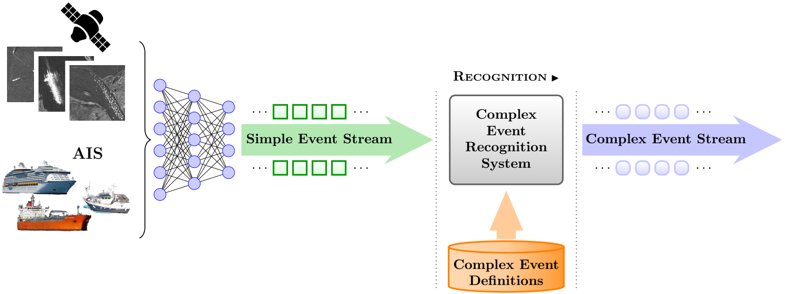
Maritime Situational Awareness in the era of Large Language/Reasoning Models

Alexander Artikis[1,2]
Andreas Kouvaras[1]

[1]University of Piraeus, Greece
[2]NCSR Demokritos, Athens, Greece

https://cer.iit.demokritos.gr

Complex Event Recognition (CER)



Large Language Models (LLMs)

What are LLMs?

Large Language Models (LLMs)

What are LLMs?

- Deep (unsupervised) learning models trained on large-scale natural language corpora.

Large Language Models (LLMs)

What are LLMs?

- Deep (unsupervised) learning models trained on large-scale natural language corpora.
- Based on transformer architectures.

Large Language Models (LLMs)

What are LLMs?

- Deep (unsupervised) learning models trained on large-scale natural language corpora.
- Based on transformer architectures.
- Predict the next token in context enabling language generation, summarisation, etc.

Large Language Models (LLMs)

What are LLMs?

- Deep (unsupervised) learning models trained on large-scale natural language corpora.
- Based on transformer architectures.
- Predict the next token in context enabling language generation, summarisation, etc.

Challenges:

- Sensitive to prompt phrasing.

Large Language Models (LLMs)

What are LLMs?

- Deep (unsupervised) learning models trained on large-scale natural language corpora.
- Based on transformer architectures.
- Predict the next token in context enabling language generation, summarisation, etc.

Challenges:

- Sensitive to prompt phrasing.
- Prone to hallucinations.

Large Language Models (LLMs)

What are LLMs?

- Deep (unsupervised) learning models trained on large-scale natural language corpora.
- Based on transformer architectures.
- Predict the next token in context enabling language generation, summarisation, etc.

Challenges:

- Sensitive to prompt phrasing.
- Prone to hallucinations.
- Limited reasoning capabilities.

LLMs in action

The ship turned off its AIS signal and ...

LLMs in action

The ship turned off its AIS signal and ... performed a suspicious activity near the port.

Large Reasoning Models (LRMs)

What are LRMs?

Large Reasoning Models (LRMs)

What are LRMs?

- Designed to think step-by-step.

Large Reasoning Models (LRMs)

What are LRMs?

- Designed to think step-by-step.
- Plan sequences, handle scheduling, solve puzzles.

Large Reasoning Models (LRMs)

What are LRMs?

- Designed to think step-by-step.
- Plan sequences, handle scheduling, solve puzzles.

Challenges:

- They still have limited reasoning capabilities.

Large Reasoning Models (LRMs)

What are LRMs?

- Designed to think step-by-step.
- Plan sequences, handle scheduling, solve puzzles.

Challenges:

- They still have limited reasoning capabilities.
- More expensive and slower than LLMs.

LLM/LRM-powered CER

Challenges in constructing composite event definitions:

- Absence of labels for composite events, prohibiting learning.

LLM/LRM-powered CER

Challenges in constructing composite event definitions:

- Absence of labels for composite events, prohibiting learning.
- Programmers lack domain knowledge.

LLM/LRM-powered CER

Challenges in constructing composite event definitions:

- Absence of labels for composite events, prohibiting learning.
- Programmers lack domain knowledge.
- Domain experts lack programming skills.

LLM/LRM-powered CER

Challenges in constructing composite event definitions:

- Absence of labels for composite events, prohibiting learning.
- Programmers lack domain knowledge.
- Domain experts lack programming skills.
- Complicated definitions (spatio-temporal constraints, background knowledge, hierarchies).

