IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, October 2026
Impact Factor: 8.2

Data Science and Machine Learning: Related Fields, Different Roles — A Conceptual Review of Their Distinctions, Intersections and Applications

Dandasi Pavan Kumar[1], Somala Vinesh Kumar[2], D. V. S. Vineela[3]
Assistant Professor, Sir C. R. Reddy College, Eluru, Andhra Pradesh, India[1,2]
MCA Scholar, Sir C. R. Reddy College, Eluru, Andhra Pradesh, India[3]

Abstract: *Data science and machine learning are closely related fields that are frequently used interchangeably in academic discussions, professional settings and public discourse. However, they differ in their objectives, methodologies and practical applications. Data science is a multidisciplinary field concerned with collecting, processing, analysing and interpreting data to generate meaningful insights, whereas machine learning is a subfield of artificial intelligence that enables computer systems to identify patterns and make predictions or decisions based on data. This conceptual review examines the definitions, foundations, workflows and professional applications of data science and machine learning. It explores their relationship within the broader field of artificial intelligence and explains how machine learning contributes to specific stages of the data science process. The paper also discusses the changing boundaries between the two fields in the context of automated machine learning, computational advances and machine learning system design. Drawing on selected academic and industry literature, the study argues that effective data-driven systems require more than sophisticated algorithms. Data quality, statistical reasoning, domain knowledge, responsible interpretation and deployment are equally important. The paper concludes that understanding the distinctions and interconnections between data science and machine learning is essential for students, educators, researchers and organisations seeking to develop reliable and useful data-driven solutions.*

Keywords: Data Science, Machine Learning, Artificial Intelligence, Data Analytics, Predictive Modelling, Automated Machine Learning, Data-Driven Decision-Making.

I. INTRODUCTION

The growing availability of digital data and advances in computational technologies have increased the importance of data science and machine learning across education, healthcare, finance, business and public administration. Both fields contribute to extracting information from data and supporting evidence-based decision-making. Their increasing presence in academic curricula, professional job descriptions and discussions about artificial intelligence has, however, created confusion regarding their meanings and respective functions.

Data science and machine learning are closely connected, but they are not identical. Data science encompasses a broad process of understanding problems, acquiring data, preparing datasets, conducting statistical analysis, developing models and communicating findings. Machine learning focuses on developing computational methods through which systems learn patterns from data and use them to make predictions, classifications or decisions.

The distinction is important because the successful application of technology depends on selecting appropriate methods for clearly defined problems. Treating machine learning as equivalent to the entire data science process may lead to an excessive emphasis on algorithms while overlooking data quality, research design, statistical interpretation and practical implementation.

Copyright to IJARSCT
www.ijarsct.co.in
DOI: 10.48175/568

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 5, October 2026 Impact Factor: 8.2

Recent discussions in the literature have also highlighted the consequences of treating artificial intelligence, machine learning and data science as interchangeable concepts. Such confusion can influence public expectations, organisational investments and policy decisions (Digital Society, 2026; IBM, n.d.). Against this background, the present paper examines the conceptual differences and practical relationships between data science and machine learning. It also considers their overlapping professional roles and the importance of data quality, system design and deployment in developing reliable data-driven applications.

1.1 Objectives of the Study

The objectives of this paper are:

1. To explain the fundamental concepts and scope of data science and machine learning.
2. To identify the major differences between their objectives, methods and applications.
3. To examine how machine learning fits within the broader data science workflow.
4. To discuss the changing professional roles associated with both fields.
5. To highlight the importance of data quality, system design and responsible implementation.

II. RESEARCH METHODOLOGY

This paper adopts a conceptual review approach based on selected academic publications, technical literature and industry resources concerning data science, machine learning and artificial intelligence. The literature considered includes discussions of the conceptual relationship between these fields, descriptions of data science workflows, and research on machine learning system design and automation.

The selected materials were examined thematically under four broad categories: conceptual definitions, analytical workflows, professional applications, and system design and deployment. The discussion synthesises these themes to explain the similarities and differences between data science and machine learning.

The paper is intended as a conceptual synthesis rather than a systematic literature review. It does not claim to provide an exhaustive search of all relevant publications or a quantitative assessment of research findings.

III. CONCEPTUAL FOUNDATIONS OF DATA SCIENCE AND MACHINE LEARNING

3.1 Artificial Intelligence as the Broader Field

Artificial intelligence (AI) refers broadly to computational systems designed to perform tasks that typically require human-like capabilities, such as reasoning, language processing, perception, planning and decision-making.

