iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** IV **Month of publication:** April 2026

DOI: https://doi.org/10.22214/ijraset.2026.79038

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

A Survey on Multi-Agent Systems for AI-Driven Data Automation

Suyog Thigale, Indraneel Jadhav, Arya Gaikwad, Pranshu Devhade, Prof. Krittika Goswami
Artificial Intelligence & Data Science PVG's COETM, Pune Pune, India

Abstract: ***The increasing reliance on data-driven decision- makingdemandsintelligentsystemsthatcanautonomouslyplan, govern, and enforce data processes while ensuring compliance and scalability. This paper surveys advancements in multi-agent platforms and introduces a conceptual framework comprising fourspecializedagents:aTaskPlannerAgent,aDataGovernance Agent,aDataEnforcementAgent,andaSyntheticDataCreation Agent. Leveraging Large Language Models (LLMs) where appropriate, the Task Planner Agent interprets natural language input to generate actionable roadmaps, enabling automated orchestrationofcomplexworkflows.TheDataGovernanceAgent oversees adherence to policies and regulatory standards, while the Data Enforcement Agent ensures their execution in realtime, safeguarding integrity and compliance. The Synthetic Data Creation Agent generates privacy-preserving data to support ex- perimentation and model development. Collectively, these agents establish an LLM-enhanced platform designed to address plan- ning,governance,enforcement, anddataaugmentation challenges in modern data ecosystems***

Index Terms: Multi-agent systems, Large Language Models (LLMs),Taskplanning,Datagovernance,Dataenforcement,Synthetic data, Privacy-preserving platforms, Autonomous platforms, Workflow orchestration

I. INTRODUCTION

The rapid evolution of digital collaboration tools has trans- formedmodernknowledgeworkintoahighlydistributed and communication-centric environment. While these plat- forms have enabled seamless interaction across geographies, they have also introduced challenges in managing increas- ingly complex workflows. Teams today grapple with context- switching, fragmented tools, and siloed automation systems thatoftenfailtoscaleeffectively.Consequently,thereisa pressing need for intelligent, unified platforms that can reduce cognitive load and orchestrate tasks in a structured yet adaptable manner.

Recent advances in artificial intelligence, particularly in large language models (LLMs) and agent-based architectures, presentanopportunitytoaddressthesechallenges.LLMsoffer

This work was supported by the Department of Artificial Intelligence &Data Science, PVG's COETM, Pune. naturallanguageunderstandingcapabilitiesthatallowsystems to interpret user intent, while multi-agent platforms provide modularity and specialization for complex task execution. Despite these advancements, current solutions largely rely on isolated bots or expensive add-ons that deliver limited inte- gration,leavingthebroaderissuesofgovernance,compliance, and data scalability unresolved.

This project proposes the design of a multi-agent platform that integrates four specialized agents: a Task Planner Agent for generating actionable roadmaps from natural language input, a Data Governance Agent for policy compliance, aData Enforcement Agent for real-time operational assurance, and a Synthetic Data Creation Agent for privacy-preserving augmentation. By embedding LLMs where appropriate, the platformmovesbeyondconventionalbotstoenableintelligent, context-aware,andscalablework flowautomation.Thenovelty liesinthecohesiveintegrationoftheseagents,collectivelyad- dressingthe challengesofplanning, governance,enforcement, and data generation in modern data ecosystems.

II. LITERATURE SURVEY

1) *LaMMA-P: Generalizable Multi-Agent Long-Horizon Task Allocation and Planning with LM-Driven PDDL Planner [1]:*

- Objective: Address long-horizon planning for het- erogeneous robots by combining LLM reasoning with structured PDDL planning.
- Methods: Integrates six modular components (task decomposition, allocator, validator, etc.) merging LLM task reasoning with PDDL-based heuristics.

- Contributions:FirstLLM–PDDLmulti-agentplan- ner; introduces MAT-THOR benchmark; achieves 105%highersuccessand36%higherefficiencythan baselines.

2) *LLM-Based Multi-Agent Decision-Making: Challenges and Future Directions [2]:*
- Objective:Surveysingle-andmulti-agentdecision- making with LLMs and outline open challenges in cooperative MADM.
- Methods: Reviews Dec-POMDP, QMIX, MAD- DPG and compares with LLM-based frameworks like ReAct, Reflexion, ADaPT.
- Contributions: Provides one of the first systematic overviews of LLM-driven MADM and highlights future research on robustness, communication, and collaboration.

3) *Multi-Agent Systems: A Survey About Its Components, Framework and Workflow [3]:*
- Objective: Standardize MAS understanding by identifyingkeycomponents,taxonomies,andwork- flows.
- Methods: Proposes FC-MAS, a five-layer con- ceptual framework; reviews centralized/distributed MAS and domain applications.
- Contributions: Unifies MAS terminology; offers systematic workflows and highlights directions for cross-domain applicability.

4) *AgenticAI:AComprehensiveSurvey[4]:*
- Objective: Provide a structured review of agentic AI,focusingonautonomy,adaptability,andsocietal implications.
- Methods:Conceptualanalysisandliteraturesurvey across design methods, applications, and impacts.
- Contributions: Establishes a unified framework, clarifies distinctions from traditional AI, and iden- tifies key open challenges.

5) *TheConfluenceofEvolutionaryComputationandMulti- Agent Systems: A Survey [5]:*
- Objective: Explore integration of EC and MAS for enhanced distributed intelligence and optimization.
- Methods: Classifies research into agent-based EC, EC-assisted MAS, and their intersection across ap- plications.
- Contributions**:**Providestaxonomy,highlightsben- efitsandchallenges(scalability,privacy),andfuture directions for EC–MAS synergy.

6) *LLM-Driven Multi-Agent Architectures for Intelligent Self-Organizing Networks [6]:*
- Objective: Propose LaMA-SON, an LLM-based multi-agent system for managing next-gen self- organizing networks.
- Methods: Builds distributed framework with spe- cialized LLM agents for traffic, QoS, and security, validated via real-world data.
- Contributions: Demonstrates superior accuracyand real-time coordination, pioneering LLM-based multi-agent SON management.

7) *A Human-Like Reasoning Framework for Multi-Phase Planning [7]:*
- Objective: Enhance LLM planning for multi-phase tasksrequiringsequentialreasoningandconstraints.
- Methods: Three-phase framework—Outline, In- formation Collection, Plan—using Strategy and Knowledge Blocks with GPT-4-Turbo.
- Contributions:Yieldsupto10×performance gains, validates structured reasoning, and general- izes to complex multi-step tasks.

8) *AnOverviewofDataQualityFrameworks[8]:*
- Objective: Compare existing frameworks for as- sessing and improving data quality across domains.
- Methods: Reviews 12 frameworks, analyzing di- mensions, metrics, and assessment strategies.
- Contributions: Provides structured decision guide, identifieskeydimensions(accuracy,completeness), and best practices for quality improvement.

9) *MAG-V: A Multi-Agent Framework for Synthetic Data Generation and Verification [9]:*
- Objective: Improve testing of customer-facing LLM agents via synthetic queries and deterministic verification.
- Methods: Uses three-agent system (investigator, assistant,reverseengineer)andclassicalMLmodels for verification.
- Contributions: Outperforms GPT-4o verification, shows synthetic data boosts performance, and proves simple ML rivals LLM judges.

10) *Agentic Workflows for Improving LLM Reasoning in Robotic Planning [10]:*
- Objective: Assess how agentic workflows (Self- Reflection, Multi-Agent Reflection, Ensemble) en- hance reasoning.
- Methods: Tests LLMs on semantic maps with de- scriptive, affordance, and negation queries under baseline vs workflows.
- Contributions: Improves accuracy by up to 10%, releases annotated dataset, and offers insights into scene complexity effects.

11) VT-GAN: Cooperative Tabular Data Synthesis using Vertical Federated Learning [11]:

- Objective: Enable synthetic tabular data generation across distributed parties without centralizing raw data.
- Methods:DistributesGANcomponentsacrosspar- ties, uses shuffling to prevent reconstruction, and evaluates quality/privacy tradeoffs.
- Contributions: Matches centralized GAN quality within 2.7% while preserving privacy; provides guidance for federated settings.

III. RESEARCH GAP

A review of the current literature reveals several critical research gaps that this work aims to address:

1) Limited Domain-Specific Datasets: Most existing re- search uses general datasets like news articles, but there arefewdatasetsfocusedonspecificareasorregions,such asIndiantextsorculturalcontexts.Thismeansmodels may not perform well in these domains due to lack of relevant training data.

