



iJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** X **Month of publication:** October 2026

DOI: <https://doi.org/10.22214/ijraset.2026.85029>

www.ijraset.com

Call: ☎ 08813907089

E-mail ID: ijraset@gmail.com

Agentic AI Systems: A Review of Multi-Agent Reasoning, Trust, Orchestration, and Autonomous Data Science

Subina S B¹, Abhijith B², Pranav Edwin³, Sania Jackson⁴, Geethu Raj⁵

^{1, 2, 3, 4}Dept. of Computer Science (Data Science), St. Thomas Institute for Science & Technology, Trivandrum, India

⁵Asst. Professor, Dept. of Computer Science (Data Science), St. Thomas Institute for Science & Technology, Trivandrum, India

Abstract: *Agentic artificial intelligence (AI) — systems of autonomous, goal-directed software agents that plan, act, and coordinate with reduced human oversight — is moving from research prototypes toward applied data-science, perception, and decision-support pipelines.*

Yet the literature remains fragmented across largely disconnected strands: trust-aware orchestration of multimodal agents for visual classification, human adoption of autonomous AI agents, automated machine-learning (AutoML) pipelines for tabular data, and fully autonomous multi-agent systems for end-to-end data engineering.

This review synthesizes four recent studies spanning these strands, together with the broader literature on multi-agent orchestration, retrieval-augmented generation (RAG), large-language-model (LLM)-based knowledge-graph construction, and explainable AI, to build a unified picture of agentic AI as applied to intelligent data science. We propose a four-part taxonomy — perception-level trust orchestration, human adoption of agentic systems, automated feature/model pipelines, and autonomous data-engineering agents — and compare the reviewed studies on automation level, reasoning mechanism, knowledge representation, retrieval use, trust treatment, and evaluation methodology. The synthesis shows that individual components of an agentic data-science stack (confidence calibration, retrieval-grounded re-evaluation, automated feature synthesis, autonomous pipeline repair) have each demonstrated measurable gains in isolation, but no reviewed system couples them with a persistent, multimodal knowledge-graph layer, and adoption research indicates that user trust is not the dominant driver of usage intention — a finding with direct design implications. We identify recurring gaps in provenance, explainability, benchmark standardization, and graph-grounded reasoning, and outline an integrated conceptual framework for the reviewed research landscape, positioning closer integration of agentic data science with knowledge-graph representation as an open research trajectory rather than a solved problem.

Index Terms: *Agentic AI, multi-agent orchestration, trust calibration, retrieval-augmented generation, knowledge graphs, AutoML, autonomous data engineering, technology adoption.*

I. INTRODUCTION

A. Research Context

The rapid diffusion of large language models (LLMs) since 2022 has driven a shift from single-purpose "AI agents" — modular, tool-augmented systems executing narrow, well-defined tasks — toward "agentic AI": ecosystems of specialized agents that decompose goals, coordinate through an orchestration layer, and act with reduced human supervision [5], [6]. This paradigm is now being explored across domains as different as agricultural diagnostics, consumer technology adoption, automated machine learning, and enterprise data engineering, yet these applications are typically studied by separate research communities using different vocabularies, evaluation practices, and notions of what it means for an agentic system to be "trustworthy" or "autonomous."

B. Problem Statement

Because agentic AI is studied along largely disconnected strands — perception-level trust orchestration, human adoption and acceptance, automated feature/model pipelines, and autonomous data-engineering agents — there is currently no unified account of what these strands collectively demonstrate, where they overlap, and where they leave open gaps.

This review addresses that fragmentation by synthesizing four recent, representative studies together with the broader supporting literature.

C. Motivation for the Review

The four studies synthesized here span (i) trust-aware multi-agent orchestration for visual classification with retrieval-augmented reasoning [1], (ii) empirical human-adoption drivers of autonomous AI agents among Generation Z users [2], (iii) an interpretable AutoML pipeline that automates feature engineering, text categorization, and hyperparameter tuning [3], and (iv) a fully autonomous, LLM-and-reinforcement-learning-driven system for end-to-end data engineering [4]. Individually, each paper makes a focused technical or empirical contribution; read together, they outline complementary layers of a single underlying problem — how to make autonomous, multi-agent systems that reason over data reliably, are trusted and adopted by users, and can eventually reason over structured, multimodal knowledge rather than only raw tabular or image inputs. This complementarity, and the absence of any existing synthesis connecting these strands, motivates the present review.

D. Research Questions

This review is guided by four questions: (RQ1) What technical mechanisms have been proposed to make multi-agent AI systems trustworthy, and how is "trust" operationalized differently across technical and human-adoption research? (RQ2) To what extent do existing automated data-science and data-engineering pipelines already exhibit agentic characteristics, as opposed to sequential automation? (RQ3) Does any reviewed approach integrate agentic reasoning with structured, multimodal knowledge representation (e.g., knowledge graphs)? (RQ4) What recurring gaps and unresolved challenges emerge when the reviewed literature is considered as a whole, rather than as isolated contributions?

E. Contributions of the Review

The contribution of this review is threefold: (1) it critically analyzes and compares four recent seed papers along a common set of technical and empirical dimensions; (2) it situates them within the broader literature on agentic AI, retrieval-augmented generation, knowledge-graph construction, and explainable AI, identifying where the reviewed work is representative of wider trends and where it is an outlier; and (3) it derives, from the reviewed literature itself, research gaps that indicate promising directions for integrating agentic data-science automation with multimodal knowledge-graph representation. The review does not present new experiments; all quantitative results discussed below are drawn from the cited sources.

F. Organization of the Paper

The remainder of this paper is organized as follows. Section II describes the review methodology. Section III introduces foundational concepts. Section IV proposes a taxonomy of existing approaches. Section V reviews and synthesizes the literature. Section VI presents a comparative analysis across three tables. Section VII discusses research gaps and open challenges. Section VIII proposes an integrated conceptual framework for the reviewed literature. Section IX outlines future research directions. Section X discusses limitations of this review, and Section XI concludes.

II. REVIEW METHODOLOGY

This review follows a structured narrative/scoping-review methodology rather than a full systematic review, since its starting point is a fixed set of four seed papers chosen a priori by the authors on the basis of thematic relevance, rather than a single systematic database query. This design is appropriate for an early-stage conceptual synthesis but should be extended toward a PRISMA-style protocol before any claim of exhaustiveness is made in a submission-ready manuscript.

A. Literature Sources

The four seed papers [1]–[4] were supplemented with literature retrieved from IEEE Xplore, arXiv, ACM Digital Library, and Information Fusion / journal-indexed sources.

B. Search Strategy

Supplementary literature was retrieved using combinations of the keywords "agentic AI," "multi-agent orchestration," "trust calibration LLM," "AutoML," "autonomous data engineering," "knowledge graph construction LLM," and "GraphRAG." The search period was not restricted at the lower bound (to allow inclusion of foundational work such as Deep Feature Synthesis [13] and the Transformer architecture [18]) but emphasized 2023–2026 publications to capture the post-ChatGPT surge in agentic-AI research.

C. Inclusion and Exclusion Criteria

Sources were included if they (a) presented an empirical, architectural, or conceptual contribution directly relevant to agentic AI, multi-agent orchestration, trust/adoption of autonomous AI systems, AutoML, autonomous data engineering, or LLM-based knowledge-graph/RAG methods, and (b) were published in a peer-reviewed venue or, for very recent work, as an arXiv preprint from an identifiable research group with a clear methodology. Sources were excluded if they were non-technical commentary, duplicate reporting of the same study, or did not provide verifiable bibliographic information.

D. Study Selection

The four seed papers were selected outside the scope of this review and were therefore not subject to database screening. Supplementary references were screened by title and abstract for topical relevance and by full-text read where available. Because a database query log was not systematically retained for this draft, the exact counts of papers identified and screened are reported as placeholders identified and included to be completed once a formal search log is produced; these values should not be fabricated in the final manuscript.

E. Data Extraction

For each of the four seed papers, the following attributes were extracted: research problem, objective, methodology/architecture, technologies used, dataset or experimental environment, main findings, contributions, limitations, and stated future work. These attributes form the basis of Table I.

F. Synthesis Method

The extracted attributes are further compared along technical dimensions (automation level, reasoning mechanism, knowledge representation, retrieval mechanism, trust mechanism, evaluation type) in Table II, and along the coverage of capabilities relevant to an integrated agentic data-science and knowledge-graph research direction in Table III.

III. BACKGROUND AND FOUNDATIONAL CONCEPTS

A. AI Agents versus Agentic AI

Sapkota et al. [5] distinguish AI Agents — modular, LLM- or vision-model-driven systems built for narrow, well-defined tasks (e.g., summarization, scheduling) — from Agentic AI, a paradigm characterized by multi-agent collaboration, dynamic task decomposition, persistent memory, and orchestrated autonomy. This distinction underlies the structure of the present review: [1] and [4] are instances of Agentic AI in this stricter sense, whereas the automated pipeline in [3] is closer to a single, tool-augmented AI Agent whose stages are chained rather than negotiated among autonomous peers, and [2] studies human responses to agentic systems generically, without proposing a new architecture.

B. Orchestration and Trust Calibration

A recurring architectural pattern is the separation of perception or execution agents from a non-perceptive orchestrator that performs meta-reasoning over agent outputs [1]. Trust in such systems is increasingly operationalized through calibration metrics — Expected Calibration Error (ECE), Overconfidence Ratio (OCR), and consistency measures — that quantify whether an agent's self-reported confidence tracks its empirical correctness [1]. This differs from human-centered notions of trust studied in technology-adoption research, where trust is a psychological construct measured through survey instruments [2].

C. Retrieval-Augmented Generation and Graph-Based Retrieval

Retrieval-augmented generation (RAG) grounds model outputs in externally retrieved evidence rather than parametric memory alone. Image-RAG, as used in [1], retrieves visually similar labeled exemplars via CLIP embeddings to support re-evaluation. A parallel and increasingly prominent line of work, GraphRAG, retrieves evidence from a structured knowledge graph rather than a flat vector index, improving multi-hop reasoning and traceability [8]–[11]. None of the four seed papers implements graph-based retrieval; [1] uses flat vector similarity search only, which is a relevant distinction developed further in Section VII.

D. Knowledge-Graph Construction with LLMs

LLMs are increasingly used to automate knowledge-graph (KG) construction — entity and relation extraction, ontology grounding, and schema alignment — reducing the traditional dependence on manual ontology engineering [7].

This capability remains largely unaddressed by the four seed papers reviewed here, which instead operate on flat tabular, textual, or image data without an explicit graph schema, despite growing interest in such integration across the broader literature.

E. Automated Machine Learning and Feature Engineering

AutoML systems automate feature engineering, model selection, and hyperparameter optimization to reduce reliance on manual data-science expertise [3], [12]. Deep Feature Synthesis (DFS) automatically constructs candidate features from relational data structures [3], [13], and post hoc explanation methods such as LIME [3], [14] provide local, model-agnostic interpretability for the resulting models.

F. Human Adoption of Agentic Systems

Complementing the technical literature, adoption research applies constructs such as perceived usefulness, perceived risk, perceived humanness, and trust to explain behavioral intention to use autonomous AI agents [2]. This body of work highlights that technical trustworthiness (Section III-B) and perceived user trust (this subsection) are related but empirically distinct constructs that a complete agentic-AI research agenda must treat separately.

IV. TAXONOMY OF EXISTING APPROACHES

Based on the content of the four seed papers and their surrounding literature, this review organizes the reviewed agentic-AI research into four categories.

- 1) Trust-Aware Multi-Agent Orchestration for Perception Tasks: systems that coordinate multiple perception agents through a meta-reasoning orchestrator and calibrate trust using quantitative metrics, exemplified by [1].
- 2) Human Adoption and Acceptance of Agentic AI: empirical, survey-based research explaining why and under what conditions users adopt autonomous agents, exemplified by [2].
- 3) Agentic/Automated Pipelines for Data Science (AutoML): systems that automate the classical data-science pipeline — preprocessing, feature engineering, model selection, tuning, and interpretation — exemplified by [3].
- 4) Autonomous Multi-Agent Systems for Data Engineering: systems that extend automation upstream of modeling, to ingestion, schema management, transformation, validation, and governed deployment, exemplified by [4].

These categories are not mutually exclusive — for instance, [1] and [4] both use an orchestrator/controller pattern with specialized sub-agents — but each paper's primary contribution maps most naturally onto one category, which structures the analysis in Section V and the comparison in Section VI.

V. REVIEW AND SYNTHESIS OF EXISTING LITERATURE

A. Trust-Aware Orchestration for Visual Classification [1]

Roumeliotis et al. [1] address the problem of trusting zero-shot vision-language agents (GPT-4o, Qwen-2.5-VL) for apple-leaf-disease classification, where no task-specific fine-tuning is assumed. Their architecture separates perception (the two vision-language agents) from meta-reasoning (a non-visual orchestrator, o3-mini), and introduces a trust evaluation layer using ECE, Overconfidence Ratio, and a Consistency Gap metric to decide when to trigger a CLIP-based Image-RAG re-evaluation loop. Across three experimental configurations — zero-shot, fine-tuned, and trust-aware with RAG — the orchestrator's accuracy rose from 48.13% (confidence-aware zero-shot) to 85.63% (trust-aware with RAG), approaching the 97.50% accuracy obtained only after supervised fine-tuning of both agents.

The study also reports that GPT-4o was better calibrated than Qwen-2.5-VL, and that when both agents confidently agreed on an incorrect label, the orchestrator had no dissenting signal to intervene on — a consensus-failure mode the authors flag as an open problem. Because the reference images used for retrieval were the same fine-tuning set, the RAG mechanism is a nearest-neighbor vector index over labeled exemplars rather than a structured knowledge graph, and the orchestrator itself lacks direct visual perception, relying entirely on agent-reported text.

B. Human Adoption of Agentic AI [2]

Hasselwander and Lah [2] shift the unit of analysis from system architecture to system acceptance, surveying 243 German Generation Z respondents and modeling intention to use and actual use of AI agents via PLS-SEM combined with Necessary Condition Analysis (NCA).

Perceived usefulness emerged as the strongest predictor of intention ($\beta = 0.560$), and perceived risk had both a direct negative effect on intention ($\beta = -0.194$) and an indirect effect via trust; critically, trust itself did not significantly predict intention to use ($\beta = 0.030$, $p > .05$), even though respondents expressed explicit concerns about data privacy and model reliability in open-ended responses. NCA further showed that perceived usefulness and perceived humanness are necessary — though not individually sufficient — conditions for higher levels of adoption intention, with bottleneck values rising as target intention levels increase. The authors interpret this as evidence that Gen Z users evaluate agentic AI pragmatically, prioritizing utility over resolved trust concerns, and argue that the resulting responsibility gap should be closed through regulation and transparent design rather than left to individual users.

C. Interpretable and Efficient AutoML Pipeline [3]

Hasan and Tausif [3] propose IDEAL, an AutoML pipeline that chains automated data-type handling, missing-value imputation, categorical encoding, LLM-based conversion of free text into compact categorical features, Deep Feature Synthesis for relational feature generation, LIME-based feature selection and interpretability, and Bayesian-optimization-based hyperparameter tuning (via BayesSearchCV).

Evaluated on ten public datasets spanning healthcare, education, telecommunications, and finance, IDEAL consistently outperformed Auto-Sklearn 2.0, with accuracy gains as large as 13 percentage points on the Telco-Customer-Churn dataset (98% versus 85%) and comparable or larger gains in precision, recall, and F1 across most datasets.

The pipeline is agentic in a limited sense: it automates a sequential, largely deterministic workflow rather than coordinating negotiating, goal-decomposing agents, and it operates purely on structured/tabular and short-text inputs without any graph-structured representation of entities or relationships.

D. Autonomous AI System for End-to-End Data Engineering [4]

Chirumamilla [4] proposes AIDE-End, a five-component architecture (ingestion, profiling/quality, transformation, self-validation, and deployment/governance agents) coordinated by an LLM multi-agent controller and a reinforcement-learning engine, backed by a metadata repository that tracks schema versions, transformation templates, lineage, and quality metrics. Evaluated conceptually against manually engineered Apache Spark pipelines and traditional ETL tools on financial-transaction, healthcare-claims, and e-commerce-catalog datasets, the system reports a 62% improvement in execution time, a 55% reduction in pipeline breaks, 47% higher anomaly-detection accuracy, a 41% reduction in missing/null values, and a 38% reduction in engineering hours. The paper is explicit that LLM-driven agents may over-generalize on rare edge cases and that continuous LLM inference carries non-trivial computational cost; it also acknowledges that specialized domains such as healthcare or finance compliance may require targeted fine-tuning. Of the four seed papers, [4] is architecturally the most extensible toward a knowledge-graph substrate, since its metadata repository already implements lineage and schema-version tracking that could plausibly be extended into a semantic graph representation, though the paper itself does not describe or evaluate such an extension.

E. Cross-Study Synthesis

Reading the four studies together, a coherent picture of a layered agentic data-science stack emerges. At the lowest layer, [4] automates ingestion, profiling, and transformation with governance and lineage tracking; above it, [3] automates feature engineering, model selection, and interpretation once data has been prepared; at the perception/decision layer, [1] demonstrates that a non-visual orchestrator can meaningfully arbitrate between imperfectly calibrated specialist agents, particularly when supplemented with retrieval-grounded re-evaluation; and orthogonal to all three, [2] demonstrates that the human acceptance of such systems is governed less by resolved trust than by perceived usefulness and humanness, a finding that should inform how any of the preceding systems are surfaced to end users.

A recurring limitation across all four is the absence of an explicit, structured, multimodal knowledge representation: [1] retrieves from a flat image index, [3] and [4] operate on tabular/relational data without semantic graph structure, and [2] does not address system architecture at all. This absence is the principal gap motivating the integrated conceptual framework proposed in Section VIII.

VI. COMPARATIVE ANALYSIS

TABLE I
OVERVIEW OF REVIEWED STUDIES

Study	Year	Domain / Problem	Method	Key Contribution	Reported Limitation
[1] Roumeliotis et al.	2026	Zero-shot apple-leaf disease classification; agent trust	Orchestrator + 2 VLM agents + CLIP Image-RAG; ECE/OCR/CG trust metrics	+37.5 pp accuracy via trust-aware orchestration and RAG without fine-tuning	No dissenting signal when both agents confidently agree on a wrong label
[2] Hasselwander & Lah	2026	Gen Z adoption of autonomous AI agents (Germany, N=243)	PLS-SEM + Necessary Condition Analysis on survey data	Perceived usefulness/humanness are necessary conditions; trust is not a significant direct predictor	Self-reported data; single-country, cross-sectional sample
[3] Hasan & Tausif	2025	Automated tabular data-science pipeline (AutoML)	DFS feature synthesis + LLM text categorization + LIME + Bayesian optimization	Outperforms Auto-Sklearn 2.0 across 10 datasets (up to +13 pp accuracy)	No built-in outlier handling by default; sequential rather than negotiated multi-agent design
[4] Chirumamilla	2024	End-to-end autonomous data engineering	LLM multi-agent controller + reinforcement learning + metadata/lineage repository	62% faster execution, 55% fewer pipeline breaks, 47% better anomaly detection vs. manual pipelines	Risk of over-generalization on rare edge cases; high continuous-inference cost; limited public benchmarking detail

TABLE II
TECHNOLOGY AND APPROACH COMPARISON

Dimension	[1] Trust Orchestration	[2] Adoption Study	[3] AutoML Pipeline	[4] Autonomous Data Engineering
Automation level	Multi-agent, orchestrated	N/A (empirical study of users)	Automated sequential pipeline	Multi-agent, orchestrated + RL
Reasoning mechanism	LLM meta-reasoning (orchestrator)	Statistical (PLS-SEM/NCA)	Rule- and search-based (Bayesian opt.)	LLM reasoning + RL policy
Knowledge representation	Flat image embeddings (CLIP)	Latent survey constructs	Tabular feature matrix	Metadata repository (schema, lineage)
Retrieval mechanism	Image-RAG (vector similarity)	None	None	None (metadata lookup only)
Trust mechanism	ECE, OCR, Consistency Gap	Perceived trust (survey construct)	LIME interpretability	Governance / lineage auditability
Evaluation type	Benchmark classification accuracy	Survey-based structural model	Cross-dataset benchmark comparison	Comparative case study on 3 domains

TABLE III
COVERAGE OF INTEGRATION-RELEVANT CAPABILITIES ACROSS REVIEWED STUDIES

Capability	[1]	[2]	[3]	[4]
Autonomous workflow execution	Partial	N/A	Partial	Yes
Multimodal data integration	Partial (image+text)	No	No (tabular/text only)	Partial (structured/semi-structured)
Knowledge-graph generation	No	No	No	No (metadata graph only)
Graph-based retrieval	No (vector only)	No	No	No
Provenance / lineage tracking	No	No	No	Yes
Explainability	Partial (rationales)	N/A	Yes (LIME)	Partial (documentation)
Trust calibration (system-level)	Yes	N/A	No	Partial (governance)
User adoption evidence	No	Yes	No	No
Reproducible public evaluation	Yes (code released)	Yes (data collection described)	Partial (public datasets)	Limited (proprietary-scale data)

VII. RESEARCH GAPS AND OPEN CHALLENGES

A. Absence of Structured, Multimodal Knowledge Representation

None of the reviewed systems constructs or reasons over a knowledge graph. [1] retrieves from a flat CLIP embedding index; [3] and [4] operate on tabular/relational structures; [4]'s metadata repository stores schema and lineage but not semantic entity-relationship knowledge. The broader literature shows LLM-driven knowledge-graph construction is increasingly feasible [7] and that graph-based retrieval (GraphRAG) improves multi-hop reasoning and traceability over flat RAG [8]–[11], but this capability has not yet been integrated with trust-aware orchestration or autonomous data engineering in the reviewed work.

B. Disconnect Between Technical Trust and User Trust

The technical trust metrics used in [1] (ECE, OCR, Consistency Gap) quantify whether an agent's confidence is empirically justified, whereas [2] shows that perceived user trust is not the dominant driver of adoption intention. This suggests a gap between system-level trustworthiness engineering and user-facing acceptance design: a well-calibrated agentic system is not automatically an adopted one, and a system optimized only for adoption (e.g., via perceived usefulness) is not automatically trustworthy.

C. Consensus-Failure and Over-Generalization Modes

Both [1] and [4] independently report a failure mode in which agents confidently converge on an incorrect output with no internal mechanism to detect it — respectively, correlated overconfidence between two vision-language agents, and LLM over-generalization on rare edge cases in data transformation. This recurring pattern across otherwise unrelated domains suggests it is a structural property of LLM-agent ensembles rather than a dataset-specific artifact, and it remains unresolved in the reviewed literature.

D. Evaluation Scale and Benchmark Standardization

Evaluation practices vary widely: [1] uses an 800-image, four-class dataset; [2] surveys 243 respondents in a single country; [3] benchmarks on ten public datasets with clear metrics; [4] reports percentage improvements over unspecified manual baselines on large but proprietary-scale datasets without public benchmark release. This heterogeneity makes cross-study quantitative comparison (as attempted in Table II) qualitative rather than statistical, and points to a need for shared, public benchmarks for agentic data-science systems.

E. Computational Cost and Latency

Trust-aware orchestration with retrieval in [1] incurs roughly $1.3\times$ the inference time of a zero-shot baseline; [4] acknowledges the computational cost of continuous LLM inference at enterprise data volumes. Neither paper provides a systematic cost-accuracy trade-off analysis, which is necessary for deployment decisions in latency-sensitive settings.

F. Governance, Explainability, and Provenance

Only [4] explicitly targets lineage and compliance; [1] provides natural-language rationales but no formal provenance trail linking a final decision to specific retrieved evidence; [3] offers local interpretability via LIME but not pipeline-level provenance. A unified agentic data-science approach would need to combine [4]'s governance orientation with [1]'s evidence-grounded reasoning and [3]'s model-level interpretability. The broader XAI literature echoes this gap, arguing that trustworthy AI requires explanations that extend beyond local, model-level interpretability to system-level provenance, robustness, and human-centred evaluation [16], [17].

VIII. INTEGRATED CONCEPTUAL FRAMEWORK FOR THE REVIEWED LITERATURE

Building on the gaps identified in Section VII, this section proposes an integrated conceptual framework for the reviewed research landscape. The framework does not describe an implemented or validated system; rather, it is an original synthesis intended to connect, without copying, the architectural patterns reported in [1] and [4] with the automated-pipeline techniques of [3] and the adoption-relevant findings of [2], into a single conceptual picture of how these currently separate research strands could plausibly relate to one another.

The proposed conceptual flow is as follows: (1) heterogeneous data sources (tabular, textual, image) enter an ingestion and profiling stage, drawing on the autonomous ingestion and quality-intelligence agent design of [4]; (2) profiled data passes to an agentic feature and model layer combining the automated feature synthesis, text categorization, and interpretable modeling techniques of [3]; (3) in parallel, a knowledge extraction and multimodal knowledge-graph construction stage — informed by LLM-based KG construction methods [7] but not present in any reviewed system — would convert entities, relations, and cross-modal links (e.g., image-derived labels linked to tabular records) into a persistent graph; (4) a trust-aware orchestration layer, generalizing the calibrated-arbitration mechanism of [1] beyond image classification to the outputs of stages 2 and 3, would decide when agent outputs are reliable and when a graph-based retrieval and re-evaluation loop (a graph-native analogue of Image-RAG, informed by GraphRAG research [8]–[11]) should be triggered; and (5) outputs would be surfaced through a human-facing interface designed according to the adoption-relevant constructs identified in [2] — emphasizing perceived usefulness and clear, humanlike communication of system behavior and limitations — before reaching insights, visualization, and decision support.

Data Sources (tabular, text, image) → Agentic Ingestion & Profiling → Feature/Entity Extraction (interpretable) → Multimodal Knowledge-Graph Construction → Trust-Aware Orchestration ↔ Graph-Based Retrieval & Re-Evaluation → Human-Facing Insights & Decision Support

Fig. 1. Integrated conceptual framework synthesizing the reviewed literature — agentic data ingestion, interpretable automated feature engineering, multimodal knowledge-graph construction, and trust-aware, retrieval-grounded orchestration — derived from [1]–[4]. (Original figure; not reproduced from any source.)

This framework is presented as a motivating conceptual direction only. No experimental evidence in the reviewed literature demonstrates that such an integrated approach achieves the gains reported individually in [1], [3], and [4], nor that it resolves the consensus-failure modes discussed in Section VII-C; these remain open empirical questions for future research.

IX. FUTURE RESEARCH DIRECTIONS

Building on the identified gaps, promising directions for future research include: (1) LLM-driven, ontology-grounded construction of multimodal knowledge graphs directly from the heterogeneous outputs of autonomous data-engineering agents [4], [7]; (2) extending image-level Retrieval-Augmented Generation [1] toward graph-based retrieval [8]–[11] so that re-evaluation loops draw on structured, multi-hop evidence rather than flat nearest-neighbor similarity; (3) unifying agent-level trust calibration (ECE/OCR-style metrics [1]) with governance-level provenance tracking [4] into a single, auditable trust layer; (4) designing agentic-AI user interfaces informed by adoption research [2], explicitly optimizing for perceived usefulness and transparent communication rather than assuming that improved technical calibration alone will drive adoption; (5) developing shared, public, cross-domain benchmarks for autonomous data-science and data-engineering agents, addressing the evaluation heterogeneity noted in Section VII-D; and (6) systematically studying the consensus-failure and over-generalization modes identified independently in [1] and [4], potentially through adversarial or ensemble-diversity mechanisms that preserve a dissenting signal even when agents agree.

X. DISCUSSION

This review is a scoping synthesis anchored on four seed papers selected for their thematic relevance to the reviewed research area, rather than the output of an exhaustive systematic search; its conclusions about the state of the field should therefore be read as illustrative rather than comprehensive. The four seed papers also differ substantially in evaluation rigor and scale — from a controlled 800-image benchmark [1] and a 243-respondent survey [2] to multi-billion-record proprietary case studies [4] — which limits the strength of direct quantitative comparison and is reflected in the qualitative, rather than statistical, nature of Tables I–III. No meta-analysis or formal PRISMA screening was performed; a future version of this manuscript intended for IEEE journal submission should formalize the search protocol, report explicit identified/screened/included counts, and expand the supplementary literature search beyond the sources located during this review. Within these bounds, the synthesis nonetheless identifies a consistent and, to our knowledge, previously unremarked pattern: technical trust engineering (as in [1]) and user trust/adoption (as in [2]) are treated as separate problems in the current literature, and no reviewed system connects automated data-engineering/AutoML pipelines to a structured knowledge-graph layer, despite each capability being independently mature enough to be combined.

