





Mathieu Ravaut

 Ravoxsg |  mravox |  ravoxsg.github.io |  mathieu.ravaut@gmail.com |  +6588947729 | French citizen

SUMMARY

Computer scientist with a strong passion for machine learning, both applied and fundamental. Versatile researcher with 7+ years of experience building machine learning pipelines. I am currently a PhD Candidate at Nanyang Technological University (NTU) and A*STAR I2R in Singapore. My research interests include natural language processing (NLP), machine learning applied to healthcare and recommender systems. I am looking for a job in industry starting mid 2024.

EDUCATION

- 2021 - 2024 **PhD** in Computer Science at **Nanyang Technological University**, Singapore. (GPA: 4.63/5.00)
Supervisors: Assoc. Prof Shafiq Joty (NTU), Assoc. Prof Aixin Sun (NTU), Dr. Nancy Chen (I2R).
Research area: improving neural abstractive summarization with sequence-level methods.
Publications: ACL '22, EMNLP '22, EACL '23, ACL '23, EACL '24 (x2), and 4 papers under review.
- 2017 - 2018 **MSc** in Applied Computing at **University of Toronto**, Canada. (GPA: 3.95/4.00)
Coursework: machine learning, deep learning, computational statistics, natural language processing.
- 2014 - 2018 **B.Eng** (2015), **M.Eng** (2018) at **Ecole Centrale Paris (CentraleSupélec)**, France. (GPA: 3.65/4.00)
Coursework: applied mathematics, computer science, quantum physics, statistical physics, business.
- 2011 - 2014 Prepa MPSI/MP* at **Lycée Montaigne**, Bordeaux, France. (Grade: A)
Intensive preparation for national competitive entrance exams to Engineering Schools.
Coursework: mathematics, physics, computer science.

WORK EXPERIENCE

- Huawei Noah Ark, Singapore - Research Intern** Jan 2023 - Jul 2023
- Search & Recommendation team, supervised by Dr. Zhang Hao and Dr. Liu Yong.
 - Research in conversational recommender systems. Paper published at EACL 2024, another paper under review.
- Layer 6 AI (TD Bank), Toronto, Canada - Machine Learning Research Scientist** May 2018 - Jul 2020
- Applied research in machine learning for healthcare. I led a collaboration between Layer 6 AI and ICES (with Prof. Laura Rosella) adapting an industry-leading software stack from the financial services industry to health administrative data. The project led to several publications in top journals: Nature, JAMA, BMJ.
 - Member of a team placing 2nd (out of 70+) at ACM RecSys Challenge 2020.
 - Insurance claim fraud detection with NLP: building a model leveraging insurance history, text and graph features. I was in charge of the model evaluation and building the text features.
- A*STAR I2R, Singapore - Research Intern** Feb 2017 - Jul 2017
- Visual Computing Lab, supervised by Dr. Vijay Chandrasekhar.
 - Research in computer vision, resulting in a workshop paper at CVPR 2017.
- Thales Solutions Asia, Singapore - Research Intern** Aug 2016 - Feb 2017
- Research & Technology department, supervised by Dr. Antoine Fagette.
 - Applied research in computer vision, leading to a paper published at IEEE Oceans 2017.

TEACHING EXPERIENCE

- Nanyang Technological University - Teaching Assistant** Jan 2021 - Nov 2022
- Graduate-level Deep Learning for NLP courses: AI6127 Deep Learning for Natural Language Processing (S1 2022), CE7455 Deep Learning for Natural Language Processing (S1 2021, S1 2022), MH6812 Deep Learning for Natural Language Processing (S2 2021), CE4045 Deep Learning for Natural Language Processing (S2 2021).
 - 1st-year CS courses: SC1003 Introduction to Computational Thinking and Programming (S2 2022), SC1007 Data Structures and Algorithms (S1 2022).
- University of Toronto - Teaching Assistant** Jan 2018 - April 2018
- 1st-year Statistics course: STA130 Introduction to Statistical Reasoning and Data Science (Winter 2018).

SKILLS

Programming Languages	Python (expert), Java, R.
Software & IDEs	LateX, Git, Github, PyCharm, VisualStudio, AWS, Docker.
Python Libraries	numpy, scipy, pandas, scikit-learn, xgboost, lightgbm, pytorch, tensorflow, keras, huggingface (transformers, datasets, accelerate, peft), pySpark, matplotlib.

PUBLICATIONS

- 17 **ChatGPT's One-year Anniversary: Are Open-Source Large Language Models Catching up?** [pdf] **NLP**
Hailin Chen, Fangkai Jiao, Chengwei Qin, Xingxuan Li, **Mathieu Ravaut**, Ruochen Zhao, Caiming Xiong, Shafiq Joty.
Preprint. Under review.
- 16 **On Context Utilization in Summarization with Large Language Models** [pdf] **NLP**
Mathieu Ravaut, Shafiq Joty, Aixin Sun, Nancy F. Chen.
Preprint. Under review.
- 15 **LOCOST: State-Space Models for Long Document Abstractive Summarization** **NLP**
Florian Le Bronnec, Song Duong, **Mathieu Ravaut**, Alexandre Allauzen, Nancy F. Chen, Vincent Guigue, Alberto Lumbreras, Laure Soulier, Patrick Gallinari.
EACL 2024.
- 14 **Parameter-Efficient Conversational Recommender System as a Language Processing Task** **NLP & Recommender Systems**
Mathieu Ravaut, Hao Zhang, Lu Xu, Aixin Sun, Yong Liu.
EACL 2024.
- 13 **Targeted COVID-19 and Human Resource for Health News Information Extraction with a Multi-Component Deep Learning Framework** [pdf] **NLP & ML for Health**
Mathieu Ravaut, Ruochen Zhao, Duy Phung, Vicky Mengqi Qin, Dusan Milovanovic, Johannes Schnitzler, Anita Pienkowska, Iva Bojic, Josip Car, Shafiq Joty.
Preprint. Under review.
- 12 **PromptSum: Parameter-Efficient Controllable Abstractive Summarization** [pdf] **NLP**
Mathieu Ravaut, Hailin Chen, Ruochen Zhao, Chengwei Qin, Shafiq Joty, Nancy F. Chen.
Preprint.
- 11 **Unsupervised Summarization Re-ranking** [pdf] **NLP**
Mathieu Ravaut, Shafiq Joty, Nancy F. Chen.
ACL Findings 2023. Acceptance rate: **39.09%**.
- 10 **A Data-centric Framework for Improving Domain-specific Machine Reading Comprehension Datasets** [pdf] **NLP**
Iva Bojic, Josef Halim, Verena Suharman, Sreeja Tar, Qi Chwen Ong, Duy Phung, **Mathieu Ravaut**, Shafiq Joty, Josip Car.
EACL 2023 Workshop on Insights from Negative Results in NLP.
- 9 **Towards Summary Candidates Fusion** [pdf] **NLP**
Mathieu Ravaut, Shafiq Joty, Nancy F. Chen.
EMNLP 2022. Acceptance rate: **22.10%**.
- 8 **SummaReranker: a Multi-task Mixture-of-Experts Re-ranking Framework for Abstractive Summarization** [pdf] **NLP**
Mathieu Ravaut, Shafiq Joty, Nancy F. Chen.
ACL 2022. Acceptance rate: **20.75%**.
- 7 **Developing Machine Learning Algorithms on Routinely Collected Administrative Health Data-Lessons from Ontario, Canada** [pdf] **ML for Health**
Vinyas Harish, **Mathieu Ravaut**, Seung Eun Yi, Jahir Mauricio Gutierrez Bugarin, Hamed Sadeghi, Kin Kwan Leung, Tristan Watson, Kathy Kornas, Tomi Poutanen, Maksims Volkovs, Laura Rosella.
International Journal of Population Data Science 2022. Impact factor (2023): **1.35**.
- 6 **Predicting Hospitalizations Related to Ambulatory Care Sensitive Conditions with Machine Learning for Population Health Planning: Derivation and Validation Cohort Study** [pdf] **ML for Health**
Seung Eun Yi, Vinyas Harish, Jahir Mauricio Gutierrez Bugarin, **Mathieu Ravaut**, Kathy Kornas, Tristan Watson, Tomi Poutanen, Marzyeh Ghassemi, Maksims Volkovs, Laura Rosella.
BMJ Open 2022. Impact factor (2023): **3.01**.
- 5 **Development and Validation of a Machine Learning Model Using Administrative Health Data to Predict Onset of Type 2 Diabetes** [pdf] **ML for Health**
Mathieu Ravaut, Vinyas Harish, Hamed Sadeghi, Kin Kwan Leung, Maksims Volkovs, Kathy Kornas, Tristan Watson, Tomi Poutanen, Laura Rosella.
JAMA Network Open 2021. Impact factor (2023): **13.37**.
- 4 **Predicting Adverse Outcomes Due to Diabetes Complications with Machine Learning Using Administrative Health Data** [pdf] **ML for Health**
Mathieu Ravaut, Hamed Sadeghi, Kin Kwan Leung, Maksims Volkovs, Kathy Kornas, Vinyas Harish, Tristan Watson, Gary F Lewis, Alanna Weisman, Tomi Poutanen, Laura Rosella.
Nature npj Digital Medicine 2021. Impact factor (2023): **15.36**.

