

Céline Budding

[Homepage](#) | [LinkedIn](#) | [GitHub](#)

Location: Rotterdam, The Netherlands

Email: c.e.budding@tue.nl

PHD CANDIDATE PHILOSOPHY OF AI

I am a third-year PhD candidate, interested in what large language models learn from their data. In particular, I focus on whether these systems learn and represent any human-interpretable algorithms that can be used to explain their behavior. To this end, I argue that we should take inspiration from methods that have been used for explaining human behavior, such as cognitive modeling and intervention studies, to better conceptualize what AI systems learn.

In addition, I have a strong interest in communicating academic research to a broader audience, for example through giving lectures to elderly, writing blogposts, and being involved in standardization of AI via the Dutch NEN.

EXPERIENCE

PhD Candidate

Philosophy & Ethics group, Eindhoven University of Technology

Eindhoven, The Netherlands

Aug 2021 - current

- Project: How-explanations for large language models, supervised by dr. C.A. Zednik and prof. dr. V.C. Müller.
- Member of the 4TU.Ethics graduate school, which offers various courses on philosophy of technology.

EDUCATION

Master of Science in Computational Neuroscience

Bernstein Center for Computational Neuroscience Berlin

Berlin, Germany

Oct 2018 - May 2021

- Thesis: Evaluating interpretability methods on structural brain MRI data with synthetic lesions, supervised by dr.rer.nat. S. Haufe and prof.dr. K. Ritter.

Bachelor of Science in Liberal Arts & Sciences

University College Roosevelt/Utrecht University

Middelburg, The Netherlands

Sep 2015 - June 2018

- Thesis: Classification of MRI scans from the ADNI database using voxel-based morphometry, principal component analysis, and support vector machine, supervised by dr. R. van den Doel and dr. G. Andringa.
- Major in mathematics, cognitive science, and life sciences.

CONFERENCE & WORKSHOP PRESENTATIONS

- Upcoming: Toward Cognitive Models in Explainable AI: Mechanistic Interpretability as a Missing Link?. Presentation together with dr. C.A. Zednik at the Biannual Conference of the European Society for Philosophy of Science. September 20-23 2023, Belgrade, Serbia.
- Tacit knowledge and how-explanations for large language models. Presentation at the 49th meeting of the Society for Philosophy of Psychology. June 20-23 2023, Pittsburgh, US.
- Do large language models have conceptual knowledge?. Poster presentation at the OZSW 2023. June 2-3 2023, Leiden, The Netherlands
- Tacit Knowledge and 'How'-Explanations for Large Language Models. Presentation at the Graduate Workshop of PhilML22. October 19 2022, Tübingen, Germany.
- Tacit Knowledge as a Target of Intervention in Neural Networks. Presentation at the Early Career Researchers Workshop in Philosophy of Mind and Cognitive Science. July 13-15 2022, Bochum, Germany [Virtual].

TEACHING EXPERIENCE

Philosophy and ethics of AI

Eindhoven University of Technology

Responsible teachers: dr. C.A. Zednik and dr. E. O'Neill

- Q2 2022-2023. Teaching assistant, partially responsible for grading intermediate assignment and exam, as well as for answering questions.

USE-BASE

Eindhoven University of Technology

Responsible teacher: dr. A. Spahn

- Q4 2021-2022. Teaching assistant, responsible for leading the tutorials and grading the assignment for one tutorial group (35 students).

ADDITIONAL ACTIVITIES

Board member WISE network

Eindhoven University of Technology

May 2023-current

- I am the PhD representative for our department (Industrial Engineering & Innovation Sciences) in the WISE board. As part of the board, I am involved in organizing events that promote women in science and engineering, as well as organizing the yearly WISE writing retreat.

Workshop on the Philosophy of Large Language Models

Eindhoven, The Netherlands

Hybrid/Eindhoven University of Technology

Jan 26 + 27 2023

- Together with Guido Löhr, I organized a two-day international workshop, with speakers (both invited and contributed) on the topics of communicating with chatbots and explaining large language models. Around 15 people attended in person, and over 20 people attended online.
- Sponsored by the Philosophy & Ethics group, Ethics of Socially Disruptive Technologies Consortium, and the Center for Philosophy of AI.

Representative PhD Council

4TU.Ethics Graduate School

Jan-Dec 2022

- As a representative in the 4TU.Ethics PhD council, I was responsible for communication between 4TU.Ethics and affiliated PhDs, organizing social events and a writing retreat, and attending the 4TU.Ethics management team meetings.
- Together with another member of the PhD council I organized a five-day writing retreat for 15 PhDs affiliated with the 4TU.Ethics graduate school.

NON-ACADEMIC ACTIVITIES

Committee member 'AI & big data'

NEN

March 2023-current

- As a committee member of the working group on AI & big data at the Dutch standardization agency (NEN), I am involved in the standardization of explainable AI at the Dutch level.

Studentdocent/Student Teacher

Stichting Oud Geleerd Jong Gedaan

Virtual/in person

Jan 2021 - current

- I developed a low-threshold and interactive lecture series (4 lectures of 1 hour) and stand-alone lecture (1 hour) about different topics in Artificial Intelligence ("Slimme computers?") for an audience of elderly people.
- I have taught the lecture series about AI three times, once in person and twice virtually, and the standalone lecture twice virtually. Attendance is usually 4 - 10 participants.