**Description of the CSV File Structure for BPMN-VQA Dataset**

| Column Name | Description |
| --- | --- |
| Image\_Id | Unique identifier of the BPMN model, extracted from the filename (e.g., 159373). |
| Model\_Name | The name of the BPMN model, extracted from the .meta.json file (e.g., "Volley Tennis Club"). |
| Lane | The name of the lane in the BPMN diagram (e.g., "Sales"). |
| Lane\_Position | The bounding box coordinates of the lane in the diagram, consisting of upperLeft and lowerRight positions. |
| Task | List of tasks associated with the lane (e.g., ["Verify Sales Quote", "Review Pricing Options"]). |
| Task\_Position | The bounding box coordinates of each task, used for spatial representation. |
| Modeling\_Language | The BPMN version used in the model (bpmn11 or bpmn20). |
| Model\_Path | The file path to the BPMN JSON file for reference. |
| Question | The generated question related to the lane and its content. |
| Answer | The corresponding answer extracted from the BPMN model data. |

Generated Questions in the BPMN-VQA Dataset

Each lane in the BPMN model is assigned four structured questions to simulate real-world process-related queries.

| Question Format | Example | Answer Format |
| --- | --- | --- |
| What is the role of "Lane Name"? | *What is the role of "Sales"?* | List of tasks inside the lane (e.g., "Verify Sales Quote, Review Pricing Options"). |
| What is the main responsibility of "Lane Name" in this BPMN model? | *What is the main responsibility of "Sales" in this BPMN model?* | Reworded task summary (e.g., "This lane is responsible for Verify Sales Quote and Review Pricing Options"). |
| Does "Lane Name" have a decision-making step (gateway)? | *Does "Sales" have a decision-making step (gateway)?* | "Yes" or "No", depending on the presence of an Exclusive Gateway. |
| Where is "Lane Name" located in the process diagram? | *Where is "Sales" located in the process diagram?* | Other lanes present in the same Pool (e.g., "Engineering, Finance"). |

**Research Applications of the BPMN-VQA Dataset**

Recent advancements in Process Modeling with Large Language Models (LLMs) focus on automatically generating BPMN models from textual descriptions, enabling non-experts to create process diagrams without deep knowledge of BPMN syntax. In contrast, our BPMN-VQA dataset takes the reverse approach, transforming structured BPMN diagrams into natural language question-answer pairs. This inversion of the paradigm has significant implications for semantic process modeling, business process analysis, and AI-driven process automation.

Semantic quality in business process models refers to the alignment of the model with domain knowledge and real-world interpretation. By converting BPMN diagrams into structured questions and answers, the BPMN-VQA dataset:

* Enhances process model interpretability by explicitly capturing roles, responsibilities, and workflow logic.
* Supports semantic validation, ensuring that the BPMN models align with intended business logic.

One of the biggest challenges in AI-generated BPMN models is the lack of explainability—models generated from text may have errors, misinterpretations, or missing elements. The BPMN-VQA dataset:

* Enables LLM-based process analysis by providing structured explanations of BPMN models.
* Helps benchmark how well AI models understand BPMN elements, tasks, and control-flow semantics.

Many business processes are implicitly embedded in BPMN diagrams without explicit documentation. Our dataset enables automated business process documentation by:

* Generating structured process descriptions from existing BPMN models.
* Supporting knowledge extraction for business analysts, where AI can translate BPMN logic into readable narratives.
* Improving process model quality assurance by ensuring that models reflect domain-specific constraints and best practices.

By transforming BPMN diagrams into structured knowledge, the BPMN-VQA dataset bridges the gap between:

1. AI-based BPMN generation (Text → Model)
2. AI-based BPMN interpretation (Model → Textual Insights)

This dataset complements AI-generated process models by enabling validation, explanation, and semantic quality assurance. Future research can leverage BPMN-VQA for ontology-driven BPMN validation, conversational AI for process analysis, and automated compliance checking in business workflows.