

The Role of AI and Modern Data Pipelines in Driving Innovation



Task 2: Aim & Objectives and Research Area

08 July 2025 12:11

Q1. Project Aim and Objectives

Aim:

To investigate how artificial intelligence (AI) and modern data pipeline technologies enable innovation across industries, with a focus on scalability, automation, and real-time decision-making.

- Objectives:
- To review literature on AI-driven innovation and the evolution of data engineering practices in the past decade.
- To analyse the architecture and capabilities of modern data pipelines such as Delta Live Tables, Apache Kafka, and Databricks.
- To explore real-world case studies demonstrating how businesses use AI and automated pipelines to drive value creation.
- To assess the challenges and limitations associated with implementing AI and modern data workflows, including data quality, ethics, and infrastructure.
- To propose best practices for organisations aiming to integrate AI and modern pipelines to accelerate digital transformation.

Q2. Research Philosophy

The most appropriate research philosophy for this project is positivism, supported by the core philosophical elements of ontology, epistemology, and axiology.

Ontology assumes a single, objective reality—that the impact of AI and modern data pipelines on innovation can be measured independently of human perception. This is appropriate when dealing with quantifiable metrics such as data throughput, latency, and system performance [1].

Epistemology in this context is based on observable and measurable evidence. Knowledge is obtained from experimental data, performance benchmarks, and system behaviour, often using real-time streaming frameworks such as Apache Kafka or Delta Live Tables [2].

Axiology within positivism stresses researcher neutrality. This project will avoid subjective interpretation and instead rely on objective findings from data pipelines, automation frameworks, and AI systems. The evaluation of tools like Databricks or Spark will focus on performance rather than preference [1].

Overall, positivism offers a structured and rigorous methodology suitable for computing projects. It ensures that the research remains evidence-based, replicable, and aligned with industry practices [1].

Q3. Search Tools and Research Strategies

To access and analyse high-quality academic information, the following search tools and databases will be used:

1. Research Databases:

- IEEE Xplore – for leading publications on AI, big data infrastructure, and computing systems [1].
- ScienceDirect – for interdisciplinary research in technology, innovation, and business

transformation [3].

- ACM Digital Library – to explore computing frameworks, software engineering, and machine learning systems [4].
- Google Scholar – to access conference proceedings and supplementary grey literature.

2. University Tools:

Federated search platforms such as EBSCOhost and ProQuest will be accessed via the university library to cross-search articles from computing, data science, and innovation-focused journals.

3. Search Strategies:

Boolean operators and keyword strings will be applied. Examples include:

- "AI" AND "data pipelines" AND "innovation"
- "Delta Live Tables" OR "Apache Kafka" AND "real-time analytics"
- "AI infrastructure" AND "automation" AND "value creation"
- "data engineering" NOT "data science" AND "enterprise innovation"

Additional techniques such as date filters (2020–2025), peer-reviewed limits, and citation snowballing will be used to identify high-impact studies.

References

- [1] M. Saunders, P. Lewis, and A. Thornhill, *Research Methods for Business Students*, 8th ed. Harlow, UK: Pearson Education, 2019.
- [2] M. Zaharia et al., "Delta Lake: High-performance ACID table storage over cloud object stores," *Proc. VLDB Endowment*, vol. 13, no. 12, pp. 3411–3424, 2020.
- [3] A. Ghodrati and M. Rahmani, "Data pipelines and AI: The engine behind scalable innovation," *J. Data Sci. and AI*, vol. 4, no. 1, pp. 22–34, 2021.
- [4] A. Raj, G. Dwivedi, and R. Srivastava, "The role of AI in business model innovation," *AI & Society*, vol. 38, no. 1, pp. 105–122, 2023.
- [5] D. Chugunov et al., "Real-time big data pipelines using Apache Kafka and Spark," *Int. J. Comput. Sci.*, vol. 10, no. 3, pp. 89–101, 2022.

