DAS Faculty of Digital & Analytical Sciences

PARIS LODRON UNIVERSITÄT SALZBURG

Einführung und Organisatorisches

UV+SE AI Werkstatt – Wintersemester 24/25

Roland Kwitt / Frank Pallas

Heute:

⟶ Vorstellung Kursorganisation / Prüfungsteile

⟶ Themenvorstellung

UV+SE AI-Werkstatt: Generelles Modell

UV und SE grds. miteinander verwoben

2er od. 3er Teams bearbeiten je ein Thema zur realistischen Anwendung von AI-Verfahren in der Praxis

Zu diesem Thema unterschiedliche Abgaben
(Vorträge, Reports, … – später mehr)

Für die Bearbeitung gilt: …

Für alle Themen gilt:

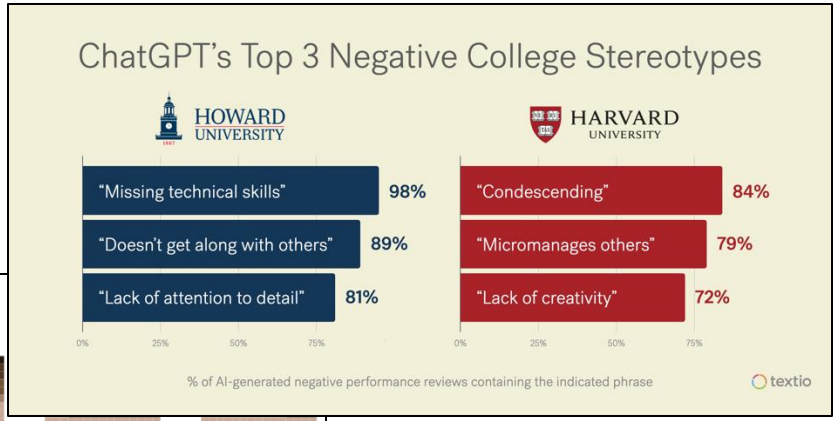
Realistisch:
Ausdrücklich **nicht** fundamentale Primitive, Basisalgorithmen etc. selbst nachimplementieren ⟶ Auf existierende Standard-Libraries, Basismodelle, usw. zurückgreifen und möglichst praxisnahe, KI-basierte Anwendungen umzusetzen

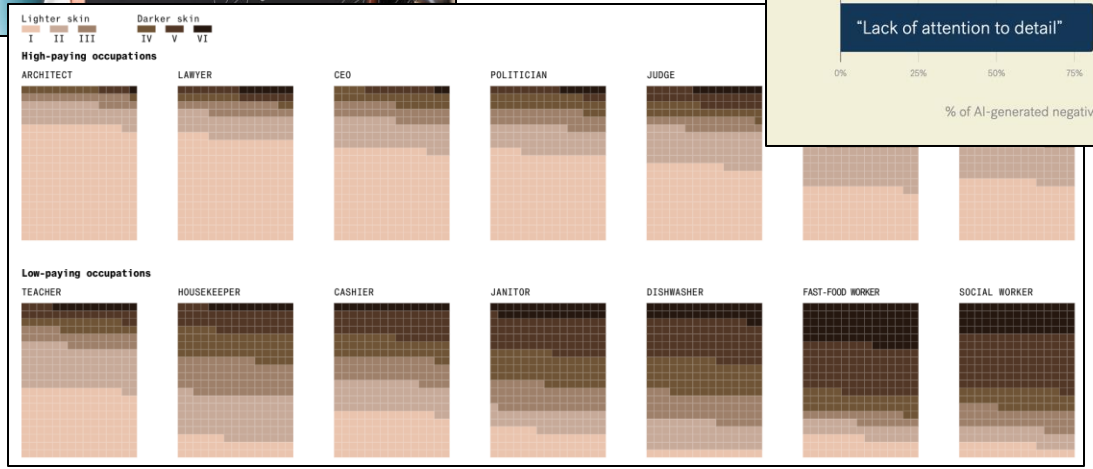
Systemkontext:
Bearbeitung im Kontext eines realistischen Systemkontextes mit konkretem Anwendungsbezug ⟶ Bearbeitung beinhaltet auch die Spezifikation des angenommenen Anwendungsszenarios. Umsetzung soll neben dem KI-Kern auch dessen stimmige Integration in größere Anwendungskontexte und entspr. Software Stacks realisieren

