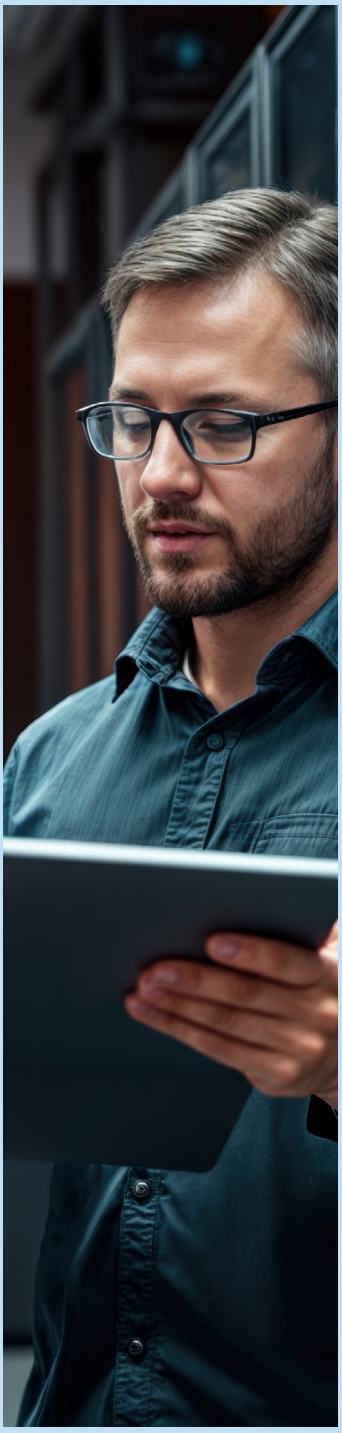
VIEW POINT



BRIDGING DATA GAPS: HOW COMPANIES CAN HARNESS AGENTIC AI TO STREAMLINE EIM OPS

Abstract

In today's digital landscape, organizations face challenges in managing complex and inconsistent data, especially in oil and gas Engineering Information Management (EIM), where issues like data gaps and varied formats complicate overall migration procedures. Agentic Artificial Intelligence (AI) is transforming this process by using autonomous intelligence to detect patterns, clean and enrich data, and automate tasks like discovery, mapping, transformation, and validation. This results in more accurate, efficient, and intelligent data migration with minimal human input and intervention.

This article delves into the capabilities of Agentic AI (Azure-based) and its potential to overcome the traditional hurdles associated with EIM data migration, offering a pathway towards a more intelligent and effective management of critical engineering assets.

Infosys®
Navigate your next

Executive Summary

Agentic AI is revolutionizing data migration in the Oil and Gas sector by addressing critical challenges in Engineering Information Management (EIM). By leveraging autonomous intelligence, it enhances data quality through intelligent cleansing, validation, and schema mapping. This AI-powered approach cuts down manual intervention, accelerates migration timelines, and ensures compliance. Seamless integration with cloud-based data stacks empowers organizations to improve operational efficiency and strategic decision-making, ultimately transforming data infrastructure. This article gives the lowdown on these challenges and the agents involved in tackling them.

Current Challenges in EIM Migration Space

Present-day industry reports reveal that organizations across the Oil and Gas sector are confronted with an unprecedented surge in data quality and complexity. This exponential growth often leads to a proliferation of outdated, duplicated, and irrelevant information, creating significant bottlenecks, and hindering efficient data processing. The challenge is particularly acute within the realm of EIM, where the intricate nature of technical data, diverse formats, and disparate systems compound these issues.

Data migration within EIM is frequently fraught with difficulties, including the need to address critical data gaps, enrich incomplete datasets to ensure usability, and consolidate information scattered across various incompatible formats into a cohesive and high-quality repository. The inability to effectively manage and migrate this vital engineering information can severely impede operational efficiency and project timelines, ultimately impacting organizational competitiveness.

Engineering industries are confronting the ever-expanding complexity of information management and migration procedures. Here's a detailed **breakdown of the challenges** in EIM migration in the domains of **People, Process, Time, and Technology.**

1. Data Potholes Outdated, Duplicated, or Irrelevant Data

People

Lack of accountability or ownership for legacy data; insufficient training to identify valuable vs. obsolete data

Process

Absence of a formal data-cleansing workflow; inconsistent or undocumented data archival/deletion policies

Time

Manual identification and cleanup of data garbage is time-consuming and slows down overall migration effort

Technology

Inadequate tools for automated data deduplication, classification, or relevance filtering

2. System Inefficiencies Broken Workflows, Inconsistent Mapping

People

Misalignment between business units and IT; lack of clarity in workflow ownership and governance roles

Process

Legacy workflows are often undocumented or misaligned with modern systems, old databases, CAD systems, or P&ID systems that don't follow modern data formats

Time

Debugging broken workflows and remapping data structures can cause significant project delays

Technology

Lack of agile, configurable platforms for adapting workflows and automating governance validation

3. Manual Efforts Repeated and Error-Prone Tasks, Resource Shortage

People

Error-prone work and burnout from repetitive manual tasks; shortage of data engineers

Process

Manual-heavy procedures in cleaning, validation, and transformation; limited adoption of best practices

Time

Manual work extends timelines drastically, increasing cost and risk of project overruns

Technology

Poor automation and tooling — legacy platforms don't support modern RPA, AI, or transformation scripting

4. Review & Approvals Ambiguity, Redundancy, Back-and-Forth Flows, Documentation, and Reporting

People

Ambiguity around decision-making authority; overloaded governance teams

Process

No clear, repeatable approval workflows; delays caused by redundant approvals

Time

Frequent back-and-forth communications and manual signoffs add to bottlenecks. Huge turnaround time involved in formulating reporting mechanism and documentation

Technology

Lack of integrated workflow management tools with approval tracking and audit trails; preparing intelligent bespoke reports aligning with business needs

5. Global Compliance / QA Privacy, Security, Quality, and Standards

People

Inadequate knowledge of regional compliance standards among global teams

Process

Disparate compliance processes for each region with no unified policy enforcement mechanism. Auditability, version control, and regulatory compliance (e.g., ISO 15926)

Time

Time-intensive audits and validation processes to meet different regional compliance requirements

Technology

Lack of tools for automated compliance checks, data masking, and policy enforcement across regions

6. Data Profiling Anomalies, Oversights, Patterns, and Unstructured Data

People

Inexperienced personnel lead to inconsistent profiling and oversight of key data anomalies. They lack skills to handle unstructured data formats

Process

Non-standardized data profiling practices across departments, engineering assets, drawings, specs, embedded tables, tags, documents, and scanned content

Time

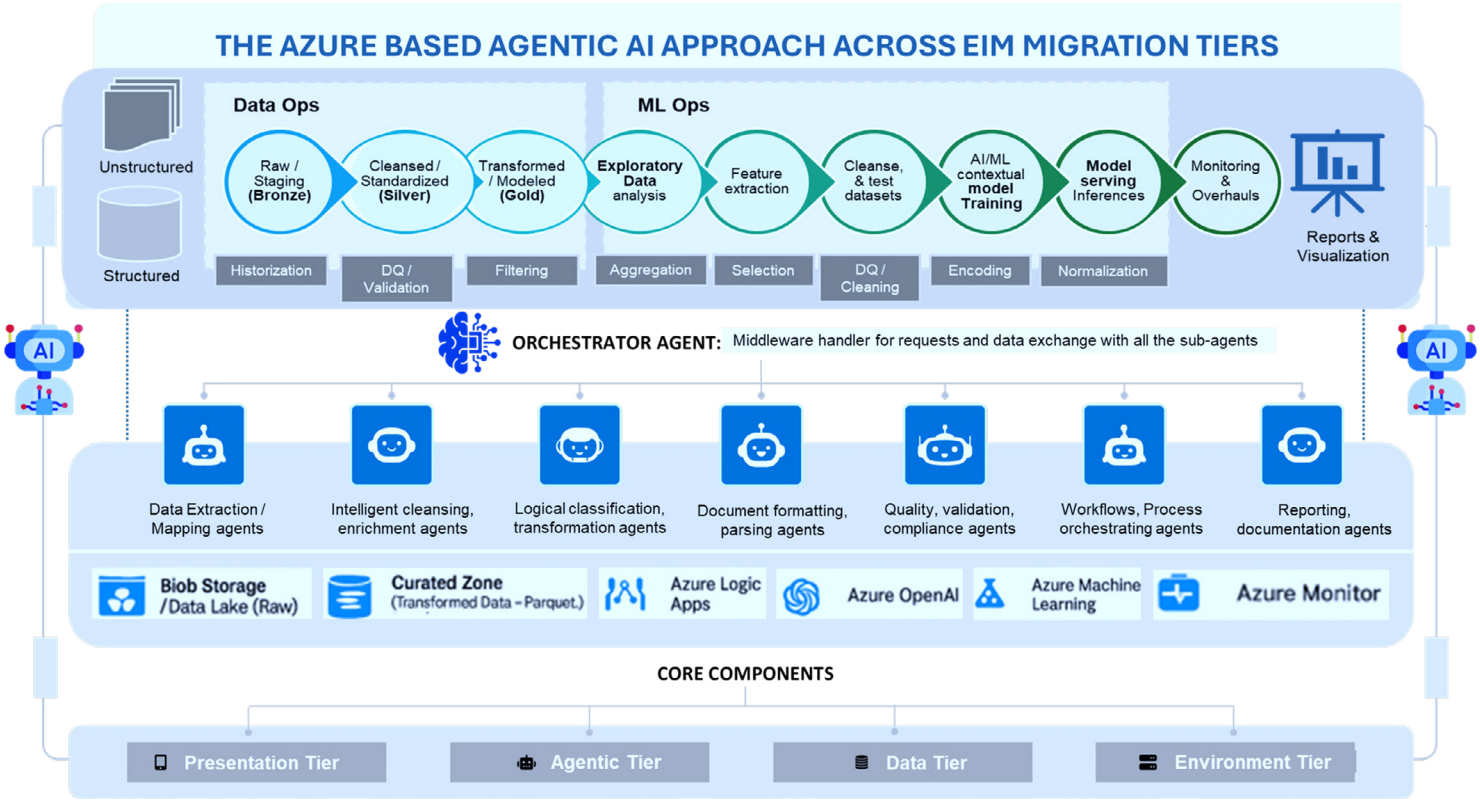
Manual profiling, feedback loops, and rework; harmonizing diverse data formats increase migration duration

