Khaoula Chehbouni

Email : kchehbouni@mila.quebec
Linkedin: https://www.linkedin.com/in/khaoulachehbouni/
Website: https://khaoulachehbouni.github.io/

Profile

I am a third year PhD Student in Computer Science at McGill University and Mila (Quebec AI Institute), supervised by Professor Golnoosh Farnadi and professor Jackie Chi Kit Cheung. I was awarded the prestigious FRQNT Doctoral Training Scholarship to research fairness and safety in large language models. Previously, I completed my Masters in Business Intelligence at HEC Montreal, where I received the Best Master Thesis award. My research interest includes responsible AI, fairness and safety issues in large language models, evaluation of natural language generation, privacy and federated learning.

Education

- **McGill University/Mila** Montreal, QC
  PhD in Computer Science; GPA: 4/4 *Aug 2023 - Aug 2027*
  Under the supervision of professor Golnoosh Farnadi and Jackie CK Cheung
- **HEC Montreal/Mila** Montreal, QC
  Master of Science in Business Intelligence; GPA: 4.07/4.3 *Aug 2020 - Aug 2022*
  Under the supervision of professors Golnoosh Farnadi and Gilles Caporossi
- **HEC Montreal** Montreal, QC
  DESS in Information Technologies; GPA: 3.99/4.3 *Aug 2019 - Aug 2020*
- **University of Montreal** Montreal, QC
  Certificate in Applied Computer Science *Jan 2017 - Dec 2017*
- **HEC Montreal** Montreal, QC
  Bachelor of Administration *Aug 2011 - Dec 2015*

Skills Summary

- **Programming Languages** Python, R, SQL, SAS
- **Languages** French, English, Arabic, Spanish

Experience

- **School of Computer Science, McGill University** Montreal, QC
  Teaching Assistant Position - Responsible AI Graduate Class *January 2025 - April 2025*
  - **Teaching Assistant:** I was a TA for Golnoosh's Farnadi graduate course on responsible machine learning. I designed and graded assignments for the course, supported students during office hours and provided guidance for the final research projects.
- **Center for AI and Digital Policy** Montreal, QC
  Research Group Member *September 2024 - December 2024*
  - **AI Policy Clinic:** I was a research group member at CAIPD and obtained my AI Policy certificate with distinctions.
- **Statistics Canada** Montreal, QC
  Senior Data Scientist *Mai 2024 - September 2024*
  - **Project:** Work in the *Research, Method and Information Extraction* section in fairness and privacy issues in large language models.
- **Mila - Quebec AI Institute** Montreal, QC
  Research Intern *Jan 2023 - June 2023*
  - **Project:** Worked on bias mitigation and privacy in sparse language models under professor Golnoosh Farnadi.
- **Statistics Canada** Montreal, QC
  Data Scientist *Sep 2022 - Jan 2023*
  - **Position:** Work in the *Methods, Quality and Research* section in responsible AI
  - **Project:** Developed a course on bias and fairness in machine learning
- **IBM x Ivado** Montreal, QC
  Research Assistant *Oct 2021 - Apr 2022*
  - **Project:** Towards Discovering and Understanding how to improve the representation of women in STEM: A case study from IBM Inventorship and Patenting
  - **Description:** Literature review and predictive modelling to understand inclusion issues in IBM's patents activities

- **La Presse x Ivado** Montreal, QC
  Data Journalist Intern *May 2021 - Sep 2021*
  - **Text Analysis:** Scrapping, topic modelling, sentiment analysis, clustering, predictive modelling
  - **Writing:** Published 9 articles on La Presse's official website, 3 of them were front-page articles
- **Canadians for Justice and Peace in the Middle East** Montreal, QC
  IT Manager *Jan 2016 - Dec 2021*
  - Managed IT bugs, transfer of data, and design improvement projects.
  - Implemented the new website on the NationBuilder platform
  - Created practical tools and tutorials for management and accounting software

Research Projects

Conferences

- **Neither Valid nor Reliable? Investigating the Use of LLMs as Judges Chehbouni, K.**, et al., In: NeurIPS 2025 - Position Paper Track (less than 8% acceptance rate)
- **Fairness in Federated Learning: Fairness for Whom?** Taik, A., **Chehbouni, K.**, Farnadi, G., In: Proceedings of the 8th AAAI/ACM Conference on AI, Ethics, and Society (AIES) 2025
- **Beyond the Safety Bundle: Auditing the Helpful and Harmless Dataset Chehbouni, K.***, Jonathan Jonathan Colaço Carr*, et al., In: Proceedings of the 2025 Conference of the North American Chapter of the Association for Computational Linguistics — Oral.
- **Enhancing Privacy in Early Detection of Online Grooming Through Federated Learning and Differential Privacy Chehbouni, K.** et al., In: Proceedings of the AAAI Conference on Artificial Intelligence, 2025. — Oral
- **From Representational Harms to Quality-of-Service Harms: A Case Study on Llama 2 Safeguards. Chehbouni, K.**, et al., In: Findings of the Association for Computational Linguistics: ACL 2024 — Poster

