

# Aufgabenblatt 11

Colmant

8. Juli 2025

## **Aufgabe 1: AI infrastructure**

## **Aufgabe 2: Justice explained by AI**

In dieser Aufgabe sollten wir einen AI-Agenten simulieren durch die Fähigkeit von modernen LLM-AIs den Kontext und Kohärenz über mehrere Fragen hinweg beibehalten. Ich habe mir den Aufgaben Teil A vorgenommen, in dem ich eine Conversation zwischen einem AI-Experten und einem Alien simuliere durch gezieltes Prompting. Der AI-Experte soll dem Alien erklären, was Gerechtigkeit ist. Ich habe mich dazu entschieden die AI einen Philosophen spielen zu lassen. Dann habe ich die Argumente so gut wie ich konnte mit meinem Philosophie-Wissen versucht zu zerlegen. Das Ergebnis ist eine interessante Diskussion, die zeigt, wie komplex und vielschichtig das Thema Gerechtigkeit ist und wie wichtig es ist in Definitionen genau und präzise zu sein. Das Gespräch kann unter folgendem Link nachgelesen werden: [AI conversation](#).

### **Evaluation**

Mir ist gleich zu Beginn aufgefallen, dass die AI eine sehr ungewöhnliche Definition von Gerechtigkeit gegeben hat. In ihrem Versuch Gerechtigkeit einem Alien ohne gesellschaftlichen Vorwissen zu erklären war die Definition sehr abstrakt und eher unpräzise und ließ viel Raum für Missinterpretationen. Fast bis zum Ende hat die AI aber Kohärenz bewahrt und hat den Charakter des Philosophen nicht gebrochen. Was ich aber schade fand ist dass, die AI nicht wirklich einen Philosophen gespielt, sondern eher alle philosophischen Ideen wieder gegeben hat. Das führte dazu dass der Charakter keine wirklichen eigenen Ideen bzw. feste Standpunkte hatte. Das führte dazu dass, die AI die Einwände des Aliens so entkräftet hat indem es einfach eine Reihe neuer Argumente gebracht hat, aus denen sich dann eins ausgesucht werden konnte. Als das Alien dann am Ende zu erkennen gab dass, es die Argumente auf eine eher unmoralische Weise interpretiert brach die AI dann aber doch etwas aus dem Charakter und es wirkt so als würde eher ein Teil des Gedankenprozesses in der Ausgabe sichtbar sein. In der aller letzten Antwort stimmte die AI dann auch noch dem Alien zu dass, die AI eigentlich gar nicht Gerechtigkeit definieren darf aufgrund der oben genannten Argumente. Für eine philosophische Diskussion ist die AI unbrauchbar sie ist sowohl zu ungenau und gleichzeitig gibt sie zu viele Argumente auf einmal so dass es sehr schwer ist von einem Argument zum nächsten zu kommen. Das kann natürlich auch an meinem Prompting liegen.

### **Github**

Wie immer sind alle meine benutzten Dateien auf meinem Github zu finden.