

SEMANTIC REPRESENTATION
P(next token | context):
= softmax ($W \cdot h_t$)
MEANING?
INTENT?

CONTEXT WINDOW
2048 / 8192 / ...

Lesen Menschen genauso wie KI-Modelle?

probability distribution

0.07
0.13
0.02
0.11
0.02

TOKEN #1523
ID: 39587

EMBEDDING VECTOR
[0.021, -0.884, 0.112, ...]

UNCERTAINTY ALIGNMENT

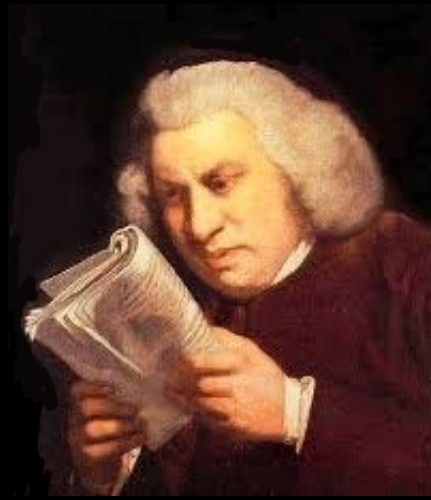
Dario Paape

LOW CONFIDENCE
0.18

TRAINING DATA DISTRIBUTION

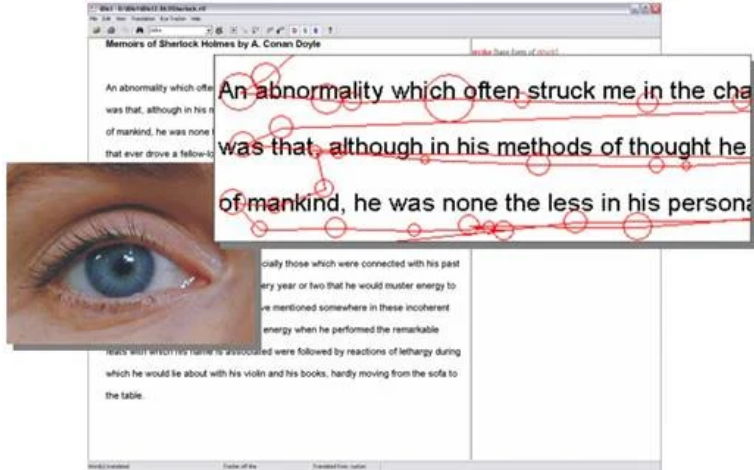
OUT OF DOMAIN?

*Ich weiß nicht,
was soll es bedeuten*



Lesezeitenmessung: Eye tracking

Lesezeitenmessung: Eye tracking



trialid	start	stop	xn	yn	dur	wordnum	word
1	1	164	56.0	54.0	164	1	In
1	186	351	180.4	105.9	166	2	der
1	378	596	378.5	125.0	219	4	römischen
1	627	770	191.2	101.1	144	3	alten
1	782	1052	145.0	86.7	271	2	der
1	1096	1246	538.6	122.0	151	5	Religion
1	1274	1566	344.1	129.7	293	4	römischen
1	1605	1885	602.3	121.0	281	6	und
1	1911	2062	747.9	104.8	152	7	Mythologie
1	2103	2261	444.1	123.3	159	5	Religion
1	2283	2374	528.9	121.6	92	5	Religion
1	2403	2585	376.7	127.2	183	4	römischen
1	2609	2898	263.9	121.6	290	3	alten
1	2934	3230	507.6	127.1	297	5	Religion
1	3254	3450	394.0	133.2	197	4	römischen

trialid	start	stop	xn	yn	dur	wordnum	word
1	1	164	56.0	54.0	164	1	In
1	186	351	180.4	105.9	166	2	der
1	378	596	378.5	125.0	219	4	römischen
1	627	770	191.2	101.1	144	3	alten
1	782	1052	145.0	86.7	271	2	der
1	1096	1246	538.6	122.0	151	5	Religion
1	1274	1566	344.1	129.7	293	4	römischen
1	1605	1885	602.3	121.0	281	6	und
1	1911	2062	747.9	104.8	152	7	Mythologie
1	2103	2261	444.1	123.3	159	5	Religion
1	2283	2374	528.9	121.6	92	5	Religion
1	2403	2585	376.7	127.2	183	4	römischen
1	2609	2898	263.9	121.6	290	3	alten
1	2934	3230	507.6	127.1	297	5	Religion
1	3254	3450	394.0	133.2	197	4	römischen

Dario mag Sprache und Computer

Dario mag Sprache und Computer findet er doof.

Dario mag Sprache und Computer findet er doof.



Dario mag Sprache und Computer findet er doof.



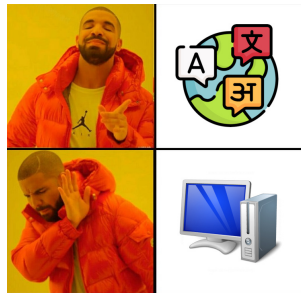
Dario mag **Sprache und Computer** findet er doof.



?



Dario mag Sprache und **Computer findet er doof.**



Zwei-Modi-Lesemodell

Zwei-Modi-Lesemodell (aus den 1980ern)



Zwei-Modi-Lesemodell (aus den 1980ern)

Zwei-Modi-Lesemodell (aus den 1980ern)

“Standard”-Modus: Erste Analyse bereits richtig 👁️ 👍

“Reanalyse”-Modus: Analyse muss korrigiert werden 👁️ 🛑 🐌

Zwei-Modi-Lesemodell (aus den 1980ern)

“Standard”-Modus: Erste Analyse bereits richtig 👁️👁️👍

“Reanalyse”-Modus: Analyse muss korrigiert werden 👁️👁️🚫🐌

- ◇ Unterschiedliches Leseverhalten je nach Modus

Zwei-Modi-Lesemodell (aus den 1980ern)

“Standard”-Modus: Erste Analyse bereits richtig 👁️👁️👍

“Reanalyse”-Modus: Analyse muss korrigiert werden 👁️👁️🚫🐌

- ◇ Unterschiedliches Leseverhalten je nach Modus
- ◇ Erste Analyse bedingt durch **sprachliche Gewohnheiten**

Zwei-Modi-Lesemodell (aus den 1980ern)

“Standard”-Modus: Erste Analyse bereits richtig 👁️👁️👍

“Reanalyse”-Modus: Analyse muss korrigiert werden 👁️👁️🚫🐌

- ◇ Unterschiedliches Leseverhalten je nach Modus
- ◇ Erste Analyse bedingt durch **sprachliche Gewohnheiten**
- ◇ **Statistik der Sprache:**

Zwei-Modi-Lesemodell (aus den 1980ern)

“Standard”-Modus: Erste Analyse bereits richtig 👁️👁️👍

“Reanalyse”-Modus: Analyse muss korrigiert werden 👁️👁️🚫🐌

- ◇ Unterschiedliches Leseverhalten je nach Modus
- ◇ Erste Analyse bedingt durch **sprachliche Gewohnheiten**
- ◇ **Statistik der Sprache:** “Wie oft steht an dieser Stelle ein Verb?”

Wie “messen” wir die Sprachstatistik?

Wie “messen” wir die Sprachstatistik?

◇ Früher:

Wie “messen” wir die Sprachstatistik?

- ◇ Früher: Auszählen in Textsammlungen

Wie “messen” wir die Sprachstatistik?

- ◇ Früher: Auszählen in Textsammlungen

NOMEN mag NOMEN und NOMEN _____

Wie "messe"

◇ Früher: Ausz

NOMEN mag N



Wie “messen” wir die Sprachstatistik?

- ◇ Früher: Auszählen in Textsammlungen

NOMEN mag NOMEN und NOMEN _____

- ◇ Heutzutage: **neuronale Sprachmodelle** 🤖

Wie “messen” wir die Sprachstatistik?

- ◇ Früher: Auszählen in Textsammlungen

NOMEN mag NOMEN und NOMEN _____

- ◇ Heutzutage: **neuronale Sprachmodelle** 🤖 (LLMs)

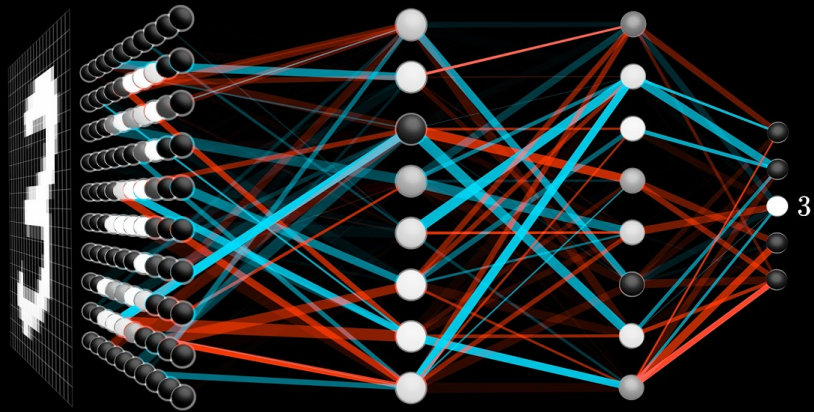
Wie “messen” wir die Sprachstatistik?

