



UNIVERSITÄT
LEIPZIG

Medizinische Fakultät

Vortrag - 1

Question Answering auf dem Lehrbuch 'Health Information Systems' mit Hilfe von unüberwachtem Training eines Pretrained Transformers

Leipzig, Juni 2023

Paul Keller

imise 

- ▶ Einleitung
 - ▶ Gegenstand und Motivation
 - ▶ Problematik
 - ▶ Ziel
 - ▶ Aufgaben
- ▶ Stand der Forschung
 - ▶ Grundlagen Forschung
 - ▶ Continual Pretraining
 - ▶ Aktuelle Modelle
 - ▶ Erforschung von Modellen
- ▶ Zeitplanung

Einleitung - Gegenstand und Motivation

- ▶ Informationsbeschaffung in der Medizininformatik
- ▶ SNIK-Ontologie
- ▶ BeLL-Arbeit (Brunch, 2022): QAS über SNIK
- ▶ ChatGPT als Wissensgrundlage?
- ▶ ChatGPT vs QAS (Omar u.a. 2023)

Einleitung - Gegenstand und Motivation

- ▶ Informationsbeschaffung in der Medizininformatik
- ▶ SNIK-Ontologie
- ▶ BeLL-Arbeit (Brunch, 2022): QAS über SNIK
- ▶ ChatGPT als Wissensgrundlage?
- ▶ ChatGPT vs QAS (Omar u.a. 2023)

Wilhelm-Ostwald-Schule, Gymnasium der Stadt Leipzig
Universität Leipzig
Medizinische Fakultät
Institut für Medizinische Informatik, Statistik und Epidemiologie

QUESTION ANSWERING AUF SNIK

BESONDERE LERNLEISTUNG

vorgelegt von

Hannes Raphael Brunsch
Besondere Lernleistung Referenzfach Informatik

Leipzig, den 20. Januar 2023

Einleitung - Gegenstand und Motivation

- ▶ Informationsbeschaffung in der Medizininformatik
- ▶ SNIK-Ontologie
- ▶ BeLL-Arbeit (Brunch, 2022): QAS über SNIK
- ▶ ChatGPT als Wissensgrundlage?
- ▶ ChatGPT vs QAS (Omar u.a. 2023)

**ChatGPT versus Traditional Question Answering for Knowledge Graphs:
Current Status and Future Directions Towards Knowledge Graph Chatbots**

Reham Omar¹, Omij Mangukiya¹, Panos Kalnis² and Essam Mansour¹

¹Concordia University, Canada, ²KAUST, Saudi Arabia

{Fname}.{Lname}@concordia.ca, panos.kalnis@kaust.edu.sa

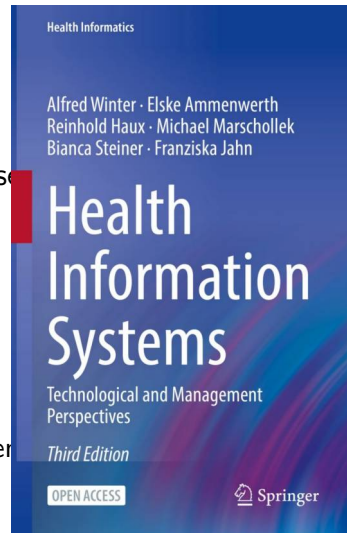


Einleitung - Problematik

- ▶ Literatur zu Informationssysteme im Gesundheitswesen
 - ▶ Blaue Buch „Health Information Systems“
 - ▶ Komplex
 - ▶ Umfangreich
 - ▶ Fragmentiert
- ▶ Problem
 - ▶ Komplexe Sachverhalte in die Praxis übersetzen
 - ▶ Extraktion von relevanten Informationen aus großem Umfang

Einleitung - Problematik

- ▶ Literatur zu Informationssysteme im Gesundheitswesen
 - ▶ Blaue Buch „Health Information Systems“
 - ▶ Komplex
 - ▶ Umfangreich
 - ▶ Fragmentiert
- ▶ Problem
 - ▶ Komplexe Sachverhalte in die Praxis übersetzen
 - ▶ Extraktion von relevanten Informationen aus großer



Einleitung - Problematik

- ▶ Literatur zu Informationssysteme im Gesundheitswesen
 - ▶ Blaue Buch „Health Information Systems“
 - ▶ Komplex
 - ▶ Umfangreich
 - ▶ Fragmentiert
- ▶ Problem
 - ▶ Komplexe Sachverhalte in die Praxis übersetzen
 - ▶ Extraktion von relevanten Informationen aus großem Umfang