Technology

Lack of robust profiling automations to detect patterns, anomalies, and quality metrics. Insufficient ETL/ELT tools to preserve multi-format data format and lineage

How Agentic AI Can Overcome These Hurdles

Agentic AI refers to AI systems capable of planning, decision-making, and operating autonomously within a goal-oriented framework. Here's an **Azure-based logical architecture** for ETL data migration in the EIM space using Agentic AI. This setup holds good for oil and gas use cases pertaining to EIM migration, plant design data transformations, tag metadata extraction from documents, and building a centralized engineering data lake.



Agents Upping the EIM Migration Game

Agent: Data Extraction, Mapping

Challenges addressed: 1,3,6

- **AI vision and NLP models** extract structured information from PDFs, P&IDs, scans, and diagrams, and link extracted data intelligently with relational systems
- **LLM agents with GPT models + open AI service codex, data catalogs, and AI scanners are used** to auto-discover schema from source and target systems. Agent learns transformation rules using prompts or from previous mappings (zero/few-shot learning).
- Uses **ontology-based AI** (like **ISO 15926 or CFIHOS standards**) to align datasets, suggest, and auto-map fields using semantic similarity, metadata analysis, and ontology matching.

Agent: Logical Classification, Transformations

Challenges addressed: 4,6

- Learns from existing mappings to recommend or automate data transformation rules — unit conversions, renaming fields, and applying logic.

  Uses **prompted AI, GPT + dbt plugins** to apply domain-specific logic dynamically by understanding business logic.
- Translates and adapts rules to the new platform, simulates transformations to verify accuracy.

Agent: Quality, Validation, and Compliance

Challenges addressed: 3,5

- Applies business/engineering rules (e.g., unit checks, tag formatting) with the aidance of synapse analytics and rule-based engines, and detects anomalies, missing data, duplication, or incorrect formats.
- Employs rule-based + ML-based models to continuously improve data validation.
- Uses reinforcement learning (RLHF) or active learning loops where AI seeks feedback for unclear mappings, improving over time with shorter loops and quicker turnarounds.
- Identifies and tags PII/PHI, suggests or applies masking/encryption strategies, and tracks data access control policies.

Agent: Intelligent Cleansing, Enrichment

Challenges addressed: 1,2

- Detects duplicates, incomplete entries, or outdated records; **uses ML or rule-based methods** to perform relationship building and taxonomies.
- Uses **NLP models** to interpret unstructured documents and extract structured metadata.
- Auto-validates asset metadata against known engineering standards.

Agent: Document Formatting, Parsing

Challenges addressed: 2,6

- Uses **Azure form recognizer, parser agents, and document intelligence models** to autonomously extract tags, assets and specs from PDFs, CAD files, scanned images, equipment datasheets, engineering documents, P&IDs, and BOMs by **using OCR + LLMs.**
- Uses **cognitive search** for deploying pre-indexed semantic search across engineering documents and content-based searches.

