CORALL: A COLREGs-Guided Risk-Aware LLM for Decision-Making in Maritime Autonomous Surface Ships

Klinsmann Agyei, Pouria Sarhadi, Wasif Naeem

Abstract

The paper introduces a novel approach to utilising Large Language Models (LLMs) for real-time COLREGs-based decision-making in collision encounters. The COLREGs (Collision Regulations), published by the International Maritime Organisation (IMO), have served as the baseline regulations for empirical maritime collision avoidance in open seas to ensure vessel safety. As COLREGs were originally defined for human seafarers, translating them into computer language for autonomous operations has been extensively investigated in recent decades, attracting significant interest but lacking a universal solution. A key aspect is the need for explainable decision-making, either for autonomous operations or as captain assistance. This study leverages the power of LLMs to develop such a decision-making algorithm. The proposed COLREGs-guided Risk-Aware LLM (CORALL) algorithm consists of a high-level, risk-aware decision-maker and an execution layer that implements path tracking and collision avoidance in the presence of external disturbances. A framework for risk-aware, COLREGs-compliant LLM-based decision-making is devised, and its performance in online operation on a high-speed ship model is demonstrated. The LLM processes navigation outputs and risk indices, identifies the COLREGs encounter type, and makes decisions with accompanying explanations. The proposed solution is tested on all the 22 Imazu's benchmark problems, evaluated using the Imazu benchmark scenarios, a widely-accepted standard in maritime collision avoidance research comprising 22 progressively complex encounter situations. To complement the study, the proposed algorithm is implemented and verified on

Klinsmann Agyei and Pouria Sarhadi are with the School of Physics, Engineering and Computer Science, University of Hertfordshire, Hatfield, United Kingdom (k.agyei@herts.ac.uk, p.sarhadi@herts.ac.uk)

Wasif Naeem is with the School of Electronics, Electrical Engineering and Computer Science Queen's University Belfast, Belfast, UK (w.naeem@qub.ac.uk)

a low-cost processor in a Hardware-in-the-Loop (HIL) setup, with end-to-end execution times measured during real-time operation. This risk-aware AI solution, thought to be the first of its kind, opens a door to future research on advanced decision-making algorithms in safe maritime transport and similar operational environments.

Index Terms

Maritime Autonomous Surface Ship, COLREGs, LLM, AI, Risk, Decision Making, Collision Avoidance.

I. Introduction

Maritime Autonomous Surface Ships (MASS) could significantly transform maritime operations, offering benefits such as enhanced safety, reduced costs, and lower carbon emissions through optimised route planning. Safe mission planning and decision-making are critical aspects for MASS, requiring global and local path planning, world modelling (including environmental conditions and spatial constraints), and consideration of vehicle dynamics and manoeuvrability limitations. MASS must also comply with the Convention on the International Regulations for Preventing Collisions at Sea (COLREGs) set by the IMO [1]. This involves safety-critical online risk assessment to detect collision risks and execute Collision Avoidance (COLAV) actions in line with these regulations. In 2022 alone, maritime accidents caused losses exceeding \$1 billion, highlighting the need for safer navigation systems [2]. The 2021 Ever Given grounding in the Suez Canal caused weekly losses of \$10 billion, while the August 2024 collision between Chinese and Philippine vessels escalated tensions over COLREGs adherence [3]. Furthermore, a 10-year global forecast projects a significant growth in the autonomous ships market, increasing from \$6.1 billion to \$14 billion [4]. Autonomous ships, especially at the IMO's full autonomy Level 4, must handle these complexities with greater reliability. However, the textual COLREGs, written for human interpretation, can be vague and challenging to apply in real-time, especially in multi-criteria decision-making situations that require rule interpretation across various scenarios. This complexity is further compounded by the need to develop decision-making capabilities that can translate qualitative COLREGs rules like 'safe distance', 'good seamanship', 'in doubt', 'look-out' and 'ample time', which require context-aware interpretation that varies with vessel characteristics and environmental conditions into actionable COLAV decisions. This challenge becomes particularly complex when making decisions in multi-vessel encounters, where MASS must assess multiple collision risks simultaneously while ensuring decisions remain compliant with COLREGs across various encounter types.

Recent advances in Machine Learning (ML) have shown promise in maritime decision-making [5], yet existing approaches often struggle to justify their COLAV decisions or maintain consistent COLREGs compliance. An early study, the COLREGs3 project [6], identified potential benefits of Large Language Models (LLMs) in this field, focusing on scene-based tests but lacking dynamic simulations. It was noted that this could lead to poor outcomes and rightly highlighted the need for re-training LLMs [6]. Similarly, research on advanced ship collision avoidance highlighted that introducing LLMs could represent a paradigm shift in ship decision-making [7]. Hybrid approaches combining MPC with COLREGs compliance have shown particular promise, such as the work by Eriksen et al. [8], which presented a three-layered collision avoidance system for ASVs compliant with COLREGs Rules 8 and 13–17, integrating energy-optimised trajectory planning with model predictive control. Our initial work [9] demonstrated the potential of LLMs in maritime decision-making through validation across four single-vessel encounter types: head-on, overtaking, crossing from starboard, and crossing from port. However, real-world maritime environments frequently require decisions involving multiple vessels with varying risk levels, demanding robust integration of risk assessment with COLREGs interpretation.

The emergence of LLMs with advanced reasoning capabilities presents an exciting opportunity to address these multi-vessel challenges. Their ability to process natural language rules and generate contextual explanations, combined with real-time risk assessment, offers a promising solution for complex maritime navigation scenarios requiring sophisticated reasoning and clear decision justification.

This paper introduces CORALL, a novel framework that advances the application of LLMs in maritime autonomy by integrating risk-aware COLREGs-compliance decision-making with single and multi-vessel encounters. The system combines the interpretative capabilities of LLMs with quantitative risk metrics to generate explainable, regulation-compliant decisions for COLAV. Our key contributions encompass real-time interpretation of maritime situations with explicit

COLREGs rule identification and risk assessment integration, handling of multi-vessel encounters through risk-prioritised decision making, clear and explainable reasoning for all decisions based on both regulatory requirements and quantitative risk metrics, and consistent performance evaluation across various encounter types, from single-vessel scenarios to complex multi-vessel interactions. Furthermore, we verify the proposed approach through Processor-in-the-Loop (PIL) simulation, a basic form of Hardware-in-the-Loop (HIL) testing commonly used to assess real-time implementability of autonomous systems at an early design stage. The PIL setup integrates the CORALL algorithm running on an embedded platform with a real-time vessel dynamics simulator, demonstrating feasibility for onboard deployment.

The remainder of this paper is organised as follows: Section II provides a review of related work across two key areas: COLREGs-based decision making, and applications of LLMs in autonomous decision systems. Section III presents our approach, beginning with CORALL's system architecture and core components, followed by the execution layer design, CORALL algorithm and decision-making framework, and implementation details that integrate these various components. Section IV presents extensive evaluation of results across the Imazu test scenarios, processor-in-the-loop validation, and an ablation study examining component contributions. Finally, Section V concludes with a discussion of implications and future research directions.

II. Related Studies

A. COLREGs-Based Decision Making

Maritime navigation safety is fundamentally governed by the International Regulations for Preventing Collisions at Sea (COLREGs), established by the IMO [10]. As illustrated in Fig. 1, the application of COLREGs during a voyage follows a hierarchical decision-making process, beginning with basic safety assessments and progressing through various encounter scenarios. The relevant COLREGs rules for autonomous COLAV are as follows [1]:

- **Rule 7— Risk of Collision:** 'Every vessel shall use all available means appropriate to the prevailing circumstances and conditions to determine if risk of collision exists. If there is any doubt, such risk shall be deemed to exist.'

- **Rule 8— Action to Avoid Collision:** 'Any action taken to avoid collision shall be positive, made in ample time and with due regard to the observance of good seamanship.'
- **Rule 13— Overtaking:** 'Any vessel overtaking any other shall keep out of the way of the vessel being overtaken.'
- **Rule 14— Head-on Situation:** 'When two power-driven vessels are meeting on reciprocal or nearly reciprocal courses so as to involve risk of collision, each shall alter her course to starboard so that each shall pass on the port side of the other.'
- **Rule 15— Crossing Situation:** 'When two power-driven vessels are crossing so as to involve risk of collision, the vessel which has the other on her starboard side shall keep out of the way and shall, if the circumstances of the case admit, avoid crossing ahead of the other vessel.'
- **Rule 16— Action by Give-way Vessel:** 'Every vessel which is directed to keep out of the way of another vessel shall, so far as possible, take early and substantial action to keep well clear.'
- **Rule 17— Action by Stand-on Vessel:** 'Where one of two vessels is to keep out of the way, the other shall keep her course and speed. The latter vessel may, however, take action to avoid collision by her manoeuvre alone, as soon as it becomes apparent to her that the vessel required to keep out of the way is not taking appropriate action.'

It should be noted that COLREGs are primarily applicable when vessels are navigating in open or moderately congested waters where collision risk can be clearly assessed between specific vessels. In highly congested or constrained areas such as ports and narrow channels, local traffic rules and Vessel Traffic Services (VTS) may override or complement COLREGs. The CORALL framework presented in this work specifically addresses open-water collision avoidance scenarios governed by COLREGs Rules 8 and 13–17, which represent the majority of encounters in international shipping lanes.

Within this open-water context, the application of COLREGs follows a hierarchical decision-making process from initial safety assessment through encounter classification to action determination, as illustrated in Fig. 1. While COLREGs define encounter types through geometric descriptions, such as Rule 13's overtaking definition of approaching “from a direction more than 22.5 degrees abaft her beam" (corresponding to $112.5°$ from ahead), these boundaries serve as guidance rather than rigid thresholds. In practice, encounter classification depends on contextual factors including relative motion, risk assessment, and environmental conditions that are not easily quantifiable into fixed rules. This inherent ambiguity in COLREGs interpretation is precisely why CORALL employs an LLM-based approach: the LLM receives the geometric relationships and risk metrics as input, but makes contextual decisions rather than applying deterministic threshold logic. The system prompt encodes COLREGs principles and the rule hierarchy from Fig. 1 as guidance, while allowing the LLM to exercise judgement in ambiguous situations.

These rules form the foundation of maritime COLAV decision-making. However, their effective implementation in MASS presents significant challenges due to three key factors. First, COLREGs rely heavily on qualitative terms which require context-aware interpretation that varies with vessel characteristics and environmental conditions. Second, in multi-vessel scenarios, MASS must prioritise and coordinate responses to potentially conflicting COLREGs requirements, particularly challenging when different rules apply simultaneously to various encounter situations. Third, decisions must maintain temporal consistency to prevent oscillatory behaviour while remaining responsive to dynamic changes in the maritime environment.

Various approaches have attempted to address these challenges. Traditional geometric approaches established foundational frameworks, with Kuwata et al. [11] developing Velocity Obstacles (VO) methods for basic COLREGs-based decision-making. Campbell et al. [12] enhanced this approach by incorporating dynamic risk assessment through fundamental collision prediction metrics. However, these methods struggled to capture the nuanced reasoning required by COLREGs in complex scenarios. Rule-based systems emerged as a natural progression, attempting to directly encode COLREGs requirements [13]–[16]. Enhanced by fuzzy logic implementations [17], these systems better handle uncertainties but lack flexibility across diverse scenarios.

Recent ML approaches have shown promise, with deep reinforcement learning demonstrating capabilities in complex navigation decisions [18]–[20]. Probabilistic models and neural networks have enabled dynamic risk assessment [5], [21], though these AI-driven systems often operate

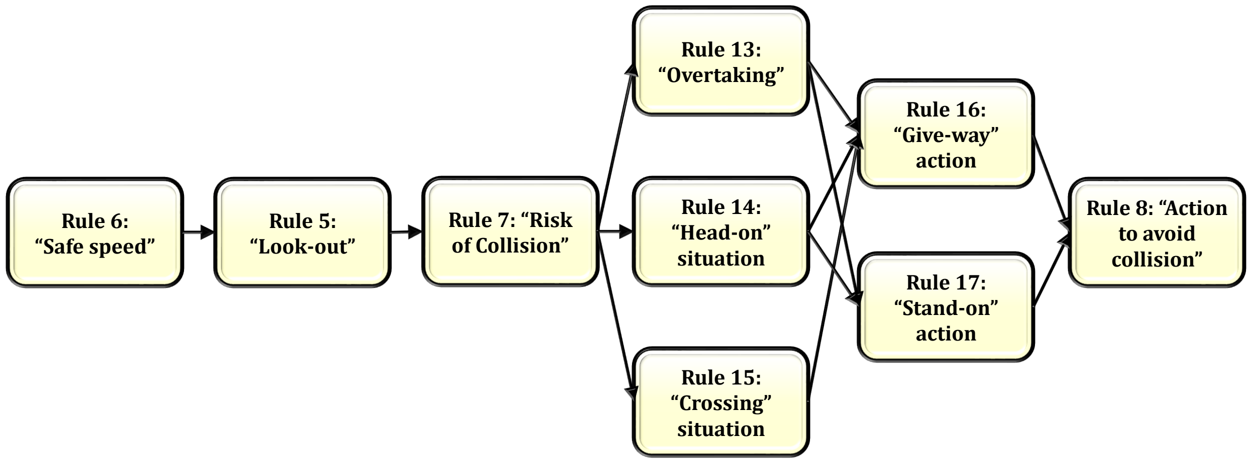

Fig. 1. Sequential flow of COLREGs, derived from IMO's textual rules [1]

as 'black boxes,' making it difficult to verify their reasoning or ensure COLREGs compliance.

Hybrid approaches have attempted to address these limitations by combining methodologies. Wang et al. [22] integrated geometric methods with ML to maintain explicit COLREGs compliance. However, challenges remain in providing human-interpretable explanations and effectively handling complex multi-vessel encounters.

Our work addresses these limitations through an LLM-based framework that bridges the gap between human-interpretable maritime regulations and autonomous control systems while maintaining acceptable risk assessment capabilities and the advantages of traditional methods in maritime navigation.

