

西安電子科技大學
XIDIAN UNIVERSITY

TLCrafter: Peeling the Layers of Natural Language to Reveal Temporal Logic Specifications

Cheng Wen[2], Wenjun Ke[2], Xiao Liang[3], Zhi Ma[1], Xiaoyong Xue[4] ,
Shengchao Qin[1], and Cong Tian[1]







Background & Motivation

Background

- Formal specifications are essential for verification and model checking.
- Formalizing NL requirements requires identifying the intended semantics and choosing appropriate operators, scopes, and predicates.
- LTL, CTL, and STL capture different temporal semantics.

Zero-Shot Prompt: You are an expert in the field of formal verification, proficient in the rules and usage of LTL formulas. Please translate the following natural language requirement into LTL formulas.

Req01: In the inertial navigation valid mode, the output will be set to reverse navigation computation, or the output will be set to low velocity low altitude navigation computation, or the output will be set to high velocity navigation computation, or the output will be set to high altitude navigation computation.	**Req02**: Globally, if the navigation system computes angles using dual inertial navigation modules, the input attitude angle shall be the weighted average of angles from module 1 and module 2 (weights 0.6 and 0.4), unless the system switches to single module or GPS signal is restored.
LTL01 by GPT-4o: G (valid_mode → (output = reverse_nav ∨ output = lowVlowH_nav ∨ output = highV_nav ∨ output = highH_nav))	**LTL02 by GPT-4o:** G ((Dual_INU_Active ∧ ¬Single_Module ∧ ¬GPS_Restored) → (Input_Attitude_Angle = 0.6 * Angle_Module1 + 0.4 * Angle_Module2))
Ground Truth: G (valid_mode → (F output = reverse_nav ∨ F output = lowVlowH_nav ∨ F output = highV_nav ∨ F output = highH_nav))	Ground Truth: G (Dual_INU_Active → (Input_Attitude_Angle = 0.6 * Angle-Module1+ 0.4 * Angle_Module2) U (Single_Module ∨ GPS_Restored))

Fig. 1. **Examples showcasing the discrepancy between LLM generated LTL and ground truth for complex requirements.** The red-circle indicates errors in the LLM's output due to missing details, contrasted with the manually corrected specification.

Overview of TlCrafter

Multi-logic support

- Recommends or accepts a target logic based on requirement semantics (LTL, CTL, or STL)
- OnionL⁺represents temporal scopes, CTL path quantifiers, STL time intervals, and atomic predicates.

Inspectable formalization

- The visual OnionL⁺tree exposes the inferred structure for inspection and localized correction.
- Syntax validation and deterministic synthesis regenerate the formula after each edit.

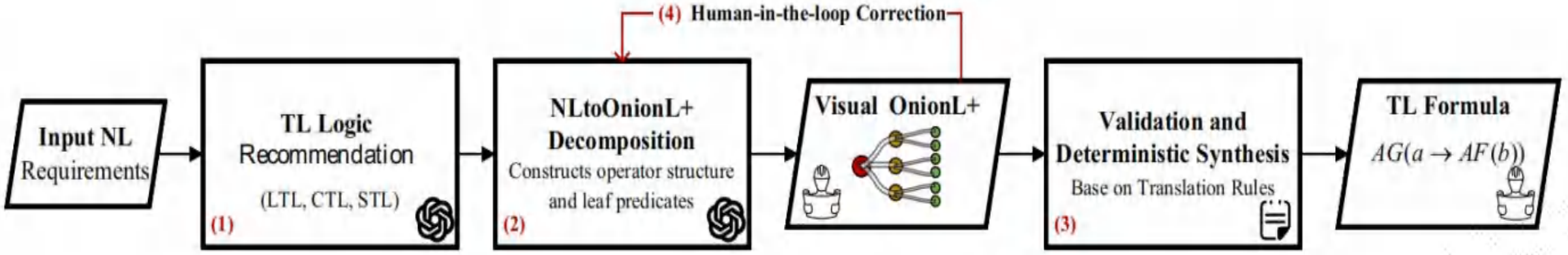


Figure 1: The High-level Workflow of TLCrafter, Built Around ONIONL⁺ as an IR

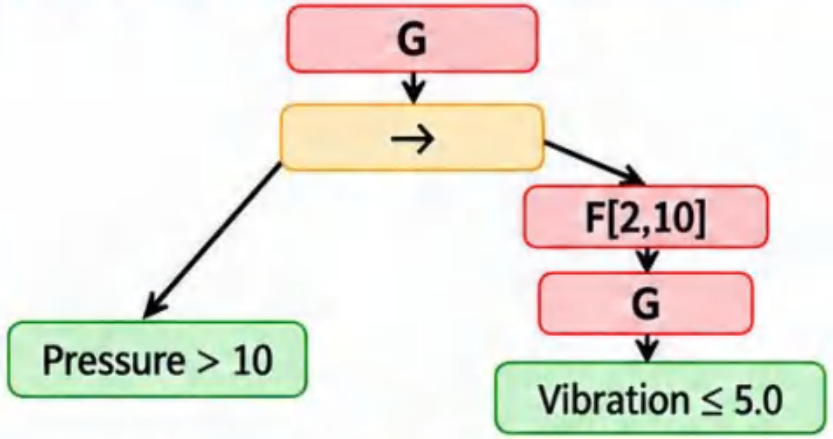
What is OnionL+?

OnionL+ is a logic-parametric, typed abstract syntax tree (AST)

$$\tau = (V, E, r, \lambda)$$

V — nodes	E — parent-child edges	r — root node	λ — typed node labels

A typed tree



operators inside • predicates at leaves

Predicate leaf

$$\langle \text{component, variable, relation, expression} \rangle$$

$$\langle \text{Chamber, Pressure, } >, \text{ } 10.0 \rangle$$

LOGIC-SPECIFIC NODES | **LTL** G F X U | **CTL** A E (path quantifiers) | **STL** F_[a,b] G_[a,b] U_[a,b] (metric intervals)

An Eaxmple

Req-NL:
Globally, in case that the Combustion Chamber Pressure signal is greater than 10.0 MPa and less than 15.0 MPa, and the value of Turbopump Speed is no less than 20,000 rpm and smaller than 25,000 rpm then all of the following conditions hold: there needs to exist a time within 2 to 10 time units, at which High-Frequency Vibration is at most 5.0 g consecutively in the future before the end of the execution.

Output: OnionL

Scopes | Relations | Atomic Propositions (APs)

Step 1: Scope Identification & Separation
Step 2: Top-Level Structure Construction

```
(G
  (In case that the Combustion Chamber Pressure signal is greater than 10.0 MPa and less than 15.0 MPa, and the value of Turbopump Speed is no less than 20,000 rpm and smaller than 25,000 rpm then all of the following conditions hold: there needs to exist a time within 2 to 10 time units, at which High-Frequency Vibration is at most 5.0 g consecutively in the future before the end of the execution.))
```

Step 3: Unary Operator Extraction
Step 4: Binary Operator Decomposition

```
(G
  (Implies
    ((Left)the Combustion Chamber Pressure signal is greater than 10.0 MPa and less than 15.0 MPa, and the value of Turbopump Speed is no less than 20,000 rpm and smaller than 25,000 rpm
    (Right) all of the following conditions hold: there needs to exist a time within 2 to 10 time units at which High-Frequency Vibration is at most 5.0 g consecutively in the future before the end of the execution) )))
```

Step 3 / Step 4 — Step 5: Atomicity Check & Termination

```
(G
  (Implies
    (AND
      (AND
        (the Combustion Chamber Pressure signal is greater than 10.0 MPa
        the Combustion Chamber Pressure signal is less than 15.0 MPa )
      (AND
        (the value of Turbopump Speed is no less than 20,000 rpm
        the value of Turbopump Speed is smaller than 25,000 rpm ))
    (F[2,10]
      (G
        (High-Frequency Vibration is at most 5.0 g ))))
```

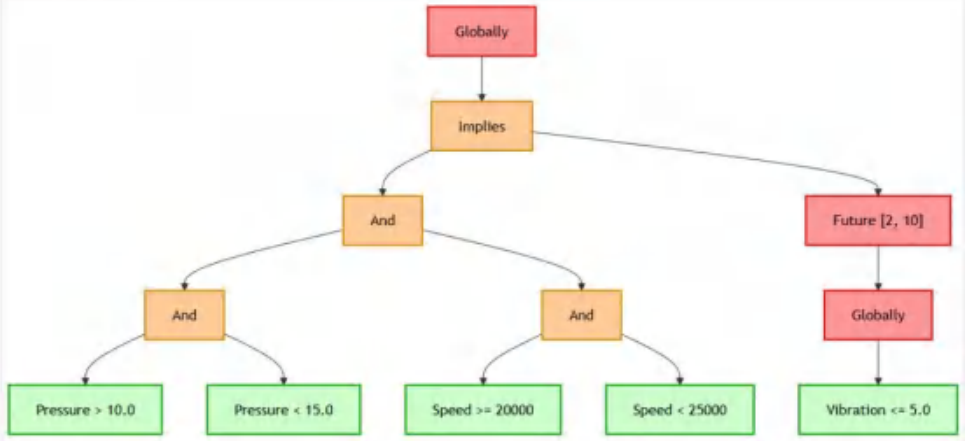
Step 6: Atomic Proposition Normalization

```
(G
  (Implies
    (AND
      (AND
        (Com="CombustionChamber", Var="Pressure", Rel=">", Formula="10.0"
        Com="CombustionChamber", Var="Pressure", Rel="<", Formula="15.0" )
      (AND
        Com="Turbopump", Var="Speed", Rel=">=", Formula="20000"
        Com="Turbopump", Var="Speed", Rel="<", Formula="25000" ))
    (F[2, 10]
      (G
        (Com="System", Var="HighFrequencyVibration", Rel="<=", Formula="5.0"))))
```

TLCrafter Pipeline → STL

ONIONL+ TREE

Visualization



GENERATED STL FORMULA

$$G\big((P > 10.0 \wedge P < 15.0) \wedge (S \geq 20000 \wedge S < 25000) \rightarrow F[2,10](G(V \leq 5.0))\big)$$

F[2,10] within 2–10 time units

G “consecutively in the future before the end of the execution”

Interactive Correction

Interactive & Auditable Correction

- Inspect the visual OnionL⁺ tree to review scopes, operators, quantifiers, intervals, and predicates.
- Use natural-language edits to correct local structures; TLCrafter then revalidates the tree and regenerates the formula.



Evaluation

Evaluation across three logic datasets

112
LTL-Aero
Aerospace requirements from two spacecraft subsystems

62
STL-62
Signal-logic pairs from DeepSTL and prior work

100
CTL-100
Robotics scenarios: navigation, recovery and contingency handling

Table 1: Validity and Formula Accuracy on Datasets Across *LTL-Aero*, *STL-62*, and *CTL-100*.

Setting: LLM + Prompt Strategy	LTL		STL		CTL	
	Syntax Validity	Formula Accuracy	Syntax Validity	Formula Accuracy	Syntax Validity	Formula Accuracy
GPT-4o + Zero-Shot	89%	43.8%	96.8%	45.1%	89%	19%
GPT-4o + NL2SPEC	91.9%	56.3%	96.8%	48.3%	93%	18%
GPT-4o + TLCrafter	100%	**88.4%**	100%	**64.5%**	100%	**74%**
DeepSeek-V3 + Zero-Shot	87.5%	34.8%	90.3%	48.3%	90%	26%
DeepSeek-V3 + NL2SPEC	91.9%	54.5%	91.9%	41.9%	93%	18%
DeepSeek-V3 + TLCrafter	100%	**86.6%**	100%	**72.6%**	100%	**65%**

100% syntax validity with TLCrafter across LTL, STL and CTL

Formula accuracy measures equivalence to the expert-validated reference formula.

XXXXX

- XXX
- XXX

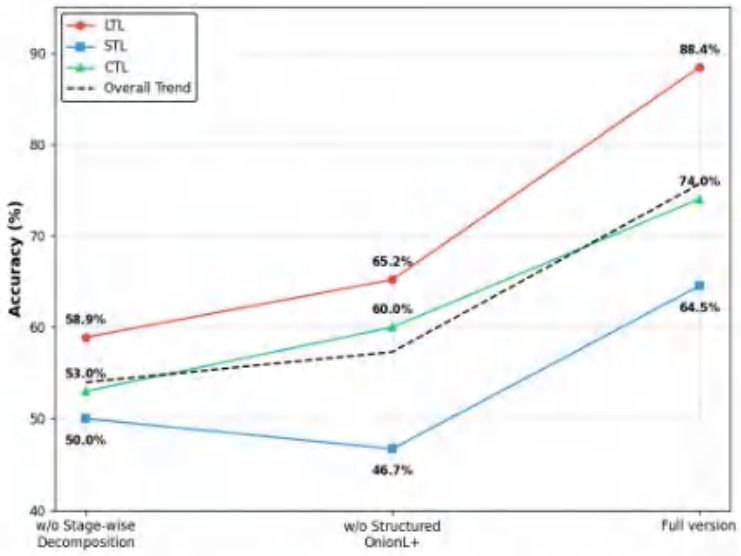


Figure 5: Ablation Study with GPT-4o

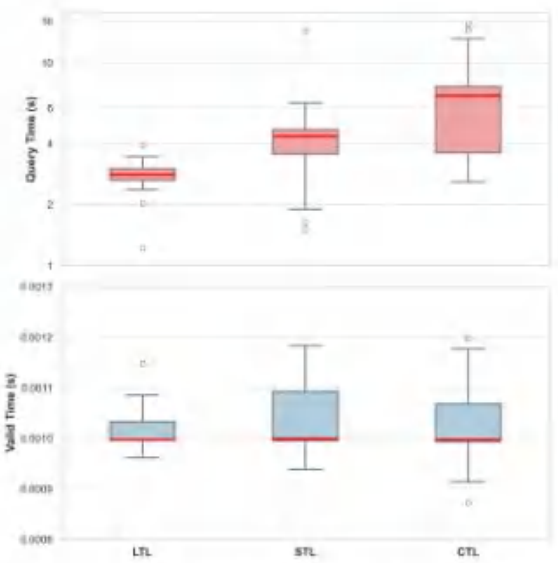


Figure 6: Runtime Overhead

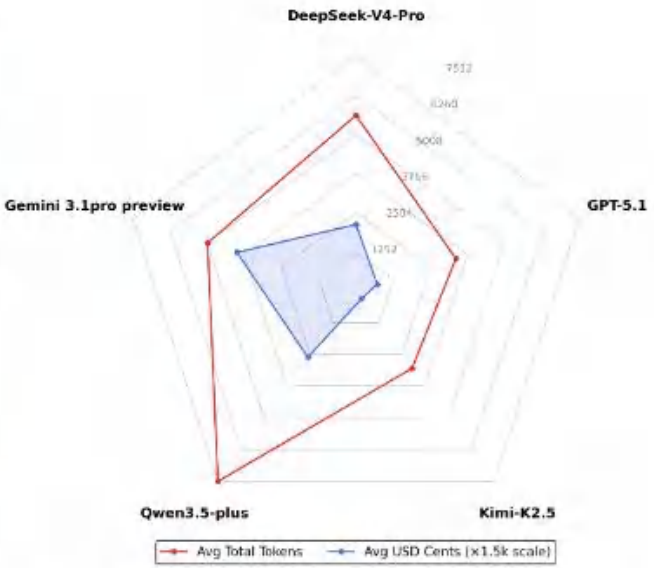
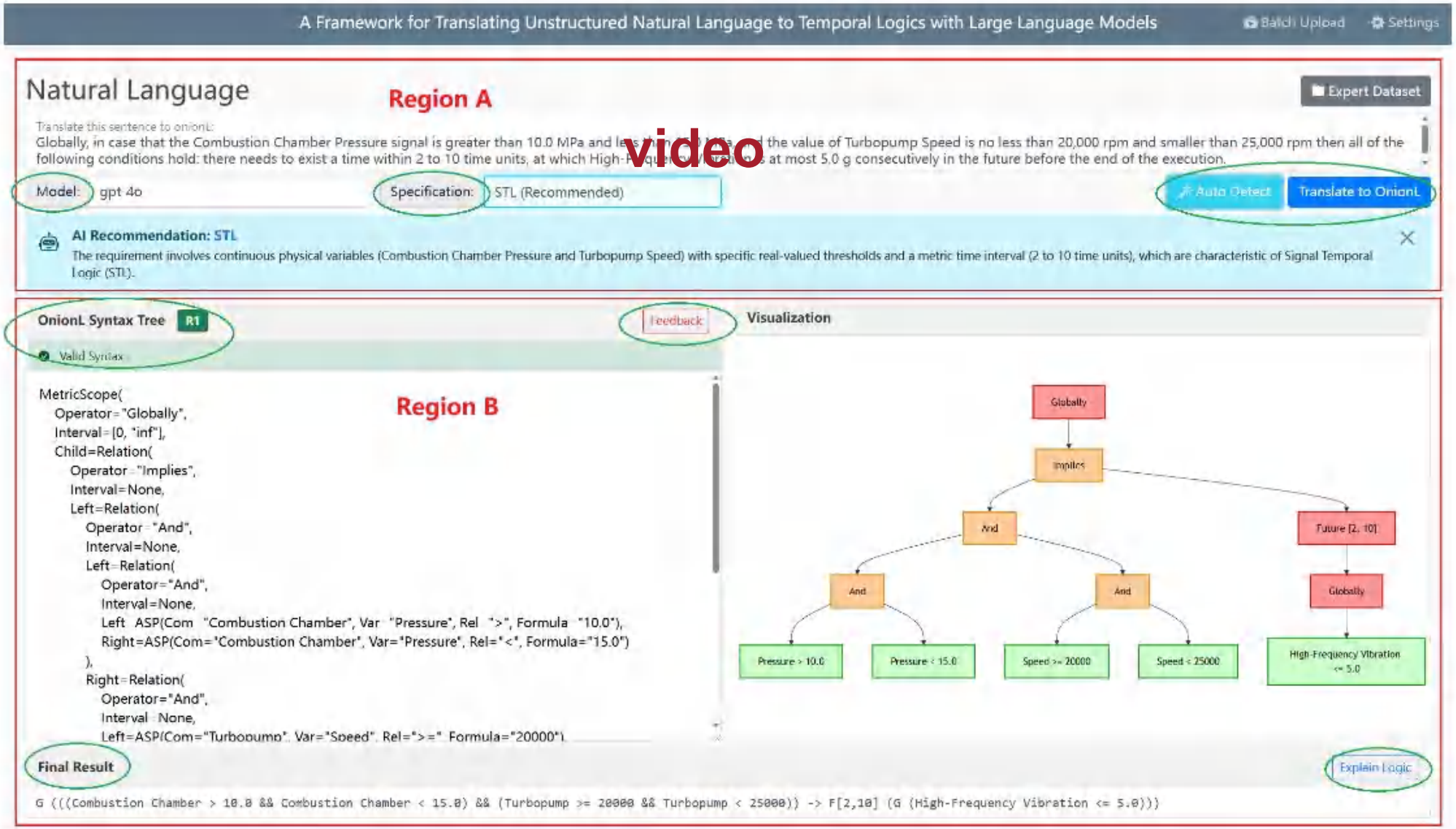


Figure 7: Token Usage and API Cost.

Demo


A Framework for Translating Unstructured Natural Language to Temporal Logics with Large Language Models
Batch Upload
Settings
Natural Language
Region A
Expert Dataset
Translate this sentence to OnionL:
Globally, in case that the Combustion Chamber Pressure signal is greater than 10.0 MPa and le
the value of Turbopump Speed is no less than 20,000 rpm and smaller than 25,000 rpm then all of the following conditions hold: there needs to exist a time within 2 to 10 time units, at which High-
at most 5.0 g consecutively in the future before the end of the execution.
video
Model:
gpt 4o
Specification:
STL (Recommended)
Auto Detect
Translate to OnionL
AI Recommendation: STL
The requirement involves continuous physical variables (Combustion Chamber Pressure and Turbopump Speed) with specific real-valued thresholds and a metric time interval (2 to 10 time units), which are characteristic of Signal Temporal Logic (STL).
OnionL Syntax Tree
R1
Feedback
Visualization
Valid Syntax
Region B
MetricScope(
Operator="Globally",
Interval=[0, "inf"],
Child=Relation(
Operator="Implies",
Interval=None,
Left=Relation(
Operator="And",
Interval=None,
Left=Relation(
Operator="And",
Interval=None,
Left=ASP(Com="Combustion Chamber", Var="Pressure", Rel=">", Formula="10.0"),
Right=ASP(Com="Combustion Chamber", Var="Pressure", Rel="<", Formula="15.0")
),
Right=Relation(
Operator="And",
Interval=None,
Left=ASP(Com="Turbopump", Var="Speed", Rel=">=", Formula="20000"),
Globally
Implies
And
Future [2, 10]
And
And
Globally
Pressure > 10.0
Pressure < 15.0
Speed >= 20000
Speed < 25000
High-Frequency Vibration <= 5.0
Final Result
Explain Logic
G (((Combustion Chamber > 10.0 && Combustion Chamber < 15.0) && (Turbopump >= 20000 && Turbopump < 25000)) -> F[2,10] (G (High-Frequency Vibration <= 5.0)))



西安電子科技大學
XIDIAN UNIVERSITY

Thanks for listening