UNIVERSITÄT
ZU KÖLN

Maschinelle Lernverfahren und “Künstliche Intelligenz”

Nils Reiter,
nils.reiter@uni-koeln.de

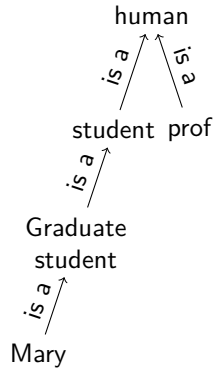
October 24, 2023

INSTITUT FÜR
DIGITAL HUMANITIES
UNIVERSITÄT ZU KÖLN

Was wissen Sie über (die technischen Hintergründe von) ChatGPT?

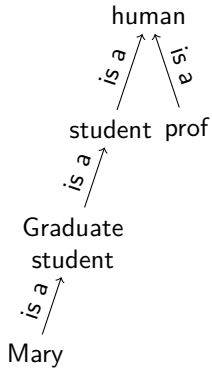
Artificial Intelligence

- Classical AI research
  - Algorithms to plan actions in order to achieve a goal
  - Represent knowledge formally (e.g., in machine-readable ontologies)
  - Draw logical conclusions from knowledge



Artificial Intelligence

- Classical AI research
  - Algorithms to plan actions in order to achieve a goal
  - Represent knowledge formally (e.g., in machine-readable ontologies)
  - Draw logical conclusions from knowledge
- Generative AI
  - No explicit knowledge representation – "neural knowledge" extracted from data sets
  - Machine learning models to make inferences



Neural Knowledge

Vector representation for “köln”

0.0539 -0.0030 0.0203 -0.1084 -0.0099 0.0705 -0.0546 -0.0433 -0.0096 0.0561 -0.0095 0.0280 0.1726 0.0190 0.0369 0.0217 -0.0002 -0.0309 0.0347 -0.0749
-0.0202 0.0151 -0.0195 0.0001 0.0232 0.0243 -0.0170 -0.0090 -0.0108 -0.0943 0.0376 0.1118 -0.0324 0.0148 -0.0033 0.0537 -0.0681 -0.0733 -0.0201 -0.0329
0.1242 0.0324 -0.0744 -0.0149 -0.0047 -0.0484 -0.0483 0.0481 0.0107 0.0101 -0.0704 0.0500 0.0112 -0.0227 0.0499 -0.0259 -0.0441 0.0712 -0.0157 -0.1271
0.0407 -0.0495 -0.0359 0.0202 0.0024 0.0764 0.0196 0.0267 -0.0117 0.0026 0.0171 -0.0121 -0.1374 -0.0370 0.0247 -0.0113 -0.0094 0.0322 -0.0347 -0.0866 0.0042
-0.0014 0.0067 0.0591 0.0009 0.0085 0.0310 0.0479 -0.0511 0.0198 -0.0886 -0.0274 -0.1364 0.0322 -0.1638 -0.0689 0.0016 -0.1039 0.0059 0.0757 -0.0034 0.1013
-0.0034 -0.0065 -0.0468 0.1577 -0.0065 -0.0478 -0.0004 0.0682 0.0045 -0.0607 -0.0590 0.0343 0.0036 -0.1014 -0.0136 -0.0063 0.0801 0.0360 0.0579 -0.0039
0.0975 0.0500 -0.0558 -0.0095 0.0057 -0.0246 0.1070 -0.0186 0.0669 -0.0781 -0.0569 -0.1286 -0.0834 0.0106 -0.0672 -0.0205 0.0613 0.0290 -0.0545 -0.0481
-0.0882 -0.0489 0.0622 -0.0730 -0.0192 -0.0415 -0.0287 0.0218 -0.0427 -0.0046 0.0255 -0.1164 0.0077 -0.0546 -0.0786 0.0000 -0.0456 0.0943 0.0157 -0.0117
-0.0441 -0.0015 -0.0556 -0.0508 0.0088 0.0418 0.0030 -0.1450 -0.0663 0.0800 0.0172 -0.0289 0.1178 -0.0973 0.0888 0.0637 -0.0295 0.0212 0.0100 -0.0860 0.0035
0.0730 0.0425 -0.0080 0.0885 -0.0166 -0.0765 0.0004 -0.0118 0.0138 -0.0093 -0.0606 -0.0447 -0.0746 0.0131 -0.0447 -0.0763 0.0032 0.1181 0.0542 0.0431
-0.0273 0.0547 0.0135 0.0006 -0.0241 -0.0418 0.0278 -0.0821 -0.0572 -0.0039 0.0214 -0.0196 0.0449 -0.0286 0.0204 0.0681 -0.0901 -0.0266 -0.0287 -0.0874
0.0797 -0.0784 -0.0920 0.0380 0.0411 0.0859 0.0369 0.0595 0.0446 0.0363 -0.0353 -0.0044 -0.0061 0.1134 0.1420 -0.0026 -0.0013 0.0033 0.0508 0.0096 -0.0757
0.0085 -0.0099 -0.0384 0.0218 -0.0259 -0.0112 -0.0212 0.0273 0.0532 -0.0278 -0.0634 0.0317 -0.0022 0.0882 -0.0240 0.0031 -0.0370 0.0747 -0.0097 -0.0315
0.0405 0.0124 -0.1416 -0.0768 0.0363 -0.1248 -0.0134 0.0702 -0.0905 -0.0387 0.0683 -0.0784 0.0886 0.0640 0.0611 -0.0199 -0.0447 -0.1331 -0.1247 0.0540
0.0499 -0.0212 -0.0544 -0.1161 -0.0729 0.0894 0.0532 0.0164 -0.0039 -0.0108 -0.0248 -0.1021 -0.0549 -0.0318 0.0309 -0.0691

Machine Learning

- What is machine learning?
  - Method to find patterns, hidden structures and undetected relations in data
- It's all around us: Stock market transactions, search engines, surveillance, data-driven research & science, text and image generation ...

Machine Learning

- What is machine learning?
  - Method to find patterns, hidden structures and undetected relations in data
- It's all around us: Stock market transactions, search engines, surveillance, data-driven research & science, text and image generation ...

Machine Learning

Classification

- Explicitly assigning *classes* to *objects/instances/items*
  - Words → parts of speech
  - Texts → genres
- Many algorithms available: Decision trees, support vector machines, naïve Bayes, neural networks, Bayesian networks, ...

Machine Learning

Classification

- Explicitly assigning *classes* to *objects/instances/items*
  - Words → parts of speech
  - Texts → genres
- Many algorithms available: Decision trees, support vector machines, naïve Bayes, neural networks, Bayesian networks, ...
- Libraries are available, not a technical challenge
- Challenges
  - Find useful training data
  - Interpret results reasonably
  - Establish a realistic and *trustworthy* test set

Machine Learning

Language Modeling

- One of the oldest NLP tasks
  - Long before predictive typing on smart phones became a thing
  - Long before "large language models" became a thing
- Language model (LM) predicts the next word, given previous words (history)
- Formally: $p(\text{word}|\text{history})$

Beispiel

Maria hat an der Universität zu Köln _____.

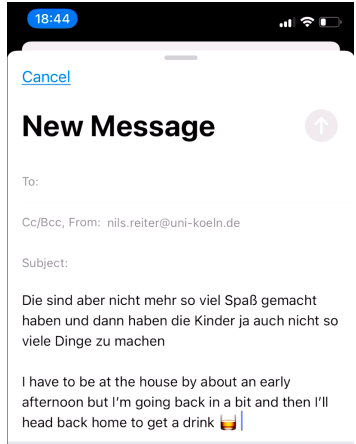
Machine Learning

Language Modeling

- One of the oldest NLP tasks
  - Long before predictive typing on smart phones became a thing
  - Long before “large language models” became a thing
- Language model (LM) predicts the next word, given previous words (history)
- Formally: $p(\text{word}|\text{history})$

Beispiel

Maria hat an der Universität zu Köln _____.



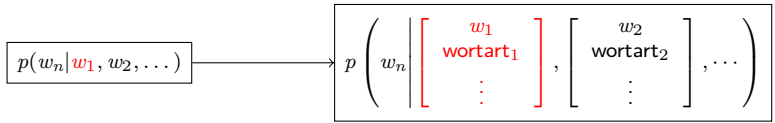
Machine Learning

From Language Models to Large Language Models

$$p(w_n|w_1, w_2, \dots)$$

Machine Learning

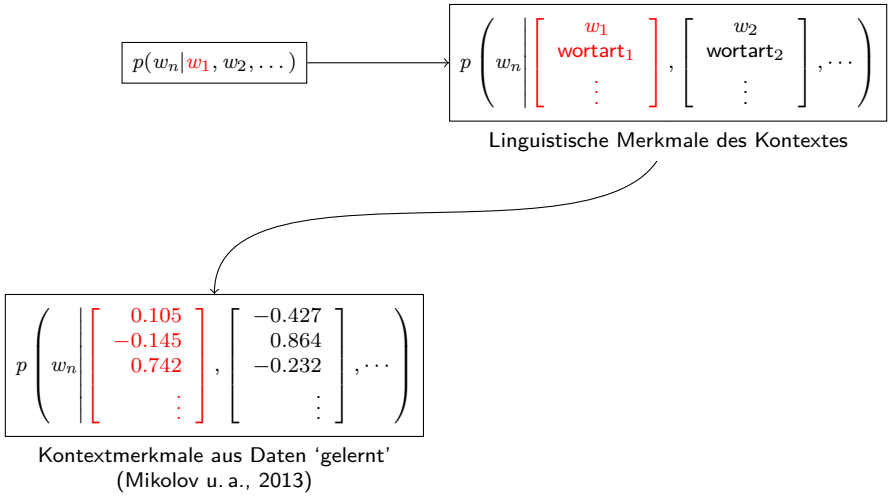
From Language Models to Large Language Models



Linguistische Merkmale des Kontextes

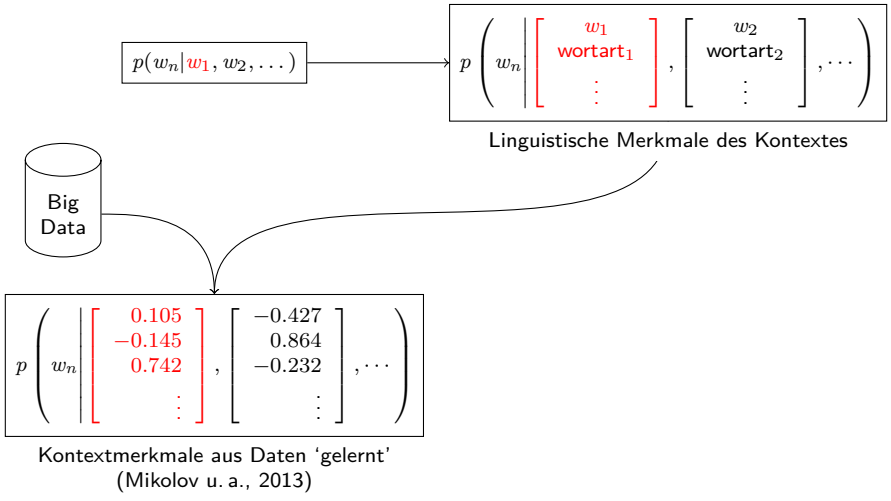
Machine Learning

From Language Models to Large Language Models



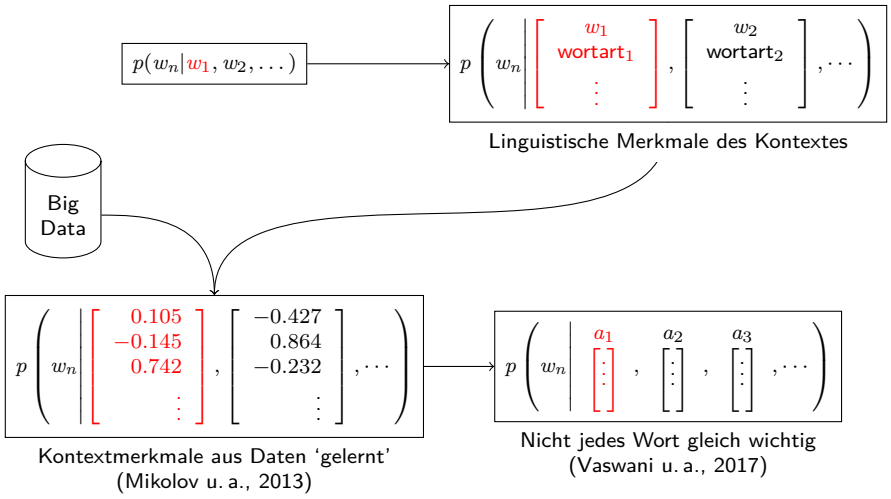
Machine Learning

From Language Models to Large Language Models



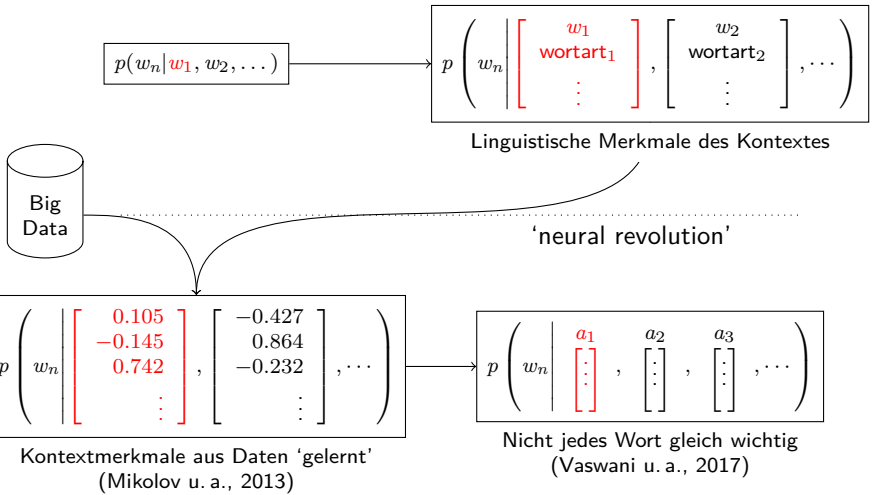
Machine Learning

From Language Models to Large Language Models



Machine Learning

From Language Models to Large Language Models



Multiple Ways of Using a Large Language Model

1. Let it generate free text ("prompting")
   - Impressive for lay persons and investors
   - Difficult to evaluate exactly: How to measure if a text is "correct" or "good"?
   - Terms are not defined explicitly, and (presumably) used with their every-day meaning (which is vague)
     - Challenge for sciences with specialised vocabulary

Multiple Ways of Using a Large Language Model

1. Let it generate free text ("prompting")
   - Impressive for lay persons and investors
   - Difficult to evaluate exactly: How to measure if a text is "correct" or "good"?
   - Terms are not defined explicitly, and (presumably) used with their every-day meaning (which is vague)
     - Challenge for sciences with specialised vocabulary
2. Fine-tune it to a specific task and let it do classification
   - Requires explicitly labeled training data
   - More labor-intensive
   - Model outputs probabilities for classes, straightforward evaluation
     - E.g., what's the percentage of correct predictions

The Role of Data

- Two purposes: Training and testing
- Training set sizes are increasing: (Bigger $\cup$ Better) $>$ Bigger $>$ Better
- Raw data is rarely used for training
  - Which preparation steps to include is a decision developers make

The Role of Data

- Two purposes: Training and testing
- Training set sizes are increasing: (Bigger $\cup$ Better) $>$ Bigger $>$ Better
- Raw data is rarely used for training
  - Which preparation steps to include is a decision developers make

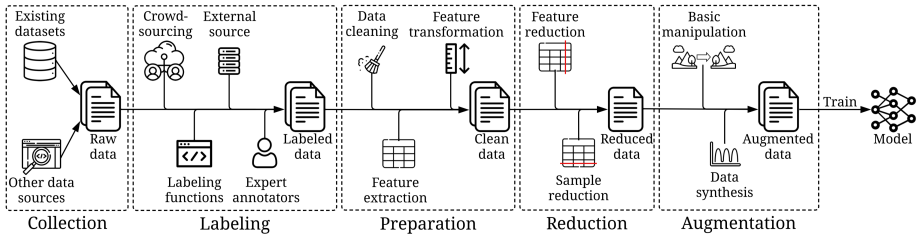


Abbildung: Training data development (Zha u. a., 2023, 8)

The Role of Data

Testing/Evaluation

- Finding out how well things work – important for scientific research
- (L)LMs hard to evaluate, because not a single correct answer
- Classification: Straightforward, because we know which classes can be given

The Role of Data I

Pitfalls and Mistakes (Roberts u. a., 2021)

- Can ML predict COVID-19, based on chest CXR and CT scans?
- 2212 studies, 415 looked at in detail
- "[...] none of the models identified are of potential clinical use due to methodological flaws and/or underlying biases"

The Role of Data II

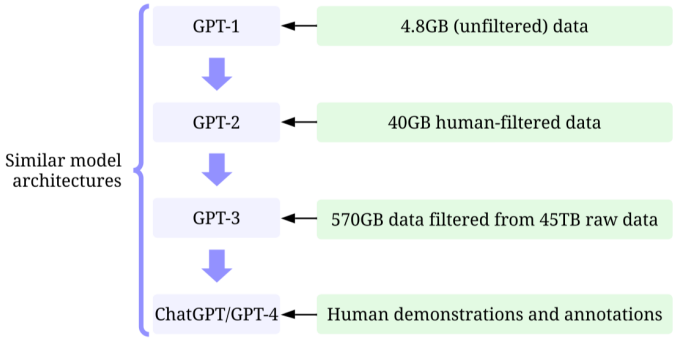
Pitfalls and Mistakes (Roberts u. a., 2021)

- Participants bias
  - Public data sets with unclear ground truth, as anyone can contribute images
  - Unclear selection criteria applied
  - Demographic differences between COVID-19 and control groups
- Predictor bias: Which features does the model look at?
  - Impossible to evaluate in deep learning scenarios
- Outcome bias sources
  - Inconsistent diagnosis of COVID-19
  - unclear definition of a control group
  - ground truths being assigned using the images themselves
  - using an unestablished reference to define outcome
  - combining public and private datasets

The Role of Data

Training data in popular models

- BERT: `https://huggingface.co/bert-base-uncased`
- GPT-2: `https://huggingface.co/gpt2`
- GPT-4: `https://arxiv.org/pdf/2303.08774.pdf`
- Image models: `https://haveibeentrained.com`



References I

Mikolov, Tomáš/Ilya Sutskever/Kai Chen/Greg S Corrado/Jeff Dean (2013). "Distributed Representations of Words and Phrases and their Compositionality". In: *Advances in Neural Information Processing Systems 26*. Hrsg. von C. J. C. Burges/L. Bottou/M. Welling/Z. Ghahramani/K. Q. Weinberger. Curran Associates, Inc., S. 3111–3119. URL: http://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality.pdf.

References II

Roberts, Michael/Derek Driggs/Matthew Thorpe/Julian Gilbey/Michael Yeung/Stephan Ursprung/Angelica I. Aviles-Rivero/Christian Etmann/Cathal McCague/Lucian Beer/Jonathan R. Weir-McCall/Zhongzhao Teng/Effrossyni Gkrania-Klotsas/Alessandro Ruggiero/Anna Korhonen/Emily Jefferson/Emmanuel Ako/Georg Langs/Ghassem Gozaliasl/Guang Yang/Helmut Prosch/Jacobus Preller/Jan Stanczuk/Jing Tang/Johannes Hofmanninger/Judith Babar/Lorena Escudero Sánchez/Muhunthan Thillai/Paula Martin Gonzalez/Philip Teare/Xiaoxiang Zhu/Mishal Patel/Conor Cafolla/Hojjat Azadbakht/Joseph Jacob/Josh Lowe/Kang Zhang/Kyle Bradley/Marcel Wassin/Markus Holzer/Kangyu Ji/Maria Delgado Ortet/Tao Ai/Nicholas Walton/Pietro Lio/Samuel Stranks/Tolou Shadbahr/Weizhe Lin/Yunfei Zha/Zhangming Niu/James H. F. Rudd/Evis Sala/Carola-Bibiane Schönlieb/AIX-COVNET (2021). "Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans". In: *Nature Machine Intelligence* 3.3, S. 199–217. DOI: 10.1038/s42256-021-00307-0.

References III

- Vaswani, Ashish/Noam Shazeer/Niki Parmar/Jakob Uszkoreit/Llion Jones/Aidan N. Gomez/Lukasz Kaiser/Illia Polosukhin (2017). *Attention Is All You Need*. arXiv: 1706.03762 [cs.CL].
- Zha, Daochen/Zaid Pervaiz Bhat/Kwei-Herng Lai/Fan Yang/Zhimeng Jiang/Shaochen Zhong/Xia Hu (2023). *Data-centric Artificial Intelligence: A Survey*. arXiv: 2303.10158 [cs.LG].