

# Writing Prompts for LLM-based Relevance Judgments

**Task Goal:** Your objective is to create prompts that guide a Large Language Model (LLM) to accurately assess the relevance of passages to a query. Each prompt you develop should direct the LLM to evaluate the passage in a way that aligns with the specified objective.

**Instructions:** You are required to complete the code for each of the following three objectives by **writing a prompt within the provided Python function templates**. There are three distinct relevance judgment tasks, and you must create a unique prompt for each one.

For each of the following 3 tasks, **fill in the provided code** structure by **specifying the system and user messages** that will instruct the LLM. Details of the tasks are provided below.

1. **Binary Relevance**
2. **Graded Relevance**
3. **Pairwise Preference**

**Try to use your own wording as much as possible.**

---

## Objective 1: Binary Relevance Decision

**Task:** Complete the function below by writing a prompt that instructs the Large Language Model (LLM) to determine if a given passage is relevant to a specific query. A passage is considered relevant if it satisfies the user's information need.

**Instructions:** You need to construct both a system message and a user message by replacing "YOUR SYSTEM MESSAGE GOES HERE." and "YOUR USER MESSAGE GOES HERE.". Use {query} and {document}. The client parameter is the representative of the LLM. You can ignore it while completing the code.

**Output:** The response should be a binary value, where "1" indicates that the passage is relevant to the query and "0" indicates that it is not relevant.

### Example of the task:

- **Query:** "do goldfish grow"
- **Passage:** ": The conditions goldfish are kept in plus their diet determine how large they will grow. I have seen goldfish grow ridiculously large in very small containers when their water was changed frequently. Goldfish will not grow if water conditions are poor. Fancy goldfish donât grow as large as Common goldfish. A good size would be around 5 inches body length for most fancy varieties, 8 inches for Comets and 12 inches for Common Goldfish. These sizes are usually only attained by pond grown fish."
- **Correct Answer:** 1

```
def binary_relevance_task(query, document, client):
    messages = [
        {
            "role": "system",
            "content": "YOUR SYSTEM MESSAGE GOES HERE."
        },
        {
            "role": "user",
            "content": f"YOUR USER MESSAGE GOES HERE"
        }
    ]

    response = client.chat.completions.create(
        model="gpt-4o",
        messages=messages,
        temperature=0
    )
    return response
```

## Objective 2 - Graded Relevance

**Task:** Complete the function below by writing a prompt that instructs the LLM to score the relevance of a passage to a query using an integer scale from 0 to 3. where:

- **[3] Perfectly relevant:** The passage is dedicated to the query and contains the exact answer.
- **[2] Highly relevant:** The passage has some answer for the query, but the answer may be a bit unclear, or hidden amongst extraneous information.
- **[1] Related:** The passage seems related to the query but does not answer it.
- **[0] Irrelevant:** The passage has nothing to do with the query.

**Instructions:** You need to construct both a system message and a user message by replacing "YOUR SYSTEM MESSAGE GOES HERE." and "YOUR USER MESSAGE GOES HERE.".

Use `{query}` and `{document}`. The `client` parameter is the representative of the LLM. You can ignore it while completing the code.

**Output:** The LLM's response should be a single integer value between 0 and 3, with no further explanation. This integer represents the passage's relevance to the query.

### Examples:

Query: "do goldfish grow"	
<b>Passage:</b> "The conditions goldfish are kept in plus their diet determine how large they will grow. I have seen goldfish grow ridiculously large in very small containers when their water was changed frequently. Goldfish will not grow if water conditions are poor. Fancy goldfish donât grow as large as Common goldfish. A good size would be around 5 inches body length for most fancy varieties, 8 inches for Comets and 12 inches for Common Goldfish. These sizes are usually only attained by pond grown fish."	<b>Passage:</b> "Actually, it is humane to keep a goldfish in a bowl, as long as the water is good, and as long as the fish can be relocated to a pond or large habitat when it becomes mature, as in too big the live in the bowl. It is not automatically cruel to keep one in a bowl."
<b>Relevance Level:</b> 3 Explanation: The answer is perfectly relevant	<b>Relevance Level:</b> 1 Explanation: The answer is related but not relevant

Please note that the explanation provided is solely to help you understand the concept of different levels of relevance. There is no need to include this explanation in the output.

```
def graded_relevance_task(query, document, client):  
    messages = [  
        {  
            "role": "system",  
            "content": "YOUR SYSTEM MESSAGE GOES HERE."  
        },  
        {  
            "role": "user",  
            "content": f"YOUR USER MESSAGE GOES HERE"  
        }  
    ]  
  
    response = client.chat.completions.create(  
        model="gpt-4o",  
        messages=messages,  
        temperature=0)  
    return response
```

## Objective 3 - Pairwise

**Task:** Complete the function below by writing a prompt that instructs the LLM to compare two passages and decide which one is more relevant to a given query. The LLM must choose one passage over the other, ensuring no ties.

**Instructions:** You need to construct both a system message and a user message by replacing "YOUR SYSTEM MESSAGE GOES HERE." and "YOUR USER MESSAGE GOES HERE.". Use {query}, {documentA} and {documentB}. The client parameter is the representative of the LLM. You can ignore it while completing the code.

**Output:** The LLM's response should clearly indicate which passage (A or B) it finds more relevant. No further explanation is required in the output.

### Example of the task:

- **Query:** "do goldfish grow"
- **Passage A:** "The conditions goldfish are kept in plus their diet determine how large they will grow. I have seen goldfish grow ridiculously large in very small containers when their water was changed frequently. Goldfish will not grow if water conditions are poor. Fancy goldfish donât grow as large as Common goldfish. A good size would be around 5 inches body length for most fancy varieties, 8 inches for Comets and 12 inches for Common Goldfish. These sizes are usually only attained by pond grown fish."
- **Passage B:** "Actually, it is humane to keep a goldfish in a bowl, as long as the water is good, and as long as the fish can be relocated to a pond or large habitat when it becomes mature, as in too big the live in the bowl. It is not automatically cruel to keep one in a bowl."
- **Correct Answer:** Passage A.

```
def pairwise_preference_task(query, documentA, documentB, client):  
    messages = [  
        {  
            "role": "system",  
            "content": "YOUR SYSTEM MESSAGE GOES HERE."  
        },  
        {  
            "role": "user",  
            "content": f"YOUR USER MESSAGE GOES HERE"  
        }  
    ]  
  
    response = client.chat.completions.create(  
        model="gpt-4o",  
        messages=messages,  
        temperature=0  
    )  
    return response
```