

UDC 004.9

*Pototskyi Maksym, Postgraduate Student,  
Gamayun Igor, D.Sc., Professor,  
Kopp Andrii, Ph.D., Associate Professor  
National Technical University «Kharkiv Polytechnic Institute»*

## **FRAMEWORK FOR TESTING SMALL LANGUAGE MODELS ABILITY TO ASSESS BPMN QUALITY**

Business process modeling using BPMN (Business Process Model and Notation) is a fundamental technique for analyzing and designing information systems. The quality of such models affects the correctness of decisions and the efficiency of workflows. Building BPMN models requires knowledge of the domain and the modeling rules. Recent studies show that large language models are capable of generating BPMN from natural language [1]. At the same time, evaluating the quality of such models remains an open problem. Existing approaches often do not consider a full set of quality criteria or rely on the language models themselves as evaluators [1]. The paper BEF4LLM proposes a multidimensional evaluation that includes syntactic, semantic, and pragmatic aspects, as well as validity [1]. Other studies [2] explore intelligent approaches to BPMN evaluation based on model element features. However, the role of Small Language Models (SLM) [3] in this problem has been scarcely investigated. The aim of this work is to develop a framework for testing the ability of small language models to evaluate BPMN quality.

The proposed framework (Fig. 1) is based on the step-by-step processing of a BPMN model and the generation of a structured representation for analysis. In the first step, the model in XML format is broken down into elements. Tasks, events, and gates are identified. A set of features is generated for each element. These features represent both structural properties and textual labels. Next, a generalized representation of the model is created in the form of a TOON (Token-Oriented Object Notation) [4], which is compact and suitable for processing by language models. Small models have limited resources, so it is important to provide them with data in a compressed form. This reduces computational costs and ensures consistent results.

Then each SLM receives the same TOON-based representation and the system prompt. The evaluation is focusing on identifying business process model tasks following the rule [1]:

$$|in(t) = 1| \wedge |out(t) = 1|.$$

Here  $in(t)$  and  $out(t)$  are sets of incoming and outgoing sequence flows of a task  $t \in T$ , while  $T$  is a set of tasks defined in the BPMN model.

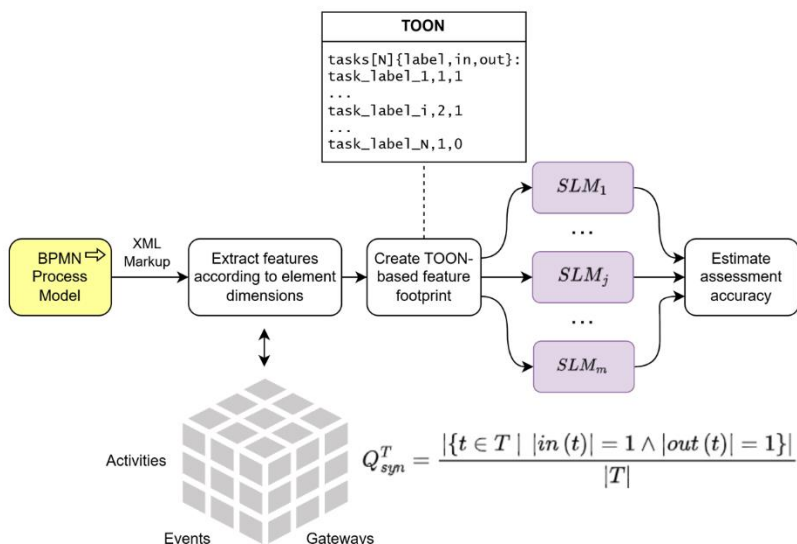


Fig. 1. Proposed framework for testing SLMs ability to assess  $Q_{syn}^T$

The syntactic quality of tasks  $Q_{syn}^T$  is computed by dividing the number of tasks following the rule, introduced above, by the total number of tasks. Then estimates obtained using SLMs are compared with the reference value calculated according to the formalized BPMN modeling rules [1].

### References:

1. Lauer C., Pfeiffer P., Rombach A., Mehdiyev N. Assessing the Business Process Modeling Competences of Large Language Models [Electronic resource]. – Access: <https://arxiv.org/abs/2601.21787>
2. Yanholenko O., Kopp A., Godlevskiy M., Orlovskiy D., Khatsko N. Towards information technology for BPMN models quality assessment based on intelligence theory [Electronic resource]. – Access: <https://ceur-ws.org/Vol-4004/paper7.pdf>
3. Johnson J. Small Language Models (SLM): A Comprehensive Overview [Electronic resource] / J. Johnson // Huggingface.co. – 2025. – Access: <https://huggingface.co/blog/jjokah/small-language-model>
4. What Is TOON: the Compact Data Format for LLMs and AI [Electronic resource]. – 2023. – Access: <https://openapi.com/blog/what-the-toon-format-is-token-oriented-object-notation>