NEURO-SYMBOLIC CONVERSATIONAL AI FOR RELIABLE PRODUCTION SERVICE INTELLIGENCE

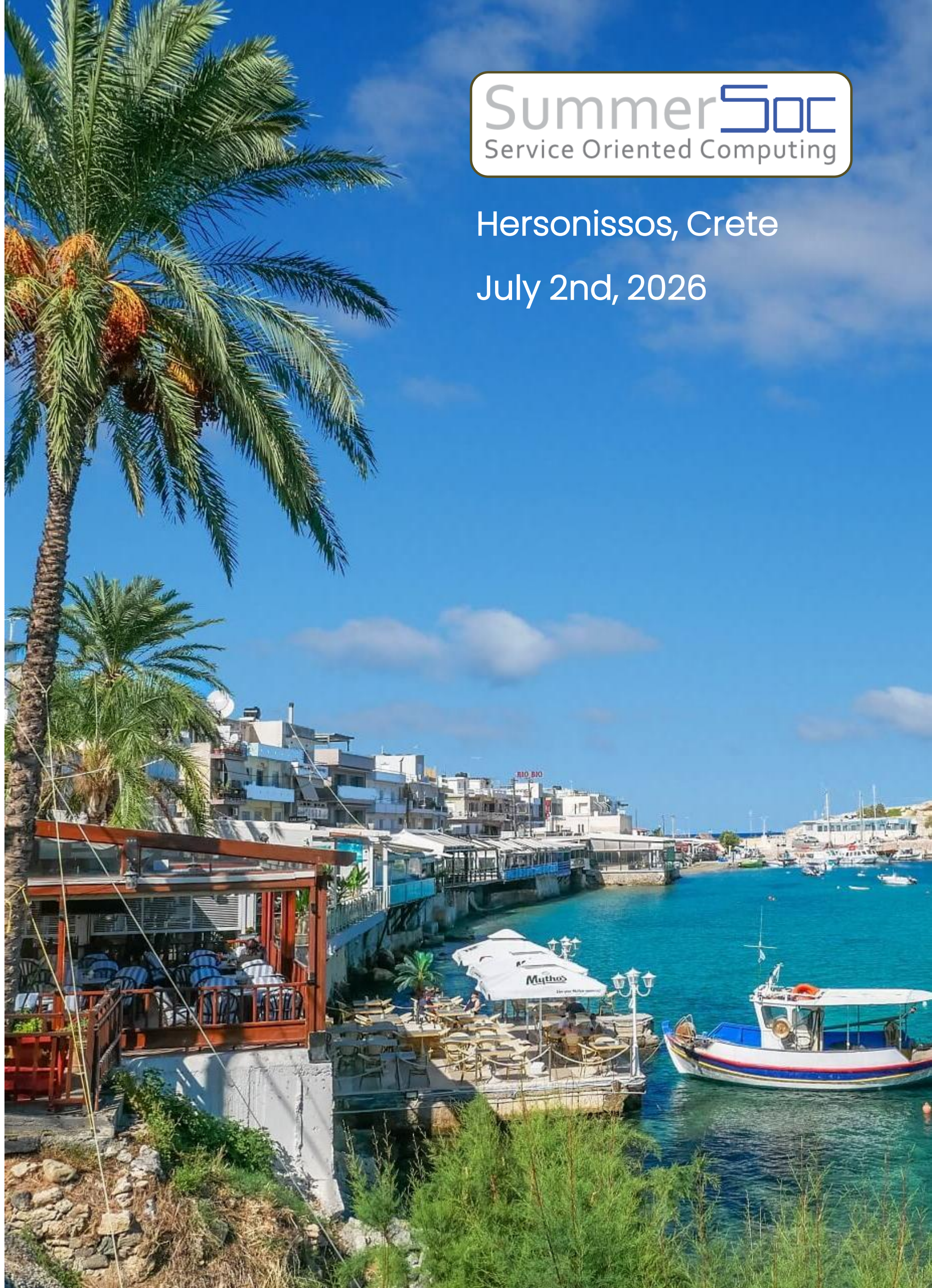
Angelo Casciani and Andrea Marrella

SAPIENZA
UNIVERSITÀ DI ROMA

Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA



SummerSoc
Service Oriented Computing

Hersonissos, Crete

July 2nd, 2026

CYBER-PHYSICAL PRODUCTION SYSTEMS

Industry 4.0 and Internet-of-Things (IoT)

- Shop-floor assets are continuously monitored and controlled by a distributed ecosystem of components, reshaping manufacturing into **Cyber-Physical Production Systems** (CPPSs)
- Organizations want to **take advantage** of continuous data streams

Production System
Software, hardware, robots, human resources, and business rules

\+

Production Process
Activities turning raw materials into finished goods

\+

Digital Models
Derived from IoT data to monitor production processes

Rossit, D.A., Tohmé, F., Frutos, M.: Production planning and scheduling in cyber-physical production systems: a review. Int. J. Comp. Int. Manuf. 32 (2019)
Agostinelli, S., Benvenuti, D., Casciani, A., De Luzi, F., Marinacci, M., Marrella, A., Rossi, J.: A Context-Aware Framework to Support Decision-Making in Production Planning. In: CAiSE 2024. (2024)

PRODUCTION SERVICE INTELLIGENCE (PSI)

Objectives

- Combine production data and digital models to **enhance decision-making**
- Foster human-production system **collaboration**

PSI tools as **cloud-native microservices**, accessed via **standardized APIs**

SIMULATION	MODEL CHECKING	PROCESS MINING
Predictive	Compliance	Descriptive

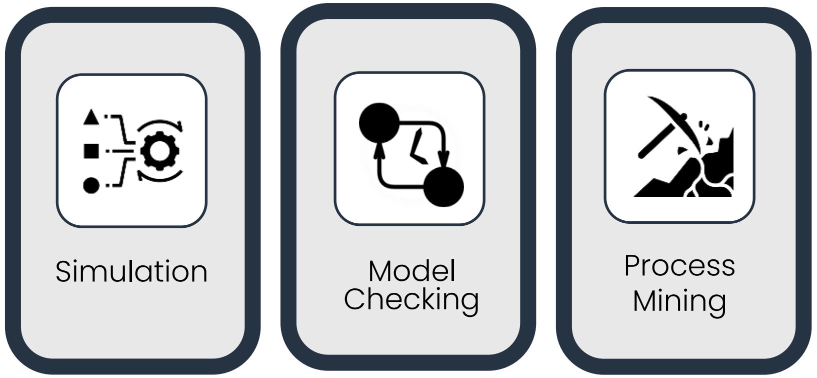
Agostinelli, S., Benvenuti, D., Casciani, A., De Luzi, F., Marinacci, M., Marrella, A., Rossi, J.: A Context-Aware Framework to Support Decision-Making in Production Planning. In: CAiSE 2024 (2024)
Keskin, Z., Joosten, D., Klasen, N., Huber, M., et al.: LLM Enhanced Human–Machine Interaction for Adaptive Decision-Making in Dynamic Manufacturing Process Environments, IEEE Access 13 (2025)

PPI: ARE WE REALLY OKAY WITH THIS?

