

Vaibhav Singh

+1 438-866-1085 | vaibhav.singh@mila.quebec | [Website](#) | [GitHub](#) | [Google Scholar](#) | [LinkedIn](#)

Education

Ph.D. in Computer Science , Mila, Montreal, Canada Advisor: Dr. Eugene Belilovsky	Sept 2023 - Present GPA: 4.3/4.0
M.S. in Computer Science , New York University (NYU), USA Thesis advisor: Dr. Anna Choromanska	Sept 2021 - May 2023 GPA: 3.96/4.0

Research Experience

Research Intern , ServiceNow Research, Montreal, Canada – Designed <i>DiffuMamba</i> (ICML 2026): a bidirectional Mamba-2 backbone (with a hybrid attention/SSM variant) for masked diffusion language models, scaled to 1.3B parameters; matched Transformer-based diffusion LMs on downstream zero-shot tasks while delivering up to $8.2\times$ higher long-context inference throughput.	Jul 2025 - Feb 2026
M.S. Thesis , under Dr. Anna Choromanska, NYU, New York, USA – Developed a <i>decoupled wake/sleep contrastive</i> training algorithm for Energy-Based Models in class incremental Continual Learning, leveraging replay-free Langevin dynamics sampling; yielded a CVPR 2024 CLVISION workshop paper. [Thesis]	Sept 2022 - May 2023
Research Assistant under Dr. Vinayak Abrol , IIIT-Delhi, India – Analyzed spectral and temporal feature integration in end-to-end deep acoustic models (Low-Rank Convolutional encoder with RNN-Transducer decoder; raw waveform vs. spectrogram inputs) via loss landscape visualization and gradient attribution; published at NeurIPS 2022 ENLSP workshop.	Mar 2021 - Sept 2021

Professional Experience

Research Assistant , Applied AI Institute, Quebec, Canada – Built a fully open-source Retrieval-Augmented Generation pipeline for the Quebec Cybersecurity Ministry using LlamaIndex, ChromaDB, and Ollama deployable in air-gapped government environments.	Jul 2024 - Oct 2024
Teaching Assistant , Mila / Concordia / NYU – Representation Learning (IFT-6150) – Deep Learning (COMP-691) – Artificial Intelligence (CS-GY 6613) – Designed and graded assignments, held weekly office hours, and led recitation sessions for graduate-level ML courses.	Jan 2022 - Dec 2024
Data Science Intern , Spectrum Labs, San Francisco, USA – Benchmarked token-level (RNN/LSTM/GRU), sentence-level (Graph CNN), and subword-level (XLM-RoBERTa) models for toxicity classification across multiple datasets; improved the production classifier by +5.0 precision and +5.6 recall.	Jun 2022 - Aug 2022
Software Engineer (ML, Voice Team) , Flipkart, Bangalore, India – Built a probabilistic context-verification model in Python over Flipkart’s 30K-product grocery catalog to support voice queries; implemented a character-level CNN for query classification and a filter-parsing technique for fashion category filters (F1 = 87.43% on 20K test queries). Shipped to the Flipkart mobile app.	Jul 2019 - Jan 2021

Peer-Reviewed Publications

Vaibhav Singh indicates first author. * denotes equal contribution.

Conferences (Main Track)

- **Vaibhav Singh**, Ostapenko, O., Noël, P.-A., Belilovsky, E., & Scholak, T. (2026). DiffuMamba: High-Throughput Diffusion LMs with Mamba Backbone. *International Conference on Machine Learning (ICML)*. [\[link\]](#)
- **Vaibhav Singh**^{*}, Janson, P.^{*}, Mehrbod, P., Ibrahim, A., Rish, I., Belilovsky, E., & Thérien, B. (2025). Beyond Cosine Decay: On the Effectiveness of Infinite Learning Rate Schedule for Continual Pre-training. *Conference on Lifelong Learning Agents (CoLLAs)*, Oral. [\[link\]](#)

Journals

- **Vaibhav Singh**, Aljundi, R., & Belilovsky, E. (2026). Dual-Phase Continual Learning: Supervised Adaptation Meets Unsupervised Retention. *Transactions on Machine Learning Research (TMLR)*. [\[link\]](#)

Workshops

- **Vaibhav Singh**, Belilovsky, E., & Aljundi, R. (2025). When Data Falls Short: Grokking Below the Critical Threshold. *NeurIPS 2025 Workshop on Continual and Compatible Foundation Model Updates (CCFM)*. [\[link\]](#)
- **Vaibhav Singh**, Khalid, Z., Oyallon, E., & Belilovsky, E. (2025). Model Parallelism with Subnetwork Data Parallelism. *ICML 2025 ES-FoMo Workshop*. [\[link\]](#)
- **Vaibhav Singh**, Choromanska, A., Li, S., & Du, Y. (2024). Wake-Sleep Energy-Based Models for Continual Learning. *CVPR 2024 CLVISION Workshop*. [\[link\]](#)
- **Vaibhav Singh**, Abrol, V., & Nathwani, K. (2022). On Spectral and Temporal Feature Encoding Behaviour in Stacked Architectures. *NeurIPS 2022 ENLSP Workshop*. [\[link\]](#)

Earlier Publications

- Kumar, A., **Vaibhav Singh**, Ali, T., Singh, J., & Pal, S. (2019). Evaluation of Shallow and Deep Classifiers for Rumor Detection. *Advances in Computing and Intelligent Systems*, Springer. [\[link\]](#)
- **Vaibhav Singh** & Sharma, K. (2019). Analysis of Shallow and Deep Architecture Classifiers on Emotion Recognition from Speech. *6th IEEE International Conference on Cyber Security (CSCloud)*. [\[link\]](#)

Selected Projects

Distributed Image Enhancement Python, TensorFlow, Ray – Encoder-decoder for super-resolution trained on a Ray cluster; PSNR 23.1, SSIM 0.72.	Spring 2022
Retail Clickstream Analysis & Prediction Python, PySpark, Streamlit – Distributed analysis of 14 GB e-commerce clickstream; deployed real-time classifier (78.4% accuracy).	Spring 2022
Topological Analysis of Prompts in CLIP Python, PyTorch – Mapper-based prompt selection algorithm; improved CLIP zero-shot performance by ~1.3% over standard prompt ensembling.	Spring 2022
Quantum Convolutional Neural Network Python, PennyLane – Independent study: 2-layer quantum circuit on Google Speech Commands - 58.2% accuracy.	Fall 2022

Fellowships, Awards, and Honours

Concordia University Doctoral Fellowship (CAD \$28,000), Concordia University	Sept 2023
NYU Dean's Merit Scholarship (USD \$16,000) for M.S. in Computer Science, New York University	Aug 2021
'Reality Stone' Award for outstanding contribution to the Context Management Service, Flipkart	Sept 2020