

ÉDUCATION

DEC (Diplôme d'étude collégiales), Sciences pures et appliquées – Marianopolis College	2021 – 2023
<i>Activités parascolaires:</i> Fondateur du Marianopolis AI Club (recrutement de 50 membres actifs; création d'un partenariat avec le AI Launch Lab ; organisation d'ateliers hebdomadaires sur Python et l'apprentissage machine et profond)	
DES (Diplôme d'études secondaires) – Collège Jean-Eudes	2016 – 2021

LANGUES

Français (langue maternelle)
Anglais (avancé)
Cantonais (avancé)
Mandarin (intermédiaire)

COMPÉTENCES TECHNIQUES

Apprentissage machine : Python, Scikit Learn, Tensorflow, Keras, OpenCV, NLTK, Pandas, Kaldi, R, Pytorch
Front-end : React, HTML, CSS, Javascript
Back-end : Node.js, Flask, ASP.NET CORE, Go, Linux

EXPÉRIENCES

Développeur University of Toronto Machine Intelligence Student Team [Site web]	2021
Développement d'un logiciel réduisant significativement le bruit visuel et l'utilisation de données durant les appels vidéos grâce à des technique d'apprentissage profond, du <i>image-to-image translation</i> avec spectrogrammes et de la reconnaissance vocale.	
<ul style="list-style-type: none">Revue de la littérature sur la résolution audioIntégration de <i>Wav2Vec</i> pour effectuer de la reconnaissance vocale en temps réel.Création d'un outil de synthèse vocale utilisant Tacotron avec des vectorisations de parole générées à partir d'un clip audio de 15 secondes.Prétraitement de données avec Pytorch et Audiomentations pour acquérir des <i>Mels Spectrograms</i> et de l'audio artificiellement dégradé.Utilisation de MaskCycleGan-VC pour débruiter de l'audio et obtenir des <i>wav2vec embeddings</i>.Création d'un site web utilisant <i>React</i> et <i>Flask</i> avec lequel les utilisateurs peuvent débruiter leurs clips sonores.	
Stagiaire en science des données Boehringer Ingelheim Canada	2021
Optimisation de l'horaire d'envoi des courriels marketing en analysant des données historiques pour augmenter l'engagement.	
<ul style="list-style-type: none">Prétraitement de données brutes avec R et Panda.Comparaison de solutions d'apprentissage profond comme XGBoost et Random Forest.Utilisation d'algorithmes d'apprentissage machine à l'état de l'art (multitarget encoding, Linear Discriminant Analysis, SMOTE) pour prédire l'horaire d'envoi optimal.Résultat: prédiction de l'heure d'envoi avec une précision de 80% (2.6 fois mieux que le modèle de base).	
Stagiaire en biostatistiques et science des données Boehringer Ingelheim Canada	2020
<ul style="list-style-type: none">Création d'une application RShiny permettant de filtrer et télécharger des essais cliniques de clinicaltrials.gov.Création d'un programme Python utilisation du traitement automatique du langage naturel (NLP) et Google Tesseract pour extraire des caractéristiques de formulaires de consentement numérisés afin de les catégoriser.Contributions à Synthea, un générateur <i>open source</i> de données patients synthétiques, en produisant des données démographiques canadiennes à partir de données provenant de Statistique Canada. [GitHub]	
Gagnant à Pharmahacks 2020 [Devpost] Hackathon (MILA, Montréal)	2019
Création d'un système prédisant le succès d'un essai clinique en utilisant le résultat d'essais similaires.	
<ul style="list-style-type: none">Expérimentation avec des algorithmes d'apprentissage machine tels que XGBoost et Naive Bayes combiné avec des techniques de prétraitement comme Word2vec, TF-IDF vectorizing et count vectorizing.Déploiement d'une solution utilisant des algorithmes d'apprentissage machine en implémentant un Flask REST API.	

PROJETS

- Création d'une chaîne **YouTube** avec plus de 40 vidéos sur l'apprentissage machine et la programmation [\[YouTube\]](#).
- Création de *Summarize And Ask*, un site web qui résume des articles et textes avec un modèle **MobileBERTSUM** et **BERT** pour répondre à des questions à propos du texte [\[GitHub\]](#).
- Création de *iNews*, dans le cadre de l'Expo-Sciences 2020 qui permet d'identifier les fausses nouvelles [\[GitHub\]](#) [Finale montréalaise annulée en raison de la pandémie].
 - Création d'un modèle d'apprentissage machine profond pour détecter la position d'une nouvelle dépassant les performances de l'état de l'art (précision de 85% au lieu de 82%) en appliquant **Word2vec embeddings** et **multi-channel convolutional architecture**.