

Johan Bjorck

Kirkland, WA

[nilsjohanbjorck@gmail.com](mailto:nilsJohanBjorck@gmail.com) | [linkedin.com/in/johan-bj%C3%B6rck-b9a6bb65/](https://www.linkedin.com/in/johan-bj%C3%B6rck-b9a6bb65/) | [Google Scholar](#) | Citations: 10560

Work Experience

Nvidia, GEAR team at Nvidia Research

Redmond, WA

Senior Research Scientist, Multimodal Foundation Models and Robotics

October 2024 - Current

- **Core contributor to Gr00t-N model series.** I have contributed to all models (1, 1.5, 1.6, and 1.7) in the Gr00t-series of models. My primary focus has been on pretraining the VLM backbone and developing instruction-following capabilities in the VLA. I led the pretraining of 1.5, as covered in this [\[blog post\]](#). Our [\[paper\]](#) has been **cited >900 times** and the [\[model weights\]](#) have been downloaded >100,000 times.
- **Technical Lead, VLMs.** I am the technical lead for VLM training at GEAR. This involves technical work like training models, producing training data, and designing evaluation. It also involves setting goals, recruiting and mentoring interns, and interfacing with external teams. We recently released our VLM [\[Vesta\]](#) which is **SOTA in both navigation, embodied reasoning, and real robot inference**.

Microsoft, Turing/GenAI

Bellevue, WA

Senior Applied Scientist

February 2022 - October 2024

- **Co-lead for Phi-3 small training.** I was the co-lead for training the 7B Phi-3 small state-of-the-art (SOTA) LLM released by Microsoft. This included selecting data, balancing multilingual capabilities, benchmarking and optimizing training throughput, and scaling up training to 1024 H100 GPUs. The model achieves SOTA results, e.g. reaching 75.5 MMLU and thus **beating Llama3, Gemma and Mixtral**. We released a [\[blog post\]](#), a [\[paper\]](#) and the [\[model weights\]](#). The 8k version of Phi-3 small has been downloaded >75,000 times in the last month.
- **LLM data.** I worked on cleaning web data from common crawl for training LLMs. I primarily worked on heuristic cleaning like deduplication and line-wise edits, scaling up our pipeline to **5 trillion tokens and >10,000 CPU cores**. After benchmarking we found that our internal dataset beat the public refinedweb dataset in small scale-ablations, e.g. giving a boost of 2.4 MMLU. This data was used for training all Phi-3 models.
- **Shipped multimodal model.** I led the development of a large CLIP-style multimodal model, from data sourcing to training and evaluation. The model was trained on billions of images using >1000 A100 GPUs and released internally in base and large size. Our model shipped to multiple scenarios in Microsoft: Bing text-to-image retrieval, Bing image-to-image retrieval, Ads retrieval, Xbox content moderation and more. In offline evaluation on production data it **improved DCG by [redacted] (English) and [redacted] (multilingual)**. We also released a [\[blog post\]](#) on the model.
- **Multimodal research.** Together with our partner team at MSRA I researched the use of masked training of vision-language models, similar to BERT but with tokenized images and text. I worked on pretraining, multilingual data, and object detection. In our [\[paper\]](#) based upon this idea we achieved SOTA results, e.g. beating **SOTA VQA by 1.7 acc** and **SOTA object detection by 0.4 AP**. In our follow-up [\[paper\]](#) we further demonstrated zero-shot and in-context capabilities for a decoder-only multimodal model.

Education

Cornell University

Ithaca, US

Ph.D. Computer Science, Committee: Kilian Weinberger, Bart Selman, and Carla Gomes

August 2015 - December 2021

- **Reinforcement Learning.** Towards the end of my PhD I was working mostly on reinforcement learning, specifically, RL applied to robotics simulated via MuJoCo. I was looking at new angles on this problem: in one [\[paper\]](#) I demonstrated that RL agents could be trained end-to-end in low precision. In another [\[paper\]](#) I showed that the high variance that RL is famous for can be decreased >3x with a few simple tricks. I also showed that spectral normalization allowed the scaling of RL models in this [\[paper\]](#).
- **Fundamentals of Deep Learning.** In the late 2010s Deep Learning was a "black-art", and some of my early research aimed to develop a principled scientific understanding of some techniques in the field. My first [\[paper\]](#) in this direction investigated batch normalization, and showed that training stability rather than "internal covariate shift" was responsible for the efficacy of batch normalization. I've also published a [\[paper\]](#) that explains how AdamW outperforms Adam due to gradient pollution.
- **Machine Learning for Science.** I researched the use of machine learning in materials science. I proposed a method for modeling X-ray crystallography measurements as a matrix factorization problem with decomposed constraints which enables solving the problem quickly. This method was used in multiple [\[papers\]](#) [\[papers\]](#) [\[papers\]](#) which were published in collaboration with materials science researchers. I also demonstrated how CNN-based neural networks could be used for analyzing elephant calls for conservation science in this [\[paper\]](#).

Chalmers University of Technology

Gothenburg, Sweden

B.S Engineering Physics, GPA : 4.9/5

August 2012 - June 2015

Selected Publications

Vesta: A Generalist Embodied Reasoning Model

Technical Report 2026

Johan Bjorck, Zhiqi Li, Yunze Man, ..., and others

World Action Models are Zero-shot Policies

In Submission 2026

Seonghyeon Ye, Yunhao Ge, ..., **Johan Bjorck**, and others

Gr00t n1: An open foundation model for generalist humanoid robots

Technical Report 2025

Johan Bjorck, Fernando Castañeda, Nikita Cherniadev, ..., and others

Phi-3 Technical Report: A Highly Capable Language Model Locally on Your Phone

Technical Report 2024

Marah Abdin, Sam Ade Jacobs, ..., **Johan Bjorck**, and others

Language Is Not All You Need: Aligning Perception with Language Models

NeurIPS 2023

Shaohan Huang, Li Dong, ..., **Johan Bjorck**, and others

Image as a Foreign Language: BEiT Pretraining for Vision and Vision-Language Tasks

CVPR 2023

Wenhui Wang, Hangbo Bao, Li Dong, **Johan Bjorck**, Zhiliang Peng, Qiang Liu, Kriti Aggarwal, Owais Khan Mohammed, Saksham Singhal, Subhojit Som, Furu Wei

Is High Variance Unavoidable In RL? A Case Study In Continuous Control

ICLR 2022

Johan Bjorck, Carla Gomes, Kilian Q. Weinberger

Towards Deeper Deep Reinforcement Learning with Spectral Normalization

NeurIPS 2021

Johan Bjorck, Carla Gomes, Kilian Q. Weinberger

Low-Precision Reinforcement Learning: Running Soft Actor-Critic in Half Precision

ICML 2021

Johan Bjorck, Xiangyu Chen, Christopher De Sa, Carla P. Gomes, Kilian Q. Weinberger 1

Understanding Batch Normalization

NeurIPS 2018

Johan Bjorck, Carla Gomes, Bart Selman, Kilian Q. Weinberger

Miscellaneous

Peer review NeurIPS, ICLR, ICML, CVPR, AAAI, JLMR.

Fulbright Scholar Merit-based scholarship. I was the youngest Swedish recipient of 2015.

Telluride Fellow Merit-based fellowship conferring room and board at the Telluride house at Cornell.