Je suis ici car Hamed a toujours eu une bonne impression de votre entreprise. Vous travaillez sur des projets variés, alliant recherche et déploiement, et votre entreprise est stable, avec plusieurs années d'existence. Pour ma part, j'ai besoin de stabilité dans ma vie professionnelle

Bonjour, merci beaucoup pour cette opportunité. Je suis ravi(e) d'être ici et de pouvoir discuter de mes compétences et de mon expérience en science des données. J'ai hâte d'échanger avec vous et de contribuer à vos projets.

Bonjour, merci beaucoup pour cette opportunité. Je suis ravi(e) d'être ici et de pouvoir discuter de mes compétences et de mon expérience en science des données. J'espère que mes compétences sont en adéquation avec vos projets et ce que vous recherchez. J'ai hâte d'échanger avec vous et de voir comment je pourrais contribuer à vos initiatives.

J'ai toujours travaillé pour des entreprises anglophones, mais j'ai toujours eu envie d'avoir l'opportunité de travailler pour une entreprise française, car j'aime la langue et la culture françaises, et je vis ici. J'espère avoir cette chance cette fois-ci. Au fil des années, j'ai accumulé une solide expérience en apprentissage automatique et je suis enthousiaste à l'idée de pouvoir l'appliquer dans de nouveaux projets."

Ma thèse de doctorat portait sur l'utilisation du traitement d'images et de l'apprentissage automatique pour classifier des frottis sanguins sous microscope, en utilisant des méthodes statistiques d'apprentissage automatique et des réseaux de neurones convolutifs (CNN)

La société Billtrust a fusionné avec une autre entreprise appelée EQT. Après cette fusion, ils n'avaient plus besoin de tous les employés. Au cours de la dernière année, ils ont licencié progressivement 60 personnes, dont moi à la mi-janvier. Tous les data scientists ont été licenciés

Une de mes faiblesses est mon niveau de français, mais je travaille pour m'améliorer. Mes points forts sont ma détermination, mon travail acharné et ma capacité à m'adapter. Je suis aussi une personne qui aime apprendre et partager des connaissances dans mon domaine

Professionnel en Science des Données et Ingénierie en Apprentissage Automatique avec un doctorat en Informatique, spécialisé en IA Générative, Big Data, Apprentissage Automatique Appliqué et Deep Learning, incluant le Traitement du Langage Naturel (NLP), le Traitement d'Images et les Données Numériques. Expérience dans la conception et le déploiement de pipelines d'apprentissage automatique de bout en bout, avec une expertise en informatique Cloud AWS. Compétences en communication en anglais et en français. Citoyen canadien.

Compétences Informatiques

Programming: Python Database and Query Service: Neo4j, MongoDB, SQL Server & Snowflake ○ ○ ○ ○ ○ ○ Machine Learning Framework: Scikit-learn, TensorFlow, Keras, PyTorch, LangChain Big data Framework: Apache Spark V2, Polars, PySpark Cloud Computing: Amazon Web Services (AWS Bedrock, S3 Bucket, Lambda, Fargate, SageMaker, EMR, ...) Model Deployment & Orchestration: Jenkins, Apache Airflow, Terraform (Terragrunt), Docker, FastAPI, Flask Version Control: Git, GitLab Data Visualization & Interactive App Development: Matplotlib, Seaborn, Bokeh, Streamlit, Chainlit

**Data Scientist Senior – Billtrust – New Jersey, USA (télétravail depuis le Canada)** **Développeur IA pour l'analyse de données structurées et non structurées à grande échelle dans des applications financières** **Novembre 2022 – Présent**

* **Reconnaissance de Factures (Deep Learning) :**
  + Détection des données et des montants des factures à partir de divers formats PDF et images bitmap pour les acheteurs et les fournisseurs.
  + Affinage de la segmentation sémantique à l'aide de la bibliothèque Layout-Parser (Detectron2 et Faster R-CNN).
  + Utilisation des modèles pré-entrainés LayoutLM.
  + Application des méthodes d'égalisation d'histogramme adaptative (AHE) et de flou gaussien pour améliorer la qualité des images.
  + Mise en œuvre du clustering agglomératif en utilisant les extracteurs de caractéristiques ResNet et réduction de la dimensionnalité avec SVD.
  + Provisionnement de l'infrastructure AWS à l'aide de Terraform, automatisation des flux de déploiement avec Terragrunt, et utilisation de Docker pour la gestion des containers.

