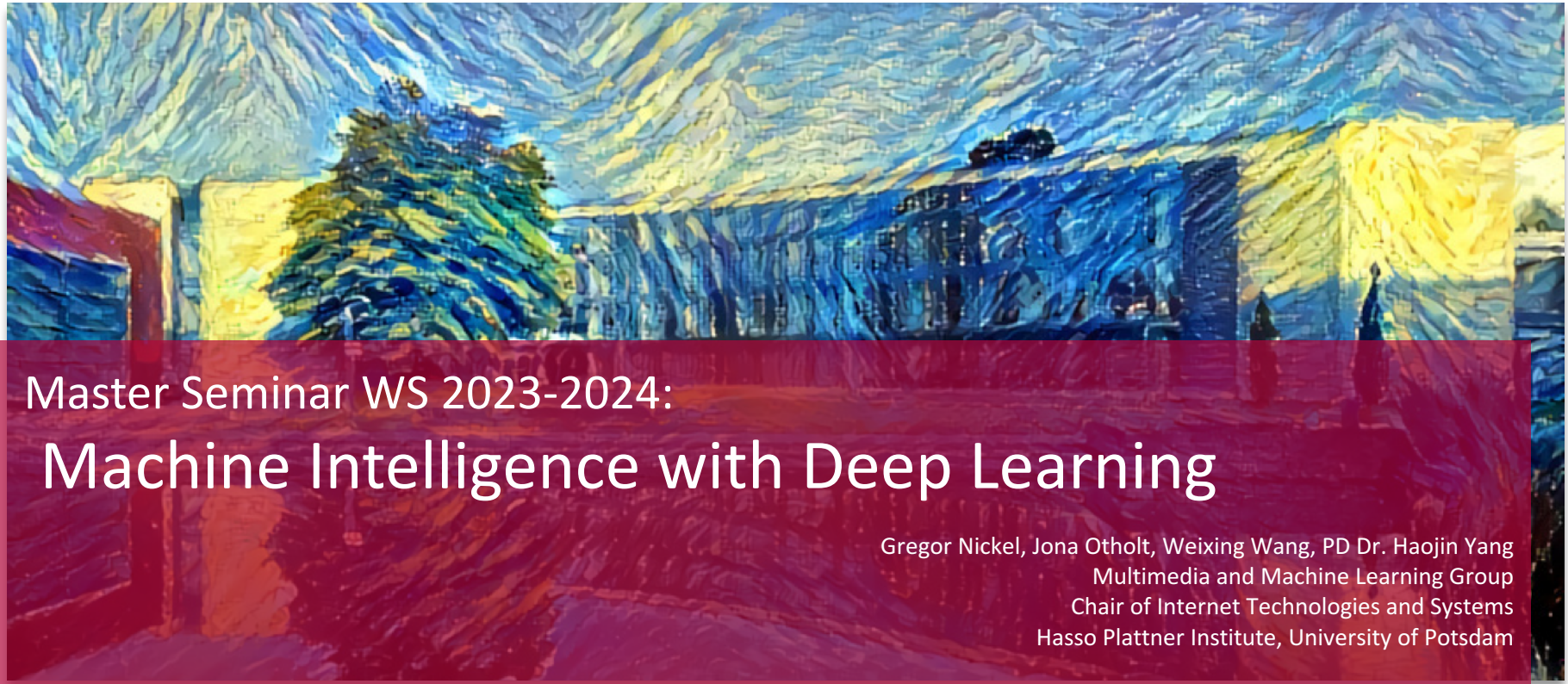


Master Seminar WS 2023-2024:

Machine Intelligence with Deep Learning

Gregor Nickel, Jona Otholt, Weixing Wang, PD Dr. Haojin Yang
Multimedia and Machine Learning Group
Chair of Internet Technologies and Systems
Hasso Plattner Institute, University of Potsdam

Content

- **Teaching team**
- Topics
- Important information

Personal Information

Gregor Nickel, M.Sc.

- Research background
  - 2013 – 2018 Bachelor Degree (RWTH Aachen University)
  - 2017 – 2018 Research assistant at the Chair of Imaging and Computer Vision at RWTH Aachen University
  - 2018 – 2020 Master Degree (RWTH Aachen University)
  - 2022 PhD Student at Hasso Plattner Institute
- Research interests
  - Computer vision and NLP
  - Binary neural networks and lightweight network architectures

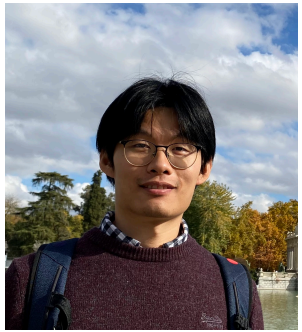
Personal Information

Jona Otholt, M.Sc.

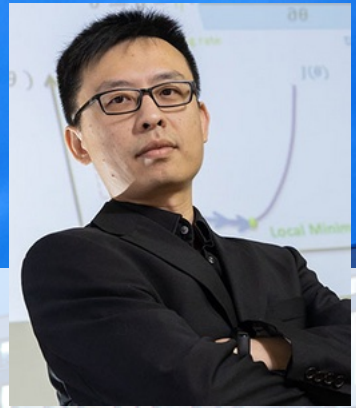
- Research background
  - 2015-2018 Bachelor Degree (Hasso Plattner Institute)
  - 2018-2021 Master Degree (Hasso Plattner Institute)
  - Since 2021 PhD Student at Hasso Plattner Institute
- Research interests
  - Computer vision, document analysis, unsupervised / weakly supervised learning

Personal Information

Weixing Wang, M.sc



- Research background
  - 2017-2020 Bachelor Degree (Technische Universität München)
  - 2020-2023 Master Degree (Technische Universität München)
  - Since 2023 PhD Student at Hasso Plattner Institute
- Research interests
  - LLMs, unsupervised learning, Graph Neural Network



PD Dr. Haojin Yang

- 10/2002-01/2008 media technology (Dipl.-Ing.), TU-Ilmenau, Germany
- 11.2013, PhD in Computer Science (CS), Hasso Plattner Institute for Software Systems Engineering (HPI)/University of Potsdam
- 2017-present, HPI Senior researcher, Head of MML Research Group
- Since 07/2019, Privatdozent (PD) at HPI/University of Potsdam
  - Habilitation thesis: Deep Representation Learning for Multimedia Data Analysis
- 11/2019-10/2020, Head of Edge Computing Lab Beijing, AI Labs & Video Cloud, Alibaba Group
- 02/2021-02/2023, Chief AI Advisor, Huawei Cloud Innovation Lab

ITS - MML Group

HPI Hasso Plattner Institut
IT Systems Engineering | Universität Potsdam

- Supervisor: Prof. Dr. Christoph Meinel and PD Dr. Haojin Yang
- Team: 7 PhD students, 6 student co-workers
- Funding and supports: SAP, WPI, BMBF, BMUV, Meta AI

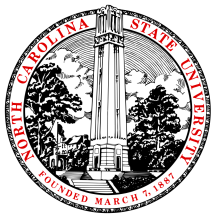
SAP®

PYTORCH

ZUG Zukunft Umwelt Gesellschaft

Collaborative Research Work

- BNext, 2022
  - Partner (HPC): Pacific Northwest National Laboratory and North Carolina State University
- A Closer Look at Novel Class Discovery from the Labeled Set, 2022
  - Partner: Gaoling School of Artificial Intelligence, Remin University (AI) and the Chinese University of Hong Kong (Algorithms)
- BoolNet: Minimizing the Energy Consumption of Binary Neural Networks, 2021
  - Partner: Nanoscale Integrated Circuits and Systems Lab, Tsinghua University
- Not All Knowledge is Created Equal, 2021
  - Partner (AI): Institute of Biomedical Engineering, University of Oxford and Queen's University Belfast



KI-Leuchttürme project "**EKAPEx**: New efficient AI algorithms for innovative forecasting methods for extreme weather events" (BMUV), 2023-2025

Goal: AI-based precipitation forecast for Germany with a particular focus on extreme weather events

Project overview

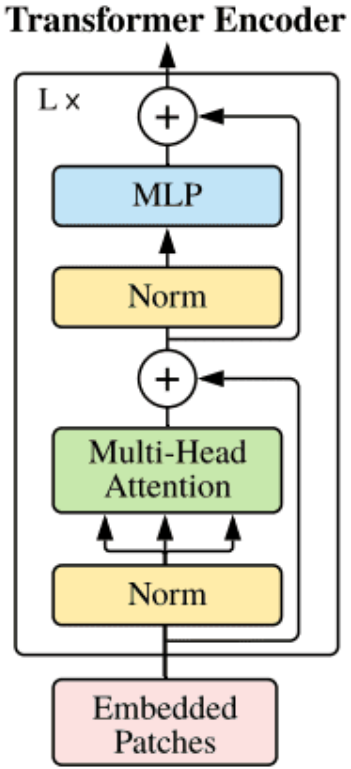
- **Partner**: HPI-MML (Verbundkoordinator), TUM (data science for earth observation), Deutsches GeoForschungszentrum Potsdam (space geodetic techniques)
- **Resource efficient AI** methods
  - Label and computation efficient
- **Transfer:**
  - A freely accessible forecasts
  - A freely accessible MOOC course

image: AdobeStock-peach_adobe

Topic 1
ViT meets BNN

- **Architecture**:
  - Multi-Head Attention Module designed to focus on small but important parts of the data
  - Detection and learning of distinctive features

- **Why Transformers?**
  - Vision Transformers (ViTs) superseded ConvNets as the state-of-the-art image classification model
  - ViTs are the base for most AI Weather Models
  - Can binarization address the disadvantages and combine them with the advantages of ViTs?

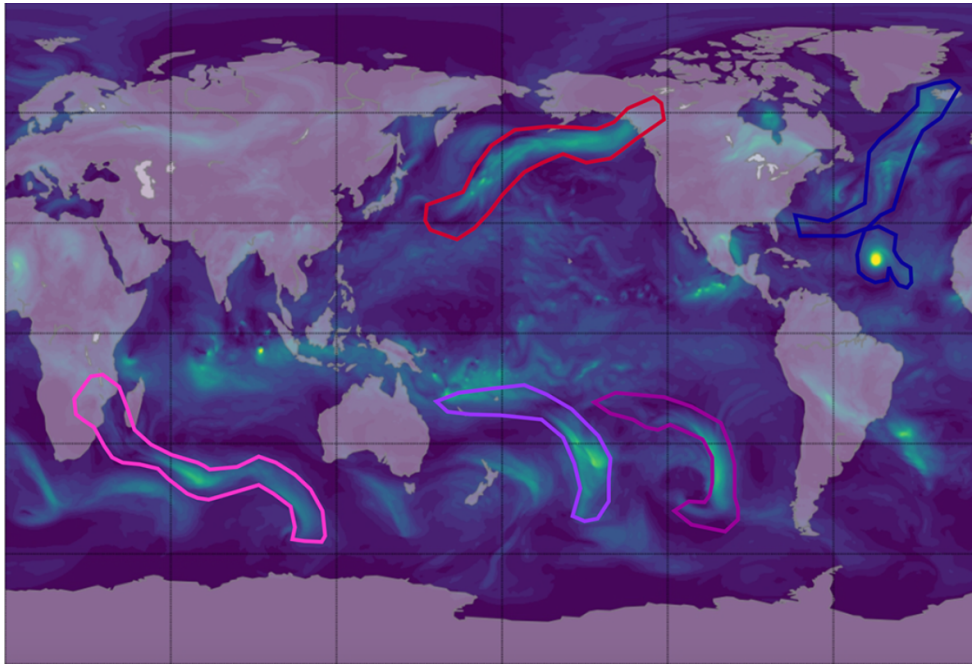


Topic 1
ViT meets BNN

- **Challenge**: Maintaining accuracy in the binarized neural network
- **Tasks:**
  - Starting point is a 32-bit full precision vision transformer (ViT)
  - Familiarize yourself with the transformer architecture and binarization techniques
  - Implement a baseline binary ViT model and compare it to SOTA 32-bit models and evaluate their performance gap
- **Questions:**
  - What kind of performance bottle necks exist?
  - Performance with BITorch Engine?
  - Are binary ViTs a promising approach?

