iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** VI **Month of publication:** June 2026

DOI: https://doi.org/10.22214/ijraset.2026.83838

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 VI June 2026- Available at www.ijraset.com

A Survey on Autonomous Data Analyst Using Artificial Intelligence

Dr. Ravindra S[1], Samanth S R[2], Sarvesh G M[3], Veeresh K H[4], Shashank T L[5]
[1]*Associate Professor,* [2,3,4,5]*Sixth Semester Students, Department of Computer Science and Engineering , Jawaharlal Nehru New College of Engineering, Shivamogga, Visvesvaraya Technological University Belagavi, Karnataka India*

Abstract: ***This investigative paper reviews the operational landscape of self-directed analytical computing entities designed to orchestrate entire data management lifecycles. Classical reporting loops frequently break down due to unstable, manually executed methods required during information formatting, layout structuring, and graphical rendering. Conversely, the framework described here natively interprets everyday human phrasing, manipulates complex multi-variable tables, and uncovers obscure correlations. Operating independently of manual supervision, this computing engine handles diverse text representations (including standard comma-separated logs, tabular sheets, and loosely structured arrays) while handling baseline requirements like value normalisation, feature selection, and document engineering. By deploying adaptive logical models and automated execution strategies, this structural blueprint accelerates calculation times, expands operational scaling, and refines validation precision, allowing enterprise teams to secure advanced analytical capabilities without needing engineering support.***

Keywords: Autonomous Data Agents, Large Language Models, Automated Analytics, Machine Learning.

I. INTRODUCTION

The contemporary architecture of global data tracking yields massive volumes of scattered information every moment across enterprise workflows, training centers, medical spaces, and digital social platforms. Extracting actionable value from these repositories traditionally demands highly technical proficiency in complex query syntax and mathematical scripting systems like SQL or Python alongside standalone visualization applications. Consequently, old-fashioned database infrastructures remain bottlenecked by high maintenance overloads, structural fragmentation, and poor usability for non-technical company staff.

To bridge this operational gap, rapid progress combining specialized artificial agents, transformer language architectures, and performance metric evaluation presents highly dynamic options. Key foundational works—specifically those investigating automated student path prediction models alongside comprehensive reviews of conversational analytical toolkits—prove that state-of-the-art software configurations can successfully parse nested matrices, run forecasting simulations, and guide corporate strategy with zero human intervention.

Specifically, transformer-powered computing analysts transform raw processing pipelines by:

1) Converting unstructured human speech loops straight into executable code snippets.
2) Normalising disparate information schemas spanning independent physical databases.
3) Generating production-ready enterprise summaries and dynamic graphical interfaces.
4) Executing advanced statistical regression checks and localized recommendation scripts.

By trading out manual engineering loops for automated, self-correcting logic models, these modern systems locate intricate underlying patterns and evaluate market velocities far more effectively than traditional business software.

II. MOTIVATION

As corporate data collection scales up exponentially, manual tracking and visualization pipelines break down under massive structural stress. Because most modern business setups lack specialized database engineers or dedicated information analysis branches, a significant portion of incoming files remains dark, unparsed, and wasted.

Simultaneously, rapid developments across the machine learning market confirm that standalone system frameworks can successfully abstract away these technological hurdles. Modern foundational models process row-and-column arrays directly, formulate multi-step execution graphs, and deliver explicit strategic recommendations instantly.

Driven by these capabilities, this project designs an automated processing framework focused on:

1) Eliminating coding barriers to entry to democratize zero-code statistical analysis.

2) Tuning ingestion mechanics to minimize update lag in business intelligence tools.
3) Maximising the validity of uncovered insights using strict system validation.
4) Constructing data-reliant operations by putting advanced analytical tools into everyday business hands.

III. LITERATURE SURVEY

In [1], the author assesses how advanced statistical models isolate customer choices and shifting retail dynamics from massive enterprise transaction logs. By tracking historical trends, geospatial population groups, and digital sentiment indices, the investigation extracts granular insights from consumer locations, financial backgrounds, and lifestyle choices. The author highlights how deep application integration increases sales projection accuracy, improves client brand interaction, and enables hyper-focused marketing campaigns. Finally, the text evaluates core ethical areas, mapping software design requirements to guard consumer privacy, protect algorithmic clarity, and prevent structural database bias.

The authors in [2], present Smart Grade, a cloud-based diagnostic infrastructure designed to automate academic tracking metrics inside educational centers. The architecture balances live interface dashboards, automatic summary generation, and instant alert modules to give administrators clear institutional oversight. Using tree-based ensemble methods like Random Forest, the platform successfully flags underperforming students early and projects final testing scores. The developer creates specialized user views for students, instructors, and deans, proving that automated computational pipelines accelerate early counseling setups and improve student retention rates.

