DHEERAJ PARKASH

Paris, Ile-de-France, France
dheerajparkash.github.io dheerajparkash36@gmail.com dheerajparkash +33 7 58 88 06 59

OBJECTIVE

Machine Learning Engineer with strong foundations in ML, NLP, and predictive modeling. Recent Master's Data Science graduate, Universit´e Paris-Saclay with 2 years of research experience, skilled in building scalable generative AI systems and turning models into impactful results. Passionate about driving innovation through AI in cross-functional teams.

EDUCATION

Universite Paris-Saclay & ENS Paris-Saclay **2024-2025**
Masters M2 Data Science, **Grade: 14.7/20** *Paris-France*
Including Disciplinary Electives **(MVA Program)**, **Grade: 17/20**

Universite Cote d'Azur **2023-2024**
Masters M1 Computer Science, **Grade: 14.8/20 Rank 5th/38** *Nice, France*

Sukkur IBA University **2018-2022**
Bachelors of Science Computer Science, **CGPA: 3.48/4** *Sukkur, Pakistan*

EXPERIENCE

ALTEN ***Paris, France***
Data Scientist Intern **April - September 2025**

- Built end-to-end large-scale multimodal data processing pipeline,validated 11.8k segments for downstream modeling.
- Designed and implemented automated pattern detection, and imputation workflows that improved label quality.
- Applied **signal transformations** to improve inter-rater consistency and data reliability.
- Evaluated annotation quality using agreement metrics like **KripAlpha, ICC, CCC, and LORO** analyses.
- Optimized multimodal Transformer models for valence/arousal prediction, increasing CCC from 0.27 $\rightarrow$ 0.62.
- Delivered insights to engineering teams and integrated improved fusion strategy into a deployable prototype.

INRIA/I3S Sophia Antipolis ***Nice, France***
Machine Learning Engineer Intern: **June - August 2024**

- Developed scalable data pipeline for French multi-party conversation to analyse behavior and interaction patterns
- Implemented classification model combining lexical & graph features for hate speech detection, achieving 78.8% F1
- Analyzed user interaction graphs and optimized feature engineering to support product content moderation insights.

Data Scientist intern **March - May 2024**

- Implemented a **structured prediction system** for **aggression** and **biases** detection in social media contents.
- Mapped comment-threads **discursive relationships** to support analytics around user behavior & content patterns.

GenAI Research Intern **October - December 2023**

- Conducted few-shot LLM (GPT-2, T5) fine-tuning and evaluation for text classification and generation tasks.
- Generated synthetic implicit hate speech samples to expand datasets and improve detection coverage.
- Analyzed model outputs to refine dataset quality and reduce false negatives in edge-case categories.

SKILLS

Programming:	Python, C++, JavaScript, SQL, Matplotlib, PyTorch, TensorFlow, Docker, Git, MLflow, MLOps
Analytics & Stats:	Hypothesis testing, regression, A/B tests, metric development, KPI analysis, cohort analysis
ML & AI:	Transformers, NLP, LLMs, multimodal classification, graph ML, Reinforcement learning
Data Workflows:	data cleaning, Data pipelines, ETL design, data quality assessment, feature engineering
Soft	Communication, Analytical thinking, Problem-solving, Cross-functional collaboration.
Language	English(C1), French(A1 -learning)

PROJECTS

Explainability of Text Classifiers in Hate Speech and Sexism Detection **2025**

- Enhanced interpretability of moderation models via token-level necessity & sufficiency metric; Adopted for auditing.
- Fine-tuned BERT on EDOS sexism dataset (14K+ samples) to analyze token importance for 7 protected groups.
- Revealed biases in 7 groups with necessity up to 0.96 and sufficiency 0.93, enhancing fairness and interpretability.

Neural Graph Generation with Specified Properties **2025**

- Explored Graph Transformers for conditional graph generation using global attention and transformer encoders.
- Used T5 and BERT embeddings to condition graph models, reducing link prediction MAE from 0.90 to 0.18.
- Implemented VAE, GAN, and diffusion-based generative models to enhance graph structure generation & diversity.

Optimized Movie Recommendation System **2024**

- Built scalable item-item collaborative filtering model using cosine similarity to recommend movies efficiently.
- Improved runtime and memory efficiency through vectorized operations and thresholding.
- Demonstrated computational optimization relevance for streaming/e-commerce recommender systems..