ETH zürich

ETH AI CENTER

Welcome to **Data Science Lab** Spring 2025 Course

Learning Objectives

The objective of this course is to **gain hands-on experience** of dealing with data science and machine learning applications **“in the wild"**. Participants work in **teams of two to three** students on projects and get **go through the full process** starting from data cleaning, modeling, execution, debugging, error analysis, and quality/performance refinement.

After this lecture, you will be able to…

- …analyze, derive, and align project goals, expectations, objectives, and KPIs from stakeholder needs
- …apply adequate data cleaning and pre-processing techniques for a real-world problem, including handling missing data, labelling, outliers, feature engineering, scaling, and normalization
- …design, develop, and evaluate machine learning models together as a team
- …debug and tune the performance of the built model, including outlier analysis
- …communicate and present the results of machine learning projects effectively, both orally and in written reports, to technical and non-technical stakeholders

Important dates and resources

Kick-off (Lecture) – Feb 18 @ 13:00-15:00
Welcome, project presentations
Where: Zoom

Submit Project preferences – Feb 25 @ 23:59
Challenge groups assigned – Feb 26
Students voice their individual challenge preferences
Where: Google Drive Excel sheet, bookmarked in Slack channel

Project Setup Lecture – Mar 4 @ 13:00-15:00
Where: ETH AI Center, OAT X 11

Check-in #1 – Mar 18 @ 13:00-15:00
Review of project scoping and setup
Where: Zoom

Check-in #2 – Apr 15 @ 13:00-15:00
Half-time progress update sharing
Where: ETH AI Center, OAT X 11

Final poster/networking session – May 27 @ 13:00-15:00
Groups present their projects & results
Where: ETH AI Center, OAT X 11
Note: Written report must be submitted by the end of the semester (May 30)

Course material and additional resources

Course material (schedule, lecture slides, recordings etc) will be made available through Slack/mail.

https://eth.slack.com
Slack channel set up with enrolled students, challenge givers, and INFK coaches.

Additional resources: in case you need additional compute or other resources, please align with your team coach

Course contact

In case you have any questions, please reach out to Junling Wang via Slack (preferred) or via email junling.wang@ai.ethz.ch

Project setup & grading

- Groups of **two to three students** work on one assigned case
- Project size: ~840 hours => 2-3 people => **~2-3 days per week per person**
- Role of **Academic Coach**: Overall scientific, technical, and team coaching
- Role of **Challenge Giver**: Provides dataset & assumes role of the customer/domain expert and can help with clarifications on the data or relevance of certain performance levels
- Grading: pass/fail with key components as follows
  - Project report submitted on time
  - Regular interactions with Academic Coach (2-4x per month) & Challenge Giver (1x per month)
  - Positive feedback from challenge mentor on stakeholder & change management
  - Clarity of project setup and organization with clear roles & responsibilities
  - Quality of presentation & project results
- Don’t forget: Outline in the report how each team member contributed
- Bonus: Participation & support of other teams in the Slack channel

Team behind the Data Science Lab Spring 2025

PD Dr. Alexander Ilic
Executive Director, ETH AI Center

Lead Lecturer

Prof. Dr. Valentina Boeva
Computational Cancer Genomics

Lecturer & Science Coach

Prof. Dr. Fanny Yang
Statistical Machine Learning

Lecturer & Science Coach

Prof. Dr. Julia Vogt
Medical Data Science

Lecturer & Science Coach

Dr. Arnout Devos
Program Manager, ETH AI Center

Lecturer & Course Coordinator

Junling Wang
PhD Fellow, ETH AI Center

Teaching Assistant

Dr. Clara Meister
MAS Data & AI coordinator, ETH CS

Course Coordinator

Prof. Dr. Elliott Ash
NLP & Human-AI Alignment

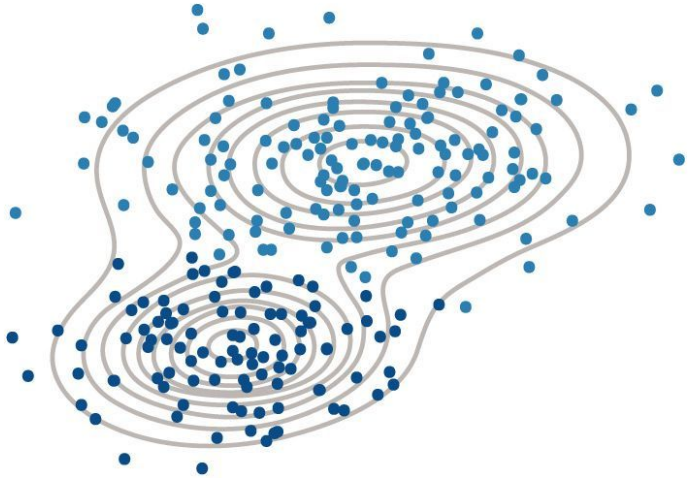
Lecturer & Science Coach

Project overview and presentation order

	Data Science Lab Kickoff online 18.02.25, 13:00-15:00	https://ethz.zoom.us/j/62701133695		
Project ID	**Topic**	**Speaker(s)**	**Time start**	**Time end**
	Welcome/Intro	Alex Ilic, Arnout Devos, Junling Wang	13:00	13:05
1	Clustering of Exam Problem Solutions	Gerd Kortemeyer	13:05	13:10
2	Text-to-Image model fine-tuning using real human preference data.	Mads Kuhlmann-Joergensen	13:10	13:15
3	Agent Simulation for Multi-Turn Survey and Evaluation	Syang Zhou	13:15	13:20
4	Developing Next-Gen 3D Human Pose Estimators	Manuel Kaufmann	13:20	13:25
5	Automated damage detection on new aircraft engine parts	Sebastian Graulich	13:25	13:30
6,7,8	Swiss AI Foundation Models	Imanol Schlag	13:30	13:35
10	Automated Research Paper Generation with LLM	Ahmed Nassar	13:40	13:45
11	Explainable AI for Detecting Synthetic Content in Phishing Emails	Lorin Schöni	13:45	13:50
12	Inference Engine for the Internet Computer	Islam El-Ashi	13:50	13:55
13	Automatic measurement of garments for a fashion retailer	Adrián González Sieira	13:55	14:00
14	Automated quality verification model generation for manual assembly	Jonas Conrad	14:00	14:05
15	Learning to Optimize for Petrol Station Replenishment Problem	Antonia Unger	14:05	14:10
16	NeRF for Ionospheric Modelling	Arno Rüegg	14:10	14:15
17	Automation Support for Corporate Reporting	Orlando Monsalve Rueda	14:15	14:20
19	Benchmarking of (privacy-preserving) Tabular Synthetic Data Generation	David Froelicher	14:25	14:30
20	Enhancing Freelancer Ratings with Pairwise Comparisons	Elliott Ash	14:30	14:35
	Questions and Answers	Junling Wang	14:35	15:00

Slide 1: Project Overview

Project ID: 1
Title: Clustering of Exam Problem Solutions
Challenge Giver: Dr. Gerd Kortemeyer, ETH AI Center & ETH Rectorate
Academic Coach: Dr. Gerd Kortemeyer
Contact Details: kgerd@ethz.ch



Overview:

- Aim: Assist in automating the grading process for STEM exams by clustering student-submitted physics- and math-exam solutions based on similarity.
- Goal: Develop an effective clustering algorithm that groups similar handwritten solutions, ensuring they receive comparable point scores.
- Tools: ADA-3 embeddings of LaTeX transcriptions, machine learning clustering techniques.
- Focus: Physics exam problem solutions consisting of mathematical derivations, sketches, calculations, and explanations.
- Support: ETH Zurich provides depersonalized exam data from 252 students with informed consent.

Data Science Lab S2025
ETH AI CENTER

Slide 2: Dataset & Challenge

Project ID: 1
Title: Clustering of Exam Problem Solutions
Challenge Giver: Dr. Gerd Kortemeyer, ETH AI Center & ETH Rectorate

Dataset:

- **Sources:** Data obtained from actual ETH exams with informed consent. Available as handwritten originals (PDF/PNG) and LaTeX transcriptions
- **Content:** Depersonalized submissions from 252 students covering four physics exam questions, including grading ground truth.
  Possible challenge in the end: new unknown math data set with new ground truth

Challenge Components:

1. Ensuring effective clustering that aligns with grading ground truth.
2. Overcoming small and ineffective clusters seen in prior ADA-3 embeddings.
3. Enhancing existing clustering methodologies to improve grading efficiency.
4. Addressing challenges of handwritten mathematical expressions in clustering.

Slide 3: Evaluation & Success Criteria

Project ID: 1
Title: Clustering of Exam Problem Solutions
Challenge Giver: Dr. Gerd Kortemeyer, ETH AI Center & ETH Rectorate

Evaluation Metrics:

1. Accuracy of clustering based on similarity of student solutions.
2. Alignment of clusters with grading ground truth.
3. Improvement over previous ADA-3 embedding results.
4. Scalability and generalizability of the clustering approach.

Successful Project:

- Produces clusters that effectively group solutions with similar point scores.
- Demonstrates clear improvements over previous methods.
- Provides insights into grading automation for STEM exams.
- Has potential for open-source contribution and academic publication.

Slide 1: Project Overview

Project ID: 2
Title:Text-to-Image model fine-tuning using real human preference data.
Challenge Giver: Mads Kuhlmann-Joergensen, Rapidata
Academic Coach: Elliott Ash
Contact Details: mads@rapidata.ai

Overview:

- Aim: Train a text-to-image model showcasing the use of Rapidata's technology in sourcing the training data.
- Goal: Fine-tune/post-train one or more text-to-image models for practical commercial use in a targeted field. (Stretch goal: deploy model for public use)
- Tools: RLHF, DPO, PyTorch, Rapidata API
- Focus: research, engineering, model evaluation
- Support: Collection of preference data, guidance from robotics/AI engineers

Slide 2: Dataset & Challenge

Title:Text-to-Image model fine-tuning using real human preference data.
Challenge Giver: Mads Kuhlmann-Joergensen, Rapidata

Dataset:

- **Sources:** Expected to be generated using selected text-to-image model as natural part of the project.
- **Content:** Generated images relevant to the selected target.

Challenge Components:

1. Identify or define a commercially promising application, such as logo creation, web graphics, illustrations, interior design, etc.
2. Define and validate metrics to evaluate model improvements. Use Rapidata's technology to source large-scale human feedback and drive iterative optimization.
3. Set up and execute a robust pipeline to train or fine-tune the model, integrating human preference data collected at scale through Rapidata's platform.

Slide 3: Evaluation & Success Criteria

Title:Text-to-Image model fine-tuning using real human preference data.
Challenge Giver: Mads Kuhlmann-Joergensen, Rapidata

Evaluation Metrics:

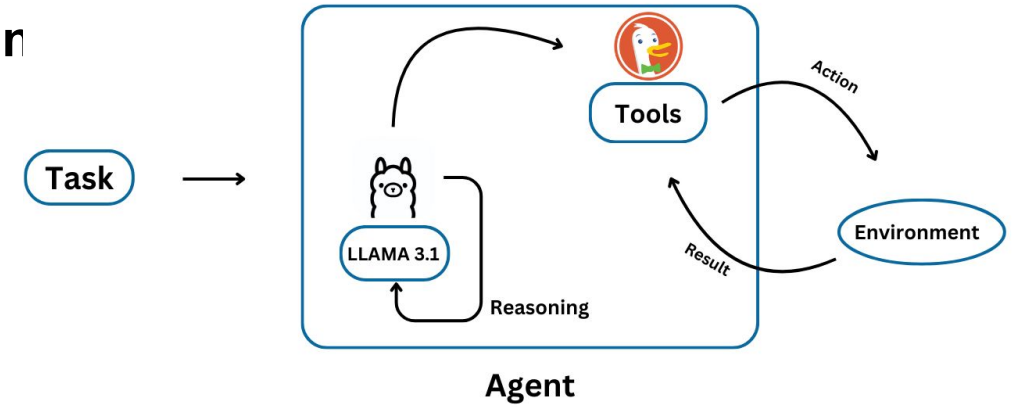
1. **Preference rating** versus existing off-the-shelf models
2. **Usage/adoption** (stretch goal)

Successful Project:

- Trains a text-to-image model based on preference data sourced through Rapidata's technology
- Evaluates model, showing significant quality improvement over off-the-shelf models in the selected area
- (Deploys model, e.g. on HF or Rapidata website)

Slide 1: Project Overview

Project ID: 3
Title: Agent Simulation for Multi-Turn Survey and Evaluation
Challenge Giver: Shijing Cai, Dr. Syang Zhou, Calvin Risk AG
Academic Coach: Prof. Dr. Elliot Ash
Contact Details: sc@calvin-risk.com, sz@calvin-risk.com



Overview:

- LLMs can be deployed as autonomous agents to tackle complex tasks and decision-making.
- Chat-style LLMs serve as a simple agent baseline for interactive dialogue.
- Key challenge: generating diverse, realistic test data sets for reliable agent evaluation.

Challenge: Create a model capable of generating realistic multi-turn human-chatbot interactions

Slide 2: Project Details

Dataset:

- **Sources:** datasets.xlsx Please feel free to find other open source datasets.
- **Content:** Various human-chatbot or human-human interactions, which should be used as training and te

  Calvin risk will support up to 3 000 CHF for azure computing or LLM tokens.

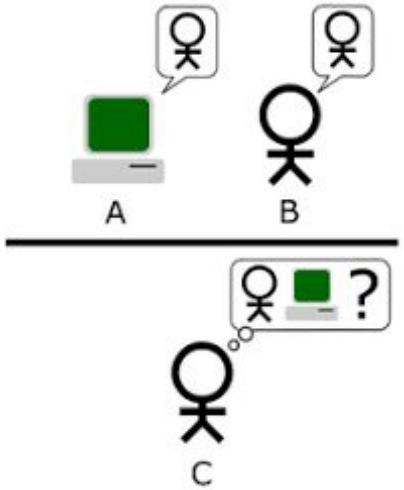
Key Components:

- **Seed interactions:** Investigate if providing seed interactions (samples of real human chatbot interactior help to generate realistic domain specific multi-turn conversations.
- **Model Comparison: Open-Source vs Proprietary:** Investigate if open-source models (e.g. LLama, Mi can match or outperform fine tuned proprietary models in generating realistic human-chatbot interaction prompt engineering etc.]
- **(Bonus) Persona-Based Conversations:** Define diverse user personas (e.g., varying socio-demograp goals) to simulate realistic interactions.

[1] LLM Roleplay: Simulating Human-Chatbot Interaction
[2] BotChat: Evaluating LLMs' Capabilities of Having Multi-Turn Dialogues

Slide 3: Evaluation & Success Criteria

Evaluation Metrics (from BotChat):



- **UniEval:** Uses LLMs as evaluators, scoring dialogues based on coherence, engagement, and relevance.
- **PairEval:** Side-by-side comparison of model outputs to determine which response is superior in realism and flow.
- **GTEval:** Focuses on goal-oriented evaluation, checking if the chatbot maintains consistency and helps users achieve their objectives.

Successful Project:

- Replication of BotChat Results – Successfully reproduce the findings from the paper on human-chatbot interactions using real-world data.
- Strong Performance on Test Dataset – Evaluate and compare open-source vs. closed-source models, analyzing their ability to generate realistic, persona-driven conversations.
- Effective Performance on RAG Bot (Calvin Risk) – Test data generation models on a Retrieval-Augmented Generation (RAG) chatbot provided by Calvin Risk to assess if the generated data is realistic (given metrics above).

Slide 1: Project Overview

Project ID: 4
Title: Developing Next-Gen 3D Human Pose Estimators
Challenge Giver: Dr. Manuel Kaufmann, ETH AI Center
Academic Coach: Dr. Manuel Kaufmann
Contact Details: manuel.kaufmann@ai.ethz.ch

From: https://microsoft.github.io/SynthMoCap/

Overview:

- Aim: Investigate quality of recent 3D human pose estimation (HPE) models
- Goal: Implement and evaluate state-of-the-art models and provide ready-to-use code
- Tools: Python, pytorch, SMPL body model
- Focus: checking reproducibility, coding / engineering, research possible
- Support: Academic coach and team members of the AIT lab.

Slide 2: Dataset & Challenge

Title: Developing Next-Gen 3D Human Pose Estimators
Challenge Giver: Dr. Manuel Kaufmann, ETH AI Center

Dataset:

- **Sources:** Synthetic data provided by the paper Look Ma, No Markers, specifically
- **Content:**
  - SynthBody: RGB images of humans with 2D/3D annotations
  - SynthHand: RGB images of (left) human hands with 2D/3D annotations

Challenge Components:

1. **Implementation:** Follow the above paper to re-implement their proposed architecture.
2. **Evaluation:** Evaluate the implementation on established benchmarks and new datasets (multi-view data) provided by the challenge giver.
3. **Engineering:** Provide the final model as an easy-to-run Python package.