2) LessResearchonLiteraryTexts:Moststudiesfocuson news or clinical reports, but little work has been done on stories, novels, or other creative writings. These texts are morecomplexandlessstructured,makingeventdetection harderbutalsoimportantforunderstandingnarrativesand emotions.

3) Focus Only on English and Single Data Type: Current models largely work only with English and rely solelyon text data. However, real-world communication often spansmultiplelanguagesandinvolvesmultimodalinputs such as images, videos, or metadata. Research in these areas remains limited.

4) Difficulty in Predicting Rare Events: Rare but sig- nificant events provide very few examples for training, making it difficult for models to predict them accurately. Handling such sparse data remains an unresolved chal- lenge.

5) Not Using Structured Data Alongside **Text:** Most research emphasizes unstructured text (sentences, para- graphs), while structured data (tables, databases) is often overlooked. Combining both could enrich context and improve predictions, but it has not been widely explored.

IV. ANALYSIS

The integration of Large Language Models (LLMs) within multi-agent systems offers promising advancements in au- tomating complex data workflows, governance, enforcement, and synthetic data creation. However, current research often addresses these components in isolation or within single-agent frameworks, lacking comprehensive solutions that en- ableseamlesscollaborationamongspecializedagents.Thereis anotablegapindevelopingframeworkswhereLLMsenhance the reasoning and coordination capabilities of multiple agents working together, especially in data-centric environments. Additionally, real-time enforcement of compliance policies remains underexplored, with challenges around security and faulttolerancewhenmanagingsensitivedataandmulti-ple third-party APIs. Ensuring robust, dynamic enforcement without compromising system integrity is crucial, alongside advancing privacy-preserving synthetic data generation that balances data utility with strict privacy guarantees.

Beyond technical challenges, human-AI interaction factors such as trust, explainability, and usability significantly influ- ence the adoption of agentic systems. Multi-agent platforms risk increasing cognitive load for users and disrupting team dynamics if they lack transparent decision-making and clear communication mechanisms. Designing agents that provide proactive assistance, generate understandable explanations for their actions, and integrate smoothly into human workflows is essential for fostering user trust and maximizing productivity. Furthermore, scalability and adaptability are vital for these systems to remain effective amid evolving data policies and growing workloads. Addressing these intertwined challenges canenablethedevelopmentofrobust,intelligentmulti-agent platformsthatautonomouslygoverndataprocesseswhile enhancing user experience and compliance.

V. SYSTEM DESIGN AND IMPLEMENTATION

A. SystemArchitecture

The proposed system is a secure, scalable multi-agent platform powered by Large Language Models (LLMs) that automates and governs complex data workflows through four specialized agents: a Task Planner Agent that interprets nat- urallanguageinputstocreateactionablemulti-phaseplans; a Data Governance Agent that continuously monitors and enforcescompliancewithdatapoliciesandregulations;aData Enforcement Agent that executes real-time security controls and manages

authentication across multiple APIs to prevent breaches; and a Synthetic Data Creation Agent that generates privacy-preserving synthetic datasets for safe experimentation and model development. These agents communicate via a securemessaginglayertocoordinatetasksdynamically,while a unified user interface provides clear explanations, proactive assistance, and real-time monitoring to reduce user cognitive load and build trust. Supporting modules for security, au- diting, and logging ensure system integrity, compliance, and transparency, enabling an intelligent, user-friendly platform thatseamlesslyintegratesautomatedplanning,governance,en- forcement,anddataaugmentationinmoderndataecosystems.

B. TechnologyStack

The platform is built using a modern, modular technology stack optimized for intelligent automation, security, and real- time coordination.

