



NLP Lab 3

Textgenerierung

# Wir erinnern uns

## Natürliche Sprache ...

... ist unendlich.

... ist nicht normiert.

... referenziert auf außersprachliche Dinge.

... verfügt über Subtext.

...

# Wir stellen fest

## Natürliche Sprache ist außerdem ...

... von Gruppe zu Gruppe unterschiedlich (Dialekte, Soziolekte, Funktiolekte).

... von Person zu Person auch innerhalb derselben Gruppe unterschiedlich.

... bei ein und derselben Person innerhalb verschiedener Kontexte unterschiedlich.

...

# Wir fixieren Person und Kontext

„Habe nun, ach! Philosophie, // Juristerei und Medizin, // Und leider auch Theologie! //  
Durchaus studiert, mit heißem Bemühn. // Da steh ich nun, ich armer Tor! // Und bin so  
klug als wie zuvor.“ (Goethe, *Faust*)

„Alles lebt, um traurig wieder zu sterben. Wir interessieren uns nur darum, wir gewinnen  
nur darum, daß wir wieder mit Schmerzen verlieren.“ (Schiller, *Räuber*)

„Ach! wär' ich nie in eure Schulen gegangen! Die Wissenschaft, der ich in den Schacht  
hinunter folgte, von der ich, jugendlich thöricht, die Bestätigung meiner reinen Freude  
erwartete, die hat mir alles verdorben. // Ich bin bei euch so recht vernünftig geworden,  
habe gründlich mich unterscheiden gelernt von dem, was mich umgibt, bin nun vereinzelt  
in der schönen Welt, bin so ausgeworfen aus dem Garten der Natur, wo ich wuchs und  
blühte, und vertrockne an der Mittagssonne.“ (Hölderlin, *Hyperion*)

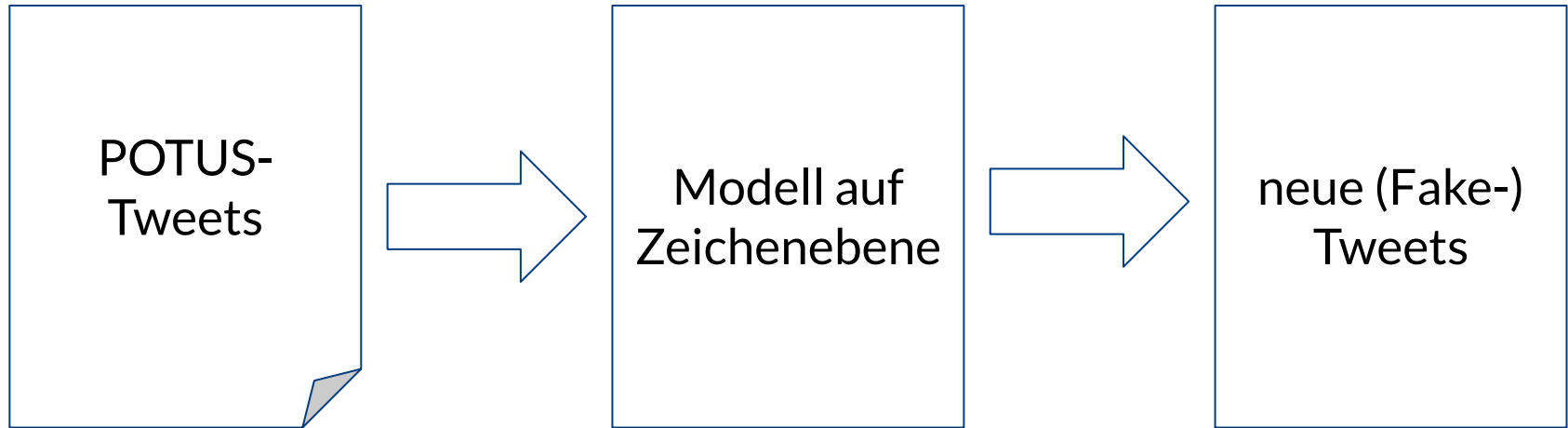
# Es geht auch moderner

Congrats to [@U S Steel](#) for investing \$1+ BILLION in America's most INNOVATIVE steel mill. 232 Tariffs make Pennsylvania and USA more prosperous/secure by bringing Steel and Aluminum industries BACK. Tariffs are working. Pittsburgh is again The Steel City. USA Economy is BOOMING!  
(@realDonaldTrump)

I am so proud of each and every student out there who's celebrating [#CollegeSigningDay](#). I know how hard you've worked and how many hurdles each of you had to clear to make it to today. I can't wait to see what you all do in the years ahead! [#BetterMakeRoom](#)  
(@MichelleObama)

Aufgabe heute: Wir bauen uns einen  
Trump-Tweet-Generator

# Vorgehen (schematisch ;-))

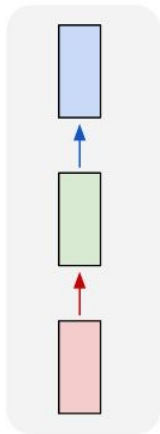


(LSTMs mit "Gedächtnis")

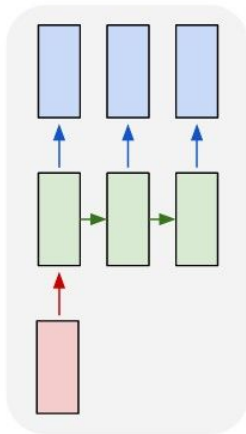
# Vorgehen in Stichworten

- **Vorverarbeitung:**
  - Tweet => Tokenliste => Zeichenliste => One-Hot-Vektor
- **Model:**
  - LSTM
  - Many-to-one: gegeben n Zeichen, sage n+1-tes Zeichen voraus

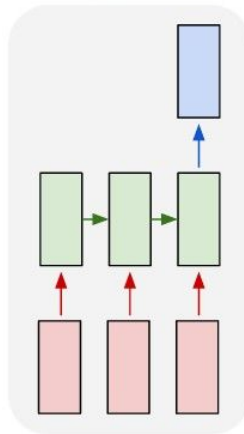
one to one



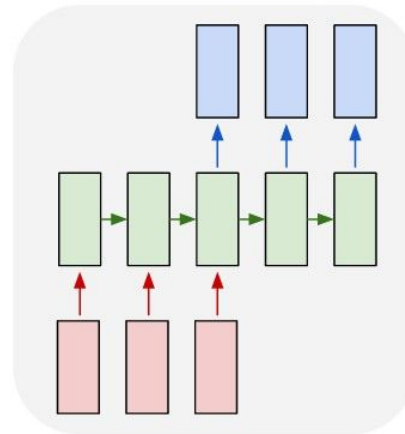
one to many



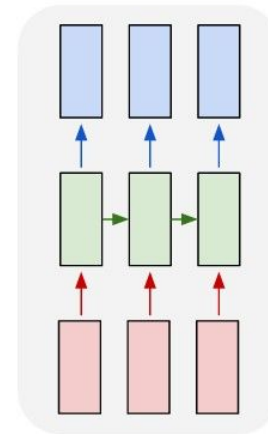
many to one



many to many



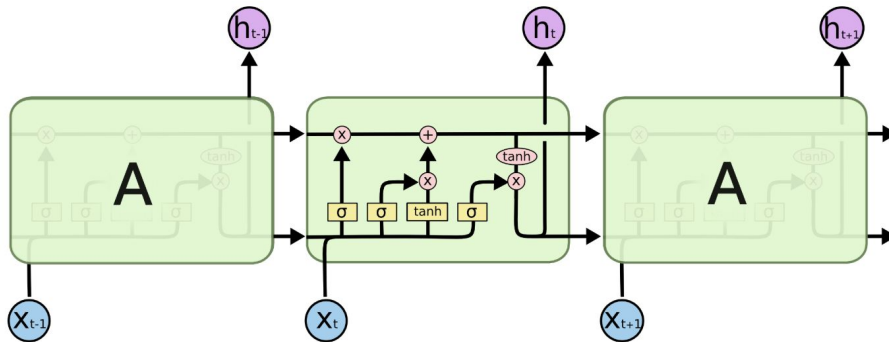
many to many



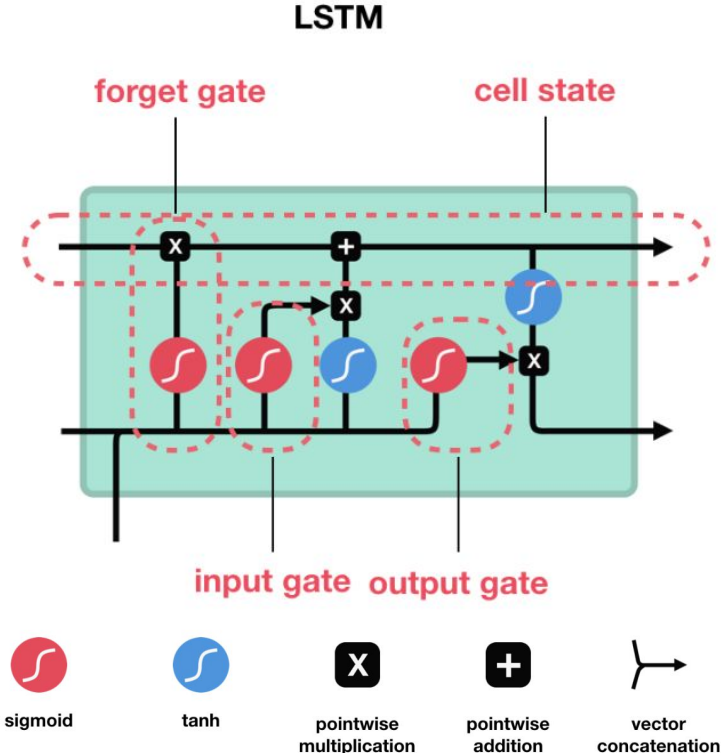


# LSTM -- Funktionsweise

- Long Short Term Memory Network
- Spezielles RNN
- Verfügt über Langzeitgedächtnis
- Beinhaltet Gates (NNs), die Zellzustand beeinflussen



The repeating module in an LSTM contains four interacting layers.



[https://github.com/hskaailabnlp/text\\_generation](https://github.com/hskaailabnlp/text_generation)

# Vielen Dank

Anna Weißhaar

[aweisshaar@inovex.de](mailto:aweisshaar@inovex.de)

Tilman Berger

[tberger@inovex.de](mailto:tberger@inovex.de)