Slide 3: Evaluation & Success Criteria

Title: Developing Next-Gen 3D Human Pose Estimators
Challenge Giver: Dr. Manuel Kaufmann, ETH AI Center

Evaluation Metrics:

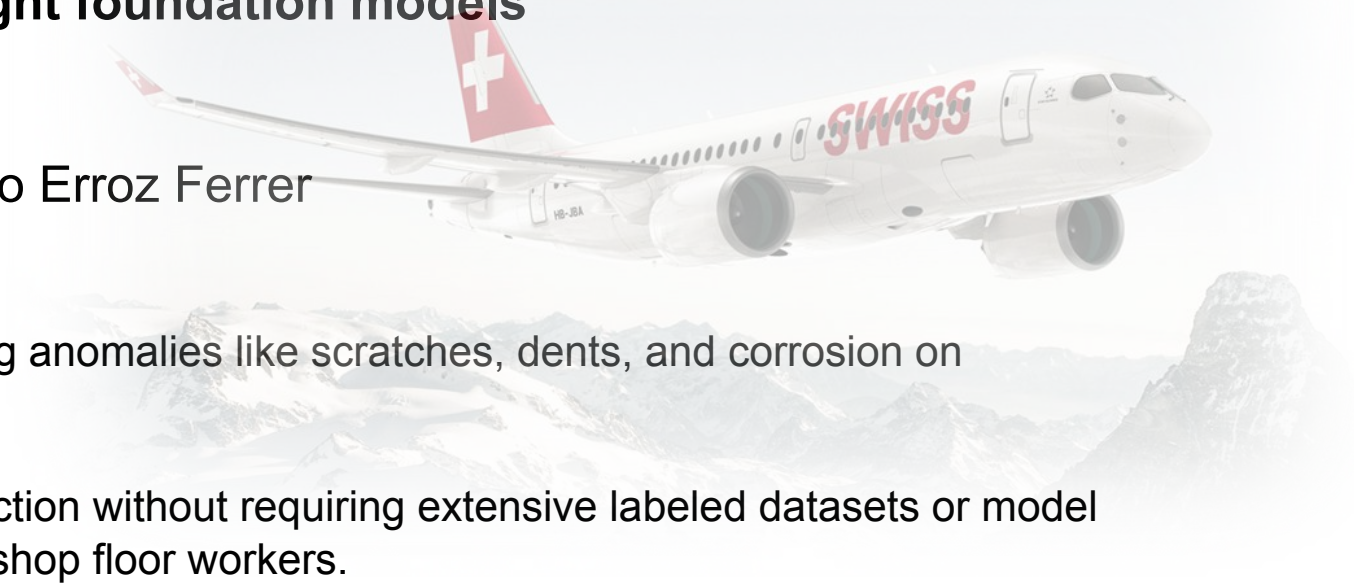
1. **MPJPE / MPVPE** on benchmarks used by the paper
2. **MPJPE / MPVPE** on new benchmarks (EMDB, internal multi-view data)
3. **Speed** of implementation

Successful Project:

- The metrics are similar to what the original paper reported.
- We know the accuracy of the method on our own data as a function of number of views.
- The final code can easily be run by a research engineer unfamiliar with the project.

Slide 1: Project Overview [Project ID: 5]

Title: Automated damage detection on new aircraft engine parts using zero-shot anomaly segmentation and open-source / open-weight foundation models
Challenge Giver: Lufthansa Technik
Academic Coach:
Contact Details: Sebastian Graulich, Santiago Erroz Ferrer
Overview:



- **Aim**: Develop a deployable pipeline for detecting anomalies like scratches, dents, and corrosion on high-resolution images of new engine parts.
- **Goal**: Enable reliable, automated anomaly detection without requiring extensive labeled datasets or model fine-tuning, ensuring usability for non-technical shop floor workers.
- **Tools:** >**Open-source models:** Segment Anything Model (SAM), YOLO, Grounding DINO >**Techniques:** Zero-shot anomaly segmentation, hybrid prompt regularization, few-shot learning (result should be commercially usable)
- **Focus**: Efficient detection of tiny anomalies on high-value parts using high-resolution images while leveraging the zero-shot capabilities of foundation models to generalize across unseen defect types.
- **Support**: Data Scientist and Project Engineer from Lufthansa Technik

Slide 2: Dataset & Challenge [Project ID: 5]

Title: Automated damage detection on new aircraft engine parts using zero-shot anomaly segmentation and open-source / open-weight foundation models
Challenge Giver: Lufthansa Technik

Dataset:

- **Sources:** Magnetic Tile Defect, GC10-DET, Due to the nature of the zero/few-shot approach, these data sets should be used for testing only. More highly realistic test images can be generated e.g. using Flux 1.1 Pro etc. (see sample images).
- **Content:** Images of shiny metal parts with small anomalies like thin scratches or dents. Sample prompt: “Aluminum part with one light scratch IMG_1018.CR2”

Challenge Components:

- **Few/Zero-Shot Learning Application:** Investigate the use of few/zero-shot anomaly detection to identify defects like scratches and dents with no/minimal labeled data, leveraging open-source models.
- **High-Resolution Image Analysis:** Develop methods to process and analyze high-resolution images efficiently for precise anomaly localization.
- **Scalability and Generalization:** Ensure the solution can generalize across diverse parts without requiring extensive retraining for each part.
- **User-Friendly Output:** Design outputs that are intuitive for shop floor workers with no IT expertise, focusing on clear defect localization.

Slide 3: Evaluation & Success Criteria [Project ID: 5]

Title: Automated damage detection on new aircraft engine parts using zero-shot anomaly segmentation and open-source / open-weight foundation models
Challenge Giver: Lufthansa Technik

Evaluation Metrics (proposal only):

- Benchmark against aircraft-specific metrics from MVTec AD/VisA
- AUROC+AUPRO jointly validate detection thoroughness at pixel/region levels
- FPR95+mFPPI balance sensitivity with operational practicality in high-stakes environments.
- VRAM Memory footprint

Successful Project:

- Can segment various types and number of damages and create corresponding bounding boxes.
- Solution should be a deployable pipeline for further testing on real engine part photos.
- Solution should be able to dissect large images into smaller square images for processing in order to optimize for memory requirements and accuracy.

Slide 1: Project Overview [Project ID: 6]

Title: Large Scale Cluster Health Utility Tool
Challenge Giver: Dr. Jonathan Coles, CSCS
Academic Coach: Dr. Imanol Schlag, ETHZ AI Center
Contact Details: jonathan.coles@cscs.ch ischlag@ethz.ch
Overview:

- **Aim**: Develop efficient methods for extracting and visualizing hardware metrics
- **Goal**: Create scalable tools that enable researchers and engineers to effectively process, analyze and visualize large-scale computational data
- **Tools**: nvidia dcgm exporter, CSCS provided tools and services

Slide 2: Dataset & Challenge [Project ID: 6]

Title: Supercomputer Data Extraction and Visualisation at Scale
Challenge Giver: Dr. Imanol Schlag, ETHZ AI Center + Dr. Jonathan Coles, CSCS

Dataset:

- **Data sources:** Hardware metrics directly extracted from the compute nodes.
- **Content:** Data consists of various measures like GPU processor/memory utilisation, temperature, bandwidth, etc.

Challenge Components:

1. **Data Collection:** Extract the information efficiently from compute nodes
2. **Storage:** Store data to allow for “playback” of a selected time window and partition of nodes.
3. **Analysis:** Examine how the data correlates with an ML model and suggest ways to provide actionable information improve a model’s use of the GPU.
4. **Visualisation:** Visualise the most salient information for CSCS engineers and users of the supercomputer to find bottlenecks or faulty hardware.
5. **Collaboration:** Work with CSCS engineers and another student team focused on visualisations.

Slide 3: Evaluation & Success Criteria [Project ID: 6]

Title: Supercomputer Data Extraction and Visualisation at Scale
Challenge Giver: Dr. Imanol Schlag, ETHZ AI Center + Dr. Jonathan Coles, CSCS

Evaluation Metrics:

1. **Analysis & Visualisation:** Useful to engineers and user alike

Successful Project:

- Deployment of a visualisation of low-level cluster hardware for CSCS engineers and users

Slide 1: Project Overview [Project ID: 7]

Title: Create an ETHZ AI Center Chatbot for Internal Use
Challenge Giver: Dr. Imanol Schlag, ETHz AI Center
Contact Details: ischlag@ethz.ch
Overview:

- **Aim**: Under the Swiss AI Initiative, various previous efforts have created inference api, data loading pipelines, and crawlers. The goal of this project is to combine everything to create a custom chatbot for internal use at the ETH AI Center.

- **Tools:** The chatbot will leverage an LLM API (e.g. our own internal API), have a RAG component (e.g. based on MMORE which was published in the Swiss AI Github org), and a crawler (e.g. similar to a previous DSL project based on scrapy). The chatbot should be integrated within our internal slack and made accessible to all ETH AI Center staff, fellows, and professors.

- **Focus:** An important feature of the described system should be to automatically crawl internal documents (e.g. our google drive, private notion page, etc), the latest ETH publications (by using a service we have prepared), and the latest ETH news (from the website, newsletter, etc).

Slide 2: Dataset & Challenge [Project ID: 7]

Title: Create an ETHZ AI Center Chatbot for Internal Use
Challenge Giver: Dr. Imanol Schlag, ETHz AI Center

Challenge Components:

- Build an “agent” with slack integration and access to public and private data
- Identify specific tasks/workflows where the agent can help operations at the ETHZ AI Center

Successful Project:

- Experimental working slack prototype
- Lessons learned on integrating data for chat use
- Lessons learned on augmenting / helping fellows and staff

Slide 1: Project Overview [Project ID: 8]

Title: Create a Blueprint for Programming with LLMs
Challenge Giver: Dr. Imanol Schlag, ETHz AI Center
Contact Details: ischlag@ethz.ch

Overview:

- **Aim**: LLMs are powerful tools. The goal of this effort is to explore "LLM Programs", i.e. classic programs with explicit subroutines implemented through a prompt and an LLM.

  The goal of this project is to kickoff a bluebrint / a collection of tricks and techniques / explore ideas on how to develop such programs. The **guiding example will be a task in collaboration with ETH Legal**. ETH legal has several hundred contracts with publishers which need to be check for AI compliance. This data will serve as an example to explore these techniques.

- **Tools:**
  https://github.com/xiaozheyao/vagent
  https://vagent.yao.sh/quick_start
  https://arxiv.org/abs/2305.05364

Slide 2: Dataset & Challenge [Project ID: 8]

Title: Create a Blueprint for Programming with LLMs
Challenge Giver: Dr. Imanol Schlag, ETHz AI Center
Contact Details: ischlag@ethz.ch

Dataset:

- Dozens of PDFs from ETHZ legal team

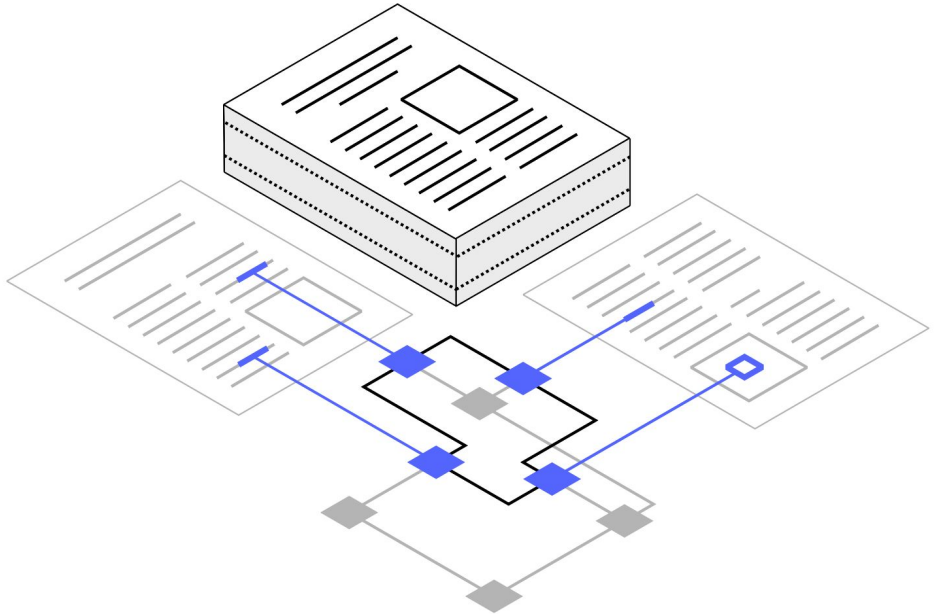
Challenge Components:

- Develop the practise of programming with LLMs
- Provide an example to identify paragraphs in ETHZ legal documents relevant to AI compliance questions.

Successful Project:

- Lessons learned and documented regarding the practise of programming with LLMs
- Provide specific result to ETH legal regarding paragraphs relevant to AI compliance

Slide 1: Project Overview



Project ID: 10
Title: Automated Research Paper Generation with LLM for Image Captioning and Citation Retrieval
Challenge Giver: Dr. Ahmed Nassar, IBM Research
Academic Coach:
Contact Details: ahn@zurich.ibm.com, taa@zurich.ibm.com, Matteo.Omenetti1@ibm.com

Overview:

- Aim: To develop an **LLM-powered framework** that automates research paper drafting by integrating agents, citation retrieval, SOTA summarization, image captioning using llms or vlms.
- Goal: an **LLM-driven research assistant** that leverages a modular agent-based framework to assist researchers in synthesizing literature, retrieving citations, and organizing technical content efficiently.
- Tools: LLMs/VLMs, Agent Frameworks, Research database APIs
- Focus: research and engineering
- Support: team of IBM researchers at ZRL lab in Rueschlikon

Slide 2: Dataset & Challenge

Title: Automated Research Paper Generation with LLM for Image Captioning and Citation Retrieval
Challenge Giver: Dr. Ahmed Nassar, IBM Research

Dataset:

- Full document AAAI, CVPR repositories converted into Docling Documents.
- Preprocessed figure captioning datasets.

Challenge Components:

1. Ensuring **generated** sections are relevant and logically structured.
2. Retrieving **accurate** and **relevant** citations.
3. Extracting and summarizing **state-of-the-art (SOTA) methods**.
4. Generating **image captions** that align with research content.
5. Addressing integration challenges between **LLMs, APIs, and research databases**.
6. Balancing model **performance** with **computational efficiency** to ensure scalability and cost-effectiveness.

Slide 3: Evaluation & Success Criteria

Title: Automated Research Paper Generation with LLM for Image Captioning and Citation Retrieval
Challenge Giver: Dr. Ahmed Nassar, IBM Research

Evaluation Metrics:

1. **Section Generation Quality:** BLEU Score / ROUGE Score to measure how well the generated aligns with real paper examples.
2. **Citation Retrieval / SOTA table:** Precision, Recall, F1-score evaluates the relevance and correctness of citations retrieved from external databases with real papers.
3. **Coherence & Readability of Generated Text:** Perplexity Score to evaluate coherence of the generated text.
4. **Image Captioning Accuracy:** CIDEr Score to compare with real captions.

Successful Project:

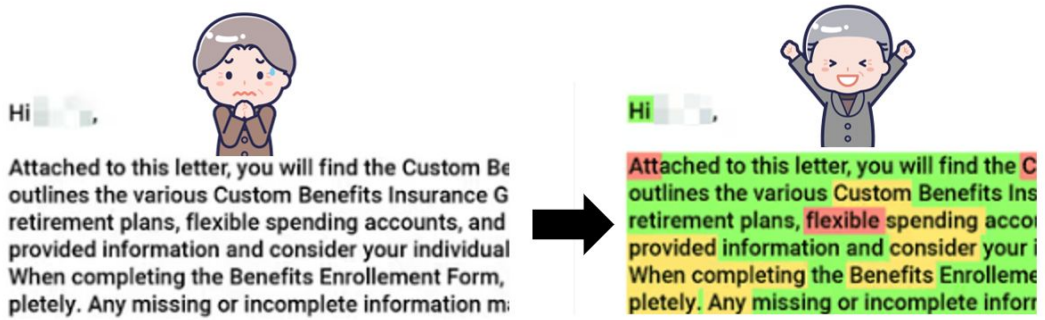
- Provides coherent, structured research paper drafts based on given title, abstract, introduction (?) figures.
- Retrieves relevant citations and references
- Generates correct SOTA comparison tables.

Slide 1: Project Overview

Project ID: 11

Title: Explainable AI for Detecting Synthetic Content in Phishing Emails

Challenge Giver: Lorin Schöni, ETH Zurich
Academic Coach: Lorin Schöni
Contact Details: lschoeni@ethz.ch



Overview:

- Aim: Detect AI-generated content in phishing emails and explain it to novice users
- Tools: LLMs, NLP, LIME, SHAP
- Focus: genAI detection, local explainability, cybersecurity, human-computer interaction
- Support: Team of the Security, Privacy and Society lab

Slide 2: Dataset & Challenge [Project ID: 11]

Title: Explainable AI for Detecting Synthetic Content in Phishing Emails
Challenge Giver: Lorin Schöni, ETH Zurich

Dataset:

- **Sources:** Greco et al. (2024), options for other datasets or generate your own!
- **Content:** Email content, brief meta information, classification

Challenge Components:

1. **NLP Extraction:** Extract linguistic and structural features, focusing on factors like length and manipulation patterns typical in phishing messages
2. **Explainability:** Use methods like LIME or SHAP to illustrate the importance of specific features. Explain them, e.g., through a visual explanation panel with extracted features and colour-coded annotations
3. **Additional Steps:** Optionally, compare and visualise trends between emails

Slide 3: Evaluation & Success Criteria [Project ID: 11]

Title: Explainable AI for Detecting Synthetic Content in Phishing Emails
Challenge Giver: Lorin Schöni, ETH Zurich

Evaluation Metrics:

1. **Performance:** Classification of AI generated phishing emails
2. **Explainability:** Stable explanations (similar classifications receive similar explanations)
3. **User Experience:** Usability of the explanation/tool

Successful Project:

- A model that detects AI generated content in phishing emails.
- Explain the classification to users: Can your friend understand it?
- Explain the classification to novice users: Can your grandma understand it?

Slide 1: Project Overview

Project ID: 12
Title: Inference Engine for the Internet Computer Protocol
Challenge Giver: Islam El-Ashi, AI Lead, DFINITY Foundation
Contact Details: islam.elashi@dfinity.org

Overview:

- The Internet Computer Protocol powers the Internet Computer network
- Think of the Internet Computer network as a decentralized cloud provider, much like AWS or Azure, but with the security and digital ownership of a blockchain network.
- Services (or smart contracts) on the Internet Computer range from cryptocurrency applications to social networks to games (and more).
- We'd like to invest more into running AI models directly within smart contracts.

Slide 2: Challenge

Title: Inference Engine for the Internet Computer Protocol
Challenge Giver: Islam El-Ashi, AI Lead, DFINITY Foundation

Challenge:

- Smart contracts on the Internet Computer are single-threaded (i.e. uses one CPU)
- Running LLMs are infeasible. Running smaller models (e.g. ~1B parameters) is feasible, but slow.

Challenge Components:

1. **Write an inference engine,** optimized for the Internet Computer.
2. **Models can be your choice** (e.g. LLMs, speech recognition)
3. **A sample implementation of an inference engine for Llama models will be provided.** You can use that as a starting point or start from scratch.

Slide 3: Evaluation & Success Criteria

Title: Inference Engine for the Internet Computer Protocol
Challenge Giver: Islam El-Ashi, AI Lead, DFINITY Foundation

INTERNET COMPUTER

Evaluation Metrics:

1. **Performance:** how fast models can run (measured in time, and in “instructions”)
   - This project has emphasis on HPC just as it does on AI
2. **Breadth:** how many different applications can be built on top of this inference engine

Successful Project:

- Implements an inference engine that can be used to run various AI models on the Internet Computer
- Easy to understand and maintain
- Can be used for various application domains

Data Science Lab S2025
ETH AI CENTER

Slide 1: Project Overview [Project ID: 13]

Title: Automatic measurement of garments for a fashion retailer
Challenge Giver: Inditex Tech
Contact details: Dr. Adrián González (adriangsi@inditex.com)
Laura Rodríguez (laurarbar@inditex.com)

Overview:

- **Aim**: Automatically measuring garments on real world images, with support for different clothing types and expert instructions.

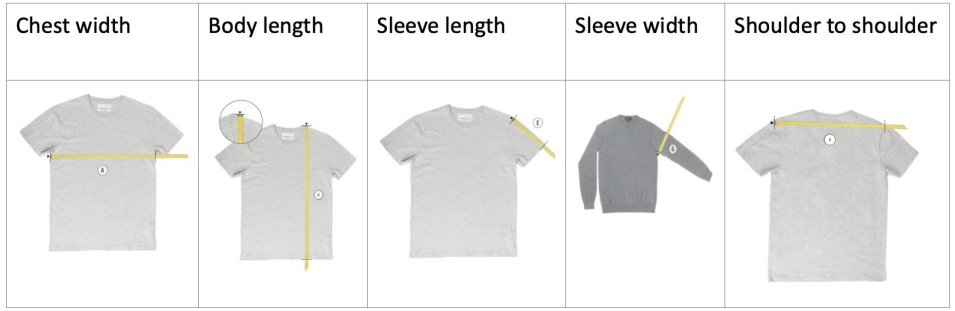
- **Goal**: Develop a robust, production-ready process, able to obtain different measurements automatically. The pipeline should be able to interpret expert instructions to know how each type of garment is measured, and adapt its behaviour without requiring data labeling or model retraining.

- **Tools**: Python, Pytorch, Segment Anything Model (SAM), Grounding Dino, Multimodal & GenAI models.
- **Focus**: coding / engineering, model evaluation
- **Support**: Data scientists from Inditex Tech

INDITEXTECH

Slide 2: Dataset & Challenge [Project ID: 13]

Title: Automatic measurement of garments for a fashion retailer
Challenge Giver: Inditex Tech



Dataset:

- **Sources:** Fashion Product Images/Fashion Clothing Segmentation/DeepFashion
- **Content:** Fashion image datasets for different computer vision tasks, i.e. segmentation, classification, landmark detection.

A proprietary dataset could also be provided for validation purposes.

Challenge Components:

1. **Image segmentation & normalization**: Background removal using segmentation models, perspective correction, and use other normalization techniques to ensure the robustness.
2. **Image classification:** Classify the garment in the picture according to certain categories.
3. **Text & image processing:** Use multimodal & GenAI models to extract measuring instructions per garment type.
4. **Keypoint extraction**: Place the key points according to the instructions that apply to each garment type, and calculate the corresponding measurements.

INDITEXTECH

Slide 3: Evaluation & Success Criteria [Project ID: 13]

Title: Automatic measurement of garments for a fashion retailer
Challenge Giver: Inditex Tech

Evaluation Metrics (suggested):

1. **F1-score:** For garment type classification.
2. **OKS:** Object keypoint similarity.
3. **Benchmarking:** Comparison between predicted metrics and provided ground truth.

Successful Project:

- Supports different garment types (tops, trousers, skirts, coats, etc.), and automatically/easily extensible to new categories.
- Uses expert instructions to determine how each garment type is measured, and it is able to support dynamic instructions.
- Returns all the expected measurements per garment type.
- Achieves high-precision results (millimetre accuracy).
- Results computed in seconds, so that the pipeline can be integrated in production processes.

Slide 1: Project Overview

Project ID: 14

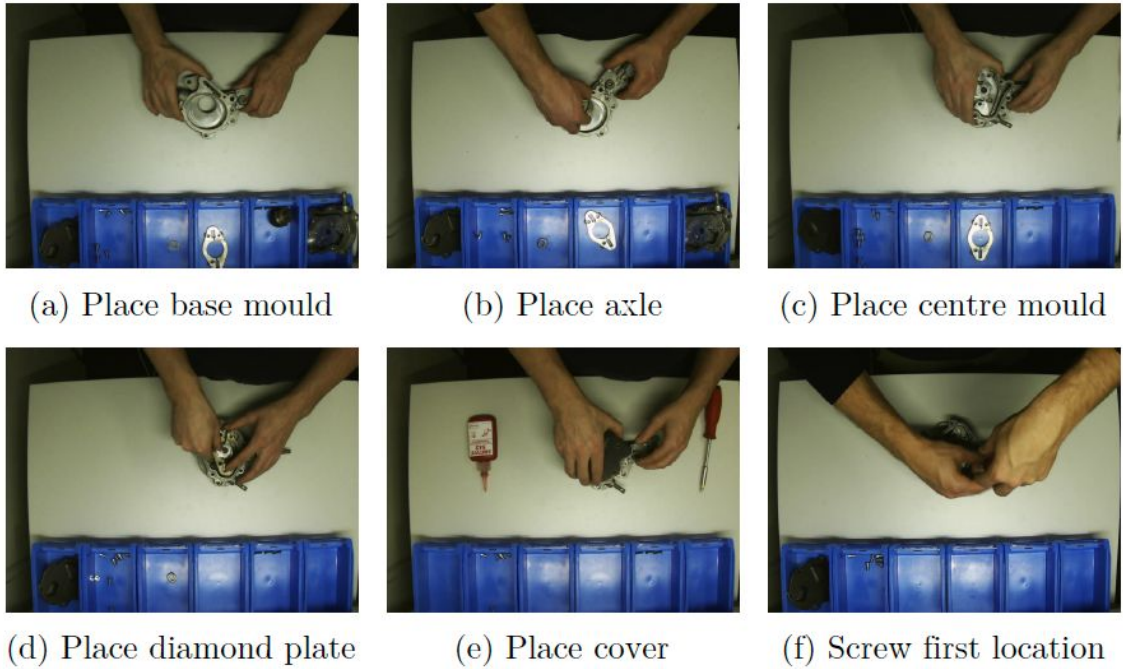
Title: Automated anomaly detection in manual assembly based on synthetic data

Challenge Giver: Constantin Herbst, Jonas Conrad, Sentinus AG

Academic Coach:

Contact Details:

constantin@sentinus.ch, jonas@sentinus.ch



(a) Place base mould (b) Place axle (c) Place centre mould

(d) Place diamond plate (e) Place cover (f) Screw first location

Overview:

- **Aim:** Perform robust anomaly detection in the manufacturing process trained on synthetic data
- Goal: A pipeline that recognizes anomalies in single parts and wrongly assembled subassemblies based on CAD data of assembly group in real time.
- **Tools:** python, blender, pytorch
- **Focus:**
  a. Robustness: needs to work on real data under realistic conditions
  b. Data-Efficiency: no need to record data to prepare inspection for assembly
  c. Semantic Scene Understanding: Correct segmentation of parts / subassemblies
- **Support:**
  a. Recording extra datasets for specific anomalies
  b. Real-time testing with sentinus smart camera hardware
  c. Guidance from Sentinus team

Slide 2: Dataset & Challenge [14]

Title: Automated anomaly detection in manual assembly based on synthetic data

Challenge Giver: Constantin Herbst, Jonas Conrad, Sentinus AG

Dataset:

- **Sources:** MA-3 dataset
- **Content:** 60 videos of manual assembly of different products; 20 operators. Focus on vacuum pump. The dataset can be enhanced by Sentinus.

