BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

Autonomous AI Agents for Automated Data Analytics

[1]Bhagyashree Basavaraj Pujari

[1]Department of Artificial Intelligence and Data Science,
[1]Ajeenkya DY Patil School of Engineering, Pune, India
Email: bhagyashreepujari08@gmail.com

[2]Prof. Ankita Pokale

[2]Department of Artificial Intelligence and Data Science,
[2]Ajeenkya DY Patil School of Engineering, Pune, India
Email: ankita.pokale@gmail.com

Abstract

Traditional data analytics face challenges due to tedious SQL-based querying, reliance on manual data cleaning, and static dashboards, making the process time-consuming and user-intensive. Autonomous AI agents provide an alternative, wherein Large Language Models (LLMs) are used for analytics ranging from natural language questioning to planning, analytical coding, and task execution requiring less human input. This paper studies autonomous AI agents for automated data analytics, describing the key components of such an agent, including the underlying LLM core, planning, reasoning, memory, and tool usage modules. Additionally, this paper describes the applications of such an agent throughout the various stages of the data analytics life cycle, including collection and cleaning, exploratory data analysis, modeling, and reporting. Following a review and analysis of seven recent works on the topic, this study identifies five core areas for future research: error handling in data cleaning, autonomous decision-making, data security and privacy, multi-modal data processing, and enhanced analytical capabilities. Finally, this paper provides a list of requirements for developing such a system for practical deployment.

Keywords: Autonomous AI agents, data analytics, exploratory data analysis, large language models, machine learning, natural language processing, reasoning, tool usage.

I. Introduction

Organizations are currently producing data at rates beyond comprehension, but leveraging this information for insights is extremely challenging. In the traditional setup, an analyst is responsible for writing the required SQL queries, cleaning and preparing the data, and designing dashboards. Overall, the task is very time-consuming for the business to get results fast. A person who is not familiar with SQL or Python programming cannot query a database; they have to rely on specialists.

Autonomous AI agents address the problems associated with the need for human intervention in the data analytics process. An agent here generally refers to an entity with an LLM at its core that has the ability

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

to understand and reason about natural language, plan, write code, and use external software to accomplish a task [5]. For the purposes of analytics, it means that a natural language request from a user can be translated into an analysis with no additional coding. An autonomous agent covers the whole data analytics lifecycle, including extracting, cleaning/preparing, analyzing, mining, and reporting [2], [3].

The objective of this paper is to

- explore the concept of autonomous AI agents for automated data analytics,
- discuss the role of LLMs, planning, reasoning, and tool use in autonomous data analysis,
- examine key operations such as cleaning and EDA that comprise practical data analytics, and how they are performed by agents, and
- elaborate on requirements and an approach for development in a realistic setting.

The remainder of the paper is structured as follows. Section II describes the background concepts. The design of an analytics agent is detailed in Section III. An overview of the analytics workflow is given in Section IV. Related work is covered in Section V, while applications, current relevance, and comparison with traditional analytics are in Section VI. Challenges and research gaps are discussed in Section VII, requirements and approach for development in Section VIII, evaluation methods in Section IX, and the conclusion in Section X.

II. Background

A. Large Language Models

LLMs are language models trained on large amounts of text. The property of LLMs that is of interest here is that they can read a question in natural language and write text or program code as an answer. This ability can be leveraged so that a user dictates what analysis to perform and gets the necessary SQL or Python to execute it. Surveys of agents based on LLMs highlight the LLM as the main “brain” behind the agent that informs its decision-making [5].

B. Autonomous AI Agents

An autonomous AI agent is an entity that acts toward a goal with minimal human intervention. An autonomous agent does not simply answer questions but makes decisions, uses tools, and assesses their impact while pursuing its goals until completion [4]. The key characteristics of an autonomous AI agent are planning, memory, reasoning, and tool use [5].

C. Traditional versus Agent-Based Analytics

In traditional analytics the human does most of the work and the software is reactive, answering specific questions, while with agent-based analytics the human sets high-level objectives and the agent decides what to do. This is a fundamental paradigm shift that promises to enable faster analyses and democratize analytics for less technically inclined employees, but it also introduces new challenges discussed in Section VII.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

D. Single-Agent and Multi-Agent Design

An analytics system can be designed as a single agent that does everything or a set of agents that each do a part of the work, for instance, cleaning data, building models, and generating reports. While single agents are easier to realize and control, multiple agents can divide work between themselves and review each other's work but are more complex to implement. Which design to choose depends on the specific analytics task.

