

# David Beauchemin

 david.beauchemin.5@ulaval.ca

 (514) 250-3616

 View GitHub

 View Profile

## Education

2020 – 2025



**PhD in Computer Science | Machine Learning | Natural Language Processing**

Université Laval, GRAIL - Group for Research in Artificial Intelligence of Laval University \*

Thesis topic: *Automatic Explanation of Automobile Insurance Contracts Using Retrieval-Augmented Generation*

2018 – 2020



**Master's Degree in Computer Science | Machine Learning | Natural Language Processing**

Université Laval, GRAIL - Group for Research in Artificial Intelligence of Laval University \*

Thesis topic: *Duplicate detection among unstructured information from external data sources*

2015 – 2018



**Bachelor Degree in Actuarial Science**

Université Laval

## Experience

### Professional

2026 –



**Research Professional**

Université Laval, Institut intelligence et données (IID) \*

2020 –



**Founding Member, AI Expert and Managing Director**

Baseline Artificial Intelligence Solidarity Cooperative

2022 – 2025



**Student Member of the Council**

Université Laval, Institut intelligence et données (IID) \*

2020 – 2022



**Project Manager for a 200-hour AI training course**

Université Laval, Faculty of Science and Engineering | Continuing education

2019 – 2025



**Founder, Researcher, and Interviewer**

OpenLayer podcast \*

AI content creation with over 40,000 views across 50 episodes

2018 – 2025



**Event Organization**

Meetup Machine Learning Québec

Co-organization of various events in Quebec City

2019 – 2020



**Developer**

Poutyne \*

Feature development

2019 – 2020



**AI Scientist**

Vooban

Technical advisor in artificial intelligence

## Teaching

2022 – 2024



### Lecturer

Université Laval, Department of Computer Science and Software Engineering  
Advanced Databases (bachelor and graduate level) (2 times)

2016 – 2022



### Course Assistant / Grader

Université Laval, School of Actuarial Science and Department of Computer Science and Software Engineering

Advanced Databases, Introduction to Actuarial Science I (2 times), Financial Risk Management I, Analysis and Collective Risk Processing, Numerical Methods in Actuarial Science (3 times), P&C Actuarial Mathematics II (2 times), Linear Models in Actuarial Science, Programming with R for Data Analysis (5 times), Internship Report Grading (8 times)

2019 – 2020



### Teaching Assistant

Université Laval, CRDM - Centre de recherche en données massives  
Machine Learning Winter School

## Research and Supervision

2020 – 2025



### Research Assistant

Université Laval

Supervision of five interns for my doctoral research project «Automatic Explanation of Automobile Insurance Contracts Using Retrieval-Augmented Generation»

Richard Khoury, Beneva

2020 – 2022



### Research Assistant

Université Laval, Institut intelligence et données (IID) ↗\*

Supervision of two interns for the project «Women facing the challenges of digital transformation: a case study in the insurance sector»

Christian Gagné, Future Skills Centre

2020



### Research Assistant

Université Laval, School of Actuarial Science

Supervision of an intern and development of the infrastructure plan for the project «Open platform for validation and benchmarking of financial and actuarial forecasts»

Vincent Goulet, Research Chair in Actuarial Science at Université Laval

2019



### Research Assistant

Université Laval, GRAIL - Group for Research in Artificial Intelligence of Laval University ↗\*

Supervision of an intern and development of a solution allowing for «Commercial risk record linkage»

Luc Lamontagne, NSERC Industrial Research Chair - Intact Financial Corporation in Machine Learning for Insurance

2019



### Research Assistant

Université Laval, GRAIL - Group for Research in Artificial Intelligence of Laval University ↗\*

Professional skills taxonomy learning

Luc Lamontagne, ENGAGE Project

2018



### Research Assistant

Université Laval, School of Actuarial Science

Measures related to conditional indexing in a defined benefit pension plan

Louis Adam, Research Chair in Actuarial Science at Université Laval

2018



### Research Assistant

Université Laval, GRAIL - Group for Research in Artificial Intelligence of Laval University ↗\*

Weighted bootstrapping method for word relation extraction

Luc Lamontagne

## Funding

---

|             |  |
|-------------|--|
| 2025        | <b>Bourse de recherche en gouvernance</b><br>Université Laval (10,000\$)   |
| 2022 – 2025 | <b>Doctoral Research Scholarship</b><br>Fonds de recherche du Québec – Nature et technologies (FRQNT) (65,000\$) |
| 2020 – 2025 | <b>Bourse de réussite FESP - FSG</b><br>Université Laval, Faculty of Science and Engineering (13,125\$)          |
| 2024        | <b>Bourse du Fonds Vision Topping entrepreneuriat</b><br>Université Laval (10,000\$)                             |
| 2020        | <b>Bourse d'admission au doctorat de l'Université Laval</b><br>Université Laval (5,000\$)                        |
| 2018        | <b>Bourse Mathieu</b><br>Université Laval (2,500\$)  |

## Publications

---

### Papers

- 2026 David Beauchemin and Richard Khoury. *Benchmarking Large Language Models for Quebec Insurance: From Closed-Book to Retrieval-Augmented Generation*. ICLR «Advances in Financial AI » Workshop (under review).
- 2026 David Beauchemin, Yan Tremblay, Mohamed Amine Youssef, and Richard Khoury. *Lost in Translation: Why SOTA LLMs Struggle with French NLU Frontiers*. ICLR «I Can't Believe It's Not Better » Workshop (under review).
- 2025 David Beauchemin. *Automatic Explanation of Automobile Insurance Contracts Using Retrieval-Augmented Generation* ↗\*. PhD Thesis, Université Laval.
- 2025 David Beauchemin, Yan Tremblay, Mohamed Amine Youssef, and Richard Khoury. *A Set of Quebec-French Corpus of Regional Expressions and Terms* ↗\*. arXiv:2510.05026. Reviewed at ACL Rolling Review of October 2025 (score: 3.0), will be submitted to ACL 2026.
- 2025 David Beauchemin, Yan Tremblay, Mohamed Amine Youssef, and Richard Khoury. *COLE: a Comprehensive Benchmark for French Language Understanding Evaluation* ↗\*. arXiv:2510.05046.
- 2025 David Beauchemin, Pier-Luc Veilleux, Richard Khoury, and Johanna-Pascale Roy. *QFrBLiMP: a Quebec-French Benchmark of Linguistic Minimal Pairs* ↗\*. arXiv:2509.25664. Accepted at EACL 2026.
- 2025 David Beauchemin and Richard Khoury. *QFrCoLA: a Quebec-French Corpus of Linguistic Acceptability Judgments* ↗\*. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 119-130, Suzhou, China. Association for Computational Linguistics.
- 2025 David Beauchemin, Michelle Albert-Rochette, Richard Khoury, and Pierre-Luc Déziel. *JUDGEBERT: Assessing Legal Meaning Preservation Between Sentences* ↗\*. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 92-118, Suzhou, China. Association for Computational Linguistics.
- 2024 David Beauchemin, Richard Khoury and Zachary Gagnon. *Quebec Automobile Insurance Question-Answering With Retrieval-Augmented Generation* ↗\*. In *Proceedings of the Natural Legal Language Processing Workshop at EMNLP*, (pp. 48-60).
- 2023 David Beauchemin, et Marouane Yassine. *Deepparse: An Extendable, and Fine-Tunable State-Of-The-Art Library for Parsing Multinational Street Addresses* ↗\*. Proceedings of the 3rd Workshop for Natural Language Processing Open Source Software (NLP-OSS 2023), pages 19–24, Singapore. Association for Computational Linguistics.
- 2023 David Beauchemin, Horacio Saggion, et Richard Khoury. *MeaningBERT: assessing meaning preservation between sentences* ↗\*. Frontiers in Artificial Intelligence. doi: 10.3389/frai.2023.1223924

