Seminar in NLP (SS24)

Parameter-Efficient Fine-Tuning in Natural Language Processing
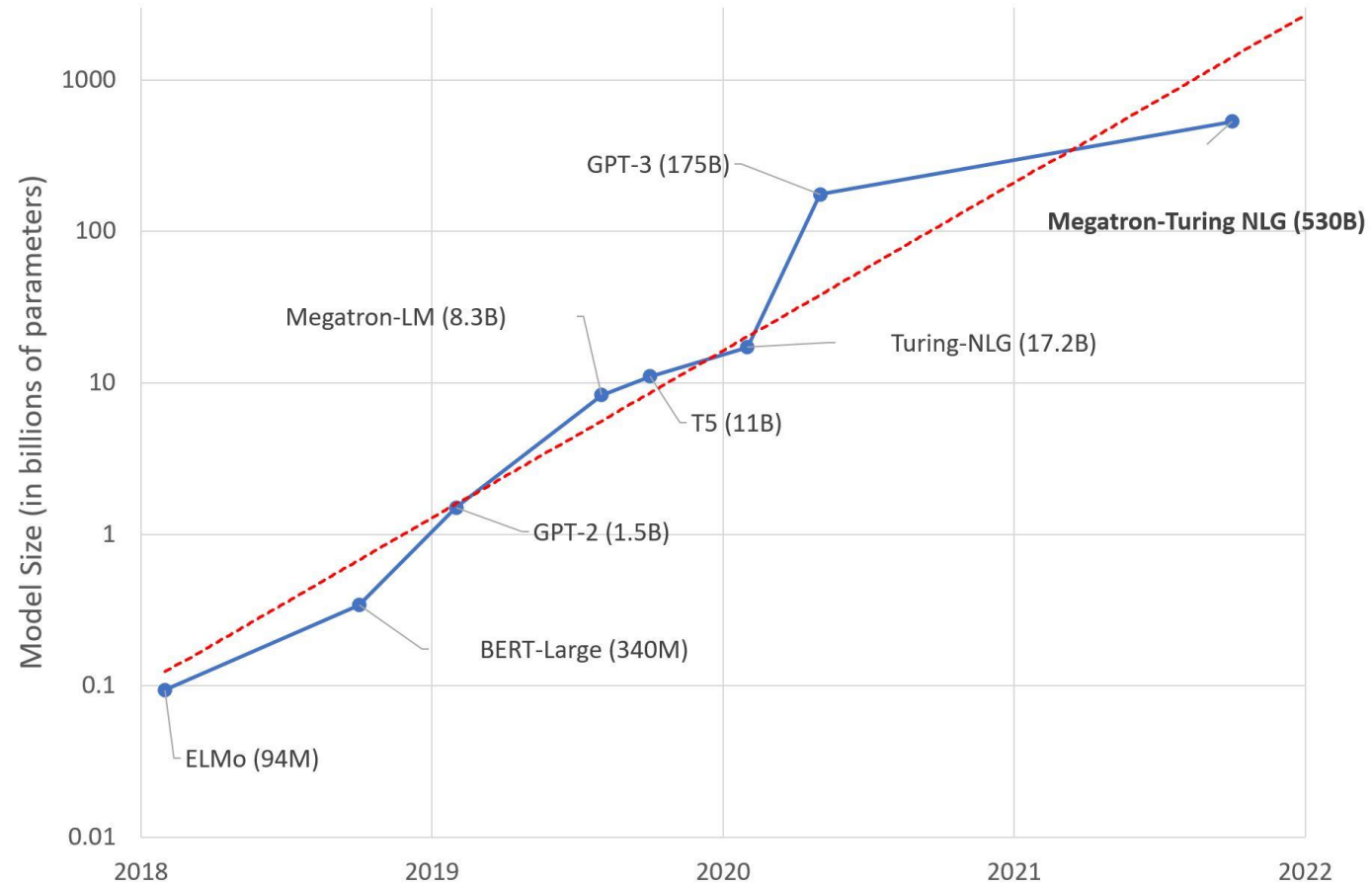
Chair XII for Natural Language Processing
Fabian David Schmidt
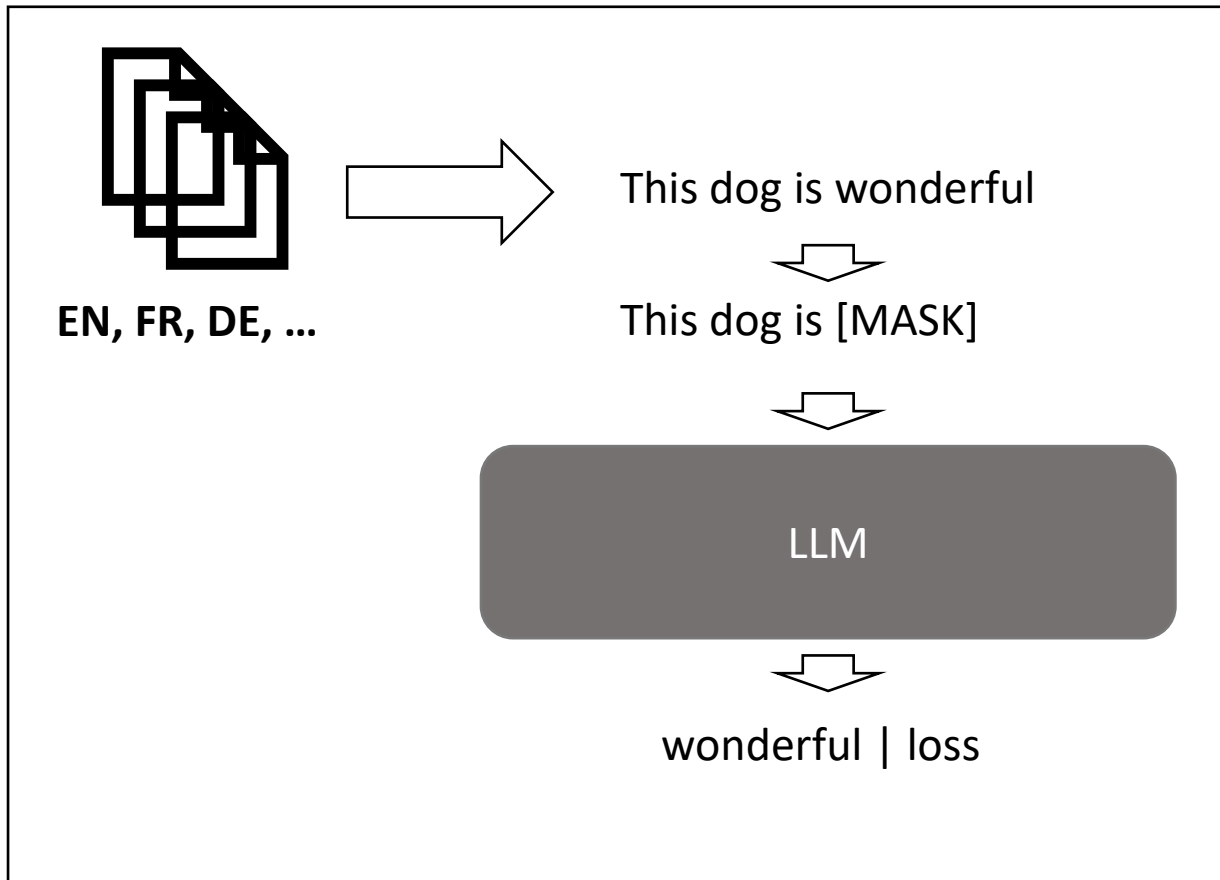Benedikt Ebing
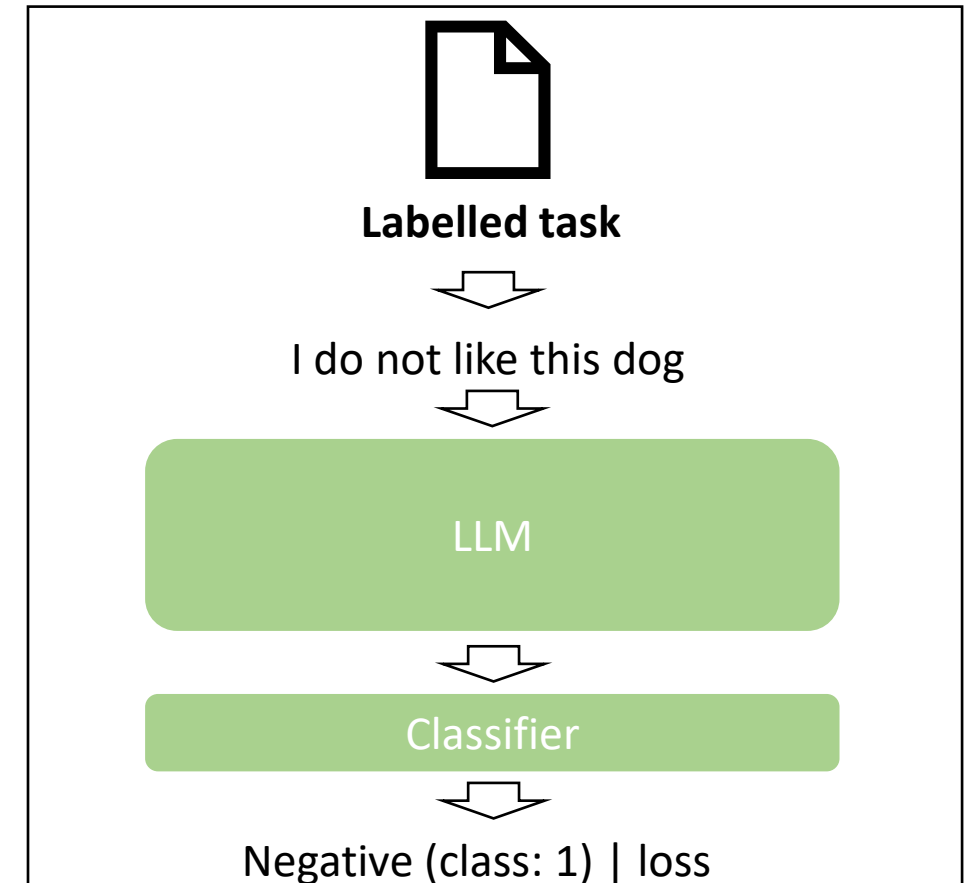Prof. Dr. Goran Glavaš

How to handle ever growing models?

https://huggingface.co/blog/large-language-models

Pretraining and fine-tuning with Large Langauge Model (LLMs)

Pretraining

EN, FR, DE, …

This dog is wonderful

This dog is [MASK]

LLM

wonderful | loss

Fine-Tuning

Labelled task

I do not like this dog

LLM

Classifier

Negative (class: 1) | loss

Pretraining and prompting with LLMs

Pretraining

This dog is wonderful

This dog is [MASK]

LLM

wonderful | loss

Prompting

Task

[x]. This sentence is [z]

This dog is awful. This sentence is [z]
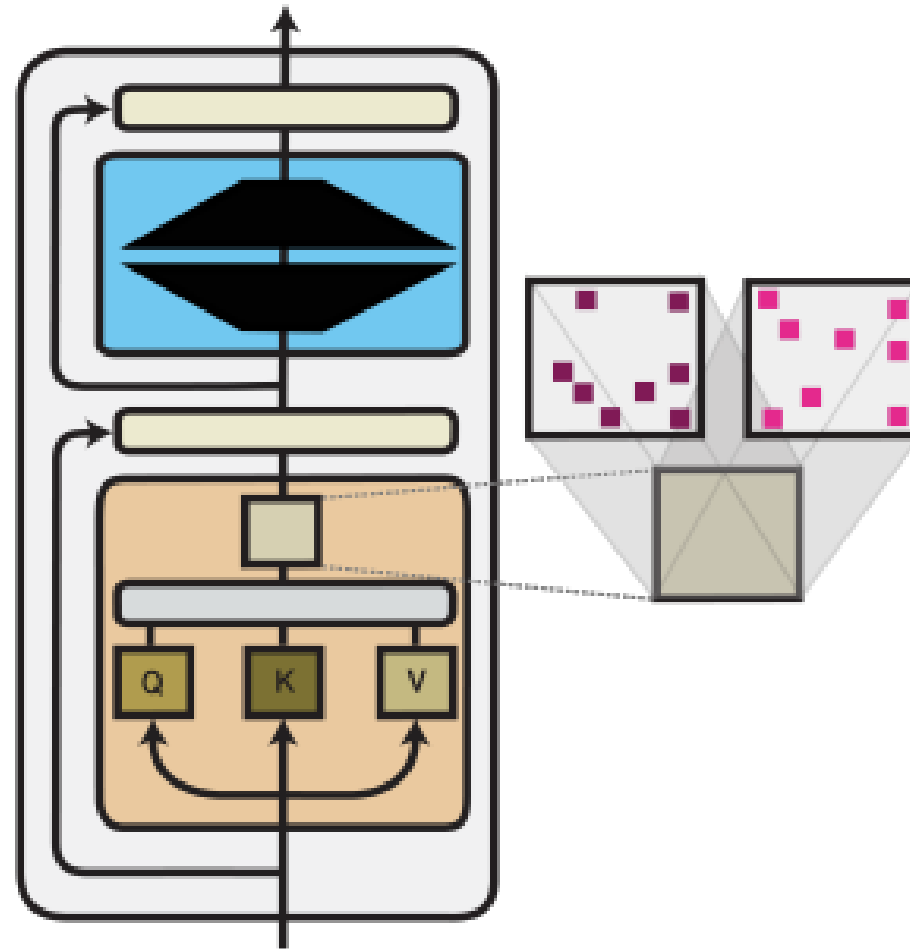
LLM

negative
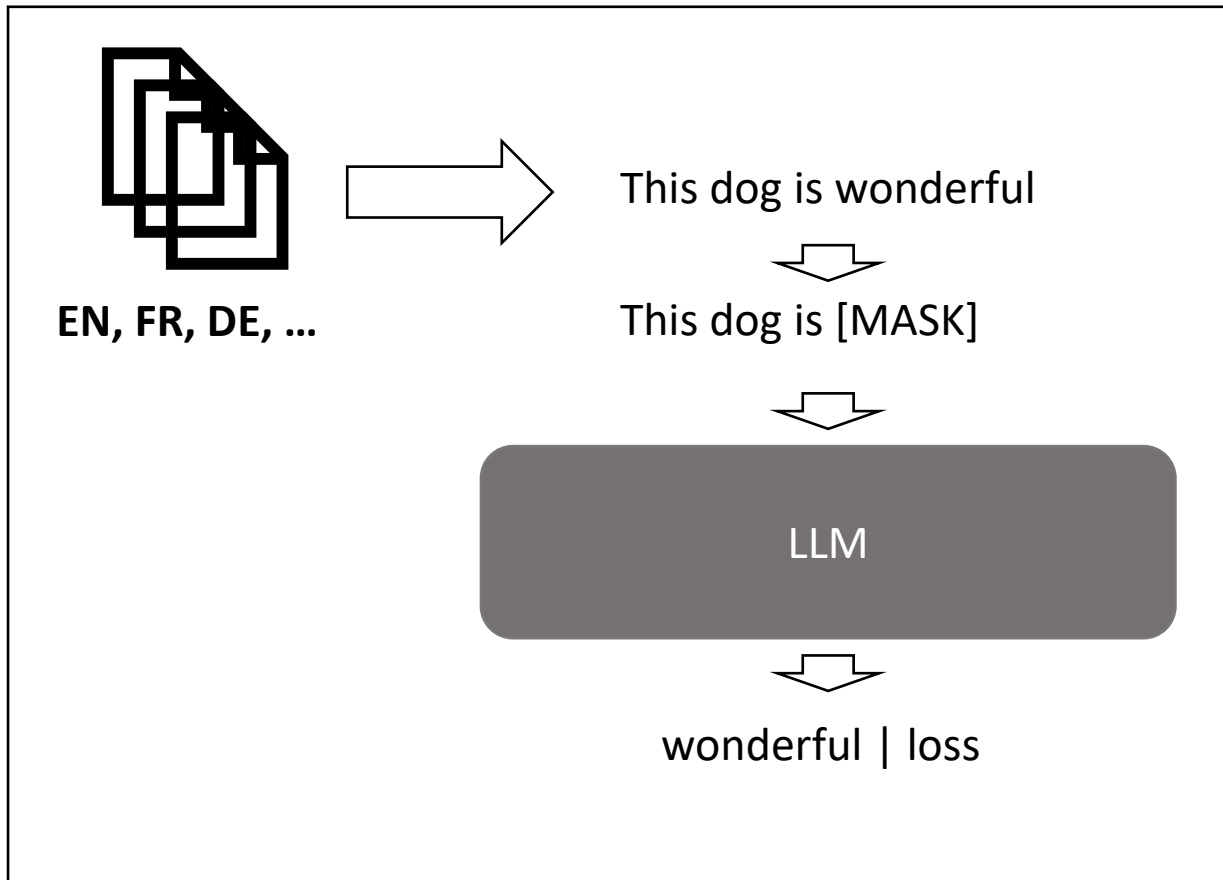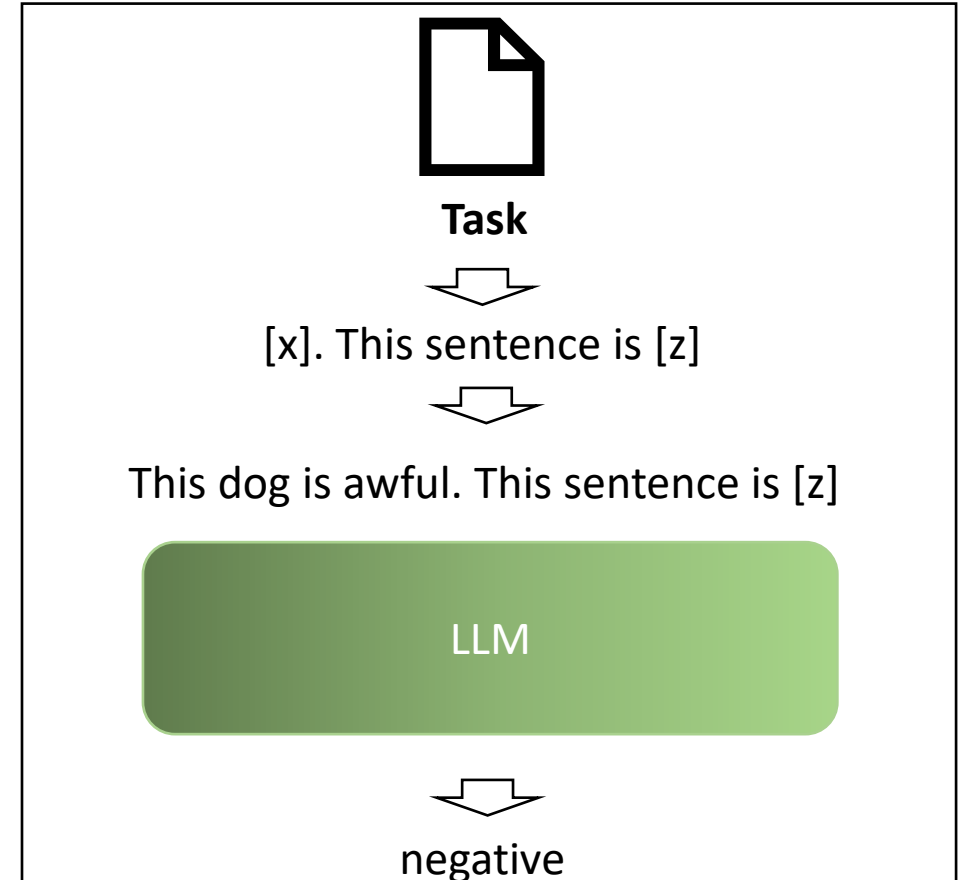
How to adapt models in efficiently?
Subtopics

- **Sparse Fine-Tuning**

- **Low-Rank Adaptation**

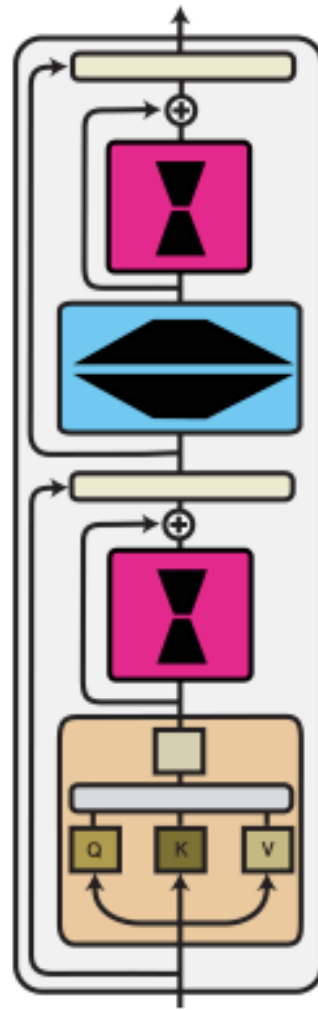- **Adapters**

- **Soft-Prompting**
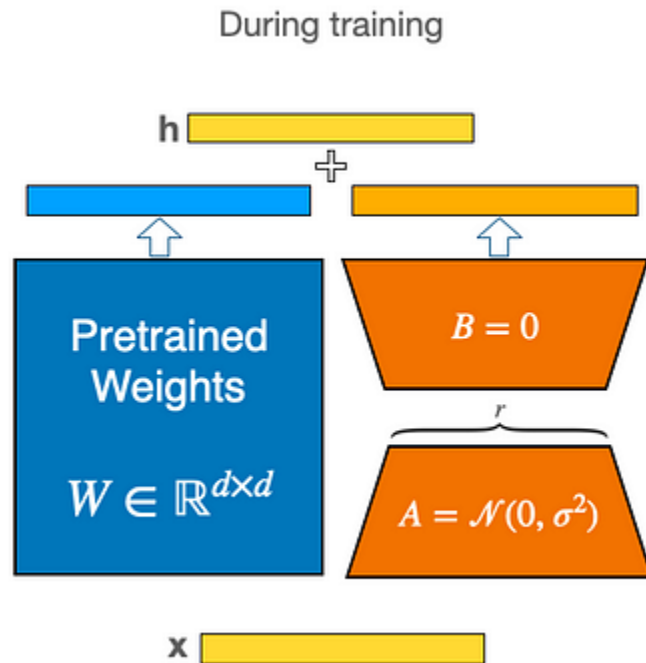
- **Hypernetworks**

Sparse-Fine-Tuning

Source

Adapter

Low-Rank Adaptation

During training

$h = Wx + BAx$

$h = \underbrace{(W + BA)}_{W_{merged}} x$

After training

Merged Weights $W_{merged} \in \mathbb{R}^{d \times d}$

Pretrained Weights $W \in \mathbb{R}^{d \times d}$

$B = 0$

$A = \mathcal{N}(0, \sigma^2)$
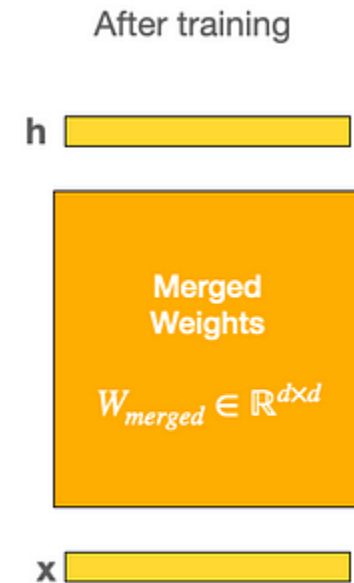
[Source](#)

Soft-Prompting

Hypernetwork

[Source](#)

Organization
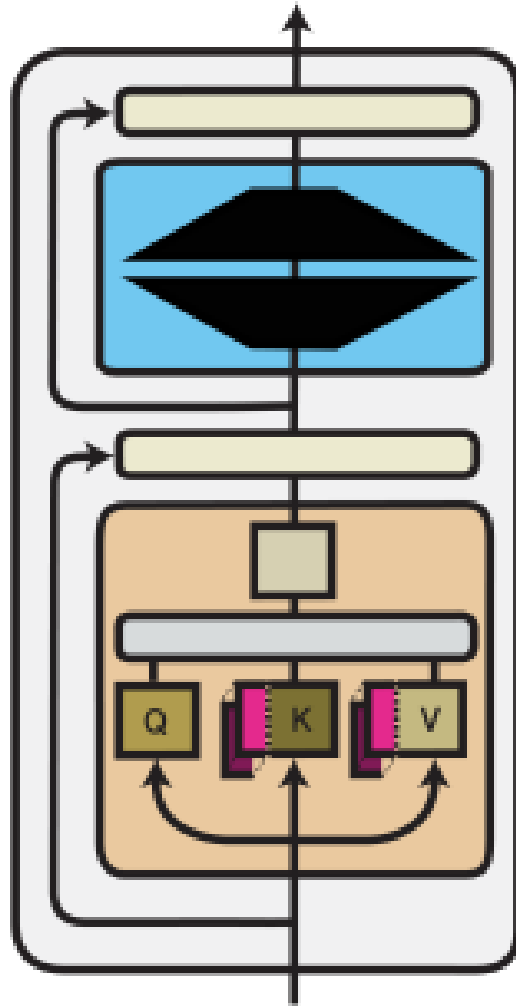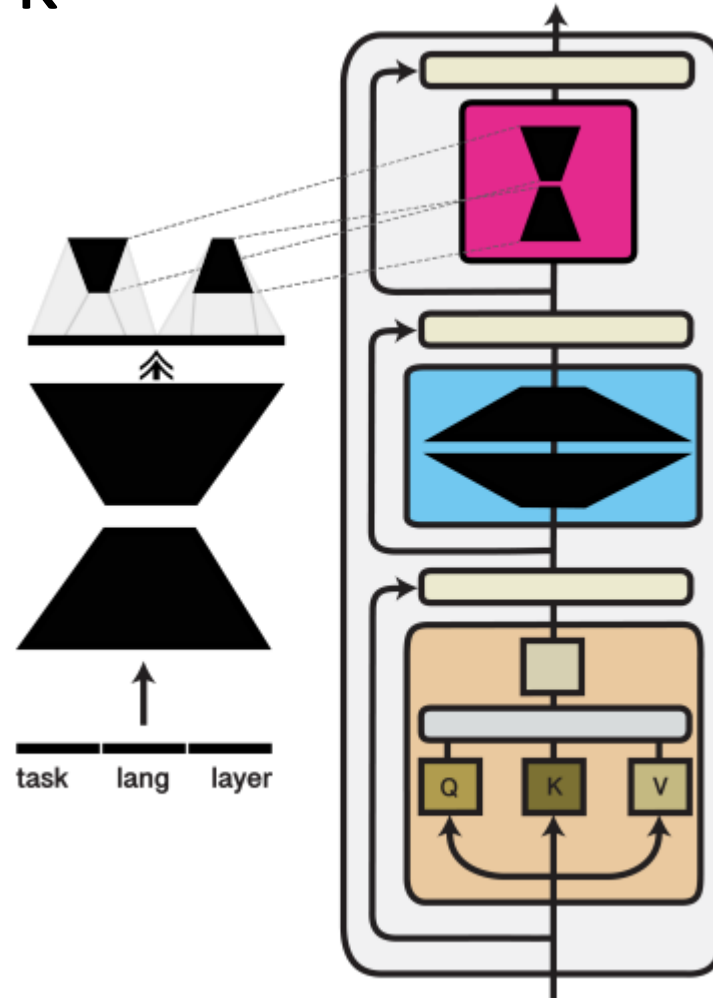
- Assignment to a subtopic
  - Send us your first and second preference via email by 29.04.24 EoD to benedikt.ebing@uni-wuerzburg.de

- Read, understand and explore subtopic

- Organize the collected knowledge for a meaningful presentation about your topic
  - **Mid-July**

- Summarize your topic in a concise report
  - **End of July**

Expectations

- Provided papers are starting points into your topic
  - Explore: e.g., papers cited by or that are cited from the provided papers, survey papers, …
- Summarize your topic including background information
  - Do not "sell" your topic or take statements for granted
  - Be critical and stay objective
- If unsure, ask us!

Presentation

- 15 minutes
- What, why, and how
- 5 minutes Q&A
- Target audience: your fellow students

Report

- [Use LaTex template](#)

- 6 – 8 pages

- Use your own words

- Follow good scientifc practice: e.g.,
  - Cite all related work, properly
  - Mark direct citations (if necessary)

- Target audience: as for the presentation

Grading

- Report and presentation are similarly important

- Do not plagiarize!

Additional Resources

- Survey on modular deep learning: „Pfeiffer, J., Ruder, S., Vulić, I., & Ponti, E. M. (2023). Modular deep learning.
  - https://arxiv.org/pdf/2302.11529.pdf

Send us your first and second subtopic preference via email by **29.04.24 EoD** to
benedikt.ebing@uni-wuerzburg.de