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)

Large Language Models in Drug Discovery: A Literature-Based Review of Applications, Challenges and Future Directions

Shubham Mhatarmare

Department of Artificial Intelligence and Data Science
Ajeenkya DY Patil School of Engineering, Pune, India
Email : mhatarmare.shubham@dypic

Abstract

Taking a new molecule to the market involves identifying a suitable biological target, designing a molecule to interact with that target, screening candidates, and demonstrating the final molecule's safety. Done the traditional way, this takes years, a great deal of money, and a lot of laboratory work. Large language models (LLMs) are the newest addition to the tools available to researchers for studying chemical and biological data with AI. They are being tested for reading biomedical literature, identifying new drug targets, finding uses for existing drugs, and designing new molecules. In this paper I review how LLMs are being used in drug discovery. I consider the main approaches, what they can and cannot offer, and highlight where the research is still lacking. I also discuss practical issues such as unreliable outputs, poor data quality, the need for laboratory validation, computing cost, and responsible use. My conclusion is that future work should aim at systems that are dependable, explainable, tested in the laboratory, and that pair language models with specialized scientific tools.

Keywords : Large Language Models, Drug Discovery, Artificial Intelligence, Generative AI, Molecular Design, Drug Repurposing, Biomedical Natural Language Processing

I. INTRODUCTION

Drug discovery sits at the heart of pharmaceutical research. The work starts with identification of the biological targets associated with a disease, followed by discovery of molecules that can interact with them and assessment of their safety and efficacy. Many candidates still end up failing late in the process, each of which requires expensive, lengthy laboratory campaigns, deep specialist knowledge and lots of capital [20], [21]. AI has started to change this picture. Machine learning and deep learning can discover patterns in huge chemical and biological datasets, foretell molecular properties and assist in prioritizing compounds for closer examination first. A good example is protein structure prediction, which has made it much easier to study targets that were previously difficult to study, such as AlphaFold [3], [4]. Large language models are a newer arrival. They can read scientific text, summarize papers, extract biomedical

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)

facts and assist researchers in pulling together evidence from various sources. Studies have begun to test them at several stages of drug discovery and development [?], [?], [7].

A word of warning: text produced by a language model is not experimental evidence. Any prediction an LLM makes has to be checked against trusted scientific sources, specialized computational methods and, in the end, laboratory experiments.

The main objectives of this paper are to:

- Provide an overview of the major applications of LLMs in drug discovery.
- Compare recent methods and approaches.
- Identify existing research gaps and highlight future directions.

II. REVIEW METHODOLOGY

A. Sources and Search

Literature can be collected from several sources such as; PubMed, IEEE Xplore, Google Scholar, SpringerLink and ScienceDirect. Drug Discovery, Molecular Drug Development, Drug Repurposing, Biomedical Information Extraction and AI aided drug discovery will be the theme of the relevant papers presented. This is a rapidly changing field and a greater emphasis has been given to more recent research. Some of the older works have been added where they are of historical interest in relation to the various techniques employed in the newer methods.

B. Screening and Comparison Framework

For the purpose of comparability, each of the studies is analysed on the basis of the same criteria which are the AI technique employed, the application area, the methodology, the main result and the limitations mentioned by the authors. Using common criteria makes it easier to compare studies that use different models and evaluation methods.

C. Scope and Limitations of the Review

This review focuses on existing research. It does not introduce a new model for training or present original laboratory results. Its main focus is on the potential applications of LLMs and how they can be combined with established computational drug discovery methods.

Readers should draw their own conclusions based on the evidence reported in the existing studies.

III. BACKGROUND

A. The Drug Discovery Process

Creating a new drug is not a simple process; it is a series of steps. The first step in research is to identify a biological target that is associated with the disease and establish that it truly is important.They then

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)

search for molecules that bind to it (hits) and then narrow down the most promising molecules to make stronger leads. These leads go through preclinical testing before any human trials can begin to make sure that they are safe and effective. At each stage, biological mechanisms, chemical structures and experimental results have to be understood.

AI can help prioritize targets and molecules, but how useful it is depends on the quality of the data, on sensible evaluation and on experimental confirmation [20].

B. Large Language Models

LLMs are deep learning models, typically based on the Transformer architecture. They are able to learn patterns from very large amounts of text and generate output from patterns they have learned as well as from the input they are given.

In biomedical research, models of this kind have been trained or adapted to handle scientific papers, patents, drug descriptions and biological information [5], [6], [10], [11]. They can help researchers look for links between drugs, diseases, genes and proteins. Related language models trained on protein sequences have also been shown to capture structural and functional information [12], [13].

C. Relevance to Drug Discovery

There is far more published science than any one person can keep up with. Relevant findings are scattered across millions of papers and dozens of databases, and important connections are easy to miss. LLMs can help by sorting through this material, pulling related pieces together and suggesting ideas that researchers may want to test.

When they are connected to specialized tools, they can also help with literature retrieval, molecular analysis and the coordination of computational workflows. Even so, a hypothesis generated by a model should never be mistaken for proof of biological activity or clinical effectiveness.

