From Trajectories to Relations: Interpreting Overtake Manoeuvres with Large Language Models and Qualitative Spatial Representations

Jana Verdoodt
Department of Geography, Ghent University, Belgium
GeoAI Research Center, Ghent University, Belgium

Changbo Zhang
Department of Geography, Ghent University, Belgium

Lars De Sloover
Department of Geography, Ghent University, Belgium
GeoAI Research Center, Ghent University, Belgium

Haosheng Huang
Department of Geography, Ghent University, Belgium
GeoAI Research Center, Ghent University, Belgium

Nico Van de Weghe
Department of Geography, Ghent University, Belgium
GeoAI Research Center, Ghent University, Belgium

Abstract

This paper explores whether qualitative spatial encodings make overtake manoeuvres more interpretable to a large language model (LLM). Using simulated traffic trajectories from a microscopic road-network model, we encode the same overtake event at three levels of complexity through Point-Descriptor-Precedence (PDP) inequality matrices. An LLM is then used in a staged, zero-shot API workflow in which longitudinal and lateral descriptor encodings derived from these matrices are interpreted separately and then merged into a final manoeuvre interpretation. Outputs are assessed through expert-based qualitative evaluation focusing on relational correctness, spatial precision, completeness, and interpretive adequacy. For the selected case, the results suggest that the model can produce meaningful natural-language interpretations at all three levels, but that these change as complexity increases. The 2-point representation yields the clearest focal manoeuvre, the 5-point representation offers the best balance between contextual richness and interpretability, and the 10-point representation produces the most nuanced structural reading, although in a less concise form. The paper should be read as an exploratory study of representation effects rather than as a benchmark of LLM performance.

2012 ACM Subject Classification Computing methodologies → Natural language generation; Computing methodologies → Spatial and physical reasoning

Keywords and phrases micro-scale traffic analysis, overtaking behaviour, symbolic movement encoding, spatial reasoning, expert evaluation

Digital Object Identifier 10.4230/LIPIcs.COSIT.2026.16

Category Short Paper

Supplementary Material *Dataset (Derived qualitative encodings, API materials, visualisations, and expert evaluations)*: `https://doi.org/10.5281/zenodo.20729343`

Acknowledgements The authors thank MINT nv for providing the simulated traffic trajectory data used in this study.

1 Introduction

How can a movement event be represented so that its spatial structure can be interpreted, rather than merely measured? This question lies at the heart of qualitative spatial reasoning (QSR), where the aim is not only to record coordinates, but to describe relations such as before, behind, left of, or alongside in a form that supports spatial reasoning [6, 3]. This relational perspective is especially relevant for traffic manoeuvres such as overtaking. An overtake is not simply the path of one vehicle, but a structured change in longitudinal and lateral relations between a focal vehicle and surrounding traffic, making it a useful event for studying qualitative movement representation [15].

Detailed trajectory data are increasingly available through video-based observation, LiDAR, drones, and automated tracking [2, 18]. Most traffic analyses nevertheless treat manoeuvres primarily as numerical patterns to cluster, predict, or score. Prior work in qualitative movement representation and micro-scale traffic analysis has shown that vehicle interactions can also be treated as structured relational events, where subtle changes in position, ordering, and grouping may be difficult to detect from raw trajectories or aggregate safety indicators alone [14, 16, 10, 12]. In parallel, recent work has evaluated the spatial reasoning abilities of large language models (LLMs) and large reasoning models, including reasoning over RCC-8 relations, cardinal directions, StepGame, and broader spatial concepts [4, 5, 8, 9, 7, 17]. However, these studies mainly rely on synthetic or text-based inputs, rather than qualitative encodings derived from movement data.

In this short paper, we ask whether an LLM can coherently interpret a qualitative spatial encoding of the same overtake manoeuvre, and how that interpretation changes as the relational context becomes more complex. Using simulated traffic trajectories from a microscopic road-network model, we represent one focal overtake event at three levels of complexity through Point-Descriptor-Precedence (PDP) inequality matrices [11]. Rather than benchmarking LLM performance at scale, we examine how representation complexity affects the interpretation of a single, controlled manoeuvre case.

