

# Mehr als nur Bahnhof verstehen

## KI für Sprachtechnologien

# Sprachtechnologie

- Mensch-Maschine Kommunikation
  - Oft:
    - Mensch muss sich an Computer anpassen



# Sprachtechnologie

- Mensch-Maschine Kommunikation
  - Oft:
    - Mensch muss sich an Computer anpassen



# Sprachtechnologie

- Mensch-Maschine Kommunikation
  - Oft:
    - Mensch muss sich an Computer anpassen
  - Ziel:
    - Natürliche Kommunikation



# Sprachtechnologie

- Mensch-Maschine Kommunikation
- Unterstützung von Mensch-Mensch Kommunikation
  - Beispiel:
    - Maschinelle Übersetzung



# Sprachtechnologie

- Mensch-Maschine Kommunikation
- Lecture translator am KIT
- Unterstützung von Mensch-Mensch Kommunikation
  - Beispiel:
    - Maschinelle Übersetzung



# Sprachtechnologie

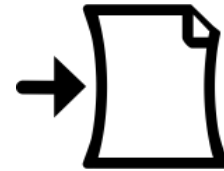
- Mensch-Maschine Kommunikation
- Unterstützung von Mensch-Mensch Kommunikation
- Informationszugriff



# Aufgaben



Spracherkennung



Automatisches Zusammenfassen



Maschineller Übersetzung



Automatisches Beantworten von  
Fragen



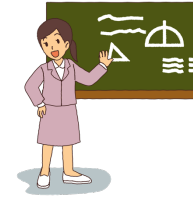
Dialogmodellierung



# Herausforderungen

## ■ Mehrdeutigkeiten

„If the baby does not like the milk, boil it“



Vorzeichen



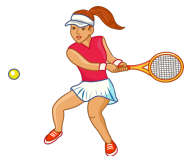
sign



omen

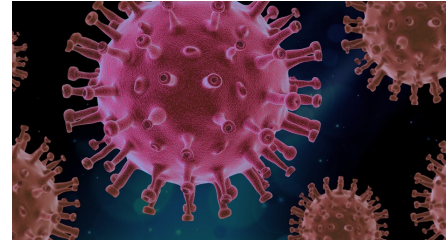


I am happy to introduce one of the best tennis players in the world.

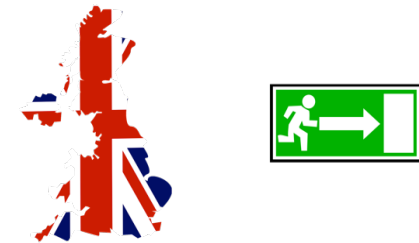


# Herausforderungen

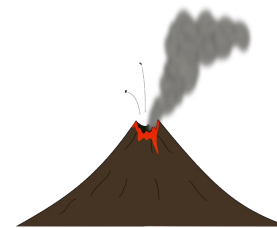
- Mehrdeutigkeiten
- Neue Wörter



Covid



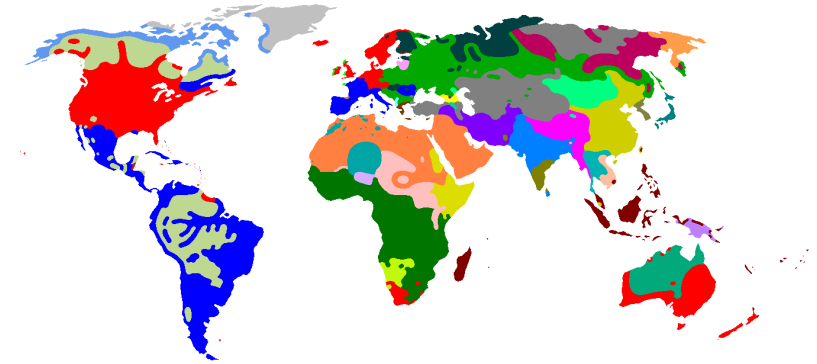
Brexit



Eyjafjallajökull

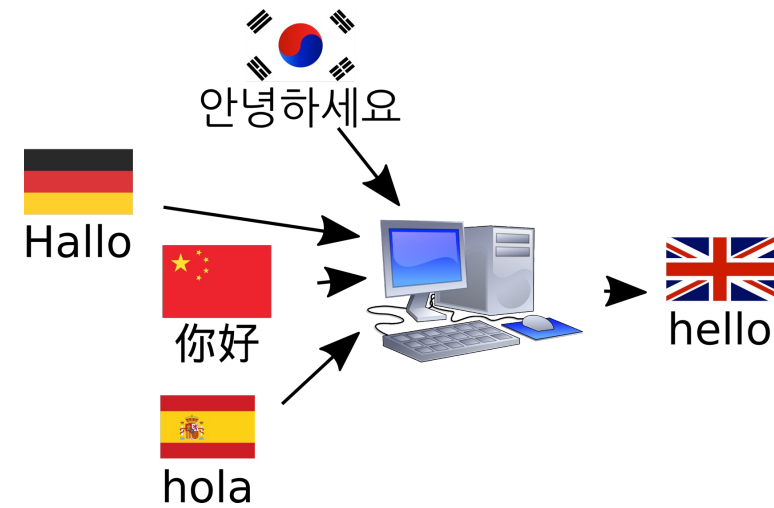
# Herausforderungen

- Mehrdeutigkeiten
- Neue Wörter
- Diversität
  - 6000-7000 Sprachen
  - Modalitäten

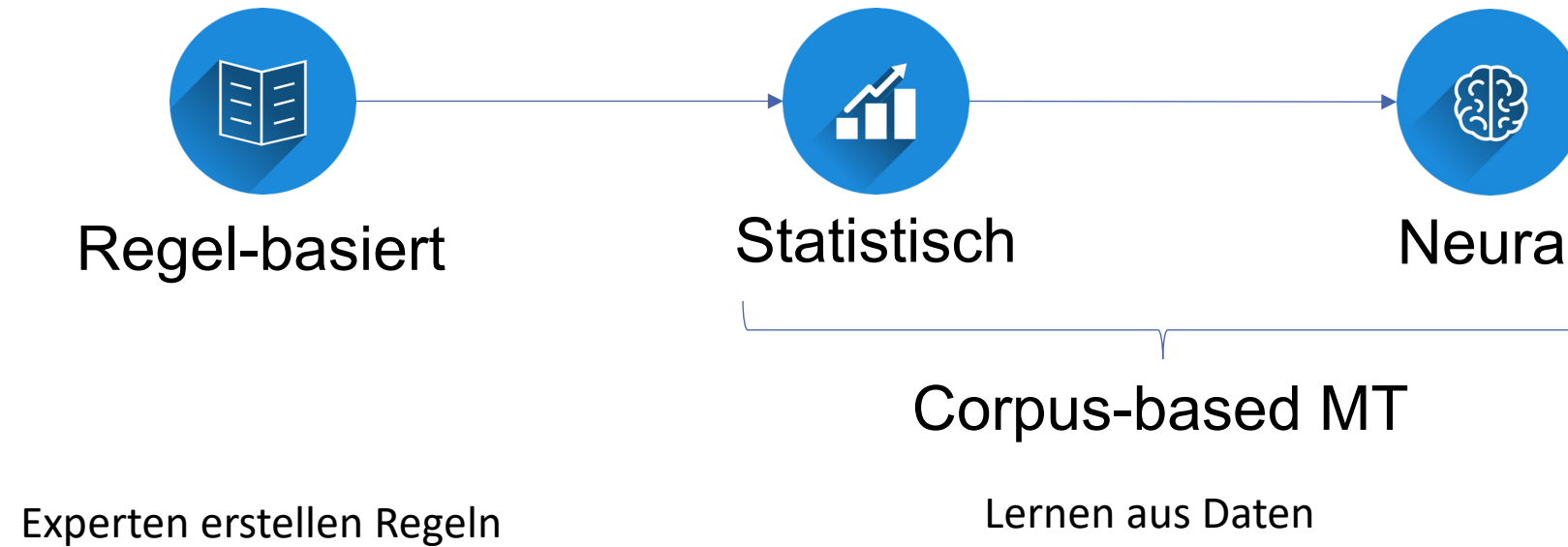


# Beispiel

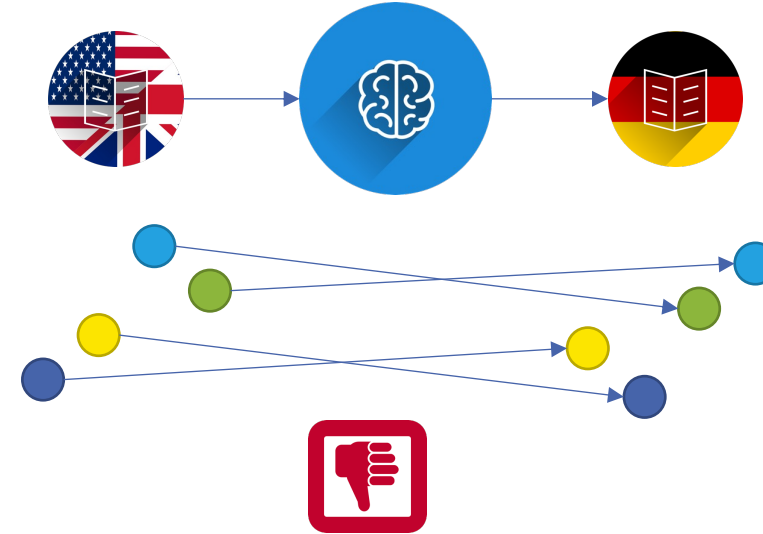
## ■ Maschinelle Übersetzung



# Ansätze der Maschinellen Übersetzung



# Corpus-basierte Maschinelle Übersetzung



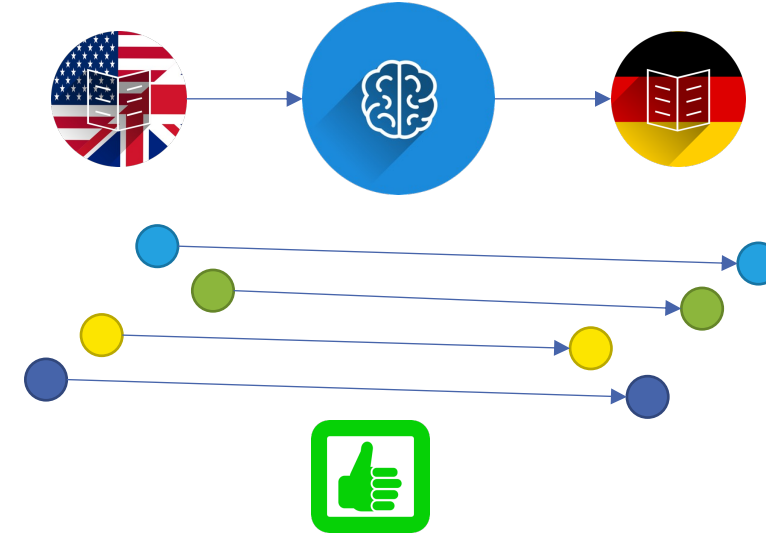
## ■ Daten

- Paralleler corpus
  - Rosetta stone
- Ein Quellsprachensatz mit Übersetzung

## ■ Machine Learning Model

- Quellsatz auf Zielsatz abbilden
- Durch Beispiele lernen

# Corpus-basierte Maschinelle Übersetzung



## ■ Daten

- Paralleler corpus
  - Rosetta stone
- Ein Quellsprachensatz mit Übersetzung

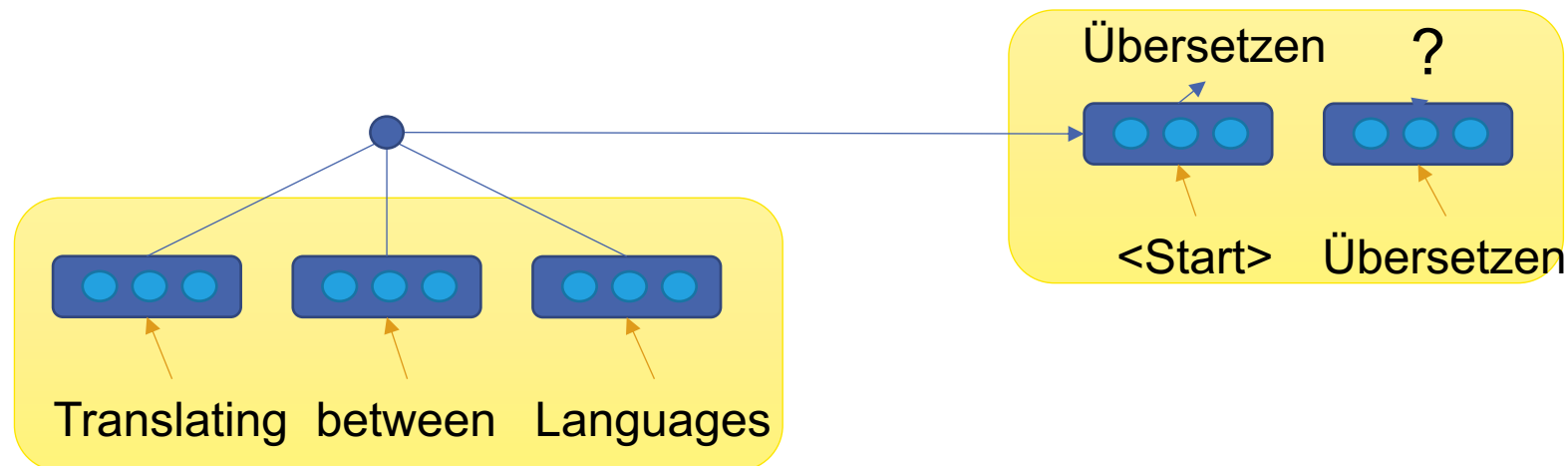
## ■ Machine Learning Model

- Quellsatz auf Zielsatz abbilden
- Durch Beispiele lernen

# Neurale Maschinelle Übersetzung

## ■ Encoder-Decoder Model

- Künstliches Neuronales Netz das die Übersetzung modelliert
- Encoder: Repräsentation des Quellsatzes
- Decoder: Generierung Wort für Wort





# Aktuelle Forschung

## ■ Dateneffizienz



< 1 Million Wörter



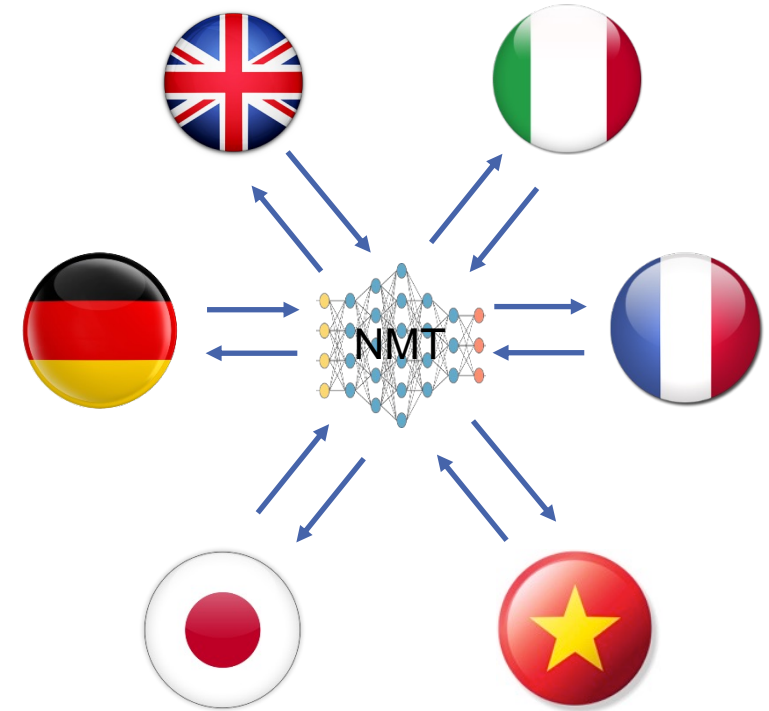
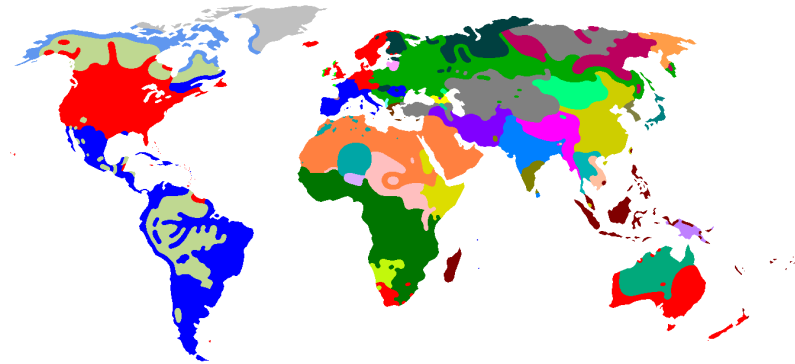
ca. 500 Million Wörter



7,5 Milliarden Wörter

# Aktuelle Forschung

- Dateneffizienz
- Multilinguality



# Aktuelle Forschung

- Dateneffizienz
- Multilinguality
- Kontinuierliches Lernen

Baseline:	Das binäre Zahlensystem ist ...	The binary payment system is
Copy	Das binäre # Zahlensystem # numeral system # ist	The binary numeral system is

# Aktuelle Forschung

- Dateneffizienz
- Multilinguality
- Kontinuierliches Lernen
- Ausgabeformate



German	Ich	melde	mich	zur	Summer	School	an
Gloss	I	register/ cancel	myself	to	summer	School	
English	I	????					

# AI for Language Technologies

## ■ Weitere Fragen?



## ■ Kontakt:

- [jan.niehues@kit.edu](mailto:jan.niehues@kit.edu)
- <https://ai4lt.anthropomatik.kit.edu/>