- 2023 David Beauchemin, et Richard Khoury. *RISC: Generating Realistic Synthetic Bilingual Insurance Contract* [↗\\*](#). Proceedings of the Canadian Conference on Artificial Intelligence. <https://doi.org/10.21428/594757db.132dae7d> (*Best Student Paper Award.*)
- 2022 David Beauchemin, et Marie-Claire Monty. *Is discrimination in artificial intelligence sufficiently regulated?* [↗\\*](#). HAL:03736828.
- 2022 Vincent Primpied, David Beauchemin, et Richard Khoury. *Quantifying French Document Complexity* [↗\\*](#). Proceedings of the Canadian Conference on Artificial Intelligence.
- 2022 David Beauchemin, Julien Laumonier, Yvan Le Ster, et Marouane Yassine. “*FIJO*”: a French Insurance Soft Skill Detection Dataset [↗\\*](#). Proceedings of the Canadian Conference on Artificial Intelligence.
- 2021 Marouane Yassine, David Beauchemin, François Laviolette et Luc Lamontagne. *Multinational Address Parsing: A Zero-Shot Evaluation* [↗\\*](#). Accepted in the International Journal of Information Science et Technology (iJIST)
- 2021 Marouane Yassine, David Beauchemin, François Laviolette et Luc Lamontagne. *Leveraging Subword Embeddings for Multinational Address Parsing* [↗\\*](#). IEEE CiSt, MNLP.
- 2020 David Beauchemin, Nicolas Garneau, Eve Gaumond, Pierre-Luc Déziel, Richard Khoury et Luc Lamontagne. *Generating Intelligible Plumitifs Summaries: Use Case Application with Ethical Considerations* [↗\\*](#). The 13th International Conference on Natural Language Generation.
- 2020 David Beauchemin. *Duplicate detection among unstructured information from different data sources* [↗\\*](#). Master’s Thesis, Université Laval.
- 2019 Nicolas Garneau, Mathieu Godbout, David Beauchemin, Audrey Durand et Luc Lamontagne. *A Robust Self-Learning Method for Fully Unsupervised Cross-Lingual Mappings of Word Embeddings: Making the Method Robustly Reproducible as Well* [↗\\*](#). REPORLANG@LREC2020.

## Non-scientific

- 2017 David Beauchemin et Vincent Goulet *Typesetting actuarial symbols easily and consistently with actuarialsymbol and actuarialangle* [↗\\*](#)

## Packages

- 2023 David Beauchemin *RISC: An open-source Python package for generating realistic synthetic insurance contracts* [↗\\*](#)
- 2022 Vincent Primpied, David Beauchemin, and Richard Khoury *Text Complexity Computer (TCC): A library for assessing text complexity in French* [↗\\*](#)
- 2020 Marouane Yassine et David Beauchemin *Deepparse: A state-of-the-art deep learning multinational addresses parser* [↗\\*](#)
- 2019 David Beauchemin *Notif: The notification package for every Python project* [↗\\*](#)
- 2017 Simon-Pierre Gadoury et David Beauchemin *nCopula: Hierarchical Archimedean Copulas Through Multivariate Compound Distributions* [↗\\*](#)
- 2017 Vincent Goulet et David Beauchemin *actuarialsymbol - Actuarial notation with LATEX* [↗\\*](#)

## Corpora & Benchmarks

- |      |   |
|------|---|
| 2026 | David Beauchemin, Pier-Luc Veilleux, Richard Khoury, and Johanna-Pascale Roy. <i>QFrBLiMP</i>  *. A dataset of 1,761 minimal pairs annotated with 20 linguistic phenomena.   |
| 2025 | David Beauchemin, Yan Tremblay, Mohamed Amine Youssef, and Richard Khoury. <i>COLE</i>  *. A benchmark composed of 23 diverse NLU tasks for LLM benchmarking, with a public benchmarking platform  *. |
| 2025 | David Beauchemin and Richard Khoury. <i>QFrCoLA</i>  *. A dataset of 25,153 in-domain and 2,675 out-of-domain grammatical and ungrammatical sentences. In partnership with the <i>Office québécois de la langue française</i> .  |
| 2025 | David Beauchemin, Michelle Albert-Rochette, Richard Khoury, and Pierre-Luc Déziel. <i>FrJUDGE</i>  *. A dataset of 297 annotated sentence pairs of legal meaning judgment assessment in French.  |
| 2023 | David Beauchemin, Horacio Saggion, and Richard Khoury. <i>CSMD</i>  *. A dataset of 1,355 English sentence triplets annotated with continuous meaning preservation ratings.  |
| 2023 | David Beauchemin, and Richard Khoury. <i>RISCBAC</i>  *. A synthetic corpus of 10,000 English and French insurance contracts.  |
| 2022 | David Beauchemin, Julien Laumonier, Yvan Le Ster, and Marouane Yassine. <i>FIJO</i>  *. A dataset of job offers annotated for soft skills detection.   |