- ◇ Früher: Auszählen in Textsammlungen

NOMEN mag NOMEN und NOMEN _____

- ◇ Heutzutage: **neuronale Sprachmodelle** 🤖 (LLMs; z.B. GPT)

Neuronales Netz ("KI")



☐ Hide

Dates re

Rereading

Now you

Date sta

2024

Add re

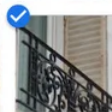
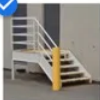
More deta



Save

Preview

Select all images with
stairs



VERIFY

API

ished to have it ad
optional)

mber 8

Post to blog

Sprachmodelle (LLMs) 🤖

Sprachmodelle (LLMs) 🤖

- ◇ Explizite Regeln?

Sprachmodelle (LLMs) 🤖

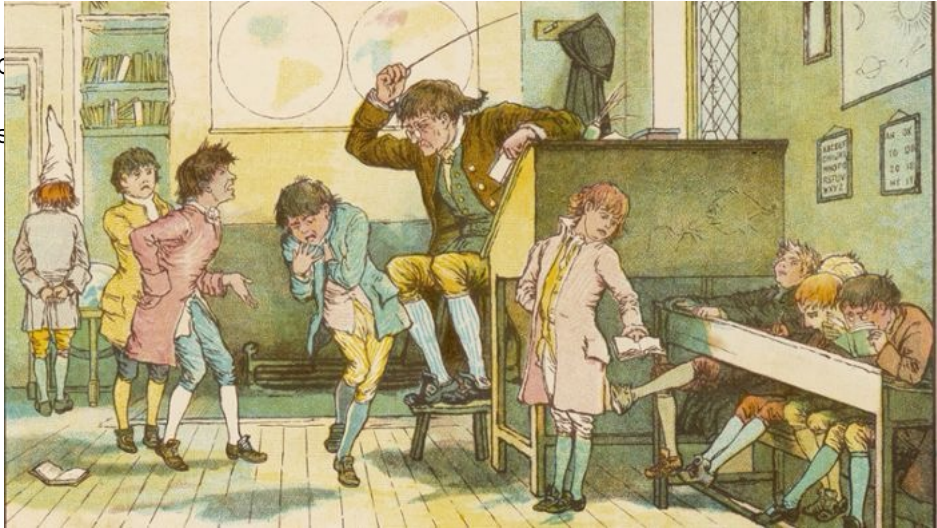
◇ Explizite Regeln?

Erst kommt das Subjekt, dann das Prädikat, dann das Objekt.

Sprachmodelle (LLMs) 🤖

◇ Exp

Ers



Sprachmodelle (LLMs) 🤖

- ◇ Explizite Regeln?

Erst kommt das Subjekt, dann das Prädikat, dann das Objekt.

- ◇ Selbstlernend:

Sprachmodelle (LLMs) 🤖

- ◇ Explizite Regeln?

Erst kommt das Subjekt, dann das Prädikat, dann das Objekt.

- ◇ Selbstlernend:

Hier ist das Internet;

Sprachmodelle (LLMs) 🤖

- ◇ Explizite Regeln?

Erst kommt das Subjekt, dann das Prädikat, dann das Objekt.

- ◇ Selbstlernend:

Hier ist das Internet; lerne selbst, welche Wörter häufig auf welche anderen Wörter folgen.

Sprachmodelle (LLMs) 🤖

◇ Explizite Regeln?

Erst kommt das Subjekt, dann das Prädikat, dann das Objekt.

◇ Selbstlernend:

Hier ist das Internet; lerne selbst, welche Wörter häufig auf welche anderen Wörter folgen.

◇ Sprache ist (zum Teil) **auch nur Statistik** 📖 📊

Sprachmodelle (LLMs) 🤖

◇ Explizite Regeln?

Erst kommt das Subjekt, dann das Prädikat, dann das Objekt.

◇ Selbstlernend:

Hier ist das Internet; lerne selbst, welche Wörter häufig auf welche anderen Wörter folgen.

◇ Sprache ist (zum Teil) **auch nur Statistik** 📖 📊

◇ Kinder lernen Sprache auch zuerst **ohne explizite Regeln** 🧒 + 📊 = 💡

Sprachmodelle (LLMs) 🤖

◇ Explizite Regeln?

Erst kommt das Subjekt, dann das Prädikat, dann das Objekt.

◇ Selbstlernend:

Hier ist das Internet; lerne selbst, welche Wörter häufig auf welche anderen Wörter folgen.

