Topics Overview

The schedule is preliminary and subject to changes!

The reading for each *lecture* is given as references to the sources the respective lectures base on. Yo are **not** obliged to read anything. However, you are strongly **encouraged** to read references marked b pin emojis 📌: those are comprehensive overviews on the topics or important works that are beneficia for a better understanding of the key concepts. For the pinned papers, I also specify the pages span fo you to focus on the most important fragments. Some of the sources are also marked with a popcorn emoji 🍿: that is misc material you might want to take a look at: blog posts, GitHub repos, leaderboard etc. (also a couple of LLM-based games). For each of the sources, I also leave my **subjective** estimation of how important this work is for **this specific** topic: from yellow 🟡 *'partially useful'* thoug orange 🟠 *'useful'* to red 🔴 *'crucial findings / thoughts'*. These estimations will be continuously updated as I revise the materials.

For the *labs*, you are provided with practical tutorials that respective lab tasks will mostly derive from. The core tutorials are marked with a writing emoji ✍️; you are **asked** to inspect them **in advance** (better yet: try them out). On lab sessions, we will only **briefly recap** them so it is up to you to prepare in advance to keep up with the lab.

Disclaimer: the reading entries are no proper citations; the bibtex references as well as detailed infos about the authors, publish date etc. can be found under the entry links.

Block 1: Intro

Week 1

22.04. *Lecture*: LLMs as a Form of Intelligence vs LLMs as Statistical Machines

That is an introductory lecture, in which I will briefly introduce the course and we'll have a warming up discussion about different perspectives on LLMs' nature. We will focus on two prominent outlooks: LL is a form of intelligence and LLM is a complex statistical machine. We'll discuss differences of LLMs with human intelligence and the degree to which LLMs exhibit (self-)awareness.

Key points:

- Course introduction
- Different perspectives on the nature of LLMs
- Similarities and differences between human and artificial intelligence
- LLMs' (self-)awareness

Core Reading:

- 📌 The Debate Over Understanding in AI's Large Language Models (pages 1-7), `Santa Fe Institute` 🟠
- Meaning without reference in large language models, `UC Berkeley & DeepMind` 🔴
- Dissociating language and thought in large language models (intro [right after the abstract, see more on the sectioning in this paper at the bottom of page 2], sections 1, 2.3 [*LLMs are predictive …*], 3-5), `The University of Texas at Austin et al.` 🔴

Additional Reading:

- Do Large Language Models Understand Us?, `Google Research` 🟠
- Sparks of Artificial General Intelligence: Early experiments with GPT-4 (chapters 1-8 & 10), `Microsoft Research` 🟡
- On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜 (paragraphs 1, 5, 6.1), `University of Washington et al.` 🟡
- Large Language Models: The Need for Nuance in Current Debates and a Pragmatic Perspective on Understanding, `Leiden Institute of Advanced Computer Science & Leiden University Medical Centre` 🟡

24.04. *Lecture*: LLM & Agent Basics

In this lecture, we'll recap some basics about LLMs and LLM-based agents to make sure we're on the same page.

Key points:

- LLM recap
- Prompting
- Structured output
- Tool calling
- Piping & Planning

Core Reading:

- A Survey of Large Language Models, (sections 1, 2.1, 4.1, 4.2.1, 4.2.3-4.2.4, 4.3, 5.1.1-5.1.3, 5.2.1-5.2.4, 5.3.1, 6) `Renmin University of China et al.` 🔴
- Emergent Abilities of Large Language Models, `Google Research, Stanford, UNC Chapel Hill, DeepMind`
- "We Need Structured Output": Towards User-centered Constraints on Large Language Model Output, `Google Research & Google`
- 📌 Agent Instructs Large Language Models to be General Zero-Shot Reasoners (pages 1-9), `Washington University & UC Berkeley`

Additional Reading:

- Language Models are Few-Shot Learners, `OpenAI`
- Chain-of-Thought Prompting Elicits Reasoning in Large Language Models, `Google Research`
- The Llama 3 Herd of Models, `Meta AI`
- Introducing Structured Outputs in the API, `OpenAI`
- Tool Learning with Large Language Models: A Survey, `Renmin University of China et al.`
- ToolACE: Winning the Points of LLM Function Calling, `Huawei Noah's Ark Lab et al.`
- Toolformer: Language Models Can Teach Themselves to Use Tools, `Meta AI`
- Granite-Function Calling Model: Introducing Function Calling Abilities via Multi-task Learning of Granular Tasks, `IBM Research`
- 🍿 Berkeley Function-Calling Leaderboard, `UC Berkeley` (leaderboard)
- A Survey on Multimodal Large Language Models, `University of Science and Technology of China & Tencent YouTu Lab`

Week 2

29.04. *Lab*: Intro to LangChain

The final introductory session will guide you through the most basic concepts of LangChain for the further practical sessions.

Reading:

- Runnable interface, `LangChain`
- LangChain Expression Language (LCEL), `LangChain`
- Messages, `LangChain`
- Chat models, `LangChain`
- Structured outputs, `LangChain`
- Tools, `LangChain`
- Tool calling, `LangChain`

01.05.

Ausfalltermin

Block 2: Core Topics

Part 1: Business Applications

Week 3

06.05. *Lecture*: Virtual Assistants Pt. 1: Chatbots

The first core topic concerns chatbots. We'll discuss how chatbots are built, how they (should) handle harmful requests and you can tune it for your use case.

Key points:

- LLMs alignment
- Memory
- Prompting & automated prompt generation
- Evaluation

Core Reading:

- 📌 Aligning Large Language Models with Human: A Survey (pages 1-14), `Huawei Noah's Ark Lab`
- Self-Instruct: Aligning Language Models with Self-Generated Instructions, `University of Washington et al.`
- A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications, `Indian Institute of Technology Patna, Stanford & Amazon AI`

Additional Reading:

- Training language models to follow instructions with human feedback, `OpenAI`
- Training a Helpful and Harmless Assistant with Reinforcement Learning from Human Feedback, `Anthropic`
- A Survey on the Memory Mechanism of Large Language Model based Agents, `Renmin University of China & Huawei Noah's Ark Lab`
- Augmenting Language Models with Long-Term Memory, `UC Santa Barbara & Microsoft Research`
- From LLM to Conversational Agent: A Memory Enhanced Architecture with Fine-Tuning of Large Language Models, `Beike Inc.`
- Automatic Prompt Selection for Large Language Models, `Cinnamon AI, Hung Yen University of Technology and Education & Deakin University`
- PromptGen: Automatically Generate Prompts using Generative Models, `Baidu Research`
- Evaluating Large Language Models. A Comprehensive Survey, `Tianjin University`

08.05. *Lab*: Basic LLM-based Chatbot

> On material of session 06.05

In this lab, we'll build a chatbot and try different prompts and settings to see how it affects the output.

Reading:

- ✍️ Build a Chatbot, `LangChain`
- ✍️ LangGraph Quickstart: Build a Basic Chatbot (parts 1, 3), `LangGraph`
- ✍️ How to add summary of the conversation history, `LangGraph`
- Prompt Templates, `LangChain`
- Few-shot prompting, `LangChain`

Week 4

13.05. *Lecture*: Virtual Assistants Pt. 2: RAG

Continuing the first part, the second part will expand scope of chatbot functionality and will teach it to refer to custom knowledge base to retrieve and use user-specific information. Finally, the most widely used deployment methods will be briefly introduced.

