LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Master Seminar

Foundation Models in AI

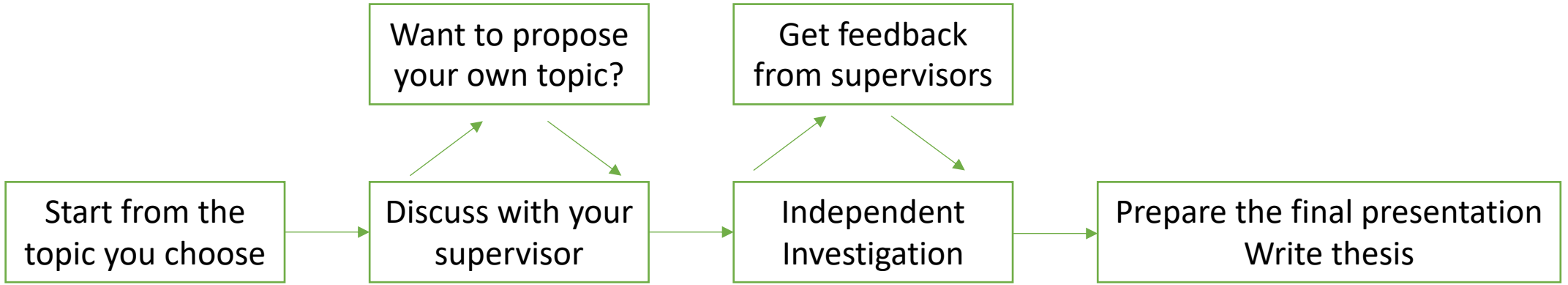
Lehrstuhl für Datenbanksysteme und Data Mining
Prof. Volker Tresp
Gengyuan Zhang

Timeline

Kick-off meeting	19.04.2023 (Mi) 16:00 - 18:00	On-site Room: Oettingenstr. 67 (C) - C 003
Topic Discussion	19.04.2023 – 03.05.2023 (2 weeks)	Hybrid (up to your supervisor)
Investigation	May-June	Online
Final Presentation	Slot 1: 26.07.2023 (Mi) Slot 2: 27.07.2023 (Do) 10:00 - 14:00	On-site (Mandatory Attendance) Room: Oettingenstr. 67 - 067
Thesis Submission	In August	Online

Guidelines

1. Kick-off meeting (today):
   - We will present a list of papers of interests for your choices
   - You will be assigned a supervisor who will take responsibilties in the whole procedure
2. In semester:



3. Final stage:
   - Final presentation
   - Submit your seminar thesis

Guidelines

- Seminar Thesis
  - On the topic you have chosen
  - ~20k Characters (main body only)
  - Deadline:

- Final Presentation
  - 20 min talk + 10 min Q&A
  - Mandatory attendence for everyone

- Key factors to succeed:
  - Follow the checkpoints of each phase
  - Get feedback from your supervisor
  - Make good and wise use of tools: google/ChatGPT etc.
  - Keep an eye of Uni2Work and DBS websites
  - The final thesis **cannot** be generated by ChatGPT or any similar tool

Topics

Power of Unimodal Generative Models *Contact: Ruotong Liao* *liao@dbs.ifi.lmu.de*	GPT-3: Language Models are Few-Shot Learners
	LLaMA: Open and Efficient Foundation Language Models
	DALL-E ： Zero-Shot Text-to-Image Generation
	Stable Diffusion ： High-Resolution Image Synthesis with Latent Diffusion Models
Multimodal learning: Vision Language Models *Contact: Yao Zhang* *yzhang@dbs.ifi.lmu.de*	BLIP: Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation
	SIMVLM: SIMPLE VISUAL LANGUAGE MODEL PRETRAINING WITH WEAK SUPERVISION
	Florence: A New Foundation Model for Computer Vision
	Flamingo: a Visual Language Model for Few-Shot Learning
Prompt Tuning in foundation models *Contact: Gengyuan Zhang* *zhang@dbs.ifi.lmu.de*	Learning to Prompt for Open-Vocabulary Object Detection with Vision-Language Model
	Conditional Prompt Learning for Vision-Language Models
	Align and Prompt: Video-and-Language Pre-training with Entity Prompts
	Visual Prompt Tuning
Reasoning on foundation models Contact: Dr. Jindong Gu Jindong.gu@outlook.com	Socratic Models: Composing Zero-Shot Multimodal Reasoning with Language
	Multimodal chain-of-thought reasoning in language models
	MM-REACT: Prompting ChatGPT for Multimodal Reasoning and Action

Resources

1. How to read a scientific paper
2. Wie halte ich einen Vortrag
3. Recommended tutorials
   - Transformers: https://jalammar.github.io/illustrated-transformer/
   - https://youtu.be/TQQlZhbC5ps

Prioritize your topics!

Q&A