

Context-Enriched Natural Language Descriptions of Vessel Trajectories

Kostas Patroumpas* Alexandros Troupiotis-Kapeliaris^{†*} Giannis Spiliopoulos^{†*}
Panagiotis Betchavas* Dimitrios Skoutas[‡] Dimitris Zissis[†] Nikos Bikakis^{§*}
*Archimedes/Athena RC, Greece [†]University of the Aegean, Greece [‡]Athena RC, Greece
[§]Hellenic Mediterranean University, Greece

Abstract—We address the problem of transforming raw vessel trajectory data collected from AIS into structured and semantically enriched representations interpretable by humans and directly usable by machine reasoning systems. We propose a context-aware trajectory abstraction framework that segments noisy AIS sequences into distinct trips each consisting of clean, mobility-annotated episodes. Each episode is further enriched with multi-source contextual information, such as nearby geographic entities, offshore navigation features, and weather conditions. Crucially, such representations can support generation of controlled natural language descriptions using LLMs. We empirically examine the quality of such descriptions generated using several LLMs over AIS data along with open contextual features. By increasing semantic density and reducing spatiotemporal complexity, this abstraction can facilitate downstream analytics and enable integration with LLMs for higher-level maritime reasoning tasks.

Index Terms—Maritime AI, Vessel Trajectory Analysis, Large Language Models, Semantic Trajectories, Spatiotemporal Data Mining, Explainable AI, Trajectory Annotation, Spatial Data Enrichment, Mobility Pattern Mining

I. INTRODUCTION

Modern sensor and mobile devices have enabled an unprecedented surge in availability and accumulated volume of trajectory data collected for various applications, calling for more advanced methods for their efficient management and analysis. In particular, tracking vessels has been greatly facilitated by the *Automatic Identification System (AIS)*, which provides their identity, position, course, heading, etc. in real time, and thus can support crucial maritime tasks such as route planning, course prediction, or anomaly detection [1], [2]. While AIS messages offer precise location information, they are inherently noisy, occasionally sparse, and crucially lack explicit semantic annotation. Since raw AIS trajectories are sequences of spatiotemporal tuples devoid of contextual grounding, relational abstractions, mobility event characterization, and domain-specific knowledge, direct interaction between unprocessed AIS data with modern *Large Language Models (LLM)* is severely hindered [3].

To address this limitation, we propose a methodology that segments AIS trajectories into distinct trips and then augments them with contextual knowledge derived from heterogeneous data sources. Our approach adheres to the paradigm introduced in [4], [5] for constructing *semantic trajectories*, but applied specifically for vessels. Each trip represents a sequence of *episodes* that abstract vessel mobility patterns (e.g., stopped,

turning, sailing). Regarding *context*, geographic layers (e.g., protected areas, capes, straits), data specific to the maritime domain (such as ports, anchorage zones, shipping lanes, traffic separation schemes, and bathymetry), as well as weather forecasts (e.g., wind force and direction) can be utilized. Thus, vessel movements get enriched with semantic *annotations* describing important interactions (e.g., port calls, strait transits, proximity to environmentally sensitive zones). By explicitly linking trajectory segments to contextual features and inferred mobility events, our framework enhances interpretability and explainability, enabling transparent reasoning about why a vessel followed a particular course or exhibited irregular behavior (e.g., a detour).

Most importantly, building upon such enriched trajectory representations, we are able to generate human- and machine-readable *natural language narratives* of vessel behavior using LLMs. Semantic trajectories embed prior knowledge about navigational corridors, port operations, sea depth limitations, and traffic regulatory zones, thereby providing structured inductive biases. In effect, input sequences of raw AIS messages are transformed into context-aware movement narratives that distill mobility features and better reflect navigational intent, operational patterns, and regulatory guidelines. Our empirical investigation with several LLMs indicates that using such contextually-enhanced trajectory descriptions can improve robustness, reduce uncertainty in case of sparse or noisy AIS data, and enable a more principled detection of deviations from expected maritime behavior. As such, semantic enrichment serves not only as an interface for LLM-based reasoning but also as a foundation for more accurate and explainable maritime analytics. This ensures compatibility with automated processing pipelines while preserving accessibility for maritime stakeholders such as port authorities, coast guards, or shipping companies.

The generated descriptions can have several applications in the maritime domain. They may assist in contextualizing possible events, such as behavior anomalies (e.g., low speed, abrupt turning) or accidents (e.g., groundings) [6]. Interestingly, such descriptions resemble *sailing route directions* consisting of safe passage guidelines or alerts, coastal or port approach suggestions, etc., based on coordinates of known landmarks (e.g., lights, buoys). For instance, detailed suggestions are stipulated

for the passage shown in Figure 1 in Baltic Sea¹. Comparing our proposed context-enriched narratives with such official suggestions can be a good indicator of possible deviations throughout a trip, and also an important step towards maintaining interoperability of suggestions crucial for maritime safety. In addition, such descriptions can automatically provide *voyage reports*², allowing shipping companies to inspect each leg of a trip, check whether the itinerary of a vessel was compliant to charter agreements, and maintain historical logs for route planning in future trips. Last, but not least, textualized semantic trajectories can enable LLMs to reliably ingest, reason over, and synthesize maritime movement information for *downstream analytical tasks*. For example, more accurate trajectory prediction [7] can be achieved by taking advantage of context to infer voyage intentions. Our context-enriched narratives ingest extra knowledge that may improve imputation of missing trajectory segments [8], [9]. Textual descriptions can also assist in navigation and collision avoidance tasks [10] that require complex reasoning to make accurate decisions.

Several recent frameworks have employed LLMs to extract [11], [7] or process [12] vessel trajectory narratives. However, their scope is mostly on maritime tasks like making collision avoidance decisions through textualized navigation scenarios [10], simply inferencing origin-destination ports [7], or apply a cluster-based trajectory segmentation to create semantic descriptions [13]. Instead, our approach detects vessel mobility events to construct semantic trajectories, annotates them with rich contextual knowledge from geographic, maritime, and weather data sources, and ultimately uses LLMs to distill them into reliable, concise, and factual trip narratives.

Our contributions can be summarized as follows:

- We develop a framework for constructing semantic trajectory representations in episodes that summarize mobility events along a vessel’s course from raw AIS locations.
- We further enrich each semantic episode with context from various geographic, maritime, and meteorological data sources in various formats to provide deep insights into the motion patterns of vessels.
- We leverage LLMs to accept such semantic representations and generate textual descriptions of vessel trips that are both human- and machine-readable.
- We conduct an empirical study over real-world AIS data enriched with context from various open data sources to generate trajectory descriptions in natural language using LLMs of varying capabilities.
- We evaluate the quality of such descriptions in terms of relevance, faithfulness, and correctness, providing strong evidence that such AI-generated narratives can help stakeholders explain and interpret vessel behavior at sea.

