

# **Data Engineering for AI Systems**

## **Data Pipelines**

Automated flows that ingest, clean, and transform data.

## **ETL vs ELT**

ETL transforms before loading; ELT transforms after loading.

## **Data Quality**

Accuracy, completeness, and consistency are critical for ML systems.

## **Storage Systems**

Data lakes, warehouses, and object storage support AI workloads.

## **Importance for RAG**

Reliable pipelines ensure high-quality documents for retrieval.