◇ Sprache ist (zum Teil) **auch nur Statistik** 📖 📊

◇ Kinder lernen Sprache auch zuerst **ohne explizite Regeln** 🧒 + 📊 = 💡

◇ Für ein LLM sind Wörter auch nur "Pixel"/Zahlenwerte 0 1 0 0

Mathematische Gewichtung vorheriger Wörter

Mathematische Gewichtung vorheriger Wörter

Vor

dem

Slam

hatte

Dario

Mathematische Gewichtung vorheriger Wörter

Vor

dem

Slam

hatte

Dario

Lampenfieber - 0.32

Muffensausen - 0.23

schlecht geschlafen - 0.09

Mathematische Gewichtung vorheriger Wörter

Vor

dem

Slam

hatte

Dario

Lampenfieber - 0.32

Muffensausen - 0.23

schlecht geschlafen - 0.09

Nach

dem

Slam

hatte

Dario

Mathematische Gewichtung vorheriger Wörter

Vor

dem

Slam

hatte

Dario

Lampenfieber - 0.32

Muffensausen - 0.23

schlecht geschlafen - 0.09

Nach

dem

Slam

hatte

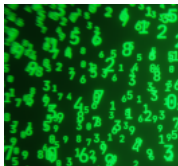
Dario

Schweißfüße - 0.12

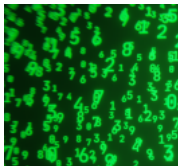
Glücksgefühle - 0.30

Durst - 0.15

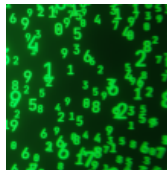
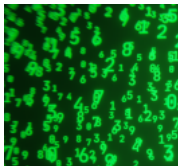
Nach $\times$ X +



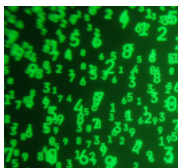
Nach $\times$ X + Slam $\times$ Y +



Nach $\times$ X + Slam $\times$ Y + hatte $\times$ Z



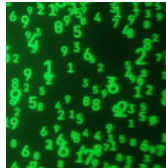
Nach $\times$ X + Slam $\times$ Y + hatte $\times$ Z = P(Schweißfüße)



+



+



=





**Perfekte Kenntnis der Sprachstatistik
kann Sprachverständnis vorgaukeln!**

Perfekte Kenntnis der Sprachstatistik kann Sprachverständnis vorgaukeln!





trialid	start	stop	xn	yn	dur	wordnum	word
1	1	164	56.0	54.0	164	1	In
1	186	351	180.4	105.9	166	2	der
1	378	596	378.5	125.0	219	4	römischen
1	627	770	191.2	101.1	144	3	alten
1	782	1052	145.0	86.7	271	2	der
1	1096	1246	538.6	122.0	151	5	Religion
1	1274	1566	344.1	129.7	293	4	römischen
1	1605	1885	602.3	121.0	281	6	und
1	1911	2062	747.9	104.8	152	7	Mythologie
1	2103	2261	444.1	123.3	159	5	Religion
1	2283	2374	528.9	121.6	92	5	Religion
1	2403	2585	376.7	127.2	183	4	römischen
1	2609	2898	263.9	121.6	290	3	alten
1	2934	3230	507.6	127.1	297	5	Religion
1	3254	3450	394.0	133.2	197	4	römischen

Dario mag Sprache und Computer **findet** er doof.

Was ist die Formel für menschliche Lesezeiten?

Was ist die **Formel** für menschliche Lesezeiten?

Lesezeit =

Was ist die **Formel** für menschliche Lesezeiten?

Lesezeit =



Wortlänge $\times$ A +

Was ist die Formel für menschliche Lesezeiten?

Les



Wo



Was ist die **Formel** für menschliche Lesezeiten?

Lesezeit =



Wortlänge $\times A +$



Alter $\times B +$

Was ist

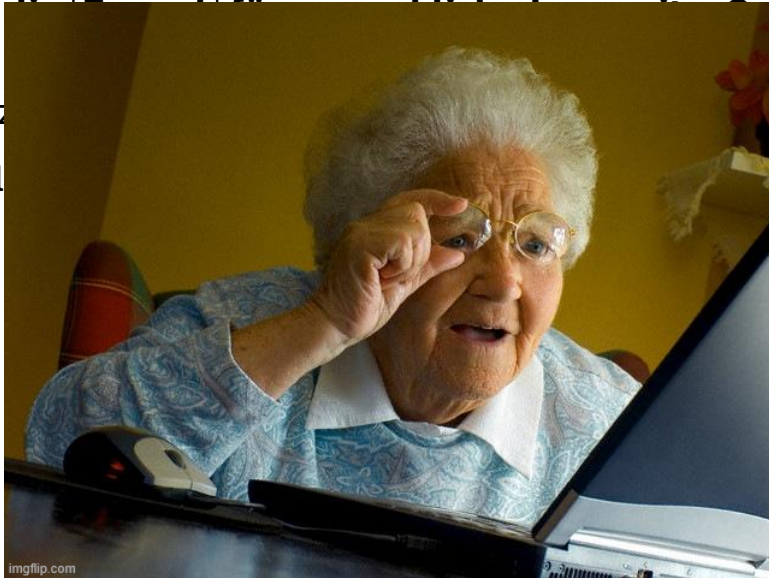
Lesez



Wortl



Alter



Was ist die **Formel** für menschliche Lesezeiten?

Lesezeit =



Wortlänge $\times A +$



Alter $\times B +$



Promille $\times C +$

Was ist die **Formel** für menschliche Lesezeiten?

Lese



Wort



Alter



Prom



Was ist die **Formel** für menschliche Lesezeiten?

Lesezeit =



Wortlänge $\times A +$



Alter $\times B +$



Promille $\times C +$



LLM-Wahrscheinlichkeit $\times F$

Was ist die **Formel** für menschliche Lesezeiten?

Lesezeit =



Wortlänge $\times A +$



Alter $\times B +$



Promille $\times C +$



LLM-Wahrscheinlichkeit $\times F$

(Wortvorhersage?)

Was ist die **Formel** für menschliche Lesezeiten?

Lesezeit =



Wortlänge $\times A +$



Alter $\times B +$



Promille $\times C +$



LLM-Wahrscheinlichkeit $\times F$ **(Wortvorhersage?)**

Sackgasse = **sehr niedrige Wahrscheinlichkeit?**

Was ist die **Formel** für menschliche Lesezeiten?

Lesezeit =



Wortlänge $\times A +$



Alter $\times B +$



Promille $\times C +$



LLM-Wahrscheinlichkeit $\times F$ **(Wortvorhersage?)**

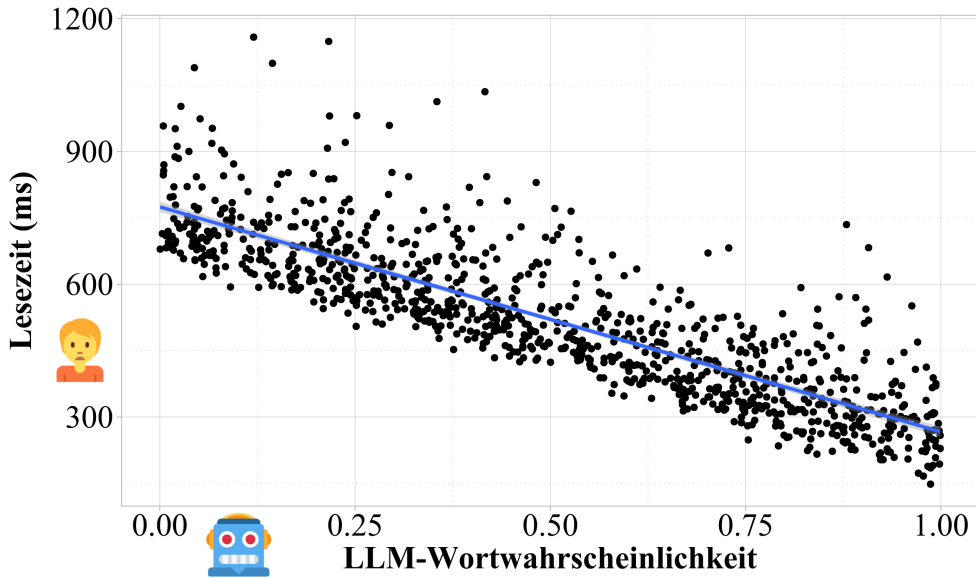
Sackgasse = **sehr niedrige Wahrscheinlichkeit?**

Sind LLMs **perfekte “Lesemodelle”**?  = 

Lesezeit = LLM-Wahrscheinlichkeit $\times$ F 
(Wortvorhersage)

Sackgasse = **sehr niedrige Wahrscheinlichkeit?**

Sind LLMs **perfekte “Lesemodelle”**?  = 



Dario mag Sprache und Computer **findet** er doof.

Dario mag Sprache und Computer **findet** er doof.

 ... [Das Wort **findet** ist hier **relativ unwahrscheinlich!**]

Dario mag Sprache und Computer **findet** er doof.

 ... [Das Wort **findet** ist hier **relativ unwahrscheinlich!**]

Dario

mag

Sprache

und

Computer

Dario mag Sprache und Computer **findet** er doof.

 ... [Das Wort **findet** ist hier **relativ unwahrscheinlich!**]

Dario

mag

Sprache

und

Computer

.	-	0.40
gern	-	0.33
total	-	0.20
nicht	-	0.06
findet	-	0.001

Brauchen wir das “Zwei-Modi”-Lesemodell
(✨ aus den 80ern ✨)
noch?