The remainder of this paper proceeds as follows. Section II surveys related work. In Section III we present our framework for constructing contextually-enriched semantic trajectories of vessels. Section IV explains how LLMs can then be used to

generate textual descriptions from such trajectory representations. Section V reports results from our empirical evaluation. Section VI offers conclusions and future research directions.

II. RELATED WORK

Trajectory data management offers methods for cleaning, storage, indexing, mobility analytics and mining, and certainly support for diverse types of queries (e.g., spatial-only, spatio-temporal, spatio-textual) over trajectories. Recent surveys [14], [15] offer a comprehensive review of relevant research trends.

Recent advances in ML models designed for Natural Language Processing have also attracted research interest in the mobility domain. Typically, discretized location identifiers act as tokens (like words) and full trajectories as sequences (like sentences) [16]. Architectures based on BERT [17] have been adapted for such tokenized trajectory sequences and develop specialized models like TrajBERT [18], Kamel [19] or BERT4Traj [20]. Yet, such models necessitate domain-specific vocabularies, which neglects the deep relational knowledge in emerging LLMs. Indeed, leveraging LLMs for direct transformation of trajectories into textual entries can take advantage of their sophisticated reasoning and contextual depth to interpret intricate relationships in input data [21]. Prompt engineering is crucial for successfully translating the mobility task and input data into semantic descriptions [22] in order to effectively answer queries like next location prediction or history summarization [23].

Specifically in the maritime domain, several models have recently been proposed. For collision risk assessment, [11] leverages a multi-scale time-series encoder, a maritime-specific text-based encoder, along with a cross-modality alignment module in order to provide numerical forecasts along with explanations in natural language. By integrating multiple heterogeneous data sources (e.g., AIS, remote sensing imagery, structured metadata) for improved maritime situational awareness, the transformer-based multimodal model proposed in [24] performs vessel detection and interpretable navigation compliance reasoning. Through trajectory segmentation into episodes, [25] creates semantic tokens, and then uses an LLM-based pipeline to interpret natural language queries and respond through semantic matching. Without generating textual trajectory descriptions but simply via natural language interfaces, [26] examines the reasoning capabilities of LLMs in querying maritime datasets. However, generating comprehensive trip descriptions can significantly enhance the contextual awareness of LLMs. As studied in [13], utilizing full narratives that integrate nearby Points-of-Interest, LLMs can be adapted to predict future movements within a rich geographic context. SEMINT [7] leverages natural language narratives to infer the vessel’s navigation intent, enabling more accurate and context-aware long-term trajectory predictions. Towards offering collision avoidance recommendations, Navigation-GPT [10] encodes real-time vessel states into the language space to align encounter scenarios with maritime regulations. Textual descriptions can also assist in interpreting vessel intent [12] and detecting anomalous behavior [27],

¹<https://msi.nga.mil/api/publications/download?key=16694491/SFH00000/Pub194bk.pdf>

²<https://support.marinetraffic.com/en/articles/9552765-voyage-report>

as demonstrated in MSCE [28], which utilizes multi-modal semantic cognition to uncover trajectory irregularities.

In this work, we propose a framework that generates textual narratives of vessel trajectories using LLMs. Our work applies approaches for mobility data segmentation and annotation towards construction of *semantic trajectories* [4], [5] in order to detect motion patterns and enable movement interpretation. According to this generic paradigm, spatio-temporal features are converted into structured trajectories (as stop and move episodes), also utilizing application-dependent geographic context (e.g., road network, Points of Interest) to annotate stops and moves with the relevant objects. In our case, from raw AIS data we identify pivotal mobility episodes and enrich them with contextual knowledge from heterogeneous data sources (geographic, maritime, meteorological). Ultimately, this distilled information can serve as a foundational layer for broader maritime AI architectures, providing such standardized semantic descriptors as input to task-specific models for vessel trajectory analysis.

III. CONSTRUCTING CONTEXT-ENRICHED SEMANTIC TRAJECTORY REPRESENTATIONS

In this section, we present our open-source framework³ that accepts as input raw AIS locations and constructs context-enriched trajectory representations. It first annotates important points along the course of each vessel according to detected mobility events (e.g., stops, turns, slow motion). Based on such annotations, it then segments each trajectory into separate trips each consisting of a sequence of episodes. Finally, it ingests extra context from various (geographical, maritime, meteorological) sources into each episode. The resulting semantic trajectory representation can be exported into several formats for further processing, analytics, or visualization in maps and charts. Note that no LLM is used in any stage of this workflow.

A. Trajectory Annotation with Mobility Events

In order to identify critical locations along a vessel’s trajectory, our method employs the *trajectory compression* framework proposed in [29], which is specifically tailored for annotating AIS data. This maintains a velocity vector over the recent history of each vessel’s movement and applies spatiotemporal filters and rules strictly over raw AIS messages by considering factors such as time, location, as well as changes in speed and heading. By examining how the motion pattern of a given vessel changes across time, this method can semantically annotate selected positions that signify several types of *mobility events*:

- *Stops* indicate time periods where the vessel remains stationary, i.e., its speed $v < 0.5$ knots. Although stops mostly occur in port areas, they can be occasionally observed offshore or even at open sea (e.g., due to mechanical failures or suspicious activities).
- *Gaps* in communication indicate that no AIS signal has been reported from a vessel over some time, e.g., in the past $\Delta T = 30$ minutes.

- *Speed change* indicates that the vessel has been decelerating or accelerating. If the rate of speed change exceeds a given threshold a (e.g., $a = 25\%$), this marks a significant change in speed, typically occurring when the vessel reaches or departs from a port or makes a maneuver.
- *Slow motion* means that the vessel consistently moves at low speed (e.g., < 1 knot) over some time. In contrast to a stop event, these successive locations usually occur along a path while the vessel is sailing.
- *Turns* are detected once the actual heading over ground deviates significantly, e.g., above a threshold $\theta > 5^\circ$. This comparison is done with respect both to the previous heading (for abrupt changes) but also with the velocity vector (for smooth, long-lasting turns typical for large ships). In case the vessel is stopped or moves slowly, any changes in heading can be ignored as they are likely caused by sea drift.

In our processing workflow, we have significantly improved this method⁴; instead of compression, it is now specifically geared towards more efficient, rich annotations of pivotal locations along each vessel’s course. Crucially, we can apply parametrizations according to ship types, e.g., specifying different angle threshold for smaller crafts and large cargo or passenger vessels. In addition, for more accuracy in spatiotemporal calculations, the velocity vector maintained per vessel can now be defined over its last few locations (e.g., 10 previous AIS messages) or over a recent time-based sliding window (e.g., 10 minutes back from now). Furthermore, instead of a single characterization, a given location can carry multiple annotations (e.g., both change in heading and slow motion), which is important in identifying successive episodes along a trajectory. Finally, this process also applies suitable filters that eliminate much of the noise inherent in AIS messages, such as duplicate positions, invalid coordinates, delayed messages, etc. Figure 1 illustrates annotated AIS locations along the trajectory of a passenger vessel that makes an itinerary between two ports in Denmark. Note that each turn includes multiple successive points, typical for large ships like this one.

