

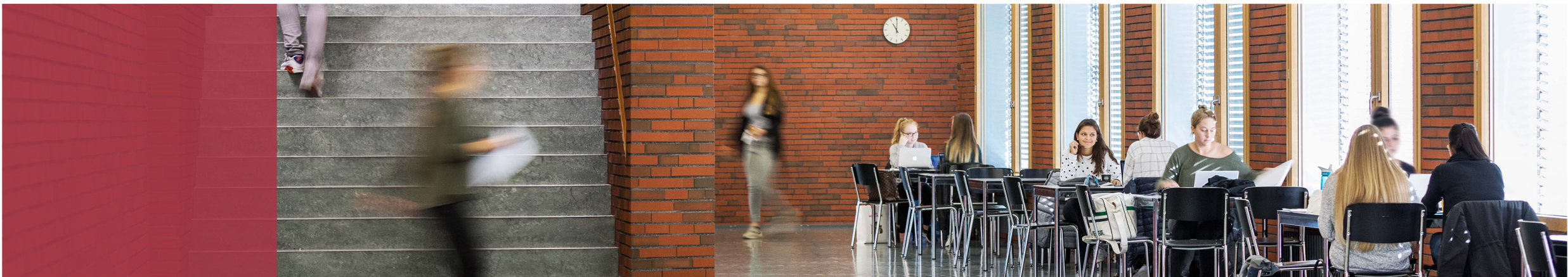


Angewandte Linguistik
ILC Institute of
Language Competence

Wie funktioniert automatische Textproduktion mit grossen Sprachmodellen?

Cerstin Mahlow, Professorin Digitale Linguistik und Schreibprozessforschung

31. Mai 2023



Automatische Textproduktion



```
graph TD; A[Automatische Textproduktion] --- B[Technische Grundlagen]; A --- C[Einsatzmöglichkeiten]; A --- D[Konsequenzen];
```

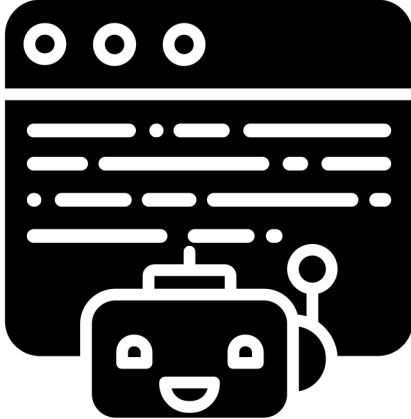
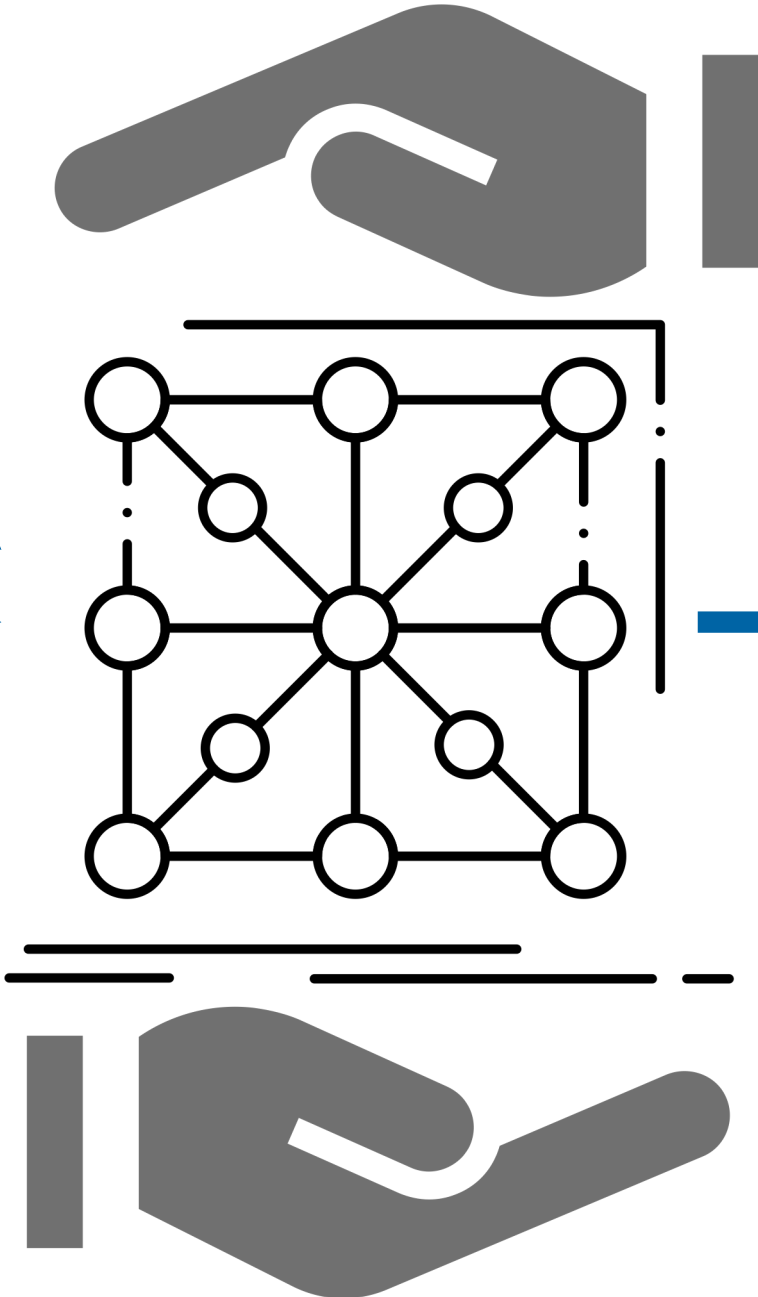
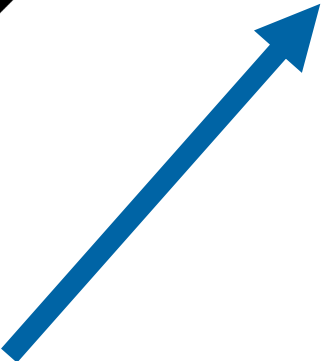
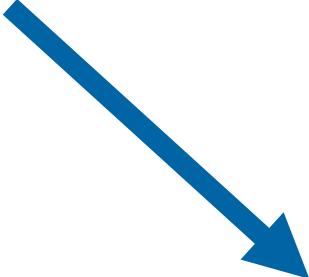
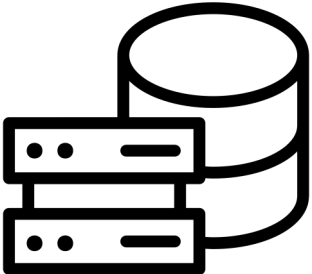
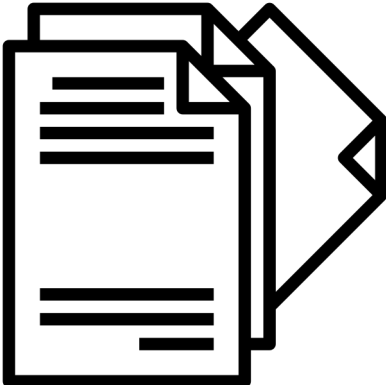
The diagram is a hierarchical tree structure. At the top is a white box with a blue border containing the text 'Automatische Textproduktion'. A vertical line descends from the bottom center of this box to a horizontal line. From the left end of this horizontal line, a vertical line goes down to a solid blue box containing the text 'Technische Grundlagen'. From the center of the horizontal line, a vertical line goes down to a white box with a light blue border containing the text 'Einsatzmöglichkeiten'. From the right end of the horizontal line, a vertical line goes down to a white box with a light blue border containing the text 'Konsequenzen'.

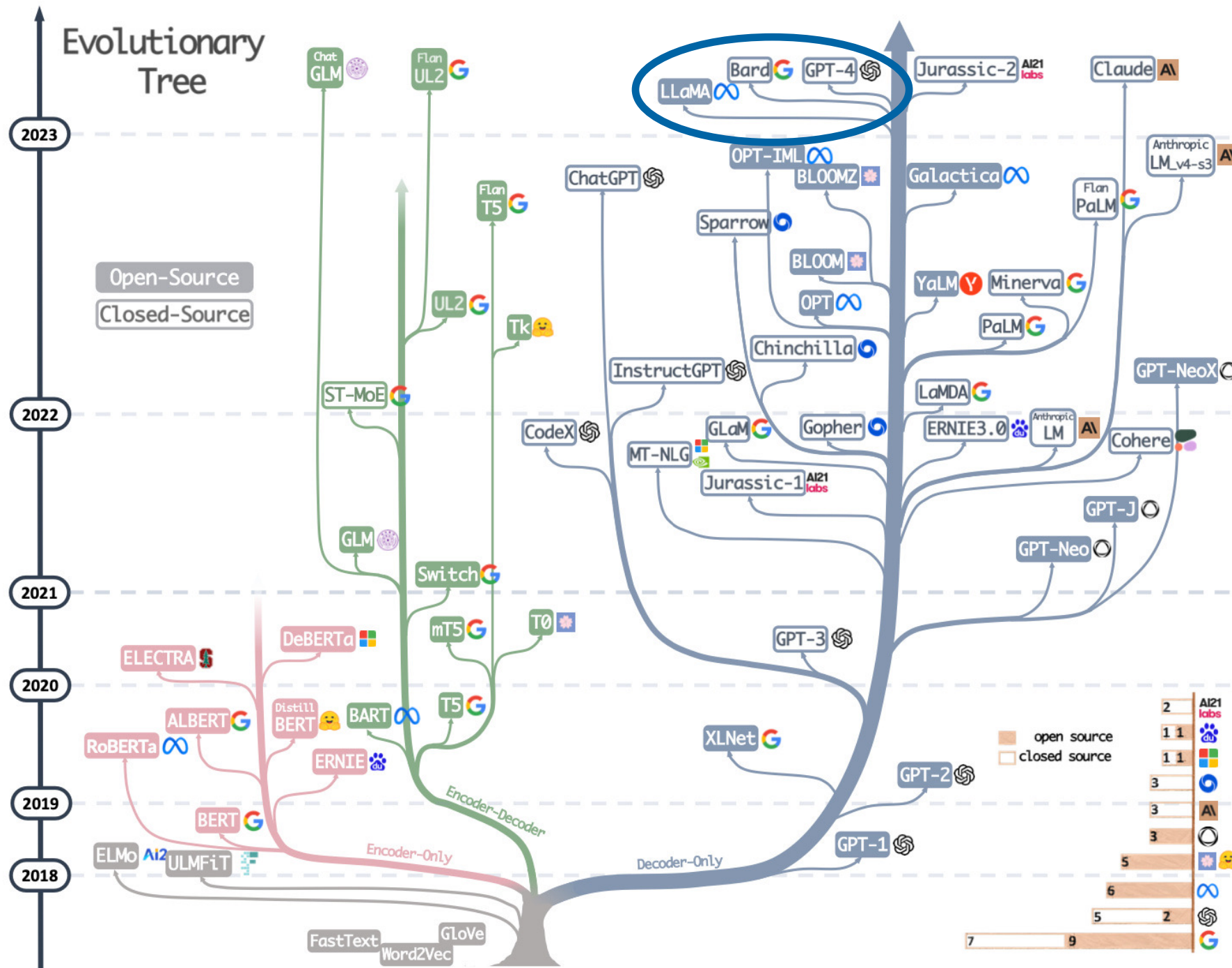
Technische Grundlagen

Einsatzmöglichkeiten

Konsequenzen

Grosse Sprachmodelle





Chatbots

Computational Linguistics

A. G. OETTINGER, Editor

1966

ELIZA—A Computer Program For the Study of Natural Language Communication Between Man And Machine

JOSEPH WEIZENBAUM
Massachusetts Institute of Technology, Cambridge, Mass.*

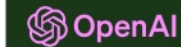
The object of this paper is to cause just such a re-evaluation of the program about to be “explained”. Few programs ever needed it more.

ELIZA Program

ELIZA is a program which makes natural language conversation with a computer possible. Its present implementation is on the MAC time-sharing system at MIT. It is written in MAD-SLIP [4] for the IBM 7094. Its name was chosen to emphasize that it may be incrementally improved by its users, since its language abilities may be continually improved by a “teacher”. Like the Eliza of Pygmalion fame, it can be made to appear even more civilized, the relation of appearance to reality, however, remaining in the domain of the playwright.

For the present purpose it is sufficient to characterize the MAC system as one which permits an individual to operate a full scale computer from a remotely located typewriter. The individual operator has the illusion that he is the sole user of the computer complex, while in fact others may be “time-sharing” the system with him. What is important here is that the computer can read messages typed on the typewriter and respond by writing on the same instrument. The time between the computer’s receipt of a message and the appearance of its response is a function of the program controlling the dialogue and of such MAC system parameters as the number of user

ELIZA is a program operating within the MAC time-sharing system at MIT which makes certain kinds of natural language conversation between man and computer possible. Input sentences are analyzed on the basis of decomposition rules which are triggered by key words appearing in the input text. Responses are generated by reassembly rules associated with selected decomposition rules. The fundamental technical problems with which ELIZA is concerned are: (1) the identification of key words, (2) the discovery of minimal context, (3) the choice of appropriate transformations, (4) generation of responses in the absence of key words, and (5) the provision of an editing capability for ELIZA “scripts”. A discussion of some psychological issues relevant to the ELIZA approach as well as of future developments concludes the paper.



Menu

Introducing ChatGPT

We’ve trained a model called ChatGPT which interacts in a conversational way. The dialogue format makes it possible for ChatGPT to answer followup questions, admit its mistakes, challenge incorrect premises, and reject inappropriate requests.

[Try ChatGPT ↗](#)

[Read about ChatGPT Plus](#)

Illustration: Ruby Chen

November 30, 2022

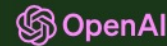
Authors
[OpenAI ↓](#)

[Product, Announcements](#)

Chatbots

Computational Linguistics

We've trained a model called ChatGPT



Menu

Introducing

Model called
interacts in a
The dialogue
possible for
ChatGPT to answer followup
questions, admit its mistakes,
challenge incorrect premises, and
reject inappropriate requests.

[Try ChatGPT ↗](#)

[Read about ChatGPT Plus](#)

Illustration: Ruby Chen

November 30, 2022

Authors
[OpenAI](#) ↓

[Product, Announcements](#)

1966

ELIZA—A Computer Program

The object of this paper is to cause just such a re-

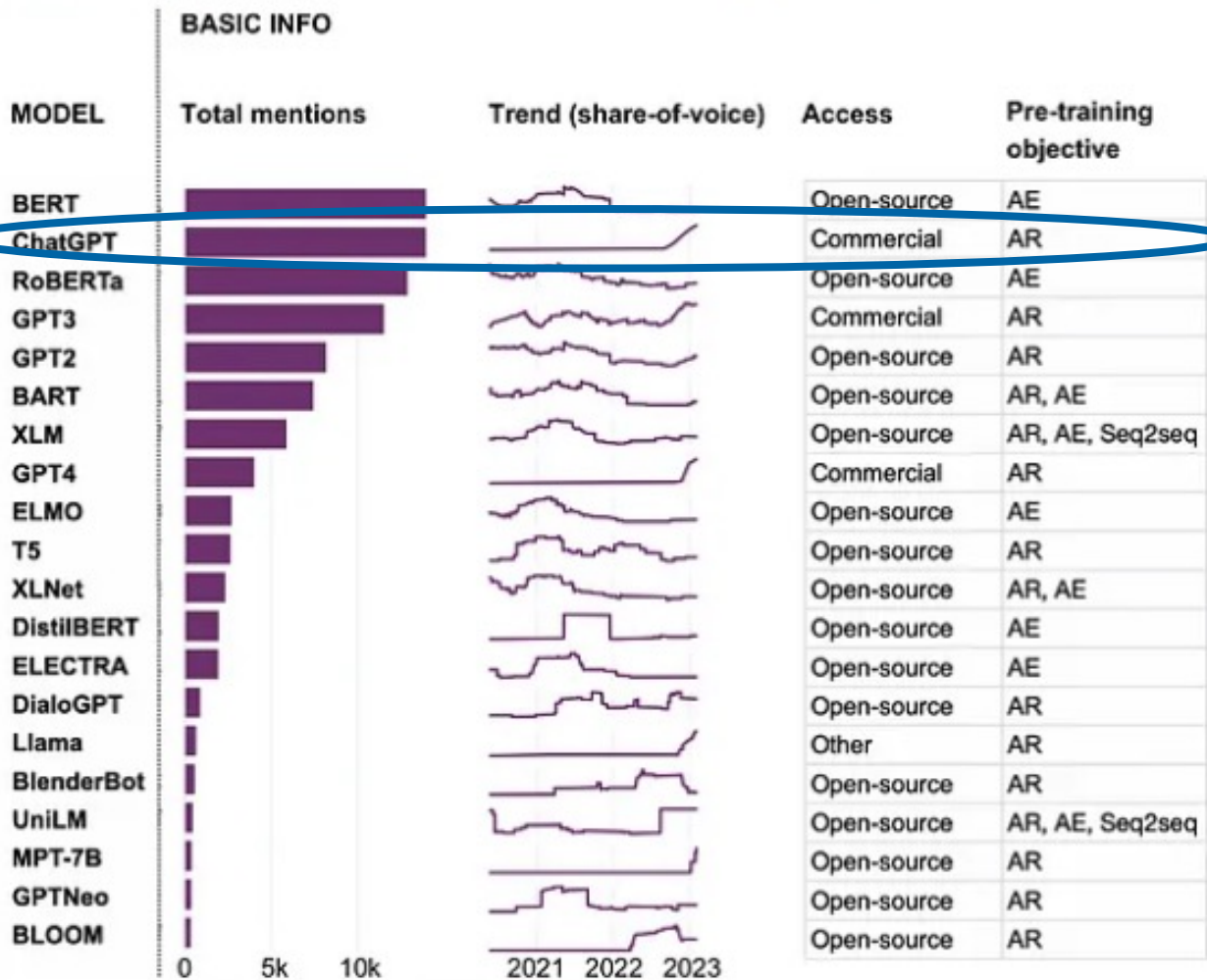
conversation between man and computer possible. Input sen-
C
A tences are analyzed on the basis of **decomposition rules** which
are **triggered by key words** appearing in the input text.

ELIZA is a program operating within the MAC time-sharing system at MIT which makes certain kinds of natural language conversation between man and computer possible. Input sentences are analyzed on the basis of decomposition rules which are triggered by key words appearing in the input text. Responses are generated by reassembly rules associated with selected decomposition rules. The fundamental technical problems with which ELIZA is concerned are: (1) the identification of key words, (2) the discovery of minimal context, (3) the choice of appropriate transformations, (4) generation of responses in the absence of key words, and (5) the provision of an editing capability for ELIZA "scripts". A discussion of some psychological issues relevant to the ELIZA approach as well as of future developments concludes the paper.

continually improved by a "teacher". Like the Eliza of Pygmalion fame, it can be made to appear even more civilized, the relation of appearance to reality, however, remaining in the domain of the playwright.

For the present purpose it is sufficient to characterize the MAC system as one which permits an individual to operate a full scale computer from a remotely located typewriter. The individual operator has the illusion that he is the sole user of the computer complex, while in fact others may be "time-sharing" the system with him. What is important here is that the computer can read messages typed on the typewriter and respond by writing on the same instrument. The time between the computer's receipt of a message and the appearance of its response is a function of the program controlling the dialogue and of such MAC system parameters as the number of user

OVERVIEW OVER THE MOST POPULAR LLMS (May 2023)



Abbreviations:
 AE = Autoencoding
 AR = Autoregression
 Seq2seq = Sequence-to-sequence

SUITABILITY FOR DOWNSTREAM TASKS*



*Scoring on a scale between 0 and 1;
 1 = highly suitable, 0 = not suitable.

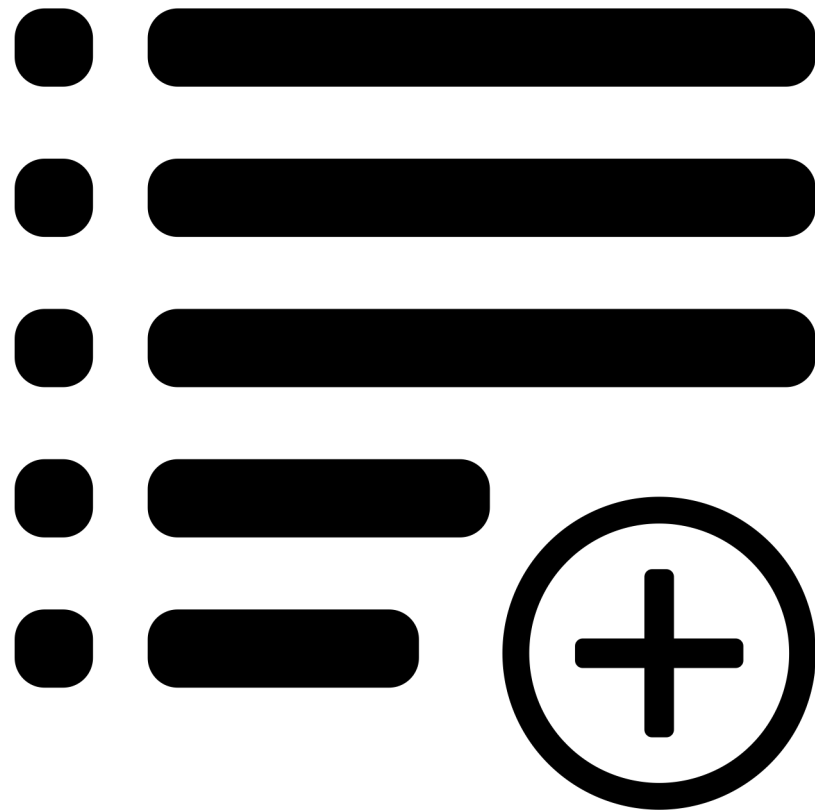
Automatische Textproduktion

Technische Grundlagen

Einsatzmöglichkeiten

Konsequenzen

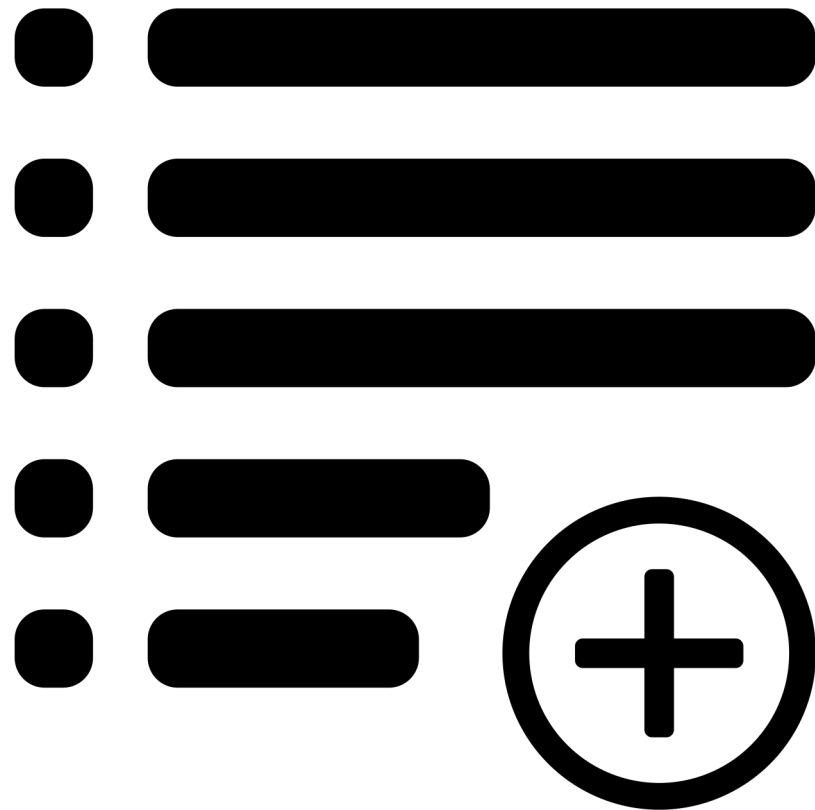
Zu vorhandenem Text passenden/plausiblen Text erzeugen



Es

handelt ...
war ...
begab ...
ist ...
könnte ...

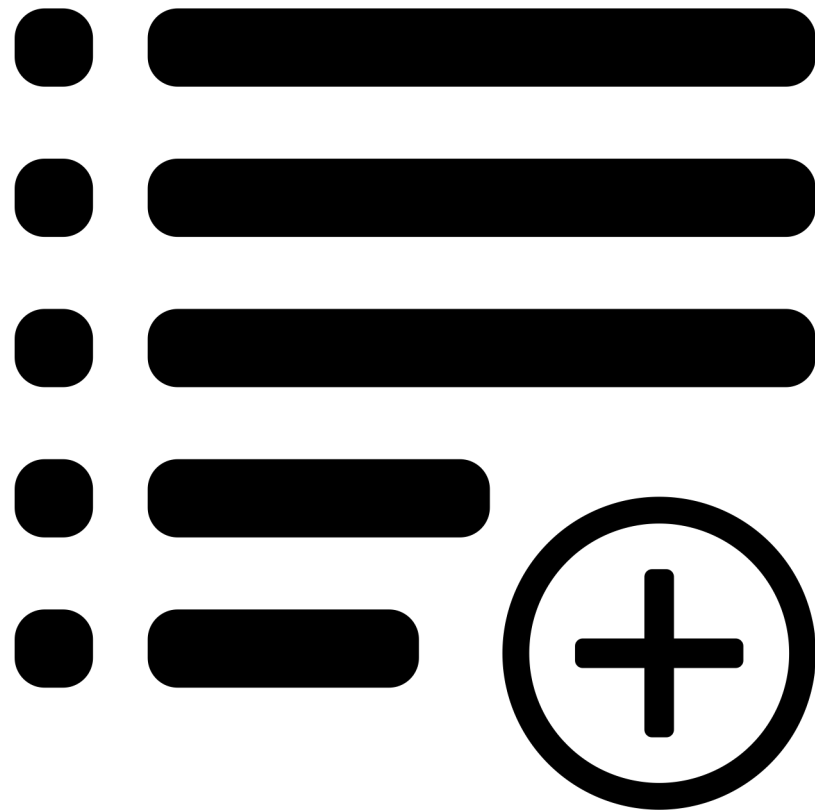
Zu vorhandenem Text passenden/plausiblen Text erzeugen



Es war

aber ...
doch ...
einmal ...
offenbar ...
kein ...

Zu vorhandenem Text passenden/plausiblen Text erzeugen

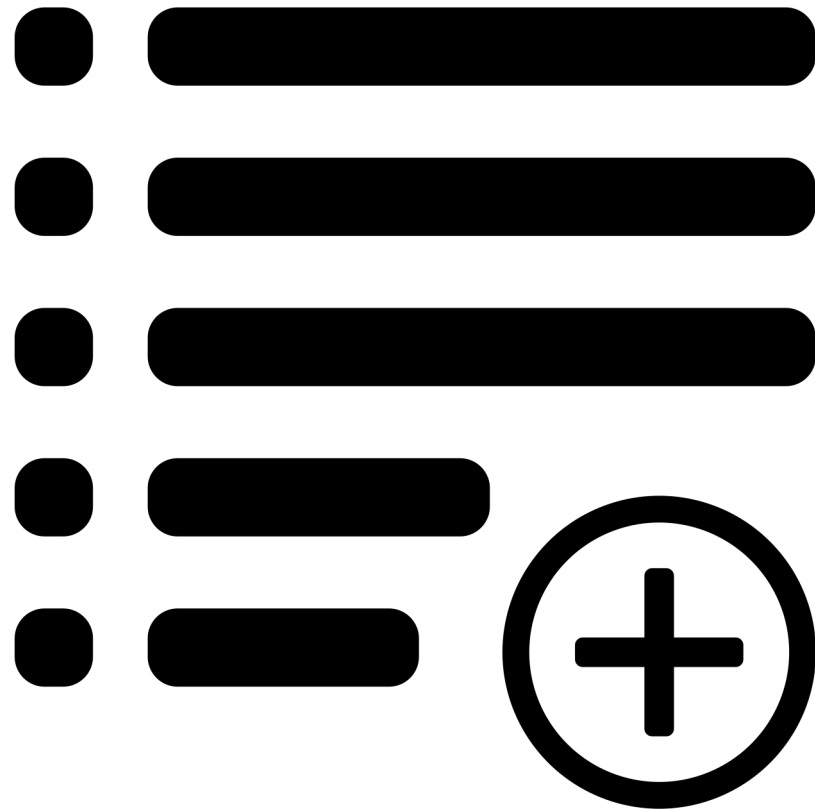


Es war einmal

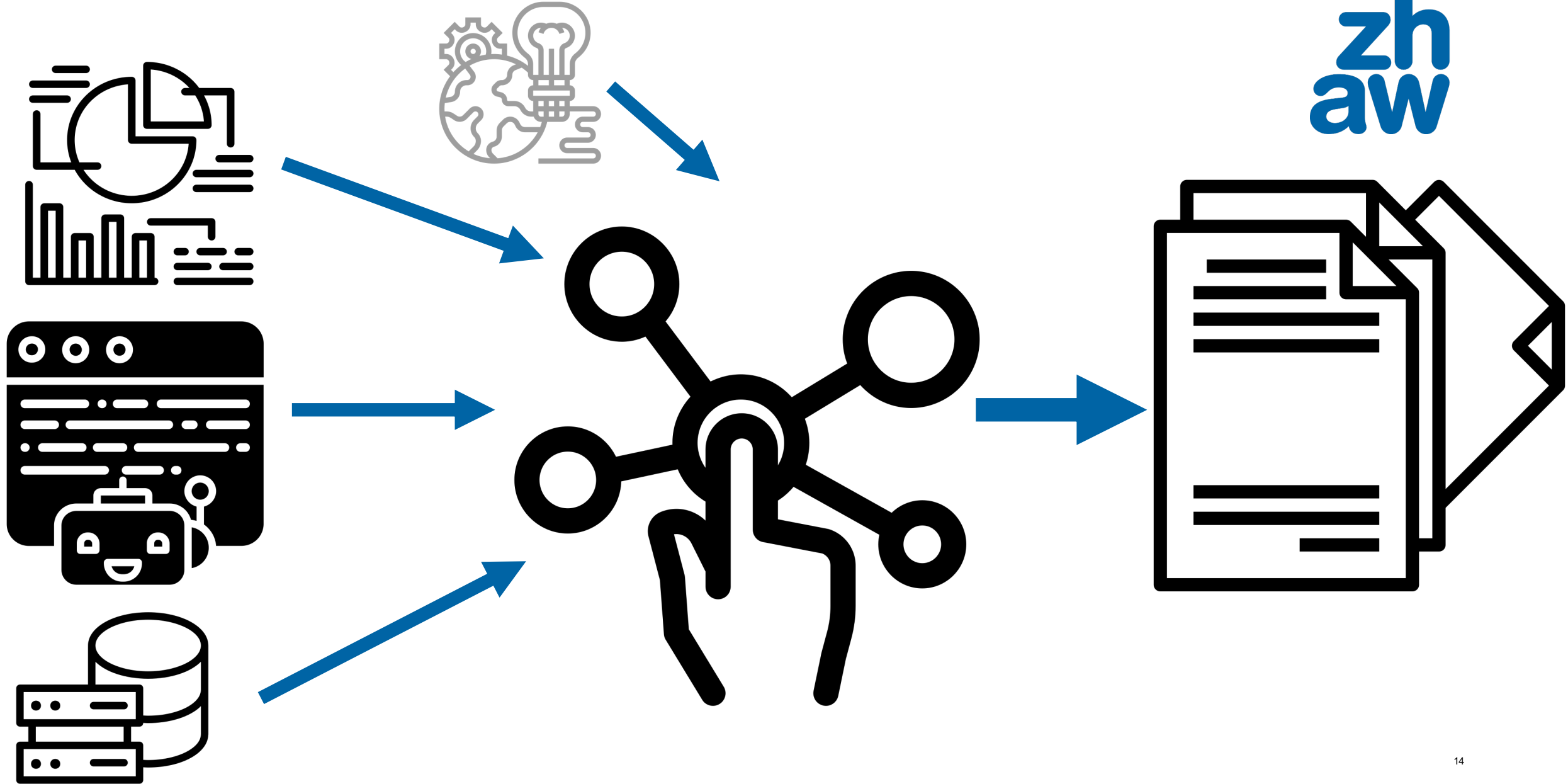
ein ...
eine ...

Zu vorhandenem Text passenden/plausiblen Text erzeugen

Musterhafte Text(teile), Textvarianten, ...



Automatische Textproduktion in der Praxis



Automatische Textproduktion



```
graph TD; A[Automatische Textproduktion] --- B[Technische Grundlagen]; A --- C[Einsatzmöglichkeiten]; A --- D[Konsequenzen];
```

A hierarchical diagram with a central box at the top containing the text 'Automatische Textproduktion'. A vertical line descends from the bottom center of this box to a horizontal line. From the left end of this horizontal line, a vertical line goes down to a box labeled 'Technische Grundlagen'. From the center of the horizontal line, a vertical line goes down to a box labeled 'Einsatzmöglichkeiten'. From the right end of the horizontal line, a vertical line goes down to a box labeled 'Konsequenzen'. The 'Konsequenzen' box is filled with a solid blue color, while the other boxes are white with a blue border.

Technische Grundlagen

Einsatzmöglichkeiten

Konsequenzen



Mensch-Maschine-Kommunikation in natürlicher Sprache

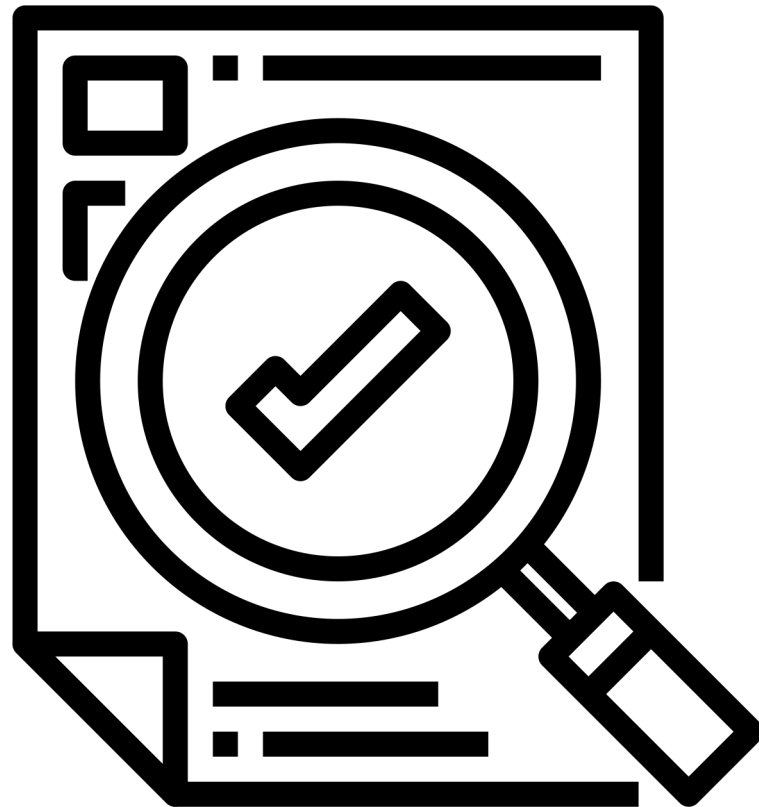
Prompt-Engineering



Theoriegeleitet, wissenschaftlich fundiert



Qualität? Autor:innenschaft?



Automatische Textproduktion

A hierarchical diagram with a central blue box at the top containing the title 'Automatische Textproduktion'. A vertical line descends from the bottom center of this box to a horizontal line. From the left and right ends of this horizontal line, two vertical lines extend downwards to the top centers of two separate blue boxes. From the top center of each of these two boxes, a vertical line extends downwards to the top center of a third blue box, which is positioned centrally below the two side boxes. This structure creates three main branches from the central title box.

Technische Grundlagen

Künstliche Neuronale
Netze

Einsatzmöglichkeiten

Fabrikation von plausiblen
Texten

Konsequenzen

KI als Werkzeug zum
Texten