The "Silo" Problem

- **Fragmented** interfaces
- **Technical expertise** required

➢ Each PSI tool: **different technologies, formats, languages**



Use Case

*"As a **plant manager**, I want to combine compliance checks with what-if analysis of the production so that I can accommodate a new order."*

Keskin, Z., Joosten, D., Klasen, N., Huber, M., et al.: LLM Enhanced Human–Machine Interaction for Adaptive Decision-Making in Dynamic Manufacturing Process Environments, IEEE Access 13 (2025)
Frigerio, N., Tan, B., Matta, A.: Simultaneous control of multiple machines for energy efficiency: a simulation-based approach, Int. J. Prod. Res. 62 (3) (2024)
Agostinelli, S., Benvenuti, D., Casciani, A., De Luzi, F., Marinacci, M., Marrella, A., Rossi, J.: A Context-Aware Framework to Support Decision-Making in Production Planning. In: CAiSE 2024 (2024)

MULTI-PERSPECTIVE ANALYSIS

Manual Integration is required!

➢ User → Manual interaction with multiple tools → Manual integration of the results

Example: A massive new order arrives.

Expert A mines historical data for station utilization

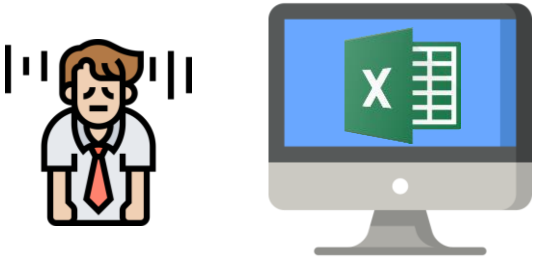
Expert B runs a simulation for the new schedule

Expert C verifies compliance constraints

Process Insights

KPIs

Boolean Result

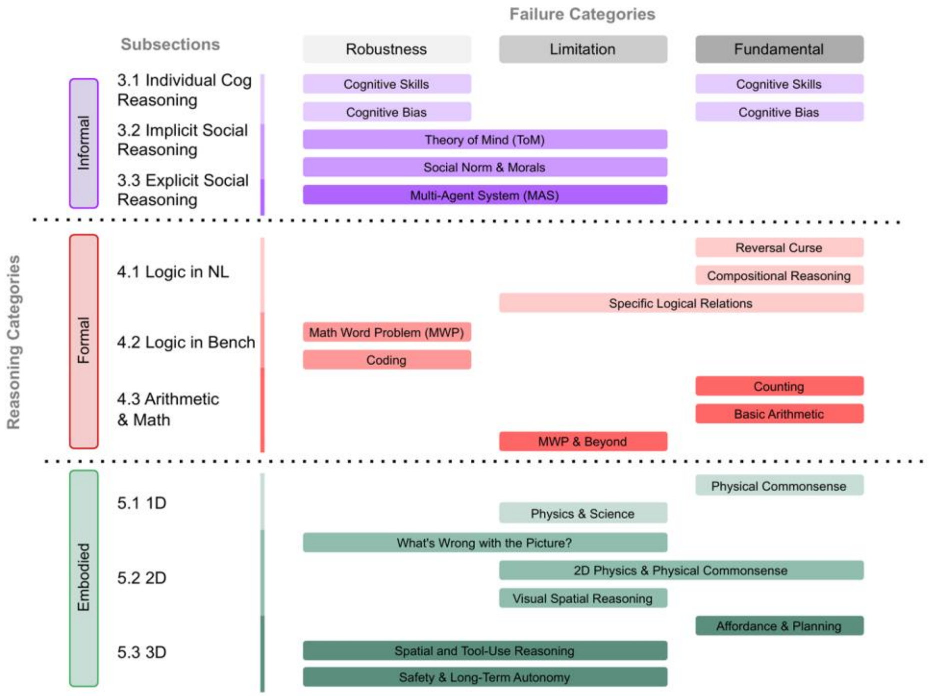


Manual, slow, and error-prone decision-making

WHY NOT JUST USE LLMS?

LLMs are probabilistic retrievers, not inference engines

- Models do not “plan" but perform **approximate retrieval of plan heuristics**
- Effectiveness **declines** when questions are **paraphrased**
- **Chain-of-Thought** traces are learned structural patterns. Answers may follow flawed/incorrect "reasoning" traces
- LLMs **lack self-verification** mechanisms

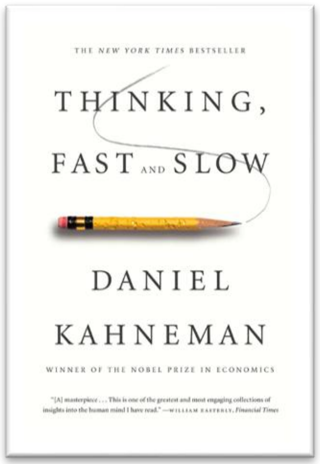


Hallucinated downtime or safety property in CPPSs → Unacceptable!

Kambhampati, S., Valmeekam, K., Guan, L., Verma, M., et al.: Position: LLMs Can’t Plan, But Can Help Planning in LLM-Modulo Frameworks, ICML 2024 (2024)
Lunardi, R., Mea, V. D. , Mizzaro, S., Roitero, K.: On Robustness and Reliability of Benchmark-based Evaluation of LLMs. ECAI 2025 (2025)
Stechly, K., Valmeekam, K., Kambhampati, S.: On the Self-Verification Limitations of Large Language Models on Reasoning and Planning Tasks, ICLR 2025 (2025)
Song, P., Han, P., Goodman, N.: Large Language Model Reasoning Failures, Trans. Mach. Learn. Res. (2026)

NEURO-SYMBOLIC INTELLIGENCE

SYSTEM 1	SYSTEM 2
Fast, intuitive, data-driven	*Slow, deliberative, reason-based*
• Good at learning, poor at reasoning • Data-hungry • Uninterpretable • Prone to hallucinations • Scalable • Grounded in language use	• Good at reasoning, poor at learning • Data-efficient • Interpretable • Hallucination-free • Hand-crafted • Grounded in facts
Generative AI	**Symbolic AI**

