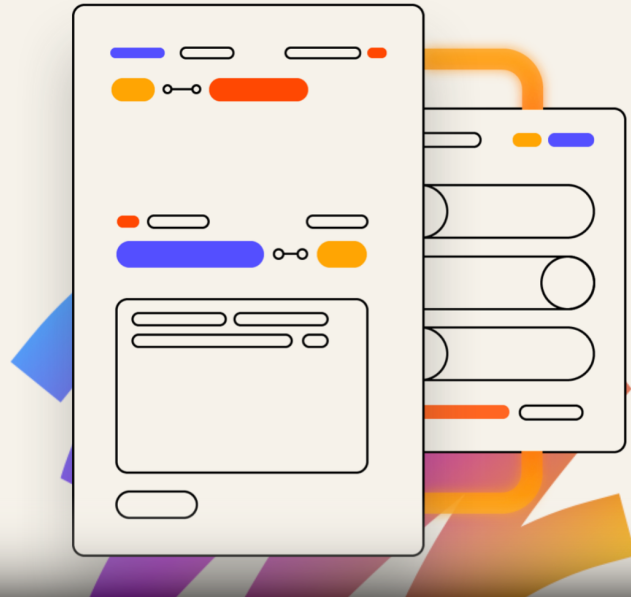


Machine Learning und die Auswertung offener Antworten

Workshop am 24. November 2025, Johannes Gutenberg-Universität Mainz



Der Workshop bietet eine Einführung in die Analyse offener Antworten aus der Perspektive qualitativer Forschung und lotet Möglichkeiten der Unterstützung durch Verfahren des Maschinellen Lernens aus. Neben der manuellen Kodierung werden exemplarisch auch Verfahren wie Clustering, Topic Modeling und der Einsatz großer Sprachmodelle vorgestellt, erprobt und diskutiert.

Im Mittelpunkt steht die Frage, wie sich eine qualitative Forschungshaltung mit KI-basierten Verfahren verbinden lässt und welche Chancen und Herausforderungen damit im Forschungsprozess entstehen.

LET'S START



C3·S
Center for
Critical
Computational
Studies

GOETHE
UNIVERSITÄT
FRANKFURT AM MAIN



Dan Verständig | verstaendig@c3s.uni-frankfurt.de

Educational Theory and Practices of Critical Computational Literacy

 [@edu_cc](https://www.instagram.com/edu_cc/)

 [edu_cc@wisskomm.social](https://m.masto.host/edu_cc/@wisskomm.social)



[c3s-frankfurt.de](https://www.c3s-frankfurt.de)

Machine Learning und die Auswertung offener Antworten

Workshop am 24. November 2025, Johannes Gutenberg-Universität Mainz



pixelspace.org/ml

Ablaufplan

10:00 – 10:30 **Einstieg:** Kennenlernen und Erfahrungen

10:30 – 11:30 **Input:** Grundlegende Begriffe und Konzepte

11:30 – 12:30 **Diskussion:** ML und die empirische Sozialforschung

12:30 – 13:30 Mittagspause

13:30 – 15:00 **Übung:** Topic Modeling und Klassifikationen

15:00 – 15:15 Pause

15:15 – 16:30 **Übung:** Auswertung Mensch vs. Maschine

16:30 – 17:00 **Abschluss:** Fazit und Feedback



Machine Learning und Qualitative Daten

Gehen Sie auf

www.menti.com

Geben Sie diesen Code ein

1562 7810



Oder verwenden Sie den QR-Code.

Machine Learning und die Auswertung offener Antworten

Workshop am 24. November 2025, Johannes Gutenberg-Universität Mainz



Der Workshop bietet eine Einführung in die Analyse offener Antworten aus der Perspektive qualitativer Forschung und lotet Möglichkeiten der Unterstützung durch Verfahren des Maschinellen Lernens aus. Neben der manuellen Kodierung werden exemplarisch auch Verfahren wie Clustering, Topic Modeling und der Einsatz großer Sprachmodelle vorgestellt, erprobt und diskutiert.

Im Mittelpunkt steht die Frage, wie sich eine qualitative Forschungshaltung mit KI-basierten Verfahren verbinden lässt und welche Chancen und Herausforderungen damit im Forschungsprozess entstehen.

LET'S START

Machine Learning und die Auswertung offener Antworten

Workshop am 24. November 2025, Johannes Gutenberg-Universität Mainz



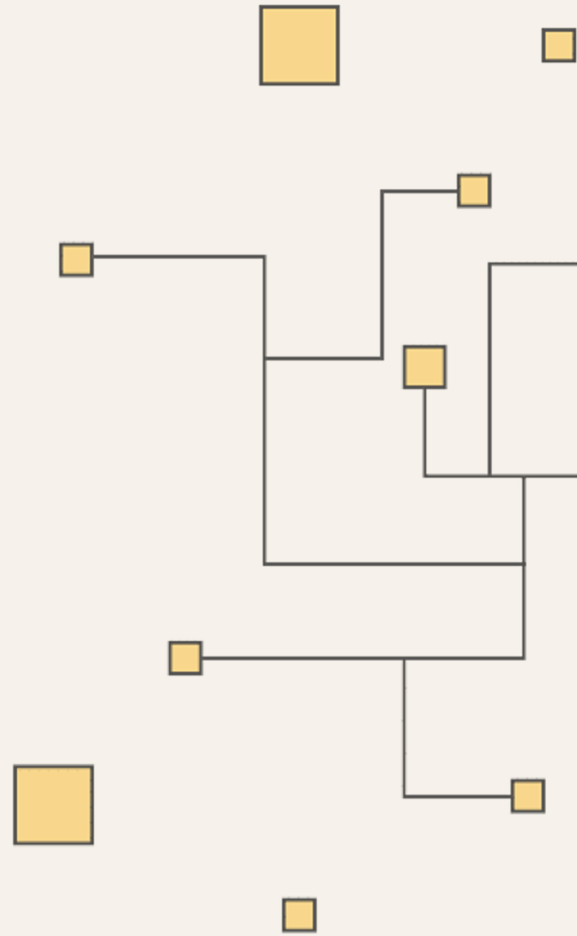
Der Workshop bietet eine Einführung in die Analyse offener Antworten aus der Perspektive qualitativer Forschung und lotet Möglichkeiten der Unterstützung durch Verfahren des Maschinellen Lernens aus. Neben der manuellen Kodierung werden exemplarisch auch Verfahren wie Clustering, Topic Modeling und der Einsatz großer Sprachmodelle vorgestellt, erprobt und diskutiert.

Im Mittelpunkt steht die Frage, wie sich eine qualitative Forschungshaltung mit KI-basierten Verfahren verbinden lässt und welche Chancen und Herausforderungen damit im Forschungsprozess entstehen.

LET'S START

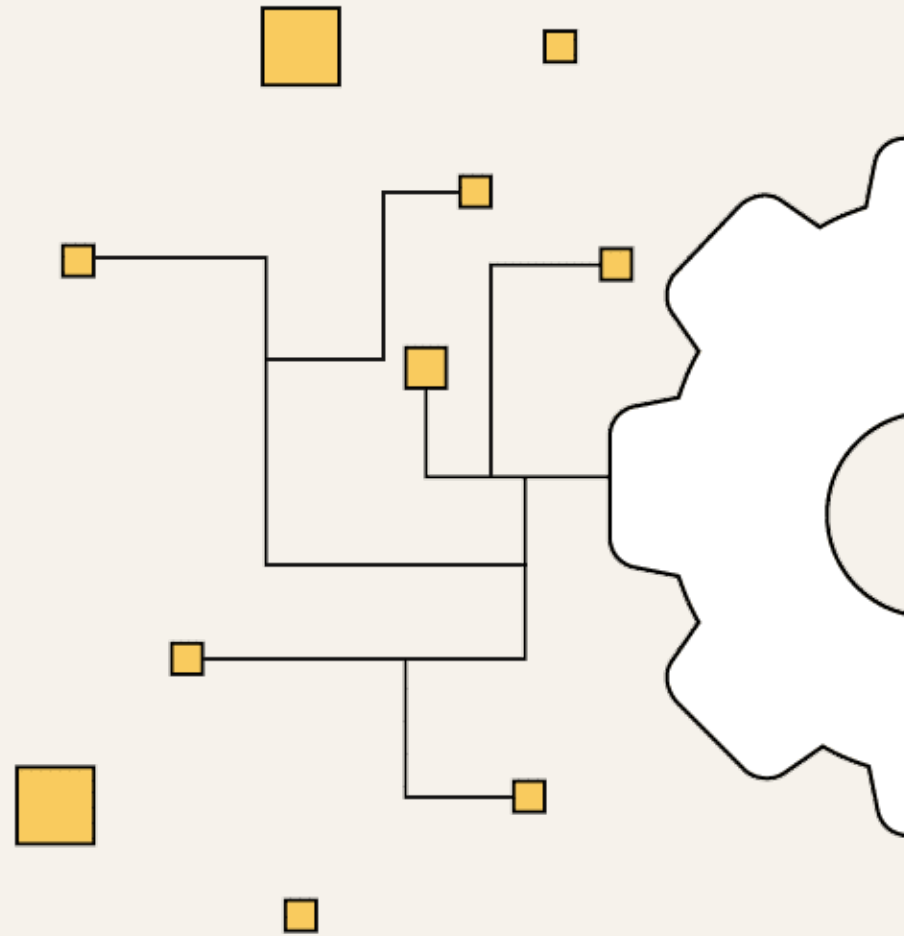
Machine Learning

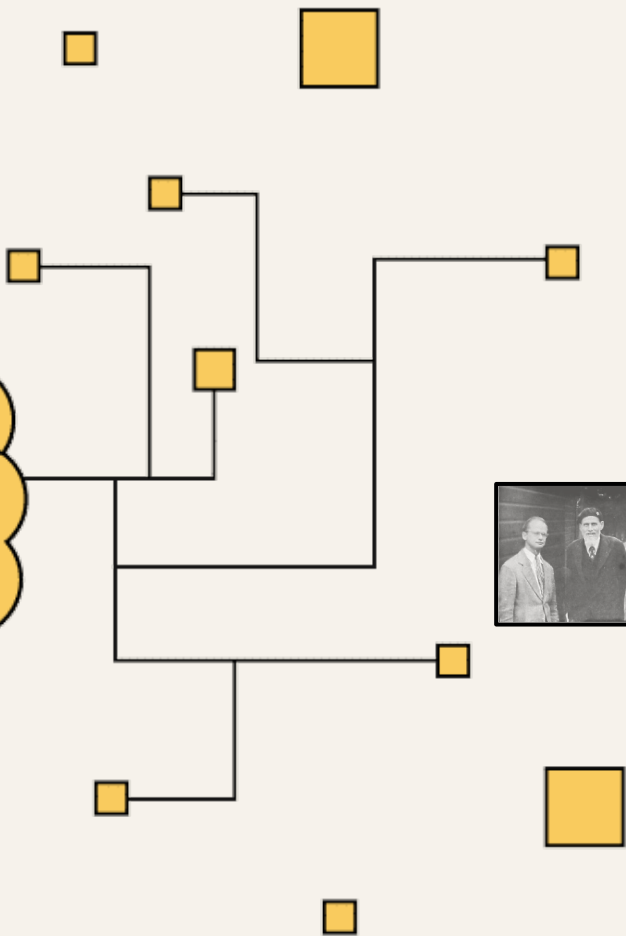
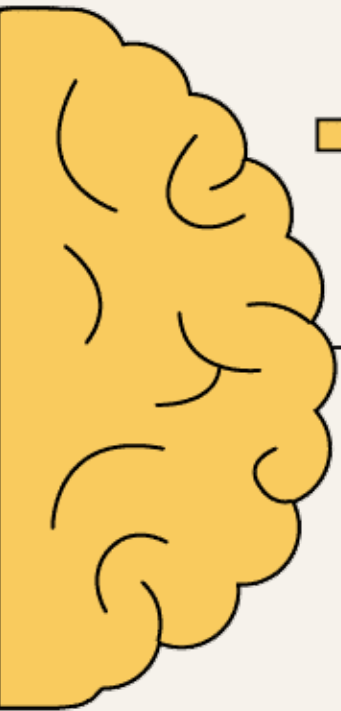
Künstliche Intelligenz



Machine Learning

Ein Teilbereich der Künstlichen Intelligenz, der darauf basiert, aus Daten Muster zu erkennen und Modelle zu erzeugen, anstatt feste Regeln vorzugeben. ML-Systeme lernen statistische Zusammenhänge, indem sie viele Beispiele analysieren, und nutzen diese Muster, um neue Daten zu klassifizieren, vorherzusagen oder zu gruppieren.

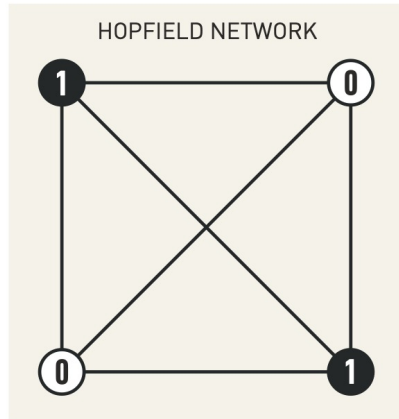




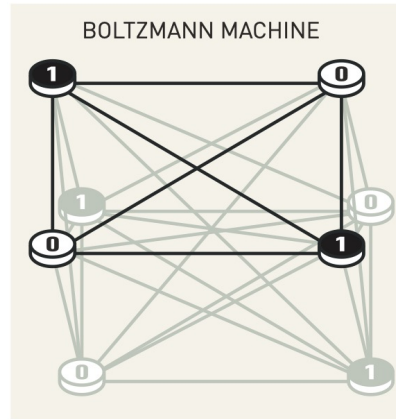
Künstliche Neuronale Netzwerke

Künstliche neuronale Netze (KNN) entstanden in den 1940er-Jahren ([McCulloch & Pitts, 1943](#)) und wurden in den 1950ern durch das [Perzeptron](#) bekannt. Nach technischen Grenzen und einem [KI-Winter](#) brachten Hopfield-Netze und Hinton's Arbeiten zu Backpropagation und [Boltzmann-Maschinen](#) in den 1980ern den Durchbruch. Sie bilden die Grundlage des heutigen [Deep Learning](#).

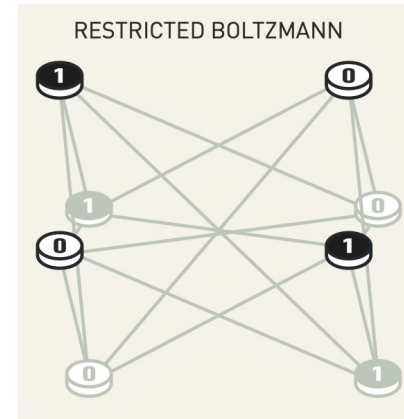
Different types of network



John Hopfield's associative memory is built so that all the nodes are connected to each other. Information is fed in and read out from all the nodes.



Geoffrey Hinton's Boltzmann machine is often constructed in two layers, where information is fed in and read out using a layer of *visible* nodes. They are connected to *hidden* nodes, which affect how the network functions in its entirety.



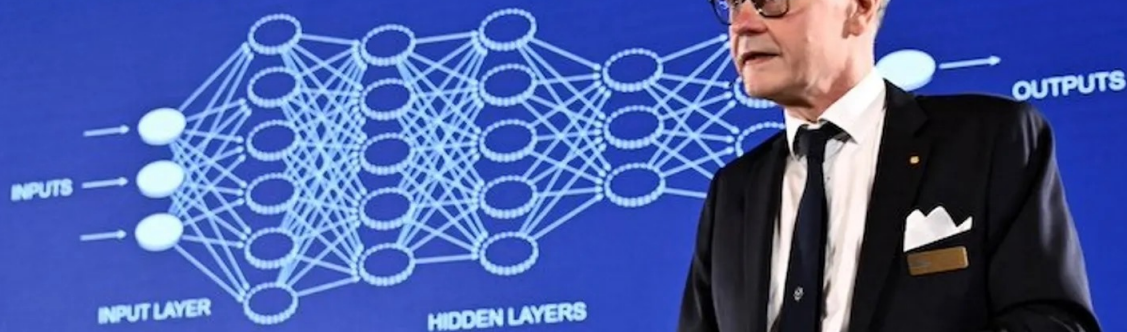
In a restricted Boltzmann machine, there are no connections between nodes in the same layer. The machines are frequently used in a chain, one after the other. After training the first restricted Boltzmann machine, the content of the hidden nodes is used to train the next machine, and so on.



NOBELPRISET I FYSIK 2024 THE NOBEL PRIZE IN PHYSICS 2024

KUNGL. VETENSKAPSKA
AKADEMIEN
THE ROYAL SWEDISH ACADEMY OF SCIENCES

Feedforward networks



KUNGL.
VETENSKAPSKA
AKADEMIEN
THE ROYAL SWEDISH ACADEMY OF SCIENCES

KUNGL.
VETENSKAPSKA
AKADEMIEN
THE ROYAL SWEDISH ACADEMY OF SCIENCES

Credit: Christine Olsson/TT via Reuters

Small Data
Raw Data
Big Data
Daten



Preprocessing

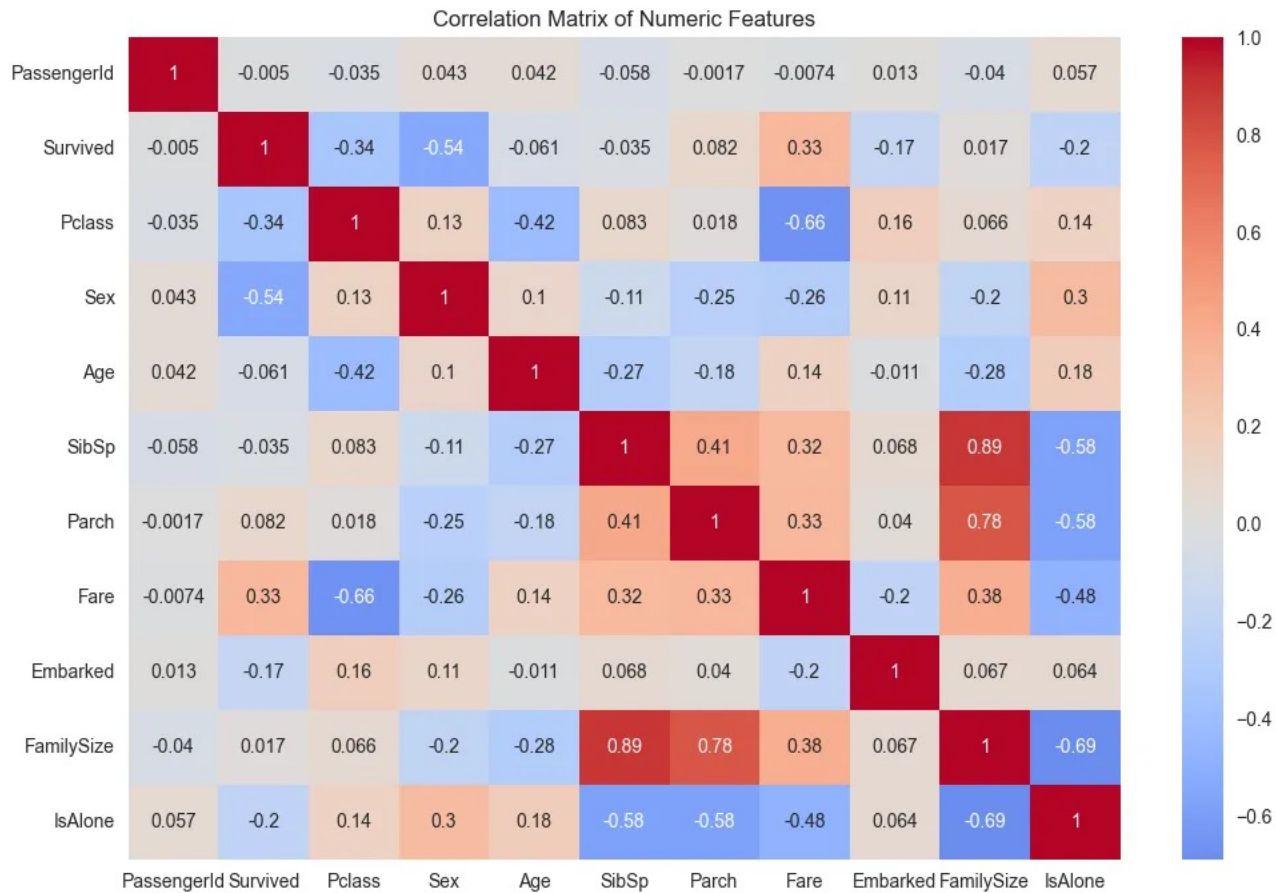
Tokenizing

Vektorizing

Clustering

Datenpipeline

Visualizing



Credit: [Kowshik Saravanan](#)





Interpretation und Komputation

MOTHERBOARD
TECHBYVICE

Flawed Algorithms Are Grading Millions of Students' Essays

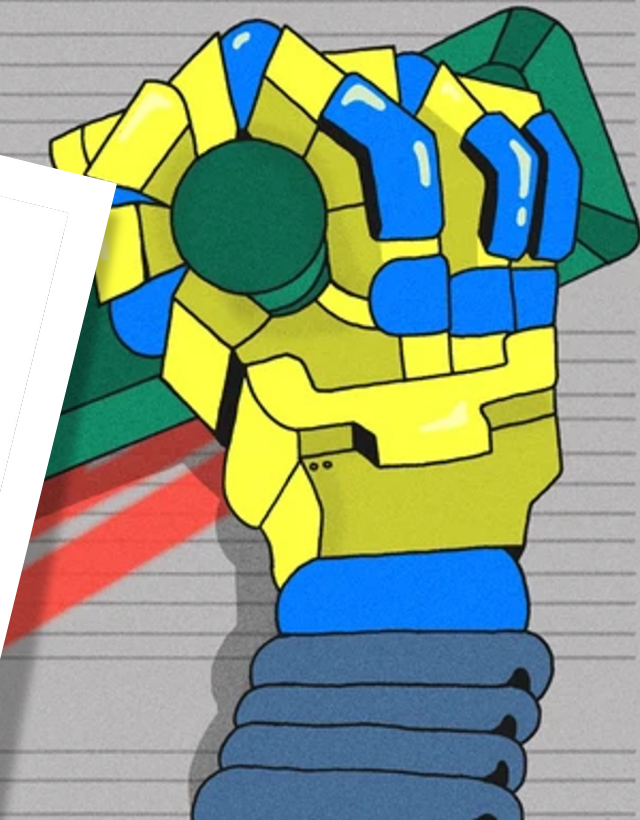
Fooled by gibberish and highly susceptible to human bias, automated essay-scoring systems are being increasingly adopted, a Motherboard investigation has found

TP By Todd Farathers

August 20, 2019 3:00pm

[f Share](#) [t Tweet](#) [s Snap](#)

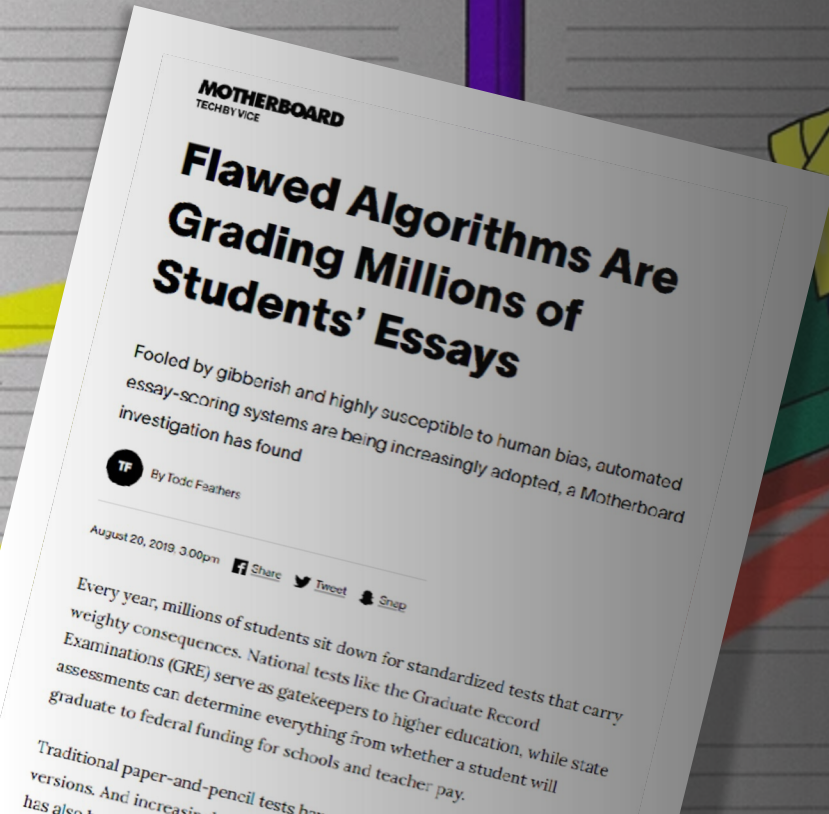
Every year, millions of students sit down for standardized tests that carry weighty consequences. National tests like the Graduate Record Examinations (GRE) serve as gatekeepers to higher education, while state assessments can determine everything from whether a student will graduate to federal funding for schools and teacher pay. Traditional paper-and-pencil tests have been around for decades. And increasing versions. And increasing versions.



1) **Biased.** Ohnehin benachteiligte Gruppen von Lernenden und Minderheiten werden algorithmisch abermals benachteiligt.



- 1) **Biased.** Ohnehin benachteiligte Gruppen von Lernenden und Minderheiten werden algorithmisch abermals benachteiligt.
- 2) **Kein Sinn.** Das System belohnt gut geschriebene Texte. Es geht nur um die Textstruktur, nicht um Bedeutung.







Pause.



Gleich geht es weiter.

Gehen Sie auf

www.menti.com

Geben Sie diesen Code ein

3159 5202



Oder verwenden Sie den QR-Code.

Data Mining Fruitful and Fun

Open source machine learning and data
visualization.



In dieser Übung betrachten wir offene Antworten mit zwei maschinellen Verfahren, dem Clustering und Topic Modeling und vergleichen die Ergebnisse mit eigenen Interpretationen.

Ziel der Übung: Sichbarmachung, wie Verzerrung (Bias) entstehen kann.

Zeit für die Übung:	30 Minuten
Arbeitsweise:	2-3er Gruppen
Diskussion:	30 Minuten

Auswertung und Bias

In dieser Übung betrachten wir offene Antworten mit zwei maschinellen Verfahren, dem [Clustering](#) und [Topic Modeling](#) und vergleichen die Ergebnisse mit eigenen Interpretationen.

Ziel der Übung: Sichbarmachung, wie [Verzerrung \(Bias\)](#) entstehen kann.

▶ ERSTE SCHRITTE MIT ORANGE

1) Orange installieren



2) Daten laden



3) Zwei Analysepfade



1) Orange installieren



2) Daten laden

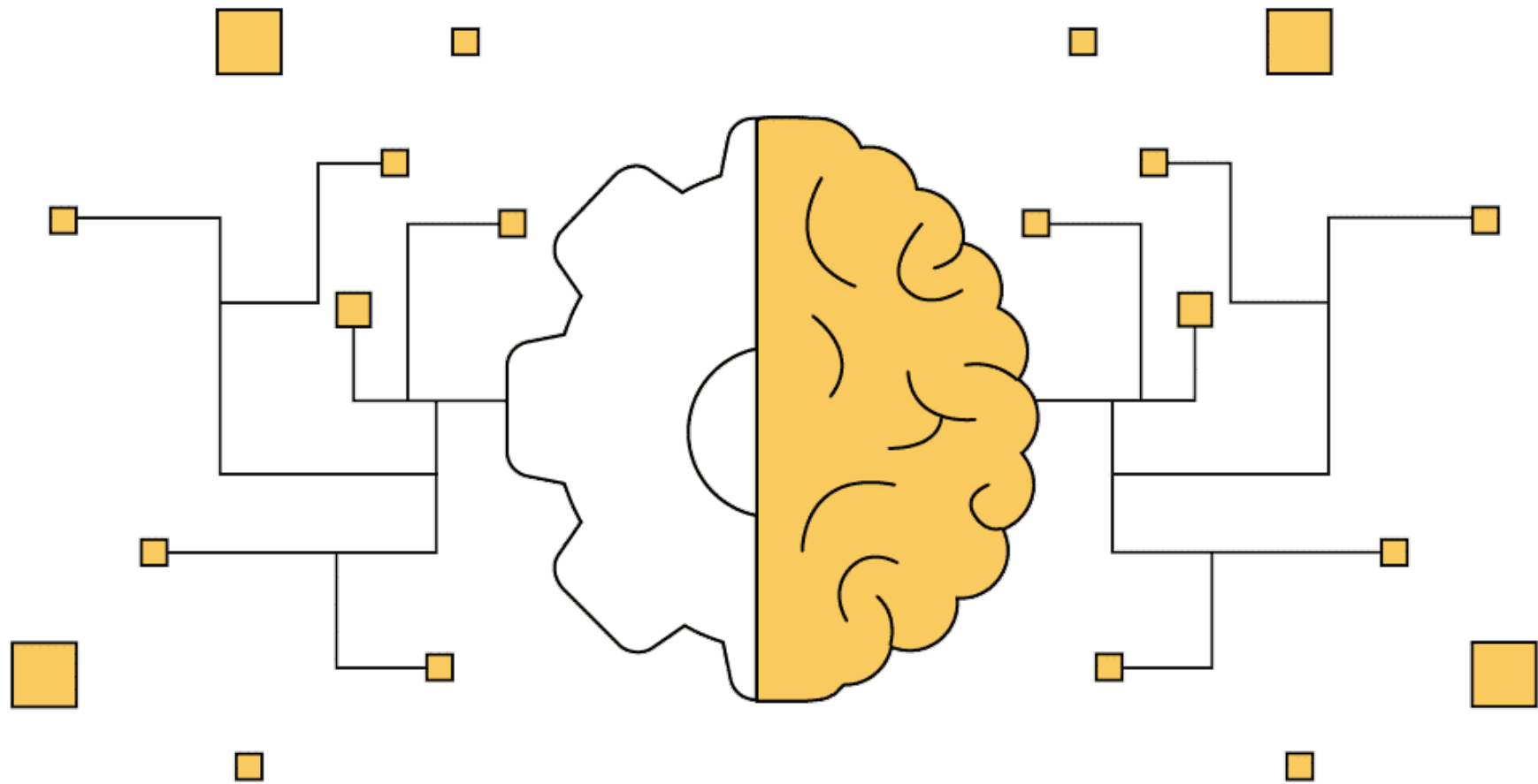


3) Zwei Analysepfade



4) Gemeinsame Reflexion

- Was wird durch Software sichtbar, was nicht?
- Welche Bedeutungen hat das Modell nicht erkannt?
- Welche Gruppen oder Probleme wurden algorithmisch unsichtbar?
- Wie beeinflusst Preprocessing (Stopwords, Tokenisierung) das Ergebnis?
- Welche Entscheidungen in der Pipeline erzeugen Bias?





Deepnote

Data notebook for the AI era



deepnote/deepnote

A Comprehensive Guide to Opening a Deepnote Account, Installation, and Usage.



Kirubasagar V

Follow

4 min read · Aug 23, 2023



2



Deepnote

Machine offline. [Restore variables](#) from last session.

Run

Create app

Topic Modeling in Deepnote

Dieses Notebook führt durch die Schritte für **Topic Modeling** mit dem generierten Datensatz

`youth_slang_200_with_metadata.csv`.

1 Bibliotheken installieren

```
1 !pip install pandas scikit-learn nltk
```

Delete this code block and [move packages "pandas", "scikit-learn", "nltk" to requirements.txt](#)

```
Requirement already satisfied: pandas in /root/venv/lib/python3.11/site-packages (2.3.3)
Requirement already satisfied: scikit-learn in /root/venv/lib/python3.11/site-packages (1.3.2)
Requirement already satisfied: nltk in /root/venv/lib/python3.11/site-packages (3.8.1)
Requirement already satisfied: numpy>=1.23.2 in /root/venv/lib/python3.11/site-packages (from pandas) (1.26.4)
Requirement already satisfied: python-dateutil>=2.8.2 in /root/venv/lib/python3.11/site-packages (from pandas) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in /root/venv/lib/python3.11/site-packages (from pandas) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in /root/venv/lib/python3.11/site-packages (from pandas) (2025.2)
Requirement already satisfied: scipy>=1.5.0 in /root/venv/lib/python3.11/site-packages (from scikit-learn) (1.11.4)
Requirement already satisfied: joblib>=1.1.1 in /root/venv/lib/python3.11/site-packages (from scikit-learn) (1.5.2)
Requirement already satisfied: threadpoolctl>=2.0.0 in /root/venv/lib/python3.11/site-packages (from scikit-learn) (3.6.0)
Requirement already satisfied: click in /root/venv/lib/python3.11/site-packages (from nltk) (8.3.0)
Requirement already satisfied: regex>=2021.8.3 in /root/venv/lib/python3.11/site-packages (from nltk) (2025.11.3)
Requirement already satisfied: tqdm in /root/venv/lib/python3.11/site-packages (from nltk) (4.67.1)
Requirement already satisfied: six>=1.5 in /root/venv/lib/python3.11/site-packages (from python-dateutil>=2.8.2->pandas) (1.17.0)
```

[notice] A new release of pip is available: 24.0 -> 25.3

[notice] To update, run: `pip install --upgrade pip`



Abschluss

Was hat gut funktioniert?

Was hat nicht geklappt?

Was hat überrascht?

Empirische Sozialforschung und Machine Learning

Evaluation

FEEDBACK GEBEN

Das Zentrum für Schul-, Bildungs- und Hochschulforschung bietet bis zum 25. November 2025 eine Evaluation des Workshops an. Ich freue mich über Feedback!

LINKS ZU EINRICHTUNGEN

Center for Critical Computational Studies der
Goethe-Universität Frankfurt am Main

Zentrum für Schul-, Bildungs- und
Hochschulforschung der JGU Mainz

GWDG KI Dienste

AG: Generative KI der Goethe-Universität
Frankfurt am Main

hessian.AI

DIENTE UND TOOLS

DokuMet QDA/AI

Deepnote

Orange data mining

huggingface.co

julius.ai

scikit-learn

KONTAKT



Dan Verständig



**Vielen Dank für die Teilnahme, Gedanken
und spannenden Diskussionen!**

LINKS ZU EINRICHTUNGEN

Center for Critical Computational Studies der
Goethe-Universität Frankfurt am Main

Zentrum für Schul-, Bildungs- und
Hochschulforschung der JGU Mainz

GWDG KI Dienste

AG: Generative KI der Goethe-Universität
Frankfurt am Main

hessian.AI

DIENTE UND TOOLS

DokuMet QDA/AI

Deepnote

Orange data mining

huggingface.co

julius.ai

scikit-learn



Dan Verständig

