Jonathan Colaço Carr

jonathan.colacocarr@mail.mcgill.ca | j-c-carr.github.io | linkedin.com/in/jonathan-colaco-carr

Education

McGill University/Mila, M.Sc in Computer Science 2023 – 2025
- Supervisors: Prakash Panangaden and Doina Precup

Stanford University, Visiting Graduate Student 2024 – 2025
- Supervisor: Benjamin Van Roy

McGill University, B.Sc in Honours Mathematics 2018 – 2023
- Minors in Computer Science and Physics

Awards

MITACS Globalink Research Award, $6,000 Sep 2024
Research award for Canadian graduate students studying abroad, determined by expected research quality.

McGill Graduate Mobility Award, $8,900 Sep 2024
Scholarship for McGill graduate students studying abroad, determined by academic standing.

NSERC Canada Graduate Scholarship – Master's, $17,500 May 2023
Graduate scholarship determined by academic excellence, research potential, and interpersonal skills.

McGill Computer Science Undergraduate Research Award, $7,000 May 2022
Research award determined by academic record and research aptitude.

McGill Space Institute Summer Undergraduate Research Award, $7,000 May 2021
Research award determined by academic record and extracurricular leadership.

RBC x Microsoft AI for Social Impact Challenge, 2nd Place Prize, $2,000 Apr 2020
National AI competition prize determined by originality, technical excellence, impact, and feasibility.

Preprints and Publications

* denotes co-first authorship.

[1] **J. Colaço Carr**, P. Panangaden, D. Precup, and B. Van Roy, *A computationally tractable extension of Nash learning from human feedback to sequential decision making*, 2025 (preprint), to be submitted to the International Conference on Machine Learning (ICML).

[2] **J. Colaço Carr**, "Learning from human preferences in sequential decision making: An economic perspective," M.S. thesis, McGill University, 2025, to be submitted to the Canadian AI Association's Masters' Thesis competition.

[3] **J. Colaço Carr**, Q. Sun, and C. Allen, "Focused skill discovery: Learning to control specific state variables while minimizing side effects," in *Reinforcement Learning Journal (RLJ)*, 2025, acceptance rate: 39.0%. [project page], [pdf].
– Preliminary version presented as an extended abstract at RLDM 2025.

[4] K. Chehbouni*, **J. Colaço Carr***, Y. More, J. C. Cheung, and G. Farnadi, "Beyond the safety bundle: Auditing the Helpful and Harmless dataset," in *Proceedings of the Annual Conference of the Nations of the Americas Chapter of the ACL (NAACL)*, 2025, acceptance rate: 22.6%, **oral presentation.** [pdf].
– Preliminary version selected for a spotlight presentation at the NeurIPS 2024 workshop on Algorithmic Fairness through the Lens of Metrics and Evaluations.

[5] **J. Colaço Carr**, P. Panangaden, and D. Precup, "Conditions on preference relations that guarantee the existence of optimal policies," in *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2024, acceptance rate: 27.6%. [pdf].

[6] J. Kennedy, **J. Colaço Carr**, S. Gagnon-Hartman, A. Liu, J. Mirocha, and Y. Cui, "Machine-learning recovery of foreground wedge-removed 21-cm light cones for high-z galaxy mapping," in *Monthly Notices of the Royal Astronomical Society (MNRAS)*, 2024, [pdf].

Research Experience

Mila May 2022 – present

- Supervisors: Prakash Panangaden, Doina Precup, and Golnoosh Farnadi
- Used human models from economics to design sequential decision-making algorithms aligned with human goals.
- Audited a human preference dataset for label quality, topic coverage, and changes in LLM behaviour after fine-tuning.

Stanford University Oct 2024 – Mar 2025

- Supervisor: Benjamin Van Roy
- Extended the maximal lottery social choice rule to sequential decision-making problems and designed game-theory algorithms for computing solutions. Implemented experiments in JAX.

UC Berkeley (Center for Human-Compatible AI) Jun 2024 – Sep 2024

- Supervisors: Cameron Allen and Stuart Russell
- Developed a simple method to improve skill discovery using state abstraction, benefiting both efficiency and safety.

McGill Space Institute May 2021 – Dec 2021

- Supervisor: Adrian Liu
- Implemented computer vision algorithms to recover signals from noisy telescope data.

Work Experience

Hortus AI, Lead Software Developer Dec 2024 – Dec 2025

- Led a team of developers (15 in total) to build a consumer reports platform for AI tools. Reported directly to startup founder, Thomas Krendl Gilbert. Platform now used by over 100 public officials across the United States.
- Designed algorithms to rank and evaluate AI tools in collaboration with the GovAI coalition.
- Wrote a policy brief on how to incentivize vendors to develop AI responsibly during AI procurement.

The Democratic Engagement Exchange, Pilot Researcher Aug 2025 – Oct 2025

- Conducted a feasibility study on the use of personalized AI assistants for voter education in Canada.

McGill University, Teaching Assistant Jan 2024 – Apr 2024

- Held weekly office hours and created assignments for a graduate-level course in machine learning (COMP 551).

AltaML, Machine Learning Intern Jan 2022 – Apr 2022

- Implemented machine learning algorithms to identify cross-selling opportunities.

Extracurricular Experience

McGill AI Safety x Law Group, Co-founder Apr 2022 – Dec 2023

- Co-founded and taught an eight-week course on AI safety for McGill Law students.

RBC x Microsoft AI for Social Impact Challenge, Team Lead Sep 2019 – Apr 2020

- Led a four-student team to design a textile recycling robot for an AI social impact challenge. Placed 2nd out of 167 teams from across Canada.

NASA Space Robotics Challenge (Phase 2), Team Lead Jul 2019 – May 2020

- Led a five-student team to design simulated localization and mapping algorithms for a space robotics competition.

Personal Interests

- Sewing: 10+ years sewing and organizing local workshops.
- Outdoor activities: Planted trees in Northern Canada for two summers (about 350,000 trees in total). Organized hiking trips for the McGill Outdoors Club.