

# Manthan Gadhia

**Contact**  
[manthan.gadhia@proton.me](mailto:manthan.gadhia@proton.me)  
www.hellomanthan.com  
+41 76 212 5643  
Zürich, Switzerland

## AI researcher

MSc student with an interdisciplinary training in the theoretical fundamentals and practical considerations of human-centred ML and data science applications. Focused on AI's impacts and potential on work and education. Interested in language-processing.

## Education

**ETH ZÜRICH, CH**  
MSc, Science, Technology, and Policy (CS focus)  
2026 (expected)

**RADBOUD UNIVERSITY, NL**  
BSc, Artificial Intelligence  
2023

## Skills

**ML:** PyTorch, River, HuggingFace, Accelerate, CUDA, CI/CD  
**LLM Training:** SFT, PPO,

## Relevant Coursework

NLP (BSc)  
Deep Learning (BSc)  
Advanced ML (MSc)  
Interactive ML (MSc)  
LLMs for Law (MSc)  
Design in EdTech (MSc)

## Experience

**MSC THESIS** ANALYSING GENDER BIAS IN LLM TRAINING  
OCT 2025 – JUN 2026 (expected, publication-track)

Setup end-to-end data pipeline to generate synthetic data from LLMs and evaluate outputs for gender bias in occupational settings at different training stages.

Future tasks: Do RLHF (via PPO) or DPO on Apertus open-models on the Alps supercomputer. [Github \(in-progress\)](#)

**AI INTERN** UNICRI CENTRE FOR AI AND ROBOTICS  
Jan 2023 – Aug 2023

Helped to catalogue, discover, and communicate about new AI tools for security and law enforcement focused on fighting online child abuse through [AI for Safer Children](#).  
Built internal software tools to manage platform access and email communication with 100+ participating officers globally.

**BSC THESIS** BAYESIAN NETWORK EXPLAINABILITY  
Feb 2022 - Jul 2022

Designed a computational-level theoretical model for encoding a user's level of domain-expertise. Context: generating explanations for Bayesian Decision Support Systems in a medical diagnostics support use case. ([Gadhia, 2022](#))

**PHILGPT** RAG  
Summer 2025

Scraped transcripts from *PhilosophizeThis!* podcast and build a RAG using sentence-transformers, and Gemma models via API. ([PhilGPT](#))