B. Application of LLMs in Autonomous Decision Making

Recently, there has been growing interest in exploiting LLMs for smart decision-making systems, with intensive research exploring their capabilities and limitations in addressing broader and more complex problems beyond Autonomous Vehicles (AVs) [23], [24]. [24] highlights comparable traits of LLMs in solving financial and psychological problems, yet observes deficiencies in human-level decision-making for medical challenges. For instance, [25] reports that GPT-3.5's average performance in medical decision-making fell 10% short of the mean passing mark over the past two years. The research highlights that, to approach human expert-level performance, further development is required. Another study [26] reveals greater shortcomings when compared to top-tier and randomly selected doctors, though GPT's performance remains notable. As LLMs are still in their infancy, researchers anticipate better results in the future and are actively exploring improved learning and deployment strategies. Our study contributes to this by proposing an LLM based real-time decision-making in COLREGs-compliant MASS navigation, a domain where LLM traits could be particularly beneficial given the written and subjective nature of these rules.

In the self-driving car sector, companies like Wayve have demonstrated LLM integration to interpret complex traffic scenarios and generate natural language explanations for vehicle decisions [27]. Similarly, NVIDIA researchers have shown how LLMs can enhance perception and decision-making in AVs through their published work on language-conditioned driving policies [28]. These implementations have shown particular promise in handling edge cases and providing human-interpretable reasoning for vehicle actions. The adoption of LLMs extends beyond self-driving cars. In robotics, researchers have demonstrated successful applications in tasks ranging from manipulation planning to human-robot interaction. Notable work includes Code as Policies [29], which show how LLMs can effectively translate natural language commands into executable robot actions. The field of Unmanned Aerial Vehicles (UAVs) has similarly embraced LLM technology, with recent research demonstrating their effectiveness in mission planning, and regulatory compliance [30], [31]. Table I summarises the landscape of some notable LLM applications in autonomous systems, reflecting the considerable activities in this field. It provides insights into key attributes of each approach, including the application, input-output structure, and the model used. The survey reveals that GPT-3.5, GPT-4, and Llama are the predominant architectures, with GPT-4 being particularly favoured for its advanced reasoning capabilities. These implementations demonstrate diverse functionalities that parallel key challenges in maritime autonomy including:

- Motion Planning: LLMs process sensor data and generate trajectory plans, similar to our need for COLREGs-compliant path planning in maritime navigation;
- Traffic Control: Interpretation of complex traffic scenarios and rule compliance, directly analogous to multi-vessel encounter situations in maritime environments;
- Policy Adaptation: Dynamic adjustment of behaviour based on environmental conditions,

reflecting the need for adaptive decision-making in varying maritime scenarios;

- Natural Language Integration: Processing rules and generating explanations, essential for interpreting qualitative COLREGs rules;
- Safety Assessment: Risk evaluation and safety-constraint verification, mirroring our requirements for maritime collision risk assessment;

Beyond LLM-specific applications, substantial progress has been made in foundational technologies that support autonomous decision-making across vehicle domains. [32] proposed a situational assessment method based on stochastic models and Gaussian distributions for intelligent vehicles, enabling collision probability estimation through trajectory prediction under uncertain conditions. [33] developed a robust lateral trajectory following control solution using model predictive control, which addresses the nonlinear dynamics and uncertainties inherent in lateral motion control. Furthermore, the interacting multiple model (IMM) approach [34] combines vehicle physics models with manoeuvre recognition using hidden Markov models to achieve accurate short-term and long-term trajectory prediction. These advances provide complementary capabilities that, when integrated with LLM-based reasoning, can enhance the overall decision-making framework for autonomous vehicles in both terrestrial and maritime domains. Both the LLM-based applications summarised in Table I and the foundational technologies discussed above share a common thread of translating high-level rules or situational understanding into actionable decisions while maintaining explainability and safety. This requirement is particularly crucial in maritime navigation, where decisions must adhere to COLREGs while remaining interpretable to human operators. Building upon these advances in autonomous systems, we present CORALL, a novel framework that integrates LLM-based decision-making with real-time risk assessment for COLREGs-compliant autonomous navigation. Our approach uniquely combines the interpretative capabilities of LLMs with quantitative risk metrics to generate explainable, regulation-compliant decisions in maritime COLAV scenarios.

III. The Proposed Method

This section presents CORALL, a novel framework that integrates LLMs with traditional maritime navigation systems to enable intelligent, risk-aware decision-making in autonomous

TABLE I
COMPARISON BETWEEN LLM-BASED ALGORITHMS IN AV APPLICATIONS

Ref.	LLM Task	Input	Output	Model
Chen et al, 2024 [27]	Explainable driving (Driving with LLM)	Vectorised representation of objects in a scene and high-level goals	Driving actions, explanations	GPT-3.5, LLaMa-7b-hf
Lin et al, 2024 [35]	Motion planner repair (Dr Planner)	Planner details, trajectories	Fixed planners	GPT-4-Turbo
Wang et al, 2024 [36]	Traffic signal control (LLM-Assisted light)	Traffic data	Signal control decisions	GPT-4
Li et al, 2024 [37]	Policy adaptation (LLaDA)	Traffic rules and scenarios	Adapted driving policies	GPT-4
Yildirim et al, 2024 [38]	Highway motion planning (HighwayLLM)	Current state and RL model output(meta level actions)	Waypoints and Trajectories	GPT-4
Zhong et al, 2024 [39]	Vision-based planning (UAVs with Dynamic Obstacle Trajectory Prediction)	Vision data, obstacle trajectories	Navigation commands	GPT-4
Ma et al, 2024 [40]	Code generation (Lampilot)	API documentation, human instructions, and driving context	Executable programs	Llama 2, PaLM 2, GPT-4
Fu et al, 2024 [41]	AV testing (LimSim++)	Visual Information, Scenario Cognition, Task Description	Driving decisions, reasoning	GPT-3.5, GPT-4, GPT-4V
Kong et al, 2024 [42]	Safe, compliant driving (Superalignment)	User requests, context, and safety rules	Safe, compliant, and potentially modified LLM responses	GPT-3.5, LLaMa2
Cui et al, 2024 [43]	Natural language driving (Drive as You Speak)	Natural language instructions and potentially contextual information	Driving commands	GPT-4
Fu et al, 2024 [44]	Human-like driving (Drive like a Human)	Scene context, past driving experience	Actions, explanations	GPT-3.5
Liang et al, 2023 [29]	Code generation (Code as Policies)	Natural language instructions, goals	Executable code	Codex code-davinci-002
Wang et al, 2023 [45]	Safety-focused planning	Scene info, actions	Safe plans	GPT-4
Wang et al, 2023 [46]	Instruction translation (Prompt a robot to walk)	Natural language prompts or instructions related to walking or locomotion	A sequence of actions or commands that the robot can execute	GPT-4
Wen et al, 2023 [47]	Knowledge-driven driving (DILU)	Sensor data, traffic information	Driving commands and policies	GPT-3.5, GPT-4
Sural et al, 2024 [48]	Context understanding (ContextVLM)	Visual, textual info	Scene context	ViLT, LLaVA
Tang et al, 2024 [49]	Scenario generation (LeGEND)	Scenario descriptions	Detailed scenarios	GPT-4
Cai et al, 2024 [50]	Regulation compliance (Driving with Regulation)	Traffic rules, action	Legality assessments	GPT-4o
Xu et al, 2024 [51]	Visual-based driving (DriveGPT4)	Images, instructions	Driving actions	GPT-4V
Deruyttere et al, 2022 [52]	Command translation (Talk2Car)	Driving commands	Physical trajectories	Sentence-BERT, RoBERT
Liang et al, 2023 [29]	Code generation (Code as Policies)	Natural language instructions, goals	Executable code	Codex code-davinci-002

surface vessels. CORALL processes real-time navigational data to generate COLREGs-compliant decisions. The system maintains state awareness across time steps, enabling consistent decision-making throughout vessel encounters while adapting to changing risk levels. The framework has been evaluated across challenging scenarios, demonstrating its capability to handle various maritime situations from simple head-on encounters to complex multi-vessel scenarios. The following subsections detail the CORALL architecture, vehicle dynamics and execution layer components including local planning and control algorithms, parameter selection rationale, the CORALL algorithm formalisation and decision-making framework, and implementation details.

A. CORALL System Architecture

The CORALL framework, illustrated in Fig. 5, comprises three distinct layers integrating navigation and perception, LLM-based decision processing, and execution control. This architecture ensures continuous risk assessment while maintaining regulatory compliance in autonomous maritime navigation. The Navigation and Perception Layer processes simulated navigational data that would typically be obtained from maritime sensors such as AIS, GNSS, IMU, RADAR, LIDAR, and Camera systems in real-world applications (not the target of this research). In our simulation environment, this layer processes parameters including Distance at Closest Point of Approach (D_{CPA}), which represents the minimum predicted distance between vessels, Time to Closest Point of Approach (T_{CPA}), indicating the time until vessels reach D_{CPA}, Range (R), measuring the current distance between ships, and relative bearings (ψ_{rel}), forming a foundation for informed decision-making. The core intelligence resides in the LLM-Based Decision Processing Layer, which consists of three primary components:

- Data Preprocessing: Performs feature extraction and data normalisation, structuring the sensory inputs for LLM interpretation
- LLM Decision Core: Serves as the central reasoning unit, performing COLREGs rule identification, risk assessment, and situation explanation
- Decision Validation: Ensures generated decisions maintain both safety constraints and COLREGs compliance

The Execution Layer implements the validated decisions through three sequential components: Local Planner, Control, and Actuator. The Local Planner generates risk-aware trajectories based on the LLM's decisions, which are then translated into control commands and finally into specific actuator signals, as will be detailed in Section III.B. This hierarchical approach ensures implementation of COLAV manoeuvres while maintaining both safety margins and regulatory compliance. For multi-vessel scenarios, CORALL employs a risk-priority queue, processing targets based on their computed risk levels $Risk(t)_i$ and temporal constraints, ensuring highest-risk encounters receive immediate attention while maintaining overall situation awareness.

B. Execution layer and vehicle model

To understand the algorithm implementation process, it is crucial to explain the vehicle model and execution layer in Fig. 5. These can be replaced by any other functional model but were selected for their proven practical effectiveness. The MASS dynamics are modelled using a nonlinear model based on [13], [53]:

$$\begin{cases} \dot{X}(t) = u(t)\cos(\psi(t)) \\ \dot{Y}(t) = u(t)\sin(\psi(t)) \\ \dot{\psi}(t) = r(t) \\ \dot{r}(t) = -(1/T_\psi)r(t) + (K_\psi/T_\psi)\tau_c(t) \\ \dot{u}(t) = -(1/T_u)u(t) + (K_u/T_u)\big(\tau(t) - d(t)\big) \\ \dot{d}(t) = -(1/T_d)d(t) + \omega_d(t) \end{cases} \tag{1}$$

The state vector $[X, Y, \psi]$ represents the vessel's pose, where X and Y denote the longitudinal/lateral positions and ψ the heading angle. The yaw rate r follows Nomoto's first-order dynamics [53] with time constant T_ψ and gain K_ψ. The control input u_c is constrained within ± 20 units to reflect realistic actuator limitations. The vessel's surge speed u is governed by a first-order transfer function with parameters T_u and K_u. Environmental disturbances d are modelled via a Markov process with Gaussian input ω_d and time constant T_d. This realistic model imitates most of the challenges in the dynamic behaviour of the MASS at this stage of design.

C. Local planning algorithm

The motion planning system integrates CORALL decisions with traditional navigation algorithms, based on [13], to execute COLREGs-compliant manoeuvres. As shown in Fig. 2, it is assumed that a global planner has generated the desired waypoints to track. The desired heading command ψ_d to follow the waypoints and avoid collisions combines three components, defined as

$$\psi_d(t) = \psi_{\mathrm{LOS}}(t) + \psi_{\mathrm{CTE}}(t) + \psi_{\mathrm{COLAV}}(t). \tag{2}$$

where ψ_{LOS} is the line-of-sight term, ψ_{CTE} penalizes cross-track error, and ψ_{COLAV} is the collision-avoidance contribution.

1) Line-of-Sight tracking: For path tracking, waypoint i is $WP_i = [X_{WP_i}, Y_{WP_i}]$. As shown in Fig. 2, the position errors relative to the active waypoint are

$$X_{e,WP_i}(t) = X_{WP_i} - X(t), \tag{3}$$

$$Y_{e,WP_i}(t) = Y_{WP_i} - Y(t). \tag{4}$$

The desired line-of-sight (LOS) angle is

$$\psi_{\mathrm{LOS}}(t) = \mathrm{atan2}\big(Y_{e,WP_i}(t),\, X_{e,WP_i}(t)\big). \tag{5}$$

2) Cross-track error elimination: To eliminate the cross-track error (CTE), denoted ϵ_{CTE} (Fig. 2), we define the track length and angle between WP_{i-1} and WP_i as

$$L_{\mathrm{TRK}_i} = \sqrt{\big(X_{WP_i} - X_{WP_{i-1}}\big)^2 + \big(Y_{WP_i} - Y_{WP_{i-1}}\big)^2}, \tag{6}$$

$$\psi_{\mathrm{TRK}_i} = \mathrm{atan2}\big(Y_{WP_i} - Y_{WP_{i-1}},\, X_{WP_i} - X_{WP_{i-1}}\big). \tag{7}$$

The along-track distance is

$$S_i(t) = \frac{\big(X_{WP_i} - X_{WP_{i-1}}\big)\, X_{e,WP_i}(t) + \big(Y_{WP_i} - Y_{WP_{i-1}}\big)\, Y_{e,WP_i}(t)}{L_{\mathrm{TRK}_i}}. \tag{8}$$

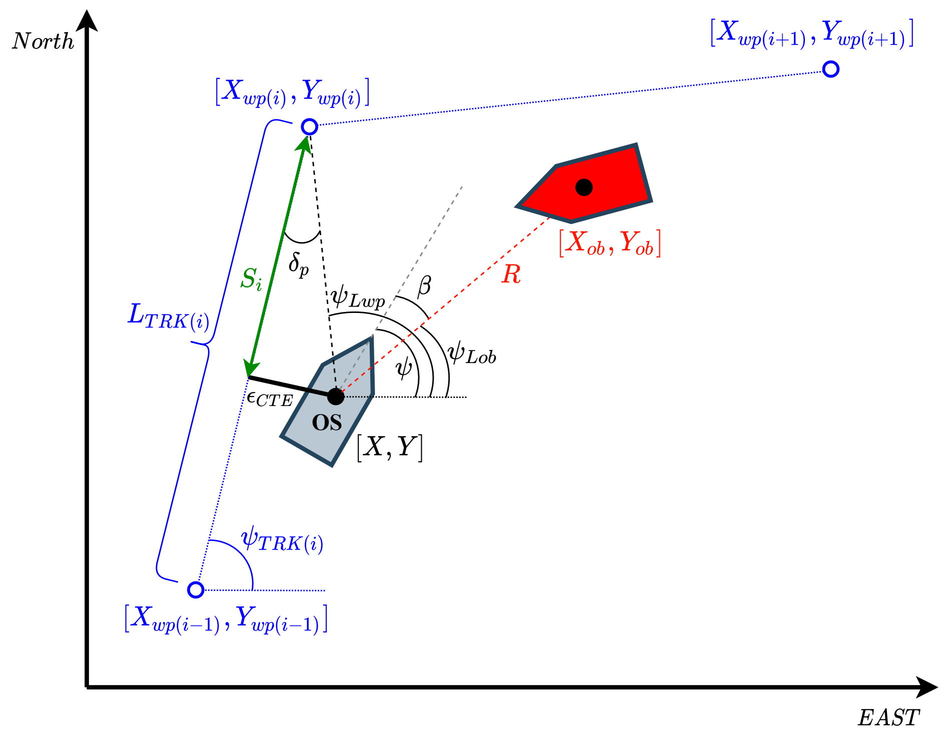

Fig. 2. Motion-planning scene showing waypoints, track line, vessel position, cross-track error, and path angles.

