

Dialog mit Daten: Von Sprachmodellen zu Wissensgraphen

Andreas Harth

mit Rene Dorsch, Zhou Gui, Daniel Henselmann, Sebastian Schmid

Fraunhofer-Institut für Integrierte Schaltungen IIS

Friedrich-Alexander-Universität Erlangen-Nürnberg

Dezember 2024

Sprachmodelle (ChatGPT, Claude, Mistral)




- können zu allem etwas sagen
- machen es einfach, beeindruckende Demos zu bauen


A gib mir eine liste der leiter der organisationen im geschaeftsbereich des bmel

Basierend auf meinen derzeit verfügbaren Informationen:

1. Bundesanstalt für Landwirtschaft und Ernährung (BLE)
 - Leiterin: Dr. Gerda Hagenmeyer
2. Bundesinstitut für Risikobewertung (BfR)
 - Präsident: Prof. Dr. Andreas Hensel
3. Julius Kühn-Institut (JKI)
 - Präsident: Prof. Dr. Frank Gottwald
4. Max Rubner-Institut (MRI)
 - Präsident: Prof. Dr. Cornelia Weigmann
5. Thünen-Institut
 - Präsident: Prof. Dr. Folkhard Isermeyer

Bitte beachten Sie, dass Führungspositionen sich ändern können. Für die aktuellsten Informationen empfehle ich, die offiziellen Websites der jeweiligen Organisationen zu konsultieren.

 Copy  Retry  

 Claude can make mistakes. Please double-check responses.

Stand:
12.12.2024

Language Models in a Nutshell

- Training
 - training data: the web
 - output: the model
 - phases: 1. next word prediction 2. task-specific training 3. human feedback
 - takes weeks/months (Llama 3.3: release Dec 6, 2024, knowledge cutoff Dec 2023)
- Inference
 - input: prompt
 - output: completion
 - model generates the next likely completion tokens based on the prompt
 - limited context window (e.g., 128k tokens)
- Veracity
 - interrogate the language model about the response

Aber ChatGPT kann auf das Web zugreifen!

gib mir eine liste der leiter der organisationen im geschaeftsbereich des bmel



Searched 5 sites ▾

Hier sind die Leiter der Organisationen im Geschäftsbereich des Bundesministeriums für Ernährung und Landwirtschaft (BMEL), basierend auf aktuellen Informationen:

1. **Julius Kühn-Institut (JKI)** – Präsident: Prof. Dr. Frank Ordon.
2. **Friedrich-Loeffler-Institut (FLI)** – Präsident: Prof. Dr. Dr. Dr. h.c. Thomas C. Mettenleiter.
3. **Max Rubner-Institut (MRI)** – Präsidentin: Prof. Dr. Tanja Schwerdtle.
4. **Johann Heinrich von Thünen-Institut** – Präsidentin: Dr. Birgit Kleinschmit.

Weitere Einrichtungen wie das Bundesinstitut für Risikobewertung (BfR) und andere Forschungsinstitute haben ebenfalls spezifische Leitungen, wobei beispielsweise das BfR aktuell von Prof. Dr. Dr. Andreas Hensel geführt wird. Informationen zu den Leitern der anderen Institutionen können je nach Aktualität variieren [BMEL](#) [DEUTSCHER BUNDESTAG](#) [BMEL](#) .

Sollten Sie spezifischere Details zu einzelnen Organisationen wünschen, lasse ich Ihnen gerne weitere Informationen zukommen!



Stand:
12.12.2024

Retrieval-augmented Generation (RAG)

- At inference time the language model can do information retrieval on documents.
- Inference with RAG
 - input: prompt
 - output: completion
 - find relevant documents for prompt
 - augment prompt with relevant portions of text from documents
 - model generates the next likely completions based on the augmented prompt
- Veracity
 - inspect the documents to check the responses of the language model

Aber wir haben Daten und nicht nur Dokumente?

```
SELECT DISTINCT ?org ?orgLabel ?director ?directorLabel
WHERE {
  wd:Q699656 wdt:P355 ?org .
  ?org wdt:P488 ?director .
  SERVICE wikibase:label {
    bd:serviceParam wikibase:language
      "[AUTO_LANGUAGE],mul,en" .
  }
}
```

[Wikidata query results](#)



Language Models on Knowledge Graphs

- At inference time the language model can access information in a knowledge graph.
- Inference with Knowledge Graphs
 - input: prompt
 - output: completion
 - find relevant portions of the graph for prompt
 - augment prompt with relevant subgraph
 - model is instructed to generate a SPARQL query as formal representation of the user question
- Veracity
 - inspect the information in the subgraphs
 - inspect the query
 - answers to queries on graphs have perfect precision and perfect recall

Sprachmodelle mit Wissensgraphen

chairpersons of the organisations that are subsidiaries of the bmel



Used agent



Sys: Answers to chairpersons of the organisations that are subsidiaries of th... are as follows.

- [Folkhard Isermeyer](#)
- [Andreas Hensel](#)
- [Frank Ordon](#)

▼ SPARQL query

Raw code



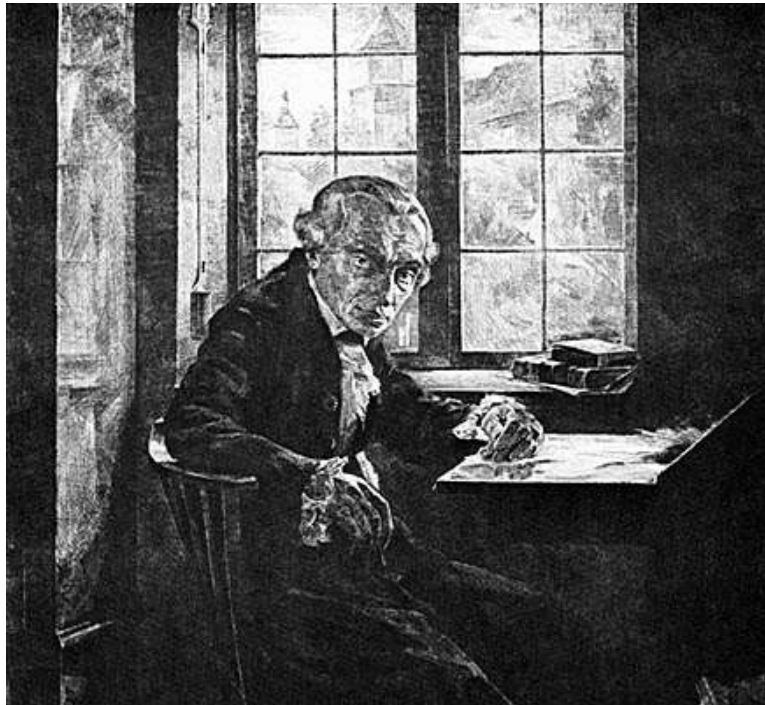
```
PREFIX wdt: <http://www.wikidata.org/prop/direct/> SELECT DISTINCT ?chairperson WHERE
{ <http://www.wikidata.org/entity/Q699656> wdt:P355 ?subsidiary . ?subsidiary wdt:P488
?chairperson . }
```

Sys: Loop ended after 4 cycles in 21 sec with 11376 prompt tokens and 154 completion tokens.