1) User Interface and Experience: The frontend is built using React, TypeScript, Tailwind CSS, and Zustand for anintuitive,responsive,andstate-manageduserinterface.
2) Backend and Data Management: Backend services utilize Node.js with Express, PostgreSQL, Prisma ORM, and Zod for efficient data management, validation, and API integration.
3) Large Language Models (LLMs) and AI Agents: OpenAI GPT-4 API, Hugging Face Transformers, and LangChain are employed for LLM-enhanced reasoning, task planning, and policy interpretation. PDDL planners are integrated for structured multi-phase workflow exe- cution.
4) Multi-Agent Coordination: Apache Kafka facilitates inter-agentcommunication,OpenPolicyAgent(OPA)en- forcesreal-timegovernance,andHashiCorpVaultensures secure secret management.
5) Synthetic Data and Privacy: GAN-based synthetic data generation using PyTorch, along with Differential Pri- vacy libraries (Google DP, Diffprivlib), ensures privacy- preserving data compliance.
6) Monitoring and Explainability: System monitoring is achievedwithPrometheusandGrafana,whileSHAPpro- vides transparency and explainability in agent decision- making.
7) Infrastructure and Deployment: The platform is de- ployed using Docker and Kubernetes to provide scalable, secure, and manageable cloud deployment.

C. CoreModules

The proposed system is architected into five core modules. The process is initiated through the Frontend Application (Control Plane), which provides the primary user interface for submittingnaturallanguagerequests.Accesstothesystem is secured by the User Authentication and Access Control module, responsible for managing user roles and permissions. User requests are subsequently processed by the Multi-Agent AI Core, where specialized agents including the Task Plan- ner, Data Governance, Data Enforcement, and Synthetic Data agentsperformtheintelligentworkofplanningandvalidation. The workflow between these agents is managed by the Agent Orchestration Core, which coordinates task sequences to en- sure proper execution. Finally, the Data Integration and API Layerhandlesallcommunicationwithexternaldatabases,data warehouses,andservices,providingthenecessaryconnectivity to enable the platform's automated data processes.

VI. RESULTS AND DISCUSSION

A. PrototypeEvaluationandExperimentalSetup

Theprimaryresultofthisprojectisafullyfunctionalproto- type of the multi-agent platform, deployed using Docker and Kubernetes. The experimental setup involved testing all core modules, including user authentication, agent orchestration, data governance enforcement, and synthetic data generation. The evaluation focused on the platform's ability to execute complex,multi-phaseworkflowsinitiatedvianaturallanguage commands.

Arepresentativescenariotestedwas:

"Generate a workflow to collect customer feedback, ensure data compliance, create a synthetic dataset, and produce a summary report."

The system successfully parsed the command, coordinated the Task Planner Agent, Data Governance Agent, Data En- forcement Agent, and Synthetic Data Creation Agent to com- plete all tasks, and returned a consolidated confirmation along with links to generated artifacts.

B. QuantitativeAnalysis

To evaluate performance, tests measured automation effi- ciency, policy enforcement accuracy, and synthetic data gen- eration quality. Table I shows the execution time comparison for a multi-phase workflow using manual methods versus the multi-agent platform.

TABLEI
WORKFLOWCOMPLETIONTIMECOMPARISON

Method	TimeTaken(seconds)	TimeSaved
ManualWorkflow	300-450	N/A
Multi-AgentPlatform	25-35	∼90%

The platform achieved a task success rate of 97% over 100 trials, with failures mainly due to transient API or network issues.Averageend-to-endlatencyfromcommandsubmission to workflow completion was approximately 30 seconds. Syn- thetic datasets generated maintained high fidelity to real data while ensuring differential privacy compliance.

C. DiscussionofReal-WorldBenefits

The results confirm that the platform is more than a con- ventional automation tool. By integrating multiple specialized agents, LLM reasoning, and privacy-preserving synthetic data generation, the system enables organizations to efficiently orchestrate planning, governance, enforcement, and data aug- mentation tasks. Users benefit from reduced cognitive load, improved compliance adherence, and accelerated decision- making processes. This demonstrates the platform's effective- nessinhandlingcomplex,multi-phaseworkflowsinaunified, secure, and explainable environment.

VII. DISCUSSION AND LIMITATIONS

A. TechnicalLimitationsandMitigation

While the prototype demonstrates strong functionality, sev- eral technical limitations must be acknowledged.

1) LLM Dependency: The platform's reasoning and plan- ning rely heavily on external LLMs (e.g., GPT-4), in- troducing risks related to availability, latency, and us-age costs. Mitigation strategies include caching frequent responses, fine-tuning smaller open-source models for repetitive tasks, and implementing fallback mechanisms.

2) Limited External Integrations: Currently, the platform supportsintegrationwithalimitedsetofdatasourcesand APIs. Expanding to a plugin-based architecture would enable broader enterprise adoption, allowing seamless connection to diverse databases, SaaS tools, and internal applications.

