Joris Postmus

joris@aisig.org ❖ jorispos.com ❖Groningen, Netherlands

WORK EXPERIENCE

AI Safety Initiative Groningen **Nov. 2022 – Present**
Co-Founder & Co-Director *Groningen, Netherlands*

- Co-founded AISIG to address AI's potential risks, focusing on safety, ethics, and alignment.
  - Organized events with hundreds of signups, including speakers from OpenAI.
  - Raised funding to support AISIG's mission and operations.
  - Organized/facilitated the AISF Technical & Governance courses for 50+ (BSc-PhD) students
- Lead/manage a team of 10+ students and engage with the wide academic/professional community.

University of Groningen **Sep. 2021 – Present**
Teaching Assistant & Student Mentor | Sep. 2021 – Mar. 2023 *Groningen, Netherlands*

- Teaching Assistant for the courses: Introduction to Artificial Intelligence and Basic Scientific Skills
  - Helped students understand the course materials in fun and engaging ways.
  - Gave tutorials & learning discussions to discuss contents, answered questions, and provided feedback.
  - Graded students work and had weekly meetings with lecturers of the courses.
- Guided a group of approx. 10 students per year to university life through group and individual mentor meetings, monitoring progress, and providing additional support.

Study Buddy | Apr. 2022 – Present

- Provide academic support and coaching to students with ASD or AD(H)D under professional supervision.

Malvern Collegiate Institute **Jan. 2018 – Jun. 2020**
President of Coding Club *Toronto, Canada*

- Taught students how to code, competed in coding competitions, and oversaw collaborative projects.

EDUCATION

University of Groningen **Sep. 2020 – Present**
BSc Artificial Intelligence (Completed), BSc Computing Science *Groningen, Netherlands*

- Final grade: 8.3 (BSc AI)
- Thesis: 9.5 (*"Steering Large Language Models using Conceptors: An Alternative to Point-Based Activation Engineering"*) [src]
- Elective courses + minor heavily geared towards mathematics, computing science, and machine learning.

Malvern Collegiate Institute **Sep. 2016 - Jul. 2020**
Ontario Secondary School Diploma (OSSD) *Toronto, Canada*

- Avg. grade of 95% (Honours, Summa Cum Laude equivalent)
- Includes highest level of: Mathematics, English, Biology, Physics, Chemistry, and Computing Science

CERTIFICATIONS, SKILLS & INTERESTS

- **Certifications/Awards:**
  - Malvern CI: Highest mark in Computer Science (2x, 2019, 2020), Honour Roll (3x, 2017-2020)
  - University of Waterloo: Cayley Math Contest (top 25% 2018), Canadian Computing Competition (CEMC) (2x, top 25% junior division, 2018, 2019)
  - Skills Ontario: Coding Competition (Gold medal 2019, Bronze medal 2020)
- **Language:** Dutch (Fluent, Native), English (Fluent, Native), French (Limited Working)
- **Skills:** Leadership, Programming (C, C++, Python, Java, JavaScript), Communication & Presenting
- **Interests:** Artificial Intelligence, Mechanistic Interpretability, Mathematics, Psychology, Personal Development
- **Publications:** NeurIPS 2024 – Workshop on Foundation Model Interventions [src]