Agent: Workflows, Process Orchestrating

Challenges addressed: 1,2,6

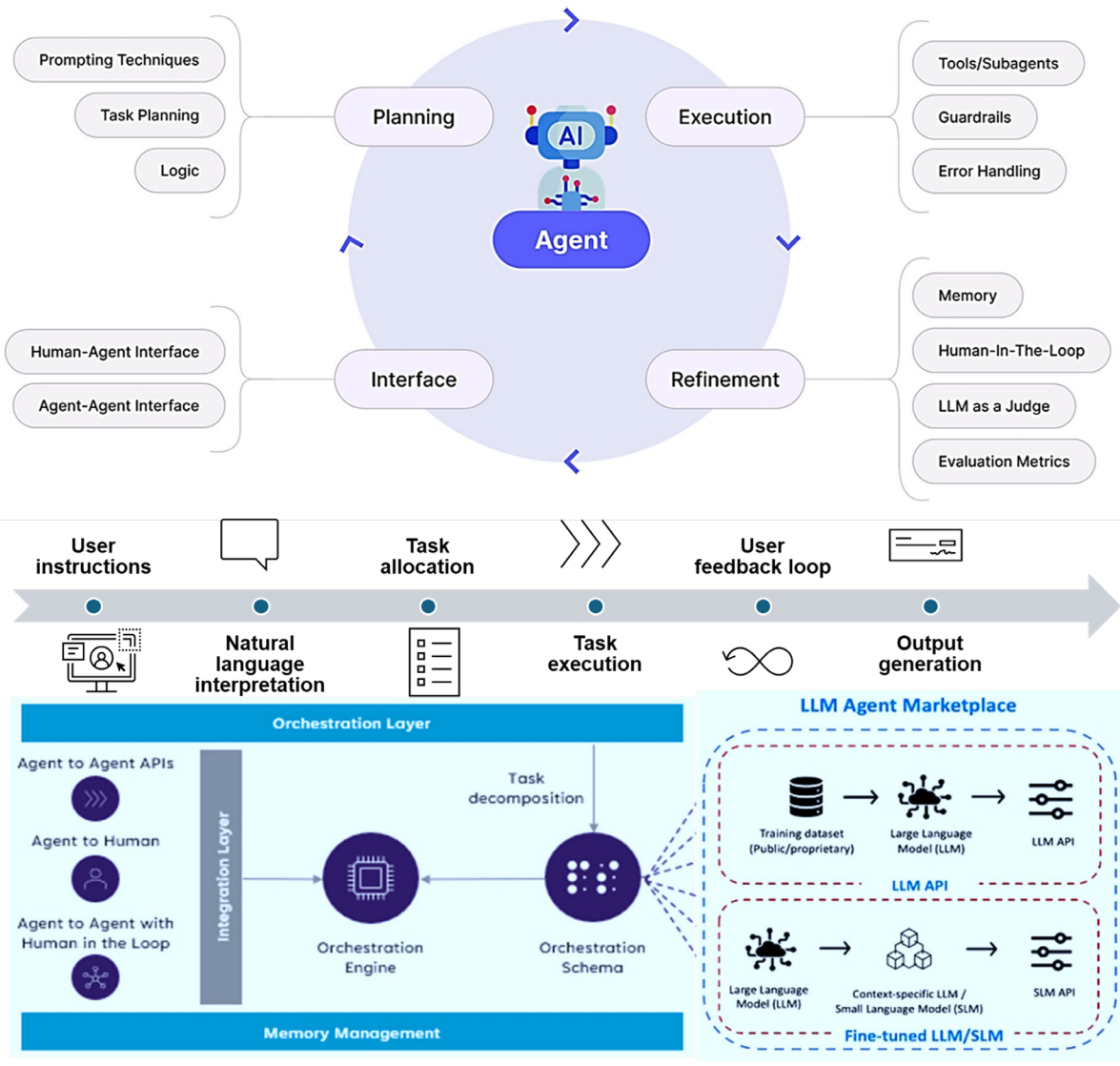
- **Orchestrator AI** can coordinate and adapt workflows (ETL pipelines) based on real-time errors, delays, or downstream feedback.
- **Multi-agent systems** orchestrate ETL workflows across different data sources and streams; adjust extraction logic without manual intervention if source system deviates.
- Coordinates multi-step flows, handles retries, notifies SMEs for human-in-the-loop with **data factory components, LangChain, semantic kernels, ML based databricks + unity catalog.**
- Combines with **RPA bots** for integration with legacy UIs or systems that lack APIs.

Agent: Reporting, User Support, and Documentation

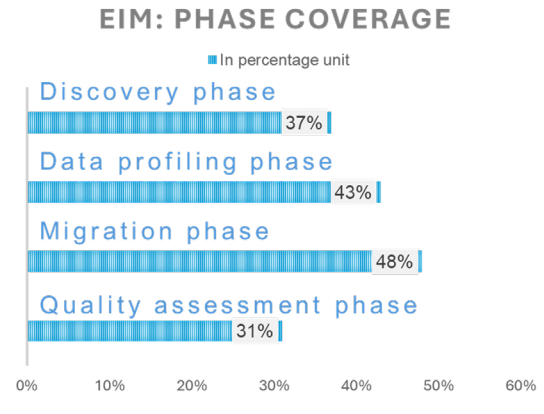
Challenges addressed: 4,5

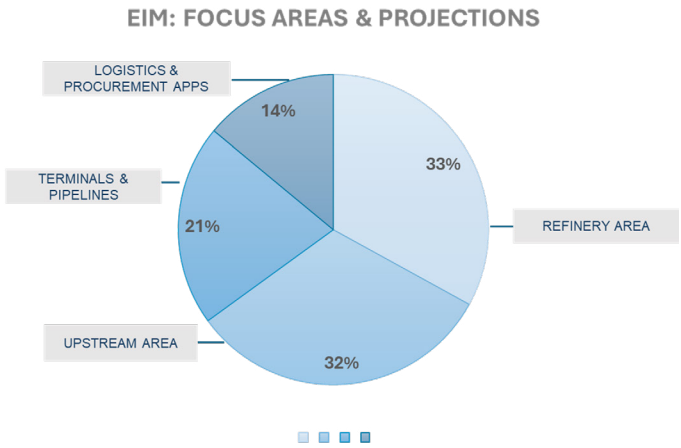
- Automatically generates migration documentation, creates dashboards showing migration progress and issues, produces audit trails and transaction logs, compares timeframe-based snapshots, keeps stakeholders informed, and provides user support with **semantic Q/A.**
- **Co-piloted Power BI, Looker AI, and Reporter AI for** intelligent documentation and producing insightful reports.

Orchestration of Agentic AI, Its Roles, and Core Capabilities



Benefits of Agentic AI in EIM Space – Roles, Areas, and Process





33% Standardization of legacy engineering apps, streamlining EIM migrations

35% Aids in building AI enabled automations across extract, transform and load phases

2x Accelerated deployment cycles, significant reduction in time-to-market and migration roll-outs

Other Benefits Include

- Autonomous schema mapping & transformation
- Context-aware migration using AI understanding of engineering terms
- Improved handling of unstructured and semi-structured data
- Enhanced data quality and consistency
- Reusable agents for future projects, plots, and ongoing data-cleansing activities
- Accelerated migration timelines
- Human-in-the-loop collaboration
- Low-code/no-code orchestration with Azure Logic Apps
- Reduces manual data entry, errors, and SME workload
- Improved scalability and reusability, better compliance & traceability
- Semantic understanding of engineering context
- Reduced downtime and cost optimization
- Seamless integration with modern data stacks

Conclusion

Agentic AI is redefining the data migration landscape for the Oil and Gas industry by combining intelligent automation with domain-specific contextual understanding. Traditional ETL migrations in this sector are often slow, resource-intensive, and highly dependent on subject matter experts due to complex data structures, engineering-specific terminologies, and a mix of structured, semi-structured, and unstructured data sources across legacy systems. Agentic AI addresses these challenges by enabling autonomous schema mapping and transformation, guided by a semantic understanding of oil and gas concepts, such as well logs, production reports, reservoir models, and engineering workflows.

This AI-driven context awareness significantly reduces errors and manual intervention during migration, while reusable agents and low-code/no-code orchestration via platforms like Azure Logic Apps accelerate timelines and reduce SME workload. Human-in-the-loop collaboration ensures expert oversight where needed, balancing automation with accountability. Additionally, enhanced data quality and consistency mechanisms — including intelligent cleansing, validation, and lineage tracking — improve compliance, traceability, and long-term data governance. These AI agents can be continuously reused for future data projects, ongoing cleansing, and system integrations, ensuring a scalable and cost-optimized approach.

Seamless integration with modern cloud-based data stacks further enhances the organization's agility, allowing oil and gas enterprises to unlock faster insights, drive operational efficiency, and support strategic decision-making across upstream, midstream, and downstream operations. Overall, Agentic AI offers a transformative leap forward, enabling digital modernization of data infrastructure with reduced downtime, improved reliability, and significantly lower total cost of ownership.

About the Authors

Pranav Gupta

Senior Principal, Infosys Consulting
London, UK

Baskar Ramamurthy

Engineering Manager,
Infosys Limited London, UK

Ramkumar Gunasekaran

Principal, Infosys Consulting
London, UK

Raghavendar Kannan

Engineering Lead, Infosys Limited
London, UK

For more information, contact askus@infosys.com

Infosys®
Navigate your next