ANDREI MURESANU

andrei.muresanu@uwaterloo.ca | +1 647-213-2129 | Google Scholar | LinkedIn | GitHub | AndreiMuresanu.com

PROFESSIONAL EXPERIENCE

Research Scientist, with Prof. Zhijing Jin, Vector Institute, Toronto, ON **January 2025 – Present**
- Leading my own research agenda on formally defining and evaluating memory for AI agents

Research Scientist, with Prof. Gillian Hadfield, Vector Institute, Toronto, ON **January 2024 – December 2024**
- **2 papers** (NeurIPS, ICLR) in a **largely unexplored field** pioneering **AI safety** via normative reasoning for human-AGI collaboration
- **Designed core** normative module and ran all experiments, achieving a **30%** boost in group norm identification and compliance
- Built a **fully expressive** text environment creation **library** with multi-agent support, reducing setup time from **5 days to 4 hours**

Research Scientist, with Prof. Nicolas Papernot, Vector Institute, Toronto, ON **May 2023 – February 2024**
- **First author** of ICML 2025 paper, proposing prompting for exact unlearning and a **new** holistic unlearning cost metric
- Designed and ran all experiments, achieving a **99.99% reduction** in exact unlearning cost vs state-of-the-art in language settings

Research Scientist, with Prof. Animesh Garg, Vector Institute, Toronto, ON **May 2023 – December 2023**
- **Built the fastest** 3D memory benchmark (**43% faster**) and the **most extensive** memory test, supporting 3 modalities and 9 tasks
- Found research gap in memory evaluation, proposed solution, wrote positioning; **first author** on TMLR submission (**top journal**)
- **Conceptualized core idea**/algorithm for **3** projects: formal memory definition, hyperbolic geometry for memory, and text-to-sim
- Conducted **5000+** experiments, using **20,000 GPU hours** across **24 GPUs**, benchmarking memory for 4 popular RL algorithms

Research Scientist, with Prof. Jimmy Ba, Vector Institute, Toronto, ON **August 2022 – April 2023**
- **First author** of ICLR paper **(1000+ citations, best paper, oral)**, proposed AI safety use and **70%** efficiency boost via binary search
- Set up **Vector's first** multi-node framework to fine-tune LLMs with **10+ billion parameters**. Edited LLM inference code using **Jax**
- Conducted **7+ day continuous** LLM fine-tuning runs, launched **2500+** experiments, and ran large-scale hyperparameter searches

AI Team Lead & CTO, WealthyPlanet, Toronto, ON **April 2023 – Present**
- Directed a **12-developer** team in engineering **Canada's best** personal finance engine, saving users **$100k by retirement**
- Defined R&D agenda, growing company from concept to beta with **100** customers, increasing valuation from **$3M to $20M**
- Managed a **40k line** full-stack codebase supporting **100k** users/month with a test framework validating a **superhuman** optimizer

Principal Investigator, Silera.ai, Montreal, QC **January 2025 – Present**
- Setting research agenda and **managing a team of 4** developers to create an automated hyperrealistic synthetic data platform

Principal Investigator, Triomics, San Francisco, CA **January 2024 – May 2024**
- **Managed a team of 5** to automate cancer trial eligibility using LLMs, reducing doctor processing time by **10 hours** per patient
- Created an MVP **20%** more accurate and **90%** cheaper than the initial prototype, featured in **The Globe and Mail** (6M readers)

Computer Vision Research Engineer, NVIDIA, Santa Clara, CA **January 2022 – April 2022**
- Served as lead developer of **Nvidia's safety-critical** FaceID system to be used in **2.2 million** luxury vehicles worldwide
- Cut cleaning time of **0.5 billion** images from **50 days to 17 hours** by developing a **new** stochastic dominant identity algorithm
- Attained a **2.3x speed-up** by redesigning the face-matching module in Meta AI's DeepFace, a library used by **50,000+ developers**
- Matched leading FaceID results within **3%** accuracy using unclean images; bias analysis showed **94%** certainty across **4** attributes

Research Scientist, under Prof. Chul Min Yeum, University of Waterloo, Waterloo, ON **September 2021 – December 2021**
- Initiated development of the **world's first** autonomous flood risk analysis system, projected to save **$10M** annually in Canada
- Proposed a photorealistic synthetic data generation system and developed a **96%** in-the-wild accurate door detection module

Machine Learning Engineer, Advanced AI & Analytics Research Team, PerkinElmer, Waterloo, ON **May 2021 – August 2021**
- Pioneered a new deep learning approach and developed a global state-of-the-art mass spectrometry analysis software
- Helped design an **original** confidence metric that was 95%+ accurate, making the model's results more interpretable

Machine Learning Engineer, Geminare, Toronto, ON **May 2020 – December 2020**
- Conceptualized an **original** object-detection process that trained on limited and mislabeled data, **saving $15,000** over 3 months
- Scaled the system to process **250,000+** images using GCP and AWS, reaching **20,000** monthly users across North America
- Designed the core algorithm for a motion-analysis app, allowing table tennis players to compare techniques with professionals