XI. CONCLUSION

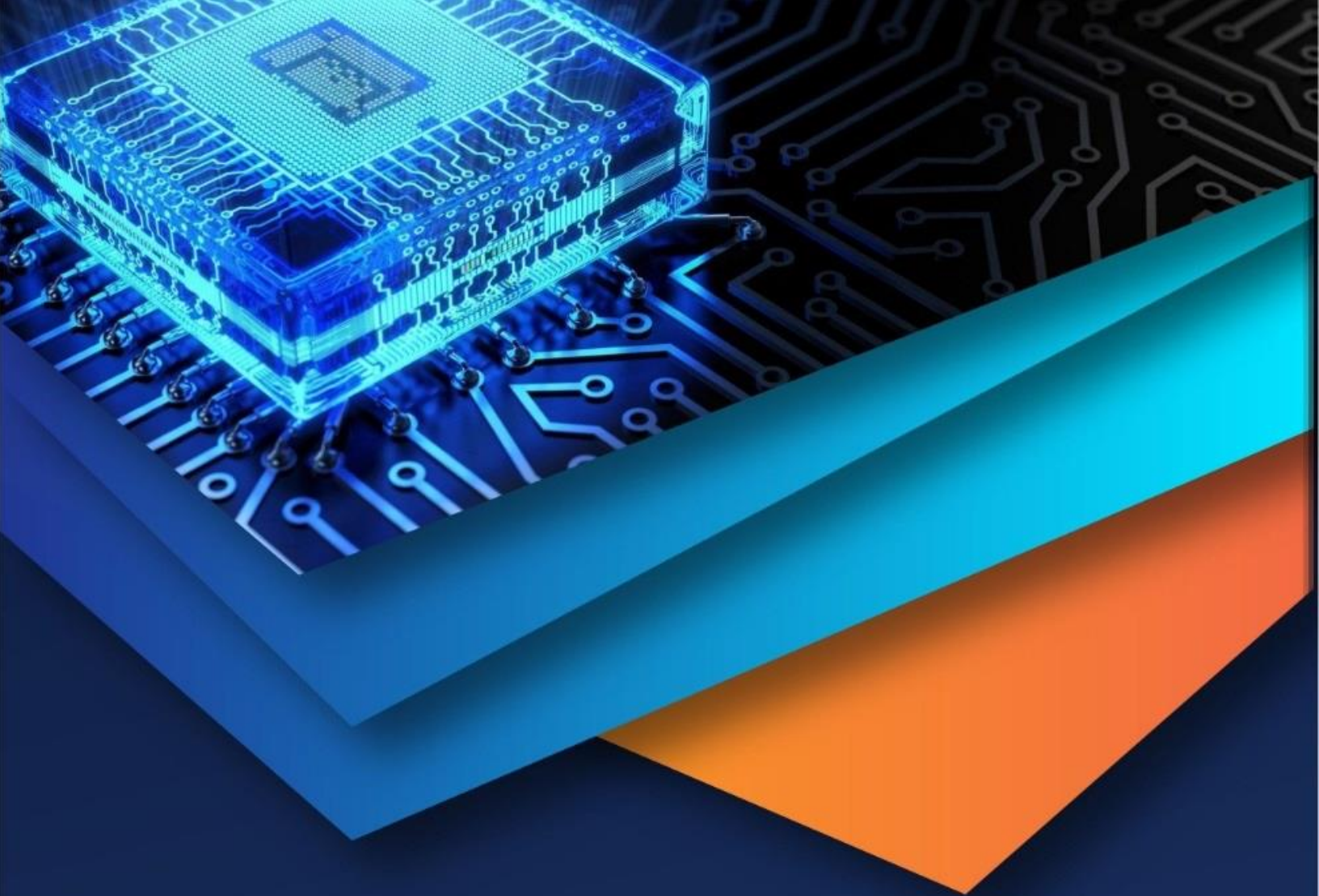
This review analyzed and synthesized four papers spanning trust-aware multi-agent orchestration for visual classification [1], human adoption of autonomous AI agents [2], an interpretable AutoML pipeline [3], and an autonomous end-to-end data-engineering system [4], situating them within the broader literature on agentic AI, retrieval-augmented and graph-based generation, LLM-driven knowledge-graph construction, and explainable AI. The comparative analysis shows that current agentic-AI research has independently matured several component capabilities — confidence-calibrated orchestration, retrieval-grounded re-evaluation, automated feature engineering, and autonomous pipeline governance — while leaving their integration with structured, multimodal knowledge representation, and their alignment with empirically grounded user-adoption principles, largely unaddressed. These gaps motivate, as a conceptual research direction rather than a demonstrated result, closer integration of autonomous agentic data science with multimodal knowledge-graph generation. Future empirical work should validate the proposed conceptual framework, formalize systematic review protocols, and develop shared benchmarks that allow the field to compare agentic data-science systems on common ground.

