International Journal of Technology and Applied Science (IJTAS)
E-ISSN: 2230-9004 ● Website: www.ijtas.com ● Email: editor@ijtas.com

Generative AI and Large Language Models in Drug Discovery and Pharmaceutical Sciences: Applications, Limitations and Ethical Considerations

Shouvik Mondal[1], Abhijeet Panigrahi[2]

[1]Assistant Professor, Department of Pharmaceutics, Netaji Subhas University, Jamshedpur, 831012
[2]Research scholar, Department of Pharmaceutics, Netaji Subhas University, Jamshedpur, 831012

Abstract:
The rise of generative artificial intelligence (AI) and large language models (LLMs) marks a revolutionary shift in pharmaceutical science. Transformer-based architectures such as GPT, BERT, Llama, and Claude have shown transformative potential throughout the entire drug discovery and development process, from target identification to clinical trial optimisation.
This review, based on five key peer-reviewed studies from 2024–2025, thoroughly explores the latest advancements in applying large language models (LLMs) to pharmaceutical sciences. The uses of LLMs in drug target identification, molecular design, virtual screening, ADMET prediction, drug repurposing, clinical trials, and analytical chemistry are compiled in this study.
Even though rentosertib (INS018_055) has advanced into Phase 2a clinical trials, there are still several obstacles to overcome, such as hallucinations, poor interpretability, problems with data quality, inadequate domain knowledge integration, and high computing demands. It is also necessary to address ethical issues including algorithmic bias, dual-use dangers, intellectual property, and patient data privacy. Multimodal LLMs, retrieval-augmented generation (RAG), parameter-efficient fine-tuning (PEFT), and responsible clinical implementation are the main areas of future study.

Keywords: Large Language Models (LLMs); Generative AI; Drug Discovery; Pharmaceutical Sciences; ADMET; Clinical Trials; AI Ethics; Drug Repurposing.

INTRODUCTION

Drug discovery and development remain one of the most complex and resource-intensive scientific endeavours. Bringing a single new molecular entity from concept to regulatory approval typically takes 10–15 years and costs over $2.5 billion USD, with a clinical success rate below 10%.[1] As of 2022, fewer than 500 empirically validated drug targets existed globally, highlighting a major bottleneck in early discovery.[2] The traditional pipeline - spanning target identification, hit generation, lead optimisation, preclinical evaluation, and multi-phase clinical trials—is marked by high attrition and rising costs.

The use of AI in pharmaceutical research has increased as a result of these difficulties. Large language models like GPT, BERT, and Llama can now effectively handle biomedical data and support a variety of drug discovery activities thanks to the development of Transformer-based architectures. [3, 4, 5, 6]
LLMs can analyse chemical and biological data for uses like drug design, clinical trial matching, protein structure prediction, and biomedical literature mining by treating molecular structures and biological sequences as structured languages. The present study is aimed at providing an objective analysis of the applications, challenges, ethical considerations, and prospects of large language models in pharmaceutical sciences through the analysis of six recent peer-reviewed articles (2024-2025). [1, 2, 7, 8, 9, 10]

2. Background: Architecture and Paradigms of Large Language Models

2.1 The Transformer Architecture and Self-Attention

Large language models today are built using the Transformer architecture. Its self-attention method permits rapid processing of extended sequences, making LLMs excellent for studying biomedical literature, molecular representations, and biological sequences. Large datasets are used to pre-train models, which can then be adjusted for specific pharmaceutical uses. [3, 9]
Large datasets are used to teach LLMs patterns during pre-training, which may then be refined for specific biomedical tasks. The quality and dependability of model responses are further enhanced by alignment strategies like Reinforcement Learning from Human Feedback (RLHF). [8, 11, 12]

2.2 Key LLM Architectures for Pharmaceutical Applications

Transformer-based models have been adapted to pharmaceutical applications. BERT-based models such as BioBERT, PubMedBERT and BioGPT have shown promising results in biomedical text mining and drug-target interaction prediction. [13, 14, 15]
General-purpose models, such as Claude, have also demonstrated promising performance in pharmaceutical applications, but models based on GPT and Llama have been adapted for molecular generation, optimisation, and interactive drug design. [7, 8, 16, 17, 18, 19]

2.3 Specialised vs. General-Purpose LLMs and Hybrid Approaches

Both general-purpose LLMs trained on large text corpora and specialised LLMs trained on biochemical data are used in drug development. While general-purpose models bring reasoning and knowledge integration capabilities, specialised models offer higher accuracy for molecular and biological activities. [9]
Hybrid techniques that integrate specialised biochemical models with general-purpose LLMs have shown enhanced performance across drug discovery tasks by merging domain-specific knowledge with superior reasoning capabilities. [9, 20]

3. Applications of LLMs Across the Drug Discovery Pipeline

3.1 Disease Mechanism Understanding and Drug Target Discovery

3.1.1 Genomics Language Models

Rational drug development begins with the identification of genes and biological pathways linked to disease. Drug target discovery is aided by genomics-based LLMs, such as DNABERT and the Nucleotide

Transformer, which examine DNA and RNA sequences to find mutations linked to disease, DNA–protein interactions, and viral evolution. [2, 9, 21]

3.1.2 Transcriptomics and Gene Network Analysis

Geneformer, an LLM built on transcriptomics and trained on millions of single-cell datasets, allows for gene regulatory network analysis and has been able to identify potential therapeutic targets that were later validated experimentally. [2, 22]

In addition to Geneformer, several other transcriptomics-based language models have improved the analysis of single-cell RNA sequencing data. Models such as scGPT, scBERT, CellPLM, and GeneCompass help researchers identify different cell types, understand gene expression patterns, and investigate gene regulatory mechanisms across different tissues. Together, these models provide valuable insights into disease biology and support the discovery of potential therapeutic targets. [23, 24, 25]

3.1.3 Protein Language Models and Target Validation

Models such as ESM, ESM2 and AlphaFold3 have made great progress in protein structure prediction, functional annotation, mutation analysis and target validation. These models enable drug target identification and speed up the discovery of novel therapeutic candidates. [2, 26, 27, 28]

Further improvements in identifying druggable proteins, designing antibodies, and generating new protein sequences with desired therapeutic properties have been achieved by other protein language models like QuoteTarget, AntiBERTa, IgFold, ProtGPT2, and ProGen2. [2, 27, 29, 30, 31, 32]

3.2 Virtual Screening and Chemical Space Exploration

LLMs have improved virtual screening, allowing rapid screening of large chemical spaces and identification of promising drug candidates. These allow for faster prediction of molecular properties compared to traditional screening methods and enable the identification of compounds with desirable pharmacological properties. [1, 33]

The integration of reinforcement learning with LLMs has further enhanced molecular generation and optimisation, allowing efficient identification of structurally similar compounds and improving virtual screening performance. [8]

3.3 Small-Molecule Design and Molecular Optimisation

3.3.1 Chemical Language Models and SMILES/SELFIES Representations

The direct use of NLP techniques in molecular design has been made possible by the idea of molecular structures as chemical "language" with atoms acting as words and SMILES strings as sentences. [1, 8] The most popular chemical notation, SMILES (Simplified Molecular Input Line Entry System), converts three-dimensional molecule structures into letter sequences. By using a self-referencing technique that ensures chemical validity, SELFIES (Self-Referencing Embedded Strings) solves the fragility of SMILES and offers more accurate representations of complicated molecules. [8]

Several chemical language models, such as 3DSMILES-GPT, FragGPT, and MolGPT, have proven to be effective in producing new drug-like compounds and refining molecular structures with better pharmacological characteristics. By generating chemically sound and structurally varied molecules, these models aid in the de novo drug creation process. [7, 8, 16, 17, 18]

3.3.2 De Novo Drug Design and Lead Optimisation

LLMs have dramatically enhanced de novo drug design by allowing for chemical optimisation, property prediction, and the development of molecules that target specific proteins. DrugAssist, Regression Transformer, TamGen, and ChemCrow are systems that help researchers create and optimise drug candidates, as well as plan synthesis and optimize reactions. [8, 19, 34, 35, 36]

3.4 ADMET Prediction and Toxicity Assessment

Accurate prediction of ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) qualities is critical for minimising medication failure throughout development. LLMs improve ADMET prediction by combining data from many biological and chemical sources, allowing for faster and more reliable assessments of medication safety and efficacy. [1, 7]

By combining multimodal molecular and clinical data, several LLM-based techniques, such as BERT-derived models, PharmaBench, GeoGNN, and CancerGPT, have enhanced the prediction of ADMET characteristics, toxicity, and adverse medication events. By allowing for the early detection of possible safety hazards, these models promote safer and more effective medication development. [7, 8, 37, 38, 39]

3.5 Drug Repurposing

Drug repurposing has become an important application of LLMs because it reduces development time and cost while building on the known safety profiles of existing drugs. Frameworks such as DrugReAlign integrate multiple biological data sources to identify new therapeutic uses for approved medicines. [7, 8, 5, 40]

By incorporating data from scