

Bachelor-/Masterarbeit

Einsatz von generativen Sprachmodellen zur Informationsgewinnung aus PDF-Dokumenten

Iserlohn,
28.07.2023

Aufgabenbeschreibung

Neuartige Dienste, wie etwa ChatPDF [1] oder PDF.ai [2] nutzen die generative KI ChatGPT [3] um PDF-Dokumente zu analysieren und zu durchsuchen. Dabei können Anfragen wie in einem Chat an das entsprechende Dokument gestellt werden, die dann von der KI in textuelle Antworten oder auch Programmcode überführt werden.

In dieser Abschlussarbeit soll ein Tool entwickelt werden, das ähnliche Funktionalitäten wie die oben genannten Dienste [1,2] aufweist. Die Herausforderungen für diese Arbeit sind u.a.:

- Extraktion des Textes aus PDF Dokumenten
- Verwendung der OpenAI API
- Formulieren von Eingabeaufforderungen (*Prompts*) für ChatGPT
- Verarbeitung langer Texte, die die maximale Kontextgröße von ChatGPT übersteigen

Referenzen

[1] <https://www.chatpdf.com>

[2] <https://pdf.ai>

[3] <https://chat.openai.com>

**Fachbereich Informatik und
Naturwissenschaften**

Prof. Dr.
Heiner Giefers

Cloud Computing

Telefon
02371 566-5252

E-Mail
giefers.heiner@fh-swf.de

Standort Iserlohn
Frauenstuhweg 31
58644 Iserlohn

Fachhochschule
Südwestfalen
Sitz: Iserlohn

**Hagen
Iserlohn
Lüdenscheid
Meschede
Soest**

www.fh-swf.de

Wir geben Impulse

