

About the job

What we are looking for :

As LLM Engineer, the purpose of this role is to work with our team for the end-to-end deployment of LLMs to benefit our products.

Please find our research page : <https://research.fynd.com/>

Key responsibilities:

Work with Python, LLM/GenAI frameworks and tools to create end to end ML pipelines from training to deployment

Responsible for the end-to-end deployment of LLMs including scoping, testing, implementation, maintenance, tracking, and optimization of predictive models.

LLM Model Fine-Tuning: Utilize your expertise in encoding and language understanding to fine-tune existing Large Language Models for optimal performance in solving domain-specific needs.

Model Training and Evaluation: Conduct rigorous testing and evaluation of LLMs to identify areas for improvement and implement solutions to enhance their performance.

Continuously research and stay up-to-date with recent advancements in NLP and large language models, applying novel techniques and methodologies to improve our models.

CI/CD pipelines development, LLM model containerizing and deployment on cloud or premise. Models testing and follow-up maintenance. All stages of ML model life cycle ensuring and support.

Design prototypes and POCs to demonstrate solution feasibility and value. Provide architecture solution.

Required Qualifications:

Bachelor's Degree (might be final course student) in Statistics, Applied Mathematics, Computer Science, or other related fields

Proficiency in Python programming, with working experience over one of the ML Frameworks - SKLearn, XGBoost, PyCaret, and Deep Learning Frameworks - Tensorflow, PyTorch, Huggingface, and Rasa

1+ years of professional experience with NLP and Large Language Models (LLM)

2+ years of experience in implementing and deploying machine learning and deep learning frameworks on cloud platforms like AWS and GCP.

Experience working with End to End Machine Learning project Life cycle .