timo Lüddecke, Sowmen Das, Kai Han, and Samuel Albanie. Gpt4geo: How a language model sees the world's geography. *arXiv preprint arXiv:2306.00020*, 2023.

[Scivos and Nebel, 2005] Alexander Scivos and Bernhard Nebel. The finest of its class: The natural point-based ternary calculus LR for qualitative spatial reasoning. In Christian Freksa, Markus Knauff, Bernd Krieg-Brückner, Bernhard Nebel, and Thomas Barkowsky, editors, *Spatial Cognition IV. Reasoning, Action, Interaction*, pages 283–303, Berlin, Heidelberg, 2005. Springer Berlin Heidelberg.

[Van de Weghe *et al.*, 2025] Nico Van de Weghe, Lars De Sloover, Anthony Cohn, Haosheng Huang, Simon Scheider, Renée Sieber, Sabine Timpf, and Christophe Claramunt. Opportunities and challenges of integrating geographic information science and large language models. *Journal of Spatial Information Science*, (30):93–116, 2025.

[van Delden and Moratz, 2014] André van Delden and Reinhard Moratz. Crossing the boundary: Two benchmarks for qualitative spatial reasoning bridging relative directions and mereotopology. In *International Conference on Spatial Cognition*, pages 299–311. Springer, 2014.

[Wolter and Latecki, 2004] Diedrich Wolter and Longin J. Latecki. Shape matching for robot mapping. In Chengqi Zhang, Hans W. Guesgen, and Wai-Kiang Yeap, editors, *PRICAI 2004: Trends in Artificial Intelligence*, pages 693–702, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg.