DANIEL J. DEUTSCH, M.Sc

AI Engineer & Data Scientist

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

Data Scientist experienced in AI research and the development of deep learning solutions. Passionate about learning and mastering AI concepts, with a demonstrated ability to navigate through technical and complex fields. Highly proficient in Python, skilled in writing clean, robust, and scalable code to design, engineer, and deploy machine learning systems.

WORK EXPERIENCES

VO2 Group Sep 2023 - Present
AI Research Engineer *Paris, FR*

- Researched and developed a multimodal transformer-based network for learning outfit-level representations, achieving 84% accuracy in FITB and 96% AUC in garment compatibility prediction.
- Engineered a hybrid attention mechanism for the transformer decoder, enhancing the model's integration of outfit-level text descriptions and garment features for more accurate and context-aware recommendations.
- Applied curriculum learning strategies for the outfit ranking loss, substantially improving recall at top-k metrics.
- Reduced overfitting by expanding the training set with 60k+ web-scraped samples, enhanced via LLaVA's image-to-text model and CLIP's zero-shot classification, improving model generalization and performance.
- Developed a Retrieval Augmented Generation (RAG) system to improve the responses of large language models in the field of Environmental, Social, and Governance data.
- **Tech stack:** Python, Pytorch, Pandas, Polars, Jupyter, NumPy, Transformers, LangChain, LlamaIndex, Git, Docker, GCP, CUDA.
- **Related topics:** Deep Learnig, GenAI, NLP, LLM, RAG, Prompt Engineering, Contrastive Learning, Data Mining.

Gravite.io May 2023 - Sep 2023
Data Scientist *Paris, FR*

- Architected a topic modeling algorithm to categorize customer feedback and identify main pain points, resulting in a 30% increase in actionable solutions to address client concerns.
- Optimized the feedback analysis pipeline through parallel processing, leading to an 80% reduction in processing time.
- Engineered a custom relevance scoring metric combining insight frequency and severity metrics, enabling stakeholders to efficiently prioritize and address the most critical customer issues.
- **Tech stack:** Python, Scikit-Learn, OpenAI API, Jupyter, Pandas, Git, Docker.
- **Related topics:** Machine Learning, Topic Modeling, NLP, Data Mining, Parallel Computing, Model Evaluation, API Integration.

Kaiko Sep 2021 - Sep 2022
Data Scientist (Intern) *Paris, FR*

- Conducted sentiment analysis research on 6M+ social media posts to verify their correlation with market trends, proposing metrics for the development of a sentiment index.
- Implemented a Variational Autoencoder to detect 13k bot accounts in social media platforms, allowing an in-depth analysis of their influence on the crypto market.
- Investigated the application of anomaly detection algorithms on live market data, paving the way for more robust risk management strategies and improved market surveillance capabilities.
- Deployed a high-performance PySpark pipeline to efficiently analyze massive volumes of historical market data, leading to an accelerated decision-making process.
- **Tech stack:** Python, TensorFlow, Scikit-Learn, Transformers, Pyspark, Jupyter, gRPC, Git, Docker.
- **Related topics:** Deep Learning, Time Series Analysis, Anomaly Detection, NLP, Big Data, Market Analysis.

BTG Pactual Dec 2019 - Aug 2020
Data Engineer (Intern) *Sao Paulo, BR*

- Designed a scalable Airflow architecture to orchestrate the usage of virtual machines for script executions, improving data-flow monitoring while reducing cloud infrastructure costs by 25%.
- Architected a serverless framework to monitor the ingestion, processing, and enhancement of trading data, ensuring near real-time feedback on data integrity.
- Implemented a CI/CD pipeline to automate the deployment of Airflow DAGs, easing updates to financial assets' data flow.
- **Tech stack:** Python, AWS (Lambda, S3, EMR, EC2, RDS, API Gateway, DynamoDB), Airflow, PySpark, Jupyter, Git, Docker.
- **Related topics:** Big Data, Cloud Computing, Workflow Orchestration, Data Monitoring, CI/CD, Distributed Computing.

EDUCATION

Master of Engineering (M.Eng) | Computer Science, Telecom Paris - Institut Polytechnique de Paris (verify)
Master of Science (M.Sc) | Financial Data Science, Universite Pantheon-Sorbonne (verify)
Bachelor of Science (B.Sc) | Electrical Engineering, Universidade de Sao Paulo (verify)

CERTIFICATES

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

ACHIEVEMENTS & AWARDS

Pantheon-Sorbonne's Best Master Thesis Award 2022
Coordination for the Improvement of Higher Education Personnel Scholarship 2021
BTG Pactual Code Challenge, 1st place 2019
Brazilian Robotics Olympiad, silver medal 2016
Kangaroo Mathematics Olympiad, silver medal 2016

PROJECTS

Effects of Social Media Bots on the Crypto Market (verify) 2022
NLP, Sentiment Analysis, Anomaly Detection, Variational Autoencoder, Deep Learning, Hypothesis Testing.

How User Information Drives Amazon's Product Recommendations (verify) 2021
Recommender Systems, Exploratory Analysis, Random Forest, Feature Importance, Web Scraping, Statistical Analysis.

Analysis of Economic Incentives of Fake Reviews for Parisian Restaurants on Yelp (verify) 2021
NLP, Regression Models, Polarity Analysis, Feature Engineering, Sentiment Analysis, Web Scraping.

SKILLS

Programming Python, C/C++, SQL.
Languages Portuguese (native), English (fluent - C2), French (fluent - C1), Spanish (intermediary - B2).