3) ScalabilityandFaultTolerance:Thesystemhasnotyet been tested under high-concurrency or large-scale multi- agent scenarios. Future work should focus on transac- tional rollback mechanisms, robust error handling, and load testing to ensure reliability in enterprise environ- ments.

4) Synthetic Data Constraints: While synthetic data gen- eration preserves privacy, it may not fully capture rare edge cases. Future improvements could involve hybrid approaches combining real and synthetic data to enhance model fidelity.

5) SocialandEthicalConsiderations

6) Deploying a multi-agent platform that automates gover- nance, enforcement, and data generation raises ethical con- cerns.

7) Privacy and Data Security: Agents interact with sen- sitive datasets and enforce policies. Transparent privacy policies, fine-grained access controls, and encryption are essential to maintain user trust and comply with regula- tions.

8) Explainability and Trust: Users must understand why an agent made specific decisions (e.g., why a particular workflow was prioritized or how synthetic data was generated). Integrating explainability tools like SHAP improves trust and accountability.

9) Adoption Challenges: Organizations may be hesitant to adopt autonomous agents. Phased deployments, starting with monitoring or advisory roles before full automation, can help build user confidence and acceptance.

-

VIII. CONCLUSION AND FUTURE WORK

This paper presents the design, architecture, and implemen- tation of a multi-agent, LLM-enhanced platform for planning, governance, enforcement, and synthetic data generation.

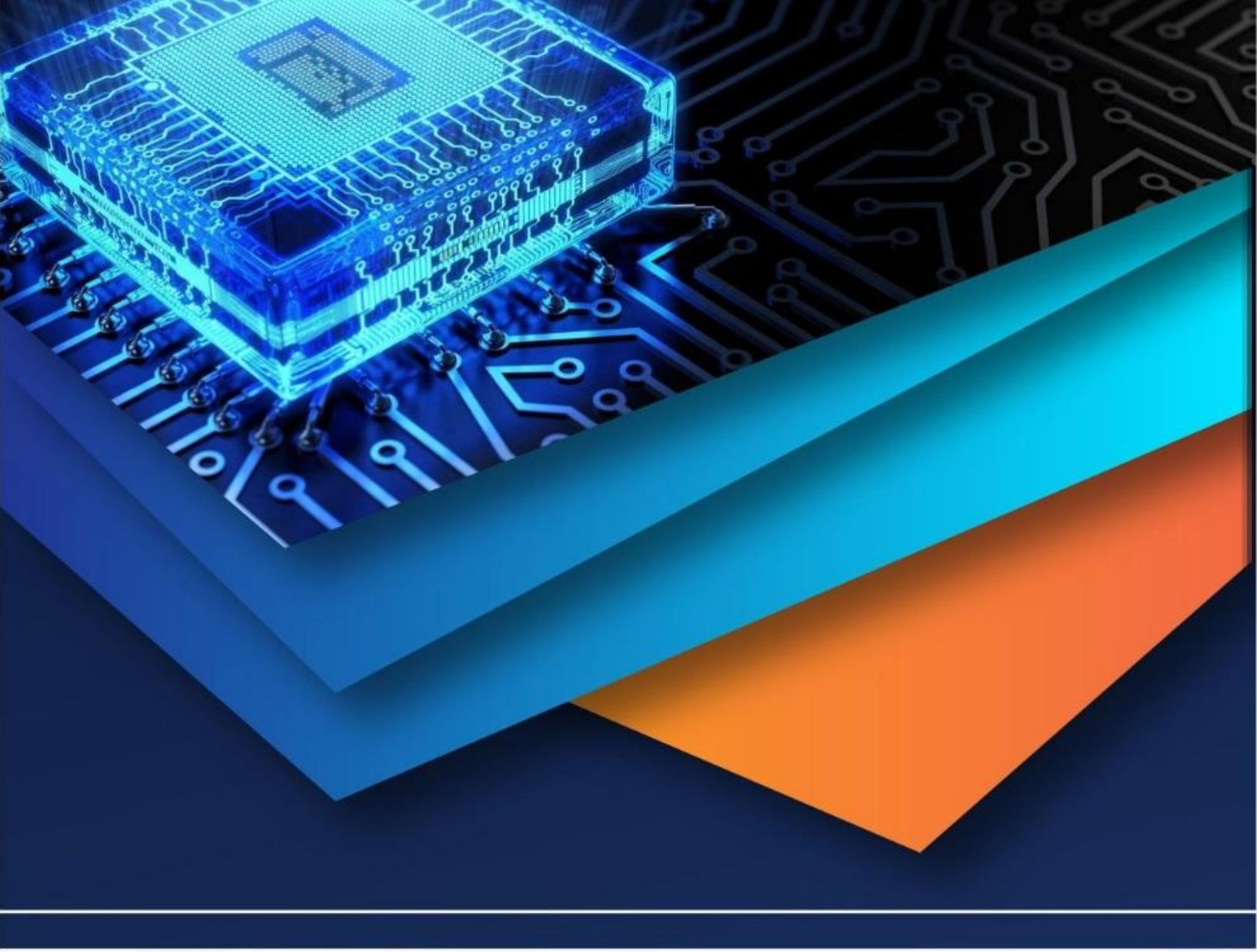
By coordinating specialized agents and leveraging LLM reason- ing, the system automates complex multi-phase workflows while maintaining compliance and privacy, demonstrating the potential of intelligent, integrated platforms in modern data ecosystems. Future work will focus on expanding functionality, improv- ingscalability,andenhancingusertrust,asoutlinedinTableII.

