

## ÉDUCATION

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

<b>Développeur</b>   University of Toronto Machine Intelligence Student Team <a href="#">[Site web]</a>	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"><li>Revue de la littérature sur la résolution audio</li><li>Intégration de <i>Wav2Vec</i> pour effectuer de la reconnaissance vocale en temps réel.</li><li>Création d'un outil de synthèse vocale utilisant Tacotron avec des vectorisations de parole générées à partir d'un clip audio de 15 secondes.</li><li>Prétraitement de données avec <b>Pytorch</b> et Audiomentations pour acquérir des <i>Mels Spectrograms</i> et de l'audio artificiellement dégradé.</li><li>Utilisation de <b>MaskCycleGan-VC</b> pour débruiter de l'audio et obtenir des <i>wav2vec embeddings</i>.</li><li>Création d'un site web utilisant <i>React</i> et <i>Flask</i> avec lequel les utilisateurs peuvent débruiter leurs clips sonores.</li></ul>	
<b>Stagiaire en science des données</b>   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"><li>Prétraitement de données brutes avec <b>R</b> et <b>Panda</b>.</li><li>Comparaison de solutions d'apprentissage profond comme <b>XGBoost</b> et <b>Random Forest</b>.</li><li>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.</li><li>Résultat: prédiction de l'heure d'envoi avec une précision de 80% (<b>2.6 fois mieux</b> que le modèle de base).</li></ul>	
<b>Stagiaire en biostatistiques et science des données</b>   Boehringer Ingelheim Canada	2020
<ul style="list-style-type: none"><li>Création d'une application <b>RShiny</b> permettant de filtrer et télécharger des essais cliniques de <a href="https://clinicaltrials.gov">clinicaltrials.gov</a>. <a href="#">[Site web]</a></li><li>Création d'un programme Python utilisation du traitement automatique du langage naturel (<b>NLP</b>) et <b>Google Tesseract</b> pour extraire des caractéristiques de formulaires de consentement numérisés afin de les catégoriser.</li><li>Création d'un algorithme NLP de prédiction de succès d'essais cliniques basé sur les critères d'inclusion et d'exclusion</li><li><b>Contributions à Synthea</b>, un générateur <i>open source</i> de données patients synthétiques, en produisant des données démographiques canadiennes à partir de données provenant de Statistique Canada. <a href="#">[GitHub]</a></li></ul>	
<b>Gagnant à Pharmahacks 2020</b> <a href="#">[Devpost]</a>   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"><li>Expérimentation avec <b>XGBoost</b> et <b>Naive Bayes</b> combiné avec des techniques de NLP comme <b>Word2vec</b> et <b>TF-IDF</b>.</li><li>Déploiement d'une solution utilisant des algorithmes d'apprentissage machine en implémentant un <b>Flask REST API</b>.</li></ul>	

## 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**.