

# Andrei Mircea Romascanu

<https://mirandrom.github.io> [mirandrom+cv@pm.me](mailto:mirandrom+cv@pm.me)

## About me

I'm passionate about enabling scientific progress with machine learning by making information more accessible. My current research centers on large language models: understanding how they represent scientific knowledge via mechanistic interpretability and enabling continual learning from the scientific literature. Also interested in HCI aspects of large language model assisted scientific research.

## Publications

- [1] **Human-Centered Information Extraction from Scientific Publications: A Case Study at the Intersection of Additive Manufacturing and ML.** 2024  
[A. Mircea\\*](#), M. Safdar, J. Xie, Y. Zhao. *In submission EACL 2024.*
- [2] **Reviewing Peer Review: Challenges and Insights from Area Chairs.** 2024  
I. Arous, N. Kennard, [A. Mircea\\*](#), A. McCallum, J. Cheung. *In submission CHI 2024.*
- [3] **Language Model Pretraining with Lexical Semantics.** [A. Mircea\\*](#), and J. Cheung. *Findings of the Association for Computational Linguistics: EMNLP 2023.* 2023
- [4] **Discourse-Aware Unsupervised Summarization for Long Scientific Documents.** 2021  
Y. Dong\*, [A. Mircea\\*](#), and J. Cheung. *EACL 2021.*  
[Paper](#) | [Code](#)
- [5] **Bridging the gap between supervised classification and unsupervised topic modelling for social-media assisted crisis management.** 2021  
M. Brunila\*, R. Zhao\*, [A. Mircea\\*](#), S. Lumley, R. Sieber. *Proceedings of the Second Workshop on Domain Adaptation for NLP.*  
[Paper](#) | [Code](#)
- [6] **Real-time Classification, Geolocation and Interactive Visualization of COVID-19 Information Shared on Social Media to Better Understand Global Developments.** [A. Mircea\\*](#). *Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020.* 2020  
[Paper](#) | [Code](#)
- [7] **Using deep learning and social network analysis to understand and manage extreme flooding.** 2020  
[A. Mircea\\*](#), H. Ker\*, R. Sieber\*, S. Greenidge, S. Lumley, D. Bush, S. Morgan, R. Zhao, M. Brunila. *Journal of Contingencies and Crisis Management* 28.3 (2020): 251-261.  
[Paper](#) | [Code](#)

## Open source and hackathons

- PyS2** Created open-source python library for working with the semantic scholar API with pydantic typing. [Code](#)
- CrisisTweetMap (McHacks 2020 winner)** Created dashboard to live classify and map tweets about covid. [Code](#)
- ArXivNet (McGill CodeJam 2019 winner)** Created interface for academic paper search/discovery in 3D. [Code](#)
- ImplementAITA (Implement AI 2019 contestant)** Created text classifier based on AITA subreddit posts. [Code](#)
- HotRL (AI for Climate Change 2019 contestant)** Reinforcement learning based thermostat control. [Code](#)

## Languages

Fluent English, French, Romanian  
Intermediate Spanish, Mandarin (HSK3)


## Education

- 1st year PhD at University of Montreal & Mila** **Present**  
Research topic: *mechanistic interpretability and continual learning of scientific knowledge in language models*  
Advisor: Irina Rish
- M.Sc. Computer Science at McGill & Mila** **2023.05**  
Thesis: *Language model pretraining with lexical semantics*  
Advisor: Jackie Cheung
- B.Eng. Software Engineering at McGill** **2020.05**  
Interrupted after acceptance into M.Sc. 2018.09
- B.Eng. Materials Engineering CO-OP at McGill** **2017.05**  
Unsatisfied with career prospects. Went backpacking. 2013.09

## Work experience and internships

- NLP Research Intern @ Public Safety Canada** **2021.09**  
Developed novel neural text segmentation model for internal bilingual documents, leveraging supervised contrastive losses to jointly segment headings, paragraphs, and sentences while also performing information extraction of non-standard span categories. Concurrently setup on-prem Ray cluster for distributed training and hyperparameter optimization. 2021.05
- ML developer @ Consilium Crypto** **2019.09**  
Created and deployed microservices-based NLP pipelines for fintech startup. Shipped internal libraries for highly parallelized preprocessing and feature engineering of LOB data, and for distributed training and inference of ML models for timeseries. 2018.05
- Materials Engineering Intern @ Tesla Motors** **2016.12**  
Provided metallurgy expertise for manufacturing, quality assurance, and design projects such as rotor failure analysis, chassis alloy selection, and implementation of fastener stress testing protocols. 2016.09
- Student Researcher @ M. Brochu Lab McGill** **2016.05**  
Created MATLAB model to predict powder recyclability in laser sintering additive manufacturing of Ti64 2016.01
- Hydrometallurgy Intern @ Canadian Royalties** **2016.05**  
Optimized reagent and process conditions to maximize metal recovery in the plant. Automated job with VBA. 2016.01
- High Freq. Radiation R&D Intern @ FPIInnovations** **2015.05**  
Set things on fire in microwaves with susceptors. 2015.01

## Academic service and teaching

- Presenter and co-organizer** **2023.05**  
CDSI training workshop series: Introduction to NLP  2023.01  
(Series renewed for 2023.09 - 2023.12)
- Teaching assistant** **2022.01**  
COMP-550 @ McGill - Natural Language Processing 2021.09
- Conference reviewer** **Various**  
ACL, EACL, EMNLP
- Conference volunteer** **Various**  
ACL, EACL, RLDM

## Awards and grants

- Mila entrepreneurship grant \$5,000 **2023**
- FRQNT M.Sc. research grant \$17,500/yr **2021-22**
- IVADO summer research grant \$5,000 **2020**
- Copper Club - Lord Bagri scholarship \$19,461 **2019**
- Sir William Dawson scholarship \$3,000 **2016**
- NSERC-USRA research grant \$5,625 **2014**
- Henry Birks scholarship \$3,000 **2013**