Hence, the cross-track error is

$$\epsilon_{\mathrm{CTE}}(t) = S_i(t)\,\tan\bigl(\delta_p(t)\bigr), \tag{9}$$

where $\delta_p(t) = \psi_{\mathrm{TRK}_i}(t) - \psi_{\mathrm{LOS}}(t)$. The CTE compensation term is

$$\psi_{\mathrm{CTE}}(t) = -\arctan\bigl(\epsilon_{\mathrm{CTE}}(t)/\mu\bigr), \tag{10}$$

where $\mu > 0$ is a tuning parameter [13].

3) CORALL-guided collision avoidance: The COLAV term incorporates CORALL's decisions and fuzzy weighting functions to prevent collisions between the OS and each TS. A generic encounter is illustrated in Fig. 3 and underpins the following equations. As shown in Fig. 4, two weighting mechanisms are used:

i) *Distance-to-obstacle weighting* (Z-shaped membership function, ZMF):

$$w_R(R) = \begin{cases} 1, & R \le a, \\ 1 - 2\left(\dfrac{R-a}{b-a}\right)^2, & a < R \le \dfrac{a+b}{2}, \\ 2\left(\dfrac{R-b}{b-a}\right)^2, & \dfrac{a+b}{2} < R < b, \\ 0, & R \ge b, \end{cases} \tag{11}$$

where $a < b$ are range breakpoints.

ii) *Bearing-to-obstacle weighting* (Gaussian):

$$w_\beta(\beta) = \exp\left(-\frac{\beta^2}{2\sigma^2}\right), \tag{12}$$

where $\sigma > 0$ is a tuning parameter and β is the relative bearing,

$$\beta(t) = \psi(t) - \operatorname{atan2}\big(Y_{ob}(t) - Y(t),\, X_{ob}(t) - X(t)\big). \tag{13}$$

The per-obstacle COLAV command combines these weights with CORALL's direction decision:

$$\psi_{\mathrm{COLAV},i}(t) = K_{\mathrm{Dir},i}\, K_{\mathrm{COLAV}}\, w_R\big(R_i(t)\big)\, w_\beta\big(\beta_i(t)\big), \tag{14}$$

where $K_{\mathrm{Dir},i} \in \{-1, 0, +1\}$ encodes port, stand-on, or starboard as interpreted from CORALL, and $K_{\mathrm{COLAV}} > 0$ is a tuning gain. For n targets, the total command is

$$\psi_{\mathrm{COLAV}}(t) = \sum_{i=1}^{n} \psi_{\mathrm{COLAV},i}(t). \tag{15}$$

This integration translates the LLM's high-level COLAV decisions into executable commands while maintaining COLREGs compliance.

D. Control algorithm

Planned manoeuvres are executed by a PD controller with speed-dependent gains:

$$\tau_c(t) = K_p(u)\left(\psi_d(t) - \psi(t)\right) - K_d(u)\, r(t), \tag{16}$$

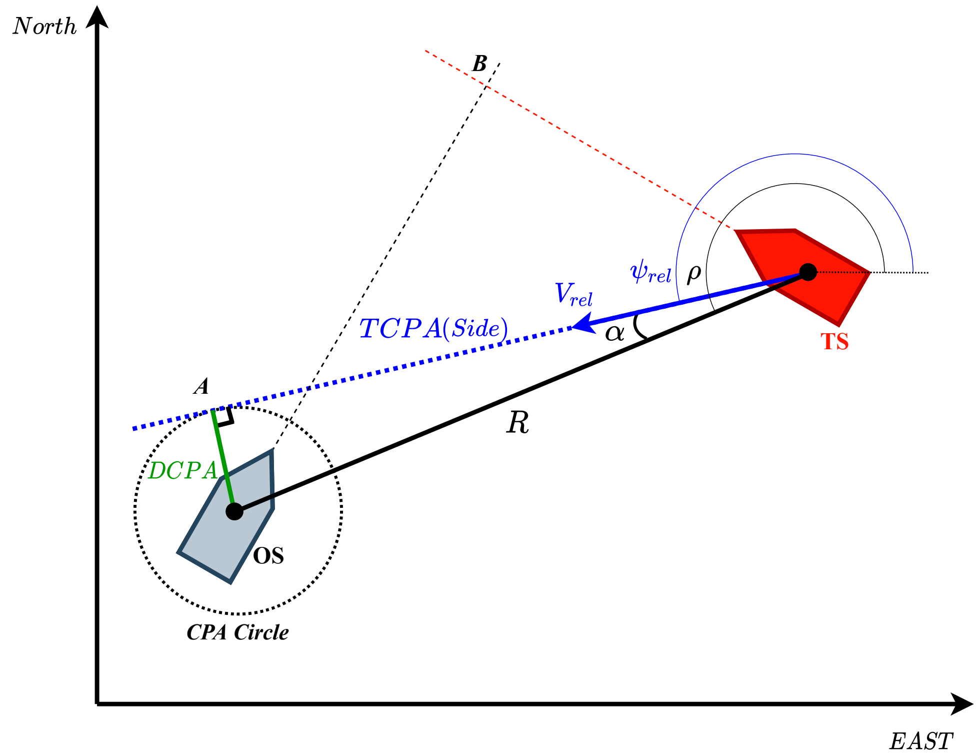

Fig. 3. A COLAV encounter illustrating OS, a TS, and CPA parameters.

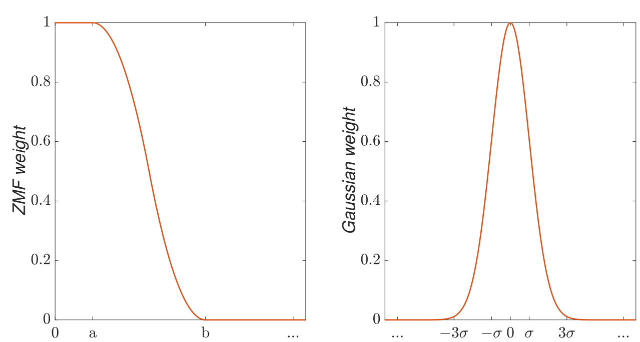

Fig. 4. Risk assessment functions: ZMF for distance-based risk and Gaussian for bearing-dependent risk.

where $K_p(u)$ and $K_d(u)$ are tuned over operating speeds. Advanced planners/controllers may be substituted as needed; see [13]. A Python implementation is available online.[1]

[1]https://github.com/Psarhadi/Autonomous_ship_planning_collision_avoidance_control

E. Parameter Selection

The system parameters, summarised in Table II, were selected based on established maritime practices, physical considerations, and empirical tuning. The environmental disturbance model in (1) captures external forces such as wind, waves, and currents. The Gaussian input ω_d was modelled with zero mean and standard deviation $\sigma_{\omega_d} = 0.1$ m/s^2, representing moderate sea conditions, while the time constant $T_d = 100$s reflects the slowly-varying nature of these forces, consistent with standard maritime simulation practices [53]. The Nomoto first-order model parameters ($T_\psi = 30$s, $K_\psi = 0.01$) represent the manoeuvring characteristics of a high-speed vessel, and the surge dynamics parameters ($T_u = 50$s, $K_u = 1.0$) reflect the vessel's speed response characteristics [53].

The ZMF thresholds for risk assessment were determined based on IMO guidelines and common maritime practice [1]. For D_{CPA}, the values $a_{D_{CPA}} = 0.2$ nmi and $b_{D_{CPA}} = 0.5$ nmi represent the range from minimum safe passing distance to comfortable separation in open waters. The T_{CPA} thresholds ($a_{T_{CPA}} = 1$ min, $b_{T_{CPA}} = 4$ min) correspond to urgent action and early warning periods, while the range thresholds ($a_R = 0.1$ nmi, $b_R = 0.3$ nmi) define critical and caution zones for immediate risk assessment. The Gaussian bearing weight parameter $\sigma = 80°$ focuses the collision avoidance response on vessels within the frontal sector, and the range-based ZMF parameters ($a = 0.1$ nmi, $b = 0.5$ nmi) ensure appropriate weighting based on proximity. The PD controller gains ($K_p = 100.0,\ K_d = 500$) were empirically tuned for responsive yet stable heading control.

Although several parameters in the proposed framework were selected empirically based on maritime operational practice, prior control experience, and iterative simulation validation, moderate variations around the selected values did not significantly alter the qualitative decision-making behaviour of CORALL within the evaluated benchmark scenarios. In particular, the COLREGs classification and manoeuvre selection remained generally consistent under small perturbations of the risk thresholds and controller gains.

TABLE II
SYSTEM PARAMETERS AND SELECTION RATIONALE

Parameter	**Value**	**Unit**	**Rationale**
Disturbance Model			
σ_{ω_d}	0.1	m/s^2	Moderate sea state (3–4)
T_d	100	s	Slowly-varying disturbance
Ship Dynamics			
T_ψ	30	s	Yaw time constant
K_ψ	0.01	–	Yaw gain
T_u	50	s	Surge time constant
K_u	1.0	–	Surge gain
Risk Assessment			
$a_{D_{CPA}}$	0.2	nmi	Minimum safe distance
$b_{D_{CPA}}$	0.5	nmi	Comfortable passing distance
$a_{T_{CPA}}$	1	min	Urgent action threshold
$b_{T_{CPA}}$	4	min	Early warning threshold
a_R	0.1	nmi	Critical range
b_R	0.3	nmi	Caution range
Fuzzy Weighting			
σ (bearing)	80	deg	Frontal sector focus
a (range ZMF)	0.1	nmi	Close-range threshold
b (range ZMF)	0.5	nmi	Far-range threshold
Control and Planning			
K_{COLAV}	2	-	Collision avoidance tuning gain
ρ	1.188	nmi	CTE lookahead distance
K_p	100.0	–	Proportional gain
K_d	500.0	–	Derivative gain

F. CORALL Algorithm Formalisation

As formalised in Algorithm 1, CORALL processes navigation data for multiple vessels through a structured decision-making pipeline. For each target vessel i, the algorithm first constructs a state vector by calculating composite risk using fuzzy membership functions that combine D_{CPA}, T_{CPA}, and range parameters. This risk assessment informs the construction of the LLM input state vector $\mathcal{S}_{LLM}(t)$, which encapsulates all critical navigation parameters. The algorithm then implements target prioritisation based on calculated risk levels and temporal constraints, maintaining situational awareness through state history $\mathcal{H}(t)$ and manoeuvre status $\mathcal{M}(t)$. These components are combined into a state representation $\mathcal{S}(t)$, which is used to construct the LLM prompt. The LLM processes this input to generate structured output $\mathcal{L}(t)$ comprising COLREGs rule identification and situation description $\mathcal{R}(t)$, action specification $\mathcal{A}(t)$, and decision explanation $\mathcal{E}(t)$. From the action component, the algorithm extracts a direction coefficient K_{Dir} that quantifies the required manoeuvre direction: +1 for starboard turn, -1 for port turn, and 0 for

Algorithm 1 CORALL: Risk-Aware LLM Decision Making for COLREGs Compliance

Require: Set of navigational data for each target $N_i = \{\psi_{rel}, R, D_{CPA}, T_{CPA}\}$ where $i \in \{1, ..., n\}$ targets
Ensure: COLREGs-compliant outputs $\mathcal{L}(t) = \{\mathcal{R}(t), \mathcal{A}(t), \mathcal{E}(t)\}$
for each target i **do**
  Calculate composite risk $Risk_i$ using fuzzy membership functions
  $Risk_i \leftarrow (f(|D_{CPA}|, a_d, b_d) + f(|T_{CPA}|, a_t, b_t) + f(R(t), a_r, b_r))/3$
  Construct state vector $\mathcal{S}_{LLM}(t) = [D_{CPA}(t), T_{CPA}(t), R(t), Risk(t), \psi(t)]$
end for
Identify critical target based on $Risk_i$ and temporal constraints
Update state history $\mathcal{H}(t)$ and manoeuvre status $\mathcal{M}(t)$
/* LLM Decision Generation */
Construct prompt with $\mathcal{S}(t) = \{\mathcal{S}_{LLM}(t), \mathcal{H}(t), \mathcal{M}(t)\}$
Query LLM for structured output $\mathcal{L}(t)$
Extract COLREGs rule $\mathcal{R}(t)$, action $\mathcal{A}(t)$, explanation $\mathcal{E}(t)$
Determine K_{Dir} from $\mathcal{A}(t)$ according to:

$$K_{Dir} = \begin{cases} +1, & \text{if starboard turn} \\ -1, & \text{if port turn} \\ 0, & \text{if stand-on} \end{cases}$$

return $\mathcal{L}(t) = \{\mathcal{R}(t), \mathcal{A}(t), \mathcal{E}(t)\}$

stand-on situations.

G. CORALL Decision-Making Framework

The design-time training process for the CORALL decision-making framework as illustrated in Fig. 6 comprises four interconnected components: prompt engineering with memory integration, CORALL decision maker, explainable action generation, and environmental feedback loop. This structure enables real-time, COLREGs-compliant navigation decisions while maintaining transparency and interoperability. The system's foundation lies in its prompt engineering framework, which incorporates both static task descriptions and dynamic memory components. The task description establishes the LLM's role as an autonomous ship's decision-making system, explicitly constraining its outputs to align with COLREGs requirements. This prompt template is augmented with a memory mechanism that maintains the temporal context of ongoing encounters, enabling the system to generate consistent decisions across multiple timesteps.

