DAS Faculty of Digital & Analytical Sciences

PARIS LODRON UNIVERSITÄT SALZBURG

Einführung und Organisatorisches

UV+SE AI Werkstatt – Wintersemester 25/26

Christine Bauer / Roland Kwitt / Frank Pallas

Heute:

⟶ Vorstellung Kursorganisation / Prüfungselemente

⟶ Themenvorstellung

UV+SE AI-Werkstatt: Generelles Modell

UV und SE sind grundsätzlich miteinander verwoben

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

Zum gewählten Thema sind mehrere, unterschiedliche Abgaben erforderlich. (Vorträge, Reports, … – später mehr)

Für die Bearbeitung gilt: …

Für alle Themen gilt:

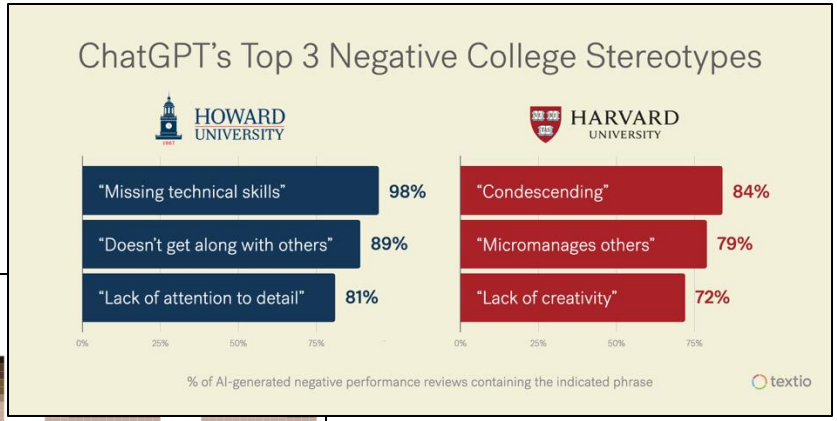
Realistisch:
Ausdrücklich **nicht** fundamentale Primitive, Basisalgorithmen etc. selbst nachimplementieren ⟶ Auf existierende Standard-Libraries, Basismodelle, usw. zurückgreifen und möglichst praxisnahe, AI-basierte Anwendungen umzusetzen

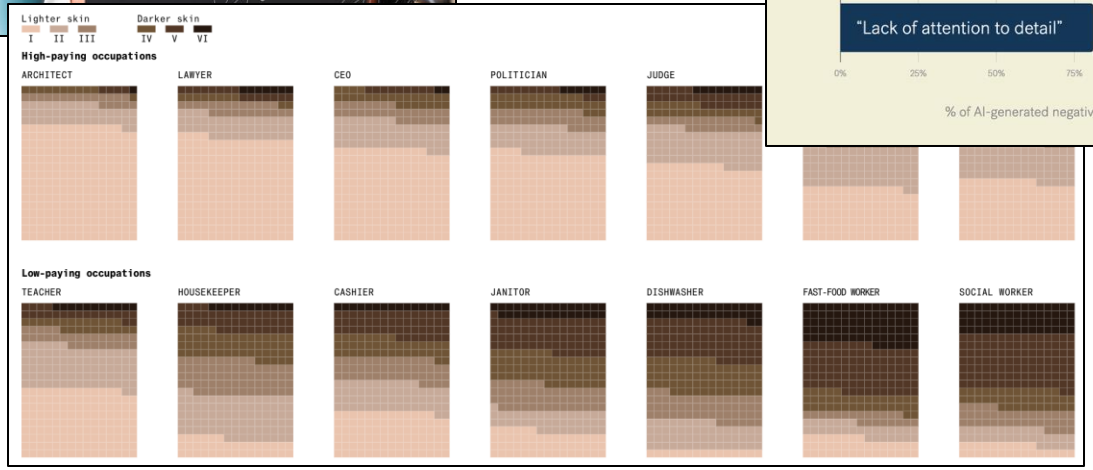
Systemkontext:
Bearbeitung im Kontext eines realistischen Systemkontextes mit konkretem Anwendungsbezug ⟶ Bearbeitung beinhaltet auch die Spezifikation des angenommenen Anwendungsszenarios. Umsetzung soll neben dem KI-Kern auch dessen stimmige Integration in größere Anwendungskontexte und entspr. Software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

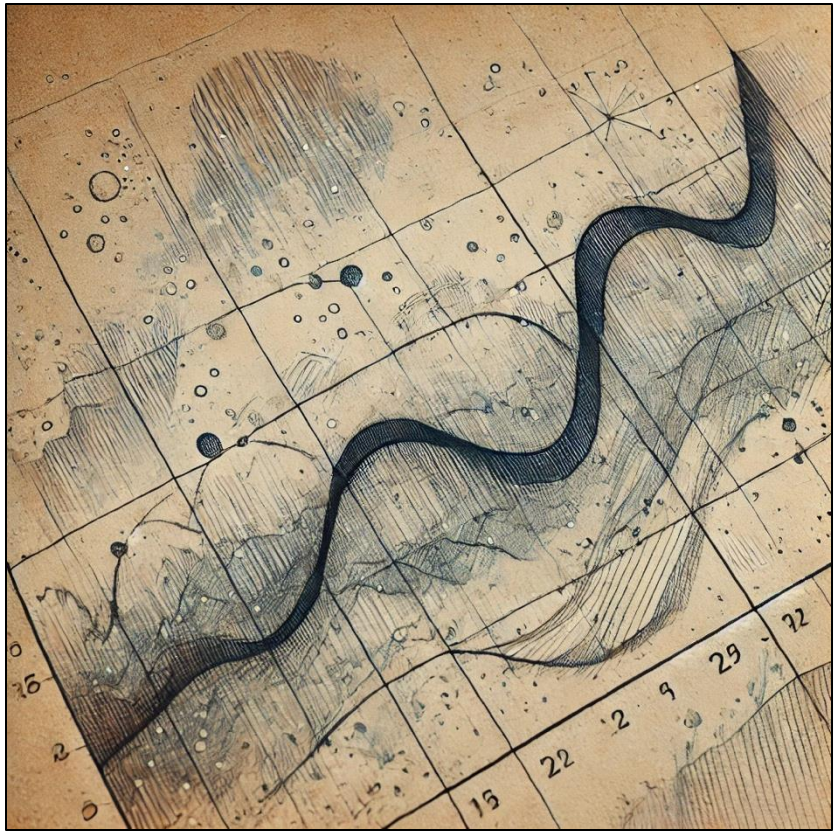
Aufgabe:

