POLITECNICO
MILANO 1863

DIPARTIMENTO DI ELETTRONICA
INFORMAZIONE E BIOINGEGNERIA

FM
2026 TOKYO

Towards an Agentic LLM-based Approach to Requirement Formalization from Unstructured Specifications

Motivation and Goals

- From **informal specifications to** NL requirements for derivation of verifiable **formal properties**
- **Flexibility** of LLM
- LLMs can **infer informal semantics**, unlike syntax-based tools
- We propose an **agentic pipeline** for controlled, verification-ready formalization

Envisioned Framework

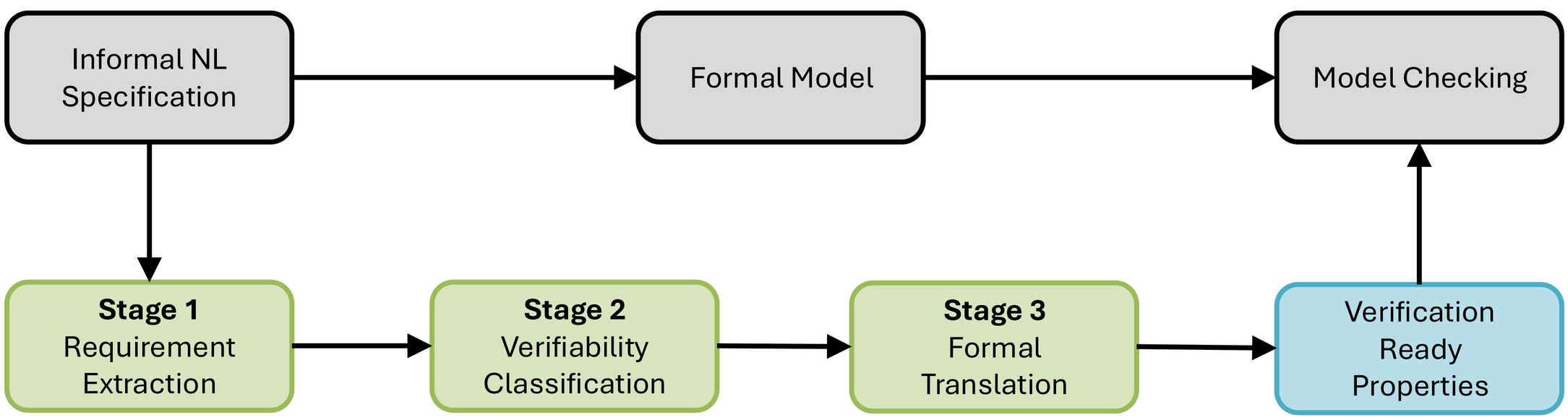
Agentic pipeline:

- Extract requirements
- Filter verifiable ones
- Translate into formal properties

LLM agents are:

- Specialized
- Guided by constraints
- Validated against GT and syntax checker

Output: Verification-ready queries



Stage 1: Requirements Extraction

- **Goal**: Translate informal NL specifications into structured atomic requirements
- **Process**: LLM isolates relevant entities, system dynamics, and safety constraints
- **Output**: Machine-readable JSON with standardized syntax (e.g., "The system must...")
- **Results**: Effectively captures core functional and navigation logic. Achieves an 81.8% match rate (exact + partial) against the GT

Stage 2: Verifiability Classification

- **Goal**: Filter out requirements unsuitable for formal verification
- **Process**: LLM evaluates each requirement against strict model boundaries, assumptions, and observable variables
- **Output**: Binary classification (verifiable: yes/no) with justifications, discards unmodeled variables and behaviors
- **Results**: Reaches 88.7% overall accuracy. High recall (>94%) ensures that valid formal constraints are preserved

Stage 3: Formal Translation

- **Goal**: Bridge the gap between ambiguous natural language and formal rigor
- **Process**: LLM translates validated requirements using mapping rules and a strict BNF grammar
- **Output**: Syntactically correct UPPAAL SMC queries, ready for direct model checking
- **Results**: Highly robust generation with 95.8% of queries compiling flawlessly in UPPAAL. When accounting for logical equivalences (rather than strict string matching), the true semantic accuracy reaches 77.8%

Alberto Tagliaferro, Bruno Guindani,
Livia Lestingi, Matteo Rossi

AI, Proof and Verification (AIPV)
Colocated with FM2026 in Tokyo
May 18-19, 2026

AIPV