LLM/LRM-powered CER

Challenges in constructing composite event definitions:

- Absence of labels for composite events, prohibiting learning.
- Programmers lack domain knowledge.
- Domain experts lack programming skills.
- Complicated definitions (spatio-temporal constraints, background knowledge, hierarchies).

Solution:

- Employ LLMs/LRMs for generating composite event definitions.

Prompting for CER

Introducing RTEC

Prompting for CER

Introducing RTEC

RTEC Predicates
(**Prompt R**)

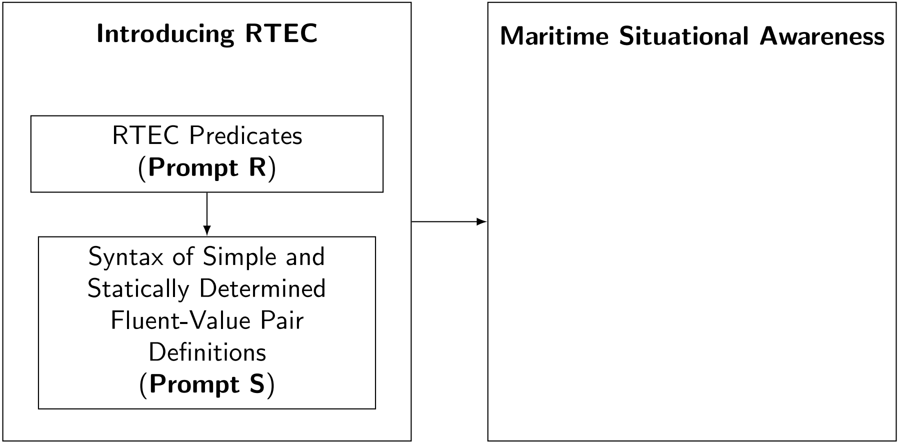
Prompting for CER

Introducing RTEC

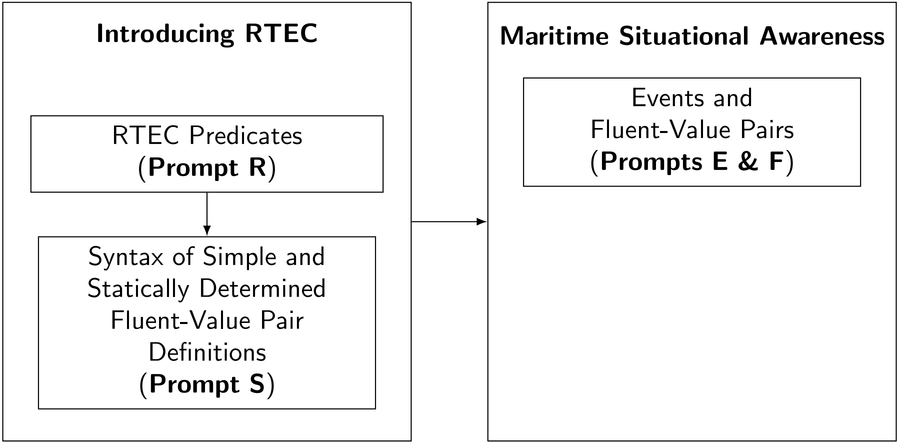
RTEC Predicates
(**Prompt R**)

Syntax of Simple and
Statically Determined
Fluent-Value Pair
Definitions
(**Prompt S**)

