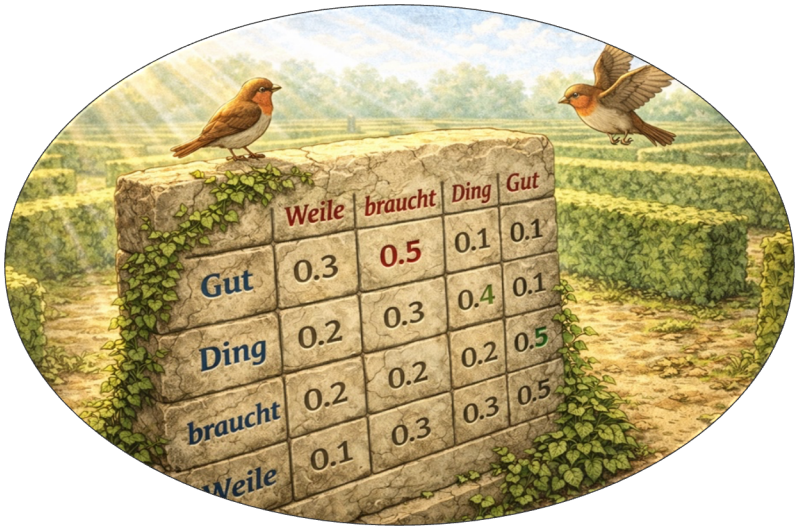


Lehrerhandreichung zu “Attention-Matrix unplugged”

Begleitdokument zu Arbeitsblatt I (Beispielsatz) und Arbeitsblatt II (generisch)

Ziel der Unterrichtseinheit: Die Lernenden verstehen das Grundprinzip des Attention-Mechanismus: Jedes Wort in einem Satz »schaut« auf alle anderen Wörter und gewichtet sie nach Relevanz. So entsteht die Attention-Matrix. Mehrere Heads tun dies gleichzeitig aus unterschiedlichen Perspektiven.
Dies ist eine didaktische Vereinfachung, kein echtes Modell.

1. Einordnung im Curriculum

Die Arbeitsblätter sind als Unplugged-Aktivität konzipiert. Sie setzen keine Programmierkenntnisse voraus und können in jeder Schulform eingesetzt werden, die sich mit Grundlagen von Sprachmodellen beschäftigt. Empfohlene Reihenfolge im Unterricht:

1. Vorab: Word-Vektor-Demo (Cosinus-Ähnlichkeit, Analogien) — schafft Verständnis für Wortdarstellungen als Vektoren
2. Arbeitsblatt I: Beispielsatz „Gut Ding braucht Weile“ — Einführung in die Attention-Matrix
3. Arbeitsblatt II: Generisches Blatt — Übertragung auf eigene Sätze
4. Arbeitsblatt III: V-Vektoren und Head-Output — Vertiefung

2. Das Punktesystem: Modell der Softmax-Normierung

Die Schülerinnen und Schüler verteilen in jeder Zeile genau 10 Punkte. Das modelliert die Softmax-Funktion im echten Transformer:

- Jede Zeile summiert sich auf denselben Gesamtwert (hier: 10, im Modell: 1)
- Die Verteilung ist kontextabhängig — jede Zeile kann ganz anders aussehen

Hinweis: In Wirklichkeit werden die Scores nicht per Hand vergeben, sondern aus Q- und K-Vektoren berechnet (Dot-Product, dann Softmax). Im echten Modell lernt jeder Head seine Spezialisierung selbst.

3. Die drei Heads: didaktische Begründung

Die drei Blickrichtungen sind didaktisch gewählt, um das Modell mehrerer paralleler Heads zu motivieren:

Head	Blickrichtung	Didaktische Funktion
Head A	Struktur	Zeigt syntaktische Bezüge — welche Wörter gehören grammatikalisch zusammen?
Head B	Bedeutung	Zeigt semantische Bezüge — welche Wörter hängen inhaltlich zusammen?
Head C	Schwerpunkt	Zeigt pragmatische Relevanz — welche Wörter tragen die Pointe?

Wichtig: Im echten Transformer sind Head-Spezialisierungen nicht vorgegeben. Sie entstehen durch Training. Die drei Perspektiven dienen hier als Gerüst für das Denken — nicht als Modellarchitektur.

