

DANIEL J. DEUTSCH, M.Sc

Data Scientist & Ingénieur IA

📍 Paris, FR 📞 +33 07 45 10 02 91 ✉ daniel.jrgde@gmail.com 🌐 danieljrgde.github.io 🔗 linkedin.com/in/danieljrgde

Data Scientist expérimenté dans la recherche en IA et le développement de solutions de deep learning. Passionné par l'apprentissage et la maîtrise des concepts d'IA, avec une capacité démontrée à naviguer dans des domaines techniques et complexes. Expert en Python, capable d'écrire du code propre et scalable pour concevoir, développer et déployer des systèmes de machine learning.

EXPÉRIENCE PROFESSIONNELLE

VO2 Group

Sept 2023 - Présent

Ingénieur de Recherche en IA

Paris, FR

- Recherche et développement d'un réseau multimodal basé sur des transformers pour l'apprentissage de représentations de tenue vestimentaire, atteignant 84% d'accuracy dans FITB et 96% d'AUC dans la prédiction de compatibilité des vêtements.
- Conception d'un mécanisme d'attention hybride pour le transformer décodeur, améliorant l'intégration par le modèle des descriptions textuelles au niveau des tenues et des vêtements pour des recommandations plus précises et cohérentes.
- Mise en place de stratégies de curriculum learning pour le ranking loss, améliorant les métriques de recall au top-k.
- Réduction de l'overfitting en élargissant le training set avec plus de 60k échantillons web-scrappés, les enrichissant via le modèle image-to-text de LLaVA et le zero-shot classification de CLIP, améliorant la généralisation et les performances.
- Développement d'un système de Retrieval Augmented Generation (RAG) pour améliorer les réponses des LLM dans le domaine des données Environnementales, Sociales, et Gouvernance.
- **Stack technique:** Python, Pytorch, Pandas, Polars, Jupyter, NumPy, Transformers, LangChain, LlamaIndex, Git, Docker, GCP.

Gravite.io

Mai 2023 - Sept 2023

Data Scientist

Paris, FR

- Conception d'un algorithme de topic modeling pour catégoriser les retours clients et identifier les principaux pain points, résultant en une augmentation de 30% des solutions actionnables pour répondre aux préoccupations des clients.
- Optimisation du pipeline d'analyse des retours via parallel processing, réduisant le temps de traitement de 80%.
- Élaboration d'un indicateur de pertinence personnalisée combinant la fréquence des insights et des métriques de gravité, permettant aux parties prenantes de hiérarchiser et de traiter efficacement les problèmes clients les plus critiques.
- **Stack technique:** Python, Scikit-Learn, OpenAI API, Jupyter, Pandas, Git, Docker.

Kaiko

Sept 2021 - Sept 2022

Data Scientist (Stagiaire)

Paris, FR

- Recherche en sentiment analysis sur plus de 6M de publications sur les réseaux sociaux pour vérifier leur corrélation avec les tendances du marché, proposant des métriques pour le développement d'un indice de sentiment.
- Implémentation d'un auto-encodeur variationnel pour détecter 13k comptes bots sur les plateformes de réseaux sociaux, permettant une analyse approfondie de leur influence sur le marché des crypto-monnaies.
- Étude de l'application d'algorithmes de détection d'anomalies sur des données de marché en temps réel, ouvrant la voie à des stratégies de gestion des risques plus robustes et à des capacités de surveillance du marché améliorées.
- Déploiement d'un pipeline PySpark haute performance pour analyser efficacement de grands volumes de données historiques de marché, conduisant à un processus de prise de décision accéléré.
- **Stack technique:** Python, TensorFlow, Scikit-Learn, Transformers, Pyspark, Jupyter, gRPC, Git, Docker.

BTG Pactual

Déc 2019 - Août 2020

Ingénieur Data (Stagiaire)

São Paulo, BR

- Conception d'une architecture Airflow scalable pour orchestrer l'utilisation de machines virtuelles pour l'exécution de scripts, améliorant le suivi des flux de données tout en réduisant les coûts d'infrastructure cloud de 25%.
- Architecture d'un framework serverless pour surveiller l'ingestion, le traitement et l'enrichissement des données de trading, assurant un retour d'information quasi temps réel sur l'intégrité des données.
- Implémentation d'une pipeline CI/CD pour automatiser le déploiement des Airflow DAGs, facilitant les mises à jour.
- **Stack technique:** Python, AWS (Lambda, S3, EMR, EC2, RDS, API Gateway, DynamoDB), Airflow, PySpark, Jupyter, Git, Docker.

ÉDUCATION

Diplôme d'ingénieur | Computer Science, Télécom Paris - Institut Polytechnique de Paris ([vérifier](#))

Master of Science (M.Sc) | Financial Data Science, Université Panthéon-Sorbonne ([vérifier](#))

Diplôme d'ingénieur | Génie Électrique, Universidade de São Paulo ([vérifier](#))

COMPÉTENCES

Programmation Python, C/C++, SQL.

Langues Portugais (langue maternelle), Anglais (courant - C2), Français (courant - C1), Espagnol (intermédiaire - B2).

CERTIFICATS

Structuring Machine Learning Projects, DeepLearning.AI (vérifier)	2024
Improving Deep Neural Networks, DeepLearning.AI (vérifier)	2023
Neural Networks and Deep Learning, DeepLearning.AI (vérifier)	2023

RÉALISATIONS & RÉCOMPENSES

Prix de la Meilleure Thèse de Master de Panthéon-Sorbonne	2022
Bourse de la Coordination pour l'Amélioration du Personnel de l'Enseignement Supérieur	2021
Code Challenge de BTG Pactual, 1ère place	2019
Olympiade Brésilienne de Robotique, médaille d'argent	2016
Olympiade de Mathématiques Kangaroo, médaille d'argent	2016

PROJETS

Effets des Bots de Réseaux Sociaux sur le Marché des Cryptomonnaies (vérifier)	2022
<i>NLP, Analyse de Sentiment, Détection d'Anomalies, Variational Autoencoder, Deep Learning, Test d'Hypothèses.</i>	
Comment les Informations Utilisateurs Alimentent les Recommandations de Produits d'Amazon (vérifier)	2021
<i>Systèmes de Recommandation, Analyse Exploratoire, Random Forest, Feature Importance, Web Scraping, Analyse Statistique.</i>	
Analyse des Incitations Économiques des Faux Avis pour les Restaurants Pariens sur Yelp (vérifier)	2021
<i>NLP, Modèles de Régression, Analyse de Polarité, Feature Engineering, Analyse de Sentiment, Web Scraping.</i>	