B. Trajectory Segmentation in Trips and Episodes

The accumulated sequence of AIS locations per vessel can eventually become very long over time, given that a ship may relay even thousands of messages per day. Splitting a trajectory into smaller parts can greatly facilitate its indexing and subsequent processing in large maritime repositories. We propose segmenting the trajectory of a vessel into distinct *trips*, each one consisting of successive *episodes*, by taking advantage of the aforementioned annotations by mobility events.

Trips. In our workflow, a *trip* comprises the subsequence of AIS locations between two successive stops or long gaps. More specifically, the segmentation process iterates chronologically over the annotated locations of a given vessel.

³<https://github.com/M3-Archimedes/AIS-semantic-trajectories>

⁴<https://github.com/M3-Archimedes/AIS-trajectory-annotation>



Fig. 1: Originally reported AIS locations (in black dots) of a vessel after annotation: green boxes signify stops, orange triangles denote turning points, and red stars indicate noise.

Noisy locations are discarded, and each annotated location concerning stop or gap events is examined as follows:

- The last location in a stop event signifies that the vessel will next be departing for a new trip.
- The first location in a stop event marks the end of the current trip.
- If a long gap occurs (e.g., lasting more than a given $\delta = 3$ hours), the current trip is abruptly terminated.
- After such a long gap, once communication with the vessel is resumed, a new trip is assigned.

In effect, locations of a vessel between two successive stops or long gaps are grouped together with a unique trip identifier. Thus, trip origin and destination are not defined strictly by stops (typically in a port, but sometimes at open sea), but also by long-lasting gaps. Note that in trip segmentation we set a longer gap duration $\delta > \Delta T$ used in annotating a gap event (as discussed in Section III-A). This is a deliberate decision (subject to a user-specified parameter δ) to account for cases where contact with a vessel is temporarily lost in the middle of its voyage, e.g., at open sea with no AIS coverage, yet it effectively continues the course to its destination. Thus, smaller communication gaps lasting less than δ may still occur in the designated trips.

Several *statistics* can be calculated per trip, such as its total *duration*, *travelled distance*, and *average speed*, accurately based on the original AIS reports.

Episodes. In principle, an *episode* abstracts a subsequence of AIS locations that are collectively characterized with a particular mobility event (i.e., a motion pattern like stop, turn, or gap), because they are highly correlated with respect to

spatio-temporal feature (e.g., speed, heading, time period). While iterating over the annotated locations per vessel, distinct episodes within a trip can be identified:

- *STOPPED*: This spans an entire stop event; the centroid of those locations is used to represent the episode.
- *TURNING*: This includes a series of successive locations: it starts from a location annotated with a significant change in heading and ends up to the last similarly annotated point before the vessel course stabilizes to a new fixed heading (i.e., after the turn). Isolated turning points are discarded as noise as they are extremely rare for most vessels, which typically take smooth turns.
- *MANEUVERING*: Such episodes also concern changes in heading, but smaller ones that denote minor corrections in the course of the vessel or local maneuvers at ports. If the overall change in heading is less than ϕ (e.g., 5°), such episodes are marked as maneuvers; otherwise, they represent more significant turns.
- *COMMUNICATION GAP*: As already discussed, this concerns gap intervals lasting less than δ time units. Since the vessel's whereabouts are unknown during such a gap, this episode is abstracted by the known start and end locations of the gap.
- *SAILING*: This characterizes every other segment where the vessel is considered moving, i.e., it is neither stopped nor turning and not in communication gap. As the vessel typically follows a straight course during this episode, it can be abstracted by its two endpoints.

For each episode, several *statistics* can be calculated, all based on originally reported AIS locations. As illustrated within the green box in Figure 2, these include duration (in seconds), travelled distance (in nautical miles), as well as average speed (in knots) during this episode. They also include the total change in heading (in degrees) and direction (e.g., south west, north) when the vessel is turning or maneuvering; in all other episodes, such values indicate the overall heading and general direction.

C. Context Enrichment

While segmenting long trajectories into trips and episodes can certainly facilitate their processing, such representations can be further enhanced with contextual information in order to provide deep insights into the motion patterns of vessels. Our workflow takes advantage of open data sources such as OpenStreetMap (OSM)⁵, OpenSeaMap⁶ or the World Port Index⁷ to extract coastal or offshore *geospatial* features using topological checks (e.g., proximity to ports or crossing environmentally sensitive areas) along or near a vessel's course and ingest them into contextual information per episode. Additional context is obtained from *array-oriented scientific data* (typically in NetCDF format⁸). This latter data (at varying

⁵<https://www.openstreetmap.org/>

⁶<https://map.openseamap.org/>

⁷<https://msi.nga.mil/Publications/WPI>

⁸<https://www.unidata.ucar.edu/software/netcdf>

t_start	t_end	geometry	episode	direction	heading	duration_sec	distance_nm	speed_knots	wind_beaufort	wind_dir	sea_depth	placemarks
1706049656	1706050187	LINestring (11.34...	SAILING	SOUTH WEST	-130	531	1,38	9,4	NULL	NULL	-5	off Lolland
1706050187	1706050277	LINestring (11.31...	TURNING	SOUTH WEST	16	90	0,28	11,3	NULL	NULL	-9	off Lolland
1706050277	1706051452	LINestring (11.31...	SAILING	SOUTH WEST	-144	1175	4,64	14,2	4	WEST	-11	off Lolland crossing Fehmarnbelt
1706051452	1706051614	LINestring (11.23...	TURNING	SOUTH	38	162	0,68	15	4	WEST	-28	crossing Fehmarnbelt
1706051614	1706052256	LINestring (11.23...	SAILING	SOUTH	175	642	2,7	15,1	4	WEST	-29	crossing Fehmarnbelt
1706052256	1706052376	LINestring (11.23...	TURNING	SOUTH WEST	23	120	0,48	14,4	4	WEST	-15	NULL
1706052376	1706052618	LINestring (11.23...	SAILING	SOUTH WEST	-152	242	0,43	6,5	4	WEST	-10	NULL
1706052618	1706052778	LINestring (11.22...	SAILING	SOUTH WEST	-154	160	0,06	1,3	NULL	NULL	NULL	NULL
1706052778	1706056727	LINestring (11.22...	STOPPED	NULL	0	3949	0	0	NULL	NULL	NULL	in PUTTGARDEN port

Fig. 2: Trip representation in successive episodes with statistics (in green box) along with contextual information from various external data sources (in red box) as extracted in CSV format suitable for map visualization (MAP).

grid resolutions) may concern *weather conditions* available from large meteorological organizations such as Copernicus⁹ or NOAA¹⁰, as well as *bathymetry* data from GEBCO¹¹. Such information is extremely important for maritime situational awareness and can help explain motion patterns, e.g., a trip detour caused by a storm, or diverse courses taken by vessels of different draught due to shallow waters.

