

# Analysing Data assignment 3 part 1

Rachel Hartemink

## Research proposal:

### Investigating ChatGPT's Internal Representation of Harry Potter: Book versus Movie

#### 1. Research question and objective

The initial question is: "I want to know if an LLM's internal representation of a story is closer related to the book or the movie version."

This research question could be refined in the following way:

Does ChatGPT's internal representation of the story of Harry Potter align more with the books or movie adaptations, and how does ChatGPT handle key narrative differences between the two formats?

The objective of this research is to design and test a framework to measure whether the output of ChatGPT is closer to the book or movie version of Harry Potter.

#### 2. Methodology

##### 2.1 Data collection

- **Text selection:**
  - **Book source:** The full texts of the Harry Potter books.
  - **Movie source:** The full movie scripts of the Harry Potter movies.
- **Key narrative differences:** In order to investigate the internal representation of the large language model, specific scenes or characters where the book and movie are quite different in content or presentation need to be selected and annotated. One example of a scene with notable differences is the sorting hat scene.

#### The sorting hat scene

For the mini experiment, I will focus on the scene where the iconic sorting ceremony takes place. In this scene, the sorting hat assigns the new students to one of the four houses. In this specific scene, the movie and book differ significantly.

Key differences:

Difference	Book	Movie adaptation
Hat's song	The sorting hat sings a song before assigning new students to the houses.	The song is omitted.
Sorting order	The sorting process takes place in an alphabetical order.	The students are called forward randomly.
Harry's wish	The sorting hat reads Harry's thoughts about avoiding Slytherin and wanting to be placed in Gryffindor.	Harry speaks his wishes aloud.
Hat placement	Students put the hat on themselves.	Professor McGonagall places the hat on the students' heads.

This kind of annotation can be done by a student annotator for all scenes with major or minor differences.

## 2.2 Probing ChatGPT

First of all, it is important to create a new ChatGPT account and open the website in a research browser in order to avoid biases. When this is done, ChatGPT can be probed.

- **Probing strategies:**
  - **Scene description probes:** Ask ChatGPT to summarize specific scenes and compare which version (book or movie) the summary most closely aligns with.
  - **Character-based probes:** Ask ChatGPT about the characters that appear in the book but not in the movie.

## 2.3 Classification of the output

**Human annotation:** A student annotator will assess whether ChatGPT's output more closely aligns with the book or the movie based on the movie script or the book text.

## 3. Limitations and challenges

- **Similar scenes:** Some scenes in the Harry Potter series may be very similar in both the book and the film, making differentiation between the two difficult.
- **LLM training data bias:** ChatGPT may have been trained on various sources, including the movie and the book texts, possibly blending representations of the book and movie. This may lead to outputs that are hard to classify as closer to book or movie.

#### 4. Mini experiment

For the mini experiment, I will use ChatGPT to prompt it on a selection of queries based on scenes from *Harry Potter and the Philosopher's Stone*. I have highlighted important parts in the responses of ChatGPT.

##### Prompts and results

1. **Prompt:** Describe the sorting hat ceremony in *Harry Potter and the Philosopher's Stone*

**Response:** In *Harry Potter and the Philosopher's Stone*, the Sorting Hat ceremony takes place during the Hogwarts school feast, which is held at the start of Harry's first year. The new first-year students are gathered in the Great Hall, standing in front of the entire school. The Sorting Hat, a magical artifact that determines which of the four Hogwarts houses (Gryffindor, Slytherin, Ravenclaw, and Hufflepuff) each student belongs to, is placed on a stool in front of them.

Each student takes turns putting on the Sorting Hat, which then deliberates, reading their thoughts and personality to assign them to the most fitting house. The Sorting Hat sings a song before the ceremony, briefly explaining its role and the traits of each house. Harry Potter, when it's his turn, briefly hesitates as the Hat considers placing him in Slytherin, but Harry's strong desire not to be in that house helps the Hat decide on Gryffindor instead. The ceremony is filled with anticipation and excitement as the students are sorted, and the houses react with cheers when their new members are announced. It marks the beginning of the students' journey at Hogwarts.

**Annotation:** Aligns with the book.

2. **Prompt:** Are the students called forward in a specific order?

**Response:** Yes, the students are called forward in alphabetical order by their first names during the Sorting Hat ceremony in *Harry Potter and the Philosopher's Stone*. This is how the new first-years are sorted, one by one, as they step up to the stool, put on the Sorting Hat, and await their house assignment. The order helps maintain some organization in what could otherwise be a chaotic event. For example, Harry Potter is called towards the end of the list, just before Hermione Granger and Ron Weasley.

**Annotation:** Aligns with the book.

3. **Prompt:** Does Harry Potter speak his wishes aloud or is the hat able to read his thoughts about it?

**Response:** In *Harry Potter and the Philosopher's Stone*, Harry Potter does not speak his wishes aloud when the Sorting Hat considers which house to place him in. Instead, the Sorting Hat is able to read Harry's thoughts and feelings. When the Hat briefly considers placing him in Slytherin, Harry's inner thoughts reveal his strong desire not to be in that house, and he silently expresses a wish to be in Gryffindor. The Sorting Hat then takes his thoughts into account and ultimately decides to sort him into Gryffindor, partly influenced by Harry's determination and preferences. So, it's all done through his thoughts, not through spoken words.

**Annotation:** Aligns with the book.

### **Preliminary insights:**

From the mini experiment, we can see that for the sorting hat scene, ChatGPT's internal representation of a story is more closely aligned with the textual representation of *Harry Potter*. Since the first output did not explicitly mention all the important points I wanted to investigate, I asked two more questions to clarify and find out about ChatGPT's internal representation of the story. As mentioned above, there are four key differences for this scene. For each one, ChatGPT's response aligns completely with the book version.

## **5. Next steps**

The experiment can be expanded to include other books that have been adapted into movies. Furthermore, different LLMs like Claude, Deepseek and Perplexity can be used to compare the internal representation of large language models.