Alexander Müller

LinkedIn: www.linkedin.com/in/alexanderkmüller
Personal Website: https://alexanderakm.github.io
Github: https://github.com/AlexanderAKM

alexanderklavermuller@gmail.com
Mobile: +31 06 22047760

Education

- **University of Groningen** — Groningen, The Netherlands
  Bachelor of Science - Artificial Intelligence — *September 2022 - July 2025*
  - **Relevant Courses**: Algorithms and Data Structures, Object-Oriented Programming, Linear Algebra, Statistics, (Multivariable) Calculus, Neural Networks, Reinforcement Learning, Uncertainty in Machine Learning, Analysis, Advanced Logic, Ethics in AI, Cognitive Psychology, Neuroscience, Behavioural Neuroscience
  - **GPA**: 8.9/10
  - **Bachelor's Thesis**: Machine Learning for Peptide Property Prediction: A Robust Nested Cross-Validation Pipeline with Mechanistic Interpretability. Conducted the first case study of mechanistic interpretability in the domain of chemistry. Received a 9/10.

  Master of Science - Artificial Intelligence — *September 2025 - July 2027*
  - **Relevant Courses**: Advanced Machine Learning, Responsible AI, Design of Multi-Agent Systems
  - **GPA**: N/A

Working Experience

- **Co-Director Artificial Intelligence Safety Initiative Groningen (AISIG)**: Aug. 2025 - Ongoing
  - **Mission**: Our mission is to raise awareness of the full spectrum of existing and potential harms from AI, inform mitigation priorities through ongoing discourse, and support the realization of effective solutions.
  - **Focus**: To achieve the aforementioned mission, we (i) provide educational resources, workshops, and seminars related to AI safety, (ii) support interdisciplinary research projects, (iii) engage with the general public to raise awareness about AI safety, and (iv) facilitate connections between students, academics, and professionals who share our vision.
- **Leading NeurAligment Research & Discussion Group**: Nov. 2024 - Ongoing
  - **Organizing and Coordinating**: Leading the NeurAlignment Research & Discussion group where the main focus is on exploring the intersection of neuroscience and AI to improve how AI systems align with human values. Under my supervision, three research projects are currently being conducted.
- **Mentor for First-Year Students**: Sep. 2023 - Feb. 2024; Sep. 2024 - Feb. 2025
  - **Student Support and Guidance**: Mentored first-year AI students, assisting them in navigating the Dutch university system and adapting to student life.
- **Teaching Assistant**: Sep. 2023 - Feb. 2025
  - **Introduction to Artificial Intelligence (Sep. - Nov. 2023; Sep. - Nov. 2024)**: Assisted in teaching foundational AI concepts, leading tutorials, labs, and assessments.
  - **Basic Scientific Skills (Nov. 2023 - Feb. 2024; Nov. 2024 - Feb. 2025)**: Supported students in developing essential scientific skills, focusing on research methodologies, data analysis, and academic writing.
  - **Cognitive Psychology (Feb. - Apr. 2024)**: Guided students through key cognitive psychology concepts, providing support with course materials and exam preparation.
  - **Introduction to the Brain (Apr. - Jun. 2024)**: Facilitated learning in the Introduction to the Brain course, covering topics related to neuroscience and its connection to AI.
- **Private Tutor for High School Students**: Mar. 2019 - Feb. 2024
  - **Customized Learning Plans**: Provided personalized tutoring in Mathematics, Chemistry, and Physics, developing tailored lesson plans and supporting students academically and psychologically for major exams.

Projects

- **AI Safety**: 2024-present
  - **AI Safety Course**: Initially completed the 'AI Safety Fundamentals - Technical' course organized by the AI Safety Initiative Groningen (AISIG).
  - **NeurAlignment Research & Discussion Group**: Together with other passionate students (BSc-PhD), we explore how we can use our knowledge of the brain to help solve the alignment problem.
- **MNIST Tutorials**: 2024
  - **Python Machine Learning**: Developed a series of tutorials on building ML models using PyTorch and various other libraries to recognize handwritten digits from the MNIST dataset.
  - **Structured Pipeline**: Designed the tutorials to focus on building a clean pipeline, allowing others to learn and adapt the code rapidly.
- **Traffic Signal Control Simulator**: 2024
  - **Reinforcement Learning**: Designed a simulator to optimize traffic light control in urban environments, using RL techniques to improve traffic flow and reduce waiting times at intersections.
  - **Various RL Agents**: Implemented DQN, A2C, and a Stochastic agent to control a 2-way traffic intersection.
- **Lunar Landing**: 2024

- **Reinforcement Learning**: Developed a RL model to safely land a lunar module on the moon's surface with the Gymnasium library, implementing Q-learning, SARSA, and Deep Q-Network (DQN) with techniques like experience replay and fixed Q-targets.

- **Role Playing Game**: 2023
  - **Object-Oriented Programming**: Collaborated with peers to develop an RPG with a graphical interface, strictly adhering to OOP principles.
  - **Model, View, Controller**: Implemented in Java, following the Model, View, Controller (MVC) design pattern.

Skills

- **Languages**: Python, C, Java, R
- **Frameworks**: PyTorch, Scikit-learn, Gymnasium, Pandas, NumPy, Keras
- **Tools**: GIT/w
- **Soft Skills**: Teaching, Group Coordination, Public Speaking, Time Management
- **Writing**: Excellent academic writing (C2 Cambridge certificate) and maintaining a newsletter on AI and neuroscience.

Awards

- **3rd Place Teaching Assistant of the Year (2025)**: Won 3rd place (out of $\tilde{5}0$ teaching assistants) at our program's TA of the year ceremony.

Talks

- **AI Safety Talk Turn.io (2025)**: One-hour AI safety talk and Q&A for 30+ staff of Turn.io where we discussed broader and more specific risks associated with LLM-powered chatbots and proposed actionable mitigations.