IV. LLM TECHNIQUES FOR DRUG DISCOVERY

A. Biomedical Literature Mining

LLMs can summarize publications, recognize biomedical entities and extract the relationships that research articles describe. This can cut down a good deal of the manual effort that a literature review normally needs.

Still, extracted information should be checked against the original sources so that unsupported conclusions do not slip through.

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. Drug Target Identification

Target identification is about finding the biological components that are involved in a disease. LLMs can be used to retrieve information on genes, proteins and pathways from the literature [1]. Explainable deep learning methods have also been used to discover novel classes of antibiotics [19].

The hypotheses that come out of this process still need confirmation from biological databases and suitable experiments.

C. Drug Repurposing

Drug repurposing is the idea of using a drug already approved for one disease to treat another disease. LLMs can make connections between information about drugs, disease mechanisms, and published studies to propose links for further investigation.

A suggested link does not show that a treatment works, and it must be validated independently.

D. Molecular Design

Generative AI can propose new molecular structures based on learned chemical representations and stated design goals. Many approaches write molecules as text strings, for example in the SMILES format, which lets language-model techniques be applied directly to chemistry [14]–[16]. Recent systems such as DrugGen and Token-Mol show how language models can be used for drug design [?], [2]. Work combining AlphaFold with AI-driven design has also led to a new small-molecule inhibitor [22].

LLM-based tools can be used to generate candidates, but each structure generated by an LLM requires chemical validation to check whether it is chemically valid, active, non-toxic and synthesizable.

E. Integration with Computational Tools

LLMs can be connected to molecular property predictors, protein structure resources, docking software and scientific databases. Studies have shown that giving a language model access to chemistry tools improves its usefulness, and that LLM-driven systems can even plan and carry out parts of chemical experiments [8], [9].

In such systems, the LLM acts as a guide to set up the research question and organize the tools used by the specialized models to perform the calculations. These results need to be evaluated on appropriate scientific standards.

V. COMPARATIVE ANALYSIS OF THE LITERATURE

The key roles of various methodologies in AI-driven drug discovery are summarized in Table I. Each strategy has pros and cons. While LLMs excel at handling language-related tasks, numerous numerical and molecular problems require more specialized computational techniques. Combining the two might

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)

enable a more unified research process, but the impact of the combination on research performance is not yet proven and needs to be demonstrated in a well-designed experiment.

TABLE I

COMPARISON OF AI APPROACHES IN DRUG DISCOVERY

Approach	Methodology	Potential Contribution	Limitation
Machine Learning	Learning patterns from structured datasets	Molecular property prediction and candidate prioritization	Depends on dataset quality and generalization
Deep Learning	Learning complex representations	Chemical and biological prediction tasks	May require substantial data and computing resources
Generative AI	Generating candidate molecular structures	Exploring possible compounds for further testing	Generated structures may be unsuitable or ineffective
LLMs	Processing and generating scientific language	Literature analysis, information extraction and hypothesis generation	May produce unsupported or incorrect information
Tool-integrated LLMs	Combining language models with external scientific tools	Organizing multi-step research workflows	Depends on tool reliability and validation

VI. CHALLENGES AND ETHICAL CONCERNS

A. Reliability and Hallucinations

LLMs can produce scientifically convincing statements without supporting evidence. Faulty statements regarding the properties of a molecule or the molecule–target relationship can lead researchers astray.Therefore,checking sources and involving expert reviewers are essential.

B. Data Quality and Bias

An AI-proposed candidate is not necessarily a good drug. Biological activity, toxicity and other properties must be evaluated by computational evaluation and laboratory testing. Experimental validation plays an important role in the process of drug discovery based on artificial intelligence.

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)

C. Experimental Validation

An AI-proposed candidate is not necessarily a good drug. Computational evaluation and laboratory testing are required to check biological activity, toxicity, and other properties. Experimental validation is a crucial step in AI-based drug discovery.

D. Computational Cost

Large models require significant computing power. These costs can be reduced using smaller, more efficient models and retrieval-based methods, depending on the task.

E. Explainability and Reproducibility

Researchers need to understand why an AI model suggests a particular drug candidate and what evidence supports its recommendation. Reproducibility is equally important. Maintaining records of the model version, data sources, prompts, and evaluation methods helps others verify and repeat the results.

VII. EVALUATION OF LLM-BASED DRUG DISCOVERY SYSTEMS

There is no single test that suits every system because each is designed for a different job and should be evaluated accordingly.A molecule generator should be evaluated based on its chemical validity, novelty and on its capacity for production of molecules with the requested properties.Biological evidence is required to corroborate a potential drug–target relationship, and any molecule that is considered a candidate must be tested in a lab.One important distinction is that performing well on a benchmark does not guarantee real world success.A high score on a test set does not prove that a candidate will become a successful drug.

VIII. RESEARCH GAPS AND FUTURE DIRECTIONS

A. Research Gaps

There are several areas which still need to be improved:

- Reduction in hallucinations and false claims.
- Extension of reliable benchmarks for drug discovery.
- Improvements in integration of text and molecular data.
- Improvements in experimental validation.
- Improvements in reproducibility and interpretability.
- Testing on a broader range of biological tasks.

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. Future Directions

LLMs can be integrated with future systems such as Retrieval-Augmented Generation (RAG), biomedical knowledge graphs, and specialized molecular prediction tools. Multimodal approaches can integrate the information of scientific texts, molecular structures, protein information, and experimental results. Systems that can be integrated into tools can assist in coordinating literature reviews, drug candidate generation and the confirmation of results. All these developments require systematic evaluation and human supervision. They should be evaluated by reliable studies and experimentally.

IX. CONCLUSION

The real promise of large language models is in their ability to support drug discovery, particularly in the analysis of scientific literature, identification of drug targets, repurposing of drugs and aiding molecular design. They can organize scientific information to suggest hypotheses and determine the next steps to investigate, and their capabilities can be expanded by connecting them with specialized computational tools. At the same time, the possibility of hallucination, restrictions on the amount of data, computational expense and the ongoing need for experimental verification are significant obstacles.

LLMs should be used as tools to assist research, rather than as a replacement for scientific verification. Future research should focus on reliable, explainable and reproducible methods, and on systems that pair language models with specialized scientific tools.

REFERENCES

[1] “Application of artificial intelligence large language models in drug target discovery,” Frontiers in Pharmacology, vol. 16, 2025. doi: 10.3389/fphar.2025.1597351.

[2] “Token-Mol 1.0: tokenized drug design with large language models,” Nature Communications, vol. 16, Art. no. 4416, 2025. doi: 10.1038/s41467-025-59628-y.

[3] J. Jumper et al., “Highly accurate protein structure prediction with AlphaFold,” Nature, vol. 596, pp. 583–589, 2021.

[4] J. Abramson et al., “Accurate structure prediction of biomolecular interactions with AlphaFold 3,” Nature, vol. 630, pp. 493–500, 2024.

[5] K. Singhal et al., “Large language models encode clinical knowledge,” Nature, vol. 620, pp. 172–180, 2023.

[6] K. Singhal et al., “Toward expert-level medical question answering with large language models,” Nature Medicine, vol. 31, pp. 943–950, 2025.

[7] A. J. Thirunavukarasu et al., “Large language models in medicine,” Nature Medicine, vol. 29, pp. 1930–1940, 2023.

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)

[8] A. M. Bran et al., "Augmenting large language models with chemistry tools," Nature Machine Intelligence, vol. 6, pp. 525–535, 2024.

[9] D. A. Boiko et al., "Autonomous chemical research with large language models," Nature, vol. 624, pp. 570–578, 2023.

[10] R. Luo et al., "BioGPT: Generative pre-trained transformer for biomedical text generation and mining," Briefings in Bioinformatics, vol. 23, no. 6, Art. no. bbac409, 2022.

[11] Y. Gu et al., "Domain-specific language model pretraining for biomedical natural language processing," ACM Trans. Comput. Healthcare, vol. 3, no. 1, pp. 1–23, 2021.

[12] Z. Lin et al., "Evolutionary-scale prediction of atomic-level protein structure with a language model," Science, vol. 379, pp. 1123–1130, 2023.

[13] A. Rives et al., "Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences," Proc. Natl. Acad. Sci. USA, vol. 118, no. 15, Art. no. e2016239118, 2021.

[14] J. Ross et al., "Large-scale chemical language representations capture molecular structure and properties," Nature Machine Intelligence, vol. 4, pp. 1256–1264, 2022.

[15] V. Bagal et al., "MolGPT: Molecular generation using a transformer-decoder model," J. Chem. Inf. Model., vol. 62, no. 9, pp. 2064–2076, 2022.

[16] R. Irwin et al., "Chemformer: A pre-trained transformer for computational chemistry," Mach. Learn.: Sci. Technol., vol. 3, no. 1, Art. no. 015022, 2022.

[17] Z. Ji et al., "Survey of hallucination in natural language generation," ACM Comput. Surv., vol. 55, no. 12, Art. no. 248, 2023.

[18] S. Pan et al., "Unifying large language models and knowledge graphs: A roadmap," IEEE Trans. Knowl. Data Eng., vol. 36, no. 7, pp. 3580–3599, 2024.

[19] F. Wong et al., "Discovery of a structural class of antibiotics with explainable deep learning," Nature, vol. 626, pp. 177–185, 2024.

[20] D. Paul et al., "Artificial intelligence in drug discovery and development," Drug Discovery Today, vol. 26, no. 1, pp. 80–93, 2021.

[21] M. K. P. Jayatunga et al., "How successful are AI-discovered drugs in clinical trials? A first analysis and emerging lessons," Drug Discovery Today, vol. 29, no. 6, Art. no. 104009, 2024.

[22] F. Ren et al., "AlphaFold accelerates artificial intelligence powered drug discovery: Efficient discovery of a novel CDK20 small molecule inhibitor," Chemical Science, vol. 14, pp. 1443–1452, 2023.