2 Data and Methods

Figure 1 presents the workflow used in this study: selected overtake cases are encoded through the Point-Descriptor-Precedence (PDP) framework [11], represented as qualitative inequality matrices, interpreted by an LLM, and then assessed through expert-based qualitative evaluation.

Figure 1 Overview of the workflow used in this study.

2.1 Traffic data and illustrative case selection

This study uses trajectory-based traffic data provided by an independent traffic and mobility consultancy. The data originate from a PTV VISSIM microscopic simulation of a three-lane highway environment, informed by representative traffic counts and regional origin–destination

matrices for the base year 2019–2020. From approximately 120 overtaking configurations available for each representation level, we focus on configuration 52 because it provides a visually clear and qualitatively well-bounded example of the same overtaking event with 2, 5, and 10 moving points. The three representation levels encode the same focal overtaking manoeuvre rather than three different traffic events. The 2-point representation contains only the two vehicles directly involved in the overtake and therefore represents the minimal focal interaction. The 5-point representation adds three surrounding vehicles to capture the immediate local traffic context, while the 10-point representation includes eight surrounding vehicles and serves as a richer complexity test. At each time step, the number of unique object pairs increases from 1 in the 2-point case to 10 and 45 in the 5- and 10-point cases. Across the full time window, the matrices also include relations between point instances at different time steps.

2.2 PDP encoding of overtake manoeuvres

To make the trajectories suitable for spatiotemporal reasoning, we use a qualitative relational representation derived from the PDP framework (Figure 2). In this representation, moving objects are expressed as points relative to selected descriptors, and their relative positions are encoded in a qualitative rather than metric form [11]. We apply the "rough" parameter only to the lateral descriptor, using a value of 1, so that small lateral coordinate differences within the same lane are treated as qualitatively equal rather than as distinct positions [10]. This prevents vehicles travelling in the same lane from being encoded as if they occupied different lateral positions solely because their coordinates are not exactly identical.

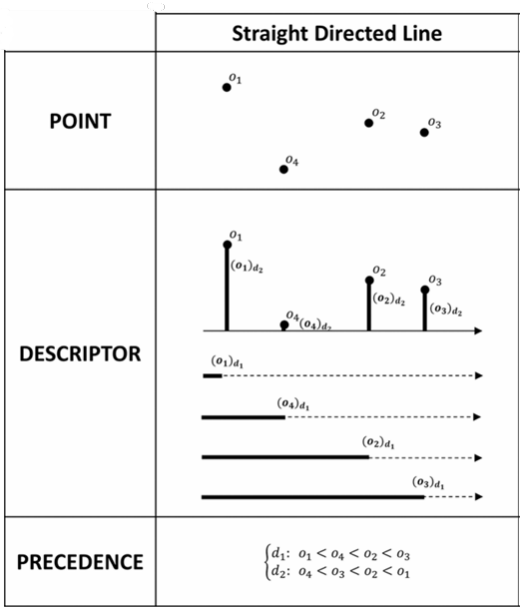

Figure 2 Conceptual illustration of the Point-Descriptor-Precedence (PDP) framework [10].

2.3 Qualitative inequality matrices

The PDP encoding results in inequality matrices that capture pairwise relations between points and provide the basis for the qualitative encodings submitted to the LLM. Separate matrices are constructed for the longitudinal x-descriptor and lateral y-descriptor. Pairwise

relations are encoded numerically, with 0, 1, and 2 corresponding to the qualitative PDP relations $<$, $=$, and $>$, respectively. Along the longitudinal descriptor, these indicate behind, equal, and ahead; along the lateral descriptor, they indicate one side, equal lane-level position, and the other side, relative to the descriptor orientation. Each matrix represents pairwise qualitative relations across a fixed time window of 50 steps covering the selected overtaking event. Increasing the number of moving points adds contextual detail but also increases representational complexity, allowing us to examine how LLM reasoning changes as the spatial encoding becomes more detailed. For expert inspection, the matrices were also rendered as colour-coded heatmaps showing the qualitative PDP relations.