REFERENCES

- [1] K. I. Roumeliotis, R. Sapkota, M. Karkee, and N. D. Tselikas, "Agentic AI With Orchestrator-Agent Trust: A Modular Visual Classification Framework With Trust-Aware Orchestration and RAG-Based Reasoning," *IEEE Access*, vol. 14, pp. 26965–26982, 2026.
- [2] M. Hasselwander and O. Lah, "Agentic AI Arrives: How Gen Z Adopts Autonomous AI Agents," *IEEE Access*, vol. 14, pp. 27083–27090, 2026.
- [3] I. Hasan and M. Tausif, "Designing an Interpretable and Efficient AutoML Pipeline for Enhanced Data Analytics," in *Proc. 3rd Int. Conf. Self Sustainable Artif. Intell. Syst. (ICSSAS)*, 2025, pp. 911–916.
- [4] K. R. Chirumamilla, "Autonomous AI System for End-to-End Data Engineering," *Int. J. Intell. Syst. Appl. Eng. (IJISAE)*, vol. 12, no. 13s, pp. 790–801, 2024.
- [5] R. Sapkota, K. I. Roumeliotis, and M. Karkee, "AI Agents vs. Agentic AI: A Conceptual Taxonomy, Applications and Challenges," *Inf. Fusion*, vol. 126, Art. no. 103599, Feb. 2026.
- [6] S. Murugesan, "The Rise of Agentic AI: Implications, Concerns, and the Path Forward," *IEEE Intell. Syst.*, vol. 40, no. 2, pp. 8–14, Mar. 2025.
- [7] Y. Zhu, X. Wang, J. Chen, S. Qiao, Y. Ou, Y. Yao, S. Deng, H. Chen, and N. Zhang, "LLMs for Knowledge Graph Construction and Reasoning: Recent Capabilities and Future Opportunities," *arXiv:2305.13168*, 2024.
- [8] B. Peng, Y. Zhu, Y. Liu, X. Bo, H. Shi, C. Hong, Y. Zhang, and S. Tang, "Graph Retrieval-Augmented Generation: A Survey," *arXiv:2408.08921*, 2024.
- [9] D. Edge, H. Trinh, N. Cheng, J. Bradley, A. Chao, A. Mody, S. Truitt, D. Metropolitanaky, R. O. Ness, and J. Larson, "From Local to Global: A Graph RAG Approach to Query-Focused Summarization," *arXiv:2404.16130*, 2025.
- [10] Q. Zhang, S. Chen, Y. Bei, Z. Yuan, H. Zhou, Z. Hong, J. Dong, H. Chen, Y. Chang, and X. Huang, "A Survey of Graph Retrieval-Augmented Generation for Customized Large Language Models," *arXiv:2501.13958*, 2025.
- [11] H. Han, Y. Wang, H. Shomer, K. Guo, J. Ding, Y. Lei, M. Halappanavar, R. A. Rossi, S. Mukherjee, X. Tang, Q. He, Z. Hua, B. Long, T. Zhao, N. Shah, A. Javari, Y. Xia, and J. Tang, "Retrieval-Augmented Generation with Graphs (GraphRAG)," *arXiv:2501.00309*, 2024.
- [12] M. Feurer, K. Eggenberger, S. Falkner, M. Lindauer, and F. Hutter, "Auto-Sklearn 2.0: Hands-Free AutoML via Meta-Learning," *J. Mach. Learn. Res.*, vol. 23, no. 1, Art. no. 261, Jan. 2022.
- [13] J. M. Kanter and K. Veeramachaneni, "Deep Feature Synthesis: Towards Automating Data Science Endeavors," in *Proc. IEEE Int. Conf. Data Sci. Adv. Analytics (DSAA)*, 2015, pp. 1–10.
- [14] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why Should I Trust You? Explaining the Predictions of Any Classifier," in *Proc. ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2016, pp. 1135–1144.
- [15] S. Afroogh, A. Akbari, E. Malone, M. Kargar, and H. Alambeigi, "Trust in AI: Progress, Challenges, and Future Directions," *Humanities Soc. Sci. Commun.*, vol. 11, no. 1, pp. 1–30, Nov. 2024.



- [16] L. Longo et al., "Explainable Artificial Intelligence (XAI) 2.0: A Manifesto of Open Challenges and Interdisciplinary Research Directions," *Inf. Fusion*, vol. 106, Art. no. 102301, Jun. 2024.
- [17] S. Ali et al., "Explainable Artificial Intelligence (XAI): What We Know and What Is Left to Attain Trustworthy Artificial Intelligence," *Inf. Fusion*, vol. 99, Art. no. 101805, Nov. 2023.
- [18] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention Is All You Need," in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2017.



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)