4. Empfohlene Sätze und Erfahrungswerte

Die folgenden vier Sätze eignen sich besonders als Beispiele für die generische Version (Arbeitsblatt II):

Satz	Didaktischer Schwerpunkt
»Gut Ding braucht Weile«	Zwei starke Wortpaare; Pointe liegt klar auf „braucht Weile“
»Dieses Buch erklärt vieles«	Subjekt-Verb-Objekt klar, Fokus auf „erklärt“
»Worte wirken oft nach«	Temporaladverb „oft“ interessant für Head A vs. C
»Auch leise Worte treffen«	Adverb „leise“ wirkt semantisch verstärkend auf „treffen“
»Nicht alles ist Gold«	Negation „Nicht“ verändert den Bedeutungsrahmen des gesamten Satzes

Alle Musterlösungen zu diesen Sätzen befinden sich auf Seite 2 von Arbeitsblatt II.

5. Durchführungshinweise

Einzelarbeit oder Partnerarbeit (ca. 15–20 Minuten):

- Jede/r füllt die drei Matrizen für Head A, B und C aus.
- Dabei müssen die Zeilensummen jeweils 10 ergeben.
- Es gibt keine einzig richtige Lösung — wichtig ist die Begründung.

Plenumsdiskussion (ca. 10 Minuten):

- Welche Wörter bekamen in Head A oft viele Punkte?
- Wo unterscheiden sich Head A, B und C deutlich?
- Was zeigt das darüber, warum mehrere Heads sinnvoll sind?

6. Differenzierung und Erweiterung

Erweiterung für leistungsstarke Gruppen:

- Lassen Sie Gruppen eigene 4-Wort-Sätze wählen und die Matrix befüllen.
- Diskussion: Kann man aus der Matrix ablesen, ob ein Satz eher „struturiert“ oder „kontextuell“ ist?

Vereinfachung für Einsteiger:

- Nur Head A und Head C (Struktur und Schwerpunkt) bearbeiten.
- Erst gemeinsam als Klasse eine Zeile ausfüllen, dann selbständig weiter.

7. Kernerkenntnis für Lehrkräfte

Merksätze für den Unterricht

- Ein Satz enthält mehrere wichtige Beziehungen gleichzeitig. Darum sind mehrere Attention-Heads sinnvoll.
- Die Attention-Matrix beantwortet: Worauf achtet dieses Wort? — Noch nicht: Was übernimmt es von dort?
- Die Antwort auf die zweite Frage liefern die V-Vektoren — Thema von Arbeitsblatt II.

Arbeitsblatt 1: Attention-Matrix – Unplugged

Beispielsatz: Gut Ding braucht Weile

Worum geht es?

Ein Sprachmodell liest einen Satz nicht Wort für Wort. Es prüft bei jedem Wort: Auf welche anderen Wörter sollte ich besonders achten?

So entsteht eine **Attention-Matrix**. Darin sieht man, wie stark jedes Wort auf die anderen Wörter schaut.

Unser Satz: Gut | Ding | braucht | Weile

Arbeitsregel

Jedes Wort verteilt in seiner Zeile insgesamt **10 Punkte** auf alle Wörter im Satz.

Viele Punkte = sehr wichtig | *Wenige Punkte = eher unwichtig*

⚠ **In jeder Zeile müssen zusammen genau 10 Punkte stehen.**

Drei Blickrichtungen

Head A

Struktur

Frage: Welche Wörter gehören im Satzbau eng zusammen, auch wenn sie inhaltlich nicht die Hauptidee tragen?

Head B

Bedeutung

Frage: Welche Wörter hängen für die inhaltliche Bedeutung eng zusammen, auch wenn sie sprachlich nicht direkt gekoppelt sind?

Head C

Schwerpunkt

Frage: Welche Wörter tragen die Pointe des ganzen Satzes am stärksten?

So lest ihr die Matrix

Beispiel: In der Zeile 'braucht' verteilt ihr 10 Punkte auf Gut, Ding, braucht und Weile. Die Frage lautet: Auf welche Wörter schaut 'braucht' besonders stark? Ein Wort darf auch auf sich selbst schauen.

Aufgabe 1 – Head A: Struktur

Q-Worte \ K-Worte	Gut	Ding	braucht	Weile
Gut				
Ding				
braucht				
Weile				

Aufgabe 2 – Head B: Bedeutung

Q-Worte \ K-Worte	Gut	Ding	braucht	Weile
Gut				
Ding				
braucht				
Weile				

Aufgabe 3 – Head C: Schwerpunkt

Q-Worte \ K-Worte	Gut	Ding	braucht	Weile
Gut				
Ding				
braucht				
Weile				

Regeln

- In jeder Zeile müssen zusammen genau 10 Punkte stehen.
- Ein Wort darf auch auf sich selbst schauen.
- Die Werte müssen nicht spiegelbildlich sein.
- Es gibt nicht nur eine richtige Lösung – wichtig ist, dass ihr eure Verteilung begründen könnt.