Experimentelle Bewertung:
Geeignete experimentelle Bewertungen (z.B. hinsichtlich Performance unter möglichst realistischen Umgebungen, Auswirkungen ausgewählter Parameter wie LLM-Größe, …) als integraler Bestandteil einer erfolgreichen Bearbeitung

Thema 1: Wie biased sind vortrainierte LLMs?

https://textio.com/blog/mindful-ai-crafting-prompts-to-mitigate-the-bias-in-generative-ai





Bloomberg testing Stable Diffusion - https://www.bloomberg.com/graphics/2023-generative-ai-bias/

Thema 1 : Wie biased sind vortrainierte LLMs?

Aufgabe:

- Finden und skizzieren Sie einen LLM-Anwendungsfall, in dem mindestens 3 möglichst unterschiedliche Bias-Risiken bestehen.
- Demonstrieren Sie die unerwünschten Auswirkungen dieser Biases in einem realistischen Anwendungsfall.
- Designen Sie einen Ansatz, mit dem sich die Biases für mindestens drei unterschiedliche LLMs experimentell bewerten lassen. Finden oder Erstellen Sie ein hierzu geeignetes Test-Datenset und führen Sie die experimentellen Bewertungen durch.

Stretch-Goal:

- Lassen sich die betrachteten Modelle durch Nachtrainieren mit geeigneten Daten “de-biasen”? Was ist dafür notwendig? Wie erfolgreich ist dies? Welche möglichen negativen Auswirkungen entstehen dadurch?

Thema 2: Datenschutzfreundliche KI-Nutzung

Thema 2: Datenschutzfreundliche KI-Nutzung

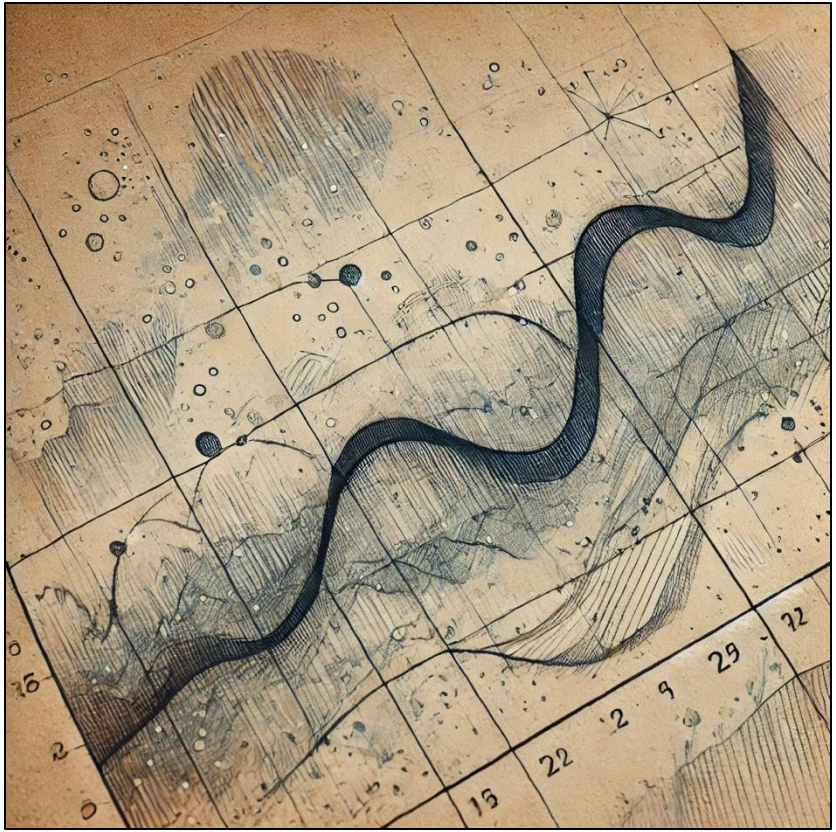
Aufgabe:

