

Part 4 - Reflection & Workflow Diagram

Q1: What was the most challenging part of the workflow? Why?

Question 1

What was the most challenging part of the workflow? Why?

The most challenging part of the AI development workflow is typically the Data Preprocessing stage. This is because:

- **Data Quality Issues:** Real-world data is often messy, incomplete, or inconsistent, requiring extensive cleaning and validation
- **Feature Engineering:** Determining which features to extract, transform, or create requires domain expertise and experimentation
- **Scale and Complexity:** Processing large datasets efficiently while maintaining data integrity can be computationally intensive
- **Domain-Specific Requirements:** Different AI applications (NLP, computer vision, time series) require specialized preprocessing techniques

Q2: How would you improve your approach with more time/resources?

With additional time and resources, I would improve the approach by:

1. **Automated Data Pipeline:** Implement automated data validation and preprocessing pipelines using tools like Apache Airflow or Prefect
2. **A/B Testing Framework:** Set up systematic experimentation to compare different preprocessing techniques and model architectures
3. **MLOps Integration:** Implement comprehensive model versioning, monitoring, and automated retraining using MLflow or Weights & Biases
4. **Cross-functional Collaboration:** Involve domain experts earlier in the process for better feature engineering and validation
5. **Scalable Infrastructure:** Utilize cloud-native solutions (Azure ML, AWS SageMaker) for distributed training and deployment

Diagram

AI Development Workflow