Under Review

- **Nothing New Under the GenAI Sun: Systematic Hazard Taxonomy Along the AI Lifecycle Chehbouni, K.**, et al., Under review: CHI 2026

Journal

- **The Effect of the Typicality of Song Lyrics on Song Popularity: A Natural Language Processing Analysis of the British Top Singles Chart. Chehbouni, K.***, Florian Carichon, F.*, Simonnot-Lanciaux, A., et al., Accepted to Psychology of Music 2025

Workshop

- **Neither Valid Nor Reliable? Investigating the Use of LLMs as Judges Chehbouni, K.**, et al., In: Third Workshop on Socially Responsible Language Modelling Research (SoLaR) 2025 — Poster
- **Leveraging AI for Natural Disaster Management: Takeaways From The Moroccan Earthquake.** Ezzine, L. **et al.**, In: 6th Workshop on Artificial Intelligence for Humanitarian Assistance and Disaster Response NeurIPS 2023. — Oral
- **Unmasking Predators: Safeguarding Vulnerable Moroccan Communities Post-Earthquake** Taïk*, A., **Chehbouni*, K.**, Jain*, K., et al., In: North Africans in Machine Learning Workshop at NeurIPS 2023. — Poster
- **Early Detection of Sexual Predators with Federated Learning. Chehbouni, K.**, Caporossi, G., Rabbany, R., De Cock, M., Farnadi, G., In: Workshop on Federated Learning: Recent Advances and New Challenges (in Conjunction with NeurIPS 2022). — Poster

Other

- **LLM explainability.** Arous, I., **Chehbouni, K.**, Cheng, Z., Dossou, B., Accepted as a book chapter in the Handbook of Human-Centered Artificial Intelligence
- **AI and Cities: Risks, Applications and Governance.** Koseki, S. **et al.**, In: 4th Urban Economy Forum Conference 2022 — White Paper

In the Media

- I was invited in the “Activists of Tech — The Responsible Tech Podcast” to talk about my research and experience in academia: How Existing Safety Mitigation Safeguards Fail in LLMs
- I was invited in the “World We Are Building” podcast to talk about my research: Khaoula Chehbouni: Rethinking AI Safety, Hidden Hierarchies in AI Development and Who Gets to Decide What Makes AI Research Worthy of Publication
- I was featured in the Fall 2023 Edition of HEC Mag: “Khaoula Chehbouni: Démasquer les prédateurs sexuels”.

Talks and Communications

- I was invited as a Panelist at McGill University—November 2025: Inaugural event of the inclusive AI Program
- I was an invited speaker for the TADA Speaker Series—November 2025: “Beyond the Safety Bundle: Auditing the Helpful and Harmless Dataset”
- I was invited as a Panelist at HEC Montréal—November 2025: “Relever les défis liés au changement de carrière”
- I was invited as a Panelist at HEC Montréal—October 2025: “Carrières en science des données”
- I was invited as a speaker for the Mila TRAIL: Responsible AI for Professionals and Leaders—October 2025:“Déconstruire les biais: garantir l’équité algorithmique dans les modèles d’IA”
- I was invited as a speaker for the Mila TRAIL: Responsible AI for Professionals and Leaders—October 2025: “Stratégies concrètes pour les évaluations d’impact de l’IA”
- I was invited as a speaker for the Morocco AI Webinar Series—July 2025, Online: “Rethinking Safety Evaluation in Large Langage Models.”
- I gave a tutorial for the AI Perspective series organized by McGill University and Ivado—June 2025: “Algorithmic Fairness: A Hands-On Introduction.”
- I was an invited speaker for Mila AI Policy Compas—June 2025: “Biais et Équité en intelligence artificielle.”
- I was invited as a speaker for the Mila Responsible AI Summer School—June 2025: “Technical Mitigation Strategies.”
- I was invited to give a workshop at the Google’s Women Techmaker Event in Montreal—April 2025: “Bias in the Machine: Can AI Ever Be Fair?”
- I was invited as a Panelist by Encode AI for an event in McGill University—March 2025.
- I was invited as a Panelist by the World Salon for an event in cooperation with KIIT University and Columbia University—January 2025: “New Applications of AI”.
- I was invited as a Panelist at the McGill Technology and Ethics Roundtables - November 2024, Montreal, Canada.
- I was an invited speaker at the StatCan-Waterloo symposium on Data Science and AI - October 2024, Online: “Understanding the Use and Limitations of Large Language Models as Natural Language Generation Evaluators”
- I was an invited speaker for Mila AI Policy Compas - August 2024: “Biais et Équité en intelligence artificielle”
- I was an invited speaker at Congrès de la Société de philosophie du Québec - June 2024: “Mitigation des risques liés aux Grands Modèles de Langage”
- I was an invited speaker at the Machine Learning Technical Series at Statistics Canada - July 2024: “From Representational Harms to Quality-of-Service Harms: A Case Study on Llama 2 Safety Safeguards”
- I participated in a live debate on the use of AI for immigration in the Responsible AI Track at the Canadian AI Conference in Guelph in May 2024.
- I was an invited speaker at Google’s Women Techmakers event - April 2024: “Towards Inclusive AI: Understanding Bias in Machine Learning Models”
- I was an expert at a roundtable about the risks of general purpose AI systems organized by Mila AI for Humanity for ISED (Innovation, Science and Economic Development Canada) in February 2024.
- I was an invited speaker at the Research IceBreaker event at Mila - October 2023: “Unmasking Predators: Safeguarding Vulnerable Moroccan Communities Post-Earthquake”