**Compréhension des Contrats Commerciaux (LLM) :**

* Automatisation de la résolution des requêtes commerciales, y compris la synthèse et l'extraction des termes financiers à partir de plusieurs contrats clients.
* Utilisation des modèles OpenAI (GPT-3.5 Turbo) et HuggingFace (Instructor-XL) pour la Recherche Augmentée par des Données (RAG) avec FAISS VectorStore, etc.
* Génie des invites et clustering de texte avec des embeddings de textes longs.
* Développement d'un cadre d'IA conversationnelle avec mémoire, intégrant des agents LLM et des outils pour transformer les demandes humaines en requêtes à travers des contrats, des données web et des calculs arithmétiques.

**Personnalisation d'e-mails pour les langues européennes (LLM) :**

* Génération automatisée d'e-mails en plusieurs langues et dialectes européens, comme le flamand et le néerlandais standard.
* Utilisation de l'API OpenAI Completion (modèles text-davinci, text-ada-001) et de LangChain pour la mise en œuvre.

**Compréhension des présentations et démonstrations commerciales (LLM) :**

* Résumé et traitement des demandes commerciales issues de réunions vidéo longues entre les représentants de l'entreprise et les clients.
* Utilisation de modèles OpenAI, d'agents d'analyse et d'intégration d'outils pour une communication efficace.
* Amélioration de la classification des produits grâce à RAG, l'apprentissage few-shot et le fine-tuning avec PEFT et LoRA sur FLAN-T5.
* Utilisation de MongoDB Query Language (MQL) pour la récupération de données et gestion des pipelines de données avec Airflow.

**Classification d'e-mails (Deep Learning) :**

* Classification du contenu des e-mails en catégories « Paiement » et « Non-Paiement ».
* Mise en œuvre de l'apprentissage zero-shot avec le framework PyTorch.
* Utilisation de Polars et de requêtes SQL pour la collecte de données à partir de SQL Server.

**Résolution d'entités commerciales (Deep Learning) :**

* Identification et résolution des doublons sémantiques parmi des millions d'enregistrements commerciaux.
* Réalisation d'une analyse de similarité sémantique avec le package Deep Learning Dedup.
* Utilisation de Polars pour une manipulation efficace des données.
* Exécution de requêtes SQL pour extraire des données de Snowflake et gestion de modèles à grande échelle avec le multiprocessing via Ray.

**Marché Total Adressable (TAM) et Valeur à Vie du Client (CLV) (Deep & Statistical Learning) :**

* Estimation du Marché Total Adressable (TAM) par l'analyse des données, des segments clients et des tendances historiques de ventes.
* Utilisation de LightGBM et du Tab Transformer pour prédire la taille du marché et le comportement des clients.

**Instructeur en informatique (en ligne) - McGill, Études continues - Canada** 2020–Présent (à temps partiel)

* Enseignement des cours pratiques en **Machine Learning**, **Deep Learning** et **réseaux de neurones**.

**Data Scientist Senior - Groupe PSL - Montréal, Canada** | Avril 2018–Nov 2022

Développeur en IA pour l'analyse de données structurées et non structurées à grande échelle dans des applications médicales.

**Détection d'attrition (Apprentissage statistique) :**Développement d'un modèle pour détecter les comportements anormaux dans les abonnements aux newsletters par e-mail en utilisant **Isolation Random Forest**.

* Intégration du modèle avec **H2O** et **Amazon EMR** pour améliorer les capacités de traitement.

**Reconnaissance d'entités médicales (Deep Learning) :**

* Identification d'entités médicales (médicaments, symptômes, etc.) à partir de rapports médicaux.
* Application de la **Reconnaissance d'Entités Nommées (NER)** avec **BioBERT** fine-tuné.
* Déploiement du modèle avec **AWS Fargate** et **Lambda** pour une exécution serverless.

**Création de profils médicaux (Deep Learning) :**

* Développement de profils médicaux pour les spécialistes en extrayant des données de diverses ressources et échantillons d'images publiques.
* Utilisation d'**Amazon Rekognition** dans un framework **AWS SageMaker** pour l'analyse d'images.
* Application de **SpaCy** et **NLTK** pour la NER afin d'extraire des entités comme les adresses et les noms de personnes.
* Utilisation de **BioBERT** fine-tuné pour la NER liée aux conditions médicales.
* Déploiement du modèle avec **AWS Lambda** pour une exécution serverless.

**Leaders d'opinion médicaux numériques (Graph et Deep Learning) :**

* Développement d'un graphe représentant les interactions entre diverses entités (professionnels de santé, hôpitaux, patients, etc.) à partir de données textuelles brutes.
* Utilisation de **Neo4j** pour construire un graphe à grande échelle pour une représentation efficace des données.
* Exploitation du **Gremlin Query Language** pour effectuer des mesures de centralité et analyser la structure du graphe.
* Application du **Neo4j Graph Data Science package** pour la détection de communautés et identifier les clusters dans le graphe.