Praktikum Wirtschaftsinformatik
12. November 2025
Prof. Dr. Mirjam Minor
M.Sc. Tolga Tel

GOETHE UNIVERSITÄT FRANKFURT AM MAIN

In this Exercise Sheet, we will start to use Generative AI to describe BPMN models. In contrast to the last exercise sheet, we will create our own small dataset. We will also compare different methods within the same pipeline and use custom components to employ these methods.
You can find all the documentation on https://haystack.deepset.ai/
BPMN files are structured in an XML format. First collections and lanes are defined. Then, each element like tasks, gateways, events and flows are defined. At last, each element is given coordinates for the visual representation. While necessary for the visual representation, the coordinates do not help in the description of the process model.

Exercise 1 (Documents) — 6 Points

a. Choose **two** of the given BPMN models and **write a short descriptive text** containing each BPMN element. (**2 Points**)
b. **Construct a BPMN model with a descriptive text** for the following topics. **(2 Points)**
   - **Too much carry-on baggage because of high occupancy. (https://www.lufthansa.com/de/en/carry-on-baggage?)**
   - **Delayed baggage (https://www.lufthansa.com/de/en/baggage-irregularities?)**
c. Put each BPMN code of the models chosen in a) in a separate .txt file with its descriptive text. The goal is to have a example BPMN file with a corresponding description in your document store.
   In the end of the BPMN files, you will find the coordinates of all elements in the BPMN model. You can remove them in the .txt file to decrease the amount of tokens in your query. (**1 Point**)
d. Initialize a document store and add your embedded .txt files. (**1 Point**)

Exercise 2 (Main RAG-System) — 6 Points

a. **Create** 3 different Prompt Builder instances to instruct the LLM to describe a given BPMN model. **(3 Points)**
   1. A ChatPromptBuilder that receives a document from the retriever and uses role prompting with a Chat Message “from_system” and a Chat Message “from_user”.
   2. A regular PromptBuilder that receives a document from the retriever.
   3. A regular PromptBuilder that does not receive any examples.
b. **Create** 3 generator instances which receive a prompt. **(1 Point)**
c. Write a custom component that receives the outputs from all generators and returns them as plain texts. **(2 Points)**

GOETHE UNIVERSITÄT FRANKFURT AM MAIN

Exercise 3 (Assemble and test the pipeline) **8 + 3* Points**

a. In exercise 1c), we removed the coordinates from the BPMN code manually. Write a function to read a BPMN file and remove the coordinates automatically for the query. *Hint: You can use regular expressions or an XML parser to remove part containing the coordinates*. **(3 Points)**
b. Build a Pipeline containing a retriever instance, all generators and Prompt Builders and run the Pipeline for the remaining BPMN models. **(1 Point)**
c. Precision and Recall are important scores when working with LLMs. For this exercise, we will evaluate the occurrence of Tasks, Gateways and Events of the BPMN model in our descriptive texts.
   - Precision is the share of relevant information in text. If our text contains x elements that are in the BPMN model and y elements that are not in the BPMN model, the Precision is
   $$Precision = \frac{x}{x + y}$$
   - Recall is the share of the information from the BPMN model that is represented in the text. If our text contains x elements that are in the BPMN model, but missed the other z elements that are in the BPMN model, the Recall is
   $$Recall = \frac{x}{x + z}$$

   Choose one of the generators in from exercise 2a) and evaluate its result on the remaining given BPMN models based on Precision and Recall metrics using a single score for tasks, gateways and events. Prepare a presentation with your scores. **(4 Points)**
d. Bonus exercise: Choose a second generator and evaluate its precision and recall scores with your previous results from exercise 3c). **(3* Points)**