



## PhD Program: Doctoral Student Gen AI Lab (Block Course):

Explore the responsible use of AI in generating scientific texts, images, audio and code for scientific purposes (4 days, 1 ECTS)

<b>Lecturers:</b>	Melanie Paschke, Jeanine Reutemann and other experts
<b>Location:</b>	ETH Zürich Center, tbd
<b>Dates:</b>	session 1: Nov 12 (Room tbc), session 2: Nov 13 (PB lab, RZ D8), session 3: Nov 18, session 4: 19 (Room tbc), 2024
<b>Time:</b>	9:00 – 17:00
<b>Credit Points:</b>	1 ECTS
<b>Nr. Participants:</b>	40

### Course Contents:

Interested in using Generative AI in your scientific work processes but in responsible and ethical responsive way? This block course for PhD students allows experimenting with generative AI to generate texts, images and audio that can be used in science from scientific presentations to publications. PhD students are invited to experiment together in a problem-based setting around concern and critical topics when using AI-based tools with research data, private data and for output generation as images, text, short videos, audio or code. Invited experts will show their knowledge for example generation of scientific illustrations or customization of AI models in several hands-on workshops.

**Targeted group:** PhD students of ETH Zurich, University of Zurich and University of Basel.

**Individual Performance and Assessment:** Ungraded semester performance. Students will actively participate during the course with group presentations at the end of each block (28 hours face-to-face) Additionally, a documentation needs to be handed in (2 hours).

### Sessions:

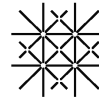
#### Day 1: Ethical Prompting Lab

Generative AI and large language models like ChatGPT have democratized interaction with AI and opened new avenues for information retrieval and management that students should be able to supervise in a meaningful, responsible and professional way. Outputs generated with AI often amplify social biases, stereotypes or inaccuracies through their filtering of training data: for example, favoring certain groups or ideas, perpetuating stereotypes, or making incorrect assumptions based on learned patterns. Through ethical prompting these shortcuts can be partly overcome together with system-based thresholds. It is in the responsibility of the PhD scientist to uphold truth, fairness, and respect for all through responsible prompting techniques and carefully evaluation of the outputs.

The day will be structured by short theoretical input on prompt engineering following the guidance of the **PROMPT AI Framework for Informed AI Use** (<https://www.turningdataintowisdom.com/responsible-ai-prompting/>): Precise, Responsible, Personalized, Culturally Sensitive, Explainable, Conversational, Troubleshooting. We will discuss the output evaluation of genAI.

Participants will be able to learn:

- To generate responsible prompts that will follow for example these principles:



- Formulating questions that are neutral and open-ended.
- Seeking objective, balanced, and different perspectives on controversial topics.
- Using specific, and non-judgemental phrasing, i.e. avoiding ambiguous or partisan language.
- Avoiding queries that accelerate stereotypes of gender, race and age.
- Learning to craft culturally sensitive prompts.
- Establishing a set of ethical rules in the prompting protocol and make the AI follow them.
- Flagging unethical content.
- Fact-checking against reliable contents.
- Withholding problematic details that could make the AI respond unethically.

### Day 2: Responsible creation of images and scientific visualization

On day 2, students are introduced into the potential and fallacies of image-based AI tools respecting originality, transparency, copyright. Further, image generation techniques for scientific visualizations are contextualized into the science historical perspective of image manipulation techniques. The applications of synthetic audiovisual media such as “deepfakes” in the realm of academic work will also be critically reflected.

How to create reproducible and accurate scientific visualizations from scientific data using e.g. stable diffusion models? What are the potentials and limitations of using diffusion models for scientific visualizations? How to prompt for original ideas and how to change the images through human expertise and own style by using creative but controlled prompting? Students will practice sketch-to-image, collage-to-image, inpainting methods to explore potential use of these tools for their own projects.

Participants will be able to:

- reflect on manipulation techniques of present and past scientific visualizations techniques.
- get introduced into synthetic audiovisual data such as deepfakes.
- explore generative AI for scientific visualizations.
- use visualization techniques in co-setup with AI-tools to create visual abstracts, thumbnails or graphical abstracts for peer-reviewed publication.

### Day 3: Customized GPTs for AI-supported scientific workflows

In the third prompting lab participants will explore prompts that can simplify, or automate certain steps in scientific workflows, from data presentation to writing. Especially we will look into customizing GPTs and bot programming for individual and automated workflows. This prompting lab should explore innovative use cases for using customized AI in the scientific working process. We are aiming in this lab also to get consistent and robust prompts for high-quality and replicable outputs, thus, will teach students a set of tests and remediation steps for testing prompts or GPTs also for their different correctness and biases.

Participants will be able to learn:

- understand how customized GPTs will help with automated workflows.
- practice collaborative workflows with genAI.
- establish their role as human in the loop.

### Day 4: Protecting your privacy

Awareness what you give away when you share (your) data in AI-based environments. What is the private, sensitive or sensible information that in different AI-based tools and instances users are leaking? We exchange on techniques and processes to avoid leakage of private or non-copyrighted but original and personally owned material, from privacy-protecting measures in institutional environments to techniques that protect images and information to be used by gen AI to new forms of depersonalizing or protecting of data with watermarks to make the origin and the copyrights transparent. In this session, we will also create awareness and sensitivity for the avoidance and protection from deep fakes from personalized data.



University of  
Zurich <sup>UZH</sup>

**ETH** zürich



University  
of Basel

Zurich-Basel Plant Science Center

---

Participants will be able to learn:

- Understand how private, sensitive or sensible information is regulated for AI use.
- Reflect on ETH compliance guidelines: how are data and information protected in general and in gen AI environments?
- Know about the risks when private, sensitive or sensible information is leaked, e.g. data misuse as deepfakes and other.
- Know about techniques of data anonymization, pseudonymization, anonymization etc. for personal data and their limitations.
- Get familiar with technologies that should help to protect personal or copyrighted data and apply them to own data.