# Steven Abreu

s.abreu@rug.nl | stevenabreu.com | +1-917-214-6102 | stevenabreu7 | stevenabreu7 | scholar

## Research Interests

My PhD research focuses on the intersection between machine learning and physics, in order to design and understand efficient machine learning systems.
I am interested in hardware-aware computing and machine learning, recurrent neural networks and state space models, efficient machine learning, always-on learning, and mechanistic interpretability.

## Education

Jan 2021 – Dec 2025 *(expected)* **Ph.D. Artificial Intelligence · University of Groningen** *Groningen, Netherlands*
"Developing and understanding novel co-designed hardware-aware machine learning systems"
Supervised by Herbert Jaeger & Elisabetta Chicca, funded by the Post-Digital EU project.

- **RNNs and state space models**: research on the effect of quantization and sparsity in state space models [W3, W2] for deployment on digital neural network accelerators.
- **Large-scale machine learning**: during my Google internship, I worked on knowledge distillation between LLMs and self-training of LLMs on synthetic, partially annotated, data [P3].
- **Brain-inspired computing theory**: review on programming brain-inspired hardware systems [C2], developed an open-source intermediate representation to unify existing software frameworks and hardware platforms [O2, J3, GitHub repo].
- **Novel hardware for AI**: reviews on physics-based computing with photonics [J2] and physical reservoir computers [J1]. Programming physics-based computers [O2] and brain-inspired chips [O1, J3, C2], flow cytometry using free-space optics and neuromorphic chip [W1]. Demonstrating sparsity as a scaling advantage for state space models on novel hardware, and implemented an LLM on an emerging AI accelerator from Intel.
- **Mechanistic interpretability**: activation steering in LLMs [W3], sparsity for interpretability.

Sep 2016 – Jun 2019 **B.Sc. Computer Science · Jacobs University Bremen** *Bremen, Germany*
Thesis on "Automated Architecture Design for Deep Neural Networks" (grade: 100%).
Courses in statistical modeling, machine learning, theoretical computer science, control engineering.

Aug 2018 – Dec 2018 **B.Sc. Artificial Intelligence (exchange) · Carnegie Mellon University** *Pittsburgh, USA*
Graduate-level courses in deep learning, artificial intelligence, language and statistics.

Oct 2015 – Jul 2016 **B.Sc. Physics (incomplete) · Karlsruhe Institute of Technology** *Karlsruhe, Germany*
Courses in theoretical physics, experimental physics, mathematics.

## Professional and Research Experience

May 2024 – Oct 2024 **Intel · Neuromorphic Algorithms Research Intern** *Munich, Germany*
Investigating novel efficient hardware to enable sparse and quantized machine learning models based on state space models [W2, W3] · co-design of dynamic routing architectures based on fine-grained MoEs · language modeling without matrix multiplications on the Loihi 2 chip.

Dec 2023 – Apr 2024 **Google · Student Researcher** *Kitchener, Canada*
Developed a "next action" prediction system for AR, based on the Ego4D dataset and using knowledge distillation from a large Gemini model into a smaller open-source Gemma model [P3]. Multi-agent systems for automatic evaluation of LLM-based embodied agents in VR. Research on novel efficient neural network architectures with binarized neural networks.

Sep 2022 – Dec 2022 **University of Gent · Visiting Researcher** *Gent, Belgium*
Joined the photonics group of Prof. Bienstmann to develop ML models for their optical setup.
Efficient real-time low-power ML for particle classification using an optical setup, event-based camera and neural network accelerator, optimized for accuracy, speed and efficiency [W1].

Mar 2022 – May 2022 **Institute of Neuroinformatics (ETH & UZH) · Visiting Researcher** *Zurich, Switzerland*
Joined the group of Prof. Indiveri to optimize energy and accuracy for inference on a low-power mixed-signal neuromorphic chip for real-time bio-signal processing.

Jun 2018 – Aug 2020 **Bloomberg LP · Machine Learning SWE Intern** *New York & London*
Three consecutive summer internships · developed ML models for real-time efficient pricing of complex financial derivatives · comprehensive benchmarking against higher-order Taylor series models.

Aug 2018 – Dec 2018 **Carnegie Mellon University · Research Assistant** *Pittsburgh, USA*
Delphi group of Prof. Rosenfeld · helped expand the group's epidemiological ML system to novel diseases · real-time illness tracking system using sensor fusion of multiple data signals.

Oct 2015 – Aug 2017 **arconsis GmbH · Mobile Software Developer** *Karlsruhe, Germany*
Designed and developed mobile apps for clients. Part-time during semester, full-time during breaks.

## Skills

**Programming**: Python & PyTorch (5+ years), JAX (2+ years), C and C++ (5+ years, not active).
**Open source**: Neuromorphic Intermediate Representation (maintainer), snnTorch, Lava-dl, Spyx (contributor).
**Languages**: English (fluent), German (fluent), Spanish (B2), Portuguese (A2), French (A2), Dutch (A2).

## Publications

See Google Scholar. * shows equal contribution. **J**ournal, **C**onference, **T**hesis, **W**orkshop, **P**re-print, **T**hesis.

2025 — Davide Zani, Felix Michalak, Steven Abreu, "Contextual Sparsity Makes Recurrent Language Models More Interpretable". In preparation (*February 2025*).

2025 — Steven Abreu, Sumit Bam Shrestha, Rui-Jie Zhu, Jason Eshraghian, "Energy-efficient LLMs without matrix multiplications on the Intel Loihi 2 chip". In preparation (*February 2025*).

2025 — Alessandro Pierro*, Steven Abreu*, Jonathan Timcheck, Andreas Wild, Sumit Bam Shrestha, "Advancing the Efficiency-Performance Pareto Front with Quantized State Space Models with Extreme Sparsity on Novel Hardware". In preparation (*January 2025*).

