

0) Was bedeutet generative KI für Hochschule und Gesellschaft?

Language Model

* Definition: Models that assign probabilities to sequences of words.

Generative Large Language Model?

* Künstliches neuronales Netzwerk, das Text passend zu einem gegebenen Kontext vorhersagt

* Erzeuge ein Wort (Token), das im Kontext der vorhergehenden Wörter (Prompt)

wahrscheinlich ist

* Wiederhole Prozess, um einen ganzen Text zu erzeugen

* Können über komplexe Muster abstrahieren

Instructions

* Neuere Modelle sind optimiert, Antworten zu Anweisungen zu generieren.

* 3 Lernphasen:

However: text on internet does not contain many conversations or questions with answers
→ it needs to learn how to react to prompts

1. Vorhersage des nächsten Wortes für immense Textsammlungen (> 3 M-d. Token)

2. Lernen aus Beispielanwendungen mit gegebenen Antworten (> 100k Beispiele)

3. Lernen aus zusätzlichem Feedback zu gewünschten Eigenschaften guter Antworten (helpful, harmless, honest)

1) Foundations of Neural Language Models

Chinese Room:

* Programming a digital computer may make it appear to understand

language but could not produce "real" understanding

2) Unhumanizing Models. Why we need to change how we think about AI

* 6 human-like capacities in AI models

- Knowing (Overfitting, context)

- Understanding (Statistical language mimicry)

- Feeling (Reflection of our own feelings)

- Creating (Needs human prompt and editing)
- Moving (moves need motives)
- Judging (Consistency, trust and complex morals)

3)

- * inconsistency in AI models
- * Octopus paper: learns only from language, doesn't know what is real / doesn't know any meaning
- * Compare Chinese room with Octopus paper

4)

- * Different architectures of LLMs that use different approaches for next word prediction
- * Transformers (actual) - quadratic in time
- * Recurrent architecture (LSTM) - linear in time \Rightarrow much more efficient

5)

- * Different sources of label variation
- * Same example annotated by different people
- * Reasons:
 - Annotator disagreement
 - Subjectivity (perspectives)
 - Multiple plausible answers
 - Annotation error

6)

- * Combination of LLMs and Structured outputs
- * Infinite ways to combine things
- * Dialogue system: Extended prompts happening in the background.

7)

1. Digitization
2. Restoration
3. Attribution
4. Linguistic Analysis
5. Textual Criticism
6. Translation

8)

* Usefulness of a system increases if more and more data exists

* Comment on plot:

- there is a point where the system only improves marginally

- At this point we should stop collecting data in order to protect privacy/user data

* Websites are collecting data for "legitimate interest". But whose legitimate interest do they mean?

- user?

- the company offering the service?

- 3rd party vendors?

- Society?