

Nageeta Kumari

 nageetaw |  Nageeta Kumari |  nageetaw@gmail.com |  Paris, France

SUMMARY

Master's degree in Data Science and Machine Learning (MVA) with over 2 years of experience in software engineering. Passionate about applied AI and data science roles that combine software engineering, and seeking a full-time position in early career roles.

WORK EXPERIENCE

Datadog - Applied AI Research Intern

April 2025 - Present

- Conducted research on **root cause analysis** methods for distributed microservices **trace data**.
- Formulate problem then designed and implemented **models** for root cause analysis.
- Proposed an unbiased sampling strategy for data collection and wrote corresponding data pipelines.
- Designed an unbiased evaluation framework to compare model's results with their model, achieving **95%** recall and **67%** precision; currently working on improvements.
- Gained hands-on experience with AWS, Trino-SQL, NBT, Mortor, Docker, and Unsupervised Learning.
- Built strong working relationships in a diverse, multicultural environment.

INRAE - Research Intern

May 2024 - July 2024

- Harmonized and analyzed **INCA2** and **INCA3** dietary survey datasets for unified classification.
- Creates Pipeline to translate nomenclature from INCA2 using GPT 3.5 and mapped it to FoodEX2.
- Developed hands-on expertise in data harmonization, ontology development, **semantic technologies**, and pre-processing real world users data.

SoundM and DLLC - Software Engineer

August 2022 - August 2023

- Built secure and user-friendly chat functionality for individual and group communication.
- Created APIs for profile management and authentication; collaborated on front-end design.
- Hands-on experience with React.js, Node.js and Fast API and deployment through **AWS**.

PROJECTS

RelevAI-Reviewer

[Link to Github](#)

- Built an AI-powered benchmark system to rank survey-paper relevance to **call-for-papers** prompts using a novel dataset of **25,164** instances.
- Fine-tuned **BERT** with thermometer label encoding, outperforming baselines (e.g., SVM) and enabling scalable high-accuracy ranking.

Research on Chain of Tables

[Link to Github](#)

- Investigated table-based QA methods, comparing [Chain-of-Table](#) vs. [ReAcTable](#).
- Conducted experiments on WikiTQ dataset with **contrastive** and **few-shot prompting** strategies.

Modeling for MNAR Data (not-MIWAE)

[Link to Report](#)

- 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 w/ Enhanced Pooling

Link to Report

 - Adapted CoVR-BLIP-2 for CIRRR dataset and tested pooling strategies (mean, max, MLP, attention).
 - Demonstrated **3% Recall@1 improvement** with attention pooling; gained hands-on experience with contrastive vision-language models.
- Evaluating Compositional Understanding in VLMs

Link to Report

 - 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.

EDUCATION

2024 - 2025	MVA (Mathématiques, Vision, Apprentissage) at ÉNS Paris-Saclay	(GPA: 16/20)
2023 - 2024	Master in Data Science at Université Paris-Saclay	(GPA: 15.5/20)
2018 - 2022	Bachelor's in Computer Science at Sukkur IBA University	(GPA: 3.68/4)

TECHNICAL SKILLS

Languages and Frameworks	Java, C++, C, Python (Hugging Face, PyTorch, scikit-learn, Seaborn, SciPy, DoWhy) , React.js, JavaScript, typescript, Node.js, Flutter, React Native, Next.js, Flask.
Databases & Big Data:	SQL, NoSQL PostgreSQL, MongoDB, Trino, Hadoop, Spark
Cloud & Deployment Tools:	Google Cloud Platform (GCP), AWS, Docker, Kubernetes
Machine Learning Models:	Self-Supervised Learning, Unsupervised Learning, LLMs (BERT, GPT, Mistral), Generative Models (VAE, GANs), Transfer Learning (Domain Adaptation)
Other:	Git, CI/CD pipelines

LANGUAGES

English	Professional working proficiency
French	Beginner