2.4 LLM interpretation procedure

We evaluated LLM-based interpretation through an instruction-based zero-shot API workflow. For each representation, the model was queried with structured qualitative encodings derived from the inequality matrices rather than raw trajectory coordinates or visualisations. The longitudinal and lateral descriptor encodings were first submitted separately, after which their partial interpretations were merged into a final manoeuvre interpretation. Each descriptor-specific prompt asked the model to interpret changes in pairwise relations over time using only the provided qualitative encoding. A final prompt then merged both outputs into a structured description of the manoeuvre, its phases, and the roles of the involved points. All calls used GPT-5.4 through the API with fixed generation settings across runs (`temperature` = 0 and `top_p` = 1). Inputs were structured JSON payloads containing configuration metadata, encoding definitions, and compressed same-time point-pair summaries derived from the qualitative inequality matrices. The complete wording of the prompts, together with the JSON schema, API settings, and representative request–response examples, is provided in the supplementary materials to support reproducibility.

2.5 Expert-based qualitative evaluation

Model outputs were assessed through a two-expert qualitative evaluation. Independently for each representation level, both experts first established a reference interpretation using the static visualisation, animation, and inequality matrices or heatmaps, without consulting the LLM output. The merged LLM interpretation was subsequently compared with these expert references. Evaluation focused on relational correctness, spatial precision, completeness, and interpretive adequacy, using the categories “Strong”, “Adequate”, “Partial”, and “Weak”. The aim was not to benchmark LLM performance quantitatively, but to assess whether the model reconstructed the main relational structure in a spatially coherent and encoding-grounded way. The expert template follows broader work on human evaluation of LLM outputs [1, 13].

3 Results

3.1 A baseline overtake case and its contextual extensions

Figure 3 shows our configuration at three levels of representational complexity. Across the three representations, the focal overtake remains unchanged, while progressively more surrounding vehicles are included. The 2-point case isolates the direct interaction between the two overtaking vehicles, the 5-point case embeds it within the immediate traffic context, and the 10-point case places it within a substantially richer interaction structure. The corresponding animations in the supplementary materials confirm that the same focal

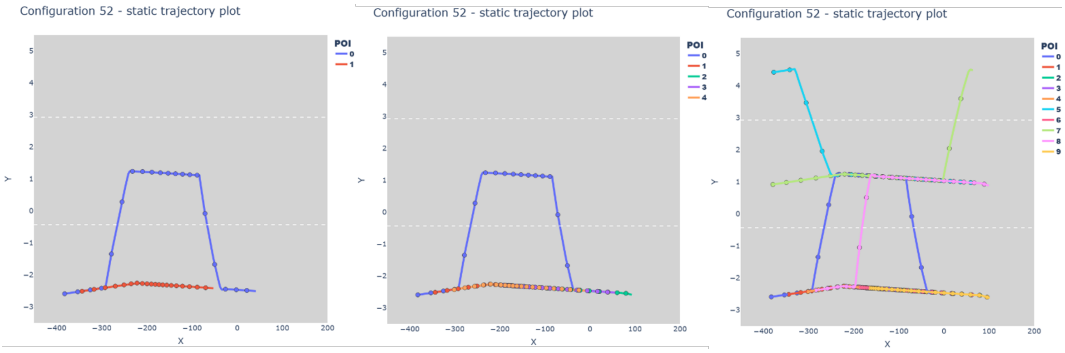

Figure 3 Configuration 52 represented at three levels of complexity (2-, 5-, and 10-point representations).

manoeuvre is preserved. This controlled design allows changes in the LLM output to be attributed primarily to representational complexity rather than to differences between manoeuvres.

3.2 From trajectory to qualitative representation

The increase in contextual detail is also visible in the qualitative encoding. Figure 4 shows the lateral-descriptor inequality matrices for configuration 52 at 2, 5, and 10 moving points. The matrices are visualised as heatmaps: green cells represent $<$, yellow cells $=$, and red cells $>$. Thick black lines mark point-specific row and column blocks, making pairwise relations easier to identify.

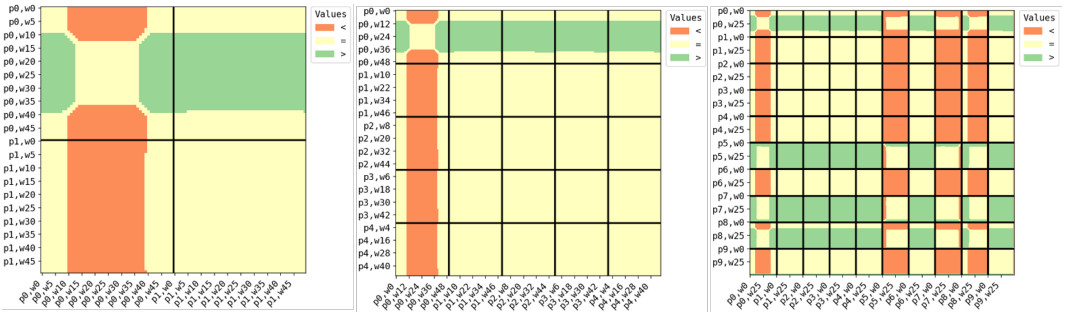

Figure 4 Heatmap representation of the lateral-descriptor inequality matrices for configuration 52 at 2, 5, and 10 moving points. Colours indicate qualitative relations ($<$, $=$, $>$), while the thick black lines mark the boundaries between point-specific row and column blocks.

In the 2-point representation, the matrix contains no surrounding traffic and captures only relations involving the two overtaking vehicles. The top-right submatrix shows how p_0 changes relative to p_1: both vehicles first occupy the same lane-level relation, then separate laterally during the overtaking phase, and finally return to the same lane-level relation. In the 5-point and 10-point representations, the same focal overtake is preserved but relations with surrounding vehicles are added. The representation therefore expands from a two-vehicle interaction to a richer local traffic configuration. In the 5-point representation, p_0 is the main changing actor, while the remaining vehicles form a comparatively stable surrounding group. In the 10-point representation, the colour pattern becomes more fragmented, especially for p_5, p_7, and p_8, making the overall structure harder to summarise at a glance.

This progression illustrates the central trade-off: with few points, the manoeuvre is readable but structurally limited; with more points, the encoding reflects a broader traffic scene but becomes harder to summarise. The inequality matrices therefore function as the key intermediate representation between raw trajectories and LLM interpretation.

3.3 LLM interpretation of the qualitative encoding

We focus on the merged manoeuvre interpretations, as these provide the most readable output of the staged API workflow. The longitudinal and lateral descriptor encodings were first interpreted separately, then combined into one final interpretation. The evaluated merged API outputs, including the corresponding JSON responses and expert evaluations, are provided in the supplementary materials.

The merged outputs were assessed against expert reference interpretations using the procedure described above. In the 2-point representation, the model produced a compact and strongly aligned overtaking narrative: p_0 starts behind p_1, overtakes it once, remains ahead, and shows a temporary lateral offset followed by realignment. This case was judged "Strong" on all four criteria by the experts.

In the 5-point representation, the output was richer and identified p_0 as the main changing actor within a comparatively stable surrounding group. It reconstructed a manoeuvre in which p_0 separates laterally, overtakes p_1 and later p_4, and then realigns. This case was judged "Strong" overall, with a minor reservation concerning spatial precision.

In the 10-point representation, the output produced the most nuanced structural reading. Rather than reducing the scene to a single overtaking label, it identified a stable lateral core and a smaller set of actors whose grouping and ordering changed over time. This case remained strong in completeness and interpretive adequacy, but was more limited in relational correctness and spatial precision than the lower-complexity cases.

