Le Ha Minh Anh

AN6001 – AI and Big Data in Business

Group B

December 8th 2024

Week 1 & 2 Assignment

1. **How do cloud platforms enable the storage and processing of large datasets required for real-world AI projects?**

Cloud platforms provide scalable database technologies and high-performance data processing to support AI projects that involve large datasets. The scalability of cloud platforms allows the system to automatically adjust resources to accommodate the large growing datasets without human intervention. Additionally, preprocessing features prepare and transform raw, complex data into high-quality and machine-learning-ready resources, allowing more efficient AI workflows. These features ensure that AI projects can handle the massive data volumes required for training, testing, and deploying models while remaining efficient and scalable.

1. **Why are web applications essential for providing user-friendly interfaces to deploy and interact with AI models?**

Web applications are essential for providing user-friendly interfaces for AI model deployments because they focus on usability through easy access and simple and instant user interaction. Using web applications, users can access AI models from any device with a browser, eliminating the need for software installation. Additionally, web applications provide simple and intuitive user interfaces and real-time request processing, allowing non-technical users to interact easily with AI models without the need to understand the fundamental workflow behind the models or advanced technical knowledge.

1. **How do cloud services support the scalability of AI projects when handling high volumes of real-time data?**

One of many features of cloud platforms is elastic computing power, allowing them to scale computing resources flexibly depending on the real-time demands. Some object storage solutions offered by these platforms such as Google Cloud and AWS are optimized to manage large datasets, ensuring they are processed efficiently without performance bottlenecks due to immediate storage need. Moreover, real-time processing capabilities of cloud services ensure continuous, low latency, and parallel processes across multiple data streams. By leveraging these cloud service capabilities, AI projects can achieve unprecedented scalability, handling massive real-time data volumes with agility and efficiency.

1. **How do web applications facilitate the integration of AI models into end-user environments for real-world applications?**

Web applications facilitate the integration of AI models into end-user environments by providing accessible, scalable, and user-friendly platforms. Web applications can be integrated with other tools and systems, allowing a widespread adoption of AI models into users’ workflow. Moreover, they include visual dashboards, charts, and interactive tools to display AI-generated insights in a simple way for users to follow. The user interface design also allows users to customize the model’s settings for domain-specific requirements without the need for advanced technical understanding, becoming more suitable for broader implementations.

1. **How do cloud-based AI services simplify model training, deployment, and monitoring in real-world scenarios?**

Cloud-based AI services simplify model training, deployment, and monitoring by providing scalable database technologies, automated machine learning tools, and integrated development environments. With powerful and scalable infrastructure, prebuilt tools, and automated training pipelines, these services enable faster training of large and complex models. Serverless deployment and tools to expose models as APIs reduce the infrastructure complexity and make them accessible to applications and end-users. Cloud services also provide tools to monitor model performance and maintain reliability over time, ensuring the long-term effectiveness and efficiency of AI models in real-world applications.