

(Version française \*\*\*The English version follows\*\*\*)

# Yuchen Hui

+1-873-992-5859 | [yuchen.hui@umontreal.ca](mailto:yuchen.hui@umontreal.ca) | [GitHub](#) | [Site Web personnel](#)

## ÉDUCATION

**Université de Montréal**  
Maîtrise de recherche en Informatique  
- Advisor: Jian-Yun Nie

Montréal, Canada  
2023.9 -

**Université de Montréal**  
B.S. en Informatique (Cheminement Honor)  
- Moyenne: 4.060/4.3 , Courses: Deep learning, NLP etc.

Montréal, Canada  
2020.9 - 2023.8

## PUBLICATION

Wu\*, Y., Chen\*, K., Zhang\*, T., Hui\*, Y., Berg-Kirkpatrick, T., & Dubnov, S. (2023). Large-scale contrastive language-audio pretraining with feature fusion and keyword-to-caption augmentation. *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP* (arXiv: <https://arxiv.org/abs/2211.06687>)

## EXPÉRIENCES DE RECHERCHE

**Stagiaire / Assistant de recherche chez Mila et LAION**  
Supervisé par Professeure Irina Rish, et Christoph Schuhmann

2022.6 - 2022.10

Collection et prétraitement des jeux de données sous forme de paires d'audio-texte pour le pré-entraînement contrastif de langage-audio à grande échelle (contrastive language-audio pre-training). les répertoires GitHub associé sont: **CLAP Audio Dataset** pour le jeu de données d'entraînement de **2 million** paires d'audio-texte et **CLAP model** pour le model de multi-modal contrastive learning.

- Télécharger des audios et ses descriptions textuelles correspondantes provenant de Sites Web (e.g. freesound.org) à l'aide de robots d'indexation ; Nettoyage et traitement de données ;
- Augmentation de données en utilisant le modèle de langage pré-entraîné T5, permettant de former des phrases a partir des mots-clé;
- Étudier l'effet de contrastive loss sur la performance du CLAP modèle.

**Compétence :** HPC Cluster (Slurm system), Git, Python (Pytorch, pandas, numpy etc.), ffmpeg.

## EXPÉRIENCES DE TRAVAIL

**Stagiaire / Scientifique de données chez Mouvement Desjardins**

2023.5 - 2023.8

Intégrer l'escouade des stagiaires au sein de la Direction Solutions Analytiques Avancées (DSAA) pour mettre en place un projet visant à Prévoir les pannes de flotte de guichets automatiques chez Desjardins en utilisant des techniques de Machine Learning afin de agir proactivement et minimiser l'impact pour les membres et clients.

- Analyser les besoins et la logique d'affaires, prétraiter les données brutes (journal du parc de guichets automatiques), Feature Engineering et créer le jeu de données.
- Formaliser le problème et sélectionner les modèles de base (AutoArima etc.) ainsi que modèles avancés (LSTM, Transformer)
- Développer la solution, affiner & valider les modèles (Hyper-parameter tuning), analyse de résultats
- Intégrer la solution sur le nuage (Azure), en collaborant étroitement avec l'ingénieur MLOps.

**Compétence :** Azure (Azure ML, Azure DevOps), Python (Pytorch, sklearn, pandas, statsmodels etc.), Power BI, communication bilingue (En & Fr).

## EXPÉRIENCES D'ENSEIGNEMENT

**Auxiliaire d'enseignement dans le Tutorat départemental du DIRO**

Hiver 2022, 2023

- Sous forme de Q&A, Couvrir les matières de tous les cours du premier cycle

**Auxiliaire d'enseignement pour le cours de Programmation Procédurale**

Automne 2023

- Enseigner la programmation procédurale en utilisant Python  
- Diriger les séances de laboratoire  
- Corriger les devoirs et projets de programmation

# BOURSES

---

## Bourse d'excellence du DIRO

2020

- \$2000
- TOP 30 Étudiants du premier cycle au DIRO

## Bourse d'exemption des droits de scolarité supplémentaires (Niveau Bacc) 2021.9-2023.4

- \$7000/trimestre
- TOP 2 Étudiants chinois du premier cycle au Québec (sélectionnés par le gouvernement du Québec)

## Bourse d'exemption des droits de scolarité supplémentaires (Niveau Maîtrise) 2023.9-2025

- \$8000/trimestre

# LANGUES

---

- **Français:** Compétence professionnelle
- **Anglais:** Compétence professionnelle
- **Chinois:** Langue maternelle

# PROJET

---

**“Slip” interpréteur**, l'interpréteur du langage de programmation “Slip” (verlan de Lisp)

- Projet pour le cours IFT2035: Concepts des langage de programmation.
- Implanté en un langage de programmation fonctionnel pure: Haskell;
- L'interpréteur d'un dialecte du langage de programmation fonctionnel pure Lisp ;
- Portée dynamique et statique supportées ;
- Lambda calcul supporté (non-typé);
- Élimination du sucre syntaxique.

# TECHNIQUES

---

- **Langages de programmation:** Python, Java, C, Julia, Haskell, Prolog, Shell, HTML5, Markdown, VHDL
- **Outils:** Git, Vim, L<sup>A</sup>T<sub>E</sub>X, Ant, Vscodé, IntelliJ, PyCharm
- **Systems d'exploitation:** Linux, Windows
- **Framework:** Pytorch
- **Autres:** PostgreSQL, Docker, modules de python (numpy, pandas, matplotlib)

(English version \*\*\*la version française précède\*\*\*)

# Yuchen Hui

+1-873-992-5859 | [yuchen.hui@umontreal.ca](mailto:yuchen.hui@umontreal.ca) | [GitHub](#) | [Personal Website](#)

## EDUCATION

---

### Université de Montréal

Master of Science in Computer Science

Montréal, Canada

2023.9 -

- Advisor: Jian-Yun Nie

### Université de Montréal

Honours Bachelor of Computer Science

Montréal, Canada

2020.9 - 2023.5

- **GPA: 4.060/4.3** Courses: Deep learning, NLP etc.

## PUBLICATION

---

Wu\*, Y., Chen\*, K., Zhang\*, T., Hui\*, Y., Berg-Kirkpatrick, T., & Dubnov, S. (2023). Large-scale contrastive language-audio pretraining with feature fusion and keyword-to-caption augmentation. *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP* (arXiv: <https://arxiv.org/abs/2211.06687>)

## RESEARCH EXPERIENCES

---

### Internship / Research Assistant at Mila and LAION

Supervised by Professor Irina Rish, and Christoph Schuhmann

2022.6 - 2022.10

Audio-text dataset collection and preprocessing for large-scale contrastive language-audio pre-training. The associated GitHub repositories are: **CLAP Audio Dataset** for the training dataset of **2 million** audio-text pairs and **CLAP model** for the multi-modal contrastive learning model.

- Download audios and their corresponding textual descriptions from websites (e.g., freesound.org) using web crawlers; Data cleaning and processing;
- Data augmentation using the pre-trained language model T5, enabling the formation of sentences from keywords;
- Investigate the effect of alternating contrastive loss on the performance of the CLAP model.

**Skills:** HPC Cluster (Slurm system), Git, Python (Pytorch, pandas, numpy, etc.), ffmpeg.

## WORK EXPERIENCES

---

### Internship / Data Scientist at Desjardins Group

2023.5.8 - 2023.8.18

Assume the data scientist in a team of 6 interns affiliated to the Advanced Analytical Solutions Department (DSAA) to implement a project aimed at Predicting ATM fleet outages at Desjardins using Machine Learning techniques. This initiative was to act proactively and minimize the impact on members and clients.

- Analyzed business needs and logic, preprocessed raw data (log from the ATM fleet), performed Feature Engineering, and created the dataset.
- Formalized the problem and selected baseline models (AutoArima etc.) as well as advanced models (LSTM, Transformer).
- Developed the solution, refined validated the models (Hyper-parameter tuning), and analyzed results.
- Integrated the solution on the cloud (Azure), closely collaborating with the MLOps engineer.

**Skills:** Azure (Azure ML, Azure DevOps), Python (Pytorch, sklearn, pandas, statsmodels etc.), Power BI, bilingual communication (En Fr).

## SELECTED HONOURS

---

### Departmental Scholarship of Excellence

2020

- \$2000
- TOP 30 undergraduate students in our Computer Science department (among all 1st, 2nd, and 3rd year students)

### Supplementary Tuition Fee Exemption Scholarship

2021.9-2023.4

- \$7000/semester
- TOP 2 Chinese Bachelor students studying in Quebec (selected and funded by the Quebec government)

## TEACHING EXPERIENCES

---

### Teaching Auxiliary in Departmental Tutoring of DIRO

Tutor, help students understand the material

2022.2 - 2023.4

- Q & A covering all undergraduate courses

## LANGUAGES

---

- **English:**      Fluent
- **French:**      Fluent
- **Chinese:**     Native

## PROJECT

---

**“Slip” interpreter**, an interpreter of a programming language “slip” (verlan of lisp)

- Implemented in a pure functional programming language Haskell;
- An interpreter of a pure functional programming language (a dialect of Lisp);
- An interpreter of a language supports both dynamic and static scope;
- Support lambda calculus (not typed!);
- Elimination of several syntactic sugars.

## TECHNICAL SKILLS

---

- **Programming Languages:**      Python, Java, C, Julia, Haskell, Prolog, Shell, HTML5, Markdown, VHDL
- **Tools:**                                Git, Vim, L<sup>A</sup>T<sub>E</sub>X, Ant, Vscode, IntelliJ, PyCharm ...
- **Operating Systems:**                Linux, Windows
- **Framework:**                         Pytorch
- **Others:**                                PostgreSQL, Docker, python's modules (e.g. numpy, pandas, matplotlib,...)