More specifically, our workflow is currently able to ingest extra context *per episode* regarding:

- *Ports*: Identifies the name of the port where the vessel is anchored during a stop episode.
- *Coastal features*: Finds whether the vessel is moving in close distance (e.g., less than 5 nautical miles) to capes, peninsulae, straits, etc.
- *Offshore areas*: Finds any polygonal regions (e.g., marine protection zones, national sea parks, fishing areas) the vessel is crossing along its course.
- *Traffic separation schemes*: Indicates whether the vessel is navigating across the designated lane for its direction in high-density areas according to traffic regulations.
- *Meteorological*: If such data is available for this time interval, it identifies the wind conditions (wind force in the Beaufort scale, wind direction) along the polyline that represents the course of a vessel during this episode.
- *Bathymetry*: Provides the seabed depth (in meters) at the grid cell of the stop event (if the vessel is anchored) or the minimum depth along the polyline of a moving episode (i.e., when the vessel is sailing or turning).

Any context obtained from geospatial features is listed as a collection of *placemarks* in the resulting representation, as shown within the red box in Figure 2. Context features from NetCDF data on weather and bathymetry are maintained as separate items (e.g., *wind_beaufort*, *sea_depth*).

Finally, more contextual information may be added concerning the *entire trip*. For instance, in case that both the *origin* and *destination ports* have been identified (i.e., in the first

<pre>{ "direction": "SOUTH WEST", "heading": -144, "duration_seconds": 1176, "distance_miles": 4.64, "speed_knots": 14.2, "movement": "SAILING", "wind_beaufort": 4, "wind_direction": "WEST", "sea_depth": -11, "placemarks": "off Lolland crossing Fehmarnbelt", "start_location": {...}, "end_location": {...} }</pre>	<pre>From 2024-01-23 22:51:17 until 2024-01-23 23:10:52 SAILING SOUTH WEST at -144 degrees off Lolland crossing Fehmarnbelt, covering 4.64 nautical miles in 1176 seconds at speed 14.2 knots at minimum sea depth of -11 meters while WEST winds of intensity 4B were blowing.</pre>
---	---

(a) JSON

(b) TXT

Fig. 3: Alternative episode representations.

and last episode of a trip), the entire trip geometry can be checked for possible match against known routes (e.g., *ferry lines* available from OSM). In case of match, *Dynamic Time Warping (DTW)* can be used to measure similarity and thus highlight any significant deviations of the current trip from the typical ferry route. For example, we can identify that the vessel in Figure 1 makes a significant detour from the typical route (shown in blue dashed line) since the estimated *DTW* = 2.91 is significantly high.

D. Output Representations

The resulting contextually enriched semantic trajectories can be exported in several output formats:

- *CSV*: A separate row is emitted per episode using a predefined attribute schema. The start and end *point* locations of each episode are listed in separate columns in Well-Known Text (WKT).
- *MAP*: This representation is also in CSV format, but it includes a single *geometry* column to facilitate map visualization (e.g., in GIS software). Each episode is represented with a linestring geometry in WKT (like the ones shown in Figure 2), which in the case of turns or maneuvers includes a sequence of vertices for more accurate illustration of such events.
- *JSON*: Each trip is serialized as a JSON array of elements, each one representing an episode. As shown

⁹<https://cds.climate.copernicus.eu/datasets>
¹⁰<https://www.noaa.gov/>
¹¹<https://www.gebco.net/>

System:

ROLE AND OBJECTIVE: You are a data scientist who studies trips of vessels, i.e., the route of ships sailing between ports or locations at open sea. You are given a list of waypoints in JSON format representing the trip of a vessel, including information about anchorage in ports, duration of each stop, sailing for varying time intervals, and occasionally turning maneuvers when a vessel changes its direction of movement. Based on this information about the trip, you must provide its description in plain text.

APPROACH AND LIMITATIONS:

- (1) All vessel locations are specified in (longitude, latitude) coordinates in WGS84 (EPSG:4326) Coordinate Reference System.
- (2) Distances are always expressed in nautical miles and are calculated using the Haversine formula on an ellipsoid.
- (3) Speed values are in knots, i.e., nautical miles per hour.
- (4) Durations are in seconds. All timestamp values are in UNIX epochs (seconds elapsed since 1970-01-01) in GMT time zone. Wind intensity is in Beaufort scale.
- (5) Placemarks indicate proximity to islands, ports, straits, capes, etc.
- (6) Note that the motion patterns and speed of vessels absolutely depend on their type (e.g., Passenger, Cargo, Tanker, Pleasure craft, etc.), and some vessels may be anchored for long intervals.
- (7) Due to COMMUNICATION GAPS, a vessel may suddenly appear far away from its previously known position.

METHODOLOGY: You must provide:

- (i) A textual description in PLAIN TEXT that outlines the trip. In a separate paragraph describe each major stage of the trip, such as port operations (e.g., anchorage), important navigational features (e.g., cruising at open sea, maneuvering, turning) as well as any proximity to placemarks (e.g., islands, straits, capes). Do NOT include any explanations or reasoning regarding any stage of the trip.
- (ii) A short summary in JSON format with statistics that include the traveled distance (in nautical miles), the total duration (in seconds), the names of the origin and destination ports (if available) and any adverse weather conditions identified across the entire trip as follows:

```
{ "traveled_distance": distance_value, "total_duration": duration_value, "origin_port":
port_name, "destination_port" : port_name, "adverse_weather_conditions": [
"wind_intensity": "beaufort_value", "wind_direction": direction_value ] }
```

For your responses, you must NOT include information from any external source; use information strictly from the input JSON.

User:

Based on this JSON array representing the simplified trip of a vessel, please provide a textual description. ``json{ ... }``

Fig. 4: Prompt template for generating trajectory descriptions. The LLM receives as input a `json array` with distilled context-enhanced representation about a vessel trip and generates: (i) semantically-rich natural language descriptions in plain text, and (ii) overall trip statistics (in `json format`).

in Figure 3a, an episode is described through a set of key-value pairs capturing the mobility characteristics and contextual information associated with that segment.

- *TXT*: An entire trip is summarized into a paragraph in plain text. We stress that this is **not** an AI-generated narrative (we discuss in the next section how this can be achieved), but rather a tedious yet factual recitation of all available mobility and contextual information using standard syntax and predefined connecting words. As illustrated in the example in Figure 3b, an episode is described by a separate sentence that simply replicates all extracted features of the episode verbatim.

