Statistische Sprachmodelle & ChatGPT

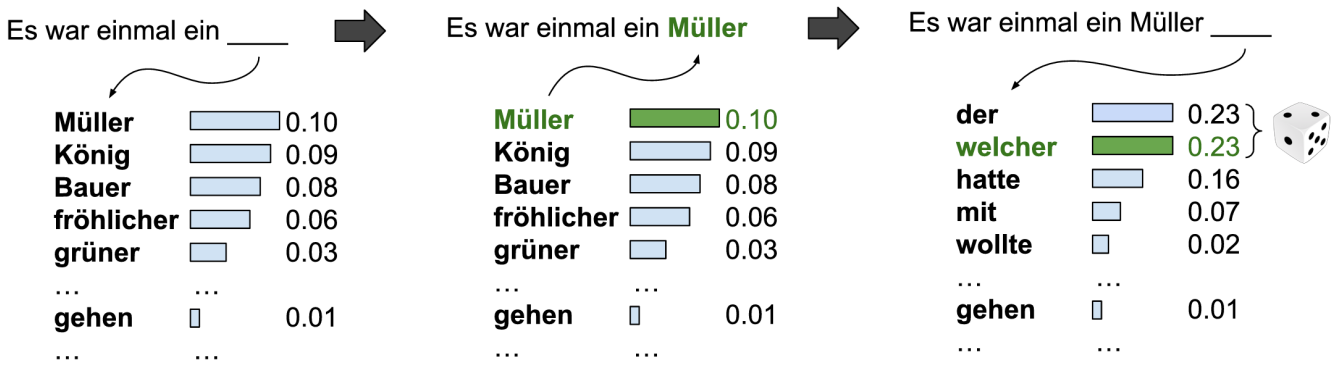
ChatGPT

Ziele

- Grundidee von **statistischen Sprachmodell** verstehen
- Selber mit einfachem Sprachmodul arbeiten
- Damit besseres Verständnis für ChatGPT erlangen

Statistische Sprachmodelle

- Basieren aus rieseigen Sammlungen von Texten: Wikipedia, Bücher, …
- Damit wird Sprachmodell *trainiert*
- Modell lernt dabei, welches Zeichen / Wort mit welcher *Wahrscheinlichkeit* als nächstes folgt
- Beispiel:
  - Gib dem Sprachmodell den Satzanfang: «Es war einmal …»
  - Im Training hat Modell gelernt, dass nach den drei Worten «Es war einmal» meistens das Wort «ein» kommt.
  - Weiter …



Text erzeugen

Soekia GPT

Schreibe mir ein Märchen.

Es war einmal ...

Wortvorschläge

Auswahl anpassen

1er und
1er .
1er die
1er sie
1er der
1er das
1er er
1er es
1er in
1er den
1er zu
1er so
1er sich
1er ein
1er dem
1er nicht
1er aber
1er auf

N-Gramme

-1 +1

1er 2er 3er 4er 5er

N-Gramme erstellen

in den Wald
. Als sie
. Und als
. Als es
. Das Schneiderlein
. Es war
. Der Dummling
. Der Königssohn
. Da gieng
. Als er
den Wald und
der Hand und
pik pik pik
. Das Mädchen
. Der Riese
. Die Frau
auf einen Streich
Es war einmal
Hänsel und Grethel
siebene auf einen

Dokumente: 12

hatte sieben Söhne und immer noch kein Töchterchen, so sehr er sichs auch wünschte; endlich gab ihm seine Frau wieder gute Hoffnung zu einem Kinde, und wies zur Welt kam, wars auch ein Mädchen. Die Freude war gross, aber das Kind war schmächtig und klein, und sollte wegen seiner Schwachheit die

wurde seine Frau krank, und als sie fühlte dass ihr Ende heran kam, rief sie ihr einziges Töchterlein zu sich ans Bett und sprach „liebes Kind, bleib fromm und gut, so wird dir der liebe Gott immer beistehen, und ich will vom Himmel auf dich herabblicken, und will um dich sein." Darauf tat sie die Augen

Dokument I

Rumpelstilzchen.

Es war einmal ein Müller, der war arm, aber er hatte eine schöne Tochter. Nun traf es sich, dass er mit dem König zu sprechen kam, und um sich ein Ansehen zu geben, sagte er zu ihm „ich habe eine Tochter, die kann Stroh zu Gold spinnen." Der König sprach zum Müller „das ist eine Kunst, die mir wohl gefällt, wenn deine Tochter so

Dokument J

Rapunzel.

Es war einmal eine Frau und ein Mann, die wünschten sich schon lange vergeblich ein Kind, endlich machte sich die Frau Hoffnung der liebe Gott werde ihren Wunsch erfüllen. Die Leute hatten in ihrem Hinterhaus ein kleines Fenster. Aus diesem konnte man in einen prächtigen Garten sehen, der voll der schönsten Blumen und Kräuter

Dokument K

Das tapfere Schneiderlein.

An einem Sommermorgen sass ein Schneiderlein auf seinem Tisch am Fenster, war guter Dinge und nähte aus Leibeskräften. Da kam eine Bauersfrau die Strasse herab und rief „gut Mus feil! gut Mus feil!" Das klang dem Schneiderlein lieblich in die Ohren, er steckte sein zartes Haupt zum Fenster hinaus und

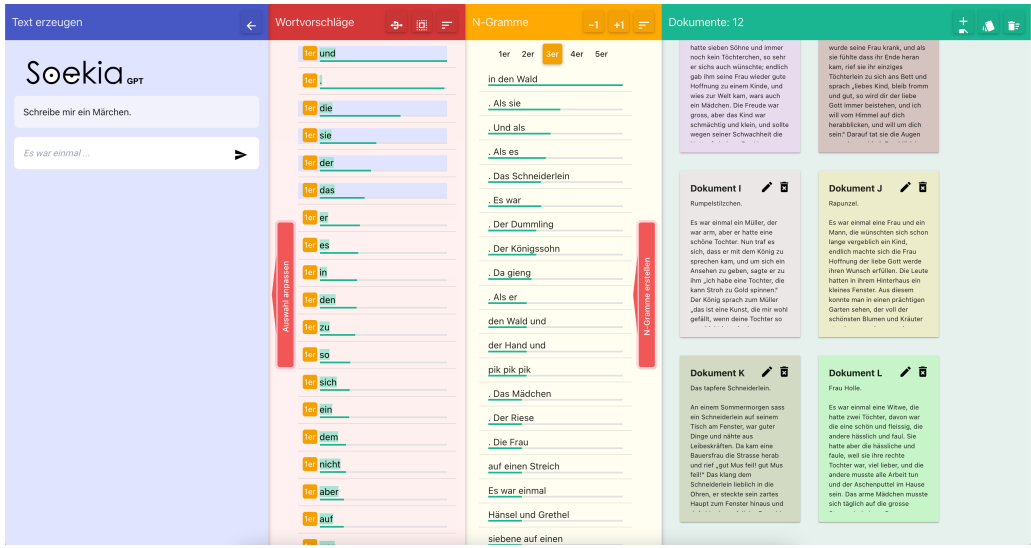
Dokument L

Frau Holle.

Es war einmal eine Witwe, die hatte zwei Töchter, davon war die eine schön und fleissig, die andere hässlich und faul. Sie hatte aber die hässliche und faule, weil sie ihre rechte Tochter war, viel lieber, und die andere musste alle Arbeit tun und der Aschenputtel im Hause sein. Das arme Mädchen musste sich täglich auf die grosse

SoekiaGPT

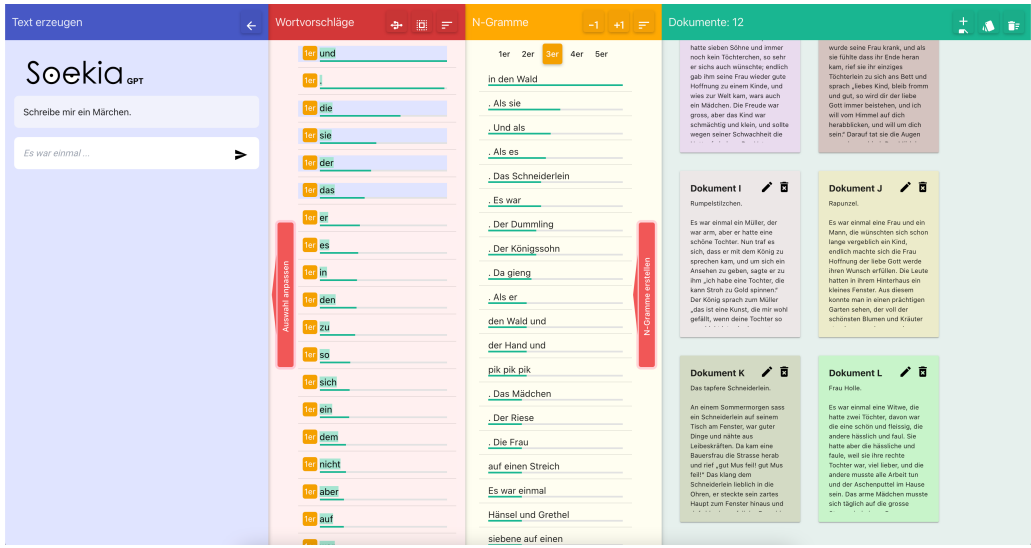
- Ist Trainiert mit insg. 12 Märchen / Geschichten
- Erstellt dabei Liste mit N-Grammen:
  - Wortfolgen mit N Worten
  - Zählt wie oft diese vorkommen in allen Texten
  - Beispiel: Häufigstes 4-Gramm ist «in den Wald und»:
    8x gefunden in 3 Dokumenten



1-Gramme	2-Gramme	3-Gramme	4-Gramme
Der	Der Müller	Der Müller hatte	Der Müller hatte eine
Müller	Müller hatte	Müller hatte eine	Müller hatte eine Tochter
hatte	hatte eine	hatte eine Tochter	hatte eine Tochter .
eine	eine Tochter	eine Tochter .	eine Tochter . Sie
Tochter	Tochter .	Tochter . Sie	Tochter . Sie war
.	. Sie	. Sie war	. Sie war sehr
Sie	Sie war	Sie war sehr	Sie war sehr schön
war	war sehr	war sehr schön	war sehr schön .
sehr	sehr schön	sehr schön .	
schön	schön .		
.			

SoekiaGPT

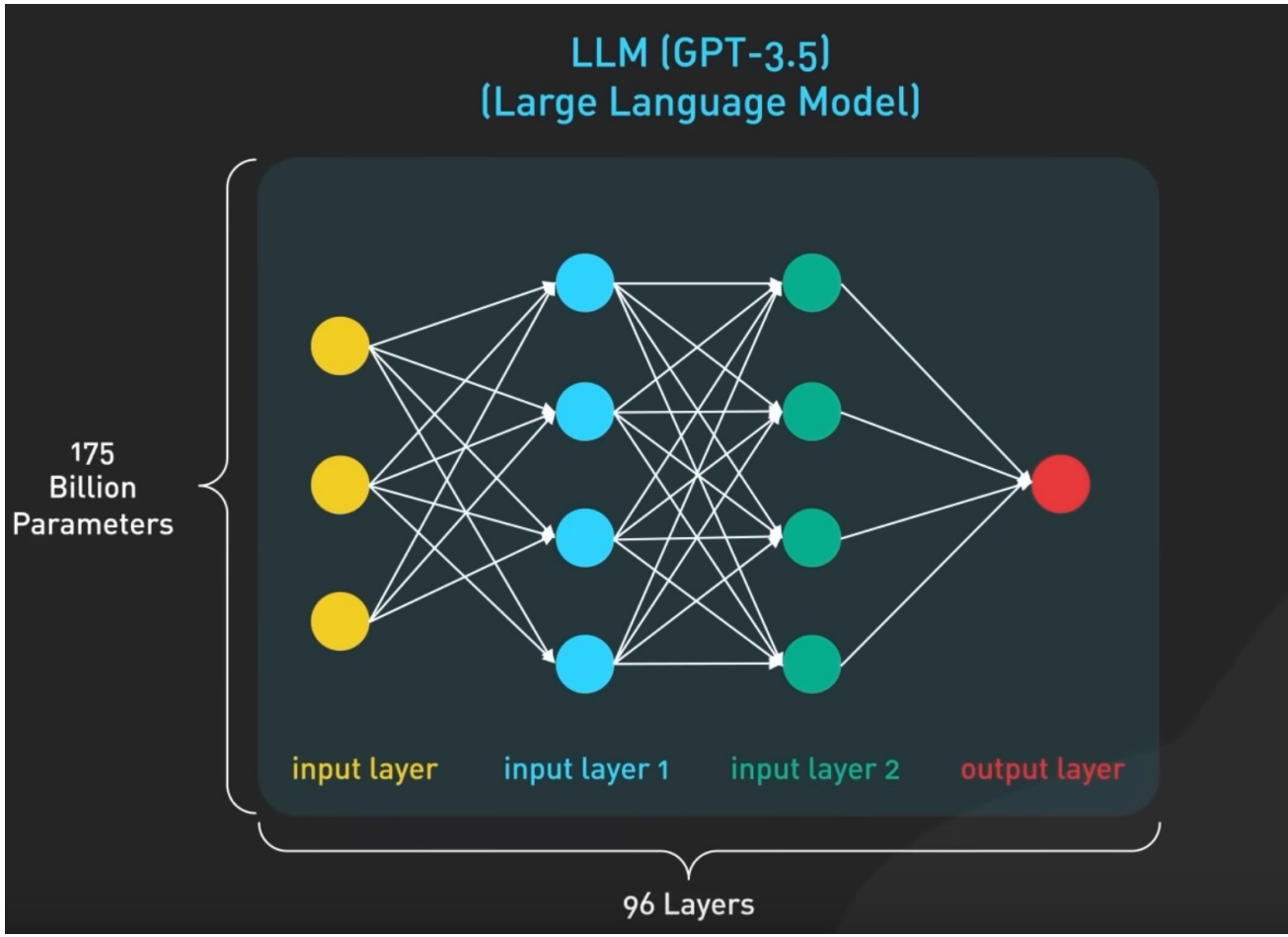
- Texterzeugung:
  - alle N-Gramme werden mit zuletzt eingegebene Wörter («Es war einmal …») verglichen
  - Alle die passen, werden in Kandidaten-Liste aufgenommen
  - Dann wird daraus eines ausgewählt
  - Kandidat hat gute Karten falls …
    - Kandidat oft Vorkommt in Dokumenten
    - N-Gramm mit höherem N
  - Aber auch Zufall (Würfeln) spielt mit



1-Gramme	2-Gramme	3-Gramme	4-Gramme
Der	Der Müller	Der Müller hatte	Der Müller hatte eine
Müller	Müller hatte	Müller hatte eine	Müller hatte eine Tochter
hatte	hatte eine	hatte eine Tochter	hatte eine Tochter .
eine	eine Tochter	eine Tochter .	eine Tochter . Sie
Tochter	Tochter .	Tochter . Sie	Tochter . Sie war
.	. Sie	. Sie war	. Sie war sehr
Sie	Sie war	Sie war sehr	Sie war sehr schön
war	war sehr	war sehr schön	war sehr schön .
sehr	sehr schön	sehr schön .	
schön	schön .		
.			

Auftrag

1. Generiere Geschichte. Spiele dabei mit:
   - «automatisch fortsetzen» vs. «selber auswählen»
   - Versch. Werte für N (N-Gramme)
   - «Auswahl anpassen»
     - «Anzahl ausgewählter Wortvorschläge»
     - «Temperatur»: Zufallsfaktor
2. Füge weitere Dokumente hinzu
   - Z.B. von «https://www.gutenberg.org/», Goethes Faust
3. Challenge: Wer lässt die beste Geschichte generieren?

ChatGPT

- Ähnlichkeiten mit statistischen Sprachmodellen …
- … basiert aber auf einem neuronalen Netzwerk (NN)
- Das NN erlaubt es, dass es nicht nur mit Statistik arbeitet, sondern auch Muster erkennen und erlernen kann