At the core of the architecture is the CORALL decision maker, implemented using LLMs. This component processes structured navigational data including relative bearing (ψ_{rel}), range (R), T_{CPA} and D_{CPA}. The decision maker analyzes these parameters along with state history and manoeuvre status within the context of COLREGs rules to evaluate the current situation and determine appropriate actions. The system generates explainable actions through a structured output format that encompasses COLREGs rule identification and situation description, specific action recommendations including stand-on and give-way decisions, manoeuvre specification when required such as starboard turns, and a risk assessment with justification for all decisions made.

An important feature of the architecture is its environmental feedback loop, which connects the control system's execution of manoeuvres back to the decision-making process. This feedback mechanism enables the LLM to adapt its decisions based on the effectiveness of previous actions and evolving navigational conditions.

The integration of these components creates a closed-loop decision-making system that maintains COLREGs compliance while providing human-interpretable explanations for all navigation decisions. This transparency is particularly valuable for the verification of autonomous maritime systems, where understanding the reasoning behind navigational choices is crucial for safety and regulatory compliance. The implementation of CORALL, including the complete simulation environment and test scenarios, is available in our GitHub repository[2].

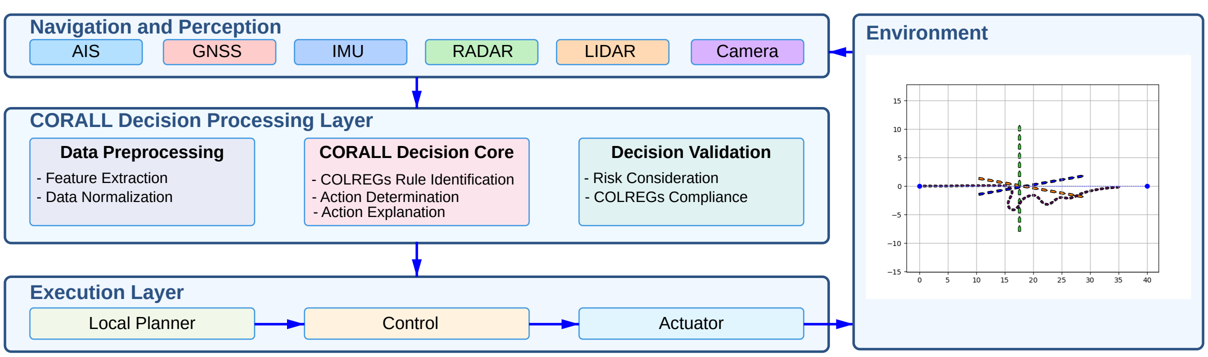

Fig. 5. Architecture of the proposed CORALL

H. Implementation

The implementation of CORALL combines the previously described components into an integrated system, bringing together real-time risk assessment with LLM-based decision-making for autonomous COLREGs-compliant navigation. While other LLMs such as LlaMA 3.1 and

[2]https://github.com/Klins101/CORALL.git

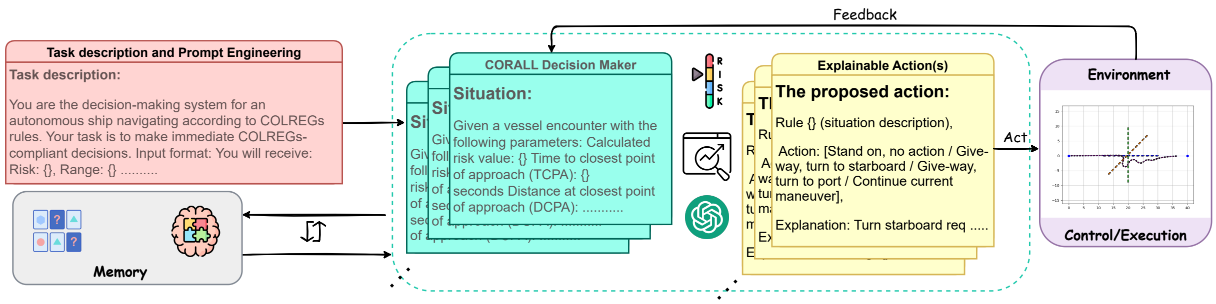

Fig. 6. The design-time architecture of CORALL

PaLM-2 were considered, GPT-4.1 Nano was selected based on initial testing of reasoning capabilities and decision consistency, though a comprehensive comparative analysis remains for future investigation. This implementation unifies the architectural elements into a cohesive framework that extends beyond traditional autonomous navigation systems by incorporating explainable decision-making while maintaining real-time performance. The algorithm underwent supervised training, with errors or improvements reported to CORALL being recorded in the LLM's memory. The run-time CORALL algorithm employs the same prompt engineering, interacting solely with the environment to generate safe actions, as illustrated in Fig. 6. Some of these prompts are revealed in Section IV.

1) Online risk assessment via CPA calculations: In maritime collision-avoidance decision making, the closest point of approach (CPA) methodology is widely used for risk analysis [13]. We adopt and extend this technique to incorporate real-time risk assessment in the LLM decision-making process. As illustrated in Fig. 3, consider a target ship approaching the own ship. The CPA point lies along the relative velocity vector $V_{\text{rel}}(t)$, tangent to a circle centred at the OS. The critical parameters D_{CPA} and T_{CPA} are

$$D_{\text{CPA}}(t) = R(t)\,\sin\bigl(\alpha(t)\bigr), \tag{17}$$

$$T_{\text{CPA}}(t) = \frac{R(t)\,\cos\bigl(\alpha(t)\bigr)}{V_{\text{rel}}(t)}. \tag{18}$$

The relative speed and its heading are

$$V_{\mathrm{rel}}(t) = \sqrt{\big(V_{x,\mathrm{ob}}(t) - V_x(t)\big)^2 + \big(V_{y,\mathrm{ob}}(t) - V_y(t)\big)^2}, \tag{19}$$

$$\psi_{v\mathrm{rel}}(t) = \mathrm{atan2}\big(V_{y,\mathrm{ob}}(t) - V_y(t),\, V_{x,\mathrm{ob}}(t) - V_x(t)\big), \tag{20}$$

where $[V_x(t), V_y(t)]$ and $[V_{x,\mathrm{ob}}(t), V_{y,\mathrm{ob}}(t)]$ are the OS and TS velocity components, respectively.

2) Risk evaluation: We compute a composite risk score using Z-shaped membership functions (ZMFs) to normalise and combine multiple factors:

$$\mathrm{Risk}(t) = \frac{\big|f_D\big(D_{\mathrm{CPA}}(t)\big)\big| + \big|f_T\big(T_{\mathrm{CPA}}(t)\big)\big| + f_R\big(R(t)\big)}{3}. \tag{21}$$

The proposed risk formulation builds upon established CPA/TCPA-based collision risk assessment methods [12], [13], with several design considerations for practical implementation. For numerical stability, a minimum velocity threshold $V_{rel,min} = 0.1$ knots is applied in (18) to prevent division instability when vessels travel at similar speeds. Unlike binary threshold approaches, the ZMF provides continuous risk gradation in $[0, 1]$, enabling proportional response while maintaining computational efficiency suitable for real-time LLM integration. In congested multi-vessel scenarios, risk scores are computed independently for each target, with the risk-priority queue ensuring highest-risk encounters are addressed first.

The component mappings are

$$f_D(x) = f\big(x;\, a_{D_{\mathrm{CPA}}}, b_{D_{\mathrm{CPA}}}\big),$$

$$f_T(x) = f\big(x;\, a_{T_{\mathrm{CPA}}}, b_{T_{\mathrm{CPA}}}\big),$$

$$f_R(x) = f\big(x;\, a_R, b_R\big),$$

where $f(\cdot;\, a, b)$ denotes the ZMF with thresholds a and b. For D_{CPA}, use $a_{D_{\mathrm{CPA}}} = 0.2\,\mathrm{nmi}$ and $b_{D_{\mathrm{CPA}}} = 0.5\,\mathrm{nmi}$; for T_{CPA}, $a_{T_{\mathrm{CPA}}} = 1\,\mathrm{min}$ and $b_{T_{\mathrm{CPA}}} = 4\,\mathrm{min}$; and for range, $a_R = 0.1\,\mathrm{nmi}$ and $b_R = 0.3\,\mathrm{nmi}$. This design normalises heterogeneous quantities such as distance and time to the range $[0, 1]$, yields smooth transitions between risk levels via ZMFs, exposes configurable thresholds for different operating contexts, and provides a unified risk signal for LLM decision making. This risk assessment directly informs the LLM, enabling quantitative evaluation while keeping the COLAV logic interpretable.

3) Data Preprocessing: The data preprocessing module transforms raw navigational data into a structured format suitable for LLM consumption. This process involves three stages. First, coordinate transformation converts the absolute positions and velocities of detected vessels from the global reference frame to the own ship's body-fixed reference frame, enabling relative motion analysis. Second, feature extraction computes the derived navigational parameters including relative bearing ψ_{rel}, range R, and the CPA metrics (D_{CPA}, T_{CPA}) using equations (17) and (20). Third, data normalisation scales the extracted features to consistent ranges: angular measurements are normalised to $[-180°, 180°]$, distances are expressed in nautical miles, and time in minutes. The composite risk score is computed as per equation (21), yielding a normalised value in $[0, 1]$. The preprocessed data is then formatted into a structured text prompt comprising the current navigational state $\mathcal{S}_{LLM}(t)$, state history $\mathcal{H}(t)$, and manoeuvre status $\mathcal{M}(t)$, as detailed in Algorithm 1.

4) Decision Integration:

State Management: The integration between CORALL's decision-making and motion planning is implemented through a state management system, as detailed in Algorithm 1. Real-time navigation parameters and risk metrics are continuously processed and structured for CORALL consumption. The system preserves temporal consistency by monitoring both current navigational states and historical encounter data, enabling the LLM to generate contextually aware decisions. This state management ensures seamless transitions between decisions while preventing oscillatory behaviour in complex scenarios.

CORALL-Motion Planner Interface: The interface facilitates bidirectional communication between CORALL's high-level decision-making and the motion planning system. LLM decisions are translated into executable commands through the direction coefficient K_{Dir}, which quantifies manoeuvre intentions into specific control inputs as shown in equation 16. The system generates risk-aware trajectories that balance COLREGs compliance with safety constraints, continuously monitoring execution performance through a feedback loop. This

TABLE III
INITIAL PARAMETERS FOR IMAZU TEST SCENARIOS

Case	Scenario Type	Target 1	Target 2	Target 3
1	Head-on	(6, 0), 180°	–	–
2	Crossing	(5, -2), 90°	–	–
3	Overtaking	(3, 0), 0°	–	–
4	Crossing	(3.44, 1.6), 295°	–	–
5	Crossing + Head-on	(5, -2.01), 90°	(6.95, 0), 180°	–
6	Multiple Crossing	(3.4, -1.36), 45°	(3, -0.35), 10°	–
7	Overtaking + Crossing	(3, 0), 0°	(3.4, -1.35), 45°	–
8	Crossing + Head-on	(5, -2.01), 90°	(7, 0), 180°	–
9	Multiple Crossing	(3.4, -1.34), 45°	(5, -2.0), 90°	–
10	Multiple Crossing	(3, 0.35), 350°	(4.4, -1.78), 90°	–
11	Multiple Crossing	(5, 2.0), $-90°$	(3.4, -1.35), 45°	–
12	Head-on + Crossing + Crossing	(7, 0), 180°	(3, 0.35), $-10°$	(3.44, -1.35), 45°
13	Head-on + Crossing + Crossing	(6, 0), 180°	(3, 0.35), 350°	(3.4, 1.47), 295°
14	Triple Crossing	(3.4, -1.35), 45°	(3, -0.35), 10°	(5, -2.0), 90°
15	Overtaking + Crossing + Crossing	(3, 0), 0°	(3.4, -1.35), 45°	(5, -2.0), 90°
16	Triple Crossing	(3.4, 1.35), $-45°$	(5, 2.0), $-90°$	(5, -2.0), 90°
17	Overtaking + Crossing + Crossing	(3, 0), 0°	(3, 0.35), $-10°$	(3.4, -1.34), 45°
18	Triple Crossing	(3.3, -0.4), 10°	(3.4, -1.35), 45°	(6.5, -1.47), 135°
19	Triple Crossing	(3, -0.35), 10°	(3, 0.35), $-10°$	(6.5, -1.42), 135°
20	Overtaking + Crossing + Crossing	(3, 0), 0°	(3, -0.35), 10°	(4.4, -1.78), 90°
21	Triple Crossing	(2.7, -0.35), 10°	(2.7, 0.32), $-10°$	(4.4, -1.78), 90°
22	Overtaking + Crossing + Crossing	(3, 0), 0°	(3.94, -1.65), 45°	(5, -2.0), 90°

monitoring enables dynamic adaptation of decisions based on the effectiveness of ongoing manoeuvres and evolving navigational conditions. The integrated framework maintains decision consistency across multiple time steps while adapting to changing risk levels. This tight coupling between decision-making and execution layers enables the system to handle complex multi-vessel encounters while maintaining clear reasoning chains for all navigation decisions.

5) LLM Configuration: The CORALL decision-making core employs OpenAI's GPT-4.1 Nano model configured with a temperature of 0.1 to ensure deterministic and consistent decisions, and a maximum output length of 50 tokens sufficient for the structured decision format. The prompt architecture follows a two-part structure comprising a system prompt and a situation input. The system prompt establishes the COLREGs decision-making framework by encoding COLREGs rules, bearing and risk assessment rules, and action protocols for head-on, crossing, and overtaking situations. The situation input provides the current navigational state in a structured format containing the number of vessels, the vessel's parameters (risk level, distance, bearing, D_{CPA}, and T_{CPA}), and recent state history to maintain temporal context. The model generates responses in a standardised format: `Rule: [number], Action: [turn to starboard/port/stand on], Explanation: [text]`. This structured input-output design ensures consistent decision parsing and enables direct extraction of the direction coefficient K_{Dir} for integration with the motion planning system described in Section III.

IV. Simulation Results

The simulation environment includes the system architecture and algorithms detailed in Section III, with the execution layer operating at 100Hz frequency for motion control and trajectory updates. The LLM-based decision-making runs at 1Hz frequency, providing sufficient responsiveness while ensuring computational feasibility for real-time implementation. The system integrates COLREGs-compliant decision-making with traditional motion planning through the architecture shown in Fig. 5.