IV. LLM-GENERATED DESCRIPTIONS OF VESSEL TRIPS

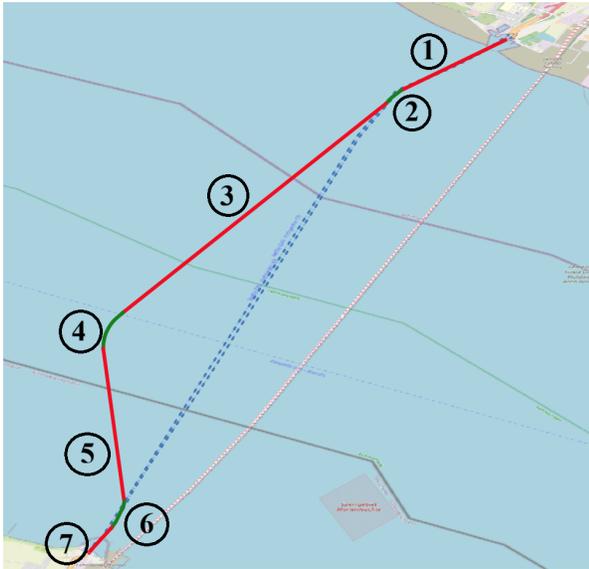
Leveraging the powerful capabilities of LLMs to extract tokens directly from continuous spatio-temporal data (like AIS) is non-trivial and is sensitive to prompt design; as a result, such trajectory descriptions generated by LLMs are not always grounded on actual information [3]. Instead, we propose to generate descriptions in natural language strictly based on contextually-enriched trajectory representations as issued from our workflow. Such descriptions can enable advanced mobility

analysis and semantic reasoning in downstream trajectory tasks (e.g., forecasting, imputation, anomaly detection).

To generate descriptions using LLMs, we construct a prompt (Figure 4) consisting of a system message that establishes the task context and a user message that provides the extracted trajectory information. This approach includes detailed instructions concerning interpretation of the various features (coordinates, timestamp, speed, distance, duration) per episode in the movement, as well as precise guidelines for measurement units (e.g., nautical miles for distance, Beaufort scale for wind force) and any available placemarks (islands, ports, capes, etc.) added from extra context to make sure that the model can interpret and use them properly. Most importantly, specific motion patterns for vessels (e.g., anchorage for long periods of time, possible communication gaps en route) are highlighted.

This prompt also includes a methodology that guides the LLM in synthesizing available information in order to provide its response in two parts:

- A *textual description* in natural language that outlines the trip and its most significant stages. Strictly based on the input trajectory representation in episodes (in JSON format), the LLM should identify port operations (e.g., anchorage), important navigational features (e.g., cruising



(a) Map illustration of input trajectory.

- ① The vessel began its voyage off the coast of Lolland, sailing south-west at about 9 knots for roughly 1.4 NM.
- ② It then performed a short turning manoeuvre still off Lolland, covering a small distance while increasing speed to around 11 knots.
- ③ Continuing south-west, the ship crossed the Fehmarn Belt under a moderate west-westerly breeze (Beaufort 4), sailing at 14 knots for more than 4 NM.
- ④ A brief turn to the south was made within the belt.
- ⑤ A steady southward leg followed at 15 knots, covering about 2.7 NM while the same moderate wind persisted.
- ⑥ The vessel executed another turning manoeuvre to the south-west. Then proceeded at reduced speed (6.5 knots) toward the approach of Putgarden.
- ⑦ A final short, very slow leg (1 knot) brought the ship to the anchorage area in Putgarden port where the vessel stopped for over an hour.

(b) LLM-generated textual description using openai/gpt-oss-120b.

Fig. 5: Trajectory episodes (depicted on map with alternating colors) given as input to LLM for generating a textual description.

at open sea, maneuvers, turns) as well as any proximity to placemarks (e.g., islands, straits, capes).

- A summary of *trip statistics* in JSON format that include the traveled distance, the total duration, and the origin and destination ports (if available). Thus, the LLM must aggregate such features from the input representation, making the necessary calculations (e.g., adding distances, averaging speed values) across the episodes of the trip. In case of any adverse weather conditions identified along the entire trip, these should be also included as well.

The textual description per episode listed in Figure 5 (concerning the same trajectory depicted in Figure 1) is indicative of the LLM responses. Comparing the obtained description of episode ③ with its respective input representations shown in Figure 3, observe how intelligently the LLM grasps information and conveys the vessel’s movement. Strictly based on the given semantic trajectory representations concerning mobility, geospatial, and context features, LLMs seem capable to synthesize a succinct summary that captures the successive episodes of each trip and highlight essential information of each motion path.

V. EMPIRICAL EVALUATION

We have conducted an empirical study by generating descriptions in natural language for contextually-enriched vessel trip representations reconstructed from real-world AIS data. We not only report performance results regarding contextualization and trip statistics, but we also assess the quality of textual descriptions generated by different LLMs. These latter tests evaluate the ability of AI models to understand mobility and spatiotemporal features pertinent to vessel trajectories and issue a succinct, yet factual account of each trip.

A. Experimental Setup

Datasets. We extracted raw AIS positions regarding *passenger* ships for a 3-month period (January–March 2024) available by the Danish Maritime Authority¹². For *context* enrichment, we extracted geospatial datasets for this region from OpenStreetMap (protected areas, capes, peninsulae, straits, islands, ferry lines), OpenSeaMap (traffic separation schemes), and World Port Index (ports). We also used NetCDF sources, namely meteorological data from Copernicus (wind features), and bathymetry features from GEBCO (sea depth).

We applied our workflow to extract context-enriched trajectory representations in JSON format. Finally, we randomly selected 460 trips between 10 specific ports in the region to be given as input to each LLM and generate their descriptions. The selected trips represent 212,203 raw AIS locations.

Models. We tested several open LLMs of varying capabilities (*llama-3.1-8B-instant*, *llama-3.3-70B-versatile*, *qwen-3-32B*, *gpt-oss-20B*, *gpt-oss-120B*) in generating textual descriptions. All LLMs are queried with a temperature 0.1 to reduce output variance. All requests are issued through Groq API¹³ using fixed model versions and the same prompt (Figure 4) to ensure a fair treatment.

B. Performance Results

Context-enriched semantic trajectories. We first examine features in semantic trajectories after applying our processing framework (Section III). Note that this does not include

¹²<https://www.dma.dk/safety-at-sea/navigational-information/ais-data>

¹³<https://groq.com/>

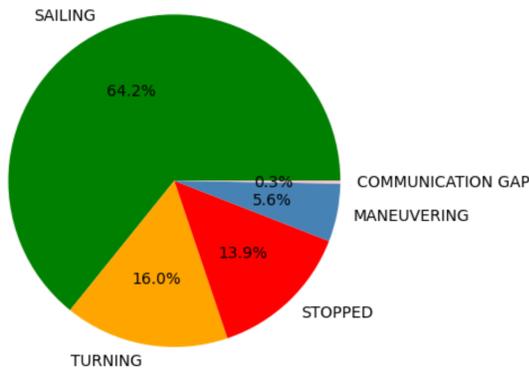


Fig. 6: Distribution of annotated episodes among all trips.

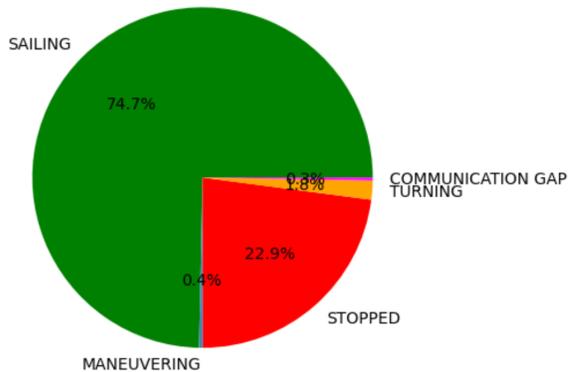


Fig. 7: Cumulative duration of annotated episodes in all trips.

TABLE I: Percentage (%) of episodes enriched with context.

Episode type	geospatial	weather	bythometry
SAILING	63	39	80
TURNING	54	37	98
MANEUVERING	41	12	64
STOPPED	100	0	37
COMMUNICATION GAP	0	0	0

TABLE II: Average response time in generating a trip description using different LLMs.

LLM Model	response time (sec)
llama-3.1-8b-instant	1.01
llama-3.3-70b-versatile	1.12
qwen/qwen3-32b	2.58
openai/gpt-oss-20b	1.45
openai/gpt-oss-120b	2.36

interaction with LLMs. Figure 6 illustrates the breakdown (%) of the type of episodes among all 460 trips detected in the AIS dataset. In total, 3,332 episodes were identified, and most of them represent sailing legs in trips. A significant number (over 21%) of episodes also concerns turns or maneuvers, while stop episodes have a fair share (13.9%). Note that very few episodes concern communication gaps in a trip; recall that each such gap spans at most $\delta = 1$ hour, because larger gap durations are used to split the trajectory into trips.

However, when considering the cumulative duration of each type of episodes, we get another interesting insight. As illustrated in Figure 7, the sailing legs and the stop periods cover most of the duration in the examined trips. Only about 2% of total trip time is spent in turns or maneuvers, although these represent more than one fifth of the detected episodes (Figure 6). The total duration of gaps is almost negligible, given that only rarely a vessel is lost during a trip in this area.

Regarding enrichment with extra context, Table I lists the percentage in each type of episodes that obtained such contextual information. Statistics are shown for each category of sources, i.e., *geospatial* (ports, straits, capes, etc.) mainly extracted from OSM, *weather* (meteorological NetCDF files), and *bathymetry* (also NetCDF data). Note that a large portion of moving episodes (i.e., sailing, turning, or maneuvering) is associated with spatial features (e.g., proximity to capes). Most importantly, the respective port is identified for all stops. Weather data was available for a significant part of moving episodes, whereas bathymetry data was associated for the vast majority of sailing and turning legs of trips. Of course, no context can be associated with communication gaps, as vessel whereabouts are unknown.

Generating trip descriptions with LLMs. Based on semantically enriched trajectory representations, we employed several LLMs to generate descriptions in natural language per trip. Next, we report performance statistics, deferring assessment of their quality for Section V-C.

Table II depicts the average time required by each LLM in order to return a trip description. Apparently, smaller models can offer their response in about a second; larger models take longer, but still manage to answer within a few seconds on average, which highlights their potential for practical use by maritime stakeholders

As each response also includes *trip statistics* (the second output requested from LLMs), we also evaluated the accuracy of the returned estimates. We report statistics (mean, standard deviation, maximum) over *deviations (%)* in *total trip duration* (Table III) and *travelled distance* (Table IV) from the respective ground truth values (computed over original AIS locations per trip). Results in **bold** indicate LLMs with the best performance. Despite specific prompt instructions on how to handle spatiotemporal measurements and units, estimates from models may diverge from actual values. Unsurprisingly, smaller models (like *llama-3.1-8B-instant*) seem less capable in accurately estimating trip statistics and incur very significant over- or underestimations. In contrast, larger models (like *openai/gpt-oss-120b*) offer much better estimates, although sometimes they can also be off the actual value by orders of magnitude (as indicated by the max % deviation). This finding underscores current limitations of LLMs in conducting accurate calculations, especially concerning mobility features crucial for trajectory analysis.

System:

ROLE AND OBJECTIVE: You are a data scientist who studies trips of vessels, i.e., the route of ships sailing between ports or locations at open sea. You are given AI-generated textual descriptions of such trips, including information about anchorage in ports, duration of each stop, sailing for varying time intervals, and occasionally turning maneuvers when a vessel changes its direction of movement. You must evaluate the quality of these descriptions on factuality, accuracy, and comprehensiveness with respect to the original input information.

APPROACH AND LIMITATIONS:

Each description in PLAIN TEXT outlines the trip of a vessel. In a separate paragraph it describes each major stage of the trip, such as port operations (e.g., anchorage), important navigational features (e.g., cruising at open sea, maneuvering, turning) as well as any proximity to placemarks (e.g., islands, straits, capes). In generating each description, the model was asked NOT to include information from any external source, but only use information strictly from the input JSON. The description does NOT include any explanations or reasoning regarding any stage of the trip.

- (1) All vessel locations are specified in (longitude, latitude) coordinates in WGS84 (EPSG:4326) Coordinate Reference System.
- (2) Distances are always expressed in nautical miles and are calculated using the Haversine formula on an ellipsoid.
- (3) Speed values are in knots, i.e., nautical miles per hour.
- (4) Durations are in seconds. Wind intensity is in Beaufort scale.
- (5) Placemarks indicate proximity to islands, ports, straits, capes, etc.
- (6) Note that the motion patterns and speed of vessels absolutely depend on their type (e.g., Passenger, Cargo, Tanker, Pleasure craft, etc.), and some vessels may be anchored for long intervals.
- (7) Due to COMMUNICATION GAPS, a vessel may suddenly appear far away from its previously known position.

EVALUATION METHODOLOGY:

Evaluate the description in each one of the following criteria:

- (i) *Relevance*: Did it actually address the user's request and provided a reliable account of the trip according to the specified guidelines?
- (ii) *Faithfulness*: Did it stick to the facts stipulated in the input instead of wandering into hallucinations?
- (iii) *Correctness*: Was it accurate from start to finish or discrepancies (overestimations or underestimations) are observed in several measurements (e.g., distance, speed, duration) along the trip?

Please rate this description with scores on a 1-5 scale for each quality criterion:

- 1 - Poor**: description below expectations with several errors and inconsistencies.
- 2 - Unsatisfactory**: description partially meets requirements, occasionally missing important aspects of the trip.
- 3 - Fair**: description meets requirements offering a consistent account of the entire trip.
- 4 - Very good**: high quality, coherent description.
- 5 - Excellent**: description significantly exceeds all expectations.

EVALUATION RESULT:

You must provide your evaluation scores as well as a brief explanation of your scoring in JSON format, as in the following example:

```
{"relevance_score": 4, "faithfulness_score": 5, "correctness_score": 3, "explanation": "The description was clear and comprehensive, accurately covering all stages of the trip but certain measurements on distance and speed deviated from actual values."}
```

User:

INPUT: This JSON array represents the trip of a vessel: ```json{ ... }```

OUTPUT: This is the AI-generated textual description of this trip: `"NL DESCRIPTION"`

TASK: Evaluate the quality of the output description in terms of relevance, faithfulness, and correctness with respect to the input JSON.

