

Email: sacha.levy@yale.edu

Phone: (203) 589-7468

Website: <https://sachalevy.fr>

GitHub: @sachalevy

LinkedIn: sacha-levy

Twitter: sachalevy3

RESEARCH INTERESTS

Generative Modeling, Active Learning, Computational Biology, Interpretability.

EDUCATION

Yale University

Ph.D., Computer Science, supervised by David van Dijk.

New Haven, CT

August 2023 – Now

McGill University

B. Eng. Computer, Co-op Program.

Montreal, Canada

September 2018 – May 2023

CONFERENCE PUBLICATIONS

INFOSHIELD: Generalizable Information-Theoretic Human-Trafficking Detection

Meng-Chieh Lee, Catalina Vajiac, Aayushi Kulshrestha, [Sacha Lévy](#), Namyong Park, Cara Jones, Reihaneh Rabbany, Christos Faloutsos.

Proceedings of the 37th IEEE International Conference on Data Engineering (ICDE), Apr. 2021. Source code: <https://github.com/mengchillee/InfoShield>.

JOURNAL

DeltaShield: Information Theory for Human- Trafficking Detection

PUBLICATIONS

Catalina Vajiac, Meng-Chieh Lee, Aayushi Kulshrestha, [Sacha Lévy](#), Namyong Park, Andreas Olligschlaeger, Cara Jones, Reihaneh Rabbany, Christos Faloutsos.

ACM Transactions on Knowledge Discovery from Data, Vol. 17, Mar. 2023. Source code: <https://github.com/catvajiac/InfoShield-Incremental>.

DEMONSTRATION

DisKeyword: Twitter Topic Keyword Discovery

PUBLICATIONS

[Sacha Lévy](#), Reihaneh Rabbany.

Proceedings of the 16th ACM International Conference on Web Search and Data Mining, Feb. 2023. Source code: <https://github.com/sachalevy/diskeyword>.

WORKSHOP

Active Keyword Selection to Track Evolving Topics on Twitter

PUBLICATIONS

[Sacha Lévy](#), Farimah Poursafaei, Kellin Pelrine, Reihaneh Rabbany.

Proceedings of the 22nd IEEE International Conference on Data Mining Workshops, Nov. 2022. Source code: <https://github.com/sachalevy/active-keyword-selection>.

Online Partisan Polarization of COVID-19

Zachary Yang, Anne Imouza, Kellin Pelrine, [Sacha Lévy](#), Jiewen Liu, Gabrielle Desrosiers-Brisebois, Jean-François Godbout, André Blais, Reihaneh Rabbany.

Proceedings of the 21st IEEE International Conference on Data Mining Workshops, Nov. 2021.

WORKSHOP

Graph Neural Networks Learn Twitter Bot Behaviour

POSTERS

Albert M Orozco Camacho, [Sacha Lévy](#), Reihaneh Rabbany.

The 34th Conference on Neural Information Processing Systems Workshops, Dec. 2020.

PREPRINTS

Operator Learning Meets Numerical Analysis: Improving Neural Networks through Iterative Methods

Emanuele Zappala, Daniel Levine, Sizhuang He, Syed Rizvi, Sacha Lévy, David van Dijk.
arXiv preprint, Oct. 2023.

Cell2Sentence: Teaching Large Language Models the Language of Biology

Daniel Levine, Syed Asad Rizvi, Sacha Lévy, Nazreen Pallikkavaliyaveetil MohammedSheriff, Ruiming Wu, Zihe Zhang, Antonio Fonseca, Xingyu Chen, Sina Ghadermarzi, Rahul M Dhopdhopkar, David van Dijk.

arXiv preprint, Sept. 2023.

Party Prediction for Twitter

Kellin Pelrine, Anne Imouza, Zachary Yang, Jacob-Junqi Tian, Sacha Lévy, Gabrielle Desrosiers-Brisebois, Aarash Feizi, Cécile Amadoro, André Blais, Jean-François Godbout, Reihaneh Rabbany.

arXiv preprint, Aug. 2023.

Ebbs and Flows of Polarization During a Political Campaign

Kellin Pelrine, Anne Imouza, Gabrielle Desrosiers-Brisebois, Zachary Yang, Sacha Lévy, Aarash Feizi, Jiewen Liu, André Blais, Jean-François Godbout, Reihaneh Rabbany.

preprint, Oct. 2021.

SGP: spotting groups polluting the online political discourse

Junhao Wang, Sacha Lévy, Ren Wang, Aayushi Kulshrestha, Reihaneh Rabbany.

arXiv preprint, Oct. 2019.