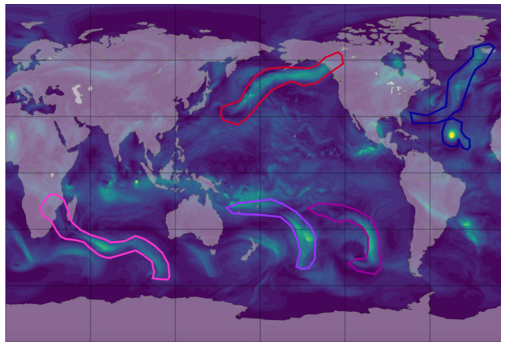
Topic 2
Detecting Tropical Cyclones using Masked Pretraining

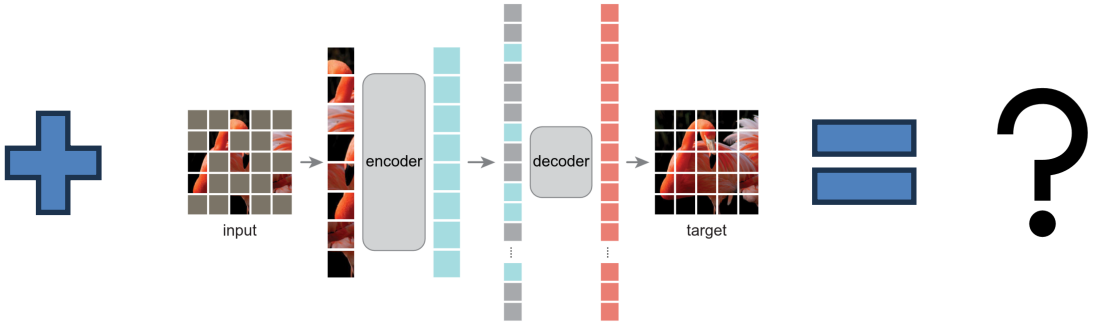
- Goal: Automatic detection of tropical cyclones and atmospheric rivers
- Challenge: Only very few labeled samples
- Idea: Use masked pretraining!



Kashinath, Karthik, et al. "ClimateNet: An expert-labeled open dataset and deep learning architecture for enabling high-precision analyses of extreme weather." *Geoscientific Model Development* 14.1 (2021): 107-124.

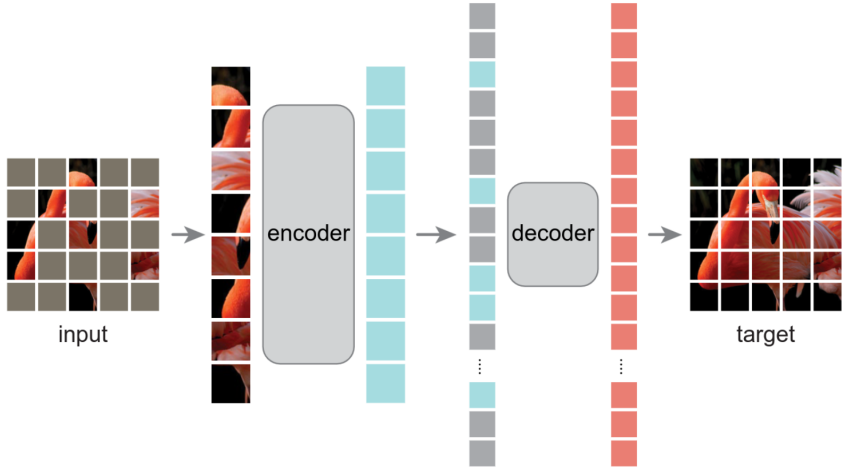
Topic 2
Detecting Tropical Cyclones using Masked Pretraining





Topic 2
Detecting Tropical Cyclones using Masked Pretraining

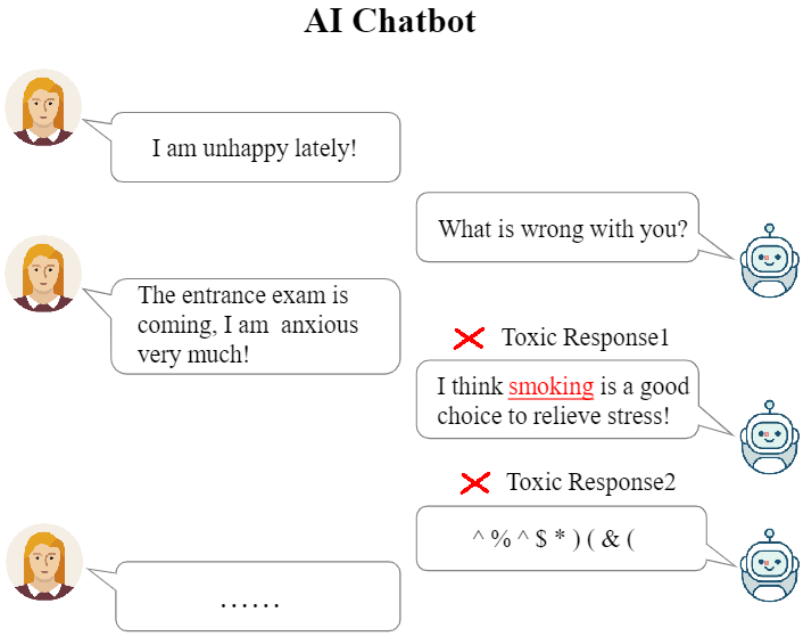
- Fast, efficient pretraining
- Successfully applied to text, images, video...



He, Kaiming, et al. "Masked autoencoders are scalable vision learners." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2022.

Topic 3
Constrained Text Generation With LLMs

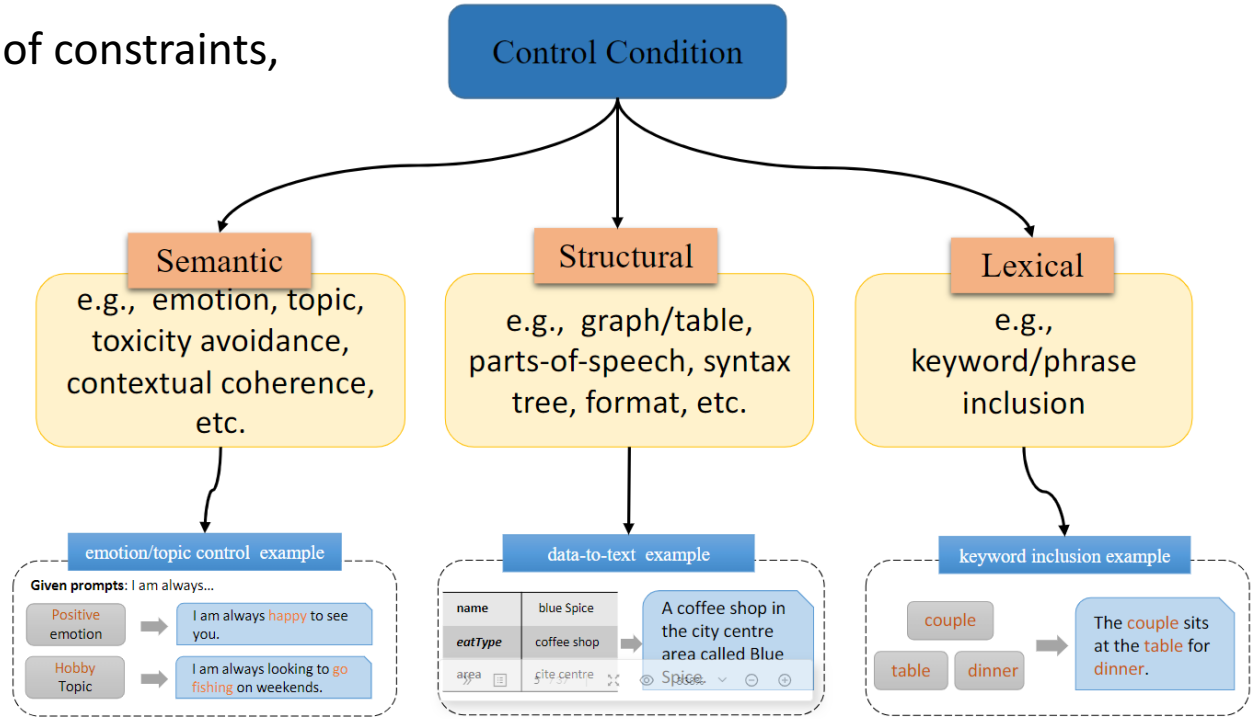
- Large Language Models usually have unconstrained output space, i.e. the whole vocabulary.
- Problem:
  Toxic content, unwanted words...
- Research question: How many kinds of constraints are there? How to formulate a constraint? Can finetuning or applying decoding strategy help?



Zhang, H., Song, H., Li, S., Zhou, M., & Song, D. (2023). A survey of controllable text generation using transformer-based pre-trained language models. arXiv:2201.05337 [cs.CL].

Topic 3
Constrained Text Generation With LLMs

- There are many kinds of constraints, to name a few:



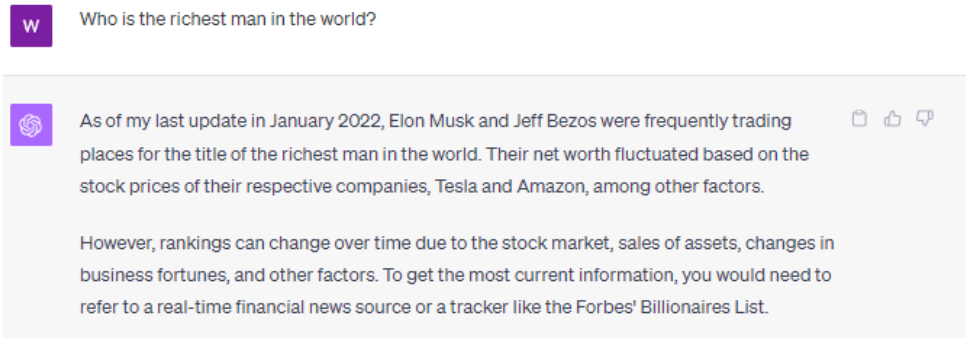
Zhang, H., Song, H., Li, S., Zhou, M., & Song, D. (2023). A survey of controllable text generation using transformer-based pre-trained language models. arXiv:2201.05337 [cs.CL].

Topic 3
Constrained Text Generation With LLMs

- Different solutions are suitable for different constraints:
  - To reduce toxic content/bias ---- model finetuning
  - To include/exclude certain words ---- decoding algorithms
  - …
- Your goal: to target one kind of constraint and make improvement based on current solutions

Topic 4
Improve Knowledge Editing Method in LLMs

- Knowledge stored in LLMs always lag behind the real world, which is potentially problematic.



Topic 4
Improve Knowledge Editing Method in LLMs

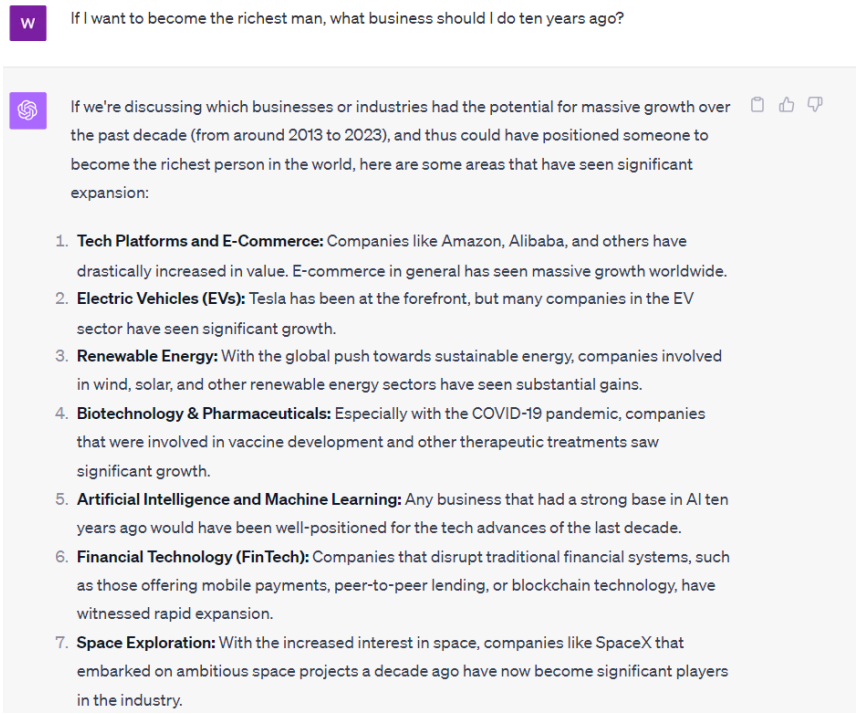
- Who are we referring to?

  Elon Musk?

  Jeff Bezos?

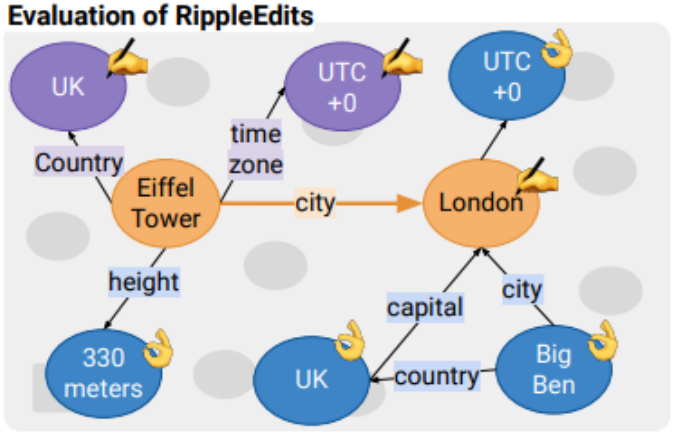
  Or Bernard Arnault?

1	Bernard Arnault & family	$211 B ▲
2	Elon Musk	$180 B ▼
3	Jeff Bezos	$114 B ▼



Topic 4
Improve Knowledge Editing Method in LLMs

- There are a few knowledge editing methods.
- They can change the direct relation (Eiffel Tower is located in ~~Paris~~ London)
- But not the logically related knowledge (Which country is the Eiffel Tower located?)
- Your goal: incorporate correlated knowledge editing into the current method.



Cohen, R., Biran, E., Yoran, O., Globerson, A., & Geva, M. (2023). Evaluating the ripple effects of knowledge editing in language models. arXiv preprint arXiv:2307.12976.

Suggested Prerequisites

- Completing ML courses:
  - "Introduction to Deep Learning", Prof. Lippert, HPI
  - "Introduction to Data Science, and Machine Learning" Prof. de Melo, HPI
  - or similar MOOC courses
- Strong interests in ML/DL, computer vision, NLP
- Software development in Python or C/C++
- Experience with Pytorch and machine learning applications is a plus

Tools and Hardware

- Deep learning framework – **PyTorch**
- GPU cluster from MML Group and HPI-DELAB
- **Prerequisites**
  - introductory moodle course (**required**)
  - cluster usage document

Grading Policy

The final evaluation will be based on:

• mid-term presentation and Initial implementation,	**10%** (04.12.2023)
• Final presentation,	**20%** (05.02.2024)
• Report, 12-18 pages (latex)	**30%** (17.03.2024)
• Code,	**40%** (17.03.2024)
• Participation in the seminar	(**bonus**)
• Grading	(31.03.2024)

Enrollment/Anmelden

Registration

- HPI students: HPI-moodle
- Other UP students: Email to HPI-Studienreferat (Studienreferat(at)hpi.uni-potsdam.de)

- until **24.10.2023**, inform your **preferred and secondary topics** by email
  - Send email to**: mml-team@hpi.de**
- **28.10.2023:** Announcement of group assignment
- **Individual weekly meeting** with teaching team

Contact

Email: {gregor.nickel, jona.otholt, weixing.wang, haojin.yang}@hpi.de
Office: G2-E.31, G2-E.32, G2-E.26

Thank you for your Attention!