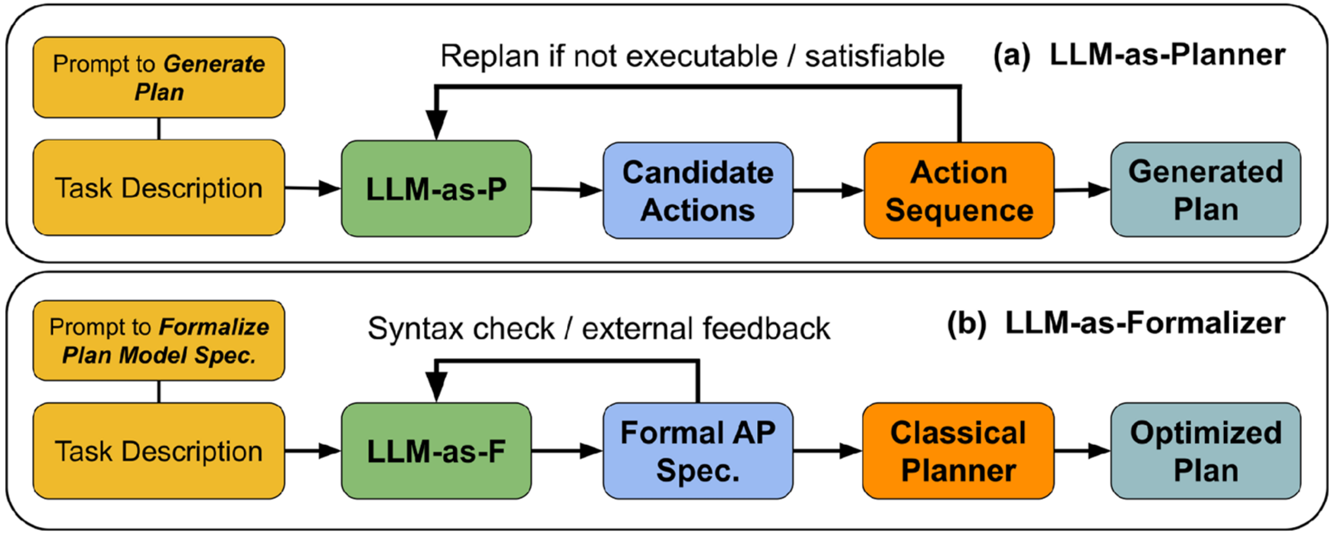


Kahneman, D.: Thinking, fast and slow. Penguin (2012)

The human mind combines **both**.

If we can figure out how to combine the best of both worlds, we can create a conversational system that is **interpretable**, **verifiable**, yet **responsive to shop-floor data streams.**

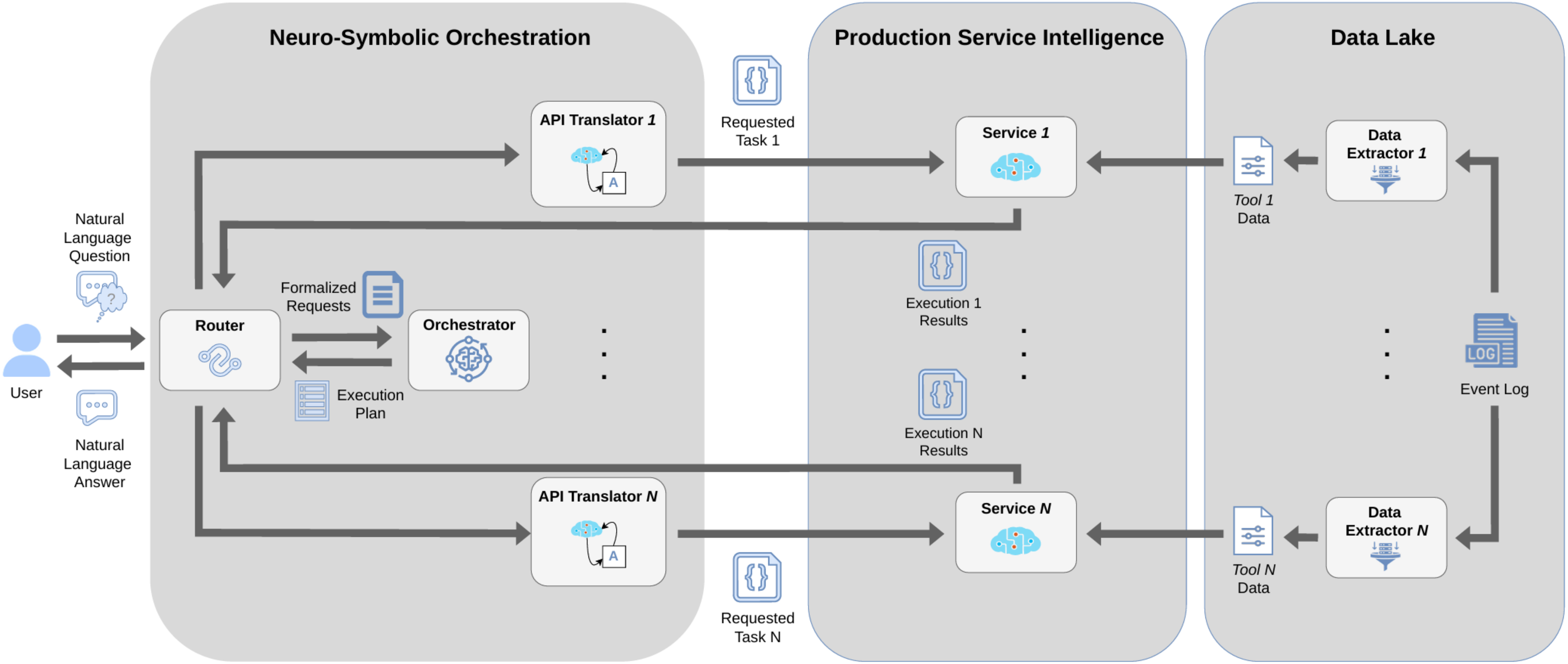
LLM SHIFT: FROM SOLVER TO FORMALIZER



APPROACH	LLM-AS-PLANNER	LLM-AS-FORMALIZER
Role	Generates plan	Translates to PDDL
Execution	Direct	Delegated to symbolic planner
Reliability	Struggles	Sound

Tantakoun, M., Muise, C., Zhu, X.: LLMs as Planning Formalizers: A survey for leveraging Large Language Models to construct automated planning models. In: Findings of the ACL 2025. (2025)
Stechly, K., Valmeekam, K., Kambhampati, S.: On the self-verification limitations of Large Language Models on reasoning and planning tasks. In: ICLR 2025 (2025)