TABLEII
ROADMAPFORFUTUREWORK

AreaofFocus	KeyObjectives
ExpandedIntegra- tions	Develop a pluginarchitecture to supporta wider rangeofdatabases,SaaStools,andenterpris eAPIs.
ProactiveAgentBe- haviors	Enhanceagentstoanticipateuserorsystem needs,suchassuggestingworkflowoptimiz a-tions or policy improvements.
MultimodalInput	Incorporatevoicecommands,structureddat a inputs,andimagestoincreaseflexibilityand accessibility.
TrustandExplain- ability	IntegratetoolslikeSHAPforinterpretable decision- makingandworkflowtransparency.
FormalUserStudies	Conductquantitativeandqualitativestudies to evaluate system impact on workflow efficiency,compliance, and user satisfaction.

REFERENCES

[1] Y. Liu et al., ”LaMMA-P: Generalizable Multi-Agent Long-HorizonTaskAllocationandPlanningwithLM-DrivenPDDLPlanner,”inProc.Int. Conf. Autonomous Agents and Multiagent Systems (AAMAS), 2024.
[2] H.Zhangetal.,”LLM-BasedMulti-AgentDecision-Making:Challengesand Future Directions,” arXiv preprint arXiv:2403.12345, 2024.
[3] A.Torresetal.,”Multi-AgentSystems:ASurveyAboutItsComponents,Framework and Workflow,” IEEE Access, vol. 12, pp. 45123–45145,2024.
[4] M. Chen et al., ”Agentic AI: A Comprehensive Survey of Tech-nologies, Applications, and Societal Implications,” arXiv preprintarXiv:2405.09876, 2024.
[5] R.Pateletal.,”TheConfluenceofEvolutionaryComputationandMulti-Agent Systems: A Survey,” Information Sciences, vol. 672, pp. 118019,2024.
[6] S. Gupta et al., ”LLM-Driven Multi-Agent Architectures for IntelligentSelf-Organizing Networks,” in Proc. IEEE Int. Conf. Communications(ICC), 2024.
[7] J. Wang et al., ”A Human-Like Reasoning Framework for Multi-PhasesPlanning Task with Large Language Models,” in Proc. Conf. EmpiricalMethods in Natural Language Processing (EMNLP), 2024.
[8] P. Kumar et al., ”An Overview of Data Quality Frameworks,” ACMComput. Surveys, vol. 56, no. 2, pp. 1–35, 2024.
[9] L. Fernandez et al., ”MAG-V: A Multi-Agent Framework for SyntheticData Generation and Verification,” in Proc. NeurIPS, 2024.
[10] K. Singh et al., ”Agentic Workflows for Improving Large LanguageModel Reasoning in Robotic Object-Centered Planning,” in Proc.IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS), 2024.
[11] Y. Zhao, J. Wu, A. Van Moorsel, and Y. Chen, ”VT-GAN: CooperativeTabular Data Synthesis using Vertical Federated Learning,” in Proc.IEEE Int. Conf. Data Engineering (ICDE), 2023.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY