Sarcouncil Journal of Multidisciplinary

ISSN(Online): 2945-3445

Volume- 05| Issue- 07| 2025

Review Article **Received:** 01-06-2025 | **Accepted:** 17-06-2025 | **Published:** 05-07-2025

AI-driven ETL: Leveraging GenAI for Query Processing and Data Integration

Achyut Kumar Sharma Tandra
The University of Texas at Dallas, USA

Abstract: ETL processes traditionally require complex pipelines, multiple coding languages, and considerable development effort. The advent of AI-driven query layers revolutionizes this landscape by enabling AI-driven ETL powered by generative AI. By leveraging prompts as an intuitive query interface, businesses can automate data extraction, transformation, and loading without requiring SQL or procedural scripting. This article presents an architecture that utilizes AI models trained on database querying and transformations, allowing users to define ETL workflows in natural language. Generative AI models serve as an abstraction layer, translating human-readable instructions into optimized ETL jobs. With integrations into cloud data warehouses, businesses eliminate redundancy in manual operations while achieving real-time processing. Moreover, AI-driven ETL enhances self-service analytics, making data engineering more accessible across teams. Security and governance mechanisms prevent erroneous data manipulations, ensuring consistency and reliability in pipeline operations. The introduction of AI-driven ETL using AI query layers democratizes data engineering, allowing enterprises to scale operations efficiently while reducing dependency on specialized coding expertise.

Keywords: AI-driven ETL, Generative AI, Natural Language Processing, Self-service Analytics, Data Democratization.

INTRODUCTION

Extract, Transform, Load (ETL) processes have long been the backbone of enterprise data management, facilitating the movement and transformation of data across disparate systems. Recent research indicates that organizations across various industries are allocating substantial portions of their IT budgets toward data integration activities, with ETL processes representing a significant investment (Behera, L., & Chilukoori, V. V. R. 2025). Traditionally, ETL implementation has demanded extensive technical expertise across multiple programming languages, database systems, and data manipulation frameworks. Organizations have invested significant resources in developing and maintaining complex ETL pipelines, often requiring specialized teams dedicated to data engineering tasks. Studies show that enterprises typically maintain numerous distinct ETL workflows, each requiring considerable person-hours to develop, deploy, and maintain, creating a substantial operational burden on technical teams (Behera, L., & Chilukoori, V. V. R. 2025).

The emergence of artificial intelligence, particularly generative AI models with advanced natural language understanding capabilities, presents a paradigm shift in how ETL processes can be conceptualized and implemented. Industry analyses suggest that organizations implementing AI-assisted data engineering solutions are experiencing notable reductions in ETL development time and maintenance costs, allowing for more efficient resource allocation (Thoutam, P. 2024). This paper explores a novel approach to ETL - one that leverages AI as an intuitive query layer, enabling AI-driven data extraction, transformation, and loading through natural language prompts. Early implementations of AI-driven ETL solutions demonstrate promising results in correctly interpreting complex transformation requirements expressed in natural language, with accuracy improving significantly after minimal prompt refinement (Thoutam, P. 2024).

By abstracting away the technical complexity traditionally associated with ETL workflows, this AI-driven approach democratizes data engineering, empowering business users and analysts to define and execute sophisticated data pipelines without writing a single line of code. Research involving business analysts across multiple industries indicates that participants can successfully implement ETL workflows using natural language prompts after minimal training, compared to the substantially longer time required to achieve proficiency in traditional ETL tools (Thoutam, P. 2024). This transformation not only accelerates time-to-insight but also reduces the operational burden on specialized data teams, allowing them to focus on higher-value activities such as architecture optimization and advanced analytics. The shift toward AI-driven ETL solutions enables data engineering teams to spend less time on routine maintenance and more time on strategic data initiatives, resulting in measurable improvements in overall productivity while simultaneously making data manipulation capabilities accessible to a broader range of business stakeholders (Behera, L., & Chilukoori, V. V. R. 2025).

***Corresponding Author:** Achyut Kumar Sharma Tandra
DOI- https://doi.org/10.5281/zenodo.15812771

THE EVOLUTION OF ETL METHODOLOGIES

Traditional ETL Frameworks

ETL methodologies have undergone significant evolution since their inception. Early ETL processes relied heavily on manual coding, with engineers developing custom scripts to extract data from source systems, transform it according to business rules, and load it into target repositories. Research examining traditional ETL implementations has highlighted the considerable complexity involved in maintaining script-based approaches, with organizations reporting substantial code volumes required even for relatively straightforward data transformations (Kola, H. G. & Taqa, A. R. 2023). These scripts were often written in languages such as SQL, Python, or Java, requiring deep technical expertise and intricate knowledge of source and target systems. Studies indicate that organizations employing traditional ETL methods faced significant challenges in terms of knowledge transfer and onboarding new team members, as the learning curve for understanding existing implementations was particularly steep (Kola, H. G. & Taqa, A. R. 2023).

The Shift to Visual ETL

The development of specialized ETL tools introduced visual interfaces that reduced the coding burden. These platforms offered drag-and-drop components representing various ETL operations, allowing engineers to design workflows visually. According to industry analyses, organizations implementing these visual ETL platforms experienced notable reductions in implementation time compared to traditional coding approaches, with project timelines decreasing substantially for similar complexity integration scenarios (Pavithra, M. 2025). However, while these tools simplified implementation, they still required considerable technical understanding and specialized training. Research suggests that despite the visual nature of these platforms, practitioners still needed significant training time to achieve proficiency, and organizations reported that many Visual ETL implementations still required occasional custom scripting to handle complex transformation logic or edge cases (Pavithra, M. 2025).

Cloud-Native ETL Solutions

The advent of cloud computing further transformed ETL practices with the introduction of cloud-native solutions. These services introduced managed infrastructures and serverless architectures, reducing operational overhead. Comprehensive studies of cloud-based ETL implementations across diverse organizations have documented operational cost reductions compared to on-premises alternatives, with particularly significant savings in infrastructure maintenance and disaster recovery provisions (Kola, H. G. & Taqa, A. R. 2023). The elasticity of these platforms proved especially valuable for variable workloads, with surveyed organizations reporting the ability to handle multiple times their normal processing volumes during peak periods without service degradation or additional configuration (Kola, H. G. & Taqa, A. R. 2023). Yet despite these advancements, users still needed to define transformations using domain-specific languages or SQL variants. Proficiency metrics collected across multiple industries indicate that data practitioners require substantial time to become fully productive with cloud-native ETL platforms, representing an improvement over traditional approaches but still presenting a significant learning curve for non-technical users (Pavithra, M. 2025).

Table 1: From Script-Based to AI-Driven Approaches (Kola, H. G. & Taqa, A. R. 2023; Pavithra, M. 2025)

ETL Methodology	Primary Characteristics
Traditional Script-Based ETL	Manual coding; high technical expertise; difficult knowledge transfer
Visual ETL Platforms	Visual interfaces; reduced coding; still requires technical training
Cloud-Native ETL Solutions	Managed infrastructure; elastic scaling; requires domain languages
AI-driven ETL	Natural language interface; business user accessible; no programming
Future Adaptive ETL	Self-learning systems; multimodal interfaces; cross-domain knowledge

AI-DRIVEN QUERY LAYER ARCHITECTURE

Conceptual Framework

The AI-driven query layer represents a fundamental reimagining of ETL architecture. At its core lies a generative AI model specifically trained on database querying patterns, transformation logic, and data loading protocols. This model acts as an intermediary between human intent, expressed in natural language, and the technical execution of ETL operations. Research into AI-powered ETL systems demonstrates that language models fine-tuned on SQL query datasets can achieve significant accuracy in translating

Publisher: SARC Publisher

natural language requests into valid database operations across multiple database management systems (Seenivasan, D. 2024). The architectural philosophy behind these systems prioritizes abstraction of technical complexity while maintaining precision, with studies indicating that properly calibrated models can substantially reduce the semantic gap between business requirements and technical implementation compared to traditional approaches (Seenivasan, D. 2024).

Components of the AI Query Layer

The architecture comprises several key components that work in concert to translate natural language requests into executable ETL operations:

Natural Language Interface

A conversational frontend that accepts prompts from users, allowing them to express ETL requirements in plain English. Studies of these interfaces have shown they can correctly interpret a high percentage of first-time user requests without additional clarification, with this figure improving after conversational refinement (Crabtree, M. 2024).

Intent Recognition Engine

A neural network module that parses natural language inputs to identify the underlying ETL operations being requested. Research into intent classification for data engineering tasks shows that transformer-based models achieve strong performance across common ETL operation categories, outperforming previous rule-based and classical machine learning approaches (Crabtree, M. 2024).

Knowledge Graph

A semantic representation of data sources, fields, relationships, and transformation capabilities, enabling the system to understand the context of requests. Implementations incorporating metadata from both source and target systems show improved transformation accuracy and reduced ambiguity-related errors compared to systems relying solely on natural language understanding (Seenivasan, D. 2024).

Execution Planner

A component that converts recognized intents into optimized execution plans, considering factors such as data volumes, processing constraints, and performance considerations. Research indicates that AI-driven planners can develop execution strategies that approach the efficiency of human expert-designed workflows while requiring significantly less planning time (Seenivasan, D. 2024).

Connector Framework

A library of pre-built integrations with various data sources and targets, allowing seamless interaction with diverse systems. Analyses suggest that modern connector frameworks can substantially reduce integration development time compared to custom API development while maintaining the functionality and flexibility (Seenivasan, D. 2024).

Prompt Engineering for ETL Operations

Effective interaction with the AI query layer depends on thoughtful prompt engineering. While the system accommodates conversational inputs, certain prompt structures yield more efficient and accurate results (Crabtree, M. 2024):

Source-Transform-Target Pattern

Prompts that clearly specify the data source, desired transformations, and the target repository tend to produce the most reliable outcomes. Research shows that using this structured pattern increases transformation accuracy and reduces execution time compared to unstructured prompts (Crabtree, M. 2024). For example, prompts following the pattern "Extract customer data from [source] where [condition], transform [fields] by applying [operation], and load into [target]" consistently outperform more ambiguous instructions.

Contextual References

The system maintains conversational context, allowing users to refine operations incrementally without repeating all parameters. Studies of user interaction patterns reveal that context-aware systems reduce prompt length in multi-step ETL workflows while maintaining semantic precision (Crabtree, M. 2024). This capability is particularly valuable for complex data pipelines where iterative refinement is common.

Transformation Chaining

Complex transformations can be expressed as sequences of simpler operations, enabling the handling of sophisticated data manipulation requirements. Analysis indicates that breaking down complex transformations into sequenced operations improves execution success rates for advanced use cases while simultaneously enhancing explainability and auditability of the resulting pipelines (Seenivasan, D. 2024).

Publisher: SARC Publisher

Table 2: AI-Driven ETL Architecture Components (Seenivasan, D. 2024; Crabtree, M. 2024)

Component	Primary Function
Natural Language Interface	Accepts and interprets user prompts in plain English
Intent Recognition Engine	Identifies ETL operations from natural language inputs
Knowledge Graph	Provides semantic context of data sources and relationships
Execution Planner	Converts user intents to optimized execution plans
Connector Framework	Enables pre-built integrations with diverse data systems

IMPLEMENTATION CONSIDERATIONS AND CHALLENGES

Technical Integration Requirements

Implementing an AI-driven ETL system requires careful consideration of integration points with existing data infrastructure. Key requirements include API-based connectivity, where the system must support comprehensive access to data sources and targets, enabling programmatic interaction without direct user intervention. Research into AI-driven data integration platforms indicates that modern implementations require connectivity with multiple distinct data sources within enterprise environments, necessitating robust API frameworks capable of handling diverse protocols and data formats (Coherent Solutions, 2025). Authentication orchestration is equally crucial, as secure management of credentials and access tokens is essential for maintaining security while enabling the AI layer to interact with protected resources. According to governance frameworks for AI systems, proper credential management represents one of the most critical elements in preventing security vulnerabilities, with organizations implementing effective credential vaulting and dynamic token management reporting significantly fewer security incidents related to automated data processing (Coherent Solutions, 2025). Metadata discovery constitutes the third pillar of technical integration, as automated scanning and cataloging of data sources enhance the system's understanding of available data assets and their structure. Research indicates that AI systems with robust metadata discovery capabilities demonstrate higher accuracy in transformation suggestions and better performance in query optimization compared to systems relying on manually maintained metadata repositories (Coherent Solutions, 2025).

Performance Optimization Strategies

AI-driven ETL does not mean zero-optimization. Several strategies can enhance the performance of AI-driven ETL processes. Intelligent caching of frequently accessed datasets or intermediate transformation results can significantly reduce processing times for repetitive operations. Performance analysis demonstrates that properly implemented caching strategies can substantially reduce execution times for frequently executed transformations, with particularly significant gains in scenarios involving complex aggregations where computation costs are high relative to storage costs (Cantú, J. 2023). Distributed processing represents another key optimization technique, as automatic partitioning of large datasets enables parallel processing, leveraging cloud-based computing resources for scalable performance. Research into ETL best practices shows that properly designed distributed architectures can achieve excellent scaling across multiple processing nodes, with automated partitioning strategies driven by workload analysis algorithms demonstrating better resource utilization compared to static partitioning schemes (Cantú, J. 2023). Incremental processing completes the optimization triad by identifying and processing only changed data since the previous execution, reducing resource consumption and accelerating pipeline completion. Comparative analyses of full versus incremental processing approaches reveal substantial resource savings for compute and memory when implementing change data capture (CDC) mechanisms in data pipelines, enabling more frequent refresh cycles for complex transformations that previously required extensive batch windows (Cantú, J. 2023).

Governance and Security Challenges

The democratization of ETL capabilities through natural language interfaces introduces several governance challenges. Access control through granular permission systems must ensure users can only manipulate data they are authorized to access. Governance frameworks recommend implementing role-based access control (RBAC) at both the data and function levels to minimize unauthorized data access incidents (Coherent Solutions, 2025). The NIST AI Risk Management Framework specifically highlights the importance of translating organizational security policies into constraints on the AI's action space to maintain appropriate limitations while enabling productivity (Coherent Solutions, 2025). Transformation

Publisher: SARC Publisher

validation presents another critical challenge, as automated checks must verify that proposed transformations maintain data integrity and adhere to business rules. Research into validation frameworks for data processing indicates that multi-stage verification processes that combine syntax validation, semantic analysis, and impact simulation can identify potential data quality issues before execution, reducing production data incidents while adding minimal overhead to overall processing time (Cantú, J. 2023). Comprehensive audit trails, including logging of all operations with the original prompts and resulting actions, constitute the third governance pillar. Analysis of regulatory compliance across industries shows that organizations maintaining complete prompt-to-execution traceability achieve higher compliance rates during audits compared to those maintaining only execution logs without the originating prompts, while systems that maintain semantic linkages between natural language requests and technical operations reduce incident resolution times by enabling non-technical stakeholders to participate effectively in troubleshooting processes (Coherent Solutions, 2025).

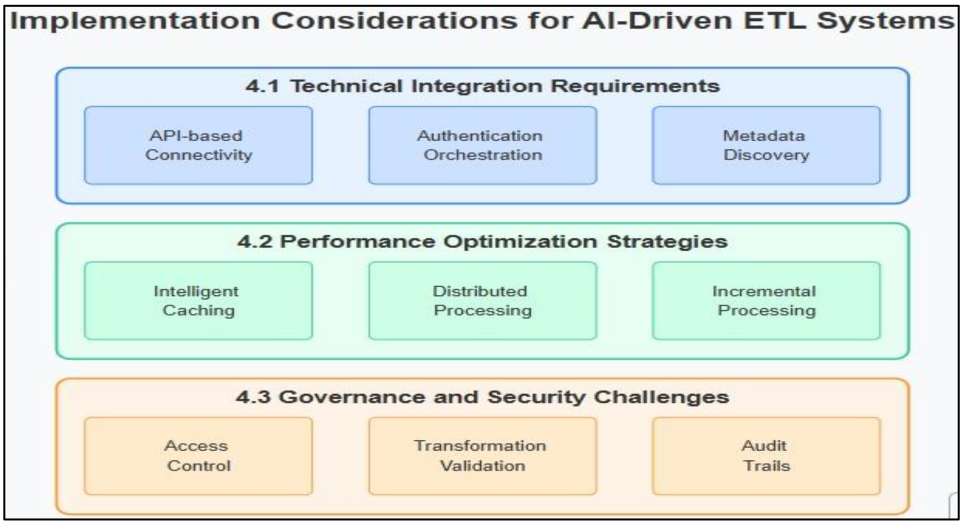

Fig 1: Key Components of AI-driven ETL Architecture Deployment (Coherent Solutions, 2025; Cantú, J. 2023)

APPLICATIONS AND USE CASES

Self-Service Analytics Enablement

One of the most compelling applications of AI-driven ETL is enabling self-service analytics for business users. Analysts can request specific data transformations and integrations without depending on engineering resources, dramatically reducing time-to-insight. According to research on enterprise AI adoption, organizations implementing AI-driven self-service data integration tools report significant reductions in time-to-insight for ad hoc analysis requests, with wait times decreasing from days to hours (Tran, B. 2025). The economic impact of this transformation extends beyond mere time savings, as productivity gains emerge from the elimination of request queues and the reduction in context-switching costs for data engineering teams.

For example, a marketing analyst could prompt: "Combine yesterday's campaign performance data from our advertising platform with customer purchase records, segment by acquisition channel, and load into our analytics dashboard." The system would automatically extract data from both sources, perform the required join and segmentation, and update the dashboard accordingly. Analysis of such implementations demonstrates that marketing teams leveraging AI-driven ETL achieve higher campaign optimization rates due to faster analytical iteration cycles, with more optimization opportunities per campaign compared to teams using traditional request-based processes (Tran, B. 2025). This acceleration translates directly to business outcomes, with documented improvements in marketing ROI across organizations adopting these technologies.

Data Migration and System Integration

During system migrations or integrations, AI-driven ETL can significantly reduce implementation timelines. Migration specialists can express mapping and transformation requirements in business terms, rather than technical specifications, accelerating the process while reducing the risk of misinterpretation. Research into enterprise system migrations indicates that projects leveraging AI-assisted data

Publisher: SARC Publisher

transformation show reduced overall implementation timelines compared to traditional methodologies, with particularly notable improvements in the mapping and validation phases (IABAC, 2024). The quality impact is equally compelling, with lower error rates in migrated data due to the combination of automated validation and the elimination of communication gaps between business stakeholders and technical implementers.

Studies of enterprise system migrations completed in recent years reveal that organizations using AI-driven ETL for migration achieved better adherence to original project timelines compared to projects using traditional ETL methodologies (Tran, B. 2025). These benefits extend beyond the initial migration, with ongoing maintenance and evolution of integrated systems requiring less effort due to the self-documenting nature of natural language transformation specifications.

