

Pen and Paper – AI Training

Beschreibung:

Große Sprachmodelle (LLMs) werden vielfältig falsch angewandt und genießen doch den Nimbus der alleinseligmachenden Lösung für alle Probleme. Um LLMs aber reflektiert und erfolgreich einsetzen zu können bedarf es eines grundlegenden Verständnisses der Wirkungsweisen und der erwartbaren Ergebnisse.

Ziel des „Pen & Paper – AI Trainings“ ist es in ca. 30 – 40 Minuten potentiellen KI Nutzern ein Basisverständnis für die Wirkungsweise eines LLMs zu vermitteln. Dies erfolgt bewusst ohne IT-Einsatz im Methodensetting des Expertenkongresses.

Ein Expertenkongress wird in zwei Runden durchgeführt. Zuerst erarbeiten sich die Teilnehmer „Experten Kenntnisse“ zu einem Teilgebiet in themengleichen Gruppen.

In Runde zwei werden die Experten themenungleich gemischt und bearbeiten ein anspruchsvolleres Thema, das eine höhere Lernzielkategorie erfüllt und das Wissen aller Expertengruppen benötigt.

Die Ergebnisse aus Runde 2 werden im Plenum geteilt. Bearbeitungszeit für Runde 1 etwa 10-15 Minuten, für Runde 2 etwa 15-20 Minuten und etwa 10 Minuten für den Austausch nach den beiden Runden. Der gesamte Zeitbedarf liegt bei etwa 45 Minuten.

Lernziele:

Die Teilnehmenden verstehen die Arbeitsweise von Large Language Models (LLMs) im Bewerbermanagement, indem sie ihre spezifischen Rollen im Input-Management, Embedding, Intention-Management und Output-Management praxisnah anwenden. Sie analysieren Textbausteine, leiten relevante Informationen und Kompetenzen ab, transformieren diese in Embeddings, definieren Zielsetzungen und bereiten Ergebnisse adressatengerecht auf. Dabei erkennen sie die Bedeutung der Rollen im Zusammenspiel und reflektieren die Potenziale und Herausforderungen der KI-gestützten Bewerberauswahl.

Rolle: Inputmanagement

Aufgabe und Bedeutung: Das Input-Management spielt eine zentrale Rolle im ersten Schritt der Verarbeitung von Text durch ein Large Language Model (LLM). Dessen Aufgabe ist es, den Rohtext in kleinere Einheiten, sogenannte Tokens, zu zerlegen. Diese Tokens können einzelne Wörter, Satzzeichen oder sogar Wortteile sein, je nach Komplexität des Modells. Die Zerlegung eines Textes in diese Tokens ist wichtig, weil das LLM mit dieser kleinsten Einheit arbeitet und die Bedeutung auf dieser Ebene lernt.

Im realen Kontext bedeutet das, dass das LLM ohne Input-Management nicht "wüsste", mit welchen Einheiten es arbeiten muss. Die korrekte Zerlegung des Textes legt den Grundstein dafür, dass das Modell später Beziehungen zwischen Wörtern herstellen kann.

Beispiel 1:

1. Eingabetext: "Der Hund bellt."
2. Tokenisierung: ["Der", "Hund", "bellt", "."] Das Input-Management trennt jedes Wort und das Satzzeichen in einzelne Einheiten, sodass das Modell den Satz analysieren kann.

Beispiel 2:

3. Eingabetext: "Katzenfreunde!"
4. Tokenisierung: ["Katzen", "freunde", "!"] Hier wird das Wort „Katzenfreunde“ in zwei Teile zerlegt, da das Modell möglicherweise lernt, dass „Katzen“ und „Freunde“ unterschiedliche Bedeutungen haben, die zusammen einen neuen Sinn ergeben.

Analysieren Sie den folgenden Satz und zerlegen Sie ihn in Tokens. Berücksichtigen Sie dabei die verschiedenen Wortarten und füllen Sie die untenstehende Tabelle.

Satz: „Die großartigen Katzenfreunde spielen fröhlich im Garten.“

Token	Wortart	Anzahl der Buchstaben
Die (Beispiel)	Artikel	3

Rolle: Embedding

Aufgabe und Bedeutung: Embedding- Spezialistinnen und Spezialisten sind dafür verantwortlich, die Tokens (z.B. Wörter oder Wortteile) in numerische Vektoren zu übersetzen, sogenannte **Embeddings**. Diese Vektoren repräsentieren die Bedeutung der Wörter in einer mathematischen Form, sodass das LLM diese „verstehen“ und weiterverarbeiten kann. Der Embedding-Spezialist übersetzt also Sprache in Zahlen, die das Modell zur Analyse der Bedeutung verwendet.

Ohne die Embeddings könnte das Modell nicht erkennen, wie Wörter zueinander in Beziehung stehen. Diese Vektoren tragen wesentlich dazu bei, dass das Modell semantische Ähnlichkeiten oder Unterschiede zwischen Wörtern „begrift“. Ein gut trainiertes Modell wird ähnliche Bedeutungen von Wörtern durch ähnliche Zahlenmuster (Vektoren) darstellen.

Beispiel 1:

1. Token: „Katze“
2. Embedding: [0.8, 0.1, 0.3, 0.5] Das Wort „Katze“ wird in einen Vektor übersetzt. Wörter mit ähnlicher Bedeutung, wie „Hund“, hätten einen Vektor, der in einem ähnlichen Bereich liegt.

Beispiel 2:

1. Token: „laufen“
2. Embedding: [0.3, 0.6, 0.7, 0.2] Ein Verb wie „laufen“ hätte einen Vektor, der sich von Substantiven wie „Katze“ unterscheidet, aber Verben wie „rennen“ oder „gehen“ hätten ähnliche Vektoren.

Beispiel mit 3 Dimensionen:

Beispieltabelle	Tier	Größe	Lebensraum	Fortbewegungsgeschwindigkeit
	Blauwal	1.0	1.0	0.3
	Karpfen	0.2	0.0	0.2
	Blauhai	0.6	1.0	0.8
	Goldfisch	0.1	0.0	0.1

Aufgabe für die Rolle Embedding:

Erstellen Sie eine Embedding-Matrix für die folgenden Tokens, die nicht nur die Ähnlichkeit, sondern auch die semantischen Beziehungen zwischen ihnen berücksichtigt. Verwenden Sie 3 Dimensionen für die Vektoren und ordne Werte von 0 bis 1 zu. Berücksichtigen Sie auch, dass ähnliche Bedeutungen ähnliche Werte in den Vektoren haben sollten.

Tokens: „Katze“, „Hund“, „laufen“, „springen“, „Vogel“

Token	Dimension und Wert		
Katze			
Hunde			
Vogel			
Token	Dimension und Wert		
laufen			
springen			

Rolle: Selbstaufmerksamkeitsmechanik (Attentionmanagement)

Aufgabe und Bedeutung: Attention-Managerinnen und Manager überwachen die **Aufmerksamkeit** des Modells. Dies ist der Teil des Modells, der festlegt, welche Wörter in einem Satz für die Interpretation besonders wichtig sind. In der Selbstaufmerksamkeit berechnet das Modell, wie stark ein Wort auf ein anderes „aufmerksam“ sein sollte, um die Bedeutung richtig zu erfassen. Es geht also darum, festzustellen, wie Wörter miteinander in Beziehung stehen und welche Wörter besonders relevant für das Verständnis eines Satzes sind.

Das Modell könnte zum Beispiel feststellen, dass in dem Satz „Die Katze sitzt auf der Matte“ die Wörter „Katze“ und „Matte“ wichtiger sind als „der“ oder „auf“, da sie die Hauptbedeutung des Satzes tragen.

Beispiel 1:

1. Satz: „Der Hund bellt laut.“
2. Wichtige Wörter: „Hund“ und „bellt“ sind hier zentral, da sie die Hauptinformation des Satzes tragen. Das Wort „laut“ ist weniger wichtig, aber es beeinflusst die Bedeutung.

Beispiel 2:

1. Satz: „Die Sonne scheint hell.“
2. Das Modell könnte seine Aufmerksamkeit auf „Sonne“ und „scheint“ fokussieren, während „hell“ eine zusätzliche Information liefert.

Analysieren Sie den Satz „Die Katze jagt die Maus und der Hund beobachtet.“ und weisen Sie den Wörtern Aufmerksamkeitswerte zu. Berücksichtigen Sie dabei den Kontext und die Rolle jedes Wortes in Bezug auf die zentrale Handlung des Satzes. Erstelle eine Tabelle mit Token, Aufmerksamkeitswert (0 bis 1.0) und einer kurzen Begründung für den Wert.

Token	Aufmerksamkeit	Begründung
Die (Beispiel)	0.1	Artikel hat geringe Relevanz für die Hauptaussage.
Katze		
jagt		
die		
Maus		
und		
der		
Hund		
beobachtet		

Rolle: Vorhersage/Output Management

Aufgabe und Bedeutung: Die Vorhersage-Managerinnen und Manager haben die Aufgabe, das nächste Wort in einem Satz oder Textabschnitt vorherzusagen. Nachdem die Tokens durch den Input-Manager zerlegt, durch den Embedding-Spezialisten in Vektoren übersetzt und durch den Attention-Manager analysiert wurden, trifft der Vorhersage-Manager die Entscheidung, welches Wort wahrscheinlich als Nächstes folgt. Dies geschieht, indem er die Wahrscheinlichkeiten für mögliche nächste Wörter berechnet.

Das Vorhersage-Management ist besonders wichtig, weil es dafür sorgt, dass das Modell „flüssig“ sprechen oder Text generieren kann. Dessen Aufgabe ist es, basierend auf dem bisherigen Kontext die beste Vorhersage zu treffen.

Beispiel 1:

1. Satzbeginn: „Die Katze...“
2. Mögliche Vorhersagen: „läuft“, „schläft“, „springt“. Der Vorhersage-Manager entscheidet auf Basis der bisherigen Wörter und der Beziehung zwischen ihnen, dass „schläft“ am wahrscheinlichsten als Nächstes kommt.

Beispiel 2:

1. Satzbeginn: „Der Vogel...“

Mögliche Vorhersagen: „fliegt“, „singt“, „ruht“. Basierend auf dem Kontext entscheidet der Vorhersage-Manager, dass „fliegt“ das wahrscheinlichste nächste Wort ist.

Analysiere Sie den Satzanfang „Die Katze sitzt auf dem“ und formulieren Sie drei plausible Vorhersagen für das nächste Wort, wobei Sie auch jeweils die Wahrscheinlichkeit, mit der dieses Wort erwartet wird, angeben sollen (von 0 bis 1). Erklären Sie die Wahl Ihrer Vorhersagen.

Die Katze sitzt auf dem...	Wahrscheinlichkeit	Erklärung

Weiterführende Aufgabe für Expertenkombi

Aufbauend auf Ihrer letzten Analyse, bei der Sie einfache Sätze mit Tieren verarbeitet haben, sollen Sie nun an einem Large Language Model zur Bewertung von Bewerbern für eine Position im Personalmanagement eines Unternehmens arbeiten. Ihr Ziel ist es, anhand von Textbausteinen aus Bewerbungsunterlagen eine KI-gestützte Analyse zu simulieren, die **Soft Skills** erkennt und die Eignung eines Kandidaten einschätzt.

Textbausteine:

1. „Frau Dr. Becker spielte mehrere Jahre als Spielführerin in der Frauenfußballmannschaft unserer Klinik.“
2. „Herr Schmidt errang mehrere bayerische Meister- und Vizemeistertitel im Schachspiel.“

Bearbeiten Sie in Ihren jeweiligen Rollen folgende Punkte und arbeiten Sie als Gruppe zusammen, um die Ergebnisse zu erzeugen:

1. Input-Management:

Zerlegen Sie die bereitgestellten Textbausteine in Tokens. Achten Sie darauf, wie die KI wichtige Informationen zu Soft Skills aus diesen Tokens extrahieren kann. Geben Sie an, welche Tokens besonders relevant für die weitere Analyse sind.

2. Embedding:

Wandeln Sie besonders relevante Tokens in Vektoren um? Erklären Sie, wie diese Vektoren Informationen über Soft Skills enthalten.

3. Intention-Management:

Welches Ziel soll die KI erreichen, und welche Schlüsselinformationen wären besonders wichtig? Geben Sie ein Beispiel für eine zu priorisierende Aussage.

4. Vorhersage- / Outputmanagement

Entwerfen Sie eine einfache Darstellungsweise (z. B. eine Übersichtstabelle oder eine Stärken-/Schwächen-Analyse) und erläutern Sie, wie dies der Personalabteilung hilft.

Bearbeitungszeit ca. 15 Minuten

Lösungsbeispiele:

Rolle Inputmanagement:

Token	Wortart	Anzahl der Buchstaben
Die	Artikel	3
großartigen	Adjektiv	11
Katzen	Substantiv	6
freunde	Substantiv	6
spielen	Verb	7
fröhlich	Adverb	9
im	Präposition	2
großen	Adjektiv	6
Garten	Substantiv	6

Rolle Embedding:

Token	Dimension und Wert		
	Geschwindigkeit.	Größe	Carnifor
Katze	0.4	0.3	1.0
Hunde	0.4	0.5	1.0
Vogel	0.6	0.1	0.0

	Dimension	Geschwindigkeit	
laufen	0.6	0.7	
springen	0.8	0.8	

Rolle Selbstaufmerksamkeitsmechanik:

Token	Aufmerksamkeitswert	Begründung
Die	0.1	Artikel hat geringe Relevanz für die Hauptaussage.
Katze	0.6	Zentral für die Handlung, da sie die jagende Figur ist.
jagt	0.8	Wichtig, da es die Hauptaktion beschreibt.
die	0.1	Geringe Relevanz, wie bei „Die“.
Maus	0.7	Zentrale Figur als Ziel des Jagens.
und	0.05	Verbindet die Teilsätze, hat aber geringe Relevanz.
der	0.1	Artikel hat geringe Relevanz.
Hund	0.4	Beobachtet und ist relevant, aber nicht aktiv.
beobachtet	0.5	Wichtig für den Kontext, trägt zur Handlung bei.

Rolle Vorhersage Manager:

- „Sofa“ (Wahrscheinlichkeit: 0.6)
Erklärung: Katzen sitzen oft auf Möbeln wie Sofas.
- „Boden“ (Wahrscheinlichkeit: 0.3)
Erklärung: Eine plausible Möglichkeit, da Katzen oft auf dem Boden sitzen.
- „Fensterbrett“ (Wahrscheinlichkeit: 0.1)
Erklärung: Weniger wahrscheinlich, da dies nicht die häufigste Position ist, aber immer noch eine Möglichkeit.

Lösungsbeispiel weiterführende Expertenaufgabe:

1. Input-Management: Tokenzerlegung

Die Aufgabe des Input-Managers besteht darin, die bereitgestellten Textbausteine in Tokens zu zerlegen und relevante Tokens zu identifizieren.

Tokenzerlegung:

Textbaustein 1:

„Frau Dr. Becker spielte mehrere Jahre als Spielführerin in der Frauenfußballmannschaft unserer Klinik.“

- Tokens: „Frau“, „Dr.“, „Becker“, „spielte“, „mehrere“, „Jahre“, „als“, „Spielführerin“, „in“, „der“, „Frauenfußballmannschaft“, „unserer“, „Klinik“.
- **Relevante Tokens:**
 - „Spielführerin“ (Führungsqualität, Teamfähigkeit)
 - „Frauenfußballmannschaft“ (Engagement im Team)
 - „Klinik“ (Bezug zum Gesundheitswesen)

Textbaustein 2:

„Herr Schmidt errang mehrere bayerische Meister- und Vizemeistertitel im Schachspiel.“

- Tokens: „Herr“, „Schmidt“, „errang“, „mehrere“, „bayerische“, „Meister-“, „und“, „Vizemeistertitel“, „im“, „Schachspiel“.
- **Relevante Tokens:**
 - „Schachspiel“ (strategisches Denken, Problemlösungsfähigkeit)
 - „Meistertitel“, „Vizemeistertitel“ (Wettbewerbsfähigkeit, Exzellenz)

Zusammenfassung des Inputs:

Wichtige Kompetenzen: Führungsqualität, Teamfähigkeit, strategisches Denken, Problemlösungsfähigkeit, Engagement.

2. Embedding: Umwandlung in Vektoren

Der Embedding-Spezialist übersetzt die relevanten Tokens in Vektoren und erläutert, wie sie die enthaltenen Informationen repräsentieren.

Relevante Tokens und Beispielvektoren:

- „Spielführerin“: 0.7,0.3,0.50.7, 0.3, 0.5 (repräsentiert Führung, Organisation, Teamarbeit)
- „Schachspiel“: 0.9,0.6,0.80.9, 0.6, 0.8 (repräsentiert Strategie, Problemlösung, Logik)
- „Meistertitel“: 0.8,0.4,0.90.8, 0.4, 0.9 (repräsentiert Exzellenz, Erfolg, Zielorientierung)

Erläuterung:

Die Vektoren codieren spezifische Eigenschaften der Tokens (z. B. Soft Skills) und ermöglichen der KI, Zusammenhänge zwischen ähnlichen Eigenschaften zu erkennen. Zum Beispiel könnten „Spielführerin“ und „Teamarbeit“ ähnliche Werte in der Dimension „Teamfähigkeit“ haben.

3. Intention-Management: Zieldefinition

Der Intention-Manager definiert das Ziel der KI und identifiziert Schlüsselinformationen.

Ziel der KI:

Die KI soll Bewerber anhand von Soft Skills bewerten, um passende Kandidaten für Führungs- oder Fachpositionen zu identifizieren.

Schlüsselinformationen und Prioritäten:

- **Führungsqualität:** Tokens wie „Spielführerin“ oder „Leiter“.
- **Strategisches Denken:** Tokens wie „Schachspiel“ oder „Planung“.
- **Erfolgsorientierung:** Tokens wie „Meistertitel“ oder „Auszeichnung“.

Beispiel:

Eine Aussage wie „Frau Dr. Becker war Spielführerin“ sollte eine hohe Priorität haben, wenn die ausgeschriebene Position Führungsqualitäten erfordert.

4. Vorhersage-/Output-Management: Ergebnisdarstellung

Der Output-Manager erstellt eine Übersicht der relevanten Ergebnisse und deren Nutzen für die Personalabteilung.

Ergebnisübersicht (Beispieltabelle):

Kandidat	Führungsqualität	Teamfähigkeit	Strategisches Denken	Erfolgsorientierung
Frau Dr. Becker	Hoch	Hoch	Mittel	Mittel
Herr Schmidt	Niedrig	Mittel	Hoch	Hoch

Erläuterung:

Die Tabelle zeigt, wie die KI die Bewerber bewertet. Personalverantwortliche können diese Informationen nutzen, um Kandidaten gezielt für Stellen auszuwählen, die ihre Stärken optimal nutzen.

Zusammenfassung

Die Aufgabe fordert die Gruppenmitglieder, ihre spezifischen Rollen zu nutzen, um die bereitgestellten Textbausteine zu analysieren, in Embeddings zu übersetzen, Intentionen zu definieren und Ergebnisse aufzubereiten. So wird das Zusammenspiel der Rollen in einem LLM praxisnah vermittelt.