DS 34 – Halluzination, n-Gramm-Modelle vs. moderne LLMs & Berufsorientierung

Informatik 10 – MINT-Zweig | Künstliche Intelligenz

Lernziel

Ich erkläre das Phänomen der Halluzination anhand von n-Gramm-Modellen, beschreibe Unterschiede zwischen n-Gramm-Modellen und modernen LLMs und recherchiere Auswirkungen generativer KI-Systeme auf Berufe.

🕘 Stundenverlaufsplan (Doppelstunde, 90 Minuten)

Orientierungsrahmen für 11 SuS: Die ersten 80 Minuten bilden den Kern. Die letzten 10 Minuten sind bewusst flexibel und werden nicht mit neuem Pflichtstoff belegt.

Phase	Zeit	Inhalt/Aktion	Sozialform
Einstieg und Aktivierung	10'	Einen plausibel klingenden, aber falschen oder unsinnigen Modellsatz beurteilen.	Plenum; kurze Einzeldenkzeit
Fragend-entwickelnde Erarbeitung	25'	Halluzination am n-Gramm-Modell erklären und die Grenzen der Analogie zu modernen LLMs klären.	Unterrichtsgespräch; Beiträge der SuS werden sichtbar festgehalten
Gemeinsame Umsetzung und Sicherung	20'	Beobachtbare Unterschiede zwischen n-Gramm-Modell und LLM gemeinsam tabellarisch sichern.	Plenum; Live-Coding, Tafelbild oder gemeinsame Simulation
Kurzer Transfer	15'	Auswirkungen generativer KI auf ein Berufsfeld anhand der Leitfragen recherchieren.	Einzelarbeit mit anschließendem Tandemabgleich (4 Tandems + 1 Trio)
Auswertung	10'	Ergebnisse kurz vorstellen und Aussagen mit konkreten Beispielen begründen.	Plenum; Lösung erklären und Fachsprache sichern
Reserve / Vertiefung	10'	Berufsbiografie, Chancen-Risiken-Abwägung oder Abschlussreflexion vertiefen.	Flexibel; bei Bedarf für Fragen, Fehlersuche oder Abschluss nutzen

🗣️ Sprachhilfe

- „Auf den Kontext … folgt mit der Häufigkeit/Wahrscheinlichkeit … das Wort …“
- „Der erzeugte Text verändert sich, weil …“
- „Im Vergleich zum n-Gramm-Modell kann das LLM …, jedoch …“

📌 Merksätze

- **Halluzination** bezeichnet das Phänomen, dass ein Sprachmodell Text erzeugt, der **grammatikalisch bzw. statistisch plausibel**, aber **inhaltlich falsch oder unsinnig** ist.
- Bei n-Gramm-Modellen entsteht Halluzination, weil das Modell **keinerlei "Verständnis"** der Bedeutung hat – es reiht Wörter nur nach **statistischer Wahrscheinlichkeit** aneinander, unabhängig davon, ob der entstehende Satz sachlich korrekt ist.
- Moderne **LLMs** unterscheiden sich von einfachen n-Gramm-Modellen u. a. in:
    - **Sprachqualität:** LLMs erzeugen deutlich flüssigeren, grammatikalisch korrekteren Text.

- **Kontextberücksichtigung:** LLMs können sehr viel längere Kontexte einbeziehen (nicht nur 1–2 vorherige Wörter).
- **Korrektheit:** LLMs erzeugen häufig zuverlässigere Antworten, können aber weiterhin halluzinieren. Das n-Gramm-Modell dient hierfür als **vereinfachte Analogie**; die Ursachen bei modernen LLMs sind wesentlich komplexer und nicht mit dem einfachen Modell gleichzusetzen.
- **Vielseitigkeit:** LLMs können unterschiedlichste Aufgaben (Übersetzen, Zusammenfassen, Programmieren …) bewältigen, während ein einfaches n-Gramm-Modell nur Text im Stil seines Trainingskorpus fortsetzt.

💡 Beispiel: Halluzination provozieren

Mit einem sehr kleinen, unpassenden Trainingskorpus erzeugt unser n-Gramm-Modell aus DS 33 leicht unsinnige, aber grammatikalisch klingende Sätze – das ist eine vereinfachte Analogie zur Halluzination bei LLMs.

```python
import random

korpus = """
der hund läuft im park der hund spielt mit dem ball
die katze sitzt auf dem baum die katze beobachtet den hund
der hund bellt laut der hund läuft zur katze
die katze springt vom baum die katze rennt weg
der hund läuft ihr hinterher der hund ist müde
"""

def erstelle_bigramm_tabelle(text):
    woerter = text.split()
    tabelle = {}
    for i in range(len(woerter) - 1):
        a, b = woerter[i], woerter[i + 1]
        tabelle.setdefault(a, {})
        tabelle[a][b] = tabelle[a].get(b, 0) + 1
    return tabelle

def naechstes_wort_random(tabelle, wort):
    folgewoerter = tabelle.get(wort)
    if not folgewoerter:
        return None
    woerter = list(folgewoerter.keys())
    gewichte = list(folgewoerter.values())
    return random.choices(woerter, weights=gewichte, k=1)[0]

def generiere_text(tabelle, startwort, laenge, strategie):
    text = [startwort]
    aktuelles_wort = startwort
    for _ in range(laenge - 1):
        naechstes = strategie(tabelle, aktuelles_wort)
        if naechstes is None:
            break
        text.append(naechstes)
        aktuelles_wort = naechstes
    return " ".join(text)

tabelle = erstelle_bigramm_tabelle(korpus)
for _ in range(5):
    print(generiere_text(tabelle, "die", 8, naechstes_wort_random))
```

```
die katze die katze rennt weg der hund
die katze springt vom baum die katze die
die katze springt vom baum die katze springt
die katze beobachtet den hund ist müde
die katze die katze beobachtet den hund spielt
```

📝 Übung 1 – Halluzinationen einordnen

Führt die Zelle oben mehrfach aus und sucht einen erzeugten Satz heraus, der **grammatikalisch plausibel, aber inhaltlich seltsam oder unsinnig** ist. Erklärt in eigenen Worten, **warum** genau dieser Fehler entsteht (Bezug zur Funktionsweise des n-Gramm-Modells).

(Euer Beispielsatz und eure Erklärung hier)

📝 Übung 2 – Vergleichstabelle

Erstellt eine Tabelle, die n-Gramm-Modelle und moderne LLMs bezüglich Sprachqualität, Korrektheit, Kontextberücksichtigung und Vielseitigkeit gegenüberstellt (Doppelklick zum Bearbeiten):

Kriterium	n-Gramm-Modell	Modernes LLM
Sprachqualität		
Korrektheit		
Kontextberücksichtigung		
Vielseitigkeit		

🧭 Berufliche Orientierung: Auswirkungen generativer KI auf Berufe

Recherchiert in Partner- oder Kleingruppenarbeit (ca. 20–25 Minuten) die Auswirkungen generativer KI-Systeme auf **ein** Berufsfeld eurer Wahl, z. B.:

- Journalismus / Content-Erstellung
- Softwareentwicklung
- Grafikdesign / Illustration
- Kundenservice / Support
- Übersetzung

Leitfragen für eure Recherche

1. Welche Tätigkeiten dieses Berufsfelds werden bereits heute durch generative KI unterstützt oder verändert?
2. Entstehen durch generative KI **neue** Aufgaben oder sogar neue Berufsbilder in diesem Bereich (z. B. "Prompt Engineering")?
3. Welche Tätigkeiten lassen sich (aus heutiger Sicht) **nicht** oder nur schwer durch KI ersetzen – und warum?
4. Wie bewertet ihr diese Entwicklung – welche Chancen und welche Risiken seht ihr?

Haltet eure Ergebnisse in Stichpunkten fest und bereitet einen kurzen mündlichen Beitrag (ca. 1 Minute) für das Plenum vor. 5. Findet (falls möglich) eine **reale Person** mit Bezug zu diesem Berufsfeld und recherchiert ihre Berufsbiografie (Werdegang: Ausbildung, erste Stelle, heutige Position) – z. B. über eine Unternehmens-Website, ein Interview oder ein Karriereportal.

(Platz für eure Rechercheergebnisse)

Berufsfeld:

Bereits veränderte/unterstützte Tätigkeiten:

Neue Aufgaben/Berufsbilder:

(Noch) nicht ersetzbare Tätigkeiten:

Chancen und Risiken – unsere Einschätzung: