FORMAS

Supporting Pragmatic Interoperability: An LLM Based Process to Analyze Distributed Intentionality (i*) Models

MoDRE:The 15th International Model-Driven Requirements Engineering Workshop

Roberto Figueiredo, Julio Cesar Leite, Célia Ralha, Rita Maciel, Daniela Claro

julioleite@ufba.br

September 2, 2025

Programa de Pós-graduação em Ciência da Computação

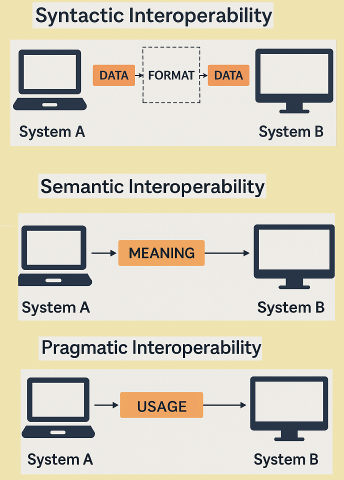
Universidade Federal da Bahia

Outline

- Context
- Motivation
- Research Problem
- Scenario
- Research Design
  - Strategic Rationale Model
  - Variation Points
  - Experiment
  - Results
- Conclusion

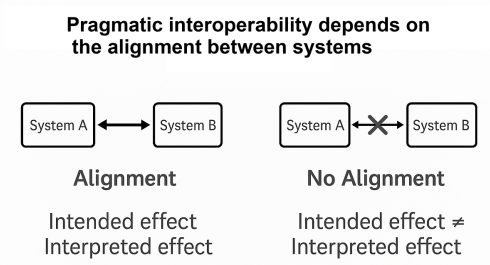
Context

- **Pragmatic Interoperability** goes beyond syntactic (format) and semantic (meaning) interoperability.
  - Ensures exchanged data is interpreted and acted upon consistently in a given **context**.
  - Misalignment arises when the **intended effect** $\neq$ **interpreted effect**.



Motivation

- Interoperability is critical for system collaboration.
- Pragmatic interoperability focuses on how exchanged data is **used**.
- Challenge: ensuring that intended and interpreted effects of messages align.
- Automatic analysis of intentional models can support this.



Comparison with Related Work

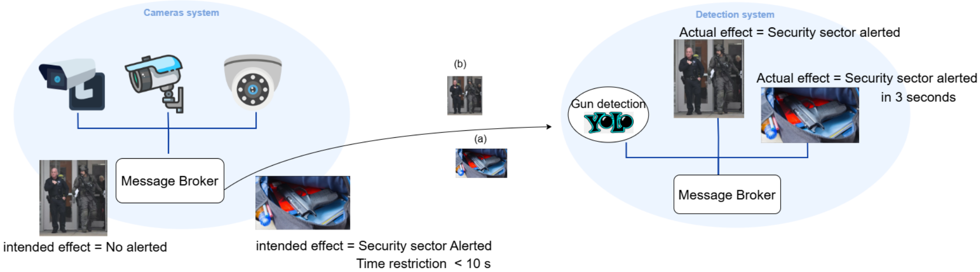
Work	Focus	Limitation
Siddeshwar et al.	Extract goals from user stories	Design-time only
Chen et al.	Build GRL models	Structural validity only
Fantechi et al.	Variability in requirements	No goal interpretation
Hassine	Traceability to goals	Security design focus
Our Work	LLM + i* SR models	Runtime focused

Table: Comparison of contributions

Research Problem

- Research question: *Does evaluating intentional models automatically improve pragmatic interoperability?*

Scenario: Surveillance system based on CCTV cameras



- (a) Gun hidden in backpack (Camera) → Detection system → True alert.
- (b) Police officer with gun (Camera) → Detection system → False alert.

Research Design

- Model → Build SR model.
- Declare → Prepare prompts.
- Analyze → Use LLM.
- Evaluate → Measure performance.

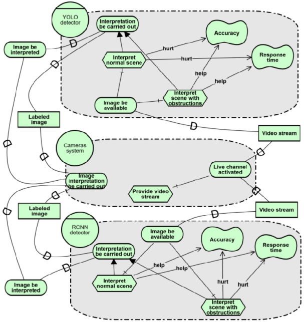
Research Design

"A Variation Point is an explicit engineering decision that allows for multiple alternative Variants to be chosen during system development, based on specific Assets."

"In some systems, the selection of one variation point can lead to dependencies on other variation points. For instance, choosing a movie player might require selecting an audio player as well."

(1)

"Means-end links represent the relationship between an 'end' such as a goal, resource, or satisfied goal and alternative 'means' to achieve it, typically expressed as tasks. In a notation, an arrow points from the means to the end."



OpenAI API

Performance evaluation of detected independent variation points

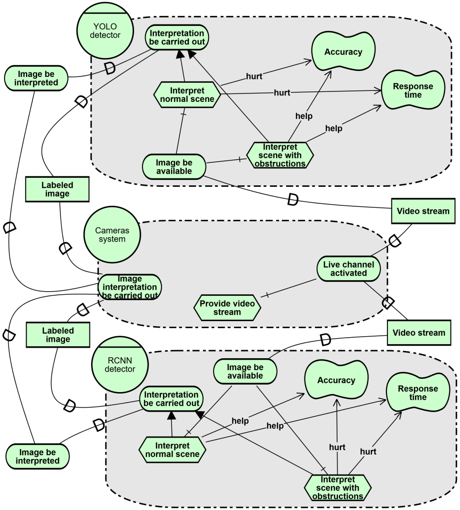
	POSITIVE	NEGATIVE
ACTUAL VALUES POSITIVE	TP	FN
ACTUAL VALUES NEGATIVE	FP	TN

$$Precision = \frac{TP}{TP + FP} \qquad Recall = \frac{TP}{TP + FN}$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Strategic Rationale (SR) Models

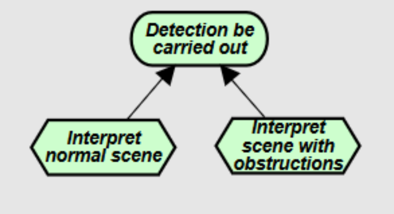
- Capture goals, tasks, softgoals, and resources.
- Model system capabilities and alternatives.



Variation point

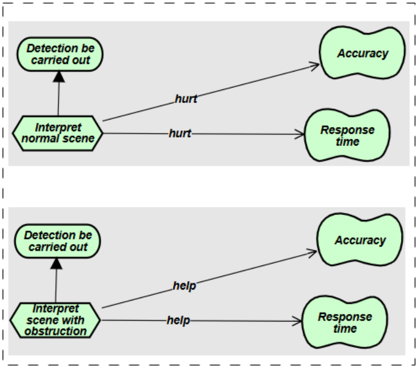
- Variation Points (VP): decision nodes
  - `(detection = interpretation)`

.

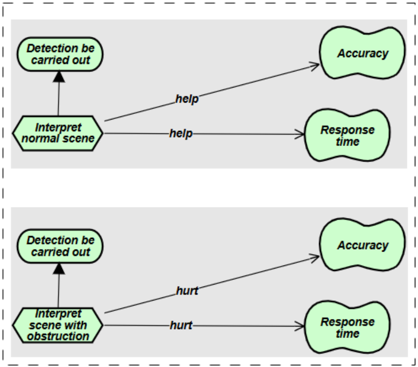


Variants for each actor

YOLO detection system variants



RNCC detection system variants

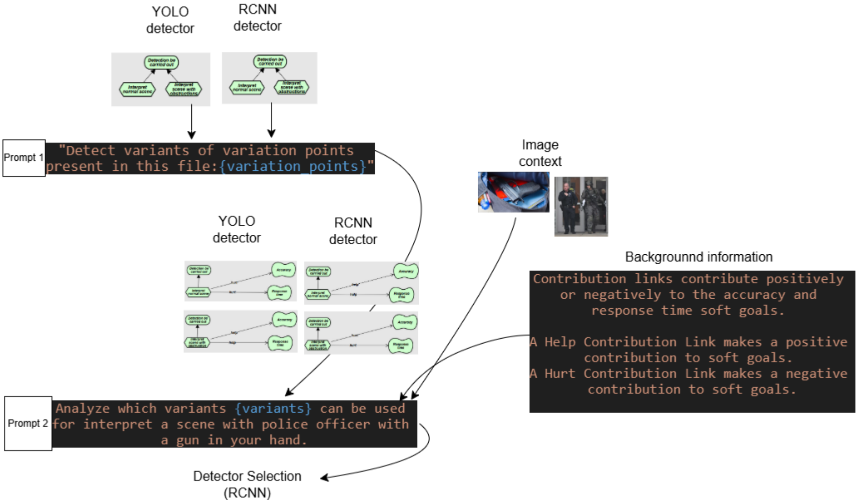


Prompt to detect variation point

- Prompt receives a txt file (raw_background) with information extracted from articles.
- Prompt receives the SR model in json file format (goal_model_str).

```
# Combine background information with the goal model and with the prompt task.
prompt = (f"Background: {raw_background}\n\nHere is a goal model: {goal_model_str}."
          " Detect independent explicit designed variation points in this goal model.")
```

Prompt to detect variants and select the one that aligns.



Outline
Context
Motivation
Comparison
Research Problem
Scenario
Research Design
Model
13 Declare
Analyze
Evaluate
Conclusion
Ack
Contacts

Experiment

- Controlled experiment with GPT-3.5.
- Zero-shot vs. contextual prompts.
- Metrics: Precision, Recall, F1-score.

Experiment steps:

- Variation point detection.
- Variant analysis for alignment.

Variation point (VP) detection

Table: Intentional elements in SR Model

Element Type	Intentional Element
A	Interpretation be carried out
B	Interpret normal scene
C	Interpret scene with obstructions
D	Accuracy
E	Response time
F	Image be available

Variation point (VP) detection

Table: Retrieved Sets for each background information

Retrieval #	Retrieved Set	Evaluation
1	{A, B, C, D, E, F}	False Positive (extra elements)
2	{B, C}	False Negative (missing A)
3	{A, B, C, D, E, F}	False Positive (extra elements)
4	{A, B, C, D, E}	False Positive (extra elements)
5	{A, B, C}	True Positive (exact match)

Element-Level Evaluation (Standard IR Way)- Results achieved in VP detection

Each retrieved *element* is evaluated against $G = \{A, B, C\}$ (partial credit allowed).

#	Retrieved Set	TP	FP	FN	P	R	F1
1	{A, B, C, D, E, F}	3	3	0	0.50	1.00	0.67
2	{B, C}	2	0	1	1.00	0.67	0.80
3	{A, B, C, D, E, F}	3	3	0	0.50	1.00	0.67
4	{A, B, C, D, E}	3	2	0	0.60	1.00	0.75
5	{A, B, C}	3	0	0	1.00	1.00	1.00

Macro-avg: P $\approx$ 0.72, R $\approx$ 0.93, F1 $\approx$ 0.78
Micro-avg: P $\approx$ 0.64, R $\approx$ 0.93, F1 $\approx$ 0.76
Reading: High recall (A,B,C usually found) but precision drops due to extra elements (D,E,F).

Prompt-Level Strict Evaluation (Exact-Match Rule) - Results achieved in VP detection

A retrieval counts as **True Positive** only if the set is *exactly* $\{A, B, C\}$. Subsets $\Rightarrow$ **FN**, supersets (extras) $\Rightarrow$ **FP**.

#	Retrieved Set	Evaluation
1	{A, B, C, D, E, F}	False Positive (extra elements)
2	{B, C}	False Negative (missing A)
3	{A, B, C, D, E, F}	False Positive (extra elements)
4	{A, B, C, D, E}	False Positive (extra elements)
5	{A, B, C}	True Positive (exact match)

Counts: TP = 1, FP = 3, FN = 1
Metrics: Precision = 0.25, Recall = 0.50, F1 = 0.33

Reading: Strict rule is unforgiving; any deviation from {A,B,C} degrades precision and F1.

Variant analysis for alignment

- LLM identifies YOLO and RCNN variants.
- Contextual choices:
  - Gun in backpack → YOLO.
  - Police officer with gun → RCNN.

Context	Detector variant chosen
Gun inside a backpack	YOLO detector variant
Police officer with a gun	RCNN detector variant

Conclusion

- Proposed process shows potential for supporting pragmatic interoperability.
- Results promising but precision must be improved.
- Future directions:
  - Larger models and tuning.
  - Retrieval-Augmented Generation (RAG).
  - Broader application.

Acknowledgements

This work was supported by:

- **FAPESB**
  (TIC 0002/2015, PIE002/2022)
- **CNPq**
  (4033361/2023-0,310505/2023-2)

Contacts

Daniela B. Claro, Rita S. P. Maciel, Celia Ghedini Ralha
Julio C. S. P. Leite, Roberto de C. Figueiredo

{dclaro, rita.suzana, julioleite, roberedo}@ufba.br
{ghedini}@unb.br

www.formas.ufba.br