

Azure AI/ML Engineer

Job description

Company Overview:

Excellanto Ventures is at the forefront of leading the way, powering next-generation enterprises with comprehensive Digital and IT Engineering end-to-end solutions. As a global provider of IT outsourcing services, our commitment to excellence, fortified by years of experience and domain knowledge proficiency, allows us to partner closely with our clients, catalyzing transformational value that grants a sustainable competitive edge.

Position Overview

We're looking for an Azure AI/ML Engineer to join our team at our India office in Kolkata and develop cutting-edge solutions for our US clients using Microsoft's latest cloud technologies.

In this role, you'll be responsible for:

- Designing, developing, and deploying AI models and software using Azure Machine Learning (AzureML) and Azure OpenAI services.
- Building RAG implementations to power real-time decision-making and knowledge management for our banking clients.
- Collaborating closely with IT engineers and professionals to understand their unique needs and tailor AI solutions to optimize operations.
- Implementing and managing embeddings, vector databases, and retrieval mechanisms for efficient data handling and retrieval.
- Maintaining and updating AI models to ensure accuracy and optimal performance under various conditions.
- Integrating AI technologies seamlessly with existing client systems and workflows through collaboration with cross-functional teams.
- Staying at the forefront of AI advancements by researching and exploring new tools and technologies to drive project success.
- Ensuring compliance with industry standards and upholding the highest levels of data security and privacy.

To be successful, you'll need:

- Proven expertise in Microsoft Azure, with hands-on experience in AzureML and Azure OpenAI services.
- A strong background in AI model implementation, particularly in RAG, Corrective RAG, and related technologies. You will be working with langchain and langroid implementations
- Experience with vector databases and embedding technologies. Including ChromaDB, Semantic Search, Pinecone, etc.

- The ability to work with complex datasets and deliver scalable, efficient solutions.
- A strong background in Python programming and associated tools such as Conda, Mamba, etc.
- Excellent experience in team programming using GitHub and Gitflow, CI/CD, Agile/Scrum etc.
- Working experience deploying to Linux OS locally (we use WSL) and in cloud-hosted environments
- Excellent problem-solving skills and the ability to thrive in a dynamic, fast-paced environment.
- Exceptional communication skills to collaborate effectively with both technical and non-technical stakeholders.

Join our team and be a part of building the future!

Job Types: Full-time, Permanent

Benefits:

- Flexible schedule

Location Type:

- In-person

Schedule:

- Day shift
- Monday to Friday
- Weekend availability

Application Question(s):

- Are you willing to relocate to Kolkata if selected?

Education:

- Bachelor's (Preferred)

Experience:

- Python: 5 years (Required)
- Total Work: 7 years (Required)

Work Location: In person