Challenge Components:

Build a CV pipeline that

1. Based on CAD data of each part / assembly stage creates a rendered dataset to train models for the following step
2. Segments all relevant assembly parts in a live video stream in real time
3. Performs anomaly detection on single parts and sub-assemblies
   a. Anomaly detection on subassembly: wrong assembly!
4. Rejects frames in which no inspection can be performed reliably due to occlusion, motion blur etc,

Slide 3: Evaluation & Success Criteria [14]

Title: Automated anomaly detection in manual assembly based on synthetic data

Challenge Giver: Constantin Herbst, Jonas Conrad, Sentinus AG

Evaluation Metrics:

1. Accuracy of recognized assembly status
2. Accuracy of segmentation (mIOU)
3. Detection rate of anomalies (screw missing, flipped part) (AUROC/PRO)

Successful Project:

- Can discern assembly status on real data
- Rejects obvious disturbances / occlusions (e.g. hand covers assembly group)
- Can pinpoint anomalies robustly: rotation, position, illumination invariant, robust against slight variation in part appearance
- Well-documented and reasonable code

Slide 1: Project Overview

Project ID: 15

Title: Learning to Optimize for Petrol Station Replenishment Problem

Challenge Giver: Askantis GmbH, Antonia Unger, Mert Erkul

Contact Details: antonia.unger@askantis.ch, mert.erkul@askantis.ch

Overview:

Aim: Given the standard NP-Hardness of the CVRPTW problem and promising results from **Learning to Optimize** paradigm, apply RL / GNN as metaheuristics for PSRP.

Goal: Generalize RL / GNN heuristics for unseen problem settings

Tools: Python, CPLEX, AI4CO

Focus: Research and Development for a real-life day-to-day operational problem

Support: Data Scientists from Askantis

Data Science Lab S2025

ETH AI CENTER

Slide 1: Project Overview

Project ID: 15

Title: Learning to Optimize for Petrol Station Replenishment Problem

Challenge Giver: Askantis GmbH, Antonia Unger, Mert Erkul

Contact Details: antonia.unger@askantis.ch, mert.erkul@askantis.ch

Data and Sources:

1. **Internal Data:** Historical operational data and replenishment logs provided by Askantis GmbH
2. **Academic Literature:** Research papers on CVRPTW, NP-Hard problems attacked with heuristics, Reinforcement Learning (RL), and Graph Neural Networks (GNN)

Challenge components:

1. **Model Generalization:** Ensuring that the developed RL and GNN models generalize well to unseen scenarios
2. **Integration Complexity:** Seamlessly incorporating the new models into existing operational frameworks
3. **Demand-as-decision:** Incorporating demand as a decision variable adds an extra layer of complexity, necessitating advanced modelling techniques to dynamically adjust to variable demand levels, potentially needing a hybrid approach

Slide 1: Project Overview

Project ID: 15

Title: Learning to Optimize for Petrol Station Replenishment Problem

Challenge Giver: Askantis GmbH, Antonia Unger, Mert Erkul

Contact Details: antonia.unger@askantis.ch, mert.erkul@askantis.ch

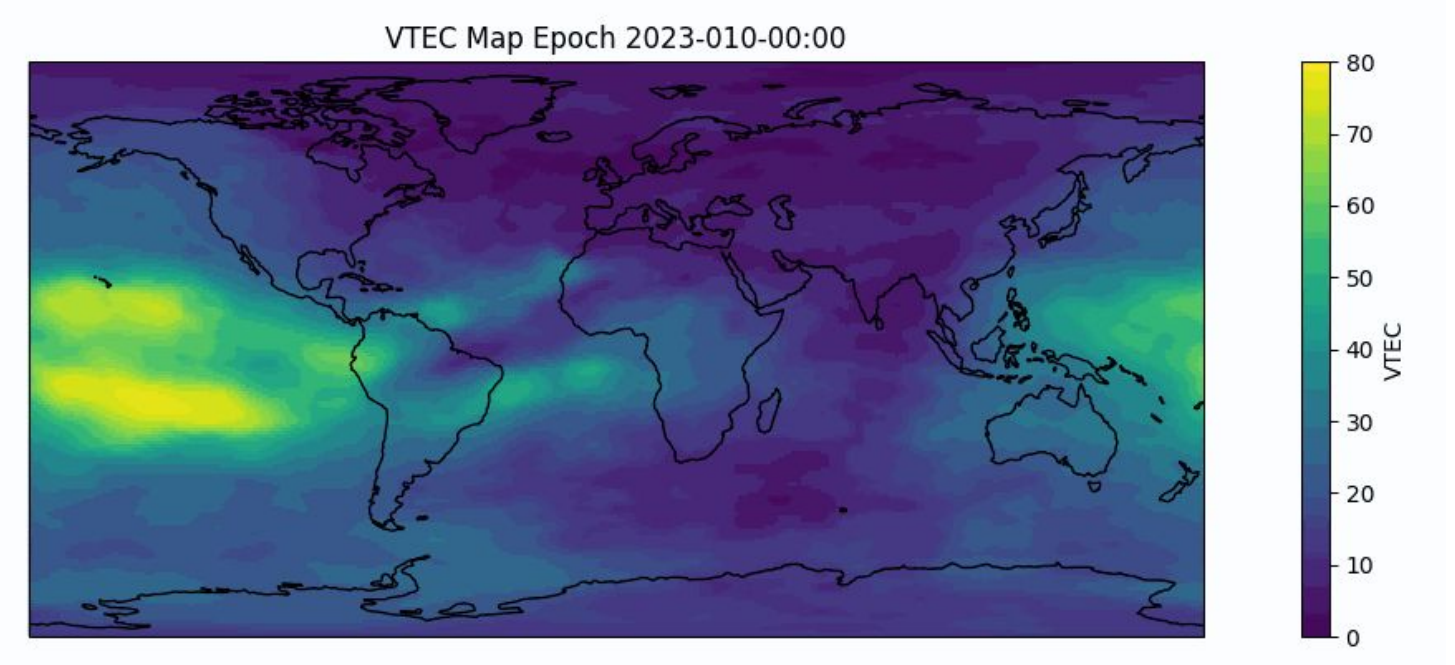
Evaluation Criteria:

1. **Robust performance** of RL/GNN models across varied and unforeseen problem instances.
2. **Integration Readiness:** Successful prototype integration with Askantis' systems, validated by data scientists.
3. **Numerical KPIs:** Better runtime, less variance among perturbed problem instances.

Successful Project:

1. **Baseline:** Integration & Clean Code + Learning as Metaheuristics proven efficient
2. **MVP:** Generalize to other regions and adapt to varied operational environments.
3. **Advanced:** Remove reliance on traditional heuristics and solvers in favor of provably successful learning techniques

Slide 1: Project Overview



Project ID: 16
Title: NeRF for Ionospheric Modelling
Challenge Giver: Arno Rüegg
Academic Coach: Prof. Benedikt Soja
Contact Details: arrueegg@ethz.ch

Overview:

- Aim: Apply NeRF like ML pipeline to GNSS observation data for ionospheric modelling.
- Goal: Learn a 3D + time ionosphere model
- Tools: Train and evaluate ML algorithm
- Focus: Machine learning application with real data for complex, physical problem
- Support: ETH Zurich Space Geodesy lab.

Slide 2: Dataset & Challenge

Title: NeRF for Ionospheric Modelling
Challenge Giver: Arno Rüegg

Dataset:

- **Sources:** International GNSS Service Station Network
- **Content:** Slant Total Electron Content (TEC) data from GNSS observations

Challenge Components:

1. **Observation Geometry:** Sparse observations in time and space
2. **Implicit Modelling:** Model 3D ionosphere with ground truth only along line of sight

Slide 3: Evaluation & Success Criteria

Title: NeRF for Ionospheric Modelling
Challenge Giver: Arno Rüegg

Evaluation Metrics:

1. **MAE:** Numeric evaluation metrics on test set
2. **Benchmarking:** Compare to baseline vertical TEC models (current standard)
3. **Independent Evaluation:** Compare to Satellite Altimetry (Jason3) data

Successful Project:

- 3D ionosphere model that gives accurate slant TEC corrections
- Meaningful 3D output, interpolating well in space and time
- Bonus: Competing with current SOTA in ionosphere modelling

Slide 1: Project Overview
Project ID: 17
Title: Finding patterns in financial Post-trading data
Challenge Giver: Orlando Monsalve & Elisabeth Wallimann, SIX Group
Academic Coach:
Contact Details: orlando.monsalverueda@six-group.com

Overview:

- Aim: Analyze financial post-trading data to uncover hidden patterns and relationships between various variables using synthetic data.
- Goal: To enhance the understanding of the interactions and dependencies between different features in post-trading activities, and improve anomaly detection in financial transactions.
- Tools: ML clustering and other unsupervised learning techniques.
- Focus: Research in Post-trading.
- Support: SIX provides synthetic data derived from post-trading activities and other anonymized variables.

Slide 2: Dataset & Challenge
Project ID: 17
Title: Finding patterns in financial Post-trading data
Challenge Giver: Orlando Monsalve & Elisabeth Wallimann, SIX Group

Dataset:

- SIX provides synthetic data with financial post-trading transactions, thus some variables are autocorrelated. The data is normalized and encrypted according to our privacy policies. Data is also cleaned, meaning that non-relevant information is excluded. Some variables are intentionally added as noise to assess the effectiveness of the technique employed.

Challenge Components:

- Utilize a range of clustering algorithms to analyze the data and identify hidden patterns and relationships among variables.
- Display the findings through comprehensive reports and visualizations to highlight the newly discovered relationships.
- Report ineffective techniques, i.e., those that did not provide insights of relationships.
- Illustrate the types of anomalies that can be detected by these effective clusters.

