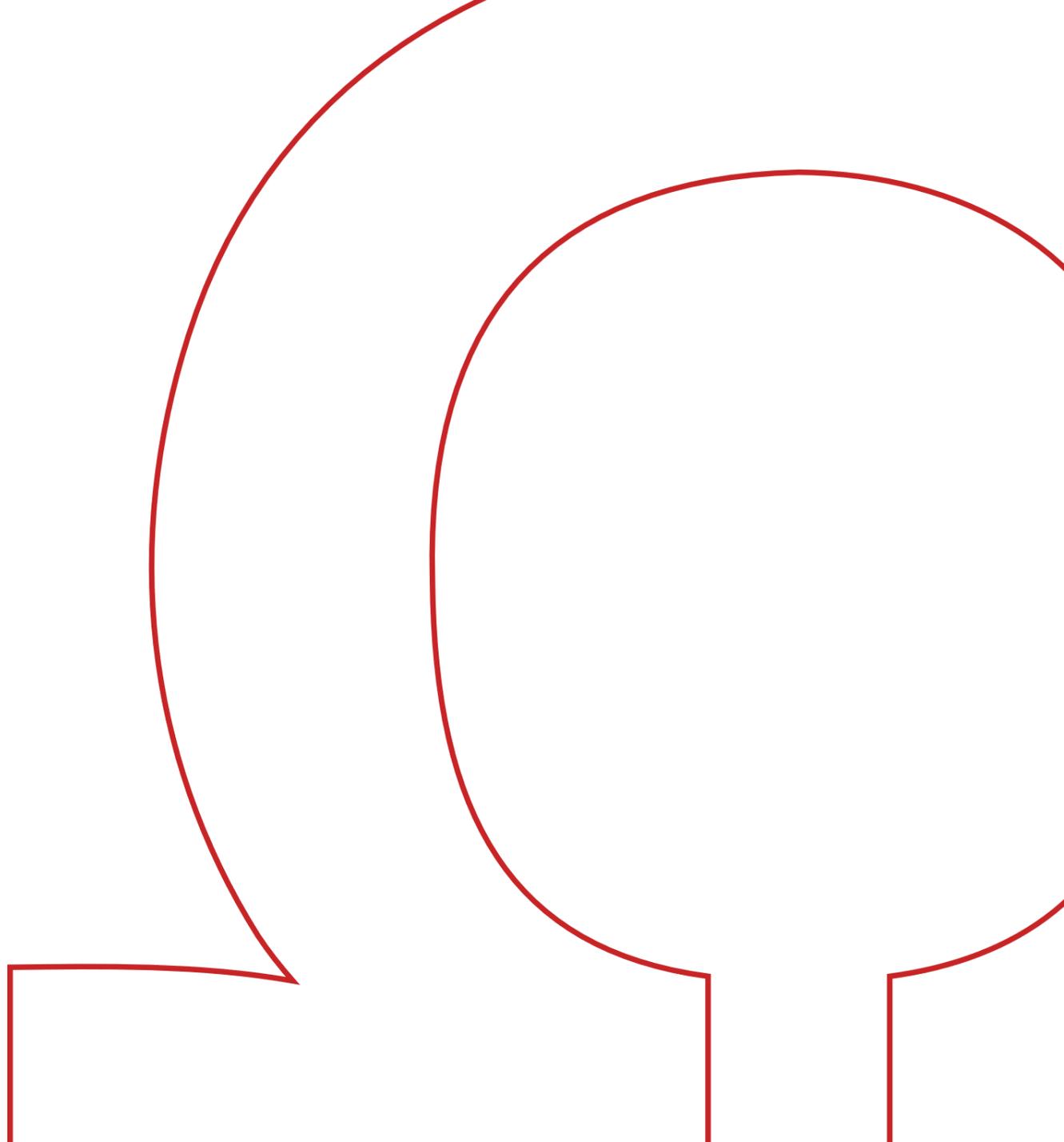


nicht-invasiven

Fortschritte in der Testdiagnostik:

Möglichkeiten der KI

Prof. Dr.-Ing. Korbinian Riedhammer  
Franziska Braun, M.Sc.



# Kognitive (Kurz-)Tests

- Mini-Mental-Status-Test (MMST)
- DemTect
- Test zur Früherkennung von Demenzen mit Depressionsabgrenzung (FTDD)
- Montreal Cognitive Assessment Test (MoCa)
- Uhrentest
- Syndromkurztest (SKT)
- CERAD(-Plus) mit Verbal Fluency, Boston Naming, u.a.

**weitgehend sprachbasiert**

# Sprachverarbeitung und pathologische Sprache

- Sprache als **nicht-invasiver Biomarker**
- **Was** wurde gesagt?
  - Automatische Spracherkennung generiert Transkript
  - Automatisierung der Testauswertung
- **Wie** wurde etwas gesagt?
  - Prosodie und Sprechgeschwindigkeit
  - Pausen, Sprecherüberlappung und Sprecherwechsel

Bocklet, Nöth und Riedhammer. [Künstliche Intelligenz für die Analyse pathologischer Sprache. Sprache-Stimme-Gehör, 2023; 47\(03\): 145 - 150](#)

# Frühe Arbeiten

- DementiaBank English Pitt Corpus (Becker, 1994)
- Kleinere spanische Datensets (u.a. zu Paisa Mutation; Perez-Toro, 2021)
- Oft nur einzelne Tests (z.B. Bildbeschreibung, MMST)
- Demenzgrad oder weitere Symptome (z.B. Depression) nicht erhoben
- Klassifikation von gesund/dement mit akustischen und wortbasierten Modellen

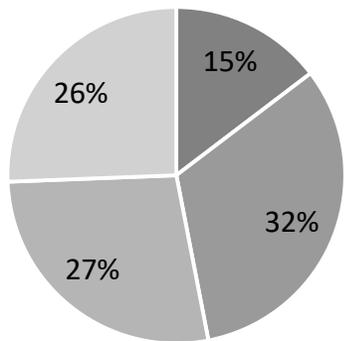
# Offene Fragen

- Können wir (sprachbasierten) Tests automatisch (und objektiv) auswerten?
- Können wir gesund/MCI/dement direkt aus der Sprache vorhersagen?
- Generalisieren die Ergebnisse auf eine breitere Population?
- Lernen wir "nur" Symptome oder das Syndrom?
- Text und Sprache: Können und sollten wir "zwischen den Zeilen" lesen?
- Sind die Entscheidungen der KI begründbar ("explainable AI")?

# Studie in der Gedächtnissprechstunde am Klinikum Nürnberg

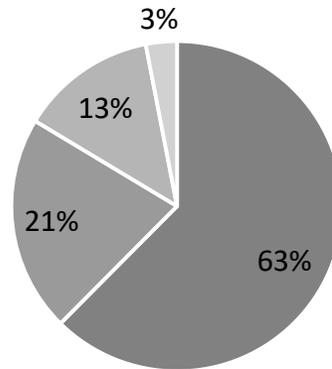
- 166 Patienten, ca. 139h Sprache
- Patient und Therapeut aufgenommen
- Psychologische & psychiatrische Diagnosen

Diagnose Demenz



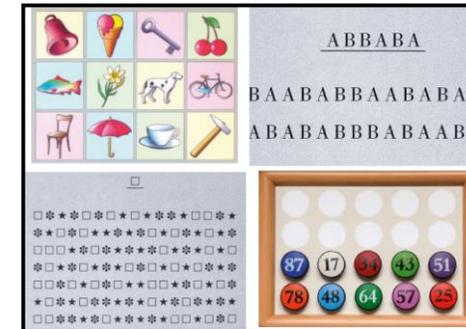
■ Keine ■ MCI ■ Leichte ■ Mittelschwere

Diagnose Depression

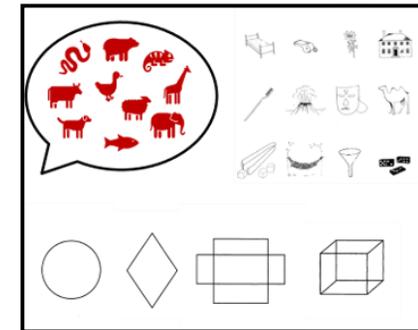


■ Keine ■ Leichte ■ Mittelschwere ■ Schwere

## Interview



SKT



CERAD

# SKT

- Kurzer Leistungstest zur Erfassung von Gedächtnis- und Aufmerksamkeitsstörungen



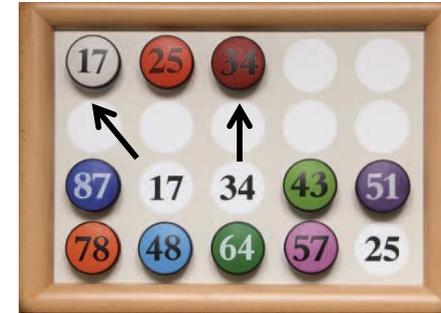
**SKT 1:**  
Objekte benennen



**SKT 2:**  
Objekte erinnern



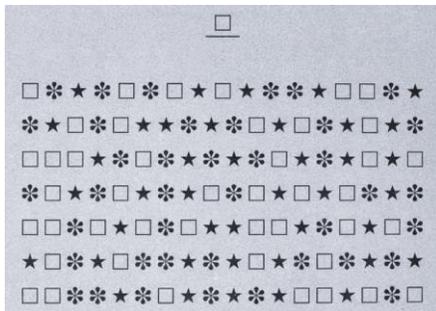
**SKT 3:**  
Zahlen vorlesen



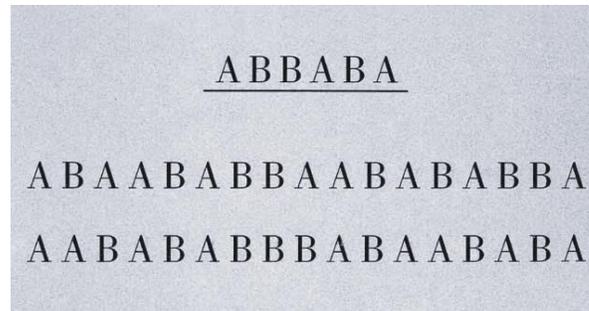
**SKT 4:**  
Zahlen ordnen



**SKT 5:**  
Zahlen zurücklegen



**SKT 6:**  
Symbole zählen



**SKT 7:**  
Interferenztest



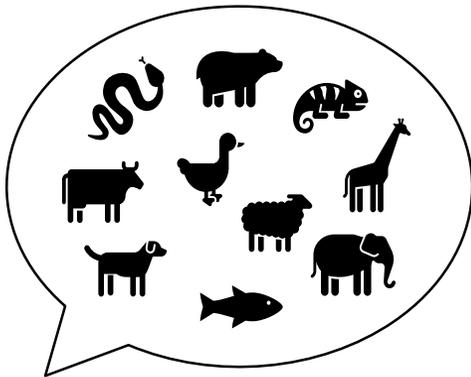
**SKT 8:**  
Objekte erinnern



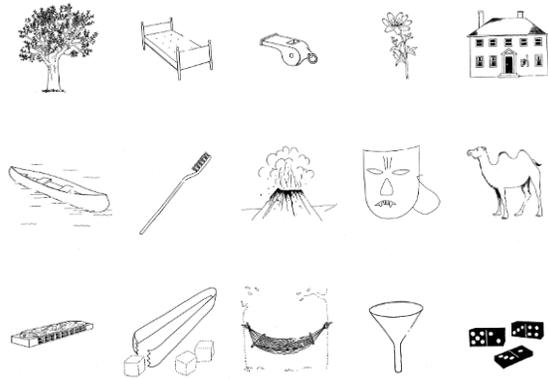
**SKT 9:**  
Objekte wiedererkennen

# CERAD

- Neuropsychologische Testbatterie für Alzheimer-Demenz

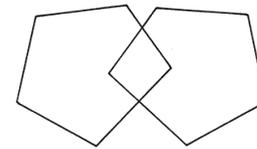


**CERAD 1:**  
Wortflüssigkeitstest



**CERAD 2:**  
Boston Naming Test

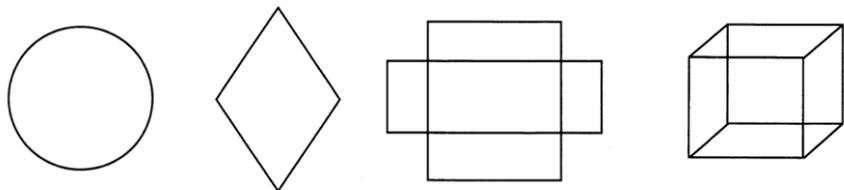
Bitte schließen Sie die Augen!



**CERAD 3:**  
Mini Mental Status Test

Butter	<input type="checkbox"/>
Arm	<input type="checkbox"/>
Strand	<input type="checkbox"/>
Brief	<input type="checkbox"/>
Königin	<input type="checkbox"/>
Hütte	<input type="checkbox"/>
Stange	<input type="checkbox"/>
Karte	<input type="checkbox"/>
Gras	<input type="checkbox"/>
Motor	<input type="checkbox"/>

**CERAD 4:**  
Wortliste lernen 3x



**CERAD 5:**  
Figuren zeichnen

**CERAD 8:**  
Figuren erinnern

**CERAD 6:**  
Wortliste erinnern

**CERAD 7:**  
Wortliste wiedererkennen

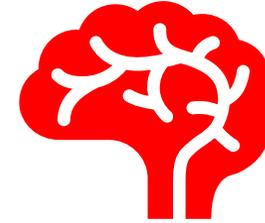
# Beispiele



- **Alter:** 80
- **Geschlecht:** weiblich
- **Score:** 30
- **Kognitiver Status:** Gesund
- **Diagnose:** Leichte Depression



- **Alter:** 67
- **Geschlecht:** männlich
- **Score:** 20
- **Kognitiver Status:** MCI
- **Diagnose:** MCI



- **Alter:** 69
- **Geschlecht:** weiblich
- **Score:** 8
- **Kognitiver Status:** Demenz
- **Diagnose:** Mittelschwere Demenz (Alzheimer's)

# Automatisierte Testauswertung

SKT

ID	Test/Task	Trans.	ASR-1	ASR-5	Top-21
1	naming objects	0.89	0.70	0.81	0.89
2	reproducing objects	1.00	0.58	0.71	0.83
3	reading numbers	0.94	0.85	0.86	0.94
6	counting symbols	0.90	0.59	0.58	0.54
7	interference test	0.99	0.97	0.98	0.99
8	naming after distraction	1.00	0.75	0.90	0.97
9	recognizing objects	0.89	0.50	0.55	0.68
-	attention score	0.92	0.84	0.82	0.85
-	memory score	0.98	0.62	0.78	0.93
-	total score	0.97	0.81	0.89	0.94

CERAD

ID	Test	Trans.	ASR-1	ASR-5	Top-21
1	verbal fluency test	0.98	0.82	0.85	0.91
2	Boston Naming Test	0.70	0.14	0.24	0.47
3	MMSE	0.71	0.07	0.35	0.52
4	word list learning	0.94	0.62	0.70	0.75
6	word list recall	0.99	0.68	0.81	0.78
-	total	0.71	0.37	0.49	0.61

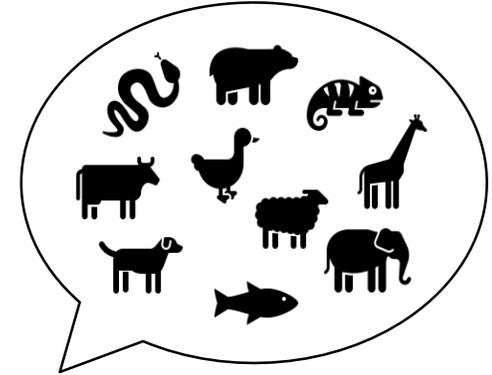
- 30 Patienten
- Hohe Korrelation zur manuellen Auswertung

# Offene Fragen

- **Können wir die (sprachbasierten) Tests automatisch auswerten?** 
- Können wir gesund/MCI/dement direkt aus der Sprache vorhersagen?
- Generalisieren die Ergebnisse auf eine breitere Population?
- Lernen wir "nur" Symptome oder das Syndrom?
- Text und Sprache: Können und sollten wir "zwischen den Zeilen" lesen?
- Sind die Entscheidungen der KI begründbar ("explainable AI")?

# PARLO Multi-Center Corpus

- 205 Patienten aus 9 Zentren in ganz Deutschland
- 7 kognitive Test (u.a. Verbal Fluency Test)
- Aufgezeichnet mit einer iPad App
- Testbasierte Eingruppierung in gesund/MCI/dement



# Experiment

- **NSC**: Nürnberg Single Center; **PMC**: Parlo Multi-Center
- 3-Klassen-Problem
- KI berechnet Akus

train	test	BERT	W2V2	PAD
<b>sVFT</b>				
NSC	NSC	54.2±10.7	61.7±9.1	45.5±8.6
PMC	PMC	46.8±8.9	55.4±6.7	44.2±10.6
NSC	PMC	49.3	43.8	35.4
PMC	NSC	56.0	51.1	38.1
MIX	MIX	56.0±3.5	56.3±3.2	48.7±3.0

# Offene Fragen

- Können wir die (sprachbasierten) Tests automatisch auswerten? ✓
- Können wir gesund/MCI/dement direkt aus der Sprache vorhersagen? ✓
- Generalisieren die Ergebnisse auf eine breitere Population? ✓
- Lernen wir "nur" Symptome oder das Syndrom?
- Text und Sprache: Können und sollten wir "zwischen den Zeilen" lesen?
- Sind die Entscheidungen der KI begründbar ("explainable AI")?

**Demenz und Depression haben ähnliche Symptome!**

# Experiment

- Daten aus Nürnberger Studie (NSC) haben Labels für Depression
- Lerne Demenz auf PMC
- Sage Demenz auf NSC voraus

	0	1	2
GT DEM 0	20	7	2
GT DEM 1	11	24	13
GT DEM 2	9	20	54
	0	1	2

PRED DEM

54 Patienten wurden korrekt als dement erkannt

	0	1	2
GT DEP 0	13	22	57
GT DEP 1	10	13	13
GT DEP 2	6	13	13
	0	1	2

GT DEM

57 Patienten sind dement, aber nicht depressiv

6 Patienten sind depressiv, aber nicht dement

# Experiment

- Daten aus Nürnberger Studie (NSC) haben Labels für Depression
- Lerne Demenz auf PMC
- Sage Demenz auf NSC voraus
- **Evaluere gegen Depression**
- **Entspricht etwa der erwarteten Verteilung**

[Braun et al., 2023. Classifying Dementia in the Presence of Alzheimers: A cross-corpus study.](#)

0	21	24	47
1	11	18	7
2	8	9	15
	0	1	2

PRED DEM

47 Patienten wurden als dement klassifiziert, sind aber nicht depressiv

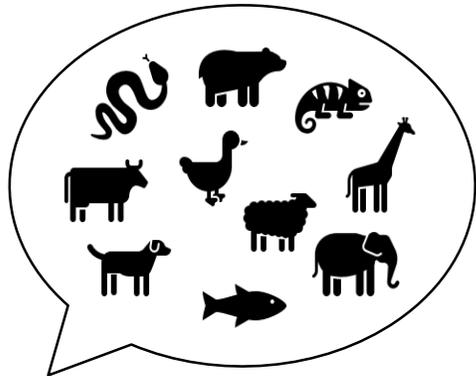
8 Patienten wurden als nicht dement klassifiziert, sind aber depressiv

# Offene Fragen

- Können wir die (sprachbasierten) Tests automatisch auswerten? ✓
- Können wir gesund/MCI/dement direkt aus der Sprache vorhersagen? ✓
- Generalisieren die Ergebnisse auf eine breitere Population? ✓
- Lernen wir "nur" Symptome oder das Syndrom? ✓
- Text und Sprache: Können und sollten wir "zwischen den Zeilen" lesen?
- Sind die Entscheidungen der KI begründbar ("explainable AI")?

# Demenz: Einfluss auf die Prosodie

- Kognitive Tests konzentrieren sich im wesentlichen auf **was** gesagt wurde (ggfs. unter Beachtung von Zeit, z.B. SKT, VFT)
- Beobachtung in verschiedenen Studien: Pausenmuster in Spontansprache können auf Demenz hinweisen
- Experiment: Auswirkung auf automatische Klassifikation bei VFT (*Geschwindigkeit und Korrektheit*) vs. Picture Description (*Inhalt und Konzepte*)



# Demenz: Einfluss auf die Prosodie

Table 2: Average AUC (in %) in 5-fold CV for the classification of NC vs. MCI using BERT and W2V2 features (best layer) from the VFT and PDT in self-attention and cross-attention with and without pause coding (P1–4) and disfluencies (Disfl.).

features	VFT		PDT	
	self-att	cross-att	self-att	cross-att
BERT	66.3	71.0	57.8	67.5
BERT P1	66.0	70.2	58.5	<b>70.3</b>
BERT P2	65.5	71.2	52.4	70.0
BERT P3	66.7	71.2	56.6	68.6
BERT P4	66.1	<b>71.4</b>	57.1	68.0
BERT P3 + Disfl.	<b>69.3</b>	71.1	62.4	67.9
W2V2	67.5	-	<b>66.2</b>	-

Table 3: Average AUC (in %) in 5-fold CV for the classification of MCI vs. AD using BERT and W2V2 features (best layer) from the VFT and PDT in self-attention and cross-attention with and without pause coding (P1–4) and disfluencies (Disfl.).

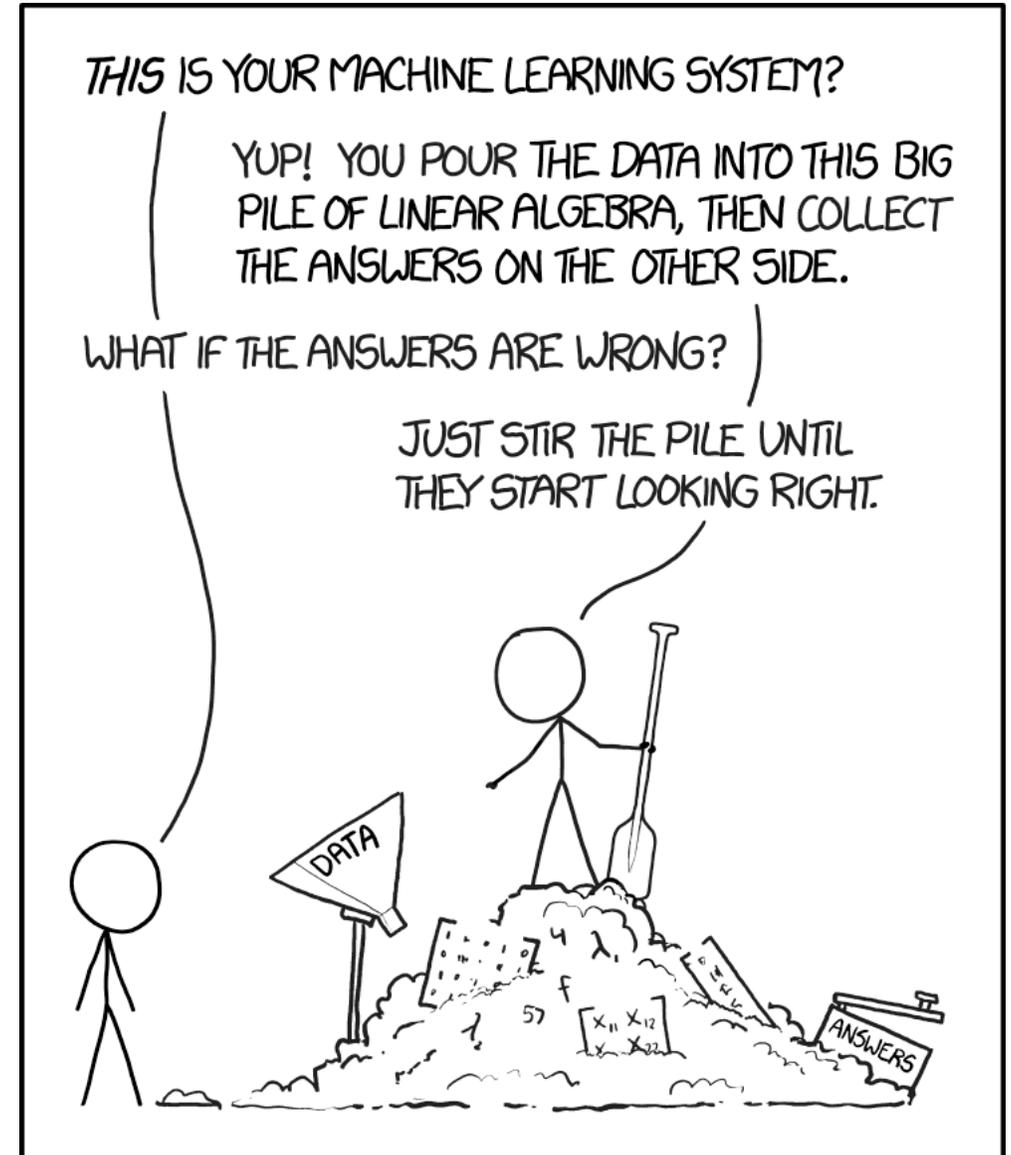
features	VFT		PDT	
	self-att	cross-att	self-att	cross-att
BERT	71.0	66.7	71.8	76.4
BERT P1	68.8	69.2	77.6	76.3
BERT P2	67.5	67.1	77.3	<b>77.8</b>
BERT P3	71.3	68.6	74.7	<b>77.8</b>
BERT P4	<b>71.7</b>	70.4	79.5	<b>77.8</b>
BERT P3 + Disfl.	70.6	<b>70.6</b>	<b>80.5</b>	77.2
W2V2	67.1	-	73.4	-

# Offene Fragen

- Können wir die (sprachbasierten) Tests automatisch auswerten? ✓
- Können wir gesund/MCI/dement direkt aus der Sprache vorhersagen? ✓
- Generalisieren die Ergebnisse auf eine breitere Population? ✓
- Lernen wir "nur" Symptome oder das Syndrom? ✓
- Text und Sprache: Können und sollten wir "zwischen den Zeilen" lesen? ✓
- Sind die Entscheidungen der KI begründbar ("explainable AI")?

# Explainable AI

- KI oft “black box”:  
Daten rein, Entscheidung raus
- Große Gefahr: System lernt zufällige Koinzidenzen im Trainingset
  - Gesund = Mikro 1, Dement = Mikro 2
  - Alter, Geschlecht, etc. (“bias”)
- Wie können wir die Entscheidungen nachvollziehbar machen?



# Interpretable AI

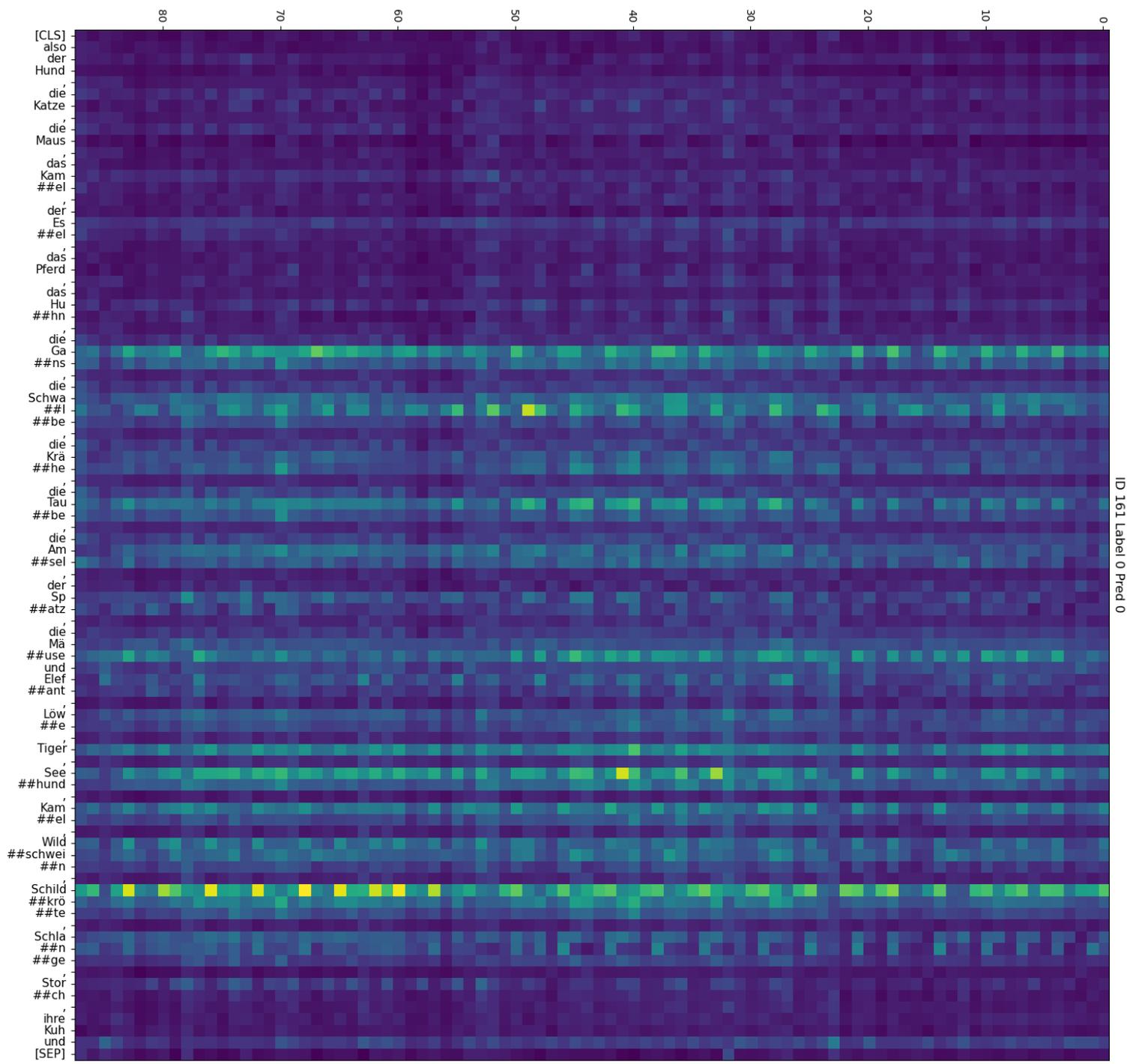
- “Erklärbar” ist ein ungünstiger Begriff: die Formeln sind klar...
- “Interpretierbar” trifft es besser: Können wir eine intuitive Darstellung finden, warum oder wie eine Entscheidung zustande gekommen ist?
- Moderne Deep Learning Verfahren verwenden (fast immer) “Attention”:
  - Netz lernt die Eingaben zueinander in Relation zu setzen
  - Je mehr ein Element Bezugspunkt wird, desto mehr trägt es zur Entscheidung bei
  - Diese Aufmerksamkeitswerte können im Nachgang visualisiert werden

# Aufmerksamkeit für Demenz im VFT

- Verbal fluency test: *Nennen Sie so viele Tiere Ihnen einfallen, in einer Minute.*
- Klassifikatorgenauigkeit gesund–dement bei 91%
- **Erfahrungen aus der Praxis die auf Zustand hinweisen:**
  - Spezifische Tiere: Gattungen/Rassen/Unterrassen
  - Exotische/ausgefallene Tiere
  - Wort- und Satzwiederholungen
  - Ablenken, Unsicherheit, Füllwörter/-phrasen
    - “Ich weiß nichts mehr”, “Habe ich schon gesagt”, ...
  - Starkes Zögern ruft Therapeuteninteraktion hervor (Hilfestellung, Nachfragen, ...)

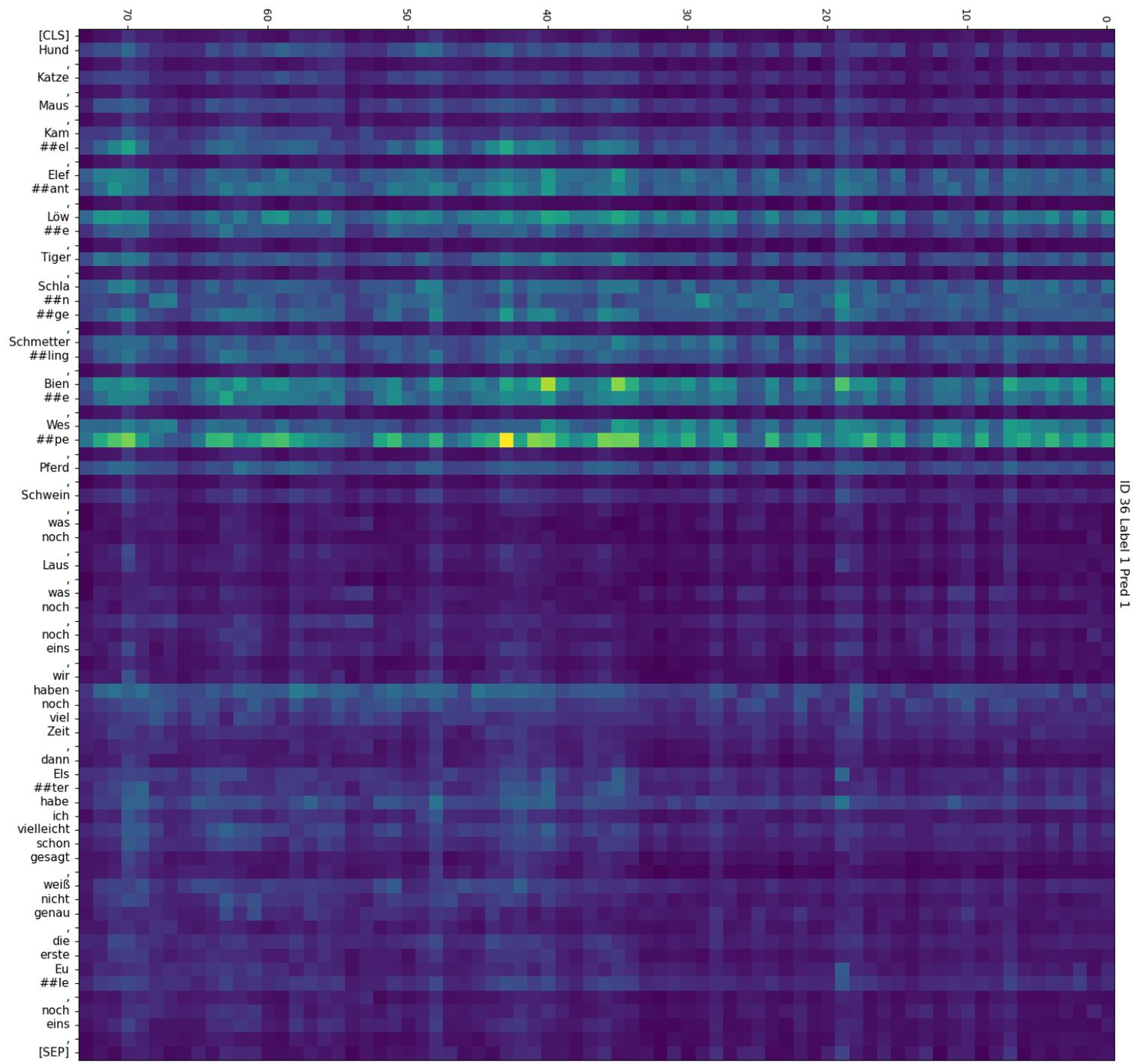
# Keine Demenz

- + Kontinuierlicher Input
- + Unterarten
- + Exotische Tiere

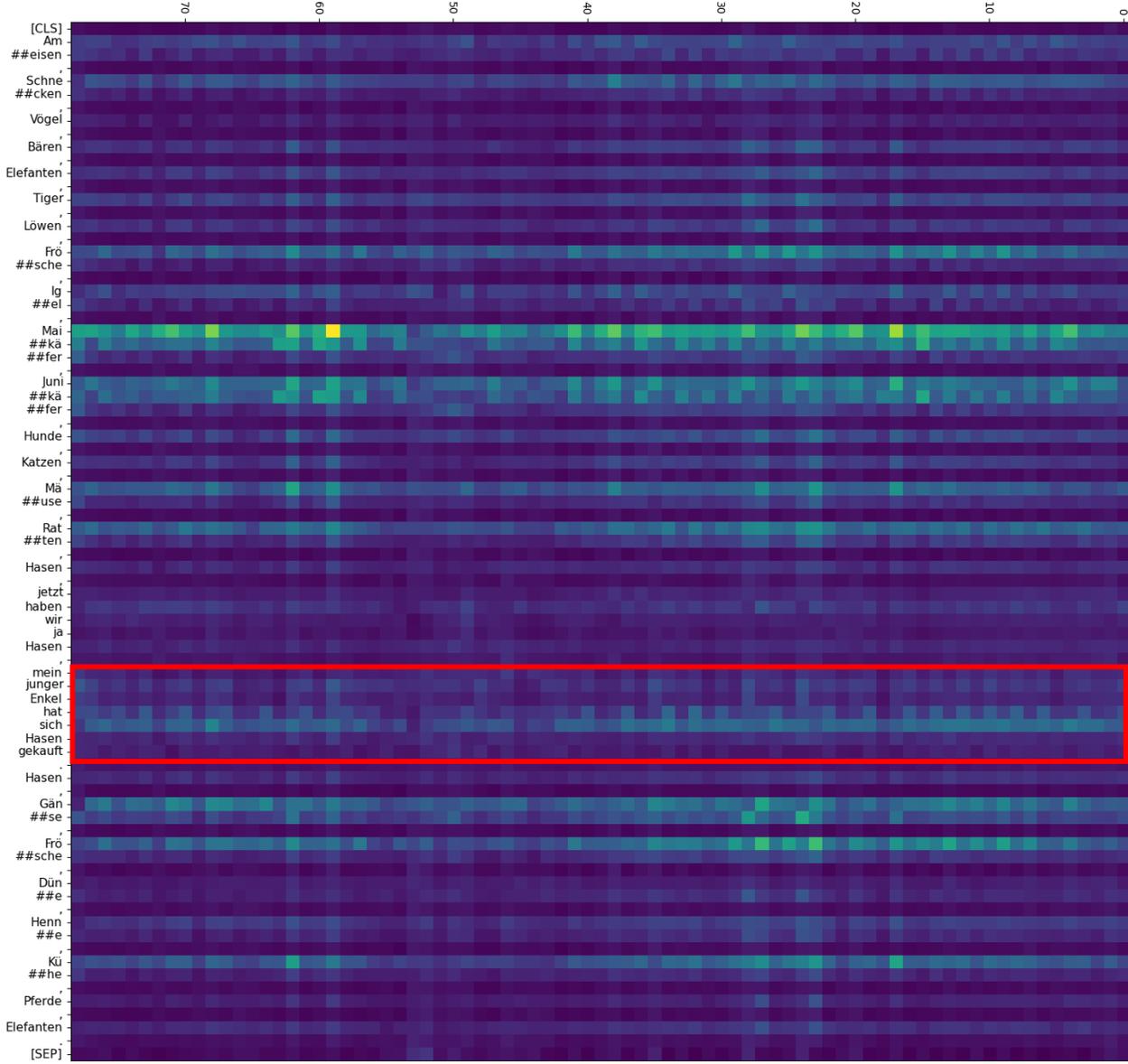


# Demenz

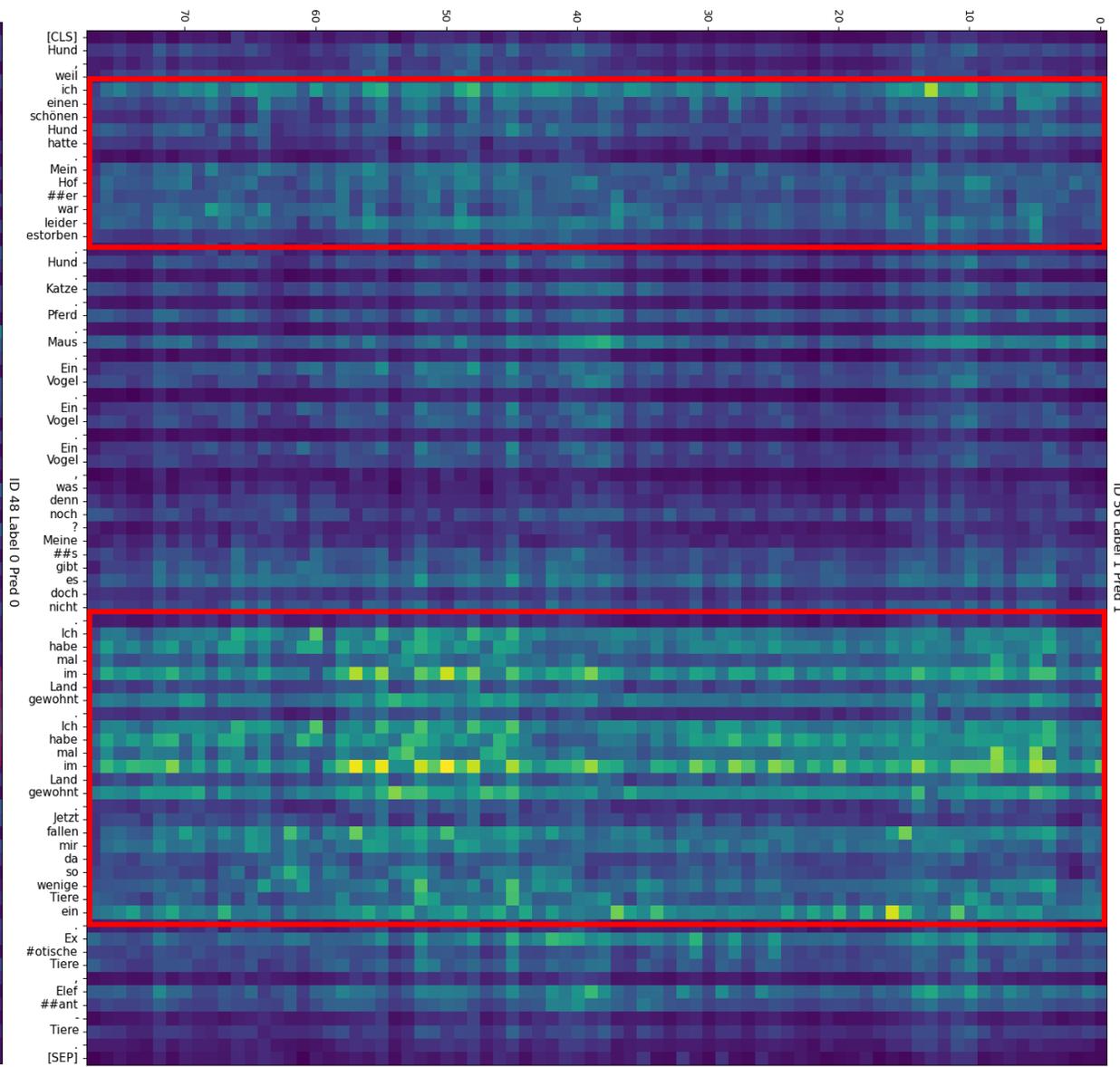
- Viel Input am Anfang
- „schon gesagt“
- „weiß nicht“
- Therapeutensprache



# Ablenkung: Keine Demenz (+Depr.)



# vs. Demenz (+Depr.)



# Offene Fragen

- Können wir die (sprachbasierten) Tests automatisch auswerten? ✓
- Können wir gesund/MCI/dement direkt aus der Sprache vorhersagen? ✓
- Generalisieren die Ergebnisse auf eine breitere Population? ✓
- Lernen wir "nur" Symptome oder das Syndrom? ✓
- Text und Sprache: Können und sollten wir "zwischen den Zeilen" lesen? ✓
- **Sind die Entscheidungen der KI begründbar ("explainable AI")?**

Hier stehen wir noch am Anfang!

# Von der Forschung in die Praxis



- Tests (insb. SKT) auf Sprachbasis sind automatisch auswertbar  
...sind sie auch automatisch durchführbar?
- Implementierung eines digitalen SKT (dSKT)
  - iPad 13" → ähnliche Größe zum Brett
  - On-device Spracherkennung → Erfüllung der DSGVO
  - Alleine oder mit geringfügiger Anleitung durchführbar
- **Aktuelle in Vorbereitung**
  - Validierungsstudie zur Usability
  - Validierungsstudie zur Sensitivity, in Kombination mit Blutmarkern (Kooperation mit UK Göttingen)

# Fazit

- Sprache kann als nicht-invasiver Biomarker verwendet werden
- Sprachbasierte Tests können i.A. automatisiert werden
- Das Sprachsignal trägt mehr Information als nur das “gesprochene Wort”
  - Prosodie, Sprechtempo
  - Stimmung, Emotion
  - KI kann diese Informationen zur Abgrenzung von Syndromen heranziehen
- Methoden zur Interpretierbarkeit sind Gegenstand der aktuellen Forschung
- Digitale Version des SKT ermöglicht niederschwellige Testung
  - (ambulante) Pflegeheime
  - Hausärzte