Key points:

- General knowledge vs context
- Knowledge indexing, retrieval & ranking
- Retrieval tools
- Agentic RAG

Core Reading:

- 📌 Retrieval Augmented Generation or Long-Context LLMs? A Comprehensive Study and Hybrid Approach (pages 1-7), `Google DeepMind & University of Michigan` 🔴
- A Survey on Retrieval-Augmented Text Generation for Large Language Models (sections 1-7), `York University` 🔴

Additional Reading:

- Don't Do RAG: When Cache-Augmented Generation is All You Need for Knowledge Tasks, `National Chengchi University & Academia Sinica` 🟠
- Self-RAG: Learning to Retrieve, Generate, and Critique through Self-Reflection, `University of Washington, Allen Institute for AI & IBM Research AI`
- Adaptive-RAG: Learning to Adapt Retrieval-Augmented Large Language Models through Question Complexity, `Korea Advanced Institute of Science and Technology`
- Auto-RAG: Autonomous Retrieval-Augmented Generation for Large Language Models, `Chinese Academy of Sciences`
- Querying Databases with Function Calling, `Weaviate, Contextual AI & Morningstar`

15.05. *Lab*: RAG Chatbot

> On material of session 13.05

In this lab, we'll expand the functionality of the chatbot built at the last lab to connect it to user-specifi information.

Reading:

- How to load PDFs, `LangChain`
- Text splitters, `LangChain`
- Embedding models, `LangChain`
- Vector stores, `LangChain`
- Retrievers, `LangChain`
- ✍️ Retrieval augmented generation (RAG), `LangChain`
- ✍️ LangGraph Quickstart: Build a Basic Chatbot (part 2), `LangGraph`
- ✍️ Agentic RAG, `LangGraph`
- Adaptive RAG, `LangGraph`
- Multimodality, `LangChain`

Week 5

20.05. *Lecture*: Virtual Assistants Pt. 3: Multi-agent Environment

This lectures concludes the Virtual Assistants cycle and directs its attention to automating everyday / business operations in a multi-agent environment. We'll look at how agents communicate with each other, how their communication can be guided (both with and without involvement of a human), and this all is used in real applications.

Key points:

- Multi-agent environment
- Human in the loop
- LLMs as evaluators
- Examples of pipelines for business operations

Core Reading:

- 📌 LLM-based Multi-Agent Systems: Techniques and Business Perspectives (pages 1-8), `Shanghai Jiao Tong University & OPPO Research Institute`
- Generative Agents: Interactive Simulacra of Human Behavior, `Stanford, Google Research & DeepMind`

Additional Reading:

- Improving Factuality and Reasoning in Language Models through Multiagent Debate, `MIT & Google Brain`
- Exploring Collaboration Mechanisms for LLM Agents: A Social Psychology View, `Zhejiang University, National University of Singapore & DeepMind`
- AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation, `Microsoft Research et al.`
- 🍿 How real-world businesses are transforming with AI — with more than 140 new stories, `Microsoft` (blog post)
- 🍿 Built with LangGraph, `LangGraph` (website page)
- Plan-Then-Execute: An Empirical Study of User Trust and Team Performance When Using LLM Agents As A Daily Assistant, `Delft University of Technology & The University of Queensland`

22.05. *Lab*: Multi-agent Environment

> On material of session 20.05

This lab will introduce a short walkthrough to creation of a multi-agent environment for automated meeting scheduling and preparation. We will see how the coordinator agent will communicate with twc auxiliary agents to check time availability and prepare an agenda for the meeting.

Reading:

- ✍️ Multi-agent network, `LangGraph`
- ✍️ Human-in-the-loop, `LangGraph`
- Plan-and-Execute, `LangGraph`
- Reflection, `LangGraph`
- ✍️ Multi-agent supervisor, `LangGraph`
- Quick Start, `AutoGen`

Week 6

27.05. *Lecture*: Software Development Pt. 1: Code Generation, Evaluation & Testing

This lectures opens a new lecture mini-cycle dedicated to software development. The first lecture overviews how LLMs are used to generate reliable code and how generated code is tested and improved to deal with the errors.