2025 — Steven Abreu, Jason Eshraghian, "Large Language Models for Ultra-Low-Power Hardware: A Review and Tutorial on Algorithm-Hardware Co-Design". Under review.

2024 W4 Joris Postmus*, Steven Abreu*, "Activation steering in large language models using compositional and semantically interpretable steering matrices". *NeurIPS Workshop (MINT).*

2024 W3 Steven Abreu*, Jens Pedersen*, Kade Heckel*, Alessandro Pierro, "Q-S5: Towards Quantized State Space Models". *International Conference for Machine Learning (ICML) Workshop (NGSM).*

2024 W2 Alessandro Pierro*, Steven Abreu*, "Mamba-PTQ: Outlier Channels in Recurrent Large Language Models.". *International Conference for Machine Learning (ICML) Workshop (ES-FOMO-II).*

2024 P3 Steven Abreu, Tiffany Do, Karan Ahuja, Eric Gonzalez, Lee Payne, Daniel McDuff, Mar Gonzalez-Franco, "PARSE-Ego4D: Personal Action Recommendation Suggestions for Egocentric Videos". Pre-print.

2024 J3 Jens Pedersen*, Steven Abreu*, et al. *(15 co-authors)*, "Neuromorphic Intermediate Representation: A Unified Instruction Set for Interoperable Brain-Inspired Computing". Accepted to *Nature Communications* (preprint).

2024 C2 Steven Abreu, Jens Pedersen, "Neuromorphic Programming: Emerging Directions for Brain-Inspired Hardware". *International Conference on Neuromorphic Systems (ICONS).*

2024 P4 Riccardo Bovo, Steven Abreu, Karan Ahuja, Eric J. Gonzalez, Li-Te Cheng, Mar Gonzalez-Franco, "EmBARDiment: an Embodied AI Agent for Productivity in XR". Pre-print.

2024 J2 Steven Abreu, et al. *(27 co-authors)*, "A Perspective on Computing with Physical Substrates". *Reviews in Physics*, Volume 12.

2024 P2 Guillaume Pourcel, Mirko Goldman, Steven Abreu, Miguel Soriano, "Two-shot learning of continuous interpolation using a conceptor-aided recurrent autoencoder". Pre-print.

2023 C1 Steven Abreu, "Developing a Framework for Programming Physical Computing Systems". *International conference on neuromorphic, natural and physical computing (NNPC).*

2023 W1 Steven Abreu, Muhammed Gouda, Alessio Lugnan, Peter Bienstman. "Flow cytometry with event-based vision and spiking neuromorphic hardware". *CVPR 2023 Workshop* (link).

2022 J1 Matteo Cucchi*, Steven Abreu*, Giuseppe Ciccone, Daniel Brunner, Hans Kleemann. "Hands-on Reservoir Computing: A Tutorial for Practical Implementation". *Neuromorphic Computing and Engineering*, Volume 2 (link). *Selected by journal as "Highlights of 2022" (link).*

2019 T1 Steven Abreu, "Automated Architecture Design for Deep Neural Networks". *B.Sc. thesis*, Jacobs University. Accessible via arXiv.

In 2022, my research was featured in the University's news magazine as making progress towards on-device efficient machine learning: "Working towards personalized intelligent computers", *University of Groningen (link).*

## Invited Talks

2024 O3 "Building blocks of efficient machine learning systems". **Machine learning for theories and theories of machine learning**, Rovinj, Croatia (link).

2024 O3 "Quantization and Sparsity in State Space Models". **Telluride neuromorphic engineering workshop**, Telluride, Colorado, USA (link).

2023 O2 "Neuromorphic intermediate representation - toward a common representation for physical computing". **Frontiers of Neuromorphic Computing**, Max Planck Institute for the Science of Light (link).

2022 O1 "Programming Physical Computers". **International School of Solid State Physics**, 82nd Workshop on Unconventional Computing (link).

## Awards

| | |
|---|---|
| Nov 2022 | **Research Member (Co-PI)**. Intel Neuromorphic Research Community. |
| Oct 2022 | **TA of the Year**. AI department, University of Groningen. |
| May 2020 | **Nanodegree** in 'Deep Reinforcement Learning', Udacity. |
| Jun 2019 | **Dean's Prize for Best Undergraduate Thesis**. School of Computing and Mathematics. |
| Oct 2016 | **Conrad Naber Scholar** Full scholarship of approx. 60,000 EUR. |
| Jan 2016 | **First Place**: Google x Udacity Android developer challenge. |

## Academic Activities

| | |
|---|---|
| Jan 2023 - present | Editor for NeuroPAC (accelerating neuromorphic research). |
| Jan 2022 - present | Thesis supervisor (8 B.Sc. students, 3 M.Sc. students), University of Groningen. |
| Jan 2021 - present | Teaching Assistant for "Machine Learning", "Neural Networks", University of Groningen. |
| Sep 2017 - Jun 2019 | Teaching Assistant for "Introduction to Computer Science", "Algorithms & Data Structures", "Machine Learning", "Statistical Modeling (M.Sc.)", Jacobs University Bremen. |

## Volunteering

Oct 2021 **Board member · PhD Day 2023** *Groningen, Netherlands*
Apr 2023 Head of sponsorship team for the PhD Day, with over 800 attendees. Raised over 25,000 EUR.

Aug 2021 **Chair & Co-founder · Science Communication Writing Club** *Groningen, Netherlands*
Mar 2023 Organized workshops, edited articles (link). Co-organized 3-Minute-Thesis 2022 (link).

Jan 2019 **Chair · Photography Club** *Bremen, Germany*
Jul 2019 Lead the university's film photography club. Taught film photography and darkroom development.