A structural categorization of the mechanical layout of modern Autonomous Data Agents is provided in [3]. The study explores how pairing language models with structural logic engines, task factorization, and localized code compilation fully automates old reporting routines. This framework is compared to manual data engineering systems, which are restricted by syntax errors and poor vertical scaling properties. The proposed framework solves these issues by dividing high-level milestones into manageable subtasks, compiling clean operational scripts, and printing final documentation autonomously. The internal agent configuration relies on connected layers handling sensory input, variable memory banks, and tool integration.

An institutional tracking system that captures learning metrics by linking varied student datasets through Graph Neural Networks (GNNs) and linguistic transformers is constructed in [4]. By matching course registries with campus spending patterns and demographic categories, the tool builds comprehensive profiles of student engagement. The processing core runs custom graph embeddings, sequence tracking modules, and mathematical entropy checks using a blended Llama 3.2 and LangChain architecture. The author highlights essential data preparation cycles—such as timestamp normalization and feature engineering—concluding that while these models optimize prediction accuracy, research hurdles remain regarding internal model visibility and cross-campus porting.

In [5], the authors evaluate how educational data analysis platforms and dynamic tutoring frameworks optimize student mastery levels. The study explores the intentional application of student tracking dashboards and predictive early-warning tools to isolate learning friction points across student groups. While validating the operational speed improvements of automated tutoring models, the writer underscores that engineering teams must watch out for over-reliance on automation, demographic data skewing, and database leakage risks, demanding strict adherence to ethical design principles during live deployments.

An extensive diagnostic overview tracking the evolution from legacy software platforms to fully self-directed analytics frameworks is completed in [6]. The survey demonstrates how current multi-modal text architectures interpret highly mismatched data arrays, including unstructured files and chaotic string lists. It evaluates contextual semantic parsing, interactive inquiry pathways, and multi-layered pipeline execution, verifying that modern agents coordinate sorting, filtration, visualization, and paper assembly without human steps.

The integration of predictive diagnostic code and machine learning routines directly within cloud-based enterprise CRM architectures is the primary focus of [7]. Using specialized Random Forest models and gradient booster frameworks, the investigator automates prospective lead conversion scoring, client attrition risks, and revenue modeling. Field metrics confirm that closed-loop, self-correcting algorithms deliver tangible improvements in pipeline closure rates, team quota achievements, and financial planning confidence while cutting sales cycle lengths compared to legacy static databases.

Finally, a high-performance classification tool that monitors student academic drop-out risks using Support Vector Machines, decision matrices, and feed-forward Neural Networks is built in [8]. Tracking digital platform login rates, class presence, and family lifestyle variables, the model crosses an 85% categorization certainty threshold. The empirical outputs establish a strong link between consistent interaction with online learning environments, high lecture attendance, and final grade rankings, while emphasizing the absolute need for operational transparency and explainability in predictive framework development.

A. Summary of Review

The evaluated literature shows that artificial intelligence breakthroughs, machine learning pipelines, and autonomous agent systems have fundamentally transformed the modern data science space. In commercial industries, these toolsets handle complex logical challenges, improve corporate cash flows, and refine raw datasets into strategic pathways. In educational settings, automated predictive software isolates academic failure indices ahead of time, driving faster data-backed student support paths.

Crucially, the current generation of autonomous AI structures removes human programming entirely. Operating via simple conversational commands, these software agents independently clean database errors, patch missing variables, write optimized database scripts, and build functional metrics dashboards. However, researchers point out that widespread industry adoption remains tied to fixing system risks like textual hallucinations, data leaks, and compute resource bottlenecks.

IV. CONCLUSIONS AND FUTURE WORK

This survey paper has evaluated the rapid structural shift toward Autonomous Data Agents across corporate and educational landscapes. Current research confirms that pairing LLMs with advanced reasoning and task-planning architectures minimizes manual coding, simplifies data engineering, and eliminates legacy software constraints. These agents lower the barrier to entry for data science, allowing non-technical managers to safely interrogate complex databases using conversation alone.

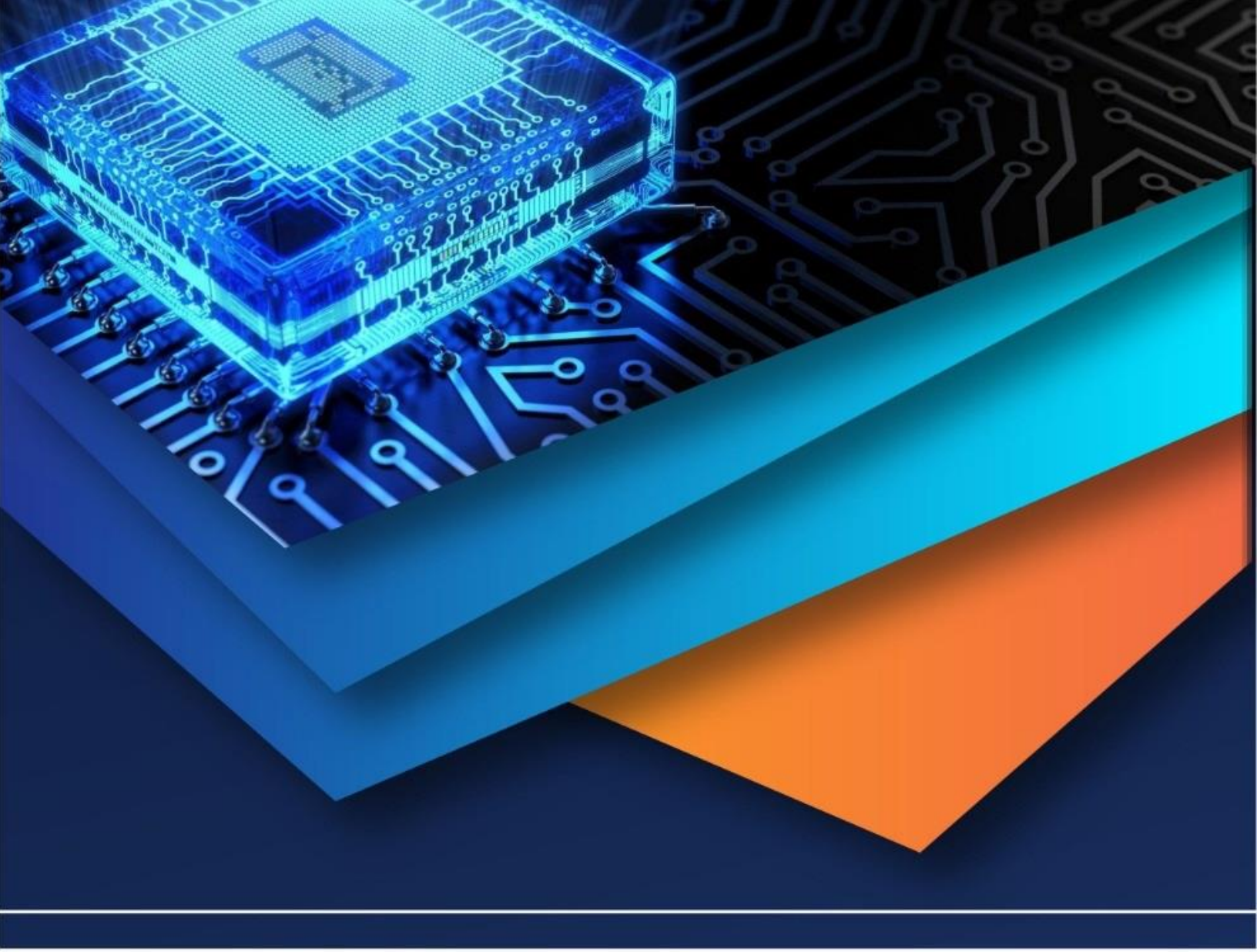
Nonetheless, moving toward fully autonomous execution requires resolving serious structural limitations. Current systems remain vulnerable to algorithmic biases, data leaks, and semantic hallucinations during automated reporting. Consequently, future research must prioritize the development of cross-validation architectures that continuously audit AI code execution. Furthermore, integrating closed-loop reinforcement learning pipelines will be essential to help autonomous agents adapt dynamically to multi-source enterprise systems without breaking.

V. ACKNOWLEDGMENT

The authors express their gratitude to Jawaharlal Nehru New College of Engineering and the Department of Computer Science and Engineering for providing the computational resources, facilities, and academic guidance required to complete this research.

REFERENCES

[1] K. R. Mudunuru, R. Remala, and S. K. S. Nagarajan, "AI-Driven Data Analytics: Unveiling Sales Insights from Demographics and Beyond," International Journal of Computer Sciences and Engineering, vol. 12, no. 5, pp. 11-18, May 2024,
[2] C. K. Gomathy, A. Thamil Neethi, and S. Sai Krishna, "Automating Student Performance Analysis Using Machine Learning,"Singaporean Journal of Scientific Research, vol. 17, no. 1, pp. 62-70, 2025.
[3] Y. Fu, D. Wang, W. Ying, X. Wang, X. Zhang, H. Liu, and J. Pei,“Autonomous Data Agents: A New Opportunity for Smart Data, "arXiv preprint, 2025.
[4] H. Li and Z. Liu, "An Intelligent Educational System: Analyzing Student Behavior and Academic Performance Using Multi-Source Data, "Electronics, vol. 14, 2025.
[5] M. Shuyeb, S. Kumar, M. Kumar, and A. Jain, "The Impact of AI and Learning Analytics on Students' Performance, "Research Paper, 2026.
[6] Z. Tang, W. Wang, Z. Zhou, et al.,"Large Language Models for Data Analytics: A Survey,"arXiv preprint, 2025.
[7] S. Thota, "Automated Sales Performance Optimization Using AI on Salesforce Platform, "Research Study, 2024.
[8] P. Somani, "Student Academic Performance Prediction Using Machine Learning Algorithms, "Research Study, 2025.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY