

GetTogetheR!

ChatGPT und R

Jonas Frost jonas.frost@studserv.uni-leipzig.de

Peter Kannewitz peter.kannewitz@uni-leipzig.de

08. Juni 2023

Motivation

- · ChatGPT und andere KI-Tools haben in letzter Zeit stark an Bedeutung gewonnen
 - · mehr Anwendungsmöglichkeiten
 - · schnellere Benutzung
 - · einfacher Zugang
 - · große Trainingsdatensätze
- · Wir können diese Tools für unsere Arbeit mit R nutzen!
- · ChatGPT kann R-Code ...
 - · generieren
 - korrigieren
 - erklären
 - kommentieren
 - · optimieren
- · Wir müssen wissen, wie wir am besten mit ChatGPT kommunizieren, um die gewünschten Ergebnisse zu bekommen (Promting)

Ablauf

- 1. Kurzeinstieg: ChatGPT
- 2. ChatGPT nutzen
- 3. Ressourcen und Hilfe
- 4. Austausch



Was ist GPT?

- Generative Pre-trained Transformer
 - · generatives Sprachmodell mit Transformer Architektur
 - · Zusammensetzung von künstlichen neuronalen Netzen
- Bezeichnet eine Familie von Large Language Models (LLM)
 - · Ziel: möglichst viel Trainingsdaten, um Sprache "generieren" zu können
- Erste Version GPT-1 2018 veröffentlicht (117 Millionen Parameter) 1
 - · aktuell: GPT-4, 2023 veröffentlicht
 - · Vorgänger: GPT-3 mit 175 Billionen Parametern

¹https://en.wikipedia.org/wiki/Generative pre-trained transformer

Intuition: Wie funktioniert ein LLM?

- · Unupervised machine learning
 - · keine explizite "Ground Truth"
 - · Muster in den Trainingsdaten werden im Netz abgebildet
- · Ziel: Gegeben eine Vorgeschichte \rightarrow Was ist das nächst Wort?

The best thing about AI is its ability to	learn	4.5%
	predict	3.5%
	make	3.2%
	understand	3.1%
	do	2.9%

Figure 1: Quelle:

https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/

Was ist ChatGPT?

ChatGPT nutzen

Möglichkeiten ChatGPT in einen R-Worflow zu integrieren

- WebApp
 - · chat.openai.com
 - · Voraussetzungen: OpenAl Konto
- · R-Packages
 - · Voraussetzungen: Konto, gültigen API-Key
 - · Einschränkung: Rate-Limit
 - · R-Package und und RStudio-Plugin: chatgpt::

Wir zeigens Euch: WebApp

Beispiel 1

"Given a dataset with 3 variables "name", "group" and "status", where name is a character, and group and status are dummy variables. Generate an R code that generates a cross table of the two variables group and status. Also, the code should filter cases with status 1 and print the names of persons with status 1. Use the tidyverse package and comment the code."

Beispiel 2

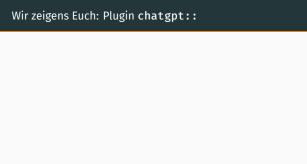
"Now, using the same dataset write an R code using the ggplot2 package that generates a barchart showing the relative frequencies of the status variable. Color the bars using the group variable."

Tipps für's Promting

- gebt ChatGPT so spezifische Anweisungen wie möglich:
 - · Welche Packages?
 - · Welche Datentypen?
 - · Wie soll das Ergebnis aussehen?
 - · Wie lautet die Fehlermitteilung?
- · Oft ist Codegeneration mit ChatGPT iterativ:
 - 1) Initialer Prompt
 - 2) erster generierter Code
 - 3) Code testen
 - 4) ChatGPT auf Fehler hinweisen oder Prompt entsprechend anpassen
 - 5) Schritt 3) und 4) wiederholen, bis Code wie erwartet läuft

OpenAl API Zugriff

- · (P)
- · Wie, Probleme, Pricing, Zeigen



· Einfaches Beispielscript mit Plugin bearbeiten

Ressourcen und Hilfe

Nützliche Ressourcen zum Thema

· (P & J)

Austausch

Austaus<u>ch</u>

- · Welche Erfahrung habt Ihr mit ChatGPT in Bezug auf R gemacht?
- · Welche Nachteile könnte ein aktiver Gebrauch von Chat-GPT haben?
- ٠..

Danke fürs Teilnehmen!