Machine learning is one of the major approaches used to develop artificial intelligence systems. Instead of relying entirely on manually specified rules, machine learning methods identify patterns in data and use them to perform particular tasks.

Data science has a different emphasis. It combines statistical reasoning, computing, analytical methods and subject-specific knowledge to extract useful information from data. Although it frequently uses machine learning, it also includes methods that do not require machine learning algorithms.

The relationship among these fields is therefore better understood through their functions and areas of overlap than through treating them as synonyms.

3.2 The Emergence of Data Science

Data science developed through the convergence of statistics, computer science, data management and applied research. Improvements in computing power, digital storage and data availability expanded the possibilities for analysing large and complex datasets.

Cleveland's (2001) proposal for expanding the field of statistics provides an important reference point in discussions about the formal development of data science. The field subsequently evolved to include data engineering, visualisation, predictive modelling and the communication of analytical findings.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, October 2026
Impact Factor: 8.2

Its central purpose is not simply to process large volumes of data but to use appropriate methods to answer questions, explain patterns and support decisions.

IV. UNDERSTANDING DATA SCIENCE

Data science is a multidisciplinary process involving the collection, preparation, analysis, interpretation and communication of data. According to IBM (n.d.), it combines methods from statistics, programming, analytics, modelling and related disciplines to generate insights and address practical problems.

A data science project typically begins with a question or problem. For example, a business may seek to understand why customers stop using its services, while a healthcare organisation may want to identify factors associated with patient risk. The data scientist must determine what information is required, assess its quality, select suitable analytical techniques and interpret the results in the relevant context.

4.1 Major Components of Data Science

Problem definition: Establishing the research question, objectives and expected outcomes.

Data collection: Obtaining relevant information from databases, surveys, digital platforms, administrative records or other sources.

Data preparation: Correcting errors, addressing missing values, removing duplicates and transforming information into a suitable format.

Exploratory data analysis: Examining distributions, relationships, trends and unusual observations through statistical methods and visualisation.

Modelling and analysis: Applying statistical techniques or machine learning algorithms where appropriate to explain patterns or make predictions.

Interpretation and communication: Presenting findings in a form that supports understanding and decision-making.

Evaluation and monitoring: Assessing the reliability of results and determining whether analytical solutions remain useful as circumstances change.

These activities demonstrate that data science extends well beyond the development of predictive models.

4.2 Importance of Data Quality and Context

The reliability of data science findings depends on the quality, relevance and limitations of the available information. Missing observations, inconsistent measurements, sampling bias and inaccurate records can affect analytical results.

Domain knowledge is equally important. A statistical relationship may not establish causation, and a model may produce misleading results when the data does not represent the population or conditions to which the findings are applied.

Consequently, data science requires critical interpretation, reproducible analytical procedures and an understanding of the social, economic or organisational context in which the data was generated.

V. UNDERSTANDING MACHINE LEARNING

Machine learning is a subfield of artificial intelligence that develops algorithms capable of learning patterns from data. These algorithms use the patterns to classify information, estimate outcomes, detect anomalies or support decisions without requiring programmers to specify every decision rule manually.

A familiar example is an email spam filter. Rather than relying exclusively on a fixed list of suspicious words, a machine learning system can learn from examples of spam and legitimate messages to identify patterns associated with each category.

5.1 Major Types of Machine Learning

Supervised learning: Algorithms learn from labelled examples to predict an outcome or classify new observations. Applications include demand forecasting, disease-risk prediction and spam detection.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, October 2026
Impact Factor: 8.2

Unsupervised learning: Algorithms identify patterns or groupings in data without relying on predefined outcome labels. Applications include customer segmentation and exploratory pattern discovery.

Reinforcement learning: Systems learn through interaction with an environment, using feedback or rewards to improve their actions. Applications include selected robotics, game-playing and control problems.

The choice of method depends on the nature of the problem, the available data and the intended outcome.

5.2 Machine Learning Applications

Machine learning is used in several sectors. In healthcare, it can assist with image analysis and risk prediction. In finance, it can support fraud detection and credit-risk assessment. In business, it can contribute to demand forecasting, recommendation systems and customer analysis.

However, the performance of a machine learning model depends on factors such as training data quality, model selection, evaluation methods and the conditions under which it is deployed. A model that performs well on historical data may not maintain the same performance when patterns or operating conditions change.

Machine learning should therefore be understood as a powerful analytical tool rather than an independent solution to every data-related problem.

VI. RELATIONSHIP BETWEEN DATA SCIENCE AND MACHINE LEARNING

Data science and machine learning overlap substantially, particularly when a project involves predictive modelling. Nevertheless, they serve different functions within a complete analytical process.

A typical data-driven project may begin with data acquisition and integration, followed by cleaning, exploration and visualisation. Machine learning may then be applied to develop a predictive or classification model. The model must subsequently be evaluated, deployed and monitored.

Research on automating end-to-end data science highlights the interconnected nature of these activities and the role of automation in selected stages of the workflow (How Can AI Automate End-to-End Data Science?, 2019).

Machine learning is therefore one component of a wider process. Data acquisition, data governance, exploratory analysis, research design and communication remain essential, whether or not a machine learning algorithm is used.

6.1 Key Differences

Data science has a broad scope and aims to generate insights and address questions using data. Machine learning is more specialised and focuses on learning patterns from data to make predictions or decisions. Data science uses statistics, visualisation, data engineering, analytics and modelling; machine learning focuses on learning algorithms, model training and evaluation. Data science may produce reports, insights, visualisations, statistical findings and models, while machine learning typically produces trained models, predictions, classifications and learned patterns. These roles overlap, and their boundaries vary across projects.

VII. PROFESSIONAL ROLES AND CAREER DIFFERENCES

The overlap between data science and machine learning is reflected in professional job descriptions. In smaller organisations, a single professional may undertake data preparation, analysis, model development and deployment. In larger teams, these responsibilities may be distributed among specialists.

7.1 Data Scientist

A data scientist generally works across the analytical lifecycle. Responsibilities may include defining research questions, collecting and preparing data, conducting exploratory analysis, applying statistical methods, building models and communicating findings to technical and non-technical audiences.

The role requires a combination of analytical reasoning, programming, statistical knowledge and communication skills. Understanding the context of a problem is particularly important when interpreting results and recommending action.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, October 2026
Impact Factor: 8.2

7.2 Machine Learning Engineer

A machine learning engineer typically concentrates on developing, testing, deploying and maintaining machine learning systems. Responsibilities may include training and tuning models, building data and model pipelines, improving computational efficiency, integrating models into software applications and monitoring performance after deployment.

The role often requires programming, knowledge of machine learning frameworks, software engineering practices and an understanding of production systems.

7.3 Areas of Overlap

Both roles may involve data preparation, programming, model evaluation and experimentation. Their principal difference is one of emphasis rather than a strict separation of tasks.

Data scientists generally focus on the broader analytical question and the insights that can be derived from data. Machine learning engineers generally focus on building and maintaining reliable systems that use learned models.

These distinctions are not universal. Actual responsibilities vary according to the organisation, project requirements and the experience of the professionals involved.

VIII. AUTOMATION AND THE CHANGING BOUNDARIES OF THE FIELDS

Automated machine learning (AutoML) has expanded the range of technical tasks that can be performed with limited manual intervention. Depending on the platform, AutoML tools may automate model selection, hyperparameter tuning, feature processing and the combination of multiple models.

Such tools can reduce repetitive work and make certain modelling techniques more accessible. However, automation does not eliminate the need to formulate appropriate questions, assess data quality, select meaningful evaluation criteria or determine whether a model's outputs are suitable for practical use.

Advances in specialised computing hardware, including graphics processing units (GPUs) and tensor processing units (TPUs), have also improved the ability to train computationally demanding models. These advances have expanded the range of machine learning applications but have not removed the fundamental importance of reliable data and careful evaluation.

The increasing availability of automated tools makes a clear understanding of the data science process more important, not less. Professionals must be able to assess the assumptions, limitations and practical consequences of automated analytical outputs.

IX. MACHINE LEARNING SYSTEMS DESIGN: DATA, DESIGN AND DEPLOYMENT

Lawrence (2019) proposes a systems-oriented perspective on machine learning organised around three elements: Data, Design and Deployment. This perspective emphasises that successful machine learning applications depend on the relationship between the data, the design of the system and the conditions in which it operates.

9.1 Data

The first element involves understanding where data originates, how it was collected, what it represents and which limitations it contains. Data quality and representativeness influence the reliability of subsequent analysis and modelling.

9.2 Design

The design stage involves selecting appropriate analytical approaches, defining evaluation criteria and developing a system that addresses the intended problem. A technically sophisticated algorithm may be unsuitable if it does not match the data or the requirements of the application.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, October 2026
Impact Factor: 8.2

9.3 Deployment

Deployment concerns integrating a model into a practical environment and maintaining its performance over time. Models may become less reliable when user behaviour, environmental conditions or underlying data patterns change. Monitoring, maintenance and periodic evaluation are therefore necessary.