NEURO-SYMBOLIC ARCHITECTURE FOR SERVICE ORCHESTRATION

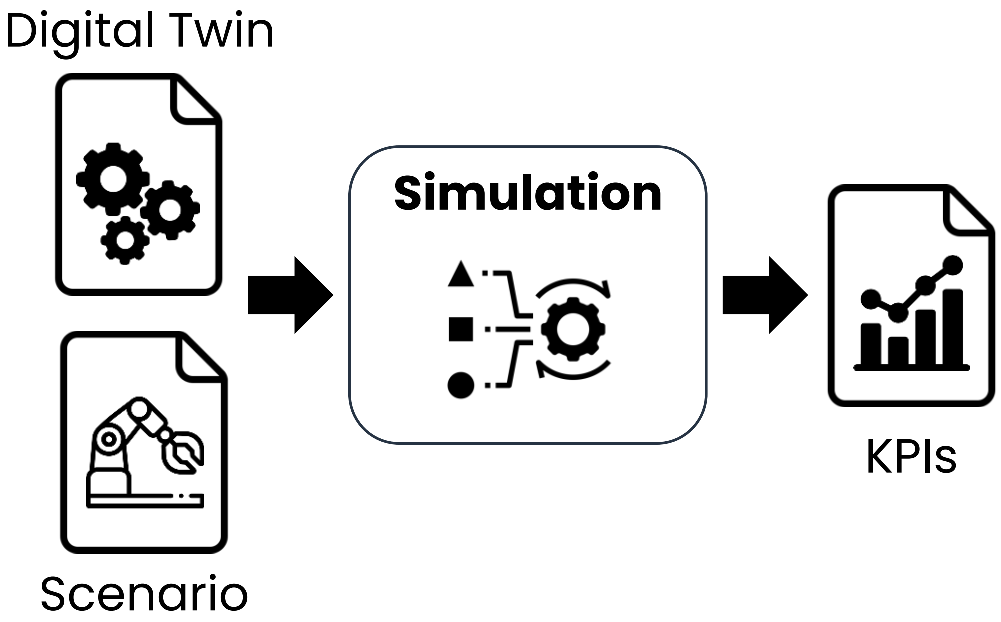


SAPIENZA
UNIVERSITÀ DI ROMA

SIMULATION AS A SERVICE

- *What-if* analysis without disrupting operations
- Implementation: SimPy

Matloff, N.: Introduction to Discrete-Event Simulation and the SimPy Language, University of California at Davis. 2 (2009)
Zhang, L., Zhou, L., Ren, L., Laili, Y.: Modeling and simulation in intelligent manufacturing. Computers in Industry 112, (2019)



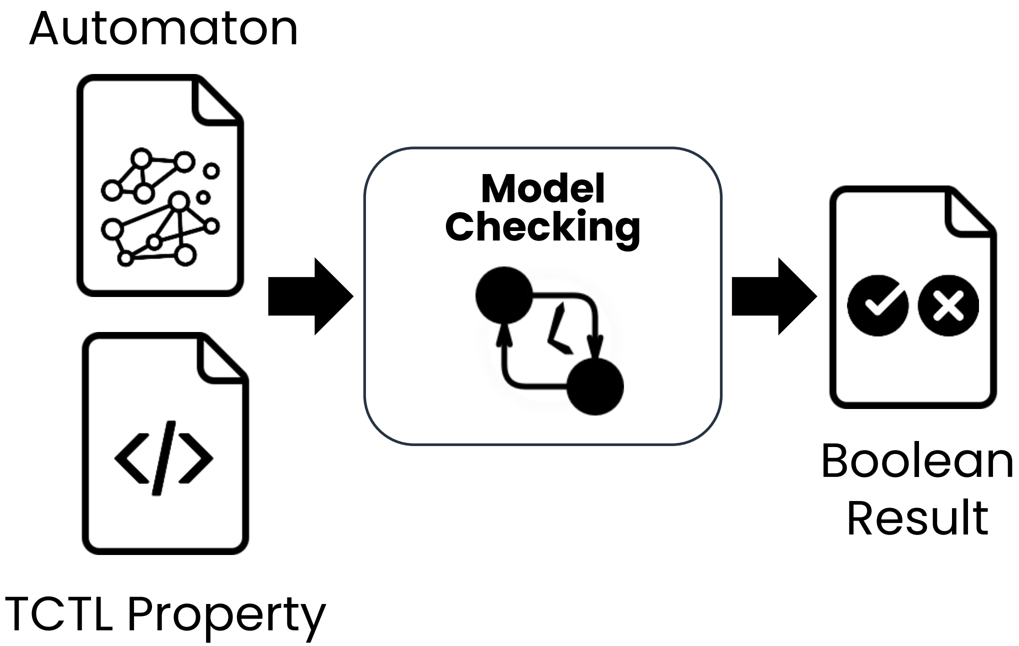
Use Case

*“As a **production manager**, I want to establish a performance baseline and assess the impact of a machine breakdown so that I can plan the production.”*

MODEL CHECKING AS A SERVICE

- **Safety guarantees** via model checking
- Implementation: **UPPAAL**

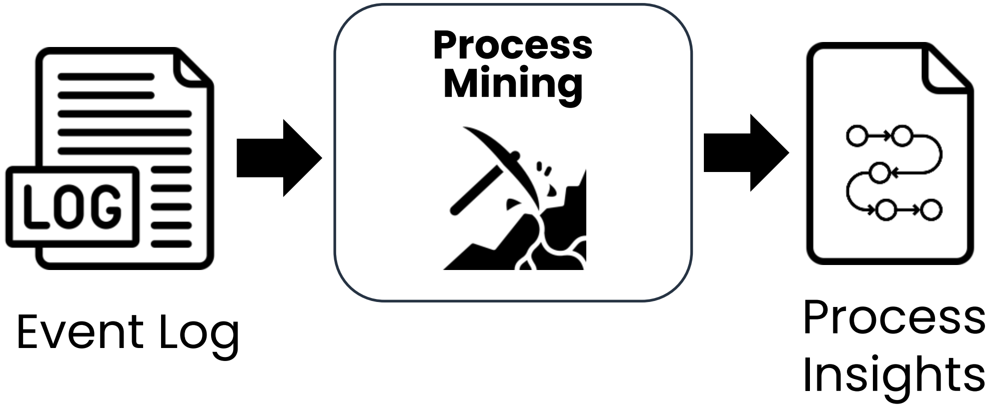
Larsen, K.G., Pettersson, P., Yi, W.: UPPAAL in a nutshell. Int. J. Softw. Tools Technol. Transf. 1 (1997)