Awards and Scholarship

- FRQNT doctoral training scholarships

  My research proposal ”Equity and Intersectionality: A New Perspective for Generative Models” was awarded one of the prestigious FRQNT doctoral training scholarships. It's an annual scholarship of $25000 per year for 4 years (for a total of $100000).

- Mila Excellence Scholarship - EDI in Research

  I was awarded one of Mila's Excellence scholarship for the category EDI in Research, PhD level. It's an annual scholarship of $8000 per year for 3 years (for a total of $24000).

- McGill University Award

  I was awarded a Women in Science grant of $6000 in 2024 and I also received $ 5000 for the Lorne Trottier Fellowship

- HEC Montreal's 2022 Best Thesis Award

  My thesis ”Mitigating Online Grooming with Federated Learning” won the 2022 Best Thesis Award at HEC Montreal. The award consisted of a scholarship of $2000.

- IVADO Excellency scholarship “Data Storytelling”

  I was awarded one of IVADO's internship scholarship to conduct an internship at La Presse in data storytelling. The award consisted of a scholarship of $5000.

- Winner of the Three Minute Thesis Competition at the Responsible AI Track of the Canadian AI Conference 2024

  I was the winner of the 3MT competition at the Canadian AI Conference — Responsible AI Track. The award was a cash prize of $600.

- Winner of the Morocco Solidarity Hackathon 2023

  Our project “Unmasking Predators: Safeguarding Vulnerable Moroccan Communities Post-Earthquake” won the Morocco Solidarity Hackathon. We were awarded virtual passes to attend NeurIPS as well as $5,000 AWS credit.

- Travel Grants
  - I was awarder $3000 by IVADO to attend a safety workshop in Berkeley in April 2025.
  - I was awarded $650 by McGill University to attend NeurIPS 2024 in Vancouver.
  - I was awarded $700 US by Women in Machine Learning to attend the workshop at NeurIPS 2024 in Vancouver, Canada.
  - I was awarded $1300 US by the AI, Ethics, and Society Conference to attend their student program in San Jose, United-States.
  - I was awarded a travel grant of $1160 to attend the Canadian AI conference 2024 at Guelph by Mila.
  - I was awarded 650$US to present my Masters work at the Women in Machine Learning Workshop at NeurIPS 2022 in New-Orleans, United-States.

Community Service

- **McGill University** Montreal, QC

  PhD Student *Present*
  - I am a member of the President's Advisory Council for Engagement (PACE).
  - I reviewed papers for the SoLaR Workshop and the AFT Workshop at NeurIPS 2023 and 2024, as well as the Responsible Generative AI workshop at CVPR 2024 and the ICML 2024 Workshop on Trustworthy Multi-modal Foundation Models and AI Agents.
  - I was a student volunteer at NeurIPS 2023 in New Orleans and at the Women in Machine Learning Workshop at NeurIPS in 2022 and in 2024.
- **Mila - Quebec AI Institute** Montreal, QC

  PhD Student *Present*
  - I organize the Fairness, Accountability, Transparency, Ethics & Society reading group at Mila: every week, I host speakers to talk about everything related to Responsible AI.
  - I am a member of the Equity, Diversity and Inclusion Committee at Mila.
  - I organized a workshop for the Responsible AI Week at Mila: (Re)Defining Responsible AI: A workshop on the characteristics and blind-spots of modern responsible AI frameworks.
  - I recently joined Mila's EDI Committee in order to help making the institute more inclusive.
  - I was a mentor at a Women at Mila event aimed at helping students with their resume.

- **Technovation Montreal** Montreal, QC
  Mentor *January-May 2024*
  - I mentor a group of high-school girls in the development of an application to help teenagers with their career choices as part of the International Technovation Challenge.
- **Canada's New Democrats (NDP) Association** Montreal, QC
  Financial Agent *2019 - 2020*
  - Managed the financial operations of candidate Miranda's Gallo campaign in the district of Saint-Laurent.
  - Reported financial results to Elections Canada.