Proceedings of Machine Learning Research vol 284:1–2, 2025 19th Conference on Neurosymbolic Learning and Reasoning

Neurosymbolic AI to Support Unstructured Data Processing in Process and Automation Engineering

Nicolai Schoch NICOLAI.SCHOCH@DE.ABB.COM
and **Mario Hoernicke** MARIO.HOERNICKE@DE.ABB.COM
Kallstadter Str. 1, 68309 Mannheim, Germany

Editors: Leilani H. Gilpin, Eleonora Giunchiglia, Pascal Hitzler, and Emile van Krieken

Abstract

Neurosymbolic AI can be an important enabler for AI-enhanced support of the unstructured data processing for process and automation engineering. Large subsymbolic models can shine with their semantic understanding of the contents of unstructured data sources, and symbolic representation can help to guide and guardrail the processing, so overall confidence in the generated outputs is reasonably high. With the *Engineering Data Funnel*, we present a neurosymbolic AI system, which uses large subsymbolic language and vision models to process unstructured engineering project data, while exploiting at the same time structured knowledge representation in engineering domain ontologies and thereon-based symbolic reasoning capabilities in order to provide reliable output for the deterministic further processing in traditional, established, mostly non-AI-based engineering tools.

1. Industry Context and Problem Statement

In the field of process and automation engineering, a wide range of software applications is already available. However, one of the biggest current challenges lies in transforming unstructured and uncontrolled multi-modal data — such as Piping and Instrumentation Diagrams (P&IDs) or Control Narratives (CNs) — into a structured, consistent, and contradiction-free format that can be directly processed by these software applications. Today, this transformation still most often requires manual efforts from experts to translate these unstructured formats into formalized, target-system-compliant representations (Hoernicke et al., 2022). This manual process is extremely time-consuming and prone to errors, even for domain experts, due to frequent updates to the original documents and to potential semantic and content-related ambiguities.

2. Neuro-Symbolic AI for the Digitization of Unstructured Multi-Modal Data in Process and Automation Engineering

At the ABB Research Center in Germany, we are therefore exploring neurosymbolic AI, which combines the strengths of knowledge-based, domain-specific ontologies with the capabilities of neural machine learning models, such as large language models (LLMs) and vision models. We present our prototype, the *Engineering Data Funnel* (EDF), a neurosymbolic AI system that converts and integrates unstructured process design information (e.g., in PDF, Excel, or Word formats) into structured formats. This enables established engineering software applications to access and utilize this data, supporting the long-envisioned goal of the "automation of automation" (see Figure 1).

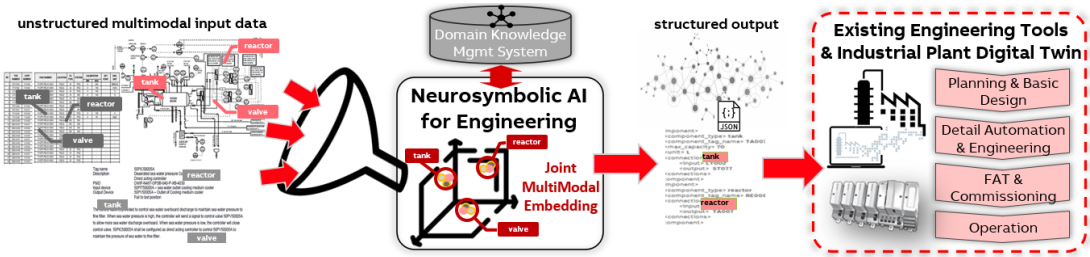

Figure 1: The Engineering Data Funnel to process unstructured multi-modal engineering design data into a structured representation (for other existing engineering tools and for the Digital Twin of the future industrial plant).

The EDF is set up as a multi-agent-system (MAS) with a group of LLM-based agents having access to a set of specialized expert ML models and tools, to thus autonomously orchestrate the processing of multi-modal data. Each expert model, such as an LLM for processing CNs, or a vision model for interpreting P&IDs, is guided and guardrailed by an engineering domain-specific ontology (Schoch et al., 2025). This ensures that domain knowledge is incorporated and respected throughout the data processing pipeline. To ensure consistency not only across all input data but also with respect to the target output format, the EDF additionally performs an alignment and transformation step, which maps uncontrolled data to controlled artifacts within our knowledge model (Schoch et al., 2023), and which thus enables the deterministic downstream processing.

3. EDF Prototype Evaluation, Discussion, and Conclusion

We will present our approach and the prototypic EDF system, demonstrating its strengths and limitations through concrete examples from the Oil & Gas, Chemicals & Pharma, and Mining industry sectors. Our proof-of-concept evaluation already now shows very satisfying and promising results, and one can expect huge time-savings through substantial automation of this so-far-manual effort of unstructured engineering design data processing. We will also provide an outlook on ongoing work to further develop the EDF and to make it more robust.

References

M. Hoernicke, K. Stark, N. Schoch, A. Markaj, and A. Fay. Modular Engineering of Conventional Plants - Using MTP for world-scale industry plants. *atp magazin*, 63(4), 2022.

N. Schoch, M. Hoernicke, and K. Stark. Semantic Facilitation and Integration Layer for Process and Automation Engineering. *Proceedings of IEEE ETFA*, 2023.

N. Schoch, M. Elsheikh, M. Hoernicke, N. Strem, K. Stark, S. Palacio, and V. Ashiwal. The Engineering Data Funnel: Knowledge-Enhanced, Agentic-AI-Based Data Processing for Automation Engineering. *Proceedings of IFAC J3C EAAS Workshop*, 2025. (Accepted).