Ansätze im Überblick

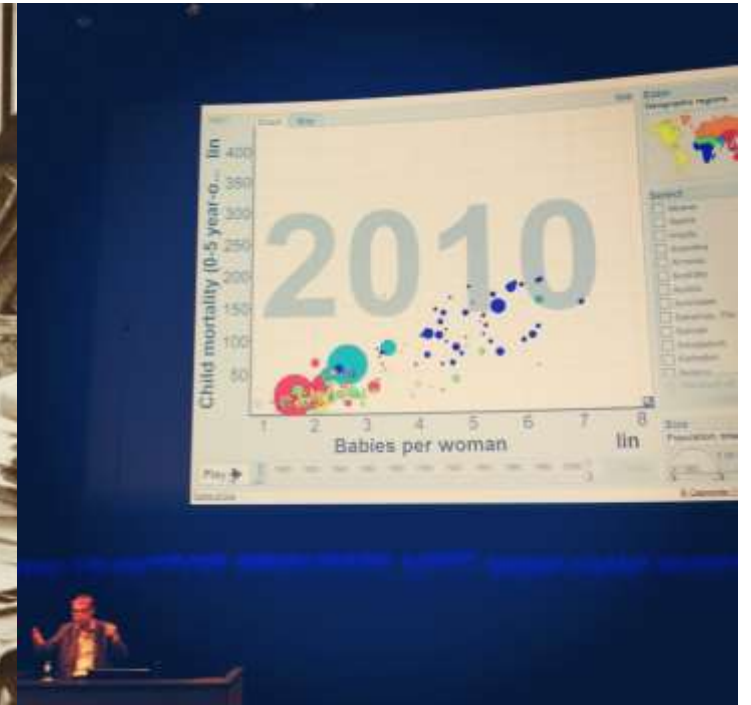
Sprachmodelle: Dialog mit neuronalen Gewichten



Sprachmodelle mit RAG: Dialog mit Dokumenten



Sprachmodelle mit Wissensgraphen: Dialog mit Daten



Welektronik: Dialog mit integrierten Daten

- [Welektronik](#) Wiki zum Thema vertrauenswürdige Elektronik als Wissensgraph
- Links zu Wikidata
- Integration der Daten via Linked Data
- Visualisierung
- Zugriff via Sprachmodell

Map of enterprises in the semiconductor industry

Created by the [Center for Applied Research on Supply Chain Services at Fraunhofer IIS](#) for the research project [Welektronik](#) (grant no. 16ME0224). Data from the [Welektronik](#) knowledge graph. Contact: [Daniel Henselmann](#)



Demo

suppliers of bosch



Used agent



Sys: Answers to `suppliers of bosch` are as follows.

- [Osram Opto Semiconductors](#)
- [Kyocera](#)
- [Integrated Silicon Solution](#)
- [Infineon Technologies](#)
- [Murata Manufacturing](#)
- [Fuji Electric](#)

► SPARQL query

Sys: Loop ended after 4 cycles in 8 sec with 13035 prompt tokens and 207 completion tokens.

Demo

sites of suppliers of suppliers of bosch



Used agent



Sys: Answers to sites of suppliers of suppliers of bosch are as follows.

- [SkyWater Technology Minnesota Fab](#)
- [SkyWater Technology Florida Fab](#)
- [SkyWater Technology Indiana Fab](#)
- [Integral Minsk Fab](#)
- [Taiwan Semiconductor Manufacturing Company Nanjing Fab](#)
- [Taiwan Semiconductor Manufacturing Company Hsinchu Fab](#)
- [Taiwan Semiconductor Manufacturing Company Tainan Fab](#)



Demo

compare key figures of intel and nvidia



Used agent



Sys: Answers to `compare key figures of intel and nvidia` are as follows.

company	year	totalrevenue	netprofit	operatingincome
Nvidia	2007		797645000.0	
Nvidia	2007			836346000.0
Nvidia	2007	4097860000.0		



Fazit

- Auswahl eines Ansatzes nach Anwendungszweck und möglichem Aufwand
 - Dialog mit neuronalen Gewichten
 - Dialog mit Dokumenten
 - Dialog mit Daten
 - Dialog mit integrierten Daten
- Die Daten sind der Schatz, nicht das Anwendungssystem

Image Credits

- Images on Slide 9 via Wikimedia Commons
- Image on Slide 9 Hans Rosling
 - Source: [https://commons.wikimedia.org/wiki/File:Hans_Rosling_2-fd0000_\(9504183446\).jpg](https://commons.wikimedia.org/wiki/File:Hans_Rosling_2-fd0000_(9504183446).jpg)
 - Author: Open Knowledge Foundation Deutschland
 - License: Creative Commons Attribution 2.0 Generic
<https://creativecommons.org/licenses/by/2.0/deed.en>