

Education

- 2025– **Visiting PhD Student, Prof. Fyshe's Lab**, *University of Alberta & Alberta Machine Intelligence Institute (Amii)*, Edmonton, Canada, Department of Computing Science.
Vision-Language Models, Compositionality, Neural Encoding, Knowledge Graph Representation Learning, Functional MRI
- 2021– **Interdisciplinary PhD Student & FWO Fellow**, *KU Leuven*, Leuven, Belgium, Departments of Neurosciences and Computer Science.
Natural Language Processing, Neurolinguistics, Neural Encoding, Alzheimer's Disease, Connected Speech, Functional MRI
- 2018–2021 **Master of Science in Computer Science**, *Bonn-Rhein-Sieg University of Applied Sciences*, Sankt Augustin, Germany.
Natural Language Processing, Bioinformatics, Alzheimer's Disease, Knowledge Graphs
- 2015–2018 **Bachelor of Science in Computer Science**, *Bonn-Rhein-Sieg University of Applied Sciences*, Sankt Augustin, Germany.
Bioinformatics, Software Engineering, Machine Learning
- 2017 **Visiting Undergraduate Student, Prof. Beiko's Lab**, *Dalhousie University*, Halifax, Canada, Department of Computer Science.
Metagenomics, Bioinformatics, Machine Learning

Research Experience

- 2018–2021 **Student Research Assistant**, *Fraunhofer Institute for Algorithms and Scientific Computing (SCAI)*, Sankt Augustin, Germany, Bioinformatics Department.
Implemented natural language processing methods in biomedical contexts and developed machine learning approaches for Alzheimer's Disease diagnosis.

Teaching Experience

- 2018–2020 **Teaching Assistant**, *Bonn-Rhein-Sieg University of Applied Sciences*, Sankt Augustin, Germany, Introduction to Probability Theory and Statistics.
Led exercise sessions for 60+ students per year covering Python programming and core statistics concepts.
- 2016–2019 **Tutor**, *Bonn-Rhein-Sieg University of Applied Sciences*, Sankt Augustin, Germany, Introductory Computer Science Courses.
Assisted students with assignments and exam preparation across multiple introductory courses.

Academic Service

- Reviewing Activities Bioinformatics, Journal of Neurology, Bioinformatics Advances, Scientific Reports, BMC Medical Informatics & Decision Making, npj Dementia, International Journal of Language & Communication Disorders, European Conference on Artificial Intelligence (ECAI)
- Master's Thesis Supervision Ella Eycken – Combining Logopedics and Natural Language Processing for Neurodegeneration.
Ali Tolga Dincer – Parameter-Efficient Fine-Tuning in Large Language Models for Brain Decoding.

Invited Talks

- 2025 **Natural Language Processing in the Human Brain**, *CLiPS Colloquium*, University of Antwerp, Antwerp, Belgium.

Awards

- 2024 Best Oral Presentation – CURE-ND Early Researchers Workshop
- 2022 PhD Fellowship Fundamental Research – Research Foundation Flanders (FWO) (120,000€ over the course of 4 years)
- 2022 Studienpreis – Anwenderforum Bonn e.V. (5,000€)
- 2016–2019 Deutschlandstipendium – Bonn-Rhein-Sieg University of Applied Sciences (300€/month)
- 2018 Best Thesis Award – Bonn-Rhein-Sieg University of Applied Sciences (1,000€)
- 2017 International Study and Training Partnerships scholarship – German Academic Exchange Service/Bonn-Rhein-Sieg University of Applied Sciences (800€/month)

Publications

For an up-to-date list of all publications, please see my google scholar profile.