Use Case

"As a ***safety engineer****, I want to understand the automaton semantics, verify global safety, and check temporal behaviors so that I can audit the production system."*

PROCESS MINING AS A SERVICE



Process Intelligence Solutions. PM4Py. Web. Jun, 2026

- Operates on **event logs** from shop floor
- Implementation: **PM4Py**

Tasks:

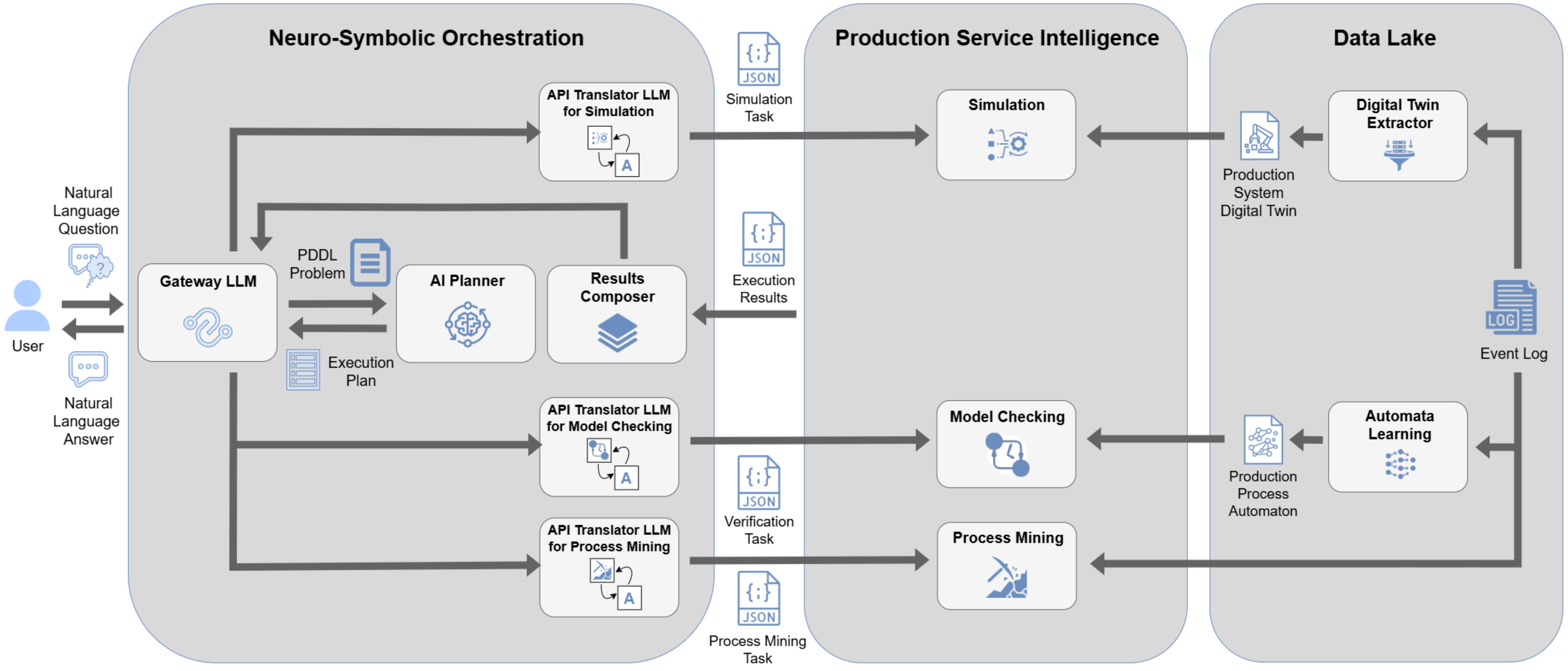
- **Process Discovery** → Petri Net
- **Conformance Checking** → Deviations
- **Time Filtering** → Filtered Event Log
- **Performance Analysis** → KPIs, Variants

Use Case

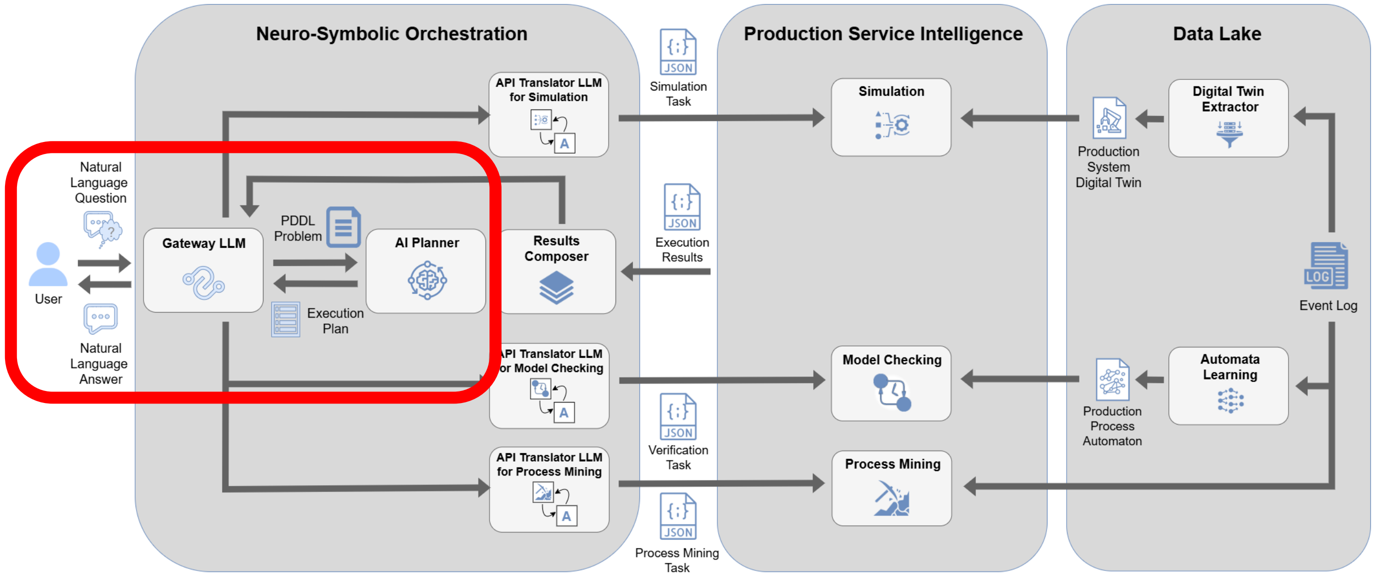
*"As a **process analyst**, I want to discover the Petri net from the event log and check if reality matches the model so that I can analyze historical executions."*

