Matthijs van der Lende

matthijs.vanderlende@gmail.com
Groningen, Netherlands
Website LinkedIn OpenReview Google Scholar

My research interests broadly concern how machine-learning systems learn, adapt, and make decisions. My specific interests include continual learning, reinforcement learning, probabilistic methods, and mechanistic interpretability.

Selected Highlights

2026 **Winner, AAMAS 2026 NS-Gym Competition**, for developing an agent that adapts to non-stationary environments.

2026 **2nd Place (Black-Box Track) and Scalability Award, Aletheia's Quest**, LLM deception-detection competition, representing SAIN Groningen.

2026 **Recognized as an Outstanding Reviewer, NLDL 2026.**

2025 **Spotlight, NeurIPS 2025 Mechanistic Interpretability Workshop** for work on interpretable feature steering for preference optimization.

2025 **First-author publication, NLDL 2025**, on Gaussian-process function approximation for reinforcement learning.

Selected Publications

[1] Jeremias Ferrao, **Matthijs van der Lende**, Ilija Lichkovski, and Clement Neo. "The Anatomy of Alignment: Decomposing Preference Optimization by Steering Sparse Features." In: *NeurIPS 2025 Mechanistic Interpretability Workshop*, **Spotlight**. 2025.

[2] **Matthijs van der Lende**, Matthia Sabatelli, and Juan Cardenas-Cartagena. "Interpretable Function Approximation with Gaussian Processes in Value-Based Model-Free Reinforcement Learning." In: *Proceedings of the 6th Northern Lights Deep Learning Conference (NLDL)*, Proceedings of Machine Learning Research, vol. 265, pp. 141–154. 2025.

Education

Sep 2024 – Jan 2027 (expected) **M.Sc. Artificial Intelligence, University of Groningen.**
Current grade average: **9/10**.
Master's thesis on curvature-aware optimization for continual learning, supervised by Gido van de Ven.

Sep 2020 – Jul 2024 **B.Sc. Artificial Intelligence, University of Groningen.**
Grade average: **8.4/10**.
Bachelor's thesis: *Interpretable Function Approximation with Gaussian Processes in Value-Based Model-Free Reinforcement Learning.* The resulting work was published at NLDL 2025.

Research Experience

Jul 2025 – Mar 2026 **Research Fellow, Apart Research** *Remote*
Co-authored a Spotlight paper at the NeurIPS 2025 Mechanistic Interpretability Workshop.

Industry Experience

Jul – Sep 2025 **Data Science Intern, Novigenix SA** *Lausanne, Switzerland*
Developed a text-to-SQL system for natural-language querying of internal biomedical databases, deployed on AWS.

Jul – Sep 2024 **Machine Learning Intern, Nuwa Innovation** *Groningen, Netherlands*
Worked with IMU data from a smart pen that digitizes handwriting and enables writing on any paper. Implemented a 2D displacement estimation algorithm for handwriting based on odometry using Extended Kalman Filtering and deep learning models.

Teaching Experience

Sep 2021 – Present **Teaching Assistant, University of Groningen** *Groningen, Netherlands*
Taught 150+ students in Machine Learning, Deep Learning, Reinforcement Learning, AI Engineering, and Trustworthy and Responsible AI. Gave lectures, led tutorials and lab sessions, and supervised research and industry-collaborative student projects. Nominated for Teaching Assistant of the Year (2021/2022).

Academic Service

2025 – 2026 **Reviewer** for NLDL 2026 and the EIML Workshops at ICML 2026 and NeurIPS 2025.

Oct 2024 – Oct 2026 **Programme Committee, University of Groningen.**
Master's student member for BSc/MSc Artificial Intelligence and MSc Computational Cognitive Science; Curriculum Committee member (Oct 2024–Oct 2026) and student chair (Sep 2025–Oct 2026).