Austausch in der Gruppe

1. Welche Wörter bekamen oft viele Punkte?
2. Wo unterscheiden sich Head A, Head B und Head C deutlich?
3. Was zeigt das über Attention?

Merksatz:
Ein Satz enthält mehrere wichtige Beziehungen gleichzeitig. Darum sind mehrere Attention-Heads sinnvoll.

Lehrerhinweise & Musterlösungen

Dieses Blatt enthält Hinweise zur didaktischen Einbettung sowie plausible Musterverteilungen. Es gibt keine einzig richtige Lösung – entscheidend ist, dass die Verteilung zur jeweiligen Blickrichtung passt.

Hinweis: Modell – kein echtes System

Diese Übung ist eine didaktische Vereinfachung. Die Klasse verteilt Punkte per Hand. Ein echtes Modell berechnet Gewichte aus Q-, K- und V-Vektoren. In Wirklichkeit werden die Scores nicht per Hand vergeben, sondern aus Q- und K-Vektoren berechnet. *Im echten Modell lernt jeder Head seine Spezialisierung selbst .*

Unterschied der drei Heads

Head	Kurz erklärt	Leitfrage
Head A – Struktur	Blick auf Satzbau und sprachliche Verbindung	Welche Wörter gehören im Satzbau eng zusammen?
Head B – Bedeutung	Blick auf den Sinn einzelner Wörter und Wortgruppen	Welche Wörter helfen zusammen, die Aussage zu verstehen?
Head C – Schwerpunkt	Blick auf die Hauptaussage des ganzen Satzes	Worauf läuft der Satz hinaus?

Plausible Musterverteilungen Die folgenden Werte zeigen eine Tendenz, keine Einheitslösung.

Head A – Struktur

schaut auf →	Gut	Ding	braucht	Weile
Gut	2	6	1	1
Ding	5	2	2	1
braucht	1	2	2	5
Weile	1	1	6	2

Head B – Bedeutung

schaut auf →	Gut	Ding	braucht	Weile
Gut	2	4	2	2
Ding	3	2	2	3
braucht	1	2	3	4
Weile	1	2	4	3

Head C – Schwerpunkt

schaut auf →	Gut	Ding	braucht	Weile
Gut	1	2	3	4
Ding	1	2	3	4
braucht	1	1	4	4
Weile	1	1	4	4

Impulsfragen für die Besprechung

- Welche Zeilen waren leicht, welche schwer zu befüllen?
- Wo sieht Head A anders aus als Head B oder Head C?
- Warum rücken bei Head C bestimmte Wörter besonders in den Vordergrund?
- Was bedeutet es, dass jeder Head eine andere Brille aufhat?

Attention-Matrix – Unplugged (Du wählst deinen Satz selbst)

Worum geht es?

Ein Sprachmodell liest einen Satz nicht einfach nur Wort für Wort. Es prüft bei jedem Wort: Auf welche anderen Wörter sollte ich besonders achten?

So entsteht eine **Attention-Matrix**. Darin sieht man, wie stark jedes Wort auf die anderen Wörter schaut.

Unser Satz / Spruch: *Wort 1 | Wort 2 | Wort 3 | Wort 4*

Arbeitsregel

Jedes Wort verteilt in seiner eigenen Zeile insgesamt **10 Punkte** auf alle vier Wörter.

Viele Punkte = dieses andere Wort ist gerade wichtig | Wenige Punkte = dieses Wort ist gerade weniger wichtig

⚠ **In jeder Zeile müssen zusammen genau 10 Punkte stehen.**

Drei Blickrichtungen

Head A — **Struktur**

Frage: Welche Wörter gehören im Satzbau eng zusammen, auch wenn sie inhaltlich nicht die Hauptidee tragen?

Head B — **Bedeutung**

Frage: Welche Wörter hängen für die inhaltliche Bedeutung eng zusammen, auch wenn sie sprachlich nicht direkt gekoppelt sind?

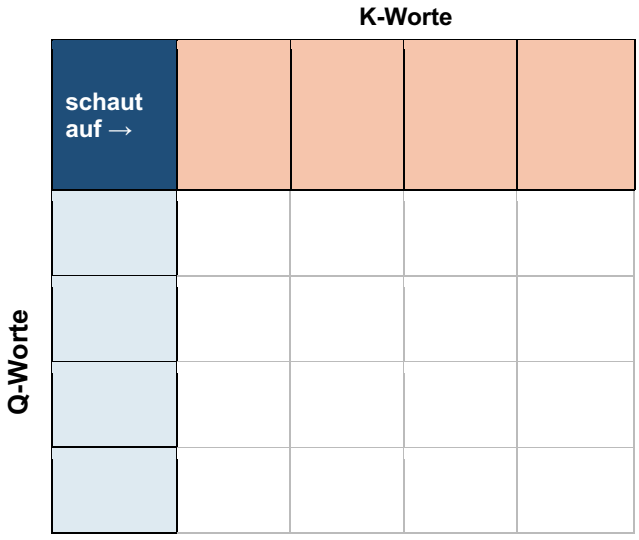
Head C — **Schwerpunkt**

Frage: Welche Wörter tragen die Pointe des ganzen Satzes am stärksten?

