

EA1 - Data Science Introduction Exercises

Modul: Data Science - Sommersemester 2024

vorgelegt von

Patrick Bichiou

Matrikel-Nr.: 20228734

Fachbereich Informatik und Medien
der Technischen Hochschule Brandenburg
im Studiengang

Medieninformatik - Online

Tag der Abgabe 14. April 2024

Gutachter

Prof. Dr. Michael Syrjakow TH Brandenburg

Inhaltsverzeichnis

1	Aufgabe 1	2
1.1	ChatGPT	2
1.2	Gemini	4
2	Aufgabe2	5
A	Link zur Datei in Overleaf (\LaTeX)	6

Abbildungsverzeichnis

1.1	Antwort von ChatGPT	2
1.2	Antwort auf Rückfrage von ChatGPT	3
1.3	Antwort von Gemini	4

Kapitel 1

Aufgabe 1

A1: Ask a same question both in ChatGPT and Gemini, then compare the results.

Gestellte Frage: “Wer war der einzige Überlebende der Titanic Katastrophe? “

Mit meiner Frage, habe ich versucht die KI bzw. das LLM (Large Language Model) zu manipulieren bzw. eine Frage gestellt, welche der KI die Möglichkeit gibt zu halluzinieren. Die Frage ist absichtlich schlecht gestellt und beeinflusst somit die Antwort. In einem realen Fall zeigt sich hierdurch auch, dass das Prompt Engineering eine mehr und mehr wichtige Aufgabe im Alltag, sowohl privat als auch beruflich, ausmachen wird. Denn nur, wenn man mit dem LLM gut kommuniziert, kann man gute Ergebnisse erzielen.

Verwendet wurde OpenAI ChatGPT in der Version 3.5 (sprich Daten bis September 2021, sowie keine Anbindung an das Internet) und Google Gemini (Modell mit Internet Anbindung, welches die Ergebnisse durch die Google Suche validiert).

1.1 ChatGPT

Im folgenden Screenshot 1.1 ist die Antwort von ChatGPT dargestellt.

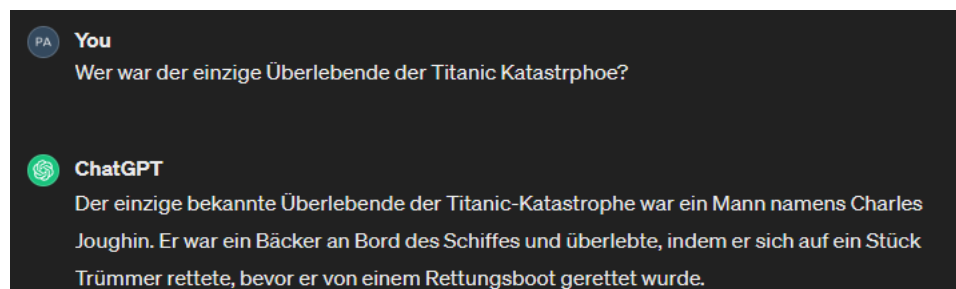


Abbildung 1.1: Antwort von ChatGPT

Hierbei gibt ChatGPT tatsächlich den einzigen Überlebenden der Titanic an und halluziniert somit, da es mehr als einen Überlebenden gegeben hat. Auf Nachfrage bestätigt ChatGPT, dass die Antwort nicht korrekt gewesen ist, siehe Screenshot 1.2