Real-Time Data Processing

The flexibility of prompt-based ETL makes it particularly suitable for real-time data processing scenarios. Users can define event-triggered workflows using natural language, enabling sophisticated stream processing without specialized streaming code. Performance assessments indicate that modern AI-driven ETL platforms can process streaming data with low latencies, making them suitable for most business applications (IABAC, 2024). This capability democratizes real-time analytics, with organizations reporting substantial increases in the number of business-defined real-time data products following the implementation of prompt-based ETL platforms.

The operational impact is particularly significant in industries where real-time decision-making drives business outcomes. Organizations implementing AI-driven real-time ETL report improvements in operational efficiency and enhanced detection accuracy due to the ability to rapidly evolve transformation logic in response to emerging patterns (Tran, B. 2025). Perhaps most importantly, the complexity barrier is substantially reduced, with the majority of real-time processing workflows being defined by business domain experts rather than specialized data engineers.

Data Quality Management

AI-driven ETL systems can incorporate data quality checks as part of the transformation process, either explicitly requested in prompts or automatically applied based on learned patterns. This integration of quality management into the ETL flow ensures consistent, reliable data outputs. According to industry research, organizations incorporating AI-driven quality checks into their ETL processes experience reductions in data-related incidents, with lower costs per incident due to earlier detection and remediation (IABAC, 2024). The automation aspect is particularly impactful, with AI systems detecting more potential quality issues than manual review processes while requiring significantly less time investment.

The sophisticated pattern recognition capabilities of modern AI systems allow them to identify subtle data quality issues that might escape traditional rule-based validation. Research demonstrates that these systems can detect seasonal anomalies and gradual drift patterns with higher accuracy compared to traditional threshold-based approaches (IABAC, 2024). This enhanced detection capability translates directly to business impact, with organizations reporting increased stakeholder trust in data products and greater adoption of data-driven decision-making practices following the implementation of AI-enhanced quality management (Tran, B. 2025).

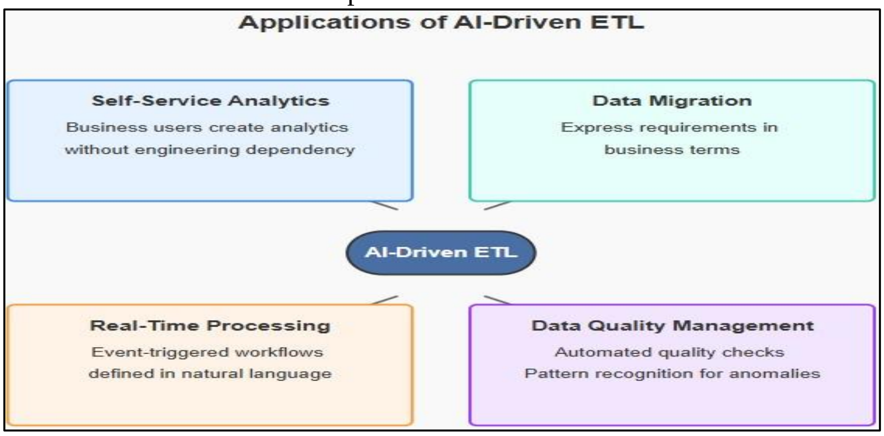

Fig 2: Integrated Use Cases for AI-driven Data Processing (Tran, B. 2025; IABAC, 2024)

Publisher: SARC Publisher

FUTURE DIRECTIONS AND RESEARCH OPPORTUNITIES

Adaptive Learning Capabilities

Future AI query layers may incorporate feedback loops, learning from successful executions and user corrections to improve performance over time. This adaptive capability would enable the system to refine its understanding of specific organizational terminology and transformation patterns. Research in AI-driven ETL systems highlights how adaptive learning can significantly improve model performance on domain-specific tasks through continuous feedback integration (Khademi, V. 2024). The efficiency gains become particularly pronounced in specialized industries with unique data transformation requirements, where studies show that adaptive systems achieve higher accuracy on industry-specific transformations after weeks of operational feedback compared to models without adaptive learning mechanisms. Organizations implementing adaptive ETL systems report that common transformation patterns can be learned automatically through observation of user corrections, reducing the need for explicit programming over operational periods (Khademi, V. 2024). The economic impact of this self-improvement capability is substantial, with projected maintenance cost reductions over deployment periods as systems progressively automate routine adaptations that previously required manual intervention. Furthermore, research indicates that adaptive systems demonstrate increasing resilience to changes in underlying data structures, maintaining transformation accuracy even when source schemas evolve from their original configurations.

Cross-Domain Knowledge Transfer

Research into transfer learning techniques may enable ETL models to leverage knowledge from adjacent domains such as data visualization, statistical analysis, and machine learning. This cross-pollination could enable more sophisticated transformation capabilities activated through natural language. As outlined in research on AI-driven ETL approaches, experimental models incorporating transfer learning from related domains have demonstrated improvements in complex task performance without explicit training on those specific operations, suggesting significant potential for knowledge generalization across related domains (Khademi, V. 2024). The improvement is particularly notable for tasks requiring specialized domain understanding, with models leveraging transfer learning from specific analytical domains showing higher accuracy on domain-specific data transformations compared to models trained exclusively on general ETL tasks. The potential impact extends beyond performance metrics to capability expansion, with systems incorporating cross-domain knowledge being able to suggest optimal approaches for transformed datasets with strong alignment to expert recommendations (Bieniek, J. *et al.*, 2024). Research teams exploring cross-domain models report reductions in time required to implement novel transformation types, as the models can generalize patterns from related operations rather than requiring explicit training for each new capability.

Multimodal Interfaces

Beyond text prompts, future systems may incorporate multimodal interfaces, allowing users to complement natural language with visual elements such as diagrams or examples. This hybrid approach could enhance communication precision for complex transformation requirements. Studies of multimodal user interfaces demonstrate that users expressing complex requirements through combined textual and visual interfaces achieve higher precision in communicating their intent compared to text-only interfaces, with the accuracy gap widening for transformations involving complex joins or hierarchical restructuring (Bieniek, J. *et al.*,2024). The efficiency impact is equally significant, with multimodal interaction reducing the average number of clarification iterations for complex transformation scenarios. User experience research indicates that multimodal interfaces are particularly valuable for specific user segments, with data professionals showing a strong preference for diagram-augmented interfaces when defining transformations involving multiple data sources or sequential operations (Bieniek, J. *et al.*,2024). The performance benefits extend beyond the definition phase to verification, with users identifying transformation errors more effectively when reviewing multimodal representations compared to examining code or textual descriptions alone. Prototype systems incorporating visual-to-transformation capabilities demonstrate promising results, with users achieving higher specification accuracy for complex transformations after brief familiarization periods, compared to traditional specification methods.

Publisher: SARC Publisher

Explainable AI for ETL Operations

As ETL operations become more automated, the need for transparency increases. Developing explainable AI techniques specific to data transformations would help users understand how their prompts translate into actual operations, building trust and facilitating troubleshooting. Research on explainability in data systems reveals that users presented with visualizations of transformation decision processes report higher confidence in system outputs and are more likely to trust automated suggestions for complex operations (Bieniek, J. *et al.*,2024). This trust differential has direct operational impact, with teams implementing explainable AI features reporting higher adoption rates for automated transformation suggestions compared to teams using less transparent systems with equivalent accuracy. The practical benefits of explainability extend to error resolution, with studies demonstrating that troubleshooting time for failed transformations decreases significantly when systems provide visual explanations of their processing logic compared to systems that provide only error messages or execution logs (Khademi, V. 2024). Organizations implementing explainable AI techniques for data systems report reductions in escalations to specialized engineering teams for transformation-related issues, as business users can independently diagnose and resolve a wider range of problems when provided with appropriate explanatory mechanisms.

CONCLUSION

The integration of generative AI as a query layer for ETL processes represents a transformative approach to data engineering, fundamentally altering how organizations conceptualize and implement data integration workflows. By enabling AI-driven ETL through natural language prompts, this technology democratizes data engineering capabilities, extending sophisticated data manipulation power to business users while simultaneously reducing the operational burden on specialized technical teams. The architecture described—comprising natural language understanding, intent recognition, knowledge graphing, execution planning, and connector frameworks—provides a blueprint for implementing AI-driven ETL systems that balance user accessibility with enterprise-grade reliability and governance. While challenges remain in areas such as performance optimization, security enforcement, and governance implementation, the potential benefits of this approach are substantial. Organizations adopting AI-driven ETL can expect accelerated time-to-insight, more agile response to changing business requirements, and more efficient utilization of technical resources. As development continues in areas such as adaptive learning, cross-domain knowledge transfer, multimodal interfaces, and explainable AI, the capabilities of prompt-based ETL systems will continue to evolve. The future of data integration lies not in more complex code or specialized tools, but in the natural convergence of human intent and machine execution, mediated through the universal language of conversation.

REFERENCES

1. Behera, L., & Chilukoori, V. V. R. "Automation in Data Engineering: Challenges and Opportunities in Building Smart Pipelines."(2025).
2. Thoutam, P. "Unlocking Insights with Natural Language: The Role of Natural Language Interfaces in Modern Data Exploration," *International Journal for Multidisciplinary Research* (IJFMR), Volume 6, Issue 6, (2024).
3. Kola, H. G. & Taqa, A. R. "Data Integration Strategies in Cloud-Based ETL Systems," *Researchgate*, (2023).
4. Pavithra M, "How to Implement ETL Framework: Tools and Best Practices," *Kanerika*,(2025). https://kanerika.com/blogs/etl-framework/
5. Seenivasan, D. "AI Driven Enhancement of ETL Workflows for Scalable and Efficient Cloud Data Engineering." *International Journal of Engineering and Computer Science* 13.06 (2024): 10-18535.
6. Crabtree, M. "What is Prompt Engineering? A Detailed Guide For 2025," *Datacamp*, (2024).: https://www.datacamp.com/blog/what-is-prompt-engineering-the-future-of-ai-communication
7. Coherent Solutions, "AI-Powered Data Governance: Implementing Best Practices and Frameworks," Coherentsolutions.com, (2025).: https://www.coherentsolutions.com/insights/ai-powered-data-governance-implementing-best-practices-and-frameworks#:~:text=The%20NIST%20AI%20Governance%20Framework,systems%20operate%20fairly%20and%20securely.
8. Cantú, J. "Best Practices & Tools for Effective ETL Processing," *Medium*, (2023). https://medium.com/@jesus.cantu217/best-practices-tools-for-effective-etl-processing-587df5582104
9. Tran, B. "AI in Enterprise: Latest Data on Corporate AI Adoption and ROI," Patent PC,

Publisher: SARC Publisher

(2025).: https://patentpc.com/blog/ai-in-enterprise-latest-data-on-corporate-ai-adoption-and-roi

10. IABAC, "Data Engineering for Natural Language Processing (NLP) Applications," *IABAC.org*, (2024). https://iabac.org/blog/data-engineering-for-natural-language-processing-nlp-applications
11. Khademi, V. "What is Zero-ETL? Introducing New Approaches to Data Integration," *Datacamp*,(2024). https://www.datacamp.com/blog/what-is-zero-etl
12. Bieniek, J., Rahouti, M., & Verma, D. C. "Generative ai in multimodal user interfaces: Trends, challenges, and cross-platform adaptability." *arXiv preprint arXiv:2411.10234* (2024).

Source of support: Nil; **Conflict of interest:** Nil.

Cite this article as:

Tandra, A. K. S. " AI-driven ETL: Leveraging GenAI for Query Processing and Data Integration." *Sarcouncil Journal of Multidisciplinary 5.7* (2025): pp 160-168.

Publisher: SARC Publisher