This perspective challenges the assumption that selecting a more advanced algorithm is sufficient to create a successful machine learning application. The entire system must be designed around the problem it is intended to address.

X. DISCUSSION AND IMPLICATIONS

The conceptual distinction between data science and machine learning has implications for education, organisational decision-making and public understanding of artificial intelligence.

10.1 Implications for Education

Educational programmes should provide students with a balanced understanding of statistics, programming, data management, research design and machine learning. An exclusive emphasis on algorithm development may leave learners insufficiently prepared to identify suitable questions, evaluate data limitations or communicate findings.

Students should learn to examine the origin and quality of data before applying analytical methods. They should also understand that predictive accuracy alone does not establish whether a model is fair, interpretable or appropriate for a particular application.

10.2 Implications for Organisations

Organisations should define their objectives before investing in machine learning technologies. Some problems can be addressed effectively through descriptive statistics, visualisation or conventional statistical analysis. Others may benefit from predictive modelling or more advanced machine learning techniques.

Choosing a method that matches the problem can help organisations avoid unnecessary complexity and improve the usefulness of analytical results. Investment in data governance, infrastructure, skilled personnel and system maintenance is also important.

10.3 Implications for Responsible AI

Confusion between data science, machine learning and artificial intelligence can lead to unrealistic expectations about what technological systems can achieve. It can also obscure the different risks associated with data collection, statistical inference, model development and automated decision-making.

Responsible implementation requires attention to privacy, data representativeness, transparency, accountability and the consequences of incorrect predictions. Human oversight and appropriate evaluation remain important, particularly in applications that may affect people's rights, opportunities or access to essential services.

10.4 Future Research Directions

Further research could examine how data science and machine learning responsibilities vary across industries, educational programmes and organisational settings. Empirical studies could also investigate the impact of AutoML tools on professional roles, the skills required by graduates and the reliability of automated analytical workflows.

Comparative studies of deployed systems could help identify which combinations of data preparation, model design and monitoring practices contribute to more dependable outcomes.

XI. CONCLUSION

Data science and machine learning are closely connected but conceptually distinct fields. Data science provides a broad framework for defining problems, collecting and preparing information, analysing patterns and communicating findings. Machine learning contributes methods that enable computational systems to learn from data and make predictions or classifications.

Their relationship is complementary rather than interchangeable. Machine learning can form an important part of a data science project, but it cannot independently replace sound research design, reliable data, statistical reasoning or

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, October 2026
Impact Factor: 8.2

contextual interpretation. Similarly, data science can generate valuable insights without necessarily using machine learning.

The growing use of automated analytical tools makes these distinctions increasingly relevant to students, educators, researchers and organisations. A balanced understanding of both fields can support better educational preparation, more appropriate technology choices and more responsible applications of data-driven systems.

Ultimately, effective data-driven solutions depend not only on the sophistication of an algorithm but also on the quality of the data, the clarity of the problem, the suitability of the methods and the reliability of implementation.

REFERENCES

1. Cleveland, W. S. (2001). Data science: An action plan for expanding the technical areas of the field of statistics. International Statistical Review, 69(1), 21–26. https://doi.org/10.1111/j.1751-5823.2001.tb00477.x
2. GeeksforGeeks. (2024). What is the role of machine learning in data science? GeeksforGeeks. [Insert the verified article URL.]
3. IBM. (n.d.). Data science vs. machine learning: What's the difference? IBM Think. https://www.ibm.com/think/topics/data-science-vs-machine-learning
4. Lawrence, N. D. (2019). Data science and digital systems: The 3Ds of machine learning systems design. arXiv preprint arXiv:1903.11241. [Verify the exact title and publication details against the original record.]
5. How can AI automate end-to-end data science? (2019). arXiv preprint arXiv:1910.14436. [Verify the authors, exact title and publication details against the original record.]
6. Sharma, M. (2025). The role of machine learning in enhancing data science workflows. International Journal of Innovative Science and Engineering Management, 4(3). [Verify the journal title, page numbers, DOI and publication details.]
7. Are artificial intelligence, machine learning, and data science all the same? The public sure thinks so, and why that's problematic. (2026). Digital Society. [Add the authors' names, volume, issue, page range or article number, and DOI after verification.]