Across the three levels, the clearest differences appeared in spatial precision and compactness. The 2-point representation yielded the clearest focal manoeuvre, the 5-point representation provided the best balance between clarity and contextual richness, and the 10-point representation supported the most nuanced but least concise structural reading. These findings are summarised in Table 1.

Table 1 Consolidated two-expert evaluation of merged LLM outputs across representation levels. Rel. corr. = relational correctness; Spatial prec. = spatial precision; Compl. = completeness; Interp. adequacy = interpretive adequacy.

Level	Rel. corr.	Spatial prec.	Compl.	Interp. adequacy	Overall judgement
2-point	Strong	Strong	Strong	Strong	Strongly aligned with expert interpretation
5-point	Strong	Adequate	Strong	Strong	Broadly aligned with minor issues
10-point	Adequate	Adequate	Strong	Strong	Broadly aligned with minor issues

4 Discussion and conclusion

This paper explored whether qualitative spatial encodings can support LLM-based interpretation of overtake manoeuvres. By representing the same overtaking event with 2, 5, and 10 moving points, the study isolates how increasing contextual complexity affects interpretation.

The staged API workflow further separates descriptor-level interpretation from the final merged manoeuvre summary, making the production of the natural-language output more transparent.

The results suggest that PDP-based inequality matrices provide a useful intermediate step between raw trajectories and natural-language description. The 2-point representation was easiest to interpret but offered limited context. The 5-point representation enriched the same manoeuvre with surrounding vehicles while remaining coherent, giving the best balance between relational richness and interpretability. The 10-point representation produced the most detailed structural reading, identifying a stable core and changing actors, but the description became less concise and less unified.

The contribution of the paper lies less in benchmarking LLM performance than in showing how the form of the encoding influences the interpretation produced. The LLM is not treated as a traffic-analysis model in its own right, but as a way to examine whether the symbolic structure captured by inequality matrices can be translated into coherent natural-language descriptions. In this setting, success means reconstructing the main relational structure of the manoeuvre in a spatially coherent and encoding-grounded way.

The study remains exploratory. Its main limitation is the use of a single illustrative configuration, which allows a controlled comparison across representation levels but does not yet show robustness across manoeuvres or interaction structures. The selected case was drawn from a broader collection of approximately 120 overtaking configurations per representation level. Extending the analysis to this larger set would support benchmark-style evaluation across cases, prompts, models, and representational variants. Future work could also extend the framework to merging, following, lane changes without passing, and other domains where moving point configurations are interpreted through changing spatial relations, such as sports or interaction-based movement analysis.

5 Data and materials availability

The supplementary materials are available via Zenodo: `https://doi.org/10.5281/zenodo.20729343`. They include the API code, representative request–response examples, evaluated merged outputs, expert evaluation materials, and derived qualitative encodings. Due to source restrictions, raw trajectory data are not shared.

References

1 C. Anghel, A. A. Anghel, E. Pecheanu, M. V. Craciun, A. Cocu, and C. Niculita. PEARL: A rubric-driven multi-metric framework for LLM evaluation. *Information*, 16(11), 2025. doi:10.3390/info16110926.

2 M. Berghaus, S. Lamberty, J. Ehlers, E. Kalló, and M. Oeser. Vehicle trajectory dataset from drone videos including off-ramp and congested traffic – analysis of data quality, traffic flow, and accident risk. *Communications in Transportation Research*, 4, 2024. doi:10.1016/j.commtr.2024.100133.

3 J. Chen, A. G. Cohn, D. Liu, S. Wang, J. Ouyang, and Q. Yu. A survey of qualitative spatial representations. *The Knowledge Engineering Review*, 30(1):106–136, 2015. doi:10.1017/S0269888913000350.

4 A. G. Cohn. An evaluation of ChatGPT-4's qualitative spatial reasoning capabilities in RCC-8, 2023. doi:10.48550/arXiv.2309.15577.

5 A. G. Cohn and R. E. Blackwell. Evaluating the ability of large language models to reason about cardinal directions. In *Leibniz International Proceedings in Informatics*, 2024. doi:10.4230/LIPIcs.COSIT.2024.28.

6 A. G. Cohn and J. Renz. Qualitative spatial representation and reasoning. In F. van Harmelen, V. Lifschitz, and B. Porter, editors, *Handbook of Knowledge Representation*, pages 551–596. Elsevier, 2008. doi:10.1016/S1574-6526(07)03013-1.

7 Eirinaios-Odysseas Gardelakos, Vasileios Kyriakopoulos, Despina-Athanasia Pantazi, Orestis-Minas Kapopoulos, Maria Tsourma, and Manolis Koubarakis. Can large reasoning models reason about spatial relations? In *Proceedings of the 8th ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery*, GeoAI '25, pages 81–91, New York, NY, USA, 2025. Association for Computing Machinery. doi:10.1145/3764912.3770841.

8 M. Hojati and R. Feick. Large language models: Testing their capabilities to understand and explain spatial concepts. In *Leibniz International Proceedings in Informatics*, volume 315 of *LIPIcs*, 2024. doi:10.4230/LIPIcs.COSIT.2024.31.

9 Fangjun Li, David C. Hogg, and Anthony G. Cohn. Advancing spatial reasoning in large language models: An in-depth evaluation and enhancement using the stepgame benchmark, 2024. doi:10.48550/arXiv.2401.03991.

10 A. Qayyum. *A Qualitative Representation of Moving Point Configurations with Illustrations in Micro Traffic Analysis*. PhD thesis, Ghent University, Faculty of Sciences, 2022. URL: http://hdl.handle.net/1854/LU-8750343.

11 A. Qayyum, B. De Baets, M. S. Baig, F. Witlox, G. De Tré, and N. Van de Weghe. The point-descriptor-precedence representation for point configurations and movements. *International Journal of Geographical Information Science*, 35(7):1374–1391, 2021. doi:10.1080/13658816.2020.1864378.

12 A. Qayyum, B. De Baets, L. De Cock, F. Witlox, G. De Tré, and N. Van de Weghe. Application of the point-descriptor-precedence representation for micro-scale traffic analysis at a non-signalized T-junction. *Geo-Spatial Information Science*, 26(3):406–430, 2023. doi:10.1080/10095020.2022.2069520.

13 T. Y. C. Tam, S. Sivarajkumar, S. Kapoor, A. V. Stolyar, K. Polanska, K. R. McCarthy, H. Osterhoudt, X. Wu, S. Visweswaran, S. Fu, P. Mathur, G. E. Cacciamani, C. Sun, Y. Peng, and Y. Wang. A framework for human evaluation of large language models in healthcare derived from literature review. *npj Digital Medicine*, 7(1), 2024. doi:10.1038/s41746-024-01258-7.

14 N. Van de Weghe. *Representing and Reasoning about Moving Objects: A Qualitative Approach*. PhD thesis, Ghent University, 2004.

15 N. Van de Weghe, A. G. Cohn, P. De Maeyer, and F. Witlox. Representing moving objects in computer-based expert systems: the overtake event example. *Expert Systems with Applications*, 29(4):977–983, 2005. doi:10.1016/j.eswa.2005.06.022.

16 N. Van de Weghe, A. G. Cohn, G. De Tré, and P. De Maeyer. A qualitative trajectory calculus as a basis for representing moving objects in geographical information systems. *Control and Cybernetics*, 35(1):97–119, 2006.

17 A. Yang, C. Fu, Q. Jia, W. Dong, M. Ma, H. Chen, F. Yang, and H. Wu. Evaluating and enhancing spatial cognition abilities of large language models. *International Journal of Geographical Information Science*, 39(9):2009–2044, 2025. doi:10.1080/13658816.2025.2490701.

18 Jiaxing Zhang, Wen Xiao, Benjamin Coifman, and Jon Mills. Vehicle tracking and speed estimation from roadside lidar. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, September 2020. doi:10.1109/JSTARS.2020.3024921.