AWARDS	Grant for Entrepreneurs (Mila)	2022
	TechAccel Grant (McGill Engine)	2022
	Accelerate Research Scholarship (Mitacs)	2020
	Alumni Prize in Engineering (McGill University)	2019

RESEARCH EXPERIENCE	Research Assistantship	Mila
	Supervisor: Professor Reihaneh Rabbany	May 2019 – April 2023

- Designed an active keyword selection method (*ICDM 2022 workshop paper*) and its dashboard (*WSDM 2023 demo paper*) to track evolving topics on Twitter.
- Implemented a Twitter data pipeline to collect over five billion tweets since 2019 and save eight hours of weekly scripting time for research assistants (fivefold speedup).
- Developed web crawling pipeline for the seven most popular online escort marketplaces in North America collecting over three million ads for the Mila Infrared project.
- Managed computational resources for the lab (i.e. servers and databases) and handled communications with IT services at McGill and Mila for three years.

Capstone Design Project

Supervisor: Professor Mark Coates

McGill University

September 2021 – April 2022

- Conducted an in-depth study of uncertainty quantification for node classification, focusing on *Graph Posterior Network: Bayesian Predictive Uncertainty for Node Classification*.
- Extended the *Graph Posterior Network* experiments with the *ogbn-mag* dataset to test the scalability of the method and its baselines (e.g. *GPN*, *VGCN-Ensemble*, *GKDE*).
- Developed scripts for configuration and dataset loading to run experiments on a Slurm-operated GPU cluster.

Summer Undergraduate Research Project (SURP) **University of Toronto**
Supervisors: Dr. Cherry Ng and Dr. Paul Scholz May 2021 – August 2021

- Optimized the CHIME radio telescope beamforming model to increase real-time Fast Radio Bursts (FRB) detections.
- Designed and ran simulations to discover optimal beam configurations using statistical FRB populations.
- Accelerated existing simulation software one hundredfold through parallelization over the beamforming parameter space.

INDUSTRY **Hulse** Montreal, Canada
EXPERIENCE Founder May 2021 – August 2022

- Engineered and launched a grid computing platform enabling teams to use idle personal computers to run inferences with NLP models from the Hugging Face model hub.
- Implemented and open-sourced *hulse-py* to run distributed inferences on Hulse clusters and *dolphin* to log hardware utilization metrics for workload orchestration purposes.
- Pitched Hulse at the McGill Dobson Cup (startup pitch competition) through the semi-finals and subsequently obtained grant funding from McGill University and Mila.
- Reached out to 150 leads and conducted 20 user interviews with ML researchers from startups and large enterprises to explore computational needs and operational bottlenecks.

Valence Discovery Montreal, Canada
Software Engineering Intern May 2020 – December 2020

- Designed and implemented an ETL pipeline to process and store five billion molecules from 20 public compound libraries into a PostgreSQL database.
- Implemented a Python client allowing ML engineers to load datasets directly to Pandas from the molecular database.
- Developed a model lifecycle management tool (e.g. MLFlow) to automate manual S3 storage of model artifacts.

McKay Brothers Paris, France
Hardware Engineering Intern May 2019 – July 2019

- Assisted engineers in developing a low-latency antenna system for the financial markets.
- Developed visualization tools to accelerate tenfold test results analysis time.

FUNDING Amazon Web Services Activate Credits for Hulse (\$5,000) 2022

TALKS **Dobson Cup Semi-Finals** March 2022
Pitched Hulse, a peer-to-peer grid computing solution, to a panel of investors and professors from the Dobson Center for Entrepreneurship at McGill University.

SURP Poster Presentation August 2021
Presented to the Dunlap Institute and CHIME affiliated faculty and staff my SURP project on beamforming optimization in a poster session.

SKILLS **Technical**
Programming languages: Python, Java, C, JavaScript, VHDL, C++, Bash.
Software tools and packages: Linux, Git, PyTorch, Pandas, PostgreSQL, MongoDB, Neo4j, Kubernetes, Docker, GitHub Actions, AWS, GCP, Slurm.

Languages

French (native), English (fluent), Chinese (basic), Spanish (basic)

SERVICE

McGill Model United Nations

Special Committee Vice Chair

Montreal, Canada

June 2019 – January 2020

- Prepared discussion on ethical usage of Artificial Intelligence for National Intelligence purposes.
- Moderated mock talks between country representatives and orchestrated policy discussions.