III. Architecture of an Analytics Agent

There is a certain amount of common design across various LLM agents, and many studies report similar components. One such decomposition [5] is shown in Fig. 1, which is based on the four-component architecture of profile, memory, planning, and action. The figure adds the think-act-observe loop of ReAct [7].

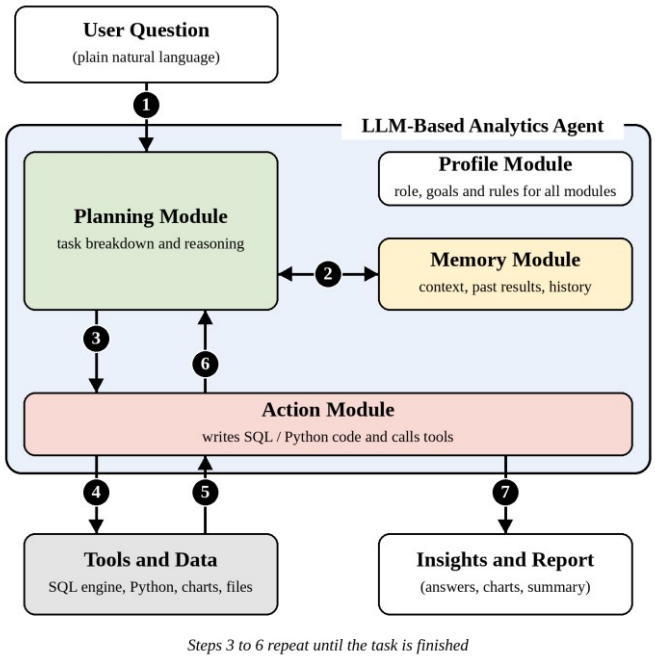


Fig. 1. Simplified architecture of an autonomous AI agent for data analytics, adapted from the module framework in [5] and the reasoning-acting loop in [7]. The numbers correspond to the processing order described in the text.

The numbers in Fig. 1 correspond to the processing order described below. First, user input is interpreted by the planning module. Second, the planner reads from and writes to the memory. Third, the planner relays instructions to the action module. Fourth, the latter performs a code or tool call. Fifth, tools return results or errors. Sixth, the actions are reported to the planner, which decides whether to continue processing or not. Steps 3–6 can be iterated as many times as needed in accordance with the ReAct loop [7]. Finally, in step 7, results are synthesized into insights and delivered to the user. The profile module is not explicitly numbered in this architecture, which reflects that it provides guidance and constraints for the other modules.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

A. LLM Core

The language model is the decision-making engine of the autonomous agent. It analyzes user queries, understands intent, and plans actions to achieve objectives. It synthesizes results into insights and reports them.

B. Profile

The profile defines the persona of the agent and instructs it on what it should and should not do and what data it has access to. It is particularly relevant for analytics to impose access and safety restrictions [5].

C. Planning

The majority of analytics queries require a multi-step approach. For instance, preparing data for analysis may involve loading the data, checking for missing values, comparing certain groups, and visualizing the results. Therefore, the planner arranges these steps in an appropriate order and instructs the action module [5].

D. Reasoning

The ability to reason enables the agent to consider intermediate results and plan the subsequent steps. The think-act-observe loop of ReAct [7] is a manifestation of this ability, which is very useful for analytics. Specifically, the model observes intermediate results, reasons about them, takes some action (e.g., advances to the next step), and updates its view of the world.

E. Memory

The agent stores input, intermediate results, and observations in memory. The memory is critical for answering follow-up questions, as otherwise the agent would forget what it did in step 3 after step 4, for instance. Another example is a follow-up question about the analysis done per region [5].

F. Actions and Tool Use

An analytics agent based on an LLM must be able to accomplish tasks such as querying databases, performing statistical analysis, and generating plots. These are generally implemented through external tools, such as an SQL database, a Python environment, and plotting software. The action module writes and executes code and reports results to the agent [5]. It is noteworthy that code execution must be done in a sandboxed environment for security.

G. Observations and Feedback Loop

The results of an action dictate further processing, which is why the code must be re-evaluated after observing results. For instance, if the agent's code failed or produced nonsensical results, it must realize this and fix the code. This is the feedback loop, and it is a defining property of an autonomous agent, manifested in steps 3–6 in Fig. 1.

IV. Automated Analytics Workflow

An analytics agent can be utilized for each of the steps in the standard analytics workflow [2], [3]. The key steps are described below.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

A. Data Collection

An agent can access relevant databases, files, spreadsheets, or other data sources and extract the necessary data to answer the question. It can formulate the necessary SQL query to extract the data, thereby saving the user from having to write it.

B. Data Cleaning

In practice, it is common to have missing or incorrect values and other issues that require additional processing. An agent can identify problems and suggest and implement a fix (e.g., write the necessary code), thereby saving a significant amount of time that would otherwise be spent on this step. At the same time, this step is error-prone, which means that mistakes made here can affect the following steps.

C. Exploratory Data Analysis

An agent can analyze the data to reveal entries, summaries, distributions, trends, groupings, and associations that are of interest for the question. It can identify which descriptive statistics and plots are needed and write the code to generate them.

D. Pattern Identification / Model Building

If applicable and relevant to the question, an agent can use an appropriate machine learning algorithm to build and evaluate a model.

E. Plotting and Report Generation

Finally, an agent can plot the obtained results and compile a report with a summary of findings relevant to the question, so that an executive summary can be formed to respond quickly to the query without having to look at the data or spend time on a query-specific dashboard.

F. Illustrative Example

As an illustrative example, consider the following query from a manager: “Which category had the biggest drop in sales in the last quarter, and what explains this decrease?” The agent would first (1) formulate an SQL query and extract the relevant data (sales per category for the last quarter). Then, the agent would (2) identify whether there are any missing or incorrect entries and remove or replace them. After that, it would perform (3) an exploratory data analysis to identify trends in sales per category over the last quarter. Using this information, the agent would generate (4) a plot that highlights the trends. Finally, the agent would compile (5) a report that states which category had the biggest drop and potential reasons for it. If the agent made a mistake at step (1), it would read the error message and adjust the SQL query accordingly. Thus, in a matter of minutes, the manager would get the answer to the question without having to write any code.

V. Literature Review

This section contains a literature review of the most relevant works related to autonomous AI agents and their role in data analytics. Table 1 below provides a summary of the reviewed articles.

Table 1: Summary of Related Work

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

Study	Method	Key finding	Limitation
Ferrag et al. (2026) [1]	Review of autonomous AI agents	Architecture, planning, memory, and tools	No practical implementation
Fikardos et al. (2026) [2]	Generative AI for data analytics	Automates data preparation, analysis, and reporting	Dependence on the data and LLM accuracy
Fu et al. (2025) [3]	Autonomous data agents for end-to-end data analytics automation	Highlights the reliability and privacy challenges	No experiments
Acharya et al. (2025) [4]	Survey of agentic AI	Describes key concepts and architecture	No experiments
Wang et al. (2024) [5]	Survey of LLM-based autonomous agents	Describes memory, planning, reasoning, and tools	No new framework
Liu et al. (2023) [6]	AgentBench evaluation of LLM-based agents	Highlights the shortcomings of agents in reasoning and planning	Focuses on evaluation
Yao et al. (2023) [7]	ReAct: reasoning plus action for autonomous agents	Improves task solving with tools	Depends on the LLM

Ferrag et al. [1] provide an overview of the evolution of reasoning in LLMs and outline the architecture, planning, memory, and tools for autonomous agents. The article reviews the current state of the practice without suggesting any new implementations. Fikardos et al. [2] highlight the potential of generative AI for automating data analytics, from data preparation to reporting, while noting the dependence on data and LLM accuracy. Fu et al. [3] describe the concept of autonomous data agents for end-to-end data analytics automation and note reliability and privacy challenges without experimenting with such agents. Acharya et al. [4] provide an overview of the autonomous agent field, covering key concepts and architecture, without experimenting. Wang et al. [5] summarize the latest developments in LLM-based agents, including memory, planning, reasoning, and tools, without suggesting a new framework. Liu et al. [6] introduce AgentBench to evaluate the ability of LLM-based agents and highlight their shortcomings in planning and reasoning. Finally, Yao et al. [7] propose ReAct, a novel approach to task solving with reasoning and action, which improves the ability of agents to use tools but still depends on the LLM.

Overall, the reviewed articles highlight the main themes and aspects of autonomous agents, with some providing excellent overviews of the field. At the same time, most of the reviewed works focus on surveys and evaluations rather than practical implementations, highlighting the need for this study.

VI. Applications and Current Relevance

A. Current Relevance

Autonomous AI agents are critically important in addressing a set of closely related challenges, including

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

- the need to adopt autonomous AI agents for a variety of business processes and analytical tasks,
- recent advances in language models that enable the development of highly capable autonomous agents,
- the growing need for automated analytics,
- the emergence of agentic AI, and
- the connections between autonomous AI agents and business intelligence.

These factors, along with the timing of this research, provide a strong foundation for this work.

B. Potential Areas of Application

There are numerous potential applications for such an agent. The unifying feature of these areas is that they involve substantial data and require little analytical expertise. For example, the following areas could benefit from an autonomous agent:

- business reporting: various questions related to sales, marketing, and operations,
- financial transaction analysis,
- medical reports or patterns,
- customer, market, and retail analytics,
- student and research assistance in education.

C. Traditional versus Agent-Based Analytics

Table 2 below compares traditional and agent-based analytics. Some of the advantages described here are explored in more detail in later sections.

Table 2: Traditional versus Agent-Based Analytics

Aspect	Traditional	Agent-Based
Question asking	Requires writing SQL or Python code	Asking in natural language
User expertise	Technical skills	Minimal
Speed	Manual, slow	Steps automated
Flexibility	Predefined dashboards	On-demand
Workflow	Human-planned	Agent-planned
Primary risks	Human errors	Wrong code, LLM issues

VII. Challenges and Research Gaps

Based on the literature review, the following research gaps were identified.

A. Data Cleaning Errors

There can be various errors during the cleaning process, which can affect later steps as well as the final results.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

B. Limited Autonomy

Typically, an autonomous agent is only fully autonomous in specific areas but requires human assistance in others.

C. Data Privacy and Security

The privacy and security of data in analytics-based AI are inadequate, which is particularly problematic in corporate settings [3].

D. Multimodal Analysis

The analysis of natural language, tabular data, and other elements, such as charts and graphs, is challenging.

E. Limited Advanced Analytical Capabilities

The capabilities of such systems in advanced analytical tools are limited, which is dictated by the mathematical and statistical abilities of the underlying LLM. The review highlights these limitations in current LLMs.

F. Hallucination and Trustworthiness

An autonomous agent based on an LLM can make mistakes when writing code or statements. This is particularly dangerous in analytics, as it can lead to incorrect results that can be disastrous for businesses that use these results to make decisions; therefore, it is critical to develop methods to verify the results of such an agent [2], [3].

G. Explainability and Accountability

For an autonomous agent to be useful, its actions must be explainable and verifiable, which can be achieved, for example, by giving access to the code that the agent writes to verify that it does not contain errors. Similarly, if the initial data was biased, this will affect the analytics results, which also must be validated.

VIII. Proposed Approach and Requirements

A. Requirements for Autonomous Analytics System

The design of an autonomous analytics system that would satisfy the requirements for practical application is seen as a goal for future research. Based on the analysis of such a system's architecture and the description of research gaps, the following requirements can be proposed:

1) **Language model:** a competent large language model that can plan and write code;

2) **Secure connections to databases:** secure connections to external databases with the possibility of managing access rights;

3) **Sandboxing:** isolating the model from the rest of the system when executing code;

4) **Context and memory:** storing context and previous results for subsequent use when interacting with the user;

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

5) **Validations:** validating the various stages, including data cleaning, and results;

6) **Human-in-loop:** the option for a human to evaluate and validate each step performed by an autonomous agent;

7) **Logging:** logging all operations for subsequent analysis and validation.

B. Proposed Implementation Roadmap

An autonomous analytics system can be implemented in stages as follows.

1) Define the use case for developing the system, as well as the data sources and users.

2) Select the language model and agent framework.

3) Develop the tool layer, including access to databases via SQL, execution of Python code, and data science and visualization libraries.

4) Implement planning capabilities, including reasoning and context management, to perform multi-step and multi-turn analytics.

5) Ensure that the system has the necessary safety features, including sandboxing, access control, validation, and logging.

6) Test the system on a sample data set with the validation tools described in Section IX.

7) Release the system to users, ensuring that they have an option to review and approve each critical step, and continuously improve the system based on user feedback.

IX. Evaluation of Analytics Agents

Analytics agents need to be evaluated to establish their usefulness and safety. While generic LLMs as autonomous agents have been benchmarked in AgentBench [6], analytics agents benefit from additional metrics:

- **Task success rate:** the percentage of tasks performed by the agent that were completed successfully according to the user query.
- **Code correctness:** whether the code executed successfully and produced the expected results.
- **Accuracy of insights:** similarity of results to those of a human expert.
- **Accuracy of data cleaning:** similarity to expert data cleaning.
- **Time and cost:** the amount of time saved compared to manual analysis, and computational costs.
- **Reliability and privacy:** consistency of results between different runs of the same task and the ability to keep data private.

X. Conclusion, Limitations and Future Scope

The paper explored the concept of autonomous AI agents for the purpose of automated data analytics. It outlined the critical aspects of such agents, including the LLM core, planning and reasoning, memory,

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

tool usage, and feedback loop. The paper demonstrated the potential of these agents in data analytics, including addressing the data pipeline from collection, preparation, and EDA to model building and report generation. A literature review of seven recent articles identified opportunities for improvement in the area of data cleaning, autonomy, data security and privacy, multimodal analysis, and advanced analytics capabilities.

Future research in this regard can be focused on prototyping an analytics agent with strong privacy guarantees, addressing data cleaning validation, and utilizing text, table, and vision analysis for enhanced capabilities. Other promising areas of research include the development of multi-agent approaches, a stronger emphasis on statistical and mathematical analysis, tighter integration with business intelligence tools, and the creation of standardized benchmarks for analytics agents.

The limitation of this paper is that it focused on seven recent articles, which might not represent the overall state of research on AI analytics agents. Another limitation is that this paper is a survey, and thus it did not include an implementation and evaluation section, meaning that the benefits of such an agent still need to be experimentally validated. Future research should therefore address these limitations by investigating more analytics agents and publishing implementation details and experiments that build upon the findings of this paper.

References

[1] M. A. Ferrag, N. Tihanyi, and M. Debbah, “From LLM reasoning to autonomous AI agents: A comprehensive review,” *IEEE Access*, vol. 14, pp. 84237–84286, 2026, doi: 10.1109/ACCESS.2026.3698694.

[2] M. Fikardos, K. Lepenioti, A. Bousdekis, D. Apostolou, and G. Mentzas, “Generative AI for autonomous data analytics,” *Intelligent Systems with Applications*, vol. 29, p. 200626, 2026, doi: 10.1016/j.iswa.2026.200626.

[3] Y. Fu, D. Wang, W. Ying, X. Zhang, H. Liu, and J. Pei, “Autonomous data agents: A new opportunity for smart data,” *arXiv preprint* arXiv:2509.18710, 2025.

[4] D. B. Acharya, K. Kuppan, and B. Divya, “Agentic AI: Autonomous intelligence for complex goals—A comprehensive survey,” *IEEE Access*, vol. 13, pp. 18912–18936, 2025, doi: 10.1109/ACCESS.2025.3532853.

[5] L. Wang et al., “A survey on large language model based autonomous agents,” *Frontiers of Computer Science*, vol. 18, no. 6, p. 186345, 2024, doi: 10.1007/s11704-024-40231-1.

[6] X. Liu et al., “AgentBench: Evaluating LLMs as agents,” *arXiv preprint* arXiv:2308.03688, 2023.

[7] S. Yao, J. Zhao, D. Yu, N. Du, I. Shafran, K. R. Narasimhan, and Y. Cao, “ReAct: Synergizing reasoning and acting in language models,” in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2023.