The performance evaluation is conducted using the standardised Imazu test scenarios [54], which are now widely accepted as benchmarks in maritime collision avoidance research [5], [55], [56]. Our system was tested with the OS starting at $(X(t_0), Y(t_0)) = (0, 0)$ nautical miles, heading $\psi(t_0) = 0$ degrees with velocity of 25 knots, and waypoint at $(X_{wp}(0), Y_{wp}(0)) = [90, 0]$ nautical miles. Target vessels move at 10 knots, with initial coordinates adapted from [16], with modifications to set initial D_{CPA} values near zero, thereby increasing collision risk (see Table III).

The trajectories across all 22 test scenarios are shown in Fig. 7, where the own ship trajectory appears in black, with target ships (TS) in blue (TS1), orange (TS2), and green (TS3) based on vessel count. Fig. 8 presents the time histories of key safety parameters (D_{CPA}, T_{CPA}, Range, and Risk), demonstrating successful collision avoidance across scenarios ranging from basic head-on encounters to complex multi-vessel interactions. The collision risk parameters were monitored within specific ranges: D_{CPA} within $[0.2, 0.5]$ nautical miles, T_{CPA} between $[2, 4]$ minutes, and obstacle distances varying from $[0.1, 0.3]$ nautical miles. Notably, the small variations in the D_{CPA} demonstrate the algorithm's robustness in handling intentionally introduced disturbances.

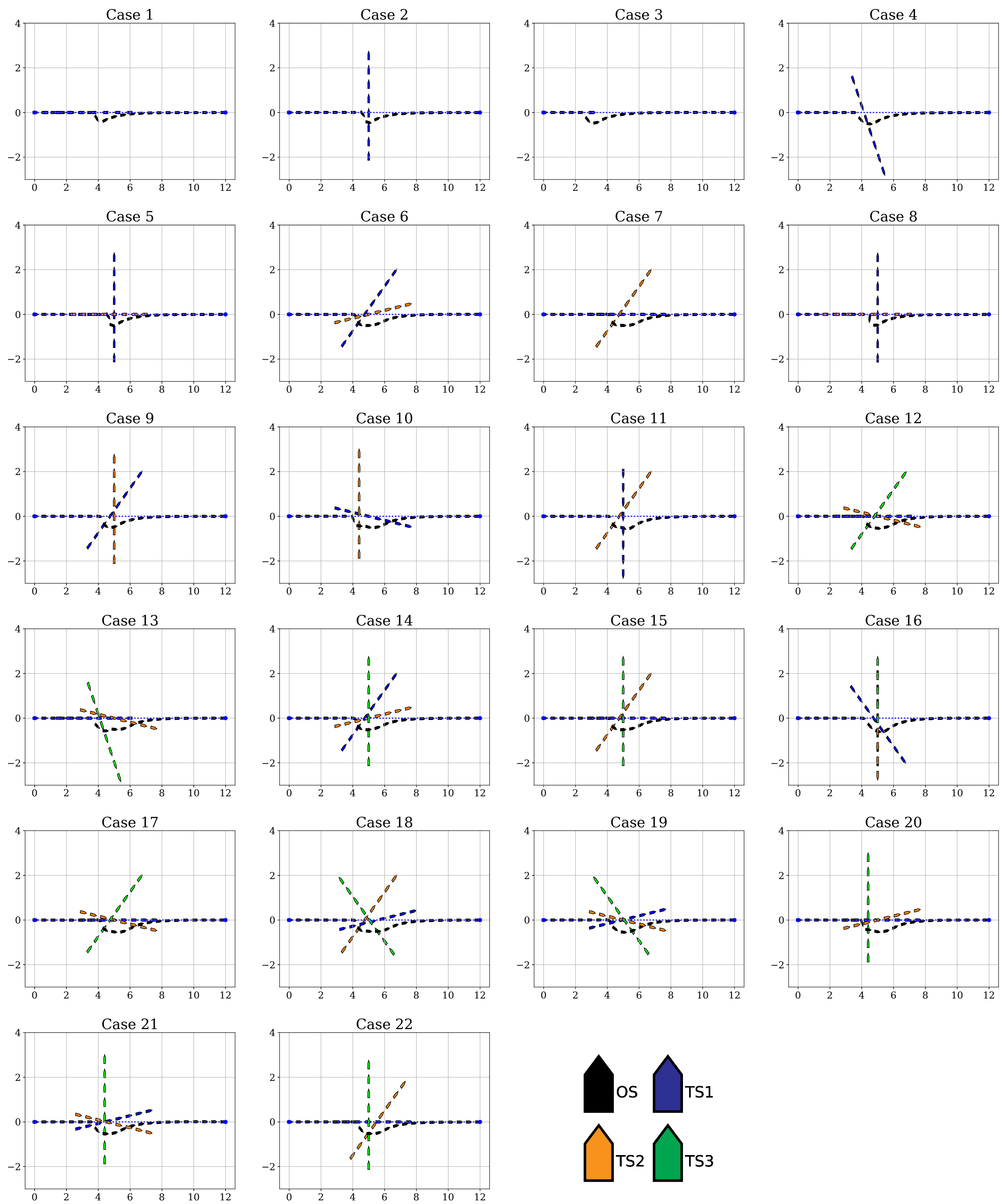

Fig. 7. Trajectories of the own ship and dynamic obstacles for all 22 Imazu cases

For a more detailed analysis, we also present three representative scenes involving the generated explanations demonstrating the performance of the CORALL.

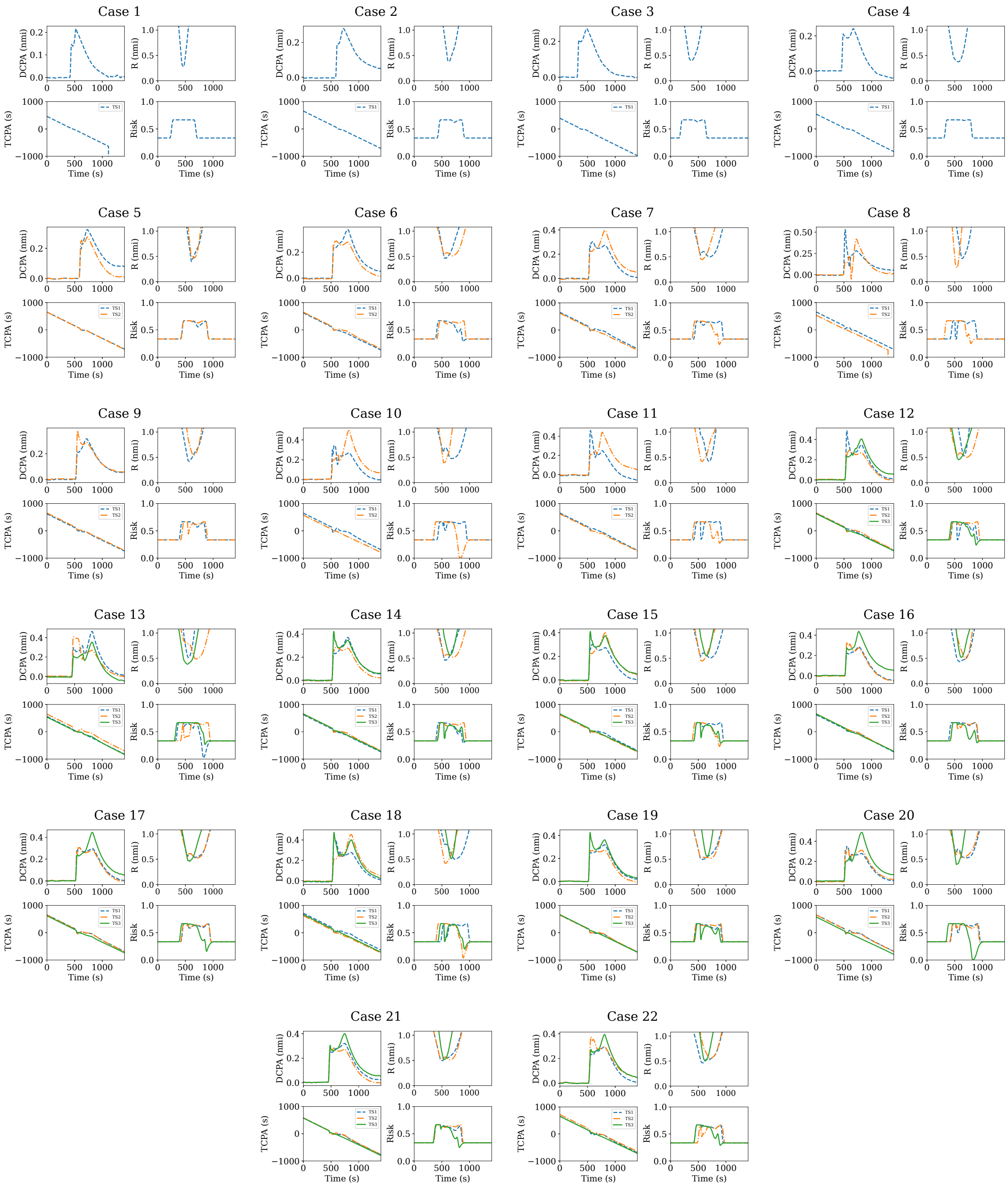

Fig. 8. Time histories of collision risk parameters in all 22 Imazu cases

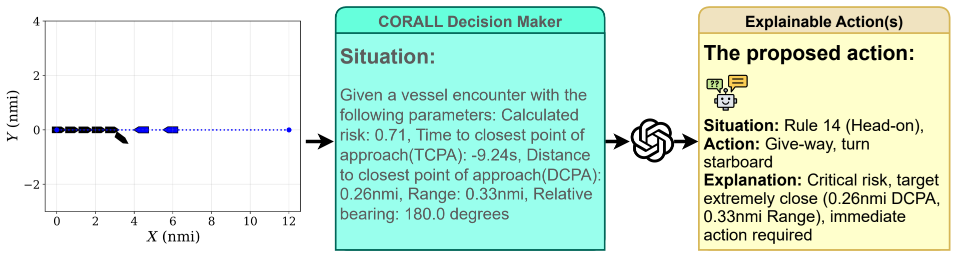

Fig. 9. LLM decision-making process for head-on encounter (Case 1)

A. *Case 1: Head-on Encounter*

In this fundamental test of COLREGs Rule 14 application, CORALL demonstrated successful decision-making throughout the encounter. The parameter evolution plots in Fig. 8 reveal key aspects of the system's performance. The D_{CPA} plot shows relatively low values initially, followed by a significant increase at around $t = 500$s as a result of the collision avoidance manoeuvre. The Risk profile shows moderate levels before the manoeuvre, with a spike during closest approach, followed by a reduction to safer levels after successful avoidance action. As shown in the decision visualisation in Fig.9 (input-output prompts), at a critical point in the encounter, CORALL correctly identified a head-on situation based on the relative bearing of 180.0^0 with high risk. The system generated a "give-way, turn starboard" decision, adhering to COLREGs Rule 14 requirements. The trajectory plot in Fig.9 (left) demonstrates the successful execution of this decision, as evidenced by the black vessel's (representing own ship) clear starboard turn away from its original course. This manoeuvre effectively prevented a collision with the target vessel, whilst maintaining COLREGs compliance, as confirmed by the subsequent improvement in risk metrics shown in Fig. 8 (case 1).

B. *Case 7: Overtaking with Crossing*

In Case 7, CORALL processes a two-vessel encounter scenario as shown in Fig.7(case 7). At the captured instance shown in Fig.10 (left), the system evaluates TS1 (blue) ahead and TS2 (orange) crossing. Although both vessels present moderate risk levels, CORALL identifies the critical nature of TS2's close proximity, generating a "give-way, turn starboard" decision in compliance with Rule 15. The trajectory plot in Fig.10 demonstrates successful execution

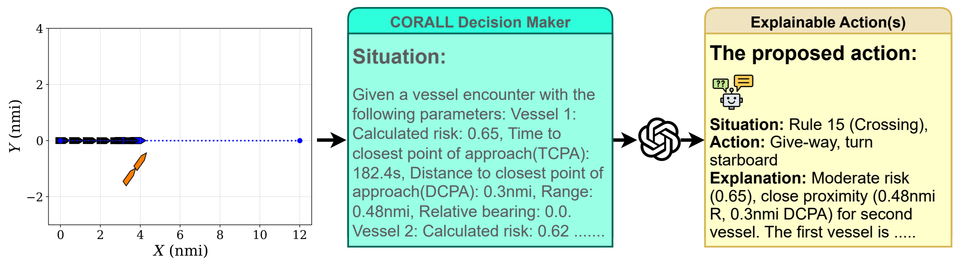

Fig. 10. LLM decision-making process for overtaking with crossing (Case 7)

of this decision, with the black vessel (OS) clearly executing the instructed starboard turn to avoid collision. The parameter evolution in Fig. 8 (case 7) shows this decision-making process. The D_{CPA} plot shows distinct patterns for both vessels, with notable improvements following the avoidance manoeuvre between $t = 500$s and $t = 1000$s. The Risk profiles demonstrate CORALL's ability to maintain separate risk assessments for each vessel while executing appropriate avoidance actions. The TCPA trends and Range plots confirm the successful management of both encounters, with safe distances maintained throughout the scenario.

C. Case 22: Complex Multi-vessel Encounter

The most complex scenario (Case 22) demonstrates CORALL's capability in handling three-vessel encounters, as shown in Fig.11 (left). At the captured instance, with three vessels in proximity (blue TS1, orange TS2, and green TS3), CORALL identifies a critical situation (input-output prompts of Fig.11) where TS3 presents moderate risk at close range (0.50 nmi Range, 0.39 nmi DCPA). The trajectory plot in Fig.11 (left) shows the black vessel (OS) executing the recommended starboard turn, successfully avoiding collision while maintaining COLREGs compliance. The parameter evolution in Fig. 8 (case 22) shows this multi-vessel management strategy. The D_{CPA} plots show coordinated patterns for all three vessels, with successful avoidance manoeuvres reflected in the improvements between $t = 500$s and $t = 1000$s. The Risk profiles, maintained below 0.7 throughout the encounter, demonstrate CORALL's ability to handle multiple simultaneous threats while executing appropriate avoidance actions. The TCPA trends confirm effective time management across all three vessel interactions.

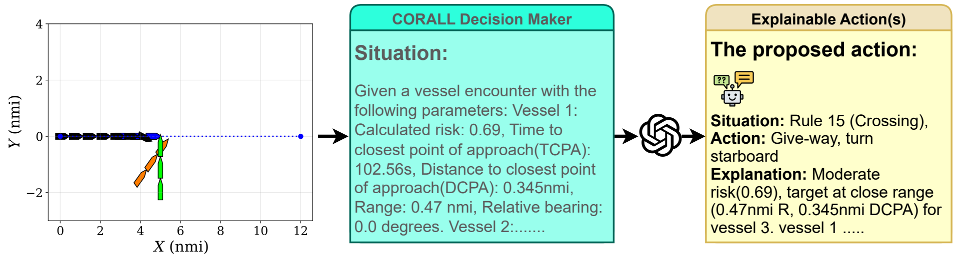

Fig. 11. LLM decision-making process for overtaking with multi-vessel crossing (Case 22)

While the simulation results demonstrate successful collision avoidance across all 22 Imazu scenarios, it should be noted that COLREGs compliance is evaluated through observed behaviour rather than formal verification methods. The system's adherence to COLREGs rules is inferred from the correct identification of encounter types, appropriate give-way/stand-on decisions, and execution of manoeuvres consistent with the regulations. Formal rule-level verification using methods such as model checking or theorem proving remains an important direction for future work to provide rigorous compliance guarantees.

Nevertheless, these results demonstrate the effectiveness of the CORALL approach across the full spectrum of maritime encounter scenarios, from simple single-vessel situations to complex multi-target encounters. The system's ability to exhibit COLREGs-compliant behaviour while providing clear, explainable decisions suggests its potential for practical maritime applications, pending further evaluation and formal verification.

D. *Processor-in-the-Loop Validation*

Processor-in-the-Loop (PIL) simulation is a basic version of Hardware-in-the-Loop (HIL) simulation, in which the implementation processor of the system is integrated with a real-time dynamic simulator, where the physical vehicle is replaced by its mathematical model [57], [58]. This test configuration is commonly used to assess the real-time implementability of autonomous systems at an early design stage [57]. In this work, we evaluate CORALL using a PIL setup developed to test the proposed LLM-based algorithm under real-time operating conditions. For this purpose, a Jetson Orin Nano (8GB) computer is selected, as it is a low-cost and widely available embedded platform suitable for real-time execution. This choice demonstrates that the proposed approach is implementable on accessible, non-specialised hardware rather than requiring high-end computing systems.

The choice of 1 Hz for LLM decision-making is grounded in the hierarchical control architecture commonly employed in autonomous vehicles. From the inner loop outward, actuation typically operates at 100 Hz, control at 10–100 Hz, guidance at 1–10 Hz, and high-level decision-making at 0.1–1 Hz (or even slower in sluggish systems). This hierarchy mirrors human decision-making, where reactive responses occur rapidly while strategic decisions unfold over longer timescales. Depending on the vehicle speed, the required update rates may change, however, in ships, due to the slow dynamics, 1 Hz and even lower frequencies are appropriate. Furthermore, COLREGs Rule 8 requires that collision avoidance manoeuvres be readily apparent to other vessels, necessitating stable, sustained actions rather than rapid course changes that could create confusion and increase collision risk. [59] demonstrated LLM-based decision-making for autonomous driving at 0.56–0.83 Hz, where vehicle dynamics are significantly faster than maritime vessels. Our PIL results show a mean LLM execution time of 0.6403s, achieving comparable or an even smaller latency to the automotive application while operating in a domain where such speed is more than sufficient.

The PIL architecture, illustrated in Fig. 12, separates the vessel simulation (plant) from the decision-making and control algorithms (processor), communicating via TCP/IP to emulate realistic deployment constraints. The Real-Time Dynamic Simulator (RTDS) runs on a laptop, simulating the vessel dynamics according to equation (1). It receives control commands (τ_c, u_c) from the processor, applies them through the actuator model, integrates the vessel dynamics at 100 Hz, and transmits the updated state vector $\mathbf{X} = [x, y, \psi, u]$ for both own ship and target ships back to the processor. The Jetson Orin Nano executes the autonomy algorithms upon receiving state information to generate the desired heading ψ_d, reactive collision avoidance, and control law computation. The LLM decision-making operates at 1 Hz while the control loop executes at 100 Hz.

The PIL simulation was conducted on Imazu Case 22, a multi-vessel encounter involving three target ships approaching from different directions. The vessel trajectories in Fig. 13 confirm successful collision avoidance, with the own ship executing a starboard manoeuvre to pass safely

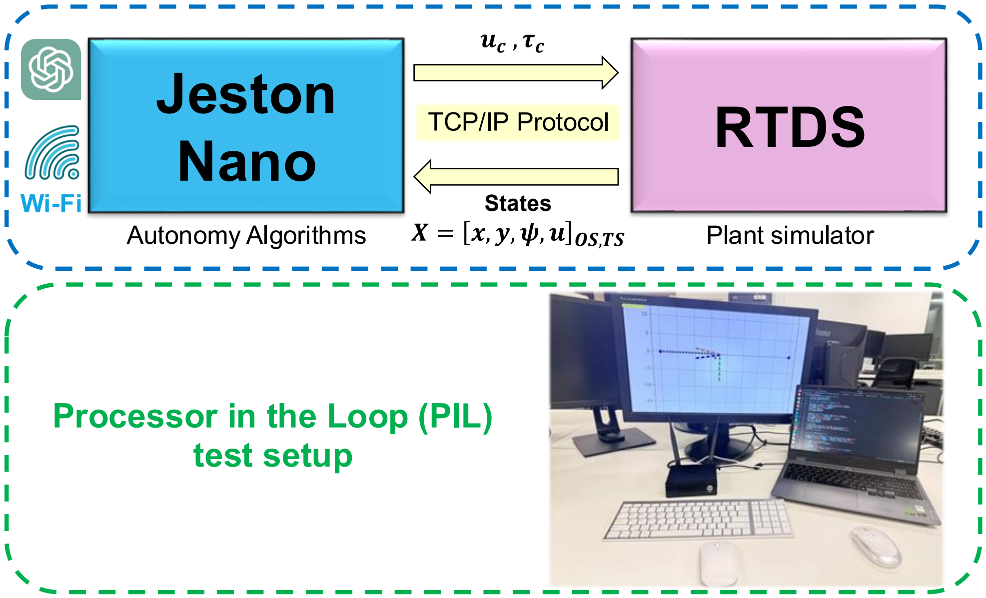

Fig. 12. Processor-in-the-Loop architecture for testing the algorithm's real-time implementability.

clear of all target vessels. Fig. 14 shows the LLM decision output K_{dir} during the encounter. The decision sequence demonstrates COLREGs-compliant behaviour, with a starboard turn initiated during the encounter phase and returning to stand-on once safe passage is established. No port turn ($K_{dir} = -1$) was commanded, as the encounter geometry required only starboard manoeuvres in accordance with COLREGs Rules 14 and 15.

Fig. 15 shows the heading response and control signals. The upper panel demonstrates that the actual heading ψ closely tracks the desired heading ψ_d commanded by CORALL throughout the avoidance manoeuvre. The lower panel displays the actuator torque τ_{ac}, confirming that the control effort remains within feasible limits for the realistic vessel model employed in the PIL environment.

Fig. 17 presents the risk assessment metrics during the PIL simulation. The results are consistent with those obtained in the pure simulation environment (Fig. 8, Case 22), with successful collision avoidance achieved for all target vessels. The minor fluctuations visible in these signals reflect the realistic sensor noise and computational delays present in the PIL environment, demonstrating the robustness of the risk assessment under practical operating conditions such as noise.

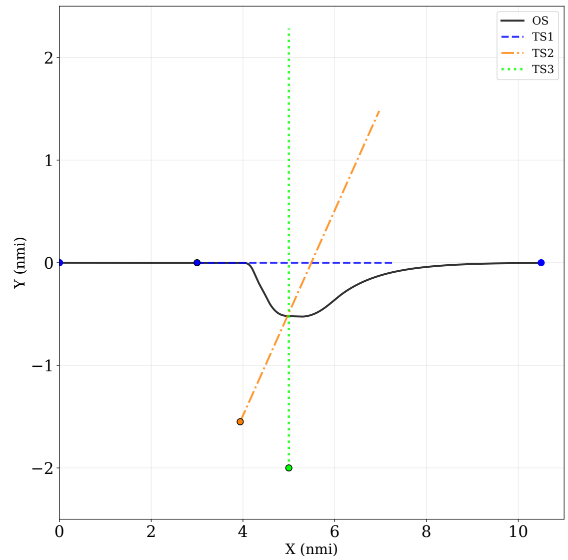

Fig. 13. Vessel trajectories from PIL simulation showing successful collision avoidance with three target ships.

To quantify real-time performance, we recorded the execution time for both the LLM decision cycle and the control loop during PIL simulation. Fig. 16 presents the execution time over approximately 140,000 iterations. The upper panel shows the LLM execution time with a mean of 0.6403 s. The occasional spikes exceeding 1 s are attributed to network latency variability and API server load. It should be noted that this study represents an early demonstrator, and a significant proportion of the latency arises from cloud-based interaction with OpenAI's servers. In practical applications, a locally deployed LLM would operate with considerably lower execution time. Furthermore, we believe that specialised LLMs or potentially Small Language Models (SLMs) should be developed for such domain-specific tasks, which would further reduce computational requirements. The lower panel shows the control loop execution time with a mean of 3.0 ms, confirming stable operation well within the 10 ms budget for 100 Hz control. These results demonstrate that CORALL operates comfortably within real-time constraints.

E. Ablation Study

To evaluate the contribution of individual components in the CORALL framework, we conducted ablation experiments by systematically removing or isolating specific elements across three categories: risk calculation metrics, prompt engineering, and vessel prioritisation. For risk

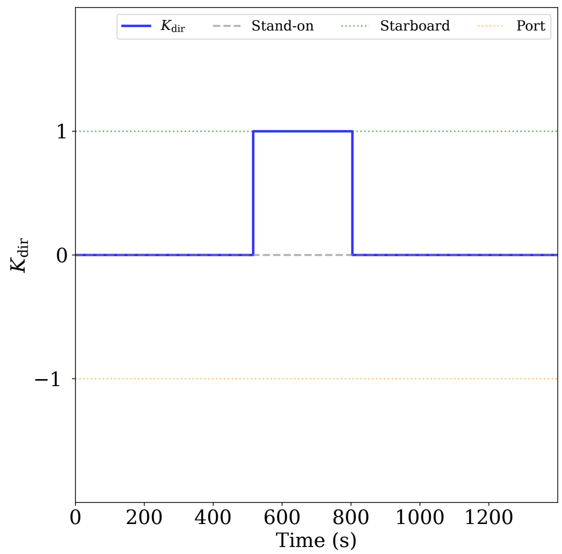

Fig. 14. LLM decision output K_{dir} during PIL simulation.

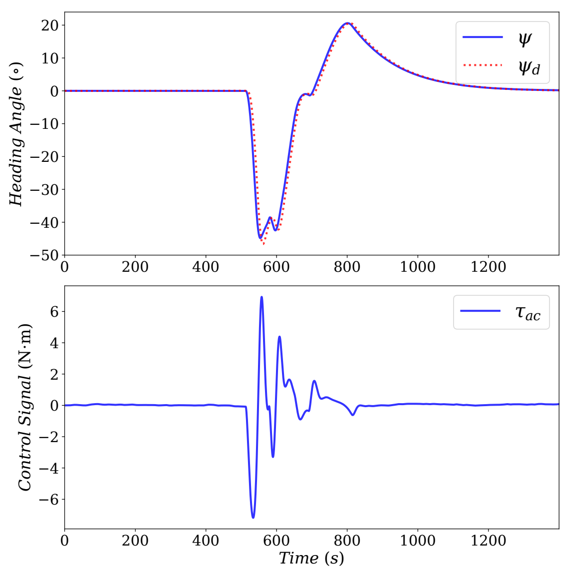

Fig. 15. PIL simulation results: heading response (upper) and actuator torque (lower).

calculation, we tested six variants: the full composite metric combining D_{CPA}, T_{CPA}, and Range; removal of individual components (without D_{CPA}, without T_{CPA}, without Range); and isolated metrics (D_{CPA}-only, T_{CPA}-only). Additionally, we evaluated a minimal prompt configuration with reduced instruction detail and an alternative vessel prioritisation scheme based on highest risk selection. The study tested these eight configurations across two representative scenarios:

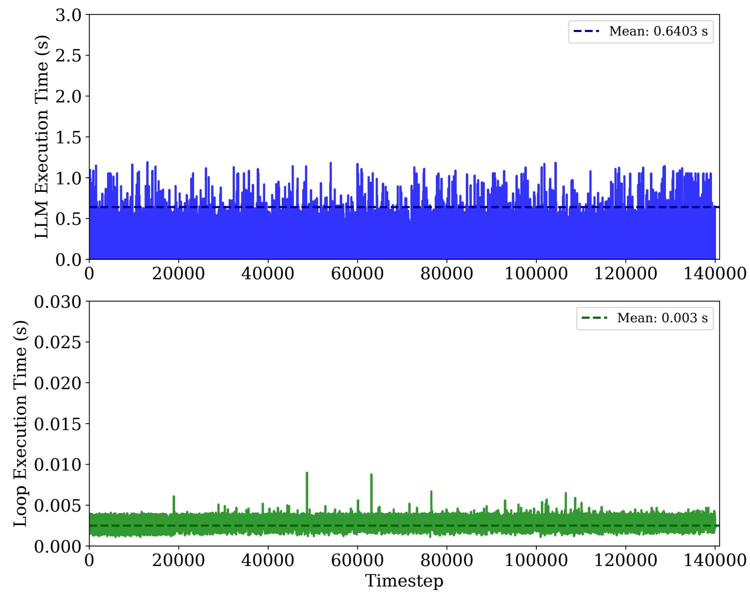

Fig. 16. Execution times during PIL simulation.

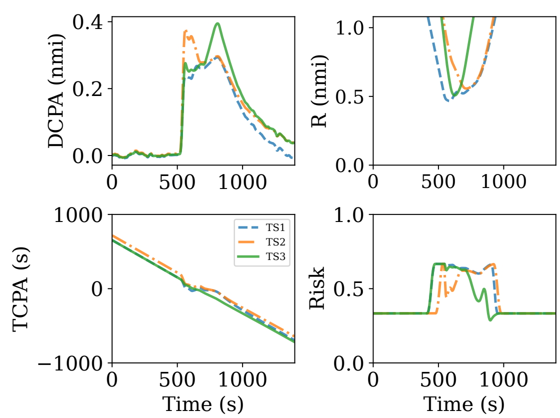

Fig. 17. Risk assessment metrics during PIL simulation.

Case 1 (head-on encounter) representing simple geometry, and Case 15 (overtaking with multiple crossing vessels) representing complex multi-vessel interactions. Accuracy is measured as the percentage of correct LLM decisions.

Fig. 19 presents the mean accuracy grouped by ablation category. Risk calculation achieves the highest mean accuracy (88.8%), followed by prompt engineering (84.2%) and vessel prioritisation (81.7%). All categories exceed the 80% acceptability threshold indicated by the dashed line, demonstrating system robustness across design variations. The error bars reveal that risk calculation exhibits the largest variance, indicating sensitivity to metric selection, while prompt engineering yields more consistent performance with smaller deviation.

Fig. 18 provides a detailed case-wise comparison across all eight configurations. Configurations C1 and C2 represent the prioritisation and minimal prompt variants respectively, while C3–C8 represent the six risk calculation variants (D_{CPA}-only, full composite, without D_{CPA}, without T_{CPA}, without Range, and T_{CPA}-only). Case 1 (blue bars) consistently achieves higher accuracy than Case 15 (orange bars) across all configurations, with the risk-based configurations reaching approximately 97% accuracy on Case 1. Notably, removing individual risk components or using isolated metrics produces nearly identical performance to the full risk formulation, suggesting the LLM can compensate for missing information using geometric relationships. The performance gap between cases (approximately 12–17%) indicates that scenario complexity is a more significant factor than configuration choice.

The relatively small performance differences observed between several ablation configurations suggest that the LLM is capable of implicitly leveraging geometric and contextual relationships present in the navigation inputs, even when certain explicit risk metrics are removed. This indicates that the reasoning capability of the LLM contributes meaningfully to the overall decision-making process beyond deterministic risk-threshold evaluation alone. However, the complete composite risk formulation provides improved consistency and robustness in more complex multi-vessel encounter scenarios, where multiple interacting risk factors must be considered simultaneously.

The accuracy heatmap in Fig. 20 visualises performance across all configuration-case combinations. Two patterns emerge: horizontal banding confirms case complexity as the dominant factor, with Case 1 achieving 88.3–96.7% and Case 15 achieving 75.0–85.0%, and vertical uniformity across risk configurations (C3–C8) demonstrates metric interchangeability, as all achieve 96.7% on Case 1. The full composite configuration (C4) achieves the best Case 15 performance at 85.0%, suggesting that while individual metrics are interchangeable for simple scenarios, the complete formulation provides marginal benefits in complex encounters.

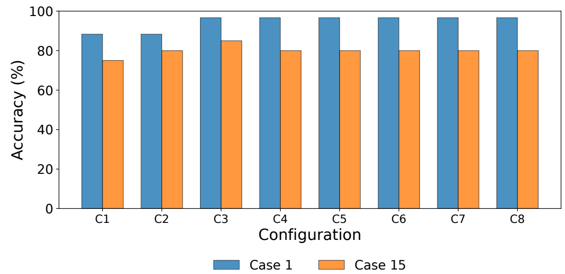

Fig. 18. Decision accuracy by configuration for Case 1 (head-on) and Case 15 (multi-vessel).

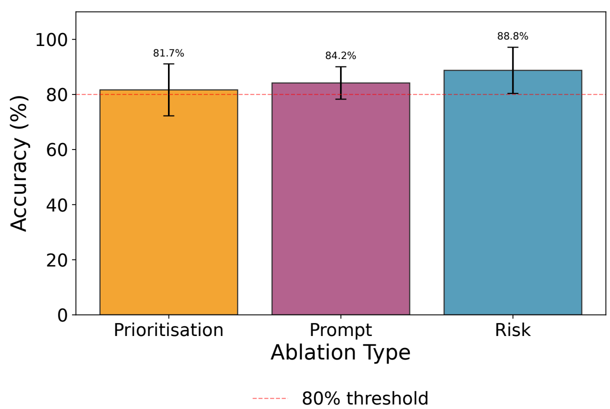

Fig. 19. Mean decision accuracy by ablation category. Error bars indicate standard deviation across configurations within each category. Dashed line represents the acceptability threshold.

V. Conclusion

This paper has presented CORALL, a novel framework that integrates large language models with risk-aware decision-making for autonomous maritime navigation. Our approach successfully addresses three fundamental challenges in autonomous vessel operation: the interpretation of qualitative COLREGs rules, the maintenance of consistent risk awareness, and the generation of explainable navigation decisions. Through comprehensive evaluation across the 22 standardised Imazu test scenarios, CORALL has demonstrated robust performance in real-time COLREGs interpretation, effective risk-aware decision-making in complex multi-vessel encounters, generation of clear explanations for all navigation decisions, and consistent behaviour across varying maritime situations. The integration of CORALL reasoning with traditional risk assessment metrics enables direct translation of qualitative maritime regulations into quantitative actions, whilst maintaining dynamic adaptation to varying risk levels. Real-time feasibility was verified

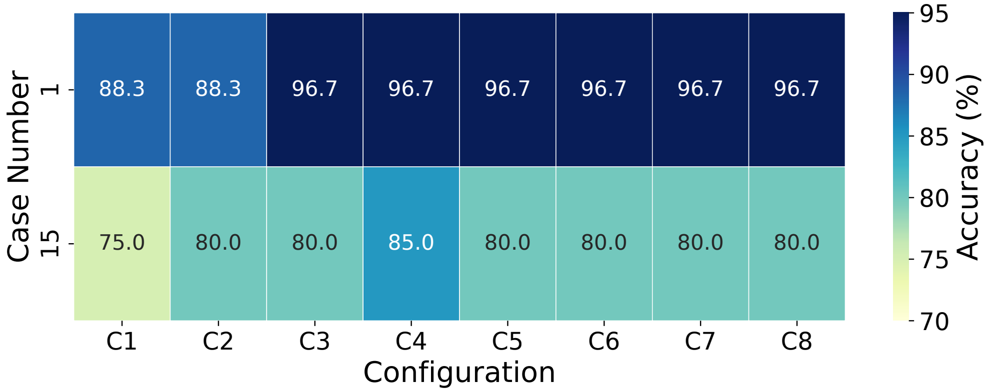

Fig. 20. Accuracy heatmap across configurations (C1–C8) and test cases. Horizontal banding indicates case complexity dominates performance and vertical uniformity demonstrates metric interchangeability.

through Processor-in-the-Loop (PIL) simulation using an embedded Jetson Orin Nano platform, achieving mean LLM execution times of 0.6403 s. Furthermore, the system provides human-interpretable explanations for all decisions, ensuring transparency in autonomous navigation whilst demonstrating scalable handling of multiple vessel encounters.

These contributions represent a meaningful step forward in bridging the gap between regulatory compliance and practical autonomous navigation systems. The CORALL framework not only demonstrates technical efficacy but also addresses the need for explainable AI in safety-critical maritime applications, establishing a foundation for future developments in autonomous vessel technology.

A. Limitations

While CORALL demonstrates promising results, several limitations should be acknowledged. First, the current implementation has been evaluated in simulation environments with idealised sensor data, and deterministic vessel behaviour. Real-world maritime operations involve additional challenges including sensor measurement uncertainty, communication latency, and unpredictable behaviour from human-operated vessels that may not strictly adhere to COLREGs.

Second, although the simulation results indicate consistent COLREGs-compliant behaviour, the evaluation relies on empirical testing rather than formal verification. The absence of formal proofs of rule compliance means that edge cases or subtle rule violations may not be detected through simulation alone.

Third, the seamanship quality of the generated manoeuvres warrants consideration. While the trajectories achieve collision-free navigation, the course alterations may not fully reflect how experienced mariners typically interpret and apply COLREGs in practice. Rule 8 emphasises that actions should be "positive, made in ample time and with due regard to the observance of good seamanship," encompassing not only collision avoidance but also predictability and acceptability to other navigators. The current implementation prioritises safety margins and rule compliance, but the resulting manoeuvres have not been validated against expert mariner judgement.

B. Future Work

Future research should address the identified limitations through several directions. First, whilst Imazu's test cases are widely accepted within the maritime COLAV community, a more rigorous evaluation approach employing Monte Carlo simulations and diverse scenario generation should be pursued to ensure robust performance testing. Second, formal verification methods such as model checking or theorem proving should be explored to provide rigorous compliance guarantees beyond empirical validation. Third, verification under realistic operational conditions incorporating sensor noise, environmental disturbances, and mixed-traffic scenarios is essential for practical deployment.

Beyond validation, several technical extensions merit investigation. Demonstrating real-time implementability with quantified latency bounds is a promising avenue, which our early studies indicate is feasible. The application of Small Language Models (SLMs) or domain-specific fine-tuned models could reduce computational burden whilst improving COLREGs-specific reasoning. Finally, incorporating feedback from experienced navigators would ensure that autonomous manoeuvres align with established bridge practices and are readily interpretable by human operators. The insights provided by this research establish a foundation for continued development of explainable and intelligent decision-making in MASS applications.

REFERENCES

[1] International Maritime Organization (IMO), *Convention on the International Regulations for Preventing Collisions at Sea, 1972 (COLREGs)*. London: IMO, 1972. [Online]. Available: https://doi.org/10.62454/KC904E_02

[2] A. McMillan, "Shipping Losses Drop to All-Time Low – Allianz Commercial," https://supplychaindigital.com/supply-chain-risk-management/allianz-commercial-safety-shipping-review-2024, 2024, accessed: 28th September 2024.

[3] C. Clark. (2024, Aug.) After Ship Collisions, Experts Say China Testing Philippine Resolve, US Ties. Accessed: 15-Sep-2024. [Online]. Available: https://breakingdefense.com/2024/08/after-ship-collisions-experts-say-china-testing-philippine-resolve-us-ties/

[4] S. Gujar and D. Vishwakarma, "Autonomous Ships Market Size - By Ship Type, By Offering, By Propulsion, By Autonomy, By End-User & Forecast, 2024-2032," https://www.gminsights.com/industry-analysis/autonomous-ships-market, Aug. 2024.

[5] P. Sarhadi, W. Naeem, and N. Athanasopoulos, "A Survey of Recent Machine Learning Solutions for Ship Collision Avoidance and Mission Planning," *IFAC-PapersOnLine*, vol. 55, no. 31, pp. 257–268, 2022, 14th IFAC Conference on Control Applications in Marine Systems, Robotics, and Vehicles CAMS 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2405896322024867

[6] L. Sanchez-Heres, R. Weber, F. Ahlgren, F. Olsson, and O. Lundström, "COLREG3: Exploring the Potential of Large Language Models in Marine Navigation Systems," *Lighthouse reports*, 2024. [Online]. Available: https://lighthouse.nu/images/Rapporter/COLREG_3.pdf

[7] Y. Wang and Y. Zhao, "Multiple Ships Cooperative Navigation and Collision Avoidance Using Multi-Agent Reinforcement Learning with Communication," *Ocean Engineering*, vol. 320, p. 120244, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0029801824035820

[8] B.-O. H. Eriksen, G. Bitar, M. Breivik, and A. M. Lekkas, "Hybrid Collision Avoidance for ASVs Compliant with COLREGs Rules 8 and 13–17," *Frontiers in Robotics and AI*, vol. 7, p. 11, 2020. [Online]. Available: http://dx.doi.org/10.3389/frobt.2020.00011

[9] K. Agyei, P. Sarhadi, and W. Naeem, "Large Language Model-Based Decision-Making for COLREGs and the Control of Autonomous Surface Vehicles," in *2025 European Control Conference (ECC)*. IEEE, 2025, pp. 1761–1766. [Online]. Available: https://ieeexplore.ieee.org/document/11186868

[10] J.-K. Kim and D.-J. Park, "Understanding of Sailing Rule Based on COLREGs: Comparison of Navigator Survey and Automated Collision-Avoidance Algorithm," *Marine Policy*, vol. 159, p. 105894, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0308597X2300427X

[11] F. Zhang, Y. Liu, L. Du, F. Goerlandt, Z. Sui, and Y. Wen, "A Rule-Based Maritime Traffic Situation Complex Network Approach for Enhancing Situation Awareness of Vessel Traffic Service Operators," *Ocean Engineering*, vol. 284, p. 115203, 2023. [Online]. Available: https://doi.org/10.1016/j.oceaneng.2023.115203

[12] Y. Huang, L. Chen, P. Chen, R. R. Negenborn, and P. Van Gelder, "Ship Collision Avoidance Methods: State-of-the-Art," *Safety science*, vol. 121, pp. 451–473, 2020. [Online]. Available: https://doi.org/10.1016/j.ssci.2019.09.018

[13] P. Sarhadi, W. Naeem, and N. Athanasopoulos, "An Integrated Risk Assessment and Collision Avoidance Methodology

for an Autonomous Catamaran with Fuzzy Weighting Functions," in *2022 UKACC 13th International Conference on Control (CONTROL)*. IEEE, 2022, pp. 228–234. [Online]. Available: https://doi.org/10.1109/Control55989.2022.9781453

[14] Y. A. Ahmed, M. A. Hannan, M. Y. Oraby, and A. Maimun, "COLREGs Compliant Fuzzy-Based Collision Avoidance System for Multiple Ship Encounters," *Journal of Marine Science and Engineering*, vol. 9, no. 8, p. 790, 2021. [Online]. Available: https://doi.org/10.3390/jmse9080790

[15] M. Lysak, G. Silva, C. Almeida, A. Martins, and J. Almeida, "COLREG Compliant Collision Avoidance System for an Unmanned Surface Vehicle," in *OCEANS 2024-Singapore*. IEEE, 2024, pp. 01–06. [Online]. Available: https://doi.org/10.1109/OCEANS51537.2024.10682357

[16] B. Singh, K. A. Dastgerdi, N. Athanasopoulos, W. Naeem, and B. Lecallard, "Provably Safe Finite-Time Guidance for Marine Vehicles," *arXiv preprint arXiv:2402.06291*, 2024. [Online]. Available: https://doi.org/10.48550/arXiv.2402.06291

[17] M. Aydin, E. Akyuz, O. Turan, and O. Arslan, "Validation of Risk Analysis for Ship Collision in Narrow Waters by Using Fuzzy Bayesian Networks Approach," *Ocean Engineering*, vol. 231, p. 108973, 2021. [Online]. Available: https://doi.org/10.1016/j.oceaneng.2021.108973

[18] L. Zhao and M.-I. Roh, "COLREGs-Compliant Multiship Collision Avoidance Based on Deep Reinforcement Learning," *Ocean Engineering*, vol. 191, p. 106436, 2019. [Online]. Available: https://doi.org/10.1016/j.oceaneng.2019.106436

[19] C. Wang, X. Zhang, H. Gao, M. Bashir, H. Li, and Z. Yang, "COLREGs-Constrained Safe Reinforcement Learning for Realising MASS's Risk-Informed Collision Avoidance Decision Making," *Knowledge-Based Systems*, vol. 300, p. 112205, 2024. [Online]. Available: https://doi.org/10.1016/j.knosys.2024.112205

[20] X. Xu, Y. Lu, X. Liu, and W. Zhang, "Intelligent Collision Avoidance Algorithms for USVs via Deep Reinforcement Learning Under COLREGs," *Ocean Engineering*, vol. 217, p. 107704, 2020. [Online]. Available: https://doi.org/10.1016/j.oceaneng.2020.107704

[21] H. Fan, H. Jia, X. He, and J. Lyu, "Navigating Uncertainty: A Dynamic Bayesian Network-Based Risk Assessment Framework for Maritime Trade Routes," *Reliability Engineering & System Safety*, vol. 250, p. 110311, 2024. [Online]. Available: https://doi.org/10.1016/j.ress.2024.110311

[22] W. Wang, L. Huang, K. Liu, X. Wu, and J. Wang, "A COLREGs-Compliant Collision Avoidance Decision Approach Based on Deep Reinforcement Learning," *Journal of Marine Science and Engineering*, vol. 10, no. 7, p. 944, 2022. [Online]. Available: https://doi.org/10.3390/jmse10070944

[23] L. Sun, Y. Huang, H. Wang, S. Wu, Q. Zhang, C. Gao, Y. Huang, W. Lyu, Y. Zhang, X. Li *et al.*, "TrustLLM: Trustworthiness in Large Language Models," vol. 235, pp. 20 166–20 270, 21–27 Jul 2024. [Online]. Available: https://proceedings.mlr.press/v235/huang24x.html

[24] B. Guo, X. Zhang, Z. Wang, M. Jiang, J. Nie, Y. Ding, J. Yue, and Y. Wu, "How Close Is ChatGPT to Human Experts? Comparison Corpus, Evaluation, and Detection," *arXiv preprint arXiv:2301.07597*, 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2301.07597

[25] A. J. Thirunavukarasu, R. Hassan, S. Mahmood, R. Sanghera, K. Barzangi, M. El Mukashfi, and S. Shah, "Trialling a Large Language Model (ChatGPT) in General Practice with the Applied Knowledge Test: Observational Study Demonstrating Opportunities and Limitations in Primary Care," *JMIR Medical Education*, vol. 9, no. 1, p. e46599, 2023. [Online]. Available: https://doi.org/10.2196/46599

[26] R. Arvidsson, R. Gunnarsson, A. Entezarjou, D. Sundemo, and C. Wikberg, "ChatGPT (GPT-4) Versus Doctors on

Complex Cases of the Swedish Family Medicine Specialist Examination: An Observational Comparative Study," *BMJ open*, vol. 14, no. 12, p. e086148, 2024. [Online]. Available: https://doi.org/10.1136/bmjopen-2024-086148

[27] L. Chen, O. Sinavski, J. Hünermann, A. Karnsund, A. J. Willmott, D. Birch, D. Maund, and J. Shotton, "Driving with LLMs: Fusing Object-Level Vector Modality for Explainable Autonomous Driving," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 14 093–14 100. [Online]. Available: doi.org/10.1109/ICRA57147.2024.10611018

[28] J. Roh, C. Paxton, A. Pronobis, A. Farhadi, and D. Fox, "Conditional Driving from Natural Language Instructions," in *Proceedings of the Conference on Robot Learning*. PMLR, 2020, pp. 540–551. [Online]. Available: https://proceedings.mlr.press/v100/roh20a.html

[29] J. Liang, W. Huang, F. Xia, P. Xu, K. Hausman, B. Ichter, P. Florence, and A. Zeng, "Code as Policies: Language Model Programs for Embodied Control," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 9493–9500. [Online]. Available: https://doi.org/10.1109/icra48891.2023.10160591

[30] S. Javaid, H. Fahim, B. He, and N. Saeed, "Large Language Models for UAVs: Current State and Pathways to the Future," *IEEE Open Journal of Vehicular Technology*, vol. 5, pp. 1166–1192, 2024. [Online]. Available: https://doi.org/10.1109/OJVT.2024.3281535

[31] J. Ioannidis, J. Harper, M. S. Quah, and D. Hunter, "Gracenote.ai: Legal Generative AI for Regulatory Compliance," in *Proceedings of the Third International Workshop on Artificial Intelligence and Intelligent Assistance for Legal Professionals in the Digital Workplace (LegalAIIA 2023)*, 2023. [Online]. Available: https://ceur-ws.org/Vol-3423/paper3.pdf

[32] H. Gao, J. Zhu, T. Zhang, G. Xie, Z. Kan, Z. Hao, and K. Liu, "Situational Assessment for Intelligent Vehicles Based on Stochastic Model and Gaussian Distributions in Typical Traffic Scenarios," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, no. 3, pp. 1426–1436, 2020. [Online]. Available: https://doi.org/10.1109/TSMC.2020.3019512

[33] H. Gao, Z. Kan, and K. Li, "Robust Lateral Trajectory Following Control of Unmanned Vehicle Based on Model Predictive Control," *IEEE/ASME Transactions on Mechatronics*, vol. 27, no. 3, pp. 1278–1287, 2021. [Online]. Available: https://doi.org/10.1109/TMECH.2021.3087605

[34] H. Gao, Y. Qin, C. Hu, Y. Liu, and K. Li, "An Interacting Multiple Model for Trajectory Prediction of Intelligent Vehicles in Typical Road Traffic Scenario," *IEEE transactions on neural networks and learning systems*, vol. 34, no. 9, pp. 6468–6479, 2021. [Online]. Available: https://doi.org/10.1109/TNNLS.2021.3136866

[35] Y. Lin, C. Li, M. Ding, M. Tomizuka, W. Zhan, and M. Althoff, "DrPlanner: Diagnosis and Repair of Motion Planners for Automated Vehicles Using Large Language Models," *IEEE Robotics and Automation Letters*, 2024. [Online]. Available: https://doi.org/10.1109/LRA.2024.3441493

[36] M. Wang, A. Pang, Y. Kan, M.-O. Pun, C. S. Chen, and B. Huang, "LLM-Assisted Light: Leveraging Large Language Model Capabilities for Human-Mimetic Traffic Signal Control in Complex Urban Environments," *arXiv preprint arXiv:2403.08337*, 2024. [Online]. Available: https://doi.org/10.48550/arXiv.2403.08337

[37] B. Li, Y. Wang, J. Mao, B. Ivanovic, S. Veer, K. Leung, and M. Pavone, "Driving Everywhere with Large Language Model Policy Adaptation," in *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 14 948–14 957. [Online]. Available: https://doi.org/10.1109/CVPR52733.2024.01416

[38] M. Yildirim, B. Dagda, V. Asodia, and S. Fallah, "HighwayLLM: Decision-Making and Navigation in Highway

Driving with RL-Informed Language Model," *Robotics and Autonomous Systems*, p. 105114, 2025. [Online]. Available: https://doi.org/10.1016/j.robot.2025.105114

[39] J. Zhong, M. Li, Y. Chen, Z. Wei, F. Yang, and H. Shen, "A Safer Vision-Based Autonomous Planning System for Quadrotor UAVs with Dynamic Obstacle Trajectory Prediction and Its Application with LLMs," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2024, pp. 920–929. [Online]. Available: https://doi.org/10.1109/WACVW60836.2024.00131

[40] Y. Ma, C. Cui, X. Cao, W. Ye, P. Liu, J. Lu, A. Abdelraouf, R. Gupta, K. Han, A. Bera *et al.*, "LaMPilot: An Open Benchmark Dataset for Autonomous Driving with Language Model Programs," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 15 141–15 151. [Online]. Available: https://doi.org/10.1109/CVPR52733.2024.01434

[41] D. Fu, W. Lei, L. Wen, P. Cai, S. Mao, M. Dou, B. Shi, and Y. Qiao, "LimSim++: A Closed-Loop Platform for Deploying Multimodal LLMs in Autonomous Driving," in *2024 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2024, pp. 1084–1090. [Online]. Available: https://doi.org/10.1109/IV55156.2024.10588848

[42] X. Kong, T. Braunl, M. Fahmi, and Y. Wang, "A Superalignment Framework in Autonomous Driving with Large Language Models," in *2024 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2024, pp. 1715–1720. [Online]. Available: https://doi.org/10.1109/IV55156.2024.10588403

[43] C. Cui, Y. Ma, X. Cao, W. Ye, and Z. Wang, "Drive as You Speak: Enabling Human-Like Interaction with Large Language Models in Autonomous Vehicles," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2024, pp. 902–909. [Online]. Available: https://doi.org/10.1109/WACVW60836.2024.00101

[44] D. Fu, X. Li, L. Wen, M. Dou, P. Cai, B. Shi, and Y. Qiao, "Drive Like a Human: Rethinking Autonomous Driving with Large Language Models," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2024, pp. 910–919. [Online]. Available: https://doi.ieeecomputersociety.org/10.1109/WACVW60836.2024.00102

[45] Y. Wang, R. Jiao, C. Lang, S. S. Zhan, C. Huang, Z. Wang, Z. Yang, and Q. Zhu, "Empowering Autonomous Driving with Large Language Models: A Safety Perspective," *CoRR*, vol. abs/2312.00812, 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2312.00812

[46] Y.-J. Wang, B. Zhang, J. Chen, and K. Sreenath, "Prompt a Robot to Walk with Large Language Models," in *2024 IEEE 63rd Conference on Decision and Control (CDC)*. IEEE, 2024, pp. 1531–1538. [Online]. Available: https://doi.org/10.1109/CDC56724.2024.10885862

[47] L. Wen, D. Fu, X. Li, X. Cai, T. MA, P. Cai, M. Dou, B. Shi, L. He, and Y. Qiao, "DiLu: A Knowledge-Driven Approach to Autonomous Driving with Large Language Models," in *The Twelfth International Conference on Learning Representations*, 2024. [Online]. Available: https://openreview.net/forum?id=OqTMUPuLuC

[48] S. Sural, R. R. Rajkumar *et al.*, "ContextVLM: Zero-Shot and Few-Shot Context Understanding for Autonomous Driving Using Vision Language Models," in *2024 IEEE 27th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2024, pp. 468–475. [Online]. Available: https://doi.org/10.1109/ITSC58415.2024.10920066

[49] S. Tang, Z. Zhang, J. Zhou, L. Lei, Y. Zhou, and Y. Xue, "LeGEND: A Top-Down Approach to Scenario Generation of Autonomous Driving Systems Assisted by Large Language Models," in *Proceedings of the 39th IEEE/ACM International Conference on Automated Software Engineering*, 2024, pp. 1497–1508. [Online]. Available: https://doi.org/10.1145/3691620.3695520

[50] T. Cai, Y. Liu, Z. Zhou, H. Ma, S. Z. Zhao, Z. Wu, and J. Ma, "Driving with Regulation: Interpretable Decision-Making for Autonomous Vehicles with Retrieval-Augmented Reasoning via LLM," *arXiv preprint arXiv:2410.04759*, 2024. [Online]. Available: https://arxiv.org/html/2410.04759v1

[51] Z. Xu, Y. Zhang, E. Xie, Z. Zhao, Y. Guo, K.-Y. K. Wong, Z. Li, and H. Zhao, "DriveGPT4: Interpretable End-to-End Autonomous Driving via Large Language Model," *IEEE Robotics and Automation Letters*, 2024. [Online]. Available: https://doi.org/10.1109/LRA.2024.3440097

[52] T. Deruyttere, D. Grujicic, M. B. Blaschko, and M.-F. Moens, "Talk2Car: Predicting Physical Trajectories for Natural Language Commands," *Ieee Access*, vol. 10, pp. 123 809–123 834, 2022. [Online]. Available: https://doi.org/10.1109/ACCESS.2022.3224144

[53] T. I. Fossen, *Handbook of Marine Craft Hydrodynamics and Motion Control*, 2nd ed. John Wiley & Sons Ltd., 2021. [Online]. Available: https://onlinelibrary.wiley.com/doi/book/10.1002/9781119994138

[54] H. Imazu, "Research on Collision Avoidance Manoeuvre," *Tokyo University of Marine Science and Technology: Tokyo, Japan*, 1987.

[55] R. Sawada, K. Sato, and T. Majima, "Automatic Ship Collision Avoidance Using Deep Reinforcement Learning with LSTM in Continuous Action Spaces," *Journal of Marine Science and Technology*, vol. 26, no. 2, pp. 509–524, 2021. [Online]. Available: https://doi.org/10.1007/s00773-020-00755-0

[56] M. Akdağ, T. A. Pedersen, T. I. Fossen, and T. A. Johansen, "Hierarchical Collision Avoidance Algorithm with Route Exchange Concept for Autonomous and Conventional Ships," *TechRxiv*, vol. 2023, no. 1222, 2023. [Online]. Available: https://doi.org/10.36227/techrxiv.170327573.35843181/v

[57] P. Sarhadi and S. Yousefpour, "State of the Art: Hardware in the Loop Modeling and Simulation with Its Applications in Design, Development and Implementation of System and Control Software," *International Journal of Dynamics and Control*, vol. 3, no. 4, pp. 470–479, 2015. [Online]. Available: https://doi.org/10.1007/s40435-014-0108-3

[58] E. Tramacere, S. Luciani, S. Feraco, A. Bonfitto, and N. Amati, "Processor-in-the-Loop Architecture Design and Experimental Validation for an Autonomous Racing Vehicle," *Applied Sciences*, vol. 11, no. 16, p. 7225, 2021. [Online]. Available: https://doi.org/10.3390/app11167225

[59] C. Cui, Y. Ma, and e. Park, "LLM4AD: Large Language Models for Autonomous Driving—Concept, Review, Benchmark, Experiments, and Future Trends," *Proceedings of the IEEE*, pp. 1–36, 2026. [Online]. Available: https://doi.org/10.1109/JPROC.2026.3678829