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 6, October 2026
Impact Factor: 8.2

A Comprehensive Overview and Analysis of Large Language Models: Trends and Challenges

More Govind Sandipan[1], Prof. P. A. Waghule[2], Prof. L. S. Bhadake[3], Dr. H. B. Jadhav[4]
Student, Department of Computer Engineering[1]
Guide, Department of Computer Engineering[2]
Co-Guide, Department of Computer Engineering[3]
Department of Computer Engineering, Adsul Technical Campus, Chas, Ahilyanagar, India

Abstract: *Large Language Models (LLMs) have become an important part of modern artificial intelligence because they can interpret and generate human language across many tasks. Built mainly on transformer-based neural networks and trained on large text collections, these models support activities such as summarisation, translation, question answering, coding assistance, and conversational interaction. This review examines the foundations of LLMs, their development, major application areas, recent trends, and unresolved challenges. It discusses pre-training, instruction tuning, retrieval-augmented generation, multimodal learning, and model adaptation, while considering concerns involving hallucinations, bias, privacy, security, computational cost, transparency, and evaluation. The review finds that LLMs can improve access to information and assist knowledge work, but their output should not automatically be treated as correct or unbiased. Responsible use requires careful testing, human oversight, appropriate data governance, and evaluation suited to each application. Future progress will depend not only on larger models, but also on more efficient systems, better grounding, clearer accountability, and stronger methods for measuring reliability.*

Keywords: Large Language Models, Artificial Intelligence, Natural Language Processing, Transformer Architecture, Generative AI, Retrieval-Augmented Generation, Model Evaluation, Responsible AI.

I. INTRODUCTION

Language is one of the main ways people share knowledge, ask questions, and solve problems. For a long time, computers handled language through rules, templates, and task-specific statistical methods. Recent progress in deep learning has changed this approach. Large Language Models (LLMs) learn patterns from very large text datasets and use those patterns to produce context-sensitive responses. Their ability to work across multiple language tasks has made them useful in education, software development, customer support, research, and content creation.

Most current LLMs build on the transformer architecture, which uses attention mechanisms to model relationships between tokens in a sequence. During pre-training, a model learns statistical relationships from large amounts of data. Additional stages, such as supervised fine-tuning and preference-based alignment, can adapt the model to follow instructions and respond in a more useful manner. These steps improve practical performance, but they do not guarantee that every answer is factual, safe, or appropriate.

The rapid adoption of LLMs has created opportunities as well as difficult questions. A model may produce fluent text that contains incorrect details, reflect harmful patterns present in training data, or expose information if a system is poorly designed. Running large models can also require substantial computing resources and energy. This paper reviews the central ideas behind LLMs, the trends shaping their development, and the technical and social challenges that must be addressed for reliable use.

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 6, October 2026
Impact Factor: 8.2

II. PROBLEM STATEMENT

LLMs are increasingly used to generate and interpret information, yet their performance varies with the task, prompt, training data, and deployment environment. A response may sound convincing while containing unsupported claims, outdated facts, biased language, or unsafe recommendations. At the same time, organisations face practical barriers related to inference cost, latency, privacy, copyright, security, and limited explainability. The main problem addressed in this review is the gap between the broad capabilities of LLMs and the level of reliability, transparency, efficiency, and accountability required for real-world use. A structured overview is therefore needed to understand current methods, identify persistent limitations, and outline directions for responsible development.

III. REVIEW OBJECTIVES

To explain the basic concepts and development of Large Language Models, including transformer-based language modelling.

To review common training and adaptation approaches, including pre-training, instruction tuning, fine-tuning, and retrieval-augmented generation.

To summarise major application areas across education, healthcare support, software engineering, business, and research.

To identify current trends such as multimodal models, smaller efficient models, tool use, and domain-specific adaptation.

To analyse challenges involving hallucination, bias, privacy, security, evaluation, computational cost, and governance.

To outline practical recommendations and future research directions for trustworthy LLM deployment.