Fig. 8: Prompt template for evaluating the generated trajectory descriptions. The LLM receives as input a `json array` with distilled context-enhanced representation about a vessel trip and evaluates the quality of the `generated NL description` in terms of relevance, faithfulness, and correctness.

TABLE III: Statistics on *deviations (%) of trip durations* estimated by LLM models w.r.t. ground truth.

LLM Model	mean	stdev	max
llama-3.1-8b-instant	113.72	1047.73	19017.24
llama-3.3-70b-versatile	13.68	16.79	85.00
qwen/qwen3-32b	7.86	17.88	206.38
openai/gpt-oss-20b	1.87	6.74	88.10
openai/gpt-oss-120b	1.53	6.47	88.10

TABLE IV: Statistics on *deviations (%) of travelled distance* per trip as estimated by LLM models w.r.t. ground truth.

LLM Model	mean	stdev	max
llama-3.1-8b-instant	13.79	26.36	248.28
llama-3.3-70b-versatile	2.44	4.52	34.55
qwen/qwen3-32b	4.92	43.48	848.22
openai/gpt-oss-20b	1.06	4.29	35.32
openai/gpt-oss-120b	0.89	4.12	35.32

TABLE V: Qualitative evaluation of textual descriptions of vessel trips generated by diverse LLM models.

LLM Model	relevance	faithfulness	correctness
llama-3.1-8b-instant	3.90	3.84	3.27
llama-3.3-70b-versatile	4.47	4.77	4.36
qwen/qwen3-32b	4.42	4.48	3.82
openai/gpt-oss-20b	4.69	4.81	4.35
openai/gpt-oss-120b	4.91	4.96	4.69

C. Qualitative Assessment of Generated Descriptions

Evaluating the quality of such LLM-generated narratives is a challenging task due to absence of ground truth. To tackle this, we opted to employ an LLM as evaluator, according to the emerging paradigm of “*LLM-as-a-Judge*” [30]. To ensure that such evaluation would be as objective, consistent, and reliable as possible, we employed a different LLM from another AI provider, namely *gemini-2.5-flash*¹⁴ using in-context learning. More specifically, we designed another prompt specifically for this evaluation (shown in Figure 8) in order to assess the quality of each generated description individually. The evaluator is given the same input representation of a trip (i.e., a context-enriched semantic trajectory in JSON format) with the same guidelines for generating descriptions. An LLM-generated description (using one of the five models mentioned in Section V-A, *unknown* to the evaluator) is also given, and the evaluator is asked to rate its quality using three *evaluation criteria* (aka *dimensions* [30]):

- *Relevance* rates whether the description offers a relevant response by giving a reliable overall account of the trip according to the stipulated guidelines.
- *Faithfulness* assesses whether the description is actually grounded on the facts given in the input JSON representation without wandering into hallucinations.
- *Correctness* evaluates the accuracy of the entire description and examines if discrepancies (over- or underestimations) are observed in measurements (e.g., distance, speed, duration) along the trip.

The evaluator assigns *scores* to each description according to the aforementioned dimensions. Each dimension is scored on a scale of 1 to 5, ranging from worst to best:

- 1) *Poor*: The description contains many errors (e.g., misinterpretation of values or units in duration of an episode) or factual inconsistencies (e.g., ignoring a leg of the trip). This score rates such severely distorted descriptions below expectations.
- 2) *Unsatisfactory*: Major discrepancies in stated values are observed or some important aspects of the trip may occasionally be missed. Thus, the description only partially meets requirements and cannot be considered reliable.
- 3) *Fair*: No significant errors are found, and all legs of the trip are mentioned. Such a description meets requirements, offering a consistent account of the entire trip.

¹⁴<https://ai.google.dev/gemini-api/docs/models/gemini-2.5-flash>

- 4) *Very good*: This is a high-quality, factual, coherent, and comprehensive description of the trip without bias.
- 5) *Excellent*: The description significantly exceeds all expectations. It creatively compiles all available information into a fluent, accurate, and reliable trip narrative.

The evaluator is also asked to justify its scores with an *explanation*; this can offer insight into the assessment process.

We performed this qualitative evaluation of all trip descriptions generated by the five models (in total, 2300 prompt requests). Table V reports averages in each score per LLM. Unsurprisingly, the most powerful model (*openai/gpt-oss-120b*) achieves the finest descriptions by all three criteria. With very few exceptions, it sticks to the input facts, and rarely its reported measurements deviate from the actual ones. In contrast, smaller models fall behind significantly, especially with respect to correctness. For instance, *lama-3.1-8b-instant* often exaggerated by orders of magnitude in its reported values on speed or duration of many episodes, due to its limited calculation capabilities as mentioned before. Other models with greater capacity to learn (like *llama-3.3-70b-versatile*, *qwen/qwen3-32b*, and *openai/gpt-oss-20b*) offer very good responses with factual narratives, but occasional inaccuracies in measurements. Again, this finding indicates the current limitations of LLMs in interpreting and correctly processing even simple spatiotemporal features like distance or speed.

VI. CONCLUDING REMARKS

In this paper, we proposed a methodology for constructing contextually-enriched representations of vessel trajectories collected from AIS messages. By semantically annotating important locations along each vessel’s course that signify mobility events like stops, turns, or communication gaps, we are able to segment long trajectories into distinct trips, each one consisting of a sequence of episodes. Geographical and maritime features from open sources (e.g., ports, capes, islands, or traffic separation zones), as well as meteorological and bathymetry data can add rich contextual information to each episode, providing important insights into vessel activity. Finally, such trajectory representations can be given as input to powerful LLMs in order to generate descriptions in natural language that effectively summarize entire trips (each possibly spanning thousands of raw AIS locations) in just a few sentences and highlight their most important mobility patterns.

We plan to improve such descriptions by taking into account vessel interactions across their trips (i.e., not each trajectory in isolation), offering richer context crucial for maritime safety. Utilizing such descriptions can be advantageous in downstream tasks, like trajectory forecasting for optimized navigation, imputation of trajectory segments missed due to gaps, or anomaly detection (e.g., identify detours), etc.

Acknowledgments. This work has been partially supported by project MIS 5154714 of the National Recovery and Resilience Plan Greece 2.0 funded by the European Union under the NextGenerationEU Program.