- 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, Rechen-Ressourcen, 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: Modelle zur Playlist Vervollständigung

Aufgabe:

- Implementieren Sie einen Ansatz zur **automatischen Vervollständigung** von Playlists mit Hilfe existierender Bibliotheken (z.B. ein Transformer Modell) und evaluieren Sie diesen auf dem Spotify 1M Playlist Datensatz.
- Wie weit sind Sie mit Ihrem Ansatz vom Leaderboard auf diesem Datensatz entfernt, wenn eines der Evaluierungsszenarien mit den entsprechenden Metriken betrachten.

Stretch-Goal:

- Implementieren Sie alle zusätzlichen Evaluierungsszenarien.


```
"name": "musical",
"collaborative": "false",
"pid": 5,
"modified_at": 1493424000,
"num_albums": 7,
"num_tracks": 12,
"num_followers": 1,
"num_edits": 2,
"duration_ms": 2657366,
"num_artists": 6,
"tracks": [
    {
        "pos": 0,
        "artist_name": "Degiheugi",
        "track_uri": "spotify:track:7vqa3sDmtEaVJ2gcvxtRID'
        "artist_uri": "spotify:artist:3V2paBXEoZIAhfZRJmo2
        "track_name": "Finalement",
        "album_uri": "spotify:album:2KrRMJ9z7Xjoz1Az4O6UML'
        "duration_ms": 166264,
        "album_name": "Dancing Chords and Fireflies"
    },
    {
        "pos": 1,
        "artist_name": "Degiheugi",
        "track_uri": "spotify:track:23EOmJivOZ88WJPUbIPjh6'
        "artist_uri": "spotify:artist:3V2paBXEoZIAhfZRJmo2
        "track_name": "Betty",
        "album_uri": "spotify:album:3lUSlvjUoHNA8IkNTqURqd'
        "duration_ms": 235534,
        "album_name": "Endless Smile"
    },
```

Thema 5: Vergleichende Analyse von Ähnlichkeits- und Diversitätsmetriken für Empfehlungssysteme

Thema 5: Vergleichende Analyse von Ähnlichkeits- und Diversitätsmetriken für Empfehlungssysteme

Aufgabe:

- Identifikation und Implementierung von Ähnlichkeits- und Diversitätsmaßen (similarity + diversity measures) für Empfehlungssysteme
  - Hinweis: Oft hängt die Implementierung von einer Ähnlichkeits-Funktion und der zugrundeliegenden „Embeddings“ ab
- Vergleichende Analyse dieser Maße anhand mehrerer Datensätze (aus möglichst unterschiedlichen Anwendungsdomänen) und mehrerer Empfehlungsalgorithmen (Baseline und mind. 2 weitere)
- Wie kongruent (oder unterschiedlich) sind die Indikationen für Ähnlichkeit oder Diversität für unterschiedliche Maße?
- Quantifizierbare Zielgröße: Ähnlichkeit oder Diversität von Empfehlungslisten (intra-list).

Thema 6: Zuverlässigkeit der Offline-Evaluierung von Empfehlungssystemen

Thema 6: Zuverlässigkeit der Offline-Evaluierung von Empfehlungssystemen

Aufgabe:

- Startpunkt ist folgender (RecSys 2025) Artikel:

  On the Reliability of Sampling Strategies in Offline Recommender Evaluation

- Untersuchen Sie, wie verschiedene **Sampling-Strategien** die *Zuverlässigkeit* der Offline-Evaluierung von Empfehlungssystemen unter unterschiedlichen **Expositionsverzerrungen (Exposure Biases)** beeinflussen.
- **Reproduzieren** Sie mindestens eine vollständige Pipeline (z.B. Uniform + Random@n + ALS).
- Arbeiten Sie aus, wann und wie die Sampling Strategien die Evaluierungsergebnisse verzerren und welche (praktikablen) Richtlinien man zur Auswahl geeigneter Strategien geben könnte um robuste Offline Vergleiche zwischen Methoden zu ermöglichen.

Prüfungselemente/Benotung

AI-Werkstatt (UV):

- Update Report: Erklären des technischen Rahmens, des Problems, des angenommenen Anwendungsfalls und des verfolgten Ansatzes
  - Ca. 3 Seiten IEEE double-column (A4),
  - Deadline: wird noch bekannt gegeben
- Full Report: Update Report + Beschreibung des eigenen Beitrags, inkl. experimenteller Bewertung
  - Ca. 6 Seiten IEEE double-column (A4),
  - Deadline: wird noch bekannt gegeben

AI-Werkstatt (SE):

- Code: Problemangemessenheit, Reproduzierbarkeit, Systematik der Evaluierung etc.
- 3x Talks (gedacht als Status Updates zu dem gewählten Thema der UV; 2x Update + 1x Final)
- Diskussionsbeteiligung, Präsentationsqualität (Klarheit, Struktur etc.)

Semesterablauf