INSTANTIATION

GitHub



INTENT ROUTING AND SERVICE COMPOSITION VIA PLANNING



Intent Routing:

- Analyzes textual query → identifies **intents**

Routing logic:

- Single-service → Direct to Translator
- Multi-perspective → Forward to Orchestrator

Orchestration:

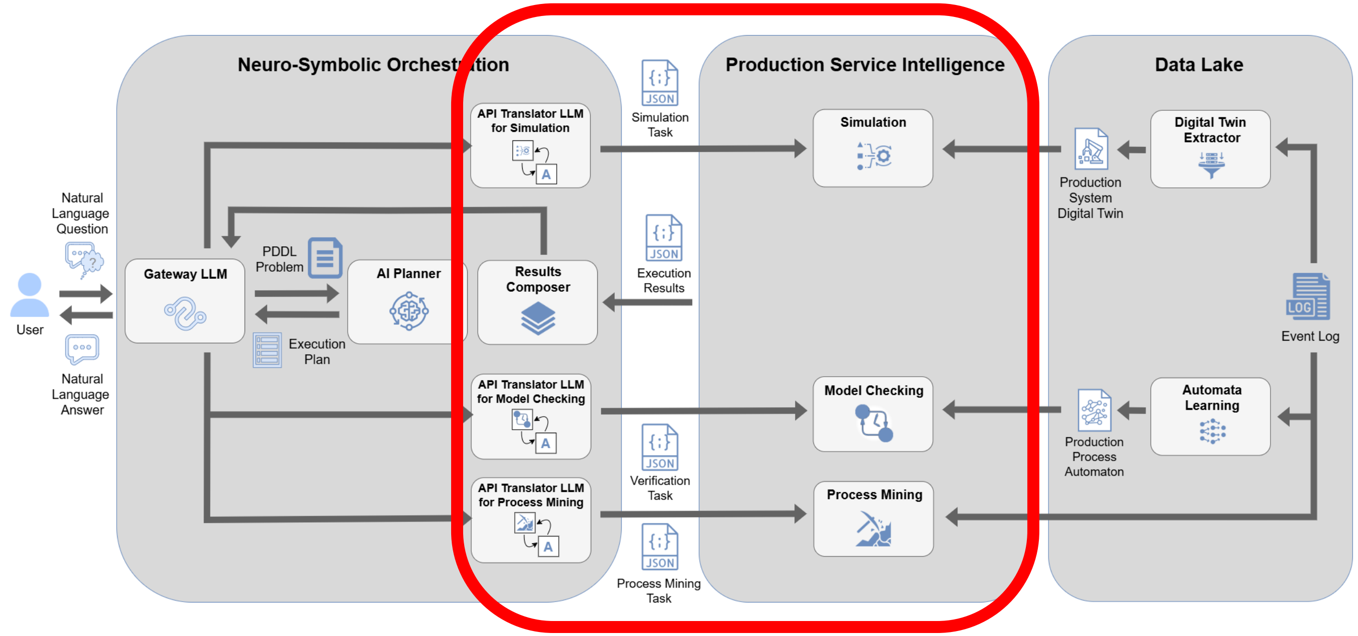
- LLM translates intents in **PDDL goals**
- Planner (Fast Downward) → **sound plan**

Why not let LLM plan?

➢ Prevent hallucinations in service composition

Sapienza Università di Roma

API TRANSLATION & RESPONSE



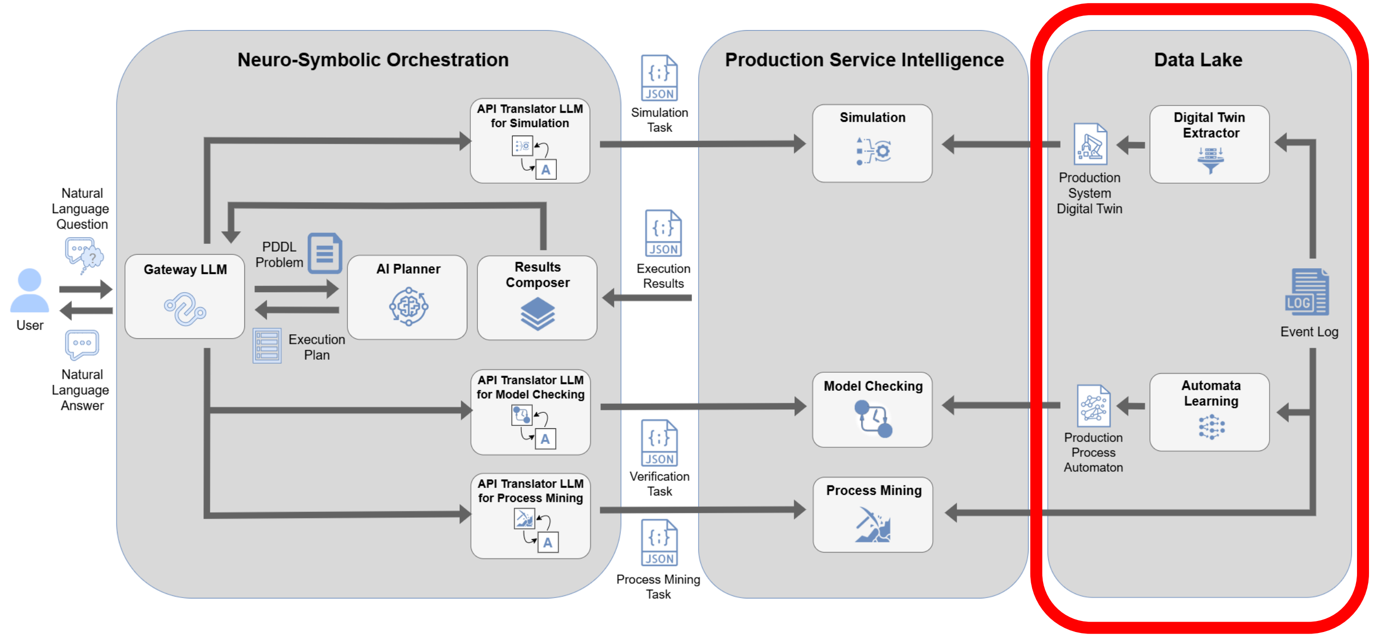
- **Specialized translator LLM** for each PSI tool (avoids interference)
- Planned actions → Encoded JSON payload → Service execution → Result

Soundness Guarantee

➢ Computation occurs **outside LLMs → Reliable results**

Gupta, A., Sheth, I., Raina, V., Gales, M.J.F., Fritz, M.: LLM task interference: An initial study on the impact of task-switch in conversational history. In: Proc. of the Conf. on Empirical Methods in NLP, EMNLP 2024. ACL (2024)