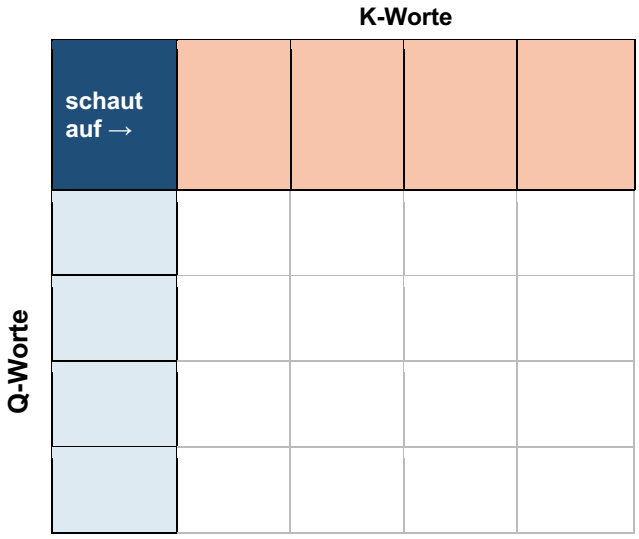
So arbeitet ihr

Bearbeitet die Matrix zeilenweise. Frage pro Zeile: Auf welche Wörter schaut dieses Wort besonders stark? Ein Wort darf auch auf sich selbst schauen. Die Matrix muss nicht spiegelbildlich sein. Es gibt nicht nur eine richtige Lösung – wichtig ist eure Begründung.

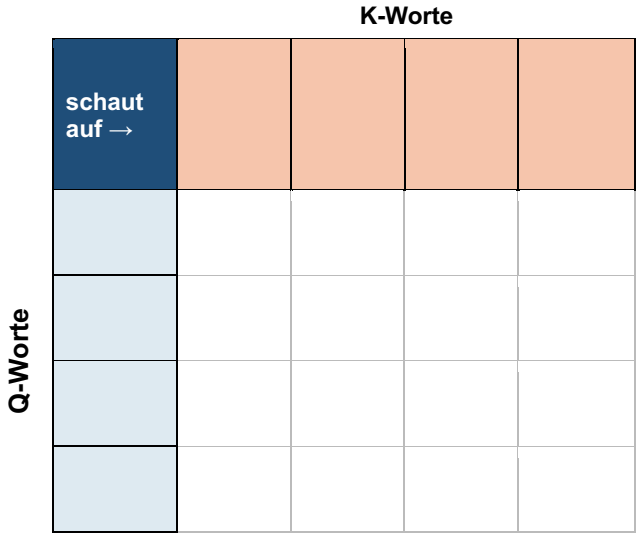
Aufgabe 1 – Head A: Struktur



Aufgabe 2 – Head B: Bedeutung



Aufgabe 3 – Head C: Schwerpunkt



Regeln

- In jeder Zeile müssen zusammen genau 10 Punkte stehen.
- Ein Wort darf auch auf sich selbst schauen.
- Die Werte müssen nicht spiegelbildlich sein.
- Es gibt nicht nur eine richtige Lösung – wichtig ist, dass ihr eure Verteilung begründen könnt.

Austausch in der Gruppe

1. Welche Wörter bekamen oft viele Punkte?
2. Wo unterscheiden sich Head A, Head B und Head C deutlich?
3. Was zeigt das über Attention?

Merksatz:
Ein Satz enthält mehrere wichtige Beziehungen gleichzeitig. Darum sind mehrere Attention-Heads sinnvoll.

Lehrerhinweise – Musterlösungen für alle Beispielsätze

Hinweis: Modell – kein echtes System

Diese Übung ist eine didaktische Vereinfachung. Die Klasse verteilt Punkte per Hand. Ein echtes Modell berechnet Gewichte aus Q-, K- und V-Vektoren. In Wirklichkeit werden die Scores nicht per Hand vergeben, sondern aus Q- und K-Vektoren berechnet. *Im echten Modell lernt jeder Head seine Spezialisierung selbst .*

Die folgenden Verteilungen zeigen plausible Tendenzen, keine Einheitslösungen. Notation: Jede Zeile summiert sich auf 10. Reihenfolge der Spalten immer: Wort 1 | Wort 2 | Wort 3 | Wort 4.

