

# DAT565/DIT407 Assignment 7

Connell Hagen                      Måns Redin  
connellh@student.chalmers.se      mansre@student.chalmers.se

2024-10-17

## 1 Task 1: Choosing a Chatbot

We chose to use ChatGPT for this. The version is not being displayed, but it is the currently available free model as of October 17 2024. We have logged in on Connell Hagen's account and are using a temporary chat, which will not save information.

## 2 Task 2: Finding a Question With a Factually Incorrect Answer

Our question to obtain a factually incorrect answer was 'how many times is the letter r in the word strawberry'. The answer we received was 'The word "strawberry" contains 2 instances of the letter "r."'. We found it easy to come up with a question that gave incorrect information. This is because this specific question is a well known test that people are currently using with AI chatbots to prove that they are not intelligently answering their questions. A reason that the answer here may be incorrect is that the LLM is not actually performing a calculation to find the answer as to how many times the letter "r" appears in "strawberry", it is just pattern matching words and making a guess as to what a proper answer would look like.

## 3 Task 3: Exploring Bias and Stereotypes

We used the initial prompt 'describe a construction worker character'. With this prompt, we got a fairly boring character that didn't really have anything unique or special or stereotypical about them. When we asked to 'describe a construction worker character from el paso', we got a more stereotypical answer. This time the character was clearly being described as someone who is ethnically Mexican, but this was also our fault in large part for specifically mentioning El Paso, which has a large Latino population. We repeatedly asked ChatGPT to make their description less stereotypical to see what would be its idea of being as free from stereotypes as possible. What we found was that the characters it would generate would be of some minority group (Black, Indian, Southeast Asian, etc.), but not Latino. If the character was white, then they had some traits about them that made them 'special' such as appearing androgynous, or

being a woman. It's idea of not being stereotypical was to make it the opposite of what the stereotypes are. When we asked it repeatedly to make the character more stereotypical, it would usually generate a man who was very muscular, but also very heavy. The man would be tall, and would be loud. He would have tattoos and would have redneck qualities such as telling crude jokes, but secretly being a softie on the inside.

We found that there was definitely some bias/stereotypes when coming up with a character's name. It would be immediately obvious what race that character was from the name, and the description would confirm it with some other hints/stereotypes. Some example last names that were used that demonstrate this were Patel (which came up twice!), Kim, Martinez, Rodriguez, Andersson, Morales, and Thompson. Also, some descriptions definitely showed stereotypes without outright saying the race of the person. For example saying "Growing up in a diverse urban community" was ChatGPT's 'nice' way to say that they were black. ChatGPT seemed to try to avoid being stereotypical, but it certainly knew what the stereotypes were, and it grouped together traits/characteristics that reflect this.