REFERENCES

- [1] D. Yang, L. Wu, S. Wang, H. Jia, and K. X. Li, "How Big Data Enriches Maritime Research—a Critical Review of Automatic Identification System (AIS) Data Applications," *Transport reviews*, vol. 39, no. 6, 2019.
- [2] Y. Yang, Y. Liu, G. Li, Z. Zhang, and Y. Liu, "Harnessing the power of machine learning for ais data-driven maritime research: A comprehensive review," *Transportation research part E: logistics and transportation review*, vol. 183, 2024.
- [3] Y. Xu, R. Yang, Y. Zhang, J. Lu, M. Zhang, Y. Wang, L. Su, and Y. Fu, "Trajectory Prediction Meets Large Language Models: A Survey," *arXiv:2506.03408*, 2025.
- [4] Z. Yan, D. Chakraborty, C. Parent, S. Spaccapietra, and K. Aberer, "Semantic Trajectories: Mobility Data Computation and Annotation," *ACM Transactions on Intelligent Systems and Technology*, vol. 4, no. 3, 2013.
- [5] C. Parent, S. Spaccapietra, C. Renso, G. Andrienko, N. Andrienko, V. Bogorny, M. L. Damiani, A. Gkoulalas-Divanis, J. Macedo, N. Pelekis, Y. Theodoridis, and Z. Yan, "Semantic Trajectories Modeling and Analysis," *ACM Computing Survey*, vol. 45, no. 4, 2013.
- [6] A. Bakdi, I. K. Glad, E. Vanem, and Ø. Engelhardtson, "Ais-based multiple vessel collision and grounding risk identification based on adaptive safety domain," *MDPI Journal of Marine Science and Engineering*, 2019.
- [7] N. Chen, A. Yang, H. Wu, L. Chen, W. Xiong, and N. Jing, "SEMINT: An LLM-Empowered Long-Term Vessel Trajectory Prediction Framework," *Journal of Geographical Information Science*, 2025.
- [8] H. Liu, T. Li, H. Wang, K. Torp, T. Zhang, Y. Li, and C. S. Jensen, "VISTA: Knowledge-driven interpretable vessel trajectory imputation via large language models," *arXiv:2601.06940*, 2026.
- [9] G. Spiliopoulos, A. Troupiotis-Kapeliaris, K. Patroumpas, N. Liapis, D. Skoutas, D. Zissis, and N. Bikakis, "Data-Driven Trajectory Imputation for Vessel Mobility Analysis," in *Intl. Conf. on Extending Database Technology (EDBT)*, 2026.
- [10] F. Ma, X.-m. Wang, C. Chen, X.-b. Xu, and X.-p. Yan, "Navigation-GPT: A Robust and Adaptive Framework Utilizing Large Language Models for Navigation Applications," *arXiv:2502.16402*, 2025.
- [11] H. Park, J. Jung, M. Seo, H. Choi, D. Cho, S. Park, and D.-G. Choi, "AIS-LLM: A Unified Framework for Maritime Trajectory Prediction, Anomaly Detection, and Collision Risk Assessment with Explainable Forecasting," *arXiv:2508.07668*, 2025.
- [12] Z. You and Y. Lin, "Interpretable Vessel Behavior Inference from AIS Trajectories via CoT-Enhanced LLMs," in *Conf. on Big Data & Artificial Intelligence & Software Engineering*, 2025.
- [13] Y. Liu, W. Xiong, N. Chen, and F. Yang, "VTLLM: A Vessel Trajectory Prediction Approach Based on Large Language Models," *Journal of Marine Science and Engineering*, vol. 13, no. 9, 2025.
- [14] S. Wang, Z. Bao, J. S. Culpepper, and G. Cong, "A Survey on Trajectory Data Management, Analytics, and Learning," *ACM Computing Survey*, vol. 54, no. 2, 2021.
- [15] G. Atluri, A. Karpatne, and V. Kumar, "Spatio-Temporal Data Mining: a Survey of Problems and Methods," *ACM Computing Surveys*, vol. 51, no. 4, 2018.
- [16] M. Musleh and M. F. Mokbel, "Let's Speak Trajectories: a Vision to Use NLP Models for Trajectory Analysis Tasks," *ACM Trans. Spatial Algorithms Syst. (TSAS)*, vol. 10, no. 2, 2024.
- [17] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding," in *Conf. of the North American chapter of the association for computational linguistics*, 2019.
- [18] J. Si, J. Yang, Y. Xiang, H. Wang, L. Li, R. Zhang, B. Tu, and X. Chen, "TrajBERT: BERT-Based Trajectory Recovery with Spatial-Temporal Refinement for Implicit Sparse Trajectories," *IEEE Transactions on Mobile Computing*, vol. 23, no. 5, 2023.
- [19] M. Musleh and M. F. Mokbel, "Kamel: a Scalable Bert-Based System for Trajectory Imputation," *VLDB Endowment (PVLDB)*, vol. 17, no. 3, 2023.
- [20] H. Yang, A. Yao, C. Whalen, and G. Mai, "BERT4Traj: Transformer Based Trajectory Reconstruction for Sparse Mobility Data," *arXiv:2507.03062*, 2025.
- [21] Y. Liang, H. Wen, Y. Xia, M. Jin, B. Yang, F. Salim, Q. Wen, S. Pan, and G. Cong, "Foundation Models for Spatio-Temporal Data Science: a Tutorial and Survey," in *ACM Intl. Conf. on Knowledge Discovery and Data Mining (KDD)*, 2025.
- [22] Z. Zhang, H. Amiri, Z. Liu, L. Zhao, and A. Züfle, "Large Language Models for Spatial Trajectory Patterns Mining," in *Workshop on Geospatial Anomaly Detection*, 2024.
- [23] X. Wang, M. Fang, Z. Zeng, and T. Cheng, "Where Would I Go Next? Large Language Models As Human Mobility Predictors," *arXiv:2308.15197*, 2023.
- [24] S. Wang, K. Wu, R. Qiao, R. Wang, and C. Cai, "Towards Maritime Safety with Industrial Multimodal Models: Compliance Reasoning from Heterogeneous Maritime Data," *Applied Soft Computing*, 2026.
- [25] X. Guo, S. Yu, J. Zhang, H. Bi, X. Chen, and J. Liu, "A Natural Language-Based Automatic Identification System Trajectory Query Approach Using Large Language Models," *ISPRS International Journal of Geo-Information*, vol. 14, no. 5, 2025.
- [26] G. Merten, G. Dejaegere, and M. Sakr, "Using LLMs for Analyzing AIS Data," *arXiv:2504.07557*, 2025.
- [27] J. K. Mbuya, D. Pfoser, and A. Anastasopoulos, "Trajectory Anomaly Detection with Language Models," in *ACM Intl. Conf. on Advances in Geographic Information Systems (SIGSPATIAL)*, 2024.
- [28] N. Chen, A. Yang, L. Chen, H. Wu, and N. Jing, "MSCE: Empowering Vessel Identity Anomaly Detection with Multimodal LLMs," *IEEE Transactions on Aerospace and Electronic Systems*, 2025.
- [29] G. Fikioris, K. Patroumpas, A. Artikis, M. Pitsikalis, and G. Paliouras, "Optimizing Vessel Trajectory Compression for Maritime Situational Awareness," *Geoinformatica*, vol. 27, no. 3, 2022.
- [30] J. Gu, X. Jiang, Z. Shi, H. Tan, X. Zhai, C. Xu, W. Li, Y. Shen, S. Ma, H. Liu, S. Wang, K. Zhang, Z. Lin, B. Zhang, L. Ni, W. Gao, Y. Wang, and J. Guo, "A survey on llm-as-a-judge," *The Innovation*, 2026.