Nageeta Kumari

nageetaw@gmail.com | linkedin.com/in/nageeta | github.com/nageetaw | nageeta.github.io

Education

Ecole normale supérieure paris saclay — Cachan, France
MVA (Mathématiques, Vision, Apprentissage) (GPA: 16/20) — *Sep 2024 – Sep 2025*

Université Paris-Saclay — Gif-sur-Yvette, France
Master in Data Science (GPA: 15.6/20) — *Aug. 2023 – Aug 2024*

Sukkur IBA University — Sukkur, Pakistan
Bachelor's in Computer Science (GPA: 3.68/4) — *Aug. 2018 – Aug 2022*

Experience

Applied AI and ML Research Intern — April 2025 – October 2025
Datadog — *Paris, France*

- Researched and deployed causal AI models for root cause analysis on distributed trace data.
- Built Trino-SQL + Python data pipelines with unbiased sampling strategy, reducing retrieval time.
- Developed an evaluation framework, achieving **95%** recall and **77%** precision on a manually annotated dataset.

Research Student and Intern — Feb. 2024 – July 2024
INRAE — *Paris, France*

- Harmonized and analyzed **INCA2** and **INCA3** dietary survey datasets for unified classification.
- Built a GPT-3.5 assisted mapping pipeline with human-in-the-loop to align nomenclature with FoodEX2.
- Developed expertise in data harmonization, ontology development, and **semantic technologies**.

Software Engineer — Aug. 2022 – Aug. 2023
SoundM and DLLC — *Florida, United States*

- Developed chat application (FastAPI, Node.js, React) with OAuth2, rate limiting, and end-to-end encryption.
- Designed REST APIs for user management and authentication, ensuring scalability and security.
- Deployed microservices on AWS using Docker, contributed to continuous monitoring, and iterative improvements.

Projects

Modeling for MNAR Data (not-MIWAE) | *PyTorch, Python* | Report — Jan. 2024 – Apr. 2024

- Re-implemented not-MIWAE generative model for handling MNAR data.
- Benchmarked on UCI & stock datasets against MIWAE, MICE, missForest, and KNN, analyzing robustness and scalability.

Composed Image Retrieval | *Vision-Language Models, PyTorch* | Report — Dec. 2023 – Feb. 2024

- Adapted CoVR-BLIP-2 for CIRR dataset; experimented with pooling strategies (mean, max, MLP, attention).
- Improved **Recall@1 by +3%** with attention pooling; gained hands-on experience with contrastive vision-language models.

Patent Match Challenge | *NLP, Transformers, TF-IDF* | Competition — Oct. 2023 – Nov. 2023

- Implemented TF-IDF with dense embeddings (sentence-transformer) for citation matching of patents, boosting recall and mean average precision.
- Fine-tuned **BERT-based models** for paragraph matching, achieving 87.6% validation accuracy.

Compositional Understanding in VLMs | *Vision-Language Models, PyTorch* | Report — Jul. 2023 – Sep. 2023

- Replicated Bags-of-Words in VLMs results with ARO benchmark, extending to new models.
- Showed **Qwen2.5-VL-3B-Instruct** outperforms CLIP/BLIP in word order, relations, and attributes, achieving **+16%** on VGA tasks.

Technical Skills

ML/AI: PyTorch, Transformers, scikit-learn, SciPy, DoWhy; NLP, Probabilistic modeling, Representation learning
Data/Backend: Python, Java, C++, .NET, SQL, NoSQL, Trino, PostgreSQL, MongoDB, Spark, Hadoop, Oracle
Web/App: FastAPI, Node.js, React.js, Next.js, React Native, typescript, Flutter, Flask
Cloud/DevOps: AWS, Google Cloud Platform, Docker, Kubernetes, CI/CD (GitHub Actions)