Key points:

- Code generation & refining
- Automated testing
- Generated code evaluation

Core Reading:

- Large Language Model-Based Agents for Software Engineering: A Survey, `Fudan University, Nanyang Technological University & University of Illinois at Urbana-Champaign`
- 📌 CodeRL: Mastering Code Generation through Pretrained Models and Deep Reinforcement Learning (pages 1-20), `Salesforce Research`
- The ART of LLM Refinement: Ask, Refine, and Trust, `ETH Zurich & Meta AI`

Additional Reading:

- Planning with Large Language Models for Code Generation, `MIT-IBM Watson AI Lab et al.`
- Code Repair with LLMs gives an Exploration-Exploitation Tradeoff, `Cornell, Shanghai Jiao Tong University & University of Toronto`
- ChatUniTest: A Framework for LLM-Based Test Generation, `Zhejiang University & Hangzhou City University`
- TestART: Improving LLM-based Unit Testing via Co-evolution of Automated Generation and Repair Iteration, `Nanjing University & Huawei Cloud Computing Technologies`
- Evaluating Large Language Models Trained on Code, `OpenAI
- 🍿 Code Generation on HumanEval, `OpenAI` (leaderboard)
- CodeJudge: Evaluating Code Generation with Large Language Models, `Huazhong University of Science and Technology & Purdue University`

29.05.

Ausfalltermin

Week 7

03.06. *Lecture*: Software Development Pt. 2: Copilots, LLM-powered Websites

The second and the last lecture of the software development cycle focuses on practical application of LLM code generation, in particular, on widely-used copilots (real-time code generation assistants) and LLM-supported web development.

Key points:

- Copilots & real-time hints
- LLM-powered websites
- LLM-supported deployment
- Further considerations: reliability, sustainability etc.

Core Reading:

- 📌 LLMs in Web Development: Evaluating LLM-Generated PHP Code Unveiling Vulnerabilities and Limitations (pages 1-11), `University of Oslo`
- A Real-World WebAgent with Planning, Long Context Understanding, and Program Synthesis, `Google DeepMind & The University of Tokyo`
- Can ChatGPT replace StackOverflow? A Study on Robustness and Reliability of Large Language Model Code Generation, `UC San Diego`

Additional Reading:

- Design and evaluation of AI copilots – case studies of retail copilot templates, `Microsoft`
- 🍿 Your AI Companion, `Microsoft` (blog post)
- GitHub Copilot, `GitHub` (product page)
- 🍿 Research: quantifying GitHub Copilot's impact on developer productivity and happiness, `GitHub` (blog post)
- 🍿 Cursor: The AI Code Editor, `Cursor` (product page)
- Automated Unit Test Improvement using Large Language Models at Meta, `Meta`
- Human-In-the-Loop Software Development Agents, `Monash University, The University of Melbourne & Atlassian`
- An LLM-based Agent for Reliable Docker Environment Configuration, `Harbin Institute of Technology & ByteDance`
- Learn to Code Sustainably: An Empirical Study on LLM-based Green Code Generation, `TWT GmbH Science & Innovation et al.`
- Enhancing Large Language Models for Secure Code Generation: A Dataset-driven Study on Vulnerability Mitigation, `South China University of Technology & University of Innsbruck`

05.06 *Lab*: LLM-powered Website

> On material of session 03.06

In this lab, we'll have the LLM make a website for us: it will both generate the contents of the website and generate all the code required for rendering, styling and navigation.

Reading:

- see session 22.05
- ✍️ HTML: Creating the content, `MDN`
- ✍️ Getting started with CSS, `MDN`

Week 8: Having Some Rest

10.06.

Ausfalltermin

12.06.

Ausfalltermin

Week 9

17.06. *Pitch*: RAG Chatbot

> On material of session 06.05 and session 13.05

The first pitch will be dedicated to a custom RAG chatbot that the *contractors* (the presenting students, see the infos about Pitches) will have prepared to present. The RAG chatbot will have to be able to retrieve specific information from the given documents (not from the general knowledge!) and use it in its responses. Specific requirements will be released on 22.05.

Reading: see session 06.05, session 08.05, session 13.05, and session 15.05

19.06.

Ausfalltermin

Week 10

24.06. *Pitch*: Handling Customer Requests in a Multi-agent Environment

> On material of session 20.05

In the second pitch, the *contractors* will present their solution to automated handling of customer requests. The solution will have to introduce a multi-agent environment to take off working load from an imagined support team. The solution will have to read and categorize tickets, generate replies and (in case of need) notify the human that their interference is required. Specific requirements will be released on 27.05.

Reading: see session 20.05 and session 22.05

26.06. *Lecture*: Other Business Applications: Game Design, Financial Analysis etc.

This lecture will serve a small break and will briefly go over other business scenarios that the LLMs are used in.

Key points:

- Game design & narrative games
- Financial applications
- Content creation

Additional Reading:

- Player-Driven Emergence in LLM-Driven Game Narrative, `Microsoft Research`
- Generating Converging Narratives for Games with Large Language Models, `U.S. Army Research Laboratory`
- Game Agent Driven by Free-Form Text Command: Using LLM-based Code Generation and Behavior Branch, `University of Tokyo`
- 🍿 AI Dungeon Games, `AI Dungeon` (game catalogue)
- 🍿 AI Town, `Andreessen Horowitz & Convex` (game)
- Introducing NPC-Playground, a 3D playground to interact with LLM-powered NPCs, `HuggingFace` (blog post)
- Blip, `bliporg` (GitHub repo)
- gigax, `GigaxGames` (GitHub repo)
- Large Language Models in Finance: A Survey, `Columbia & New York University`
- FinLlama: Financial Sentiment Classification for Algorithmic Trading Applications, `Imperial College London & MIT`
- Equipping Language Models with Tool Use Capability for Tabular Data Analysis in Finance, `Monash University`
- LLM4EDA: Emerging Progress in Large Language Models for Electronic Design Automation, `Shanghai Jiao Tong University et al.`
- Assisting in Writing Wikipedia-like Articles From Scratch with Large Language Models, `Stanford`
- Large Language Models Can Solve Real-World Planning Rigorously with Formal Verification Tools, `MIT, Harvard University & MIT-IBM Watson AI Lab`

Part 2: Applications in Science

Week 11

01.07. *Lecture*: LLMs in Research: Experiment Planning & Hypothesis Generation

The first lecture dedicated to scientific applications shows how LLMs are used to plan experiments and generate hypothesis to accelerate research.

Key points:

- Experiment planning
- Hypothesis generation
- Predicting possible results

Core Reading:

- 📌 Hypothesis Generation with Large Language Models (pages 1-9), `University of Chicago & Toyota Technological Institute at Chicago`
- 📌 LLMs for Science: Usage for Code Generation and Data Analysis (pages 1-6), `TUM`
- Emergent autonomous scientific research capabilities of large language models, `Carnegie Mellon University`

Additional Reading:

- Improving Scientific Hypothesis Generation with Knowledge Grounded Large Language Models, `University of Virginia`
- Paper Copilot: A Self-Evolving and Efficient LLM System for Personalized Academic Assistance, `University of Illinois at Urbana-Champaign, Carnegie Mellon University & Carleton College`
- SciLitLLM: How to Adapt LLMs for Scientific Literature Understanding, `University of Science and Technology of China & DP Technology`
- Mapping the Increasing Use of LLMs in Scientific Papers, `Stanford`

03.07: *Lab*: Experiment Planning & Hypothesis Generation

> On material of session 01.07

In this lab, we'll practice in facilitating researcher's work with LLMs on the example of a toy scientific research.

Reading: see session 22.05

Week 12

08.07: *Pitch*: Agent for Code Generation

> On material of session 27.05

This pitch will revolve around the *contractors'* implementation of a self-improving code generator. The code generator will have to generate both scripts and test cases for a problem given in the input prompt, run the tests and refine the code if needed. Specific requirements will be released on 17.06.

Reading: see session 27.05 and session 05.06

10.07. *Lecture*: Other Applications in Science: Drug Discovery, Math etc. & Scientific Reliability

The final core topic will mention other scientific applications of LLMs that were not covered in the previous lectures and address the question of reliability of the results obtained with LLMs.

Key points:

- Drug discovery, math & other applications
- Scientific confidence & reliability

Core Reading:

- 📌 Can LLMs replace Neil deGrasse Tyson? Evaluating the Reliability of LLMs as Science Communicators (pages 1-9), `Indian Institute of Technology`

Additional Reading:

- A Comprehensive Survey of Scientific Large Language Models and Their Applications in Scientific Discovery, `University of Illinois at Urbana-Champaign et al.`
- Large Language Models in Drug Discovery and Development: From Disease Mechanisms to Clinical Trials, `Department of Data Science and AI, Monash University et al.`
- LLM-SR: Scientific Equation Discovery via Programming with Large Language Models, `Virginia Tech et al.`
- 🍿 Awesome Scientific Language Models, `yuzhimanhua` (GitHub repo)
- CURIE: Evaluating LLMs On Multitask Scientific Long Context Understanding and Reasoning, `Googl et al.`
- Multiple Choice Questions: Reasoning Makes Large Language Models (LLMs) More Self-Confident Even When They Are Wrong, `Nanjing University of Aeronautics and Astronautics et al.`

Block 3: Wrap-up

Week 13

15.07. *Pitch*: Agent for Web Development

> On material of session 03.06

The *contractors* will present their agent that will have to generate full (minimalistic) websites by a prompt. For each website, the agent will have to generate its own style and a simple menu with workin navigation as well as the contents. Specific requirements will be released on 24.06.

Reading: see session 03.06 and session 05.06

17.07. *Lecture*: Role of AI in Recent Years

The last lecture of the course will turn to societal considerations regarding LLMs and AI in general and will investigate its role and influence on the humanity nowadays.

Key points:

- Studies on influence of AI in the recent years
- Studies on AI integration rate
- Ethical, legal & environmental aspects

Core Reading:

- 📌 Protecting Human Cognition in the Age of AI (pages 1-5), The University of Texas at Austin et al.
- 📌 Artificial intelligence governance: Ethical considerations and implications for social responsibilit (pages 1-12), `University of Malta`

Additional Reading:

- Augmenting Minds or Automating Skills: The Differential Role of Human Capital in Generative AI's Impact on Creative Tasks, `Tsinghua University & Wuhan University of Technology`
- Human Creativity in the Age of LLMs: Randomized Experiments on Divergent and Convergent Thinking, `University of Toronto`
- Empirical evidence of Large Language Model's influence on human spoken communication, `Max-Planck Institute for Human Development`
- 🍿 The 2025 AI Index Report: Top Takeaways, `Stanford`
- Growing Up: Navigating Generative AI's Early Years – AI Adoption Report: Executive Summary, `AI a Wharton`
- Ethical Implications of AI in Data Collection: Balancing Innovation with Privacy, `AI Data Chronicles`
- Legal and ethical implications of AI-based crowd analysis: the AI Act and beyond, `Vrije Universiteit`
- A Survey of Sustainability in Large Language Models: Applications, Economics, and Challenges, `Cleveland State University et al.`

Week 14

22.07. *Pitch*: LLM-based Research Assistant

> On material of session 01.07

The last pitch will introduce an agent that will have to plan the research, generate hypotheses, find the literature etc. for a given scientific problem. It will then have to introduce its results in form of a TODO or a guide for the researcher to start off of. Specific requirements will be released on 01.07.

Reading: see session 01.07 and session 03.07

24.07. *Debate*: Role of AI in Recent Years + Wrap-up

> On material of session 17.07

The course will be concluded by the final debates, after which a short Q&A session will be held.

Debate topics:

- LLM Behavior: Evidence of Awareness or Illusion of Understanding?
- Should We Limit the Usage of AI?

Reading: see session 17.07