Task 3: Research Strategies

08 July 2025 12:11

Q1. Validity, Reliability, Generalisability, and Ethical Considerations

Validity

Validity ensures the research measures what it is intended to measure (Creswell & Creswell, 2018).

This project will:

- Define terms like “innovation,” “modern data pipeline,” and “AI adoption” precisely to avoid misinterpretation.
- Triangulate data from academic research, industry case studies, and practitioner interviews.
- Use pilot-tested surveys and interview guides to ensure clarity and relevance.

Reliability

Reliability ensures research results are consistent over time and across contexts. To maintain reliability:

- Apply standardised questions for surveys and interviews.
- Maintain a consistent coding framework for qualitative data analysis.
- Use version control for collected datasets (e.g., Git or secure cloud storage).

Generalisability

Generalisability assesses whether findings can apply to other contexts. For this project:

- Include examples from multiple industries (e.g., finance, healthcare, manufacturing).
- Target respondents from organisations of different sizes and technical maturity.

Ethical Considerations

- Obtain **informed consent** from all participants.
- Anonymise all identifying details and store data securely.
- Seek university ethical approval before starting data collection.
- Comply with **British Computer Society (BCS) Code of Conduct** for responsible handling of data.

Q2. Research Strategies and Pilot Data Collection

Research Strategy

A **mixed-methods** approach will be used:

- **Qualitative:** Semi-structured interviews with AI engineers, data pipeline architects, and innovation leads.
- **Quantitative:** Online surveys to measure adoption rates, perceived innovation impact, and organisational readiness.

Data Sources

- Academic databases: IEEE Xplore, ACM Digital Library, ScienceDirect.
- Industry reports: Gartner, McKinsey, Databricks technical documentation.
- Technical blogs, conference papers, GitHub repositories.

Pilot Study

- **Aim:** Test clarity and relevance of survey questions.
- **Participants:** 3 industry professionals (2 data engineers, 1 AI solutions architect).
- **Process:** Distribute pilot survey and conduct brief interviews for feedback.
- **Findings:** Simplified technical jargon, added a question on “data governance,” removed a redundant question on “legacy infrastructure.”

See Appendix A for the pilot questionnaire and feedback notes.

Q3. Quantitative and Qualitative Analysis

Quantitative Analysis

- Descriptive statistics (mean, median, mode) for survey responses.
- Correlation analysis between AI adoption maturity and reported innovation metrics.
- Visualisation of trends using Python (Matplotlib, Seaborn).

Qualitative Analysis

- Thematic coding of interview transcripts to identify themes like “automation benefits,” “skills gap,” “data quality challenges.”
- Use NVivo for coding to maintain transparency and audit trail.
- Compare qualitative themes with quantitative patterns for integrated insights.

Integration of Findings

A **convergent parallel design** will be applied: both qualitative and quantitative data analysed independently, then merged to form a unified interpretation of how AI and modern data pipelines drive innovation.

Task 3 – References

[1] J. W. Creswell and J. D. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 5th ed. Thousand Oaks, CA, USA: Sage, 2018.

[2] British Computer Society, *Code of Conduct*. London, UK: BCS, 2022. [Online]. Available: <https://www.bcs.org/membership/become-a-member/bcs-code-of-conduct/>

[3] Databricks, *Delta Live Tables Documentation*. San Francisco, CA, USA: Databricks, 2024. [Online]. Available: <https://docs.databricks.com/en/delta-live-tables/index.html>

[4] Gartner, *Emerging Technologies and Innovation Insights*. Stamford, CT, USA: Gartner, 2023. [Online]. Available: <https://www.gartner.com/en/information-technology/insights/emerging-technology>

[5] McKinsey & Company, *The State of AI in 2023*. New York, NY, USA: McKinsey, 2023. [Online]. Available: <https://www.mckinsey.com/business-functions/mckinsey-analytics>