




# Mathieu Ravaut

 [Ravoxsg](#) |  [mravox](#) |  [ravoxsg.github.io](#) |  [mathieu.ravaut@gmail.com](mailto: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.

## EDUCATION

---

- 2021 - 2024 **PhD** in Computer Science at **Nanyang Technological University**. (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 2022, EMNLP 2022, EACL 2023, ACL 2023, and 4 papers under review.
- 2017 - 2018 **MSc** in Applied Computing at **University of Toronto**. (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**. (GPA: 3.65/4.00)  
Coursework: applied mathematics, computer science, quantum physics, statistical physics, business.
- 2011 - 2014 Prepa MPSI/MP\* at **Lycée Montaigne, Bordeaux**. (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. Two papers are currently 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 Stats course: STA130 Introduction to Statistical Reasoning and Data Science (Winter 2018).

## SKILLS

---

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

## PUBLICATIONS

---

- 15 **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.
- 14 **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.
- 13 **On Context Utilization in Summarization with Large Language Models** [pdf] NLP  
**Mathieu Ravaut**, Shafiq Joty, Aixin Sun, Nancy F. Chen.  
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. Under review.
- 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**.
- 3 **Predicting Twitter Engagement with Deep Language Models** [pdf] NLP & Recommender Systems  
Maksims Volkovs, Zhaoyue Cheng, **Mathieu Ravaut**, Hojin Yang, Kevin Shen, Jin Peng Zhou, Anson Wong, Saba Zuberi, Ivan Zhang, Nick Frosst, Helen Ngo, Carol Chen, Bharat Venkitesh, Stephen Gou, Aidan N Gomez.  
*ACM RecSys Workshop on the Recommender Systems Challenge 2020*.
- 2 **Truly Multi-modal Youtube-8M Video Classification with Video, Audio and Text** [pdf] NLP & Computer Vision  
Zhe Wang, Kingsley Kuan, **Mathieu Ravaut**, Gaurav Manek, Sibong Song, Yuan Fang, Seokhwan Kim, Nancy F. Chen, Luis Fernando D'Haro, Luu Anh Tuan, Hongyuan Zhu, Zeng Zeng, Ngai Man Cheung, Georgios Piliouras, Jie Lin, Vijay Chandrasekhar.  
*CVPR Workshop on YouTube-8M Large-Scale Video Understanding 2017*.

## 1 Deep Learning Applied to Underwater Mine Warfare [pdf] [Computer Vision](#)

Killian Denos, **Mathieu Ravaut**, Antoine Fagette, Hock-Siong Lim.

*IEEE Oceans 2017.*

## PATENTS

---

### 2 System and Method for Dynamically Predicting Fraud using Machine Learning.

Xiao Shi Huang, Sandra Aziz, Juan Felipe Perez Vallejo, Jean-Christophe Bouëtté, Jennifer Bouchard, **Mathieu Ravaut**, Maksims Volkovs, Tomi Johan Poutanen, Joseph Pun, Ghaith Kazma, Olivier Gandouet.

*US Patent App.* 17/207,221, 2022.

### 1 Regularization of Recurrent Machine-Learned Architectures.

Maksims Volkovs, **Mathieu Ravaut**, Kin Kwan Leung, Hamed Sadeghi.

*US Patent App.* 16/435,213, 2020.

## ACADEMIC SERVICES

---

### Reviewer:

ACL Rolling Review (ARR)	2023 -
CoNLL	2023
SIGDIAL	2023
AACL	2023
IEEE/ACM Transactions on Audio, Speech, and Language Processing (TASLP)	2021-

## AWARDS

---

### 9 Singapore Data Science Consortium (SDSC) PhD Fellowship 2023

Granted by *Singapore Data Science Consortium* in 2023.

10k SGD cash award to support my PhD research.

### 8 Singapore International Graduate Award (SINGA)

Granted by *A\*STAR* in 2020.

Full tuition fees waiver and monthly 2.7k SGD stipend to support my PhD studies at NTU.

### 7 TD Data Analytics Helix Award "Best Partnership - Business and Analytics"

Granted by *TD Bank* in 2020.

Recognition for our project applying NLP to an insurance business case.

### 6 CodaLab MIND News Recommendation Challenge 2020 (Microsoft) - 3rd prize

Granted by *Microsoft* in 2020.

1k USD cash award for a machine learning challenge in news recommendation.

### 5 ACM Twitter RecSys Challenge 2020 - 2nd place team

Granted by *ACM* in 2020.

Worldwide machine learning challenge in recommender systems.

### 4 Kaggle Competitions Expert

Granted by *Kaggle* in 2019.

I achieved several solo Bronze and Silver medals in worldwide machine learning competitions on the popular Kaggle platform, see my profile [here](#).

### 3 TD Principles in Practice Achievement Award

Granted by *TD Bank* in 2018.

Recognition for my novel research work during internship at Layer 6 AI.

### 2 Mitacs Accelerate Program

Granted by *Mitacs* in 2018.

30k CAD scholarship to support my 8-months internship at Layer 6 AI.

### 1 Singapore International Pre-Graduate Award (SIPGA)

Granted by *A\*STAR* in 2017.

Monthly 1.5k SGD stipend to support my 5-months internship at A\*STAR I2R Visual Computing lab.