Slide 3: Evaluation & Success Criteria
Project ID: 17
Title: Finding patterns in financial Post-trading data
Challenge Giver: Orlando Monsalve & Elisabeth Wallimann, SIX Group

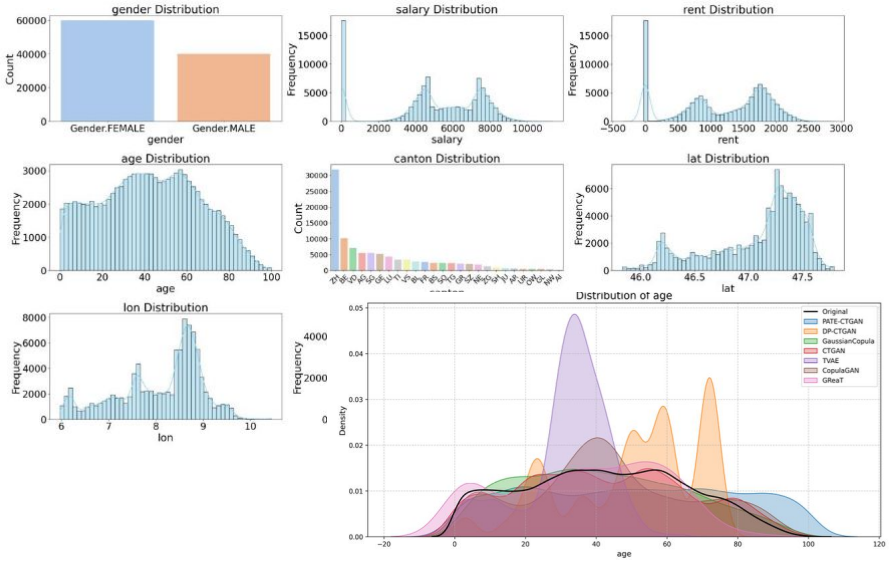
Evaluation Metrics:

1. Effectiveness of unsupervised learning techniques based on the number of variables that can be clustered.
2. The ability of the technique to identify and classify noise variables accurately.
3. Stability of the unsupervised learning technique in terms of how clusters adapt when new data is introduced.
4. Scalability and generalizability of the approach.

Successful Project:

- Offers valuable insights into the different techniques utilized and assess their effectiveness, based on number of variables clustered, their relationships and other measures proposed by the participants.

Slide 1: Project Overview



Project ID: 19
Title: Benchmarking of (privacy-preserving) Tabular Synthetic Data Generation
Challenge Giver: Dr. David Froelicher & Dr. Daniel Dobos
Academic Coach:
Contact Details: david.froelicher2@swisscom.com

Overview:

- Aim: Comprehensive benchmark for synthetic tabular data generation
- Goal: Design and implement a benchmarking approach for synthetic data generators evaluating their privacy, utility and performance tradeoffs.
- Tools: Synthetic data generators (GANs, LLMs, …) and benchmarking libraries.
- Focus: research and engineering
- Support: team from the Swisscom Digital Lab

Data Science Lab S2025
ETH AI CENTER

Slide 2: Dataset & Challenge

Title: Benchmarking of (privacy-preserving) Tabular Synthetic Data Generation
Challenge Giver: Dr. David Froelicher & Dr. Daniel Dobos

Dataset:

- **Sources & content:**

  a. exploratory benchmarking library https://github.com/MagMueller/Synthetic-Data-Generation & report on first exploratory projects

  b. Outcomes of previous exploratory projects in Swisscom

Challenge Components:

1. **Evaluation of existing approach:** Research existing methods and identify improvements
2. **Practical Insights:** Explore practicality of existing approaches, their practicality, usability, and design of new approach
3. **Advanced:** Recommendation of the best approach to use in an industry context

Slide 3: Evaluation & Success Criteria

Title: Benchmarking of (privacy-preserving) Tabular Synthetic Data Generation
Challenge Giver: Dr. David Froelicher & Dr. Daniel Dobos

Successful Project:

- Identifies improvements with respect to existing approaches.
- Implementation and design of a novel approach
- Provides insights and recommendations based on the novel approach

Project ID 20 – 1. Overview

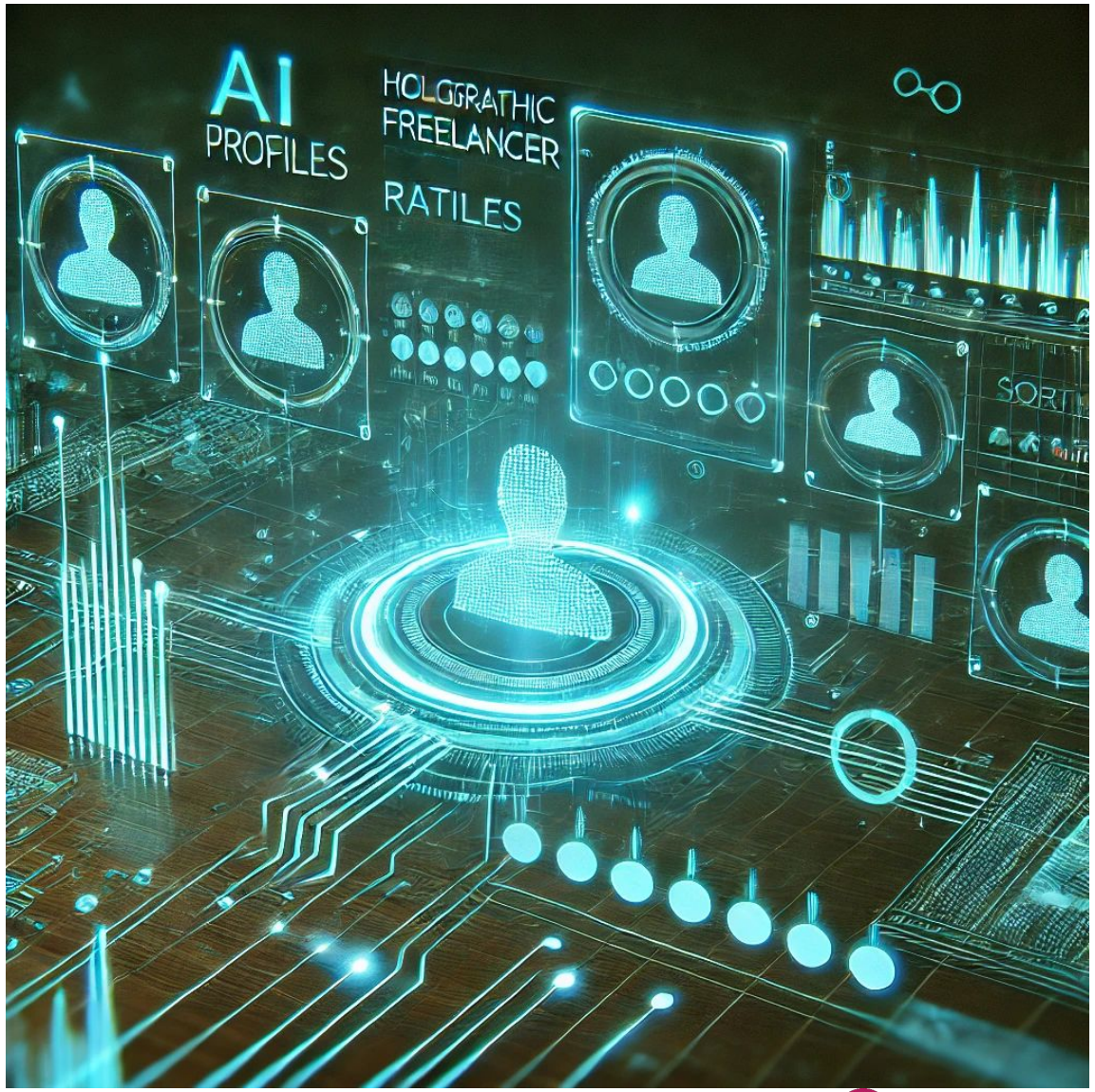
Enhancing Freelancer Ratings with Pairwise Comparisons

Challenge Giver / Academic Coach: Prof. Elliott Ash

Contact Details: ashe@ethz.ch

Overview: We're creating an AI-driven rating system that replaces the flawed five-star approach with pairwise comparisons. The end result is fairer, more accurate freelancer evaluations.

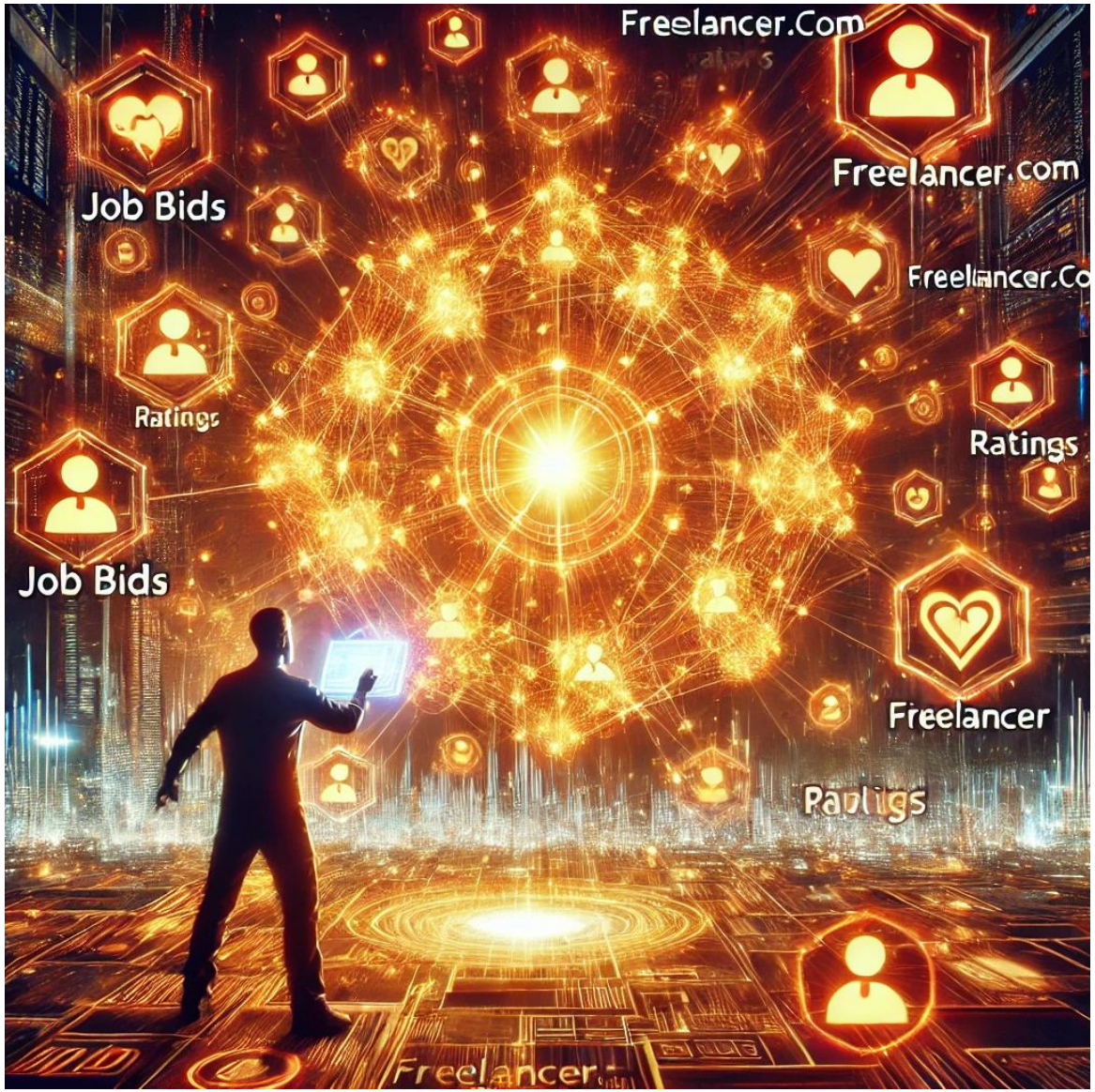
- **Aim**: Develop a comparative rating mechanism that reduces subjectivity and guilt around giving low scores.
- **Goal**: Build a scalable platform leveraging game theory, latent variable models, Bradley-Terry estimation, and causal machine learning to produce less biased freelancer rankings.
- **Tools**: Python, statistical modeling frameworks, cloud computing resources, causal inference methods.
- **Focus**: Both research (publishable findings) and engineering (functional prototype).
- **Support**: You'll collaborate with a professor, a postdoc, and two PhD students who offer guidance on modeling, data processing, and publication.



Data Science Lab S2025

ETH AI CENTER

Project ID 20
2. Dataset & Challenge



- **Data Sources**: a web scrape of Freelancer.com and other gig-economy datasets.
- **Content**: Information on jobs, bids, and the current star-based ratings.
- **Challenge Components**:
  1. **Data Wrangling**: Extracting and organizing large-scale gig data (projects, bids, ratings).
  2. **Game Theory:** Model the interaction between workers and bidders based on the rating mechanism. Solve the equilibrium.
  3. **Model Building**: Implementing and testing latent variable models and Bradley-Terry regressions with pairwise comparisons.
  4. **Bias & Confounding**: Factoring in contextual aspects such as project complexity and employer behavior through causal methods.
  5. **Engineering**: Designing a system that can handle real-world data volume while remaining user-friendly.

Data Science Lab S2025
ETH AI CENTER

Project ID 20

3. Evaluation & Success

- **Evaluation Metrics**:
  - Demonstrate theoretically and empirically the bias in current star-based ratings; in particular, top-coded ratings being uninformative.
  - Show how pairwise comparisons are better calibrated and more predictive of future performance.
  - Simulate scenarios to learn worker quality faster using comparative methods.
  - Identify conditions (e.g., too few comparable tasks) that undermine model accuracy, and adapt latent variable models and causal ML methods to mitigate these problems.
  - Assess the benefits of integrating both absolute and pairwise ratings into the platform.
- **Deliverables**:
  - Research paper on theory, methods, results, and implications.
  - A startup concept to bring this improved rating system to real-world freelancing platforms.

This project offers a challenging and entertaining environment for aspiring data scientists aiming for grad school or entrepreneurial ventures. You'll learn skills in game theory, statistical modeling, causal inference, and machine learning. You'll contribute to novel research. And the work could have real-world impact on freelancing platform design.

Data Science Lab S2025

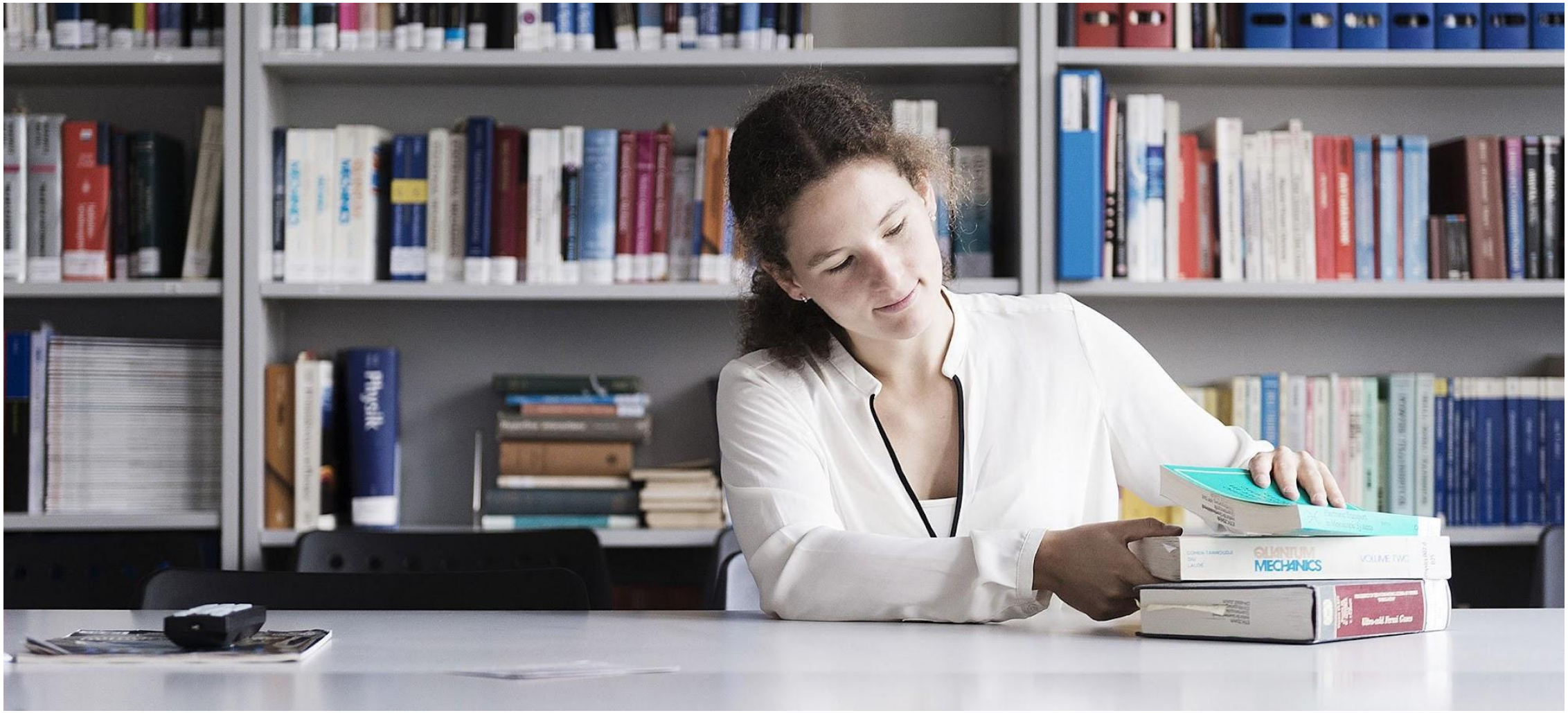
ETH AI CENTER

Project Preference and Group Formation

Data Science
Lab S2025

ETH AI CENTER

Q & A



Data Science
Lab S2025

ETH AI CENTER