Einleitung - Zielsetzung

1. Beantwortung von Fragen zu Informationssystemen im Gesundheitswesen in natürlicher Sprache durch eine Konversations-KI mit Hilfe von Winter u.a. (2023)
2. Lösung einer Beispielklausur des Moduls „Architektur von Informationssystemen im Gesundheitswesen“ mit Hilfe einer Konversations-KI.
3. Das Ziel ist kein Produktivsystem, sondern soll lediglich die Machbarkeit der Beantwortung von Fragen mit Hilfe einer Konversations-KI

Einleitung - Aufgabenstellung

1. Ziel 1

- 1.1 Vergleich von vortrainierten Sprachmodellen aus Nutzbarkeit und Leistung
- 1.2 Datenkuration von Winter u.a. (2023)
- 1.3 Continual Pretraining eines Modells

2. Ziel 2

- 2.1 Bewertung von Klausurfragen und deren generierte Antworten
- 2.2 Evaluation vor und nach dem Training
- 2.3 Vergleich mit SOTA (GPT-4)

- ▶ Erstmalige Transformer-Architektur: Vaswani u. a. (2017)
 - ▶ Weiterentwicklung Deep Residual Connections: He u. a. (2016)
 - ▶ Weiterentwicklung Dropout: Srivastava u. a. (2014)
 - ▶ Weiterentwicklung Byte Pair Encoding: Sennrich u. a. (2016)
- ▶ Transformer Taxonomie: Kalyan u. a. (2022)

Attention Is All You Need

► Erstmalige Tra

► Weiterentv

► Weiterentv

► Weiterentv

► Transformer T.

Ashish Vaswani*
Google Brain
avaswani@google.com

Noam Shazeer*
Google Brain
noam@google.com

Niki Parmar*
Google Research
nikip@google.com

Jakob Uszkoreit*
Google Research
usz@google.com

Llion Jones*
Google Research
llion@google.com

Aidan N. Gomez* †
University of Toronto
aidan@cs.toronto.edu

Lukasz Kaiser*
Google Brain
lukaszkaiser@google.com

Illia Polosukhin* ‡
illia.polosukhin@gmail.com

- ▶ Erstmalige Transformer-Architektur: Vaswani u. a. (2017)
 - ▶ Weiterentwicklung Deep Residual Connections: He u. a. (2016)
 - ▶ Weiterentwicklung Dropout: Srivastava u. a. (2014)
 - ▶ Weiterentwicklung Byte Pair Encoding: Sennrich u. a. (2016)
- ▶ Transformer Taxonomie: Kalyan u. a. (2022)

Deep Residual Learning for Image Recognition

Kaiming He Xiangyu Zhang Shaoqing Ren Jian Sun
Microsoft Research
{kahe, v-xiangz, v-shren, jiansun}@microsoft.com

Journal of Machine Learning Research 15 (2014) 1929-1958

Submitted 11/13; Published 6/14

▶ Erstmalige Transformer-Architektur: Vaswani

▶ Weiterentwicklung Deep Residual Connections: Dropout: A Simple Way to Prevent Neural Networks from Overfitting

▶ Weiterentwicklung Dropout: Srivastava u. a.

▶ Weiterentwicklung Byte Pair Encoding: Sennrich

Nitish Srivastava
Geoffrey Hinton
Alex Krizhevsky
Ilya Sutskever

NITISH@CS.TORONTO.EDU
HINTON@CS.TORONTO.EDU
KRIZ@CS.TORONTO.EDU
ILYA@CS.TORONTO.EDU
RSALAKHU@CS.TORONTO.EDU

▶ Transformer Taxonomie: Kalyan u. a. (2022)

Ruslan Salakhutdinov
Department of Computer Science
University of Toronto
10 Kings College Road, Rm 3302
Toronto, Ontario, M5S 3G4, Canada.

Editor: Yoshua Bengio

Neural Machine Translation of Rare Words with Subword Units

Rico Sennrich and Barry Haddow and Alexandra Birch

School of Informatics, University of Edinburgh
{rico.sennrich, a.birch}@ed.ac.uk, bhaddow@inf.ed.ac.uk

- ▶ Erstmalige Transformer-Architektur: Vaswani u. a. (2017)
 - ▶ Weiterentwicklung Deep Residual Connections: He u. a. (2016)
 - ▶ Weiterentwicklung Dropout: Srivastava u. a. (2014)
 - ▶ Weiterentwicklung Byte Pair Encoding: Sennrich u. a. (2016)
- ▶ Transformer Taxonomie: Kalyan u. a. (2022)

AMMUS : A Survey of Transformer-based Pretrained Models in Natural Language Processing

Katikapalli Subramanyam Kalyan, Ajit Rajasekharan, and Sivanesan Sangeetha

Continual Pretraining

- ▶ Leistungssteigerung durch Continual Pretraining
 - ▶ GPT: Radford und Narasimhan (2018)
 - ▶ BioBERT: Lee u. a. (2019)
 - ▶ Gururangan u. a. (2020)
- ▶ Leistungssteigerung durch Größere Modelle: Kaplan u. a. (2020)
- ▶ Leistungssteigerung durch Fine-Tuning: Ziegler u. a. (2019)

Continual Pretraining

- ▶ Leistungssteigerung durch Continual
 - ▶ GPT: Radford und Narasimhan (2019)
 - ▶ BioBERT: Lee u. a. (2019)
 - ▶ Gururangan u. a. (2020)
- ▶ Leistungssteigerung durch Größere Modelle
- ▶ Leistungssteigerung durch Fine-Tuning

Improving Language Understanding by Generative Pre-Training

Alec Radford OpenAI alec@openai.com	Karthik Narasimhan OpenAI karthikn@openai.com	Tim Salimans OpenAI tim@openai.com	Ilya Sutskever OpenAI ilyasu@openai.com
---	---	--	---

Data and text mining

BioBERT: a pre-trained biomedical language representation model for biomedical text mining

Jinhyuk Lee ^{1,†}, Wonjin Yoon ^{1,†}, Sungdong Kim ², Donghyeon Kim ¹, Sunkyu Kim ¹, Chan Ho So ³ and Jaewoo Kang ^{1,3,*}

¹Department of Computer Science and Engineering, Korea University, Seoul 02841, Korea, ²Clova AI Research, Naver Corp, Seong-Nam 13561, Korea and ³Interdisciplinary Graduate Program in Bioinformatics, Korea University, Seoul 02841, Korea

*To whom correspondence should be addressed.

[†]The authors wish it to be known that the first two authors contributed equally.

Associate Editor: Jonathan Wren

Received on May 16, 2019; revised on July 29, 2019; editorial decision on August 25, 2019; accepted on September 5, 2019

Don't Stop Pretraining: Adapt Language Models to Domains and Tasks

Suchin Gururangan[†] Ana Marasović^{†,◇} Swabha Swayamdipta[†]
Kyle Lo[†] Iz Beltagy[†] Doug Downey[†] Noah A. Smith^{†,◇}

[†]Allen Institute for Artificial Intelligence, Seattle, WA, USA

[◇]Paul G. Allen School of Computer Science & Engineering, University of Washington, Seattle, WA, USA
{suchin, anam, swabhas, kylel, beltagy, dougd, noah}@allenai.org

Scaling Laws for Neural Language Models

► Leistungssteigerung durch Continual

- GPT: Radford und Narasimhan (2019)
- BioBERT: Lee u. a. (2019)
- Gururangan u. a. (2020)

► Leistungssteigerung durch Größere Modelle: Kaplan u. a. (2020)

► Leistungssteigerung durch Fine-Tuning: Ziegler u. a. (2019)

Jared Kaplan *
Johns Hopkins University, OpenAI
jaredk@jhu.edu

Sam McCandlish*
OpenAI
sam@openai.com

Tom Henighan OpenAI henighan@openai.com	Tom B. Brown OpenAI tom@openai.com	Benjamin Chess OpenAI bchess@openai.com	Rewon Child OpenAI rewon@openai.com
Scott Gray OpenAI scott@openai.com	Alec Radford OpenAI alec@openai.com	Jeffrey Wu OpenAI jeffwu@openai.com	Dario Amodei OpenAI damodei@openai.com

Continual Pretraining

- ▶ Leistungssteigerung durch Continual Pretraining
 - ▶ GPT: Radford und Narasimhan (2018)
 - ▶ BioBERT: Lee u. a. (2019)
 - ▶ Gururangan u. a. (2020)
 - ▶ Leistungssteigerung durch Größere Modelle: Kaplan u. a. (2020)
 - ▶ Leistungssteigerung durch Fine-Tuning: Ziegler u. a. (2019)
-

Fine-Tuning Language Models from Human Preferences

Daniel M. Ziegler* Nisan Stiennon* Jeffrey Wu Tom B. Brown
Alec Radford Dario Amodei Paul Christiano Geoffrey Irving
OpenAI

{dmz,nisan,jeffwu,tom,alec,damodei,paul,irving}@openai.com

- ▶ GPT-2: Radford u. a. (2019)
- ▶ GPT-4: OpenAI (2023)
- ▶ GPT-NeoX: Black u. a. (2022)
- ▶ LLaMa: Touvron u. a. (2023)

Language Models are Unsupervised Multitask Learners

Alec Radford^{*1} Jeffrey Wu^{*1} Rewon Child¹ David Luan¹ Dario Amodei^{**1} Ilya Sutskever^{**1}

GPT-4 Technical Report

OpenAI*

GPT-NeoX-20B: An Open-Source Autoregressive Language Model

Sid Black*

Stella Biderman*

Eric Hallahan*

Quentin Anthony

Leo Gao

Laurence Golding

Horace He

Connor Leahy

Kyle McDonell

Jason Phang

Michael Pieler

USVSN Sai Prashanth

Shivanshu Purohit

Laria Reynolds

Jonathan Tow

Ben Wang

Samuel Weinbach

- ▶ GPT-2: Radford u. a. (2019)
- ▶ GPT-4: OpenAI (2023)
- ▶ GPT-NeoX: Black u. a. (2022)
- ▶ LLaMa: Touvron u. a. (2023)

LLaMA: Open and Efficient Foundation Language Models

Hugo Touvron*, Thibaut Lavril*, Gautier Izacard*, Xavier Martinet
Marie-Anne Lachaux, Timothee Lacroix, Baptiste Rozière, Naman Goyal
Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin
Edouard Grave*, Guillaume Lample*

Meta AI

- ▶ Adapter: Houlby u. a. (2019)
- ▶ Transformer als Informationsquelle
 - ▶ Petroni u. a. (2019)
 - ▶ Wissensneuronen: Dai u. a. (2022)
- ▶ AI Safety: OpenAI (2023)

Parameter-Efficient Transfer Learning for NLP

Neil Houlsby¹ Andrei Giurgiu^{1*} Stanisław Jastrzębski^{2*} Bruna Morrone¹ Quentin de Laroussilhe¹
Andrea Gesmundo¹ Mona Attariyan¹ Sylvain Gelly¹

- ▶ Adapter: Houlsby u. a. (2019)
- ▶ Transformer als Informationsquelle
 - ▶ Petroni u. a. (2019)
 - ▶ Wissensneuronen: Dai u. a. (2022)
- ▶ AI Safety: OpenAI (2023)

Language Models as Knowledge Bases?

**Fabio Petroni¹ Tim Rocktäschel^{1,2} Patrick Lewis^{1,2} Anton Bakhtin¹
Yuxiang Wu^{1,2} Alexander H. Miller¹ Sebastian Riedel^{1,2}**

¹Facebook AI Research

²University College London

{fabiopetroni, rockt, plewis, yolo, yuxiangwu, ahm, sriedel}@fb.com

Erforschung von Modellen

- ▶ Adapter: Houlsby u. a. (2019)
- ▶ Transformer als Informationsquelle
 - ▶ Petroni u. a. (2019)
 - ▶ Wissensneuronen: Dai u. a. (2022)
- ▶ AI Safety: OpenAI (2023)

Knowledge Neurons in Pretrained Transformers

Damai Dai^{†‡*}, Li Dong[‡], Yaru Hao[‡], Zhifang Sui[‡], Baobao Chang[‡], Furu Wei[‡]

[†]MOE Key Lab of Computational Linguistics, Peking University

[‡]Microsoft Research

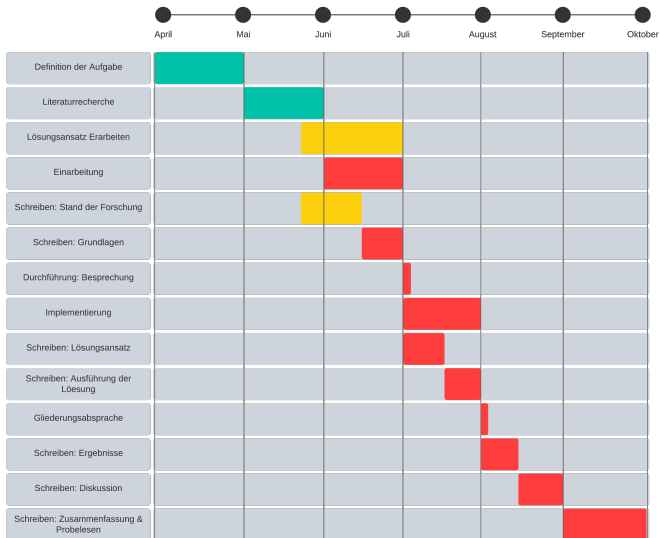
{daidamai, szf, chbb}@pku.edu.cn

{lidong1, yaruhao, fuwei}@microsoft.com

Erforschung von Modellen

- ▶ Adapter: Houlsby u. a. (2019)
- ▶ Transformer als Informationsquelle
 - ▶ Petroni u. a. (2019)
 - ▶ Wissensneuronen: Dai u. a. (2022)
- ▶ AI Safety: OpenAI (2023)

Zeitplanung



- ▶ Ben Wang (2021). „Mesh-Transformer-JAX: Model-Parallel Implementation of Transformer Language Model with JAX“. In: <https://github.com/kingoflolz/mesh-transformer-jax>
- ▶ Black, Sidney u. a. (Mai 2022). „GPT-NeoX-20B: An Open-Source Autoregressive Language Model“. In: Proceedings of BigScience Episode 5 – Workshop on Challenges Perspectives in Creating Large Language Models. Association for Computational Linguistics, S. 95–136. doi: 10.18653/v1/2022.bigscience-1.9.
- ▶ Brown, Tom u. a. (2020). „Language Models are Few-Shot Learners“. In: Advances in Neural Information Processing Systems. Hrsg. von H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan und H. Lin. Bd. 33. Curran Associates, Inc., S. 1877–1901. url: https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfc4967418bfb8ac142f64a-Paper.pdf
- ▶ Brunsch, Hannes Raphael (2022). „Question Answering auf SNIK“. Besondere Lernleistung. Leipzig, Germany: Wilhelm-Ostwald-Schule. url: <https://www.snik.eu/public/bell-hrb.pdf>.
- ▶ Dai, Damai, Li Dong, Yaru Hao, Zhifang Sui, Baobao Chang und Furu Wei (2022). Knowledge Neurons in Pretrained Transformers. arXiv: 2104.08696 [cs.CL].
- ▶ Dehouche, Nassim (März 2021). „Plagiarism in the age of massive Generative Pre-trained Transformers (GPT-3): “The best time to act was yesterday. The next best time is now.”“ In: Ethics in Science and Environmental Politics 21. doi: 10.3354/esep00195
- ▶ Gururangan, Suchin, Ana Marasovi´c, Swabha Swayamdipta, Kyle Lo, Iz Beltagy, Doug Downey und Noah A. Smith (Juli 2020). „Don't Stop Pretraining: Adapt Language Models to Domains and Tasks“. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Online: Association for Computational Linguistics, S. 8342–8360. doi: 10.18653/v1/2020.acl-main.740. url:<https://aclanthology.org/2020.acl-main.740>.
- ▶ He, Kaiming, Xiangyu Zhang, Shaoqing Ren und Jian Sun (2016). „Deep Residual Learning for Image Recognition“. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), S. 770–778. doi: 10.1109/CVPR.2016.90.
- ▶ Hounsby, Neil, Andrei Giurgiu, Stanislaw Jastrzebski, Bruna Morrone, Quentin De Laroussilhe, Andrea Gesmundo, Mona Attariyan und Sylvain Gelly (2019). „Parameter-Efficient Transfer Learning for NLP“. In: Proceedings of the 36th International Conference on Machine Learning. Hrsg. von Kamalika Chaudhuri und Ruslan Salakhutdinov. Bd. 97. Proceedings of Machine Learning Research. PMLR, S. 2790–2799. url: <https://proceedings.mlr.press/v97/hounsby19a.html>.

Quellen 2

- ▶ Jiang, Zhengbao, Antonios Anastasopoulos, Jun Araki, Haibo Ding und Graham Neubig (Nov. 2020). „X-FACTR: Multilingual Factual Knowledge Retrieval from Pretrained Language Models“. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). Online: Association for Computational Linguistics, S. 5943–5959. doi: 10.18653/v1/2020.emnlp-main.479.
- ▶ Kalyan, Katikapalli Subramanyam, Ajit Rajasekharan und Sivasenan Sangeetha (2022). „AMMU: A survey of transformer-based biomedical pretrained language models“. In: Journal of biomedical informatics 126, S. 103982. doi: 10.1016/j.jbi.2021.103982.
- ▶ Kaplan, Jared, Sam McCandlish, Tom Henighan, Tom B. Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu und Dario Amodei (2020). Scaling Laws for Neural Language Models. arXiv: 2001.08361 [cs.LG].
- ▶ Lee, Jinhyuk, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So und Jaewoo Kang (2019). „BioBERT: a pre-trained biomedical language representation model for biomedical text mining“. In: Bioinformatics 36.4, S. 1234–1240. issn: 1367-4803. doi: 10.1093/bioinformatics/btz682.
- ▶ Omar, Reham, Omij Mangukiya, Panos Kalnis und Essam Mansour (2023). ChatGPT versus Traditional Question Answering for Knowledge Graphs: Current Status and Future Directions Towards Knowledge Graph Chatbots. Version 1. arXiv: 2302.06466 [cs.CL].
- ▶ OpenAI (2023). „GPT-4 Technical Report“. In: arXiv: 2303.08774[cs.CL].
- ▶ Petroni, Fabio, Tim Rocktäschel, Sebastian Riedel, Patrick Lewis, Anton Bakhtin, Yuxiang Wu und Alexander Miller (Nov. 2019). „Language Models as Knowledge Bases?“ In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong, China: Association for Computational Linguistics, S. 2463–2473. doi: 10.18653/v1/D19-1250.
- ▶ Pfeiffer, Jonas, Andreas Rücklé, Clifton Poth, Aishwarya Kamath, Ivan Vulic, Sebastian Ruder, Kyunghyun Cho und Iryna Gurevych (Okt. 2020). „AdapterHub: A Framework for Adapting Transformers“. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. Online: Association for Computational Linguistics, S. 46–54. doi: 10.18653/v1/2020.emnlp-demos.7. url: <https://aclanthology.org/2020.emnlp-demos.7>
- ▶ Radford, Alec und Karthik Narasimhan (2018). „Improving Language Understanding by Generative Pre-Training“.
- ▶ Radford, Alec, Jeff Wu, Rewon Child, David Luan, Dario Amodei und Ilya Sutskever (2019). „Language Models are Unsupervised Multitask Learners“. In: url: <https://d4mucfpxsywv.cloudfront.net/better-language-models/language-models.pdf>.

Quellen 3

- ▶ Sennrich, Rico, Barry Haddow und Alexandra Birch (Aug. 2016). „Neural Machine Translation of Rare Words with Subword Units“. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Berlin, Germany: Association for Computational Linguistics, S. 1715–1725. doi: 10.18653/v1/P16-1162. url: <https://aclanthology.org/P16-1162>.
- ▶ Srivastava, Nitish, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever und Ruslan Salakhutdinov (2014). „Dropout: A Simple Way to Prevent Neural Networks from Overfitting“. In: Journal of Machine Learning Research 15.56, S. 1929–1958. url: <http://jmlr.org/papers/v15/srivastava14a.html>.
- ▶ Touvron, Hugo u. a. (2023). LLaMA: Open and Efficient Foundation Language Models. arXiv: 2302.13971 [cs.CL].
- ▶ Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser und Illia Polosukhin (2017). „Attention is All you Need“. In: Advances in Neural Information Processing Systems. Hrsg. von I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan und R. Garnett. Bd. 30. Curran Associates, Inc. url: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf
- ▶ Winter, Alfred, Elske Ammenwerth, Reinhold Haux, Michael Marschollek, Bianca Steiner und Franziska Jahn (2023). Health Information Systems. 3. Aufl. Health Informatics. Springer Cham. isbn: 978-3-031-12310-8. doi: 10.1007/978-3-031-12310-8.
- ▶ Ziegler, Daniel M., Nisan Stiennon, Jeffrey Wu, Tom B. Brown, Alec Radford, Dario Amodei, Paul F. Christiano und Geoffrey Irving (2019). „Fine-Tuning Language Models from Human Preferences“. In: CoRR abs/1909.08593. url: <http://arxiv.org/abs/1909.08593>.



UNIVERSITÄT
LEIPZIG

Medizinische Fakultät

VIELEN DANK!

Paul Keller

Institut für Medizinische Informatik, Statistik und Epidemiologie (IMISE)

www.imise.uni-leipzig.de