

International Conference on Artificial Intelligence and Cybersecurity (ICAIC 2025)

Conference Proceedings

November 27-28, 2025 - Japan

Artificial Intelligence in Data Engineering: Use Cases, Challenges, and Future Directions

Joseph Aaron Tsapa

Birla Institute of Technology and Science, Pilani, USA

Abstract

AI is reshaping data engineering from batch plumbing to adaptive, feedback-driven systems. The talk maps concrete points of impact across the pipeline: LLM-assisted ingestion and schema inference; automatic ELT code synthesis and transformation validation; contract-aware quality checks with anomaly and drift detection; lineage reconstruction and metadata enrichment; cost- and carbon-aware orchestration; and policy-as-code access controls. Case patterns highlight measurable outcomes—faster onboarding of new data sources, lower incident rates, and improved reliability of downstream analytics and ML.

Adoption remains nontrivial. Key risks include hallucinated mappings, brittle prompts, hidden data leakage, unclear accountability in human-AI handoffs, and regulatory exposure under privacy and sovereignty rules. An evaluation-first approach is outlined: task-level metrics (e.g., precision/recall for anomaly detection, F1 for mapping), shadow and canary deployment, rollback strategies, and red-teaming of prompts and models. A reference architecture separates a governed control plane (model/prompt registry, guardrails, policy engine, audit) from an execution plane (DAG compiler, vectorized retrieval for context, observability with lineage-aware alerts), with human-in-the-loop checkpoints anchored in data contracts and SLAs.

Looking ahead, the field is trending toward agentic pipelines that negotiate data contracts, privacy-preserving transformation (differential privacy, federated execution, synthetic data), multimodal ETL, standardized benchmarks for DE tasks, and greener scheduling objectives. Attendees gain a decision framework—what to automate, how to measure it, and how to scale it safely—without sacrificing governance, cost control, or trust.