Berner
Fachhochschule

Potenzial und Risiken von KI Anwendungen

Prof. Dr. Mascha Kurpicz-Briki
Applied Machine Intelligence
Bern University of Applied Sciences, Switzerland
http://www.bfh.ch/ami

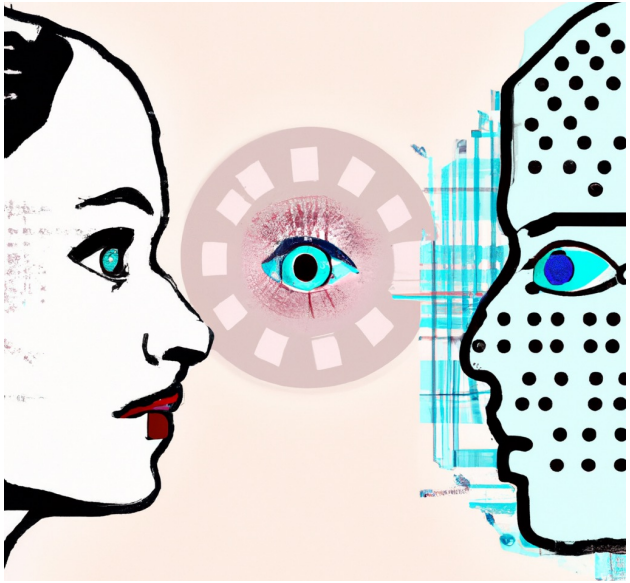
Über mich

Prof. Dr. Mascha Kurpicz-Briki

Berner Fachhochschule
Department Technik und Informatik
Institute for Data Applications and Security IDAS
Co-Leiterin Applied Machine Intelligence AMI Research Group

Forschungsinteressen:

- Anwendung digitaler Methoden auf gesellschaftliche Herausforderungen
- Natural Language Processing (NLP)
- Sprachmodelle
- Fairness in maschinellem Lernen



Created by Mascha&DALL-E

Generative AI

Viele neue Möglichkeiten durch schnellen technologischen Fortschritt

Sprachmodelle, Anwendungen wie ChatGPT, etc.

Traditionelle Software

Zutaten + spezifische Anleitung = Resultat

Artificial Intelligence

„machine learning“

Image Source: pixabay.com

Überwachtes Machine Learning

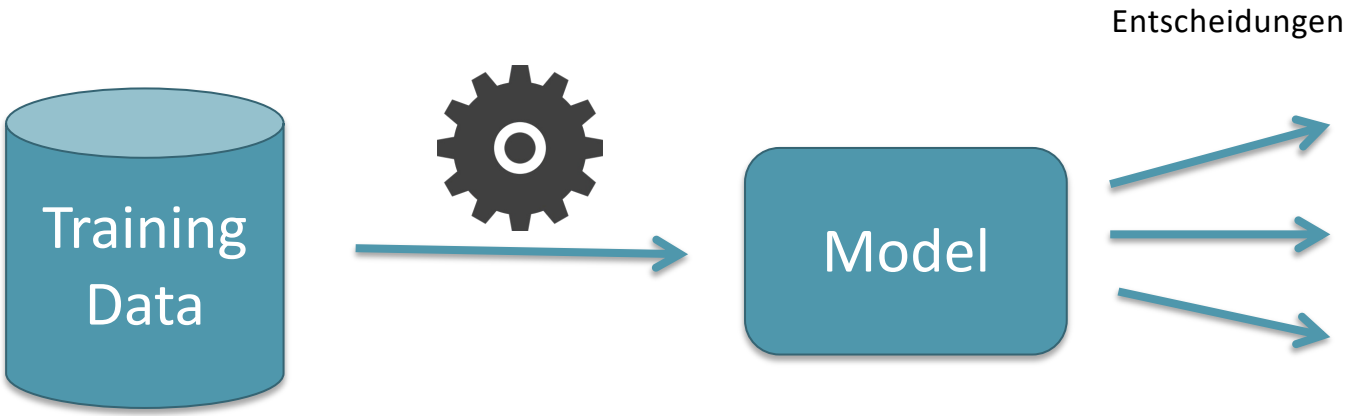
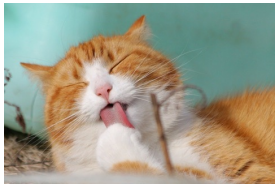


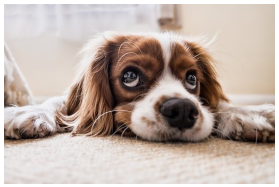
Image Source: pixabay.com

Katze



Katze

Hund



Hund

...

??

Katze

Training Data

Model

Image Source: pixabay.com

Wo KI aktuell schon erfolgreich eingesetzt wird:

- Automatische **Übersetzung** von Texten und Dokumenten
- **Generation** einfacher, repetitiver Texte (komplexere Texte in Zusammenarbeit mit Menschen)
- **Informationsextraktion** aus Dokumenten, **Klassifizierung** von Dokumenten
- **Chatbots** bei vielen gleichen Kundenanfragen (oft in Kombination mit Menschen)

Ein Blick hinter die Kulissen von KI zur Textverarbeitung

Word Embeddings

Für automatische Verarbeitung:
Mathematischer Vektor, z.B. 300 Dimensionen

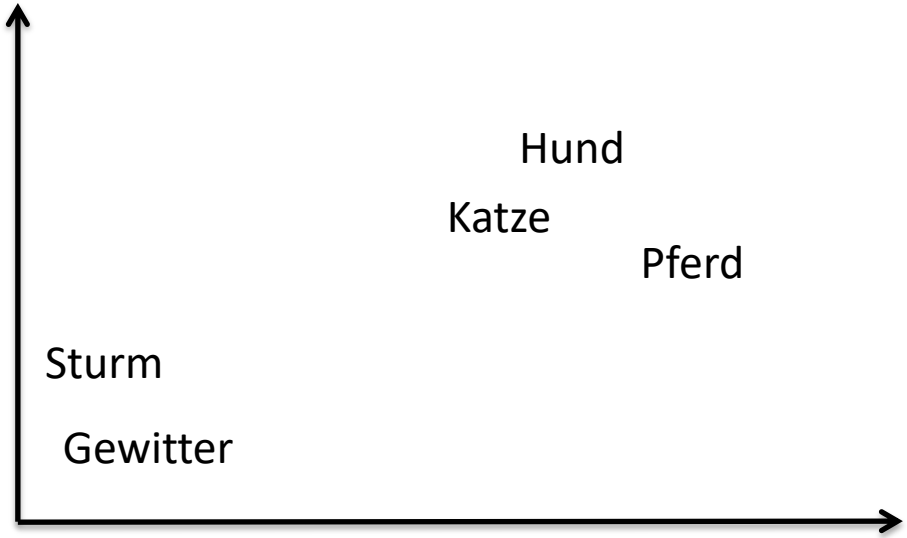
„Katze“ =

$$\begin{bmatrix} 11.2 \\ 3.4 \\ 4.5 \\ \dots \\ 6.7 \end{bmatrix}$$

Für Menschen: Wort in natürlicher Sprache, z.B. Deutsch

Image Source: pixabay.com

Word Embeddings



Wörter mit ähnlicher Bedeutung haben Vektoren, die näher beieinander sind

Image Source: pixabay.com

Eigenschaften von Word Embeddings

Diese Differenz zwischen den Vektoren kann genutzt werden:

„Man is to King, as Woman is to X“ X=Queen

weil

$$\overrightarrow{\text{Man}} - \overrightarrow{\text{Woman}} \approx \overrightarrow{\text{King}} - \overrightarrow{\text{Queen}}$$

→ Sehr nützlich für vielerlei Anwendungen!

Reference: Bolukbasi, Tolga, et al. "Man is to computer programmer as woman is to homemaker? debiasing word embeddings." *Advances in neural information processing systems*. 2016.

Risiken und Limitationen von KI Software: Stereotypen

News Headlines

« Tay: Microsoft issues apology over racist chatbot fiasco »

« Why your voice assistant might be sexist »

« Google apologises for Photos app's racist blunder »

« Amazon scraps secret AI recruiting tool that showed bias against women »

Beispiel: Bilderkennung

«Women with darker skin are more than twice as likely to be told their photos fail UK passport rules when they submit them online than lighter-skinned men, according to a BBC investigation.»

UK passport photo checker shows bias against dark-skinned women

By Maryam Ahmed
BBC News

8 October 2020

https://www.bbc.com/news/amp/technology-54349538?__twitter_impression=true

Beispiel: Automatische Auswahl von CVs

- Automatische Auswahl von Bewerbungen in Tech Firma
- Trainiert auf den Lebensläufen der letzten 10 Jahre
- Die Anwendung bekam einen Bias gegen Frauen
- Auf Grund der männlich dominierten Trainingsdaten

Source: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

Beispiel: maschinelle Übersetzung

Englisch:

*The **expert** and the **secretary** went to the bank. The **nurse** and the **doctor** went to the park.*

Maschinelle Übersetzung mit gängigen Tools zu Deutsch:

*Der **Experte** und die **Sekretärin** gingen zur Bank. Die **Krankenschwester** und der **Arzt** sind in den Park gegangen.*

Stereotypen bei Wortvektoren

Die Beziehungen zwischen den Wortvektoren sind nützlich für viele Anwendungen, aber können auch **Stereotypen** enthalten:

$$\overrightarrow{\text{father}} - \overrightarrow{\text{mother}} \approx \overrightarrow{\text{doctor}} - \overrightarrow{\text{nurse}}$$

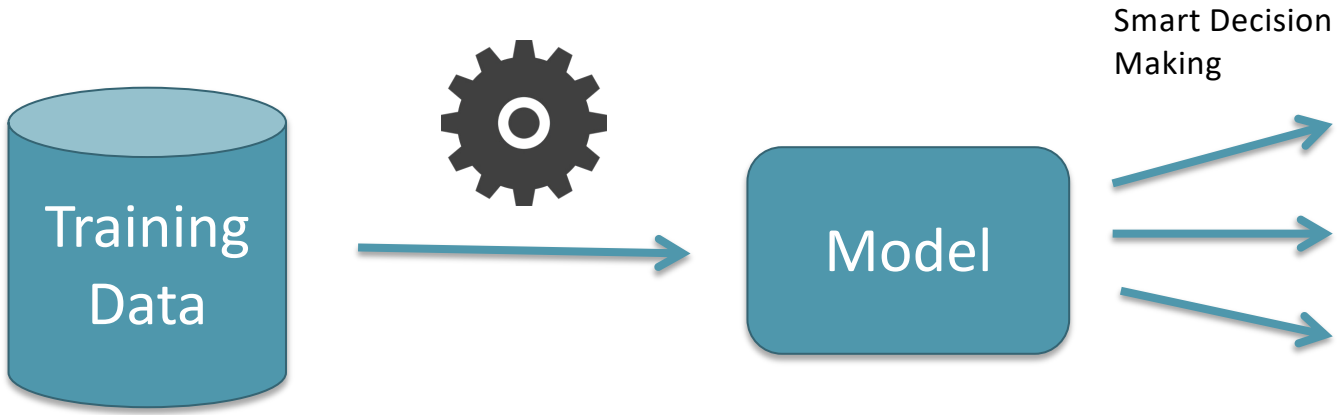
„Father is to Doctor, as Mother is to Nurse“ ??

Reference: Bolukbasi, Tolga, et al. "Man is to computer programmer as woman is to homemaker? debiasing word embeddings." *Advances in neural information processing systems*. 2016.

Stereotypen bei DALL-E

„generate a photo of a computer programmer“

Created by Mascha&DALL-E



Die Sprachmodelle enthalten Stereotypen!

Was bedeutet das für die Entscheide oder generierte Texte?

→ Kritisches Hinterfragen durch den/die Benutzer*in erforderlich

Image Source: pixabay.com

BIAS: Mitigating Diversity Biases in the Labor Market

BIAS Mitigating biases of AI in the labour market

www.biasproject.eu

What is our mission?

Empower the Artificial Intelligence (AI) and Human Resources Management (HRM) communities by addressing and mitigating algorithmic biases.

- Wie werden KI Anwendungen auf dem Arbeitsmarkt eingesetzt?
- Wie wird menschlicher Bias in der KI und Sprachmodellen reflektiert?
- Wie kann solcher Bias gemessen und reduziert werden?

Horizon Europe (HORIZON)

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Staatssekretariat für Bildung, Forschung und Innovation SBFI

BIAS: Mitigating Diversity Biases in the Labor Market

BIAS Mitigating biases of AI in the labour market

www.biasproject.eu

What is our mission?

Empower the Artificial Intelligence (AI) and Human Resources Management (HRM) communities by addressing and mitigating algorithmic biases.

Horizon Europe (HORIZON)

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Staatssekretariat für Bildung, Forschung und Innovation SBFI

Jetzt mitmachen, jede Stimme zählt!

Umfrage Erfahrungen/ Meinungen zu KI auf dem Arbeitsmarkt

https://www.biasproject.eu/surveys/

Infos erhalten über Ergebnisse, Workshops und andere Aktivitäten: Anmelden bei den **National Labs** https://www.biasproject.eu/nationallabs/

Wie soll die digitale Gesellschaft der Zukunft aussehen?

Image Source: pixabay.com

Zwei Arten von AI

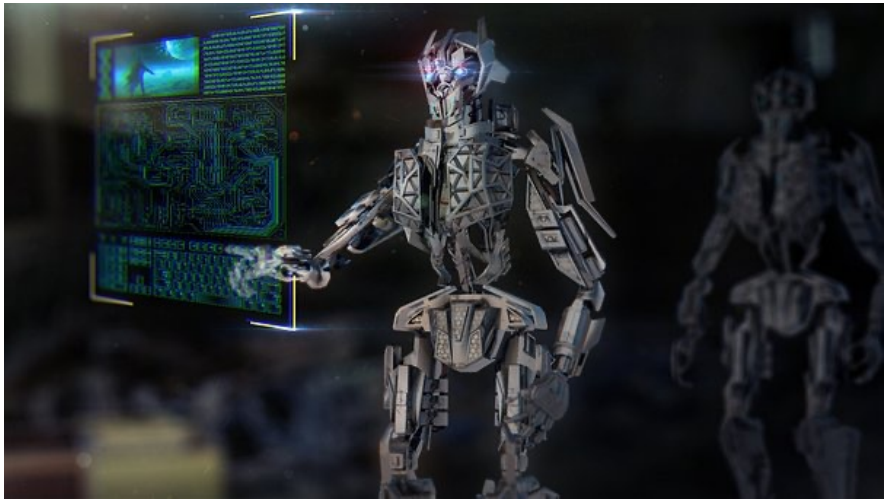


Image Source: pixabay.com

Augmented Intelligence, statt Artificial Intelligence

- Menschen unterstützen, anstelle sie zu ersetzen

- AI als Tool zur Unterstützung bei repetitiven Arbeiten, damit der Mensch mehr Zeit hat für andere Aufgaben

- Nützliches Werkzeug, Menschen zu ergänzen

Image Source: pixabay.com

Fazit

- Die KI-Software als Werkzeug sehen und verantwortungsvoll einsetzen

- Sich den Limitationen bewusst sein
  - Ist der Anwendungsfall geeignet?
  - Generierte Inhalte auf Korrektheit prüfen
  - Kritisch hinterfragen (Stereotypen?)

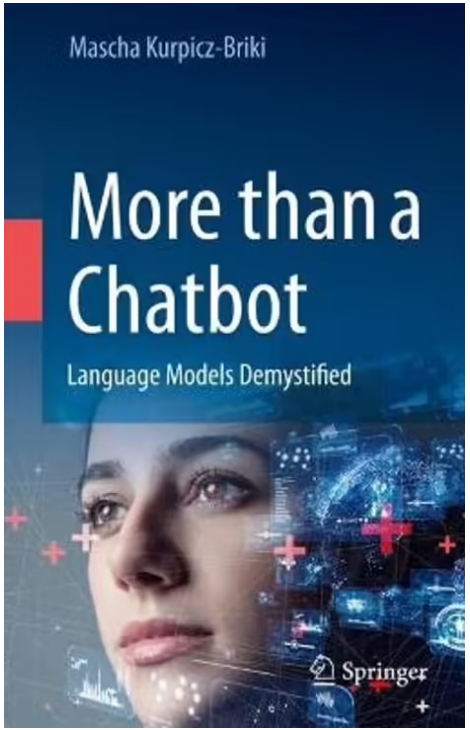
Image Source: pixabay.com

Berner
Fachhochschule

Prof. Dr. Mascha Kurpicz-Briki
Applied Machine Intelligence
Bern University of Applied Sciences
http://www.bfh.ch/ami

mascha.kurpicz@bfh.ch

@SocietyData



Erscheint Nov. 2023
Jetzt vorbestellen:

(Ausgabe auf Deutsch geplant 2024)