DATA LAKE

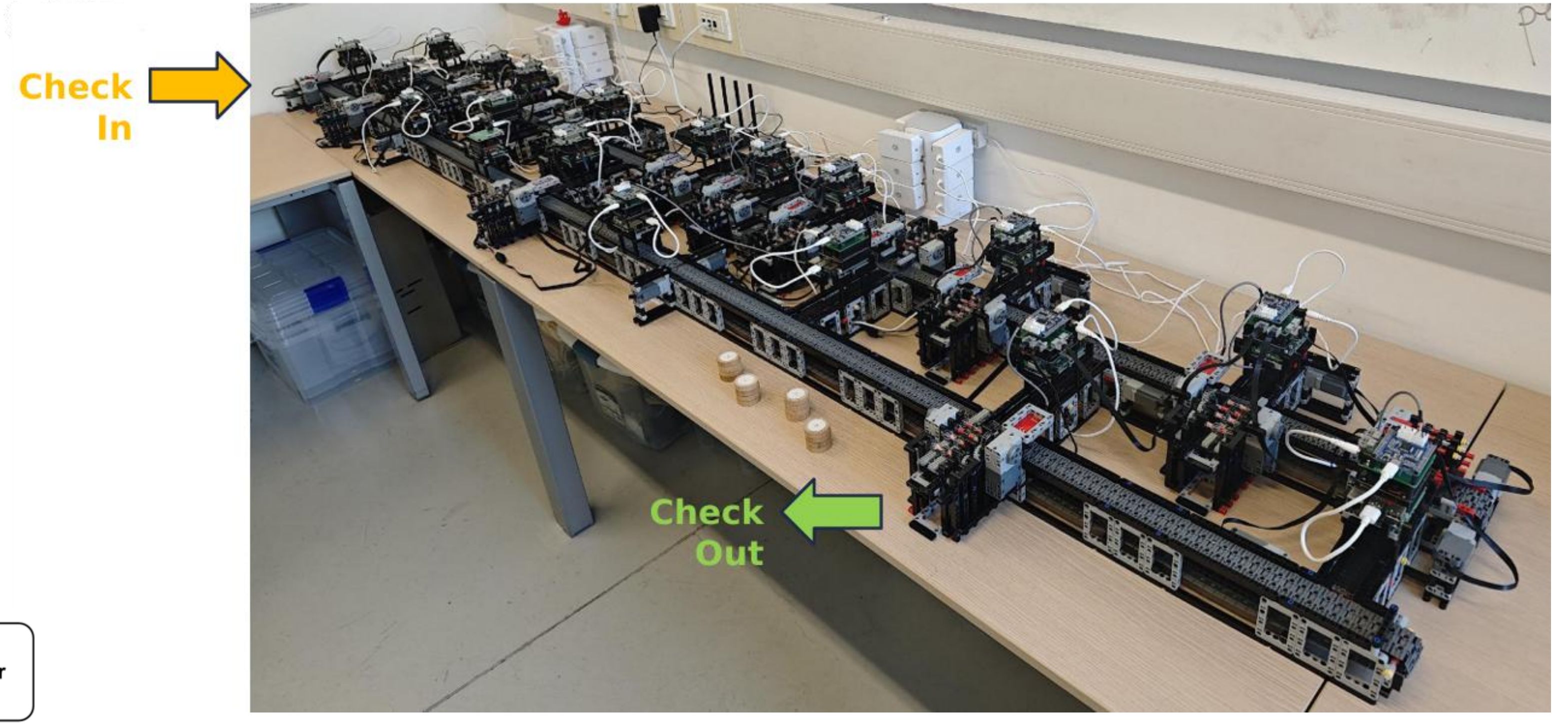


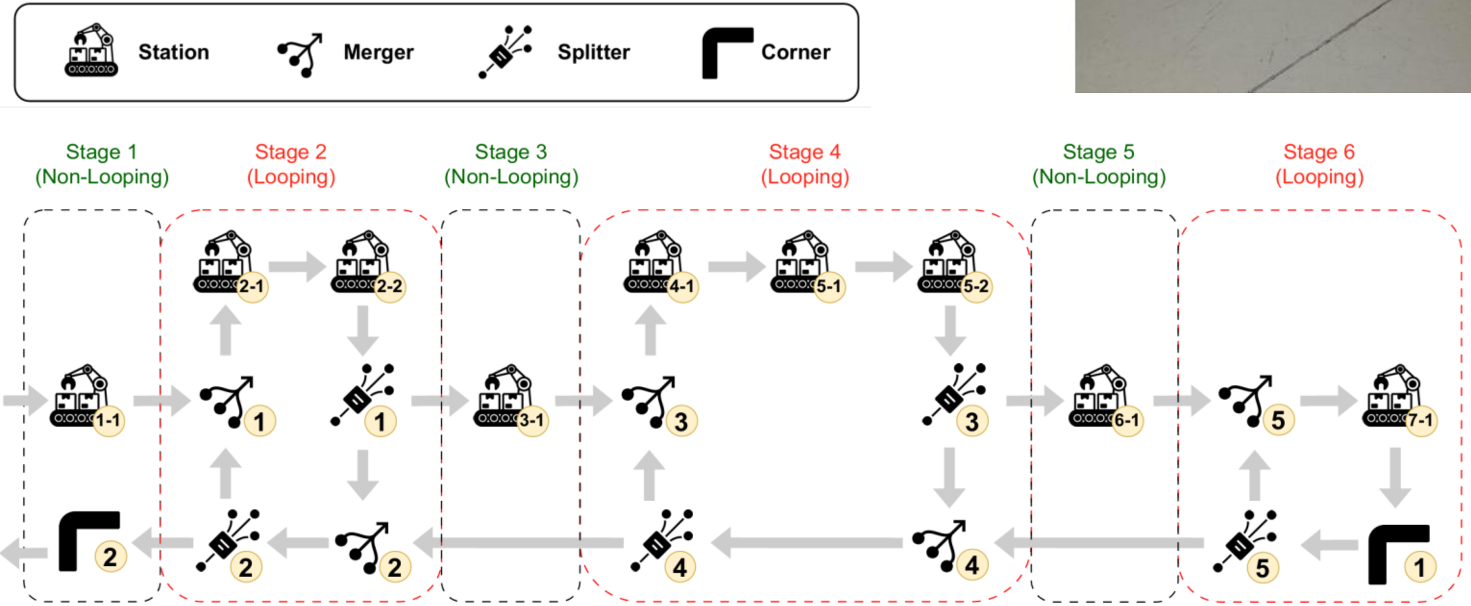
- **Event log** from shop-floor sensors
- **Automated models extraction** (no manual modeling)

EXTRACTOR	OUTPUT	USED BY
Digital Twin	Scenario Parameters	Simulation
Automata Learning	Probabilistic Timed Automaton	Model Checking
-	Raw Event Log	Process Mining

SAPIENZA UNIVERSITÀ DI ROMA

CASE STUDY: LEGO FACTORY





MOTOWN: http://www.open.diag.uniroma1.it/project_detail/28097. Web. Jul, 2026

PRELIMINARY EVALUATION SETUP

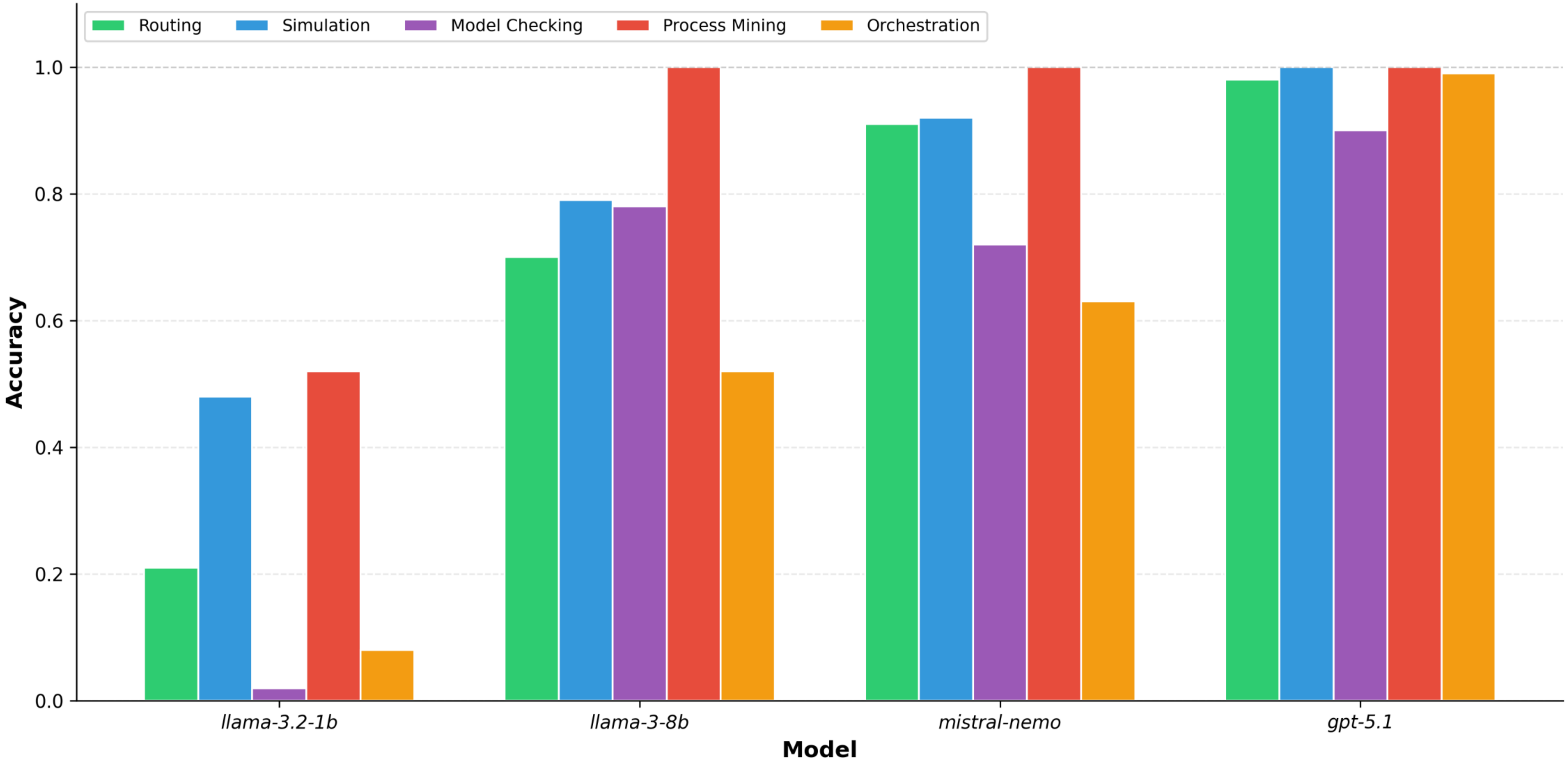
Dataset Composition

- **1.5k** <natural-language request, expected answer> pairs (300 samples/task)
- **Five tasks**:
  - *routing* of a request; and
  - *API translation* for *simulation*, *model checking*, *process mining*, and *multi-perspective orchestration*.

Performance Metric

- We measured **accuracy** (#correct LLM-generated outputs/#questions)
  - For routing: **correct** if answer includes expected service label
  - For API translations: **correct** if answer is a valid JSON that matches syntax and parameters for the target service

RESULTS


Routing
Simulation
Model Checking
Process Mining
Orchestration
1.0
0.8
0.6
0.4
0.2
0.0
Accuracy
llama-3.2-1b
llama-3-8b
mistral-nemo
gpt-5.1
Model

THREATS TO VALIDITY

Internal Validity

LLM Stochasticity

- *Temperature = 0*
- Performed 5 *runs/test*

Prompt Quality

- *In-context learning*
- *Few-shot* prompting

External Validity

Domain Bias

- Domain-independent architecture

Model Diversity

- Benchmarked 4 LLMs to avoid single-provider bias

Reproducibility

- Documented model versions against risks of commercial API updates/deprecation

Construct Validity

Metric Focus

- Accuracy used to validate encoding and orchestration

System Boundaries

- Utility is bounded by the capabilities of PSI layer

Data Grounding

- Logical consistency is guaranteed
- Reliability of PSI results depends on data quality

CONCLUSION

A Conversational Methodology for Production Service Intelligence

- Introduced a **domain-agnostic neuro-symbolic conversational AI framework** for reliable, multi-perspective PSI on CPPSs.
- Transitioned LLMs from **flawed PSI executors to flexible formalizers**:
  - the LLM encodes and redirect;
  - the symbolic AI planner handle orchestration and services perform computations.

Future Work

- Fine-tuning to **mitigate encoding hallucinations** and **streamline the modeling effort**
- Evaluation with industrial practitioners
- Real-time IoT stream processing and state-aware service composition

SummerSoc
Service Oriented Computing

Hersonissos, Crete

July 2nd, 2026

THANK YOU

Presented by Angelo Casciani

Mail: casciani@diag.uniroma1.it

GitHub

About Me

SAPIENZA
UNIVERSITÀ DI ROMA

Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA