

# Understanding and Training Language Models: Outlook and Conclusion

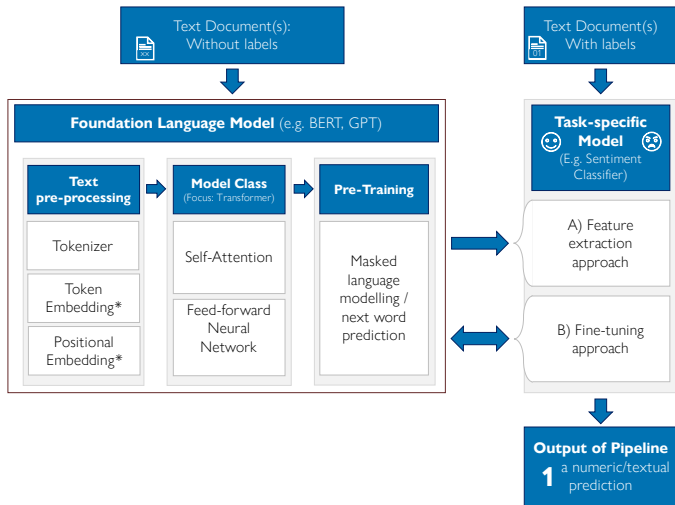
Erik-Jan Senn

Faculty of Mathematics and Statistics, University of St. Gallen

---

CSH Autumn School at University of Hohenheim  
September/October 2024

# What did we learn?



Language Modelling Pipeline (own illustration)

# Areas of Research in LLMs

## Improving LMs capabilities

- ▶ Reasoning / logical capabilities: LMs currently are not "smart", e.g. r's in Strawberry, river crossing riddle, math.
- ▶ Multimodal models: process different information sources and combine information e.g. visuals, audio, text.
- ▶ Retrieval-augmented generation: retrieve external information during inference (e.g. google search).
- ▶ Pretraining data quality.

# Areas of Research in LLMs

## Improving LM computation

- ▶ Smaller models with same capabilities.
- ▶ increasing context window (computational issue)
- ▶ efficient fine-tuning
- ▶ architecture adjustments

# Areas of Research in LLMs

## LM usage

- ▶ Interpretability of LM predictions
- ▶ Alignment of LMs with human needs: e.g. reinforcement learning from human feedback, instruction fine-tuning.
- ▶ Robustness: fairness / biases, adversarial attacks
- ▶ Data protection
- ▶ Remark: LMs as automated researcher?

# Questions ?

# Contact

Open for contact, questions, bugfixing, collaborations etc. after the course.

- ▶ [erik-jan.senn@unisg.ch](mailto:erik-jan.senn@unisg.ch)

- ▶ [eriksenn.github.io](https://eriksenn.github.io)

Thank you Johannes and CSH!



## Final Words

Thank you everyone!

# References