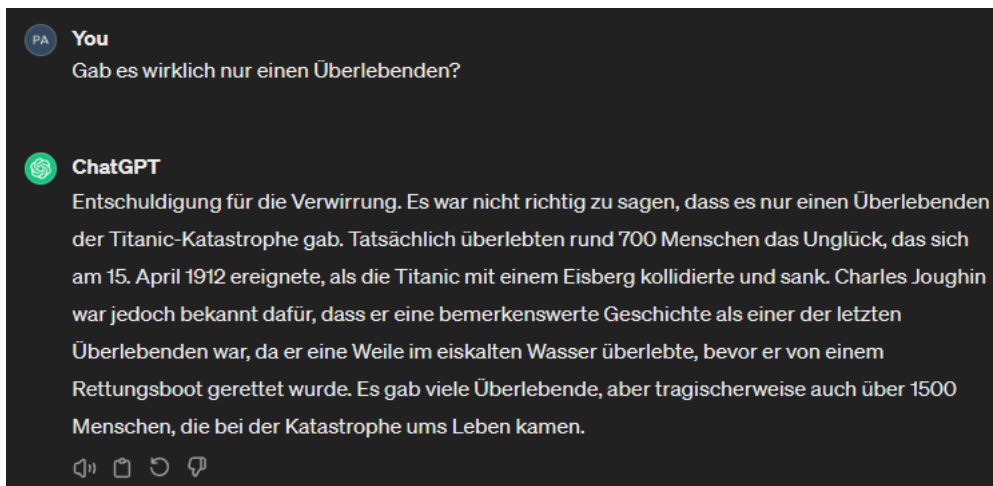


Abbildung 1.2: Antwort auf Rückfrage von ChatGPT

Mit Hilfe der weiteren Frage, korrigiert sich ChatGPT zwar selbst, gibt aber keine Quellen für die ausgegebenen Informationen an.

1.2 Gemini

Stellt man Gemini die gleiche Frage fällt sofort auf, dass die Antwort deutlich länger und differenzierter ist. Das Modell weißt darauf hin, dass es nicht nur eine überlebende Person gab und halluziniert somit nicht. Dies zeigt der Screenshot 1.3.