Bra



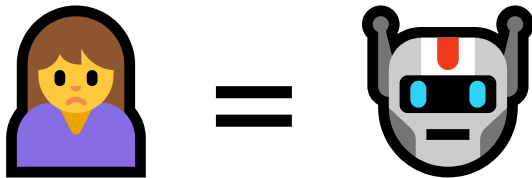
lell

**Brauchen wir das “Zwei-Modi”-Lesemodell
(✨ aus den 80ern ✨)
noch?**

... oder sind Menschen auch nur LLMs?

**Brauchen wir das “Zwei-Modi”-Lesemodell
(✨ aus den 80ern ✨)
noch?**

... oder sind Menschen auch nur LLMs?



Bin ich ein LLM?





STOP

I REFUSE TO BELIEVE THAT

Mein ✨ neues ✨ Modell:

Mein ✨neues✨ Modell:

Drei Lesemodi + mathematische Modellierung

Mein ✨neues✨ Modell:

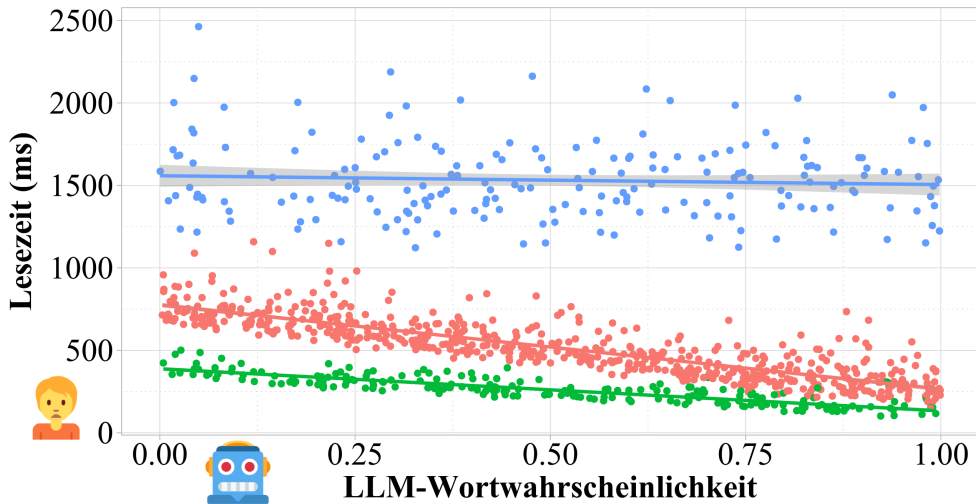
Drei Lesemodi + mathematische Modellierung

“Standard”-Modus: Erste Analyse bereits richtig 👁️👁️👍

“Reanalyse”-Modus: Analyse muss korrigiert werden 👁️👁️🚫🐌

🙈-Modus: Schnelles, unaufmerksames Lesen

Modus  Standard  Unaufmerksam  Reanalyse



◇ LLMs können nicht **unaufmerksam/besoffen sein** 🤖🍺🤖

- ◇ LLMs können nicht **unaufmerksam/besoffen sein** 🤖🍺🤖
- ◇ LLMs haben keinen bewussten **Sackgassen-Effekt** 🤖!?

- ◇ LLMs können nicht **unaufmerksam/besoffen sein** 🤖🍺🤖
- ◇ LLMs haben keinen bewussten **Sackgassen-Effekt** 🤖!?

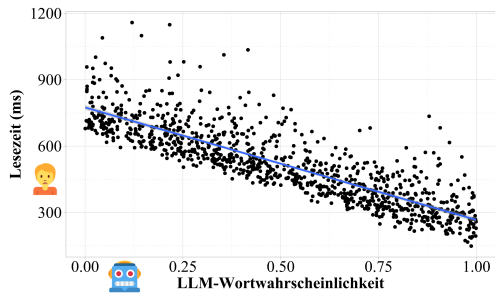
Aber:

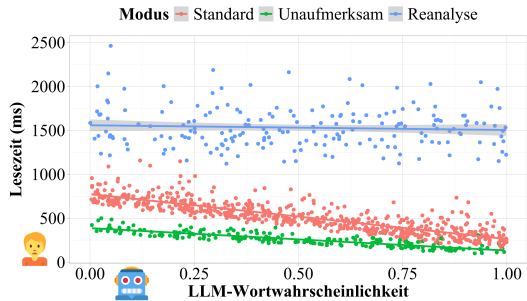
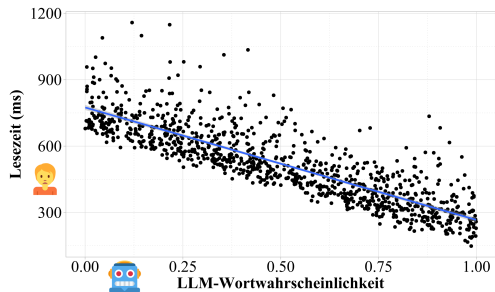
- ◇ LLMs können nicht **unaufmerksam/besoffen sein** 🤖🍺🤖
- ◇ LLMs haben keinen bewussten **Sackgassen-Effekt** 🤖!?

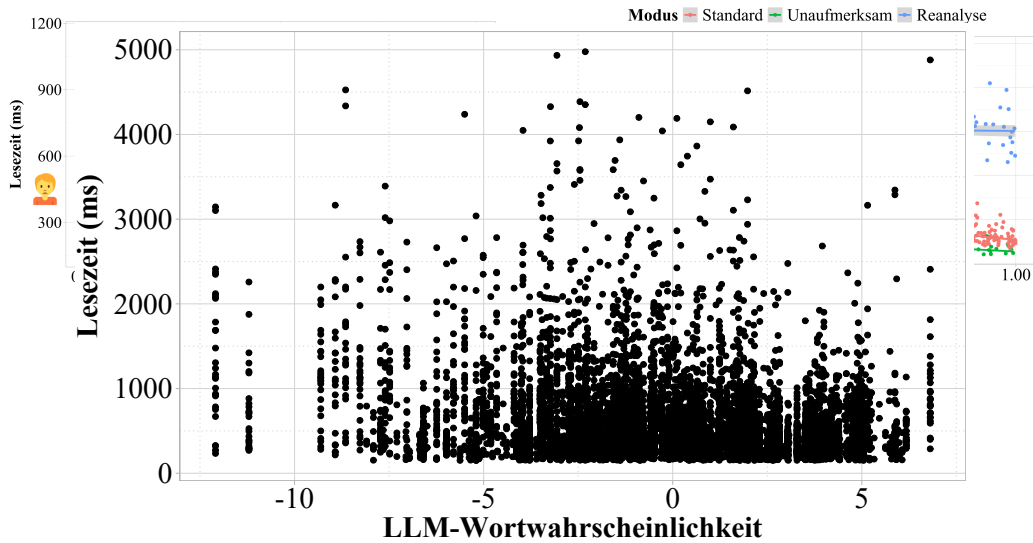
Aber:

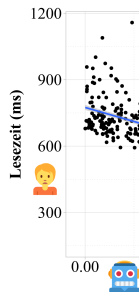


“Nimm im Zweifel das einfachste Modell”

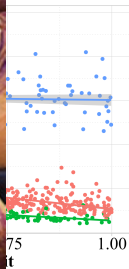


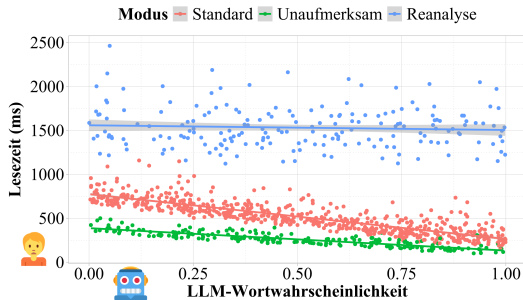
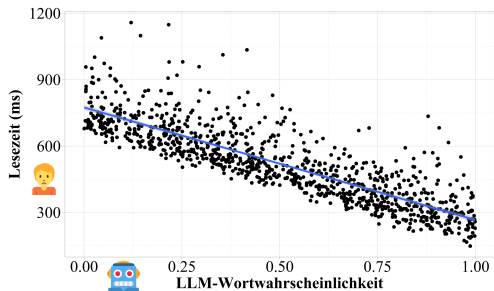




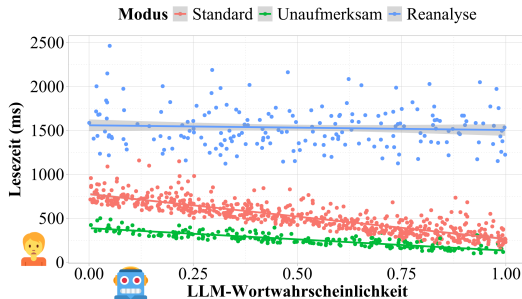
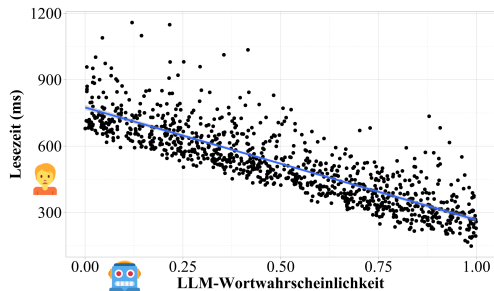


Reanalyse



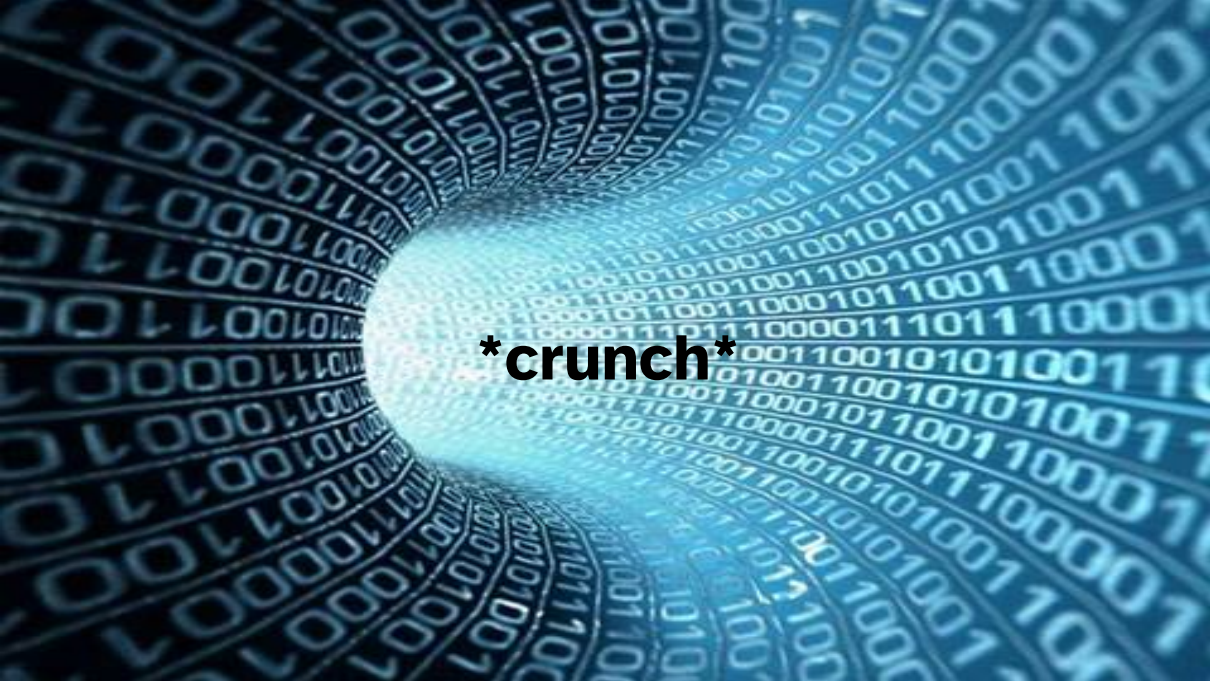


Welches Modell ist näher an den Daten?



Welches Modell ist näher an den Daten?

→ Abweichungen aufsummieren

The image features a perspective view of a tunnel formed by concentric, glowing blue rings of binary code (0s and 1s). The rings curve inward, creating a sense of depth and movement towards a bright, hazy light at the far end. The overall color palette is a range of blues, from deep navy to bright cyan. In the center of the tunnel, the text ***crunch*** is displayed in a bold, black, sans-serif font. The asterisks are slightly larger than the letters, and the text is perfectly centered both horizontally and vertically within the frame.

crunch





Mein **Drei-Modi-Modell** erklärt menschliche Lesezeiten
ca. **viermal** besser als das reine LLM-Modell!



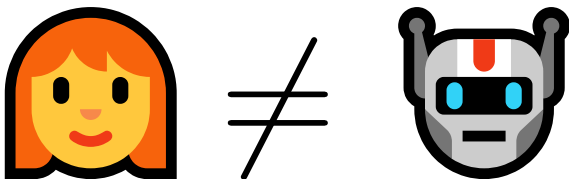
Mein **Drei-Modi-Modell erklärt menschliche Lesezeiten
ca. **viermal** besser als das reine LLM-Modell!**

(und das ist kreuzvalidiert, also kein Overfitting! 😊)



Mein **Drei-Modi-Modell** erklärt menschliche Lesezeiten
ca. **viermal** besser als das reine LLM-Modell!

(und das ist kreuzvalidiert, also kein Overfitting! 😊)



Was haben wir gelernt?

Was haben wir gelernt?

**MENSCHEN SIND
KEINE LLMS**

Was haben wir gelernt?

**MENSCHEN SIND
KEINE LLMS**

!!!1

Was haben wir gelernt?



Was haben wir gelernt?

**MENSCHEN SIND
KEINE LLMS**

!!!1

**(... aber 🤖 können helfen, menschliches Lesen
besser zu verstehen)**

DANKE!



DANKE!



Fanmail  und Jobangebote  gerne an
paape@uni-potsdam.de