1. »Gut Ding braucht Weile«

Head A – Struktur

Zwei starke Paare: Gut–Ding und braucht–Weile.

schaut	Gut	Ding	braucht	Weile
Gut	2	6	1	1
Ding	5	2	2	1
braucht	1	2	2	5
Weile	1	1	6	2

Head B – Bedeutung

Inhaltlich etwas breiter verteilt, aber braucht–Weile bleibt stark

schaut	Gut	Ding	braucht	Weile
Gut	2	4	2	2
Ding	3	2	2	3
braucht	1	2	3	4
Weile	1	2	4	3

Head C – Schwerpunkt

Fokus auf braucht und Weile als eigentliche Pointe.

schaut	Gut	Ding	braucht	Weile
Gut	1	2	3	4
Ding	1	2	3	4
braucht	1	1	4	4
Weile	1	1	4	4

2. »Dieses Buch erklärt vieles«

Head A – Struktur

Dieses–Buch und erklärt–vieles sind die stärksten sprachlichen Verbindungen.

schaut	Dieses	Buch	erklärt	vieles
Dieses	2	6	1	1
Buch	5	2	2	1
erklärt	1	2	2	5
vieles	1	1	6	2

Head B – Bedeutung

Buch und erklärt hängen inhaltlich besonders eng zusammen.

schaut	Dieses	Buch	erklärt	vieles
Dieses	2	4	2	2
Buch	2	2	4	2
erklärt	1	4	2	3
vieles	1	2	5	2

Head C – Schwerpunkt

Fokus auf erklärt als Kern des Satzes, vieles als Ziel der Erklärung.

schaut	Dieses	Buch	erklärt	vieles
Dieses	1	2	5	2
Buch	1	2	5	2
erklärt	1	1	5	3
vieles	1	1	6	2

.

3. »Worte wirken oft nach«

Head A – Struktur

Worte–wirken und oft–nach bilden die stärksten Strukturbezüge.

schaut	Worte	wirken	oft	nach
Worte	2	6	1	1
wirken	4	2	2	2
oft	1	2	2	5
nach	1	2	5	2

Head B – Bedeutung

Worte–wirken–nach greifen inhaltlich stark ineinander.

schaut	Worte	wirken	oft	nach
Worte	2	4	1	3
wirken	3	2	1	4
oft	1	3	2	4
nach	2	3	2	3

Head C – Schwerpunkt

Schwerpunkt auf wirken und nach, weil da die eigentliche Aussage liegt.

schaut	Worte	wirken	oft	nach
Worte	1	4	1	4
wirken	1	4	1	4
oft	1	3	2	4
nach	1	3	2	4

4. »Auch leise Worte treffen«

Head A – Struktur

leise–Worte und Worte–treffen sind die stärksten sprachlichen Bindungen.

schaut auf →	Auch	leise	Worte	treffen
Auch	2	4	2	2
leise	1	2	5	2
Worte	1	3	2	4
treffen	1	2	5	2

Head B – Bedeutung

leise und treffen hängen über die Aussage eng zusammen; Worte bleibt zentral.

schaut auf →	Auch	leise	Worte	treffen
Auch	1	3	3	3
leise	1	2	3	4
Worte	1	2	2	5
treffen	1	3	4	2

Head C – Schwerpunkt

Fokus auf Worte und treffen; dort liegt die Pointe.

schaut auf →	Auch	leise	Worte	treffen
Auch	1	1	4	4
leise	1	2	3	4
Worte	1	1	3	5
treffen	1	1	5	3

5. »Nicht alles ist Gold«

Head A – Struktur

Nicht–alles und ist–Gold sind die stärksten Strukturpaare.

schaut auf →	Nicht	alles	ist	Gold
Nicht	2	6	1	1
alles	5	2	2	1
ist	1	2	2	5
Gold	1	1	6	2

Head B – Bedeutung

Nicht verändert die Bedeutung des ganzen Satzes; Gold bleibt der zentrale Bedeutungsanker.

schaut auf →	Nicht	alles	ist	Gold
Nicht	2	3	2	3
alles	2	2	2	4
ist	1	2	3	4
Gold	2	2	3	3

Head C – Schwerpunkt

Fokus auf Nicht und Gold, weil dort Kontrast und Pointe sitzen.

schaut auf →	Nicht	alles	ist	Gold
Nicht	3	2	1	4
alles	3	2	1	4
ist	2	1	2	5
Gold	3	1	2	4