PUBLICATIONS

Large Language Models Are Human-Level Prompt Engineers

Yongchao Zhou*, **Andrei Muresanu***, Ziwen Han*, Keiran Paster, Silviu Pitis, Harris Chan, Jimmy Ba
International Conference on Learning Representations (ICLR). 2023
Best Paper Award at NeurIPS 2022 ML Safety Workshop
Oral presentation at NeurIPS 2022 Foundation Models for Decision Making Workshop
1003 citations as of February 27, 2025

Benchmarks for Physical Reasoning AI

Andrew Melnik, Robin Schiewer, Moritz Lange, **Andrei Muresanu**, Mozhgan Saeidi, Animesh Garg, Helge Ritter
Transactions on Machine Learning Research (TMLR). 2023
Awarded the **Exceptional Survey Certificate**

Unlearnable Algorithms for In-context Learning

Andrei Muresanu, Anvith Thudi, Michael Zhang, Nicolas Papernot
International Conference on Machine Learning (ICML). 2025

Normative Modules: A Generative Agent Architecture for Learning Norms that Supports Multi-Agent Cooperation

Atrisha Sarkar, **Andrei Muresanu**, Carter Blair, Aaryam Sharma, Rakshit S Trivedi, Gillian K Hadfield
Submitted to the Conference on Neural Information Processing Systems (NeurIPS). 2024
Accepted to the workshop on Foundation Models and Game Theory (FMGT). 2024

Altared Environments: The Role of Normative Infrastructure in AI Alignment

Rakshit Trivedi, Nikhil Chandak, **Andrei Muresanu**, Shuhui Zhu, Atrisha Sarkar, Joel Leibo, Dylan Hadfield-Menell, Gillian Hadfield
Submitted to the International Conference on Learning Representations (ICLR). 2025

In Preparation:

ReMEMber: Defining And Assessing The Memory Capabilities Of AI Agents

Andrei Muresanu, Marta Skreta, Kourosh Darvish, Alán Aspuru-Guzik, William Cunningham, Animesh Garg
Submitted to the Conference on Neural Information Processing Systems (NeurIPS). 2024
Will be resubmitted to Transactions on Machine Learning Research (TMLR). 2025

Which LLM Parameters Are Important For Fine-Tuning?

Andrei Muresanu*, Ziwen Han*, Jimmy Ba
Will be submitted to workshops at the International Conference on Machine Learning (ICML). 2025

Using Hyperbolic Geometry To Create Memory With Self-Assembling Hierarchical Structure

Andrei Muresanu*, Marta Skreta*, Kourosh Darvish, Alán Aspuru-Guzik, Animesh Garg
Will be submitted to Transactions on Machine Learning Research (TMLR). 2025

EDUCATION

Bachelor of Computer Science — **September 2019 – April 2024**
University of Waterloo — Waterloo, Ontario, Canada
Recipient of Research Certificate

Bachelor of Statistics — **April 2024 – February 2025**
University of Waterloo — Waterloo, Ontario, Canada

SELECTED PROJECTS

Superhuman Poker AI — **March 2021 – March 2022**

- Recreated Facebook AI's 2019 "Pluribus" project **from scratch** and corrected 5+ errors in one of the supporting papers

Unity Neural Network Library — **February 2019 – April 2019**

- Constructed the **first-ever** Unity neural network library from scratch, used to create backpropagation neural networks

Indie Game Developer — **September 2013 – June 2019**

- Built **40+ games over 6 years**. Released on desktop, mobile, and in the browser. Primarily developed with Unity in C#

AWARDS, FELLOWSHIPS, & GRANTS

- (2025) ($17,500) Vector Master's Scholarship (**1 of 2** recipients for the Computer Science category at the University of Toronto)
- (2024) ($8,000) Vector Institute Research Grant
- (2024) ($33,000) Vector Institute Research Grant
- (2023) ($8,000) Georgia Tech Research Grant
- (2023) ($8,000) University of Toronto Research Grant
- (2022) ($7,500) Vector Institute Research Grant
- (2021) ($1,000) University of Waterloo Undergraduate Research (URA) Grant
- (2019) ($2,000) University of Waterloo President's Entrance Scholarship

Competition Awards:

- (2019) Top 20 finalist (Top 0.00045%) in C1 Terminal International AI Programming Competition
- (2019) 2nd Place in Toronto Police Hackathon. Presented our idea to the mayor in a televised board meeting
- (2019) Won 1st place and $5,000 in the DMZ Basecamp pitch competition as co-founder of a non-invasive insulin patch startup

SERVICES

- Reviewer for the Association for Computational Linguistics (ACL) international conference 2025
- NLP and LLM Workshop Lead at MacHacks | McMaster University 2023
- Computer Vision Workshop Lead at MacHacks | McMaster University 2022

ADDITIONAL INFORMATION

- **Coding Languages:** Proficient: **C**, **C#**, **C++**, **Python**, **MIPS**, **ARM,** and **Scheme/Lisp**; working knowledge: **SQL**, **R**, **MATLAB**, and **Java**
- **Languages**: Fluent in English and Romanian; professional working proficiency in French
- **Skills:** Git, Docker, NumPy, SciPy, Pandas, OpenCV, CUDA, Scikit-Learn, CNN, data mining, data visualization, computer vision, web scraping, big data, data analytics, deep learning, GPU, parallel programming, simulation, reinforcement learning, PyTorch, TensorFlow, algorithms, GCP, Azure, AWS