- 2025 Compositional Complexity in Text and Images. *Neurobiology of Language [Stage 1 In-Principle Acceptance]*. Helena Balabin, Antonietta Gabriella Liuzzi, Kevin Statz, Marie-Francine Moens, Patrick Dupont, Rik Vandenberghe.
- 2025 Evaluating Language Model Embeddings for Parkinson's Disease Cohort Harmonization Using a Novel Manually Curated Variable Mapping Schema. *Scientific Reports*. Yasamin Salimi, Tim Adams, Mehmet Can Ay, Helena Balabin, Marc Jacobs, Martin Hofmann-Apitius.
- 2025 Natural Language Processing-Based Classification of Early Alzheimer's Disease from Connected Speech. *Alzheimer's & Dementia*. Helena Balabin, Bastiaan Tamm, Laure Spruyt, Nathalie Dusart, Ines Kabouche, Ella Eycken, Kevin Statz, Steffi De Meyer, Hugo Van Hamme, Patrick Dupont, Marie-Francine Moens, Rik Vandenberghe.
- 2024 Semantic Harmonization of Alzheimer's Disease Datasets Using AD-Mapper. *Journal of Alzheimer's Disease*. Philipp Wegner, Helena Balabin, Mehmet Can Ay, Sarah Bauermeister, Lewis Killin, John Gallacher, Martin Hofmann-Apitius, Yasamin Salimi.
- 2024 A Comparative Analysis of Language Models for the Classification of Alzheimer's Disease Based on Connected Speech. *AAAI Workshop on Artificial Intelligence for Brain Encoding and Decoding*. Helena Balabin, Laure Spruyt, Ella Eycken, Ines Kabouche, Bastiaan Tamm, Jolien Schaefferbeke, Patrick Dupont, Marie-Francine Moens, Rik Vandenberghe.
- 2023 Investigating Neural Fit Approaches for Sentence Embedding Model Paradigms. *ECAI*. Helena Balabin, Antonietta Gabriella Liuzzi, Jingyuan Sun, Patrick Dupont, Rik Vandenberghe, Marie-Francine Moens.
- 2022 How Relevant Is Selective Memory Population in Lifelong Language Learning? *AACL/IJCNLP*. Vladimir Araujo, Helena Balabin, Julio Hurtado, Alvaro Soto, Marie-Francine Moens.
- 2022 Pre-trained Speech Representations as Feature Extractors for Speech Quality Assessment in Online Conferencing Applications. *Interspeech*. Bastiaan Tamm, Helena Balabin, Rik Vandenberghe, Hugo Van Hamme.
- 2022 Deep Learning-based Detection of Psychiatric Attributes from German Mental Health Records. *IJMI*. Sumit Madan, Fabian Julius Zimmer, Helena Balabin, Sebastian Schaaf, Holger Froehlich, Juliane Fluck, Irene Neuner, Klaus Mathiak, Martin Hofmann-Apitius, Pegah Sarkheil.

- 2022 STonKGs: A Sophisticated Transformer Trained on Biomedical Text and Knowledge Graphs. *Bioinformatics*. Helena Balabin, Charles Tapley Hoyt, Colin Birkenbihl, Benjamin M Gyori, John Bachman, Alpha Tom Kodamullil, Paul G Ploeger, Martin Hofmann-Apitius, Daniel Domingo-Fernandez.
- 2022 ProtSTonKGs: A Sophisticated Transformer Trained on Protein Sequences, Text, and Knowledge Graphs. *SWAT4HCLS*. Helena Balabin, Charles Tapley Hoyt, Benjamin M Gyori, John Bachman, Alpha Tom Kodamullil, Martin Hofmann-Apitius, Daniel Domingo-Fernandez.
- 2020 The COVID-19 Ontology. *Bioinformatics*. Astghik Sargsyan, Alpha Tom Kodamullil, Shounak Baksi, Johannes Darms, Sumit Madan, Stephan Gebel, Oliver Keminer, Geena Mariya Jose, Helena Balabin, Lauren Nicole DeLong, Manfred Kohler, Marc Jacobs, Martin Hofmann-Apitius.

Skills

Technical Python (huggingface, pytorch, scikit-learn); MLOps (git, docker, mlflow, pytest, hydra); fMRI scanning & processing (BIDS, psychopy, fMRIPrep, GLMSingle); Linux (HPC, slurm, remote workflows)

Languages

German Native speaker
 English Fluent
 Dutch Fluent
 Russian Conversational

Interests

Artificial Intelligence Natural Language Processing, Compositionality, Vision-Language Models, Knowledge Graphs, Neural En-/Decoding
 Neuroscience Neurolinguistics, Compositionality, Cognitive Neurology, Alzheimer's Disease, Functional MRI