- Finden und skizzieren Sie einen KI-Anwendungsfall, in dem personenbezogene Daten notwendigerweise für das Training genutzt werden. Implementieren Sie den Anwendungsfall (incl. Training mit einem geeigneten Datensatz) und skizzieren Sie das Privatheitsproblem anschaulich.
- Führen Sie dann parallel auf dem Datensatz eine geeignete Anonymisierung durch und trainieren Sie das genutzte KI-Modell auf dem anonymisierten Datensatz.
- Wie wirkt sich die Anonymisierung auf die Qualität der Ergebnisse aus? Welchen Einfluss haben unterschiedliche Anonymisierungparameter?

Stretch-Goal:

- Sind im Rahmen der vorgeschalteten Anonymisierung evtl. noch Optimierungen möglich, etwa durch geschickte Auswahl der zur Anonymisierung genutzten Attribute? Für einige Standardverfahren und -implementierungen (insb. scikit-learn) existieren außerdem mittlerweile prototypische Implementierungen zB für “differentially private learning”. Wie schneiden diese im Vergleich ab?

Thema 3: Bewertung/Evaluierung moderner ML Ansätze zur Zeitreihenvorhersage



Thema 3: Bewertung/Evaluierung moderner ML Ansätze zur Zeitreihenvorhersage

Aufgabe:

- Identifikation eines geeigneten (large-scale) Benchmark Datensatzes (z.B. M4 Competition); univariant od. multivariat (je nach Präferenz)
- Recherche und Auswahl geeigneter Performanz Maße für das jeweilige Problem
- Evaluierung einer geeigneten Auswahl an (2-3) aktuellen ML Ansätzen zur Zeitreihenvorhersage hinsichtlich Vorhersagequalität, Resourcenbedarf, Anpassungsmöglichkeiten.

Stretch-Goal:

- Ist es möglich die verschiedenen Ansätze zu kombinieren, um bessere Vorhersagen zu erhalten? Welche Möglichkeiten gibt es hier? Was ist sinnvoll, was nicht? Ist es möglich, bevor eine Vorhersage getroffen wird, zu entscheiden ob überhaupt eine Vorhersage getroffen werden soll?

Thema 4: ML Methoden zur Analyse & Modellierung kollektiven Verhaltens

Thema 4: ML Methoden zur Analyse & Modellierung kollektiven Verhaltens

Aufgabe:

- Nutzen Sie existierende Bibliotheken zur Simulation kollektiven Verhaltens mittels verschiedener etablierter Modelle
- In 2D/3D würden Sie z.B. sich über die Zeit verändernde Punktewolken erhalten
- Variieren Sie die Modellparameter und generieren Sie so eine Menge an „Trainingsdaten“
- Folgend identifizieren Sie passende Ansätze solche Sequenzen von Punktwolken verarbeiten zu können (z.B. rekurrente neuronale Netze mit entsprechender Funktionalität Punktewolken als Input zu erhalten
- Trainieren und evaluieren Sie folgend, wie gut es möglich ist mittels solcher Ansätze die Parameter (meist 1-4) des Simulationsmodells vorherzusagen
- Welche Performanz Maße zur Evaluierung sollen verwendet werden?

Stretch-Goal:

- Die Aufgabenstellung kann beliebig erweitert werden, vor allem hinsichtlich der Anzahl an simulierten Modellen und möglichen Ansätzen zur Verarbeitung der Sequenzen.

Prüfungselemente

UV:

- TechReport: Erklären des technischen Rahmens, des Problems, des angenommenen Anwendungsfalls und des verfolgten Ansatzes
  - Ca. 3 Seiten IEEE double-column (A4),
  - 15. Dezember
- Short Paper / Report: TechReport + beschreiben des eigenen Beitrags, incl. exp. Bewertung
  - Ca. 6 Seiten IEEE double-column (A4),
  - 29. Jänner

SE:

- Code: Problemangemessenheit, Qualität, …
- 3x Talks
- Diskussionsbeteiligung etc.

Semesterablauf