IV. REVIEW METHODOLOGY

This paper uses a narrative review approach to organise established research and technical discussions about LLMs. The review themes include model architecture, training methods, alignment, grounding, application areas, evaluation, and responsible deployment. Foundational work on transformers and language-model scaling is considered alongside research on instruction following, retrieval-based systems, model evaluation, and risk management. The material is compared conceptually rather than through a new experimental benchmark. As a result, this paper does not claim to train a new model or report original performance measurements. Reported results from individual studies should be interpreted in the context of their datasets, tasks, and evaluation settings.

V. FOUNDATIONS OF LARGE LANGUAGE MODELS

A. Transformer Architecture

The transformer processes sequences through attention layers that help the model relate different tokens to one another. Text is first divided into tokens, which are represented numerically. The model then repeatedly transforms these representations to learn contextual patterns. Unlike earlier sequence models that process information strictly step by step, transformer attention can connect distant parts of an input efficiently during training. The architecture is a major reason modern language models can be trained at large scale, although model size alone does not ensure reasoning ability or factual reliability.

B. Pre-training and Adaptation

During pre-training, models commonly learn to predict missing or subsequent tokens from examples in a large corpus. This process develops broad language representations. A pre-trained model can then be adapted through supervised fine-tuning on task examples, instruction tuning, or preference optimisation. Fine-tuning can make a model more suitable for a domain or style, but it may also introduce overfitting or reduce some general capabilities if training is poorly balanced.

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 6, October 2026
Impact Factor: 8.2

C. Prompting and In-Context Learning

Prompting provides instructions, context, constraints, and examples at inference time without necessarily changing the model's parameters. Clear prompts can improve consistency, but prompt wording is not a substitute for verification. In-context examples may guide format and task behaviour, while complex tasks often benefit from breaking work into smaller steps and checking intermediate results.

D. Retrieval-Augmented Generation

Retrieval-Augmented Generation (RAG) combines a language model with a search or retrieval component. Relevant documents are retrieved from an external collection and supplied as context for the answer. This can help a system use current or organisation-specific information and make responses easier to trace to sources. However, RAG does not eliminate errors: retrieval may miss the best document, sources may conflict, and the model may misinterpret retrieved content. Source quality and retrieval evaluation remain essential.

VI. MAJOR APPLICATION AREAS

A. Education and Learning Support

LLMs can explain concepts at different levels, generate practice questions, provide writing feedback, and support language learning.

They can help students explore a topic, but answers may be inaccurate and should be checked against course materials. Educators also need clear rules for disclosure, assessment integrity, and appropriate use.

B. Healthcare and Scientific Research

In healthcare settings, LLMs may help summarise documents, draft patient-friendly explanations, or support administrative workflows. They should not replace qualified clinical judgement, and any system handling health information must meet strict privacy and validation requirements. In research, models can assist with literature exploration and drafting, but generated references, quotations, and scientific claims require independent verification.

C. Software Development

LLMs can explain code, suggest implementations, create tests, and help developers understand unfamiliar libraries. Their suggestions may contain security weaknesses, incorrect assumptions, or outdated APIs. Code generated by a model should be reviewed, tested, and checked for licensing and security implications before use.

D. Business and Public Services

Organisations use LLM-based assistants for document search, customer communication, summarisation, and workflow support.

Benefits depend on integration quality, access controls, and the accuracy of the information supplied. Systems that influence employment, finance, public benefits, or other high-impact decisions require heightened oversight and clear routes for human review.

VII. CURRENT TRENDS

A. Multimodal Models

Many newer systems process more than text, combining language with images, audio, video, or structured inputs. Multimodal capability can support accessibility, document understanding, and richer human-computer interaction. It also expands the range of possible errors and privacy risks, so evaluation must cover each supported input and output type.

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 6, October 2026
Impact Factor: 8.2

B. Smaller and More Efficient Models

Quantisation, pruning, distillation, and parameter-efficient adaptation can reduce memory or computation requirements. Smaller models may be easier to deploy locally or on specialised hardware, offering potential benefits for latency and data control. Their performance still depends on the task and available resources, and efficiency claims should be measured on realistic workloads.

C. Tool Use and Agentic Workflows

LLM systems increasingly connect to calculators, databases, search engines, code interpreters, and business applications. Tool use can ground some responses in external results and enable multi-step workflows. Yet a tool-using system may select the wrong tool, misread results, or take an unintended action. Permission boundaries, confirmation steps, logging, and limited access are important safeguards.

D. Domain Adaptation and RAG

Specialised deployments combine general language models with domain-specific documents, terminology, and evaluation sets.

Retrieval-based grounding can reduce dependence on information stored only in model parameters. Organisations must still keep knowledge sources updated, protect sensitive documents, and test performance on representative queries.

VIII. KEY CHALLENGES

A. Hallucination and Factual Reliability

A language model generates likely sequences rather than directly checking every statement against reality. It can therefore invent facts, citations, or explanations, especially when a question is ambiguous or outside its knowledge. Grounding, calibrated uncertainty, source citations, and independent verification can reduce risk, but no single method guarantees correctness.

High-impact uses need human review and clear limits on what the system is allowed to decide.

B. Bias and Fairness

Training data can contain stereotypes, under-represented languages, and unequal coverage of social groups. Models may reproduce these patterns or perform differently across dialects and communities. Fairness assessment should use diverse test cases, document known limitations, and involve affected users where appropriate. Filtering alone cannot resolve every source of bias.

C. Privacy and Data Protection

Prompts and uploaded documents may contain personal, confidential, or proprietary information. Poorly designed systems can expose such material through logs, access-control failures, or unintended memorisation. Data minimisation, clear retention policies, encryption, access controls, and appropriate consent are important. Users should understand what data is processed and whether it may be used to improve a service.

D. Security and Misuse

LLM applications can be targeted by prompt injection, data exfiltration attempts, malicious documents, and unsafe tool requests.

Models can also be misused to generate deceptive or harmful content. Security needs to be designed across the full application rather than delegated to the model alone. Sandboxed tools, least-privilege access, input and output checks, monitoring, and red-team testing can reduce exposure.

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 6, October 2026
Impact Factor: 8.2

E. Cost, Energy, and Access

Training and serving large models may require specialised hardware, electricity, and substantial financial resources. These costs can concentrate capability among a small number of organisations and limit access for smaller institutions. Efficient architectures, caching, quantisation, model selection, and transparent reporting of resource use can help. Evaluations should consider quality alongside latency, cost, and energy consumption.

F. Evaluation and Transparency

A model may perform well on a benchmark yet fail on unfamiliar or real-world tasks. Benchmark contamination, changing model versions, prompt sensitivity, and differences in scoring can make comparisons difficult. Evaluation should use task-relevant datasets, test for robustness and safety, report uncertainty, and include human assessment when needed. Documentation about training data, limitations, intended use, and updates improves accountability.

IX. COMPARISON OF APPROACHES

Approach	Main Strength	Important Limitation
Prompting	Fast to try; no parameter updates needed	Sensitive to wording; limited control
Fine-tuning	Can adapt style or task behaviour	Needs suitable data and careful testing
RAG (Retrieval-Augmented Generation)	Uses external, potentially current sources	Retrieval and source quality can fail
Tool Use	Can perform calculations or access systems	Requires strict permissions and monitoring
Smaller/Quantised Models	Lower resource and deployment cost	May lose capability on difficult tasks

No single approach is best for every application. Prompting is useful for rapid experimentation, fine-tuning for adapting model behaviour, RAG for grounding answers in a knowledge base, and tools for actions that require external computation or systems. A reliable design often combines methods and evaluates the complete application rather than the language model in isolation.

X. RESPONSIBLE DEPLOYMENT FRAMEWORK

A practical deployment process should begin with a clear definition of the task, intended users, expected benefits, and foreseeable harms. Teams should select a model based on task performance, privacy requirements, cost, latency, and maintenance needs. Before release, testing should cover factuality, robustness, bias, security, refusal behaviour, and performance across relevant user groups.

After release, the system should be monitored for failures, misuse, changing data, and user feedback. Important outputs should be traceable to evidence where possible, and users should be told when they are interacting with AI. High-impact decisions need meaningful human oversight, a way to challenge errors, and a plan to disable or roll back unsafe functionality. Governance is not a one-time checklist; it must continue throughout the system's lifecycle.

XI. FUTURE SCOPE

Future research should focus on models that are more dependable, efficient, and easier to evaluate. Promising directions include improved grounding and citation verification, better uncertainty estimation, robust multilingual performance, privacy-preserving adaptation, and stronger defences against prompt injection. Research into compact models may make advanced language tools more accessible on personal devices and within local environments.

Another important direction is evaluation that reflects real use. Benchmarks should be refreshed, tested against contamination, and supplemented with domain-specific human assessment. Multimodal systems need testing for cross-modal errors and accessibility.

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 6, October 2026
Impact Factor: 8.2

Further work is also needed on energy reporting, copyright-respecting data practices, user consent, and governance that matches the risk of each application. Progress should be measured not only by model size or benchmark scores, but also by usefulness, reliability, fairness, and the ability to recover safely from errors.

XII. CONCLUSION

Large Language Models represent a major development in artificial intelligence and natural language processing. Their transformer-based foundations and large-scale training enable them to support a wide range of language tasks, while techniques such as instruction tuning, retrieval-augmented generation, multimodal learning, and tool use continue to broaden their capabilities. These developments create practical opportunities in education, research, software development, healthcare support, and business workflows.

At the same time, fluent output should not be confused with verified knowledge. Hallucination, bias, privacy risks, security vulnerabilities, resource demands, and inconsistent evaluation remain significant barriers to dependable deployment. Addressing these issues requires more than improving the model alone: developers and organisations must test the full system, protect data, document limitations, monitor real-world behaviour, and retain human oversight where consequences are important. Overall, the long-term value of LLMs will depend on combining technical progress with responsible design, transparent evaluation, and careful use.

ACKNOWLEDGMENT

The authors express their sincere gratitude to the project guide and co-guide for their guidance, constructive suggestions, and encouragement during the preparation of this review. The authors also thank the Department of Computer Engineering and Adsul Technical Campus, Chas, Ahilyanagar, for providing an academic environment that supported this work. Finally, the authors acknowledge the researchers whose published studies on transformers, language modelling, retrieval-based systems, and responsible artificial intelligence informed the discussion presented in this paper.

REFERENCES

[1] A. Vaswani et al., “Attention Is All You Need,” in Advances in Neural Information Processing Systems, vol. 30, 2017.
[2] T. B. Brown et al., “Language Models are Few-Shot Learners,” in Advances in Neural Information Processing Systems, vol. 33, pp. 1877–1901, 2020.
[3] J. Wei et al., “Finetuned Language Models Are Zero-Shot Learners,” in International Conference on Learning Representations, 2022.
[4] L. Ouyang et al., “Training language models to follow instructions with human feedback,” in Advances in Neural Information Processing Systems, vol. 35, pp. 27730–27744, 2022.
[5] P. Lewis et al., “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks,” in Advances in Neural Information Processing Systems, vol. 33, pp. 9459–9474, 2020.
[6] R. Bommasani et al., “On the Opportunities and Risks of Foundation Models,” arXiv:2108.07258, 2021.
[7] S. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell, “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?” in Proceedings of the ACM Conference on Fairness, Accountability, and Transparency, pp. 610–623, 2021.
[8] E. M. Bender and A. Koller, “Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data,” in Proceedings of ACL, pp. 5185–5198, 2020.
[9] N. Carlini et al., “Extracting Training Data from Large Language Models,” in USENIX Security Symposium, pp. 2633–2650, 2021.
[10] National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST AI 100-1, 2023.