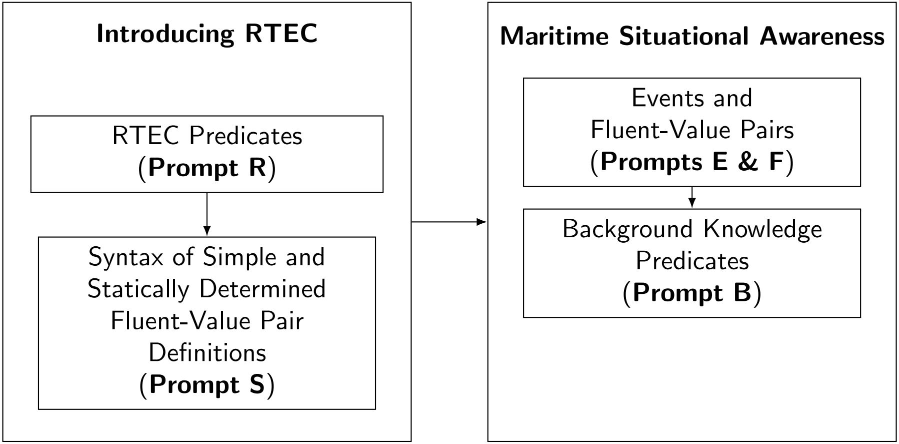
Prompting for CER



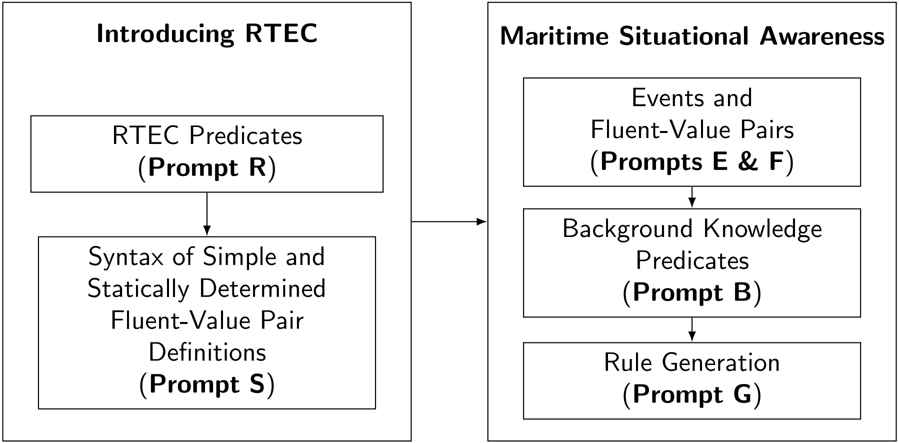
Prompting for CER



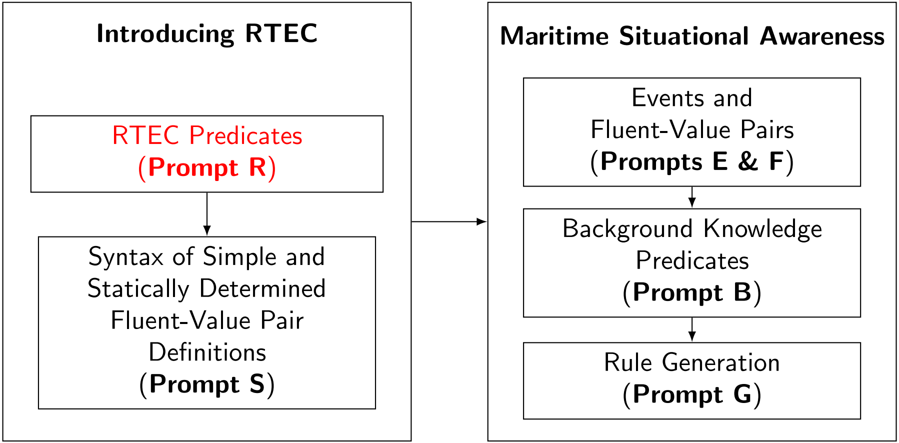
Prompting for CER



Prompting for CER



Prompting for CER



Prompt R

You are an assistant in constructing rules in the language of the Run-Time Event Calculus (RTEC), given a composite activity description in natural language. The Event Calculus is a logic-based formalism for representing and reasoning about events and their effects. The Run-Time Event Calculus (RTEC) is a Prolog programming implementation of the Event Calculus, that has been optimised for composite activity recognition. Below, we summarise the language of RTEC.

Prompt R (... continued)

Below are the predicates of RTEC.

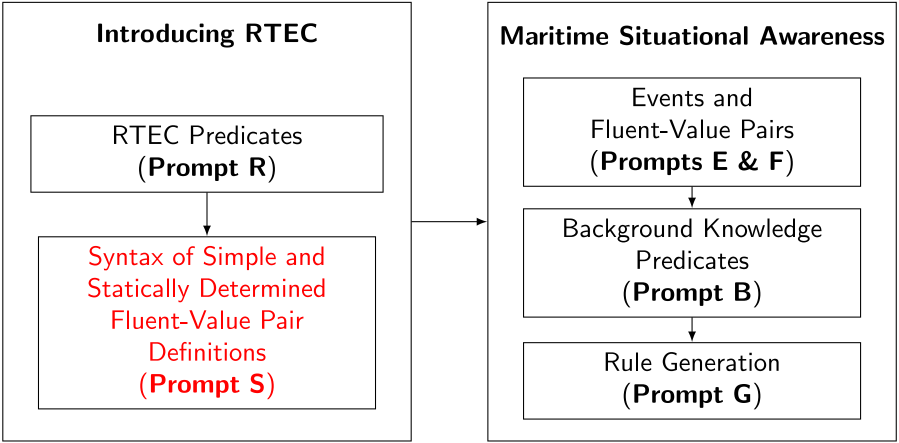
RTEC - Predicate 1: happensAt(E,T)
Meaning: Event E occurs at time T

RTEC - Predicate 2: holdsAt(F=V,T)
Meaning: The value of fluent F is V at time T

RTEC - Predicate 3: holdsFor(F=V,I)
Meaning: I is the list of the maximal intervals during which F=V holds continuously

RTEC - Remaining Predicates

Prompting for CER



Syntax of Simple Fluent-Value Pair Definitions

Prompt S

Example 1: Given a composite maritime activity description, provide the rules in the language of RTEC.

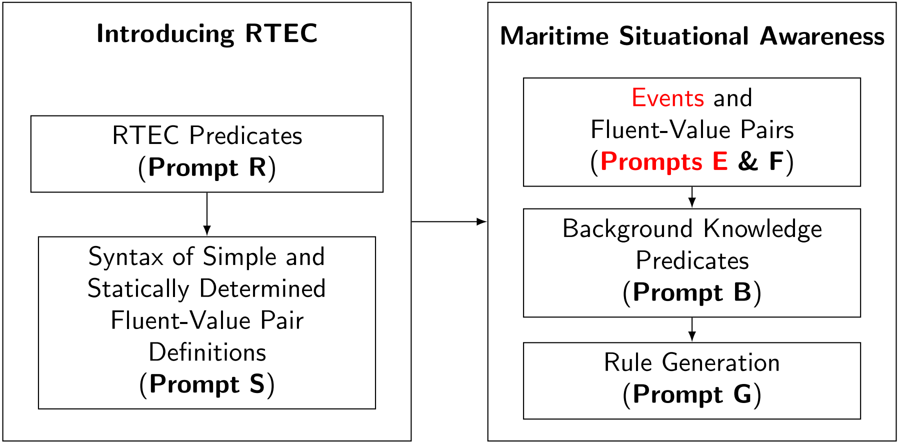
Composite Maritime Activity Description: 'withinArea'. This activity starts when a vessel enters an area of interest. The activity ends when the vessel leaves the area that it had entered, or when the vessel stops transmitting its position, since we can no longer assume that the vessel remains in the same area in the case of transmission gaps.

Prompt S (... continued)

The activity 'withinArea' is expressed as a Boolean simple fluent with two arguments, i.e., 'Vessel' and 'AreaType'. This activity starts when a vessel enters an area of interest. We use an 'initiatedAt' rule to express this initiation condition. The body literals of this rule are an event labelled 'entersArea' with two arguments, 'Vessel' and 'Area', and a background knowledge predicate named 'areaType' with two arguments, 'Area' and 'AreaType'. Below is rule in the language of RTEC:

$$
\begin{aligned}
&\mathrm{initiatedAt}(\mathit{withinArea}(\mathit{Vessel}, \mathit{AreaType}) = \mathrm{true}, T) \leftarrow \\
&\qquad \mathrm{happensAt}(\mathit{entersArea}(\mathit{Vessel}, \mathit{Area}), T), \\
&\qquad \mathit{areaType}(\mathit{Area}, \mathit{AreaType}).
\end{aligned}
$$

Prompting for CER



Prompt E

In addition to the built-in events of RTEC, you may use the following Maritime Situational Awareness (MSA) events:

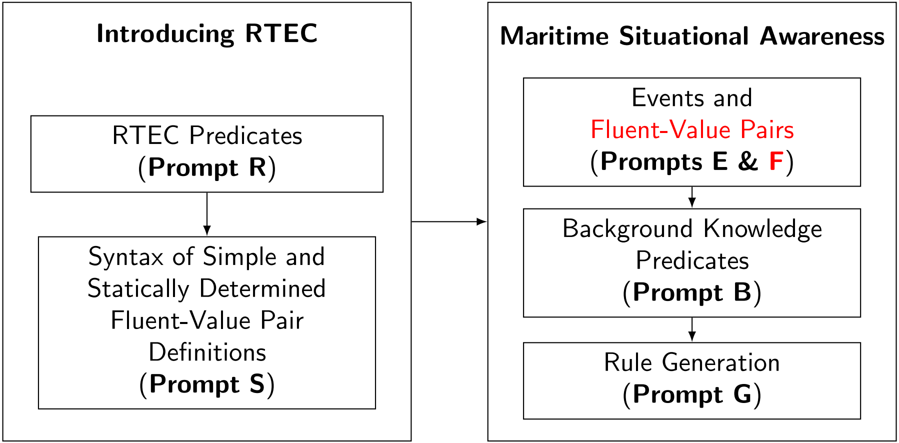
MSA - Event 1: change_in_speed_start(Vessel)
Meaning: 'Vessel' started changing its speed.

MSA - Event 2: change_in_speed_end(Vessel)
Meaning: 'Vessel' stopped changing its speed.

MSA - Event 3: change_in_heading(Vessel)
Meaning: 'Vessel' changed its heading.

MSA - Remaining Events

Prompting for CER

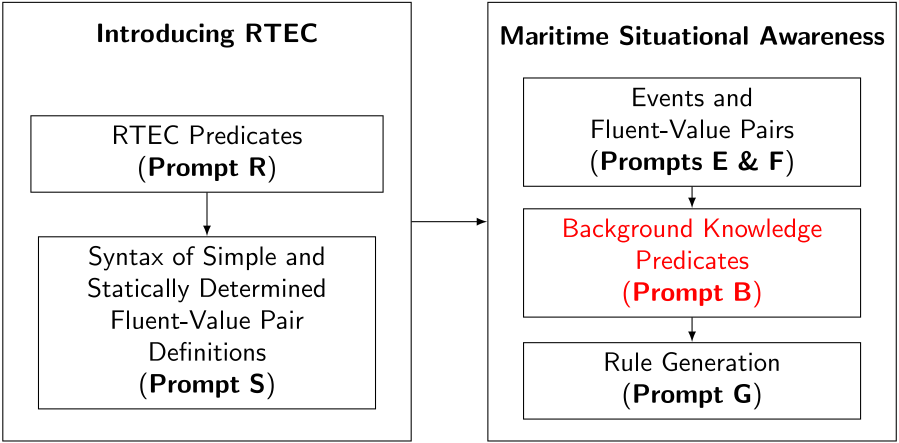


Prompt F

You may also use the following Maritime Situational Awareness (MSA) input fluent:

MSA - Input Fluent 1: proximity(Vessel1, Vessel2) = true
Meaning: 'Vessel1' and 'Vessel2' are close to each other.

Prompting for CER



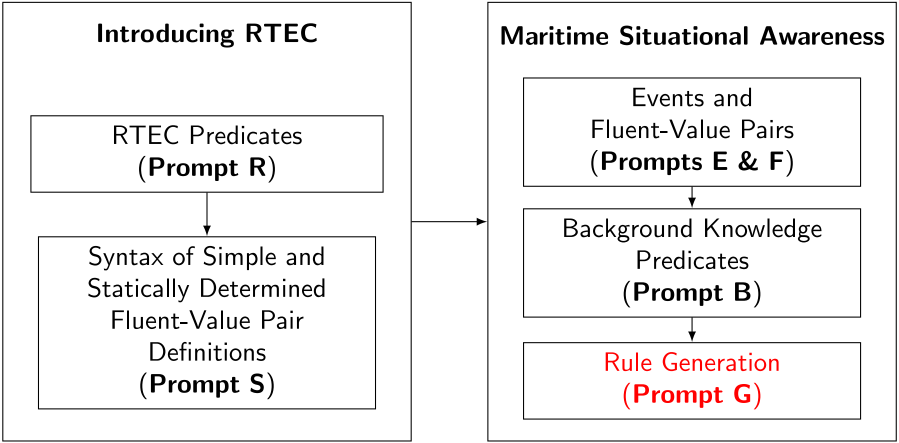
Background Knowledge Predicates

Prompt B
You may also use the MSA background knowledge predicates. MSA includes various such predicates. We start with the predicates named 'thresholds' which have two arguments. The first argument refers to the threshold type and the second one to the threshold value.

MSA Background Knowledge - Predicate 1:
thresholds(nearCoast, MaxSpeed)

Meaning: The maximum sailing speed that is safe for a vessel to have in a coastal area.
MSA Background Knowledge - Remaining Predicates

Prompting for CER



Prompt G

Given a composite activity description, provide the rules in the language of RTEC. You may use the built-in events of RTEC, the MSA events, the MSA input fluents, and the MSA background knowledge predicates. You may also use any of the MSA output fluents that you have already learned.

Composite Maritime Activity Description - 'gap': A communication gap starts when we stop receiving messages from a vessel. We would like to distinguish the cases where a communication gap starts (i) near some port and (ii) far from all ports. A communication gap ends when we resume receiving messages from a vessel.

Evaluating the responses of LLMs/LRMs

- Compute the syntactic similarity between
  - LLM/LRM-generated composite event definitions and
  - hand-crafted composite event definitions.

Evaluating the responses of LLMs/LRMs

- Compute the syntactic similarity between
  - LLM/LRM-generated composite event definitions and
  - hand-crafted composite event definitions.
- Compute the predictive accuracy of the LLM/LRM-generated composite event definitions.

Syntactic Similarity*

Hand-crafted rule:

$$\begin{aligned}&\mathrm{initiatedAt}(\mathit{withinArea}(\mathit{Vessel}, \mathit{AreaType}) = \mathrm{true}, T) \leftarrow\\&\qquad \mathrm{happensAt}(\mathit{entersArea}(\mathit{Vessel}, \mathit{Area}), T),\\&\qquad \mathit{areaType}(\mathit{Area}, \mathit{AreaType}).\end{aligned}$$

LLM/LRM-generated rule:

$$\begin{aligned}&\mathrm{initiatedAt}(\mathit{withinArea}(\mathit{Vessel}, \mathit{AreaType}) = \mathrm{true}, T) \leftarrow\\&\quad \mathrm{happensAt}(\mathit{entersArea}(\mathit{Vessel}, \mathit{Area}), T),\\&\quad \mathit{areaType}(\mathit{AreaType}, \mathit{Area}).\end{aligned}$$

*Kouvaras et al, Generating Activity Definitions with Large Language Models. EDBT, 2025.

Syntactic Similarity*

Hand-crafted rule:

$$\begin{aligned}&\mathrm{initiatedAt}(\mathit{withinArea}(\mathit{Vessel}, \mathit{AreaType}) = \mathrm{true},\ T) \leftarrow\\ &\qquad \mathrm{happensAt}(\mathit{entersArea}(\mathit{Vessel}, \mathit{Area}),\ T),\\ &\qquad \mathit{areaType}(\mathit{Area}, \mathit{AreaType}).\end{aligned}$$

LLM/LRM-generated rule:

$$\begin{aligned}&\mathrm{initiatedAt}(\mathit{withinArea}(\mathit{Vessel}, \mathit{AreaType}) = \mathrm{true},\ T) \leftarrow\\ &\qquad \mathrm{happensAt}(\mathit{entersArea}(\mathit{Vessel}, \mathit{Area}),\ T),\\ &\qquad \mathit{areaType}(\mathit{AreaType}, \mathit{Area}).\end{aligned}$$

* Kouvaras et al, Generating Activity Definitions with Large Language Models. EDBT, 2025.

Qualitative Analysis: Logical Inconsistencies

Prompt: ‘trawling movement’

This activity indicates that the vessel is sailing in a manner that is typical for trawling. The activity starts when the vessel changes its heading while sailing within a fishing area.

Qualitative Analysis: Logical Inconsistencies

Prompt: 'trawling movement'

This activity indicates that the vessel is sailing in a manner that is typical for trawling. The activity starts when the vessel changes its heading while sailing within a fishing area.

Gemini 2.5 Pro generated the following rule:

$$\begin{aligned}
&\mathrm{initiatedAt}(\mathit{trawlingMovement}(\mathit{Vessel}) = \mathrm{true},\ T) \leftarrow \\
&\quad \mathrm{happensAt}(\mathit{changeInHeading}(\mathit{Vessel}),\ T), \\
&\quad \mathrm{holdsAt}(\mathit{withinArea}(\mathit{Vessel},\ \mathit{FishingArea}) = \mathrm{true},\ T).
\end{aligned}$$

Qualitative Analysis: Weaker Definitions

Prompt: 'trawling speed'
We are interested in computing the intervals during which a vessel is sailing in a fishing area and its speed is within the expected bounds for trawling. The activity ends when the speed of the vessel falls outside the expected bounds. Moreover, when there is a gap in signal transmissions, we can no longer assume that the vessel's speed remains within the aforementioned bounds. The activity also ends when the vessel leaves the fishing area.

Qualitative Analysis: Weaker Definitions

Prompt: 'trawling speed'
We are interested in computing the intervals during which a vessel is sailing in a fishing area and its speed is within the expected bounds for trawling. The activity ends when the speed of the vessel falls outside the expected bounds. Moreover, when there is a gap in signal transmissions, we can no longer assume that the vessel's speed remains within the aforementioned bounds. The activity also ends when the vessel leaves the fishing area.

Gemini 2.5 Pro did not generate the following rule:

$$\begin{aligned}&\mathsf{terminatedAt}(\mathit{trawlingSpeed}(\mathit{Vessel}) = \mathsf{true},\ T) \leftarrow \\ &\quad \mathsf{happensAt}(\mathit{end}(\mathit{withinArea}(\mathit{Vessel}, \mathit{fishing}) = \mathsf{true}),\ T).\end{aligned}$$

Qualitative Assessment: Weaker Rules

Prompt: ‘moving speed’
We are interested in computing the intervals during which a vessel of a certain type is moving at a speed that is ‘below’ the threshold of the minimum speed expected for a vessel of this type [..].

Qualitative Assessment: Weaker Rules

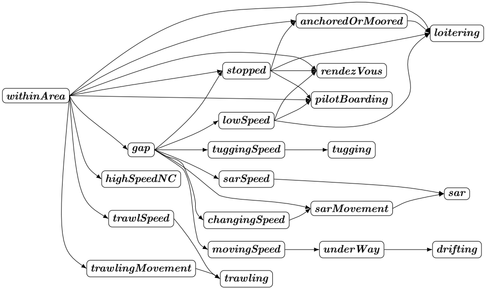
Prompt: 'moving speed'
We are interested in computing the intervals during which a vessel of a certain type is moving at a speed that is 'below' the threshold of the minimum speed expected for a vessel of this type [..].

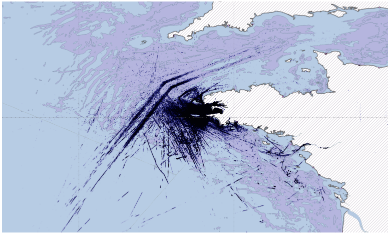
GPT-4o generated the following rule:

$$\begin{aligned}
&\mathsf{initiatedAt}(\mathit{movingSpeed}(\mathit{Vessel}) = \mathit{below}, T) \leftarrow \\
&\quad \mathsf{happensAt}(\mathit{velocity}(\mathit{Vessel}, \mathit{Speed}, _, _), T), \\
&\quad \mathit{vesselType}(\mathit{Vessel}, \mathit{Type}),\ \ \mathit{typeSpeed}(\mathit{Type}, \mathit{Min}, _, _), \\
&\quad \mathit{thresholds}(\mathit{movingMin}, \mathit{MovingMin}), \\
&\quad \mathit{Speed} > \mathit{MovingMin}, \\
&\quad \mathit{Speed} < \mathit{Min}.
\end{aligned}$$

Predictive Accuracy

- o4-mini-high, o3 and GPT-4o exhibit F1-score more than 80%.





Summary

What we've seen so far:

- LLM/LRM-powered CER.

Summary

What we've seen so far:

- LLM/LRM-powered CER.
- Prompting Guidelines for CER.

Summary

What we've seen so far:

- LLM/LRM-powered CER.
- Prompting Guidelines for CER.
- Qualitative and Quantitative Assessment.

Summary

What we've seen so far:

- LLM/LRM-powered CER.
- Prompting Guidelines for CER.
- Qualitative and Quantitative Assessment.

Next: Open Issues and Topics not covered