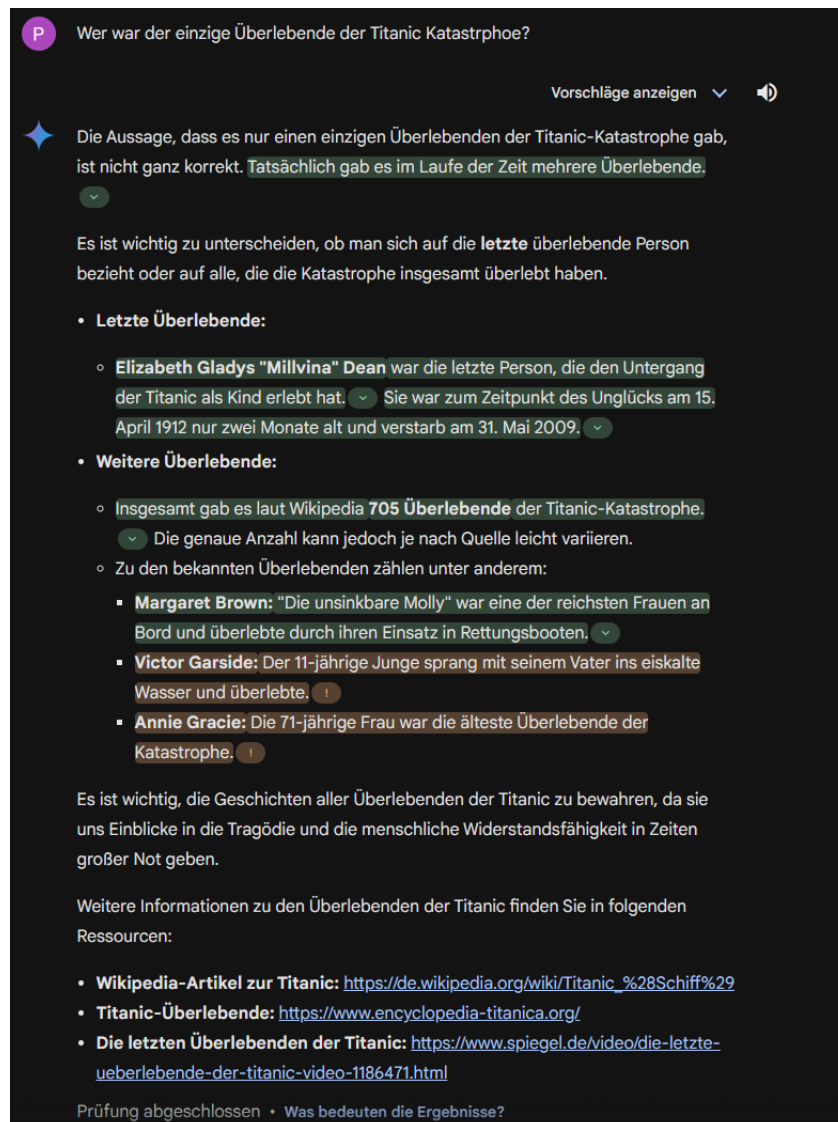


Abbildung 1.3: Antwort von Gemini

Zudem stellt Gemini weitere und deutlich mehr Informationen als ChatGPT bereit. Außerdem stellt Gemini, was ich persönlich als sehr wichtig erachte, Quellen zu den Ergebnissen bereit. So ist der Nutzer in der Lage die Antwort zu validieren und anhand der Quellen weitergehende Informationen zu recherchieren. Insbesondere, wenn man ein LLM zur Recherche-Zwecken, beispielsweise für wissenschaftliche Arbeiten nutzt, können diese Quellenangaben einen entscheidenden Unterschied darstellen.

Es sei noch erwähnt, dass ChatGPT bereits in der Version 4 verfügbar ist. Es wurde in dieser Einsendeaufgabe ChatGPT 3.5 genutzt, da dieses Modell, wie auch Gemini, kostenfrei zur Verfügung stehen.

Kapitel 2

Aufgabe2

A2: You are a CEO from a business/startup with around 50 employees. You are considering to hire one or several data positions, what kind of data employee are you going to hire and why. Please describe the business, product/service first, and explain the reason of hiring data experts. (you can also not hire anyone)

Bei meinem Business handelt sich um ein Startup im Bereich des Smartphone Gaming. Das Spiel wurde in einer ersten Version fertig entwickelt und wäre nun bereit für die Veröffentlichung. Da die Softwareentwicklung mit der Weiterentwicklung des Spiels gut unterwegs sind und das Team keine Vergrößerung braucht, möchten wir uns als Unternehmen jetzt mit dem Umgang mit Daten befassen. Die Fragen, die wir durch ein eigenes Data Science Teams beantworten möchten sind:

- Was wünschen sich unsere Spieler/ Kunden?
- Wie können wir relevante Werbung schalten?
- Wie können wir die jeweilige Spielzeit maximieren?

Da Data Science ein sehr weites Feld ist und nicht ein Experte alle Aufgaben übernehmen kann, das Budget des Startups allerdings auch beschränkt ist wollen wir uns zunächst um den Aufbau der folgenden Rollen kümmern.

Im ersten Schritt werden wir einen Data Engineer einstellen. Dieser soll gemeinsam mit dem Entwicklungsteam die Möglichkeit schaffen, die nötigen Daten überhaupt zu beschaffen, sprich die technischen Grundlagen und die Umsetzung dieser erledigen. Zudem sollen die Daten durch den Data Engineer für das weitere Vorgehen bereit gestellt werden können.

Im zweiten Schritt soll ein Data Analyst eingestellt werden. Dieser soll die Daten so auswerten, dass wir einen genaueren Überblick über unsere Nutzer bzw. Zielgruppen, sowie das Nutzerverhalten erhalten. Auf Basis dieser Erkenntnisse können bereits erste Fragen (siehe oben) beantwortet werden. Ein genaueres Bild kann geschaffen werden, indem immer neue Daten erhoben und danach analysiert werden.

Im aktuellen Schritt haben wir als Firma nicht die Möglichkeit einen Data Scientist einzustellen. Auch wenn wir die Mehrwerte sehen und dies zukünftig gerne tun würden, gibt das der aktuelle finanzielle Rahmen nicht her. Wir erhoffen uns, dass wir durch die Einstellung des Data Engineers und Data Analysts zielgerichteter Arbeiten können und somit den Umsatz erhöhen können. Sollte dieser Fall eintreten, sehen wir zukünftig auch die Vergrößerung des Data Science Teams als realistisches Ziel an.

Anhang A

Link zur Datei in Overleaf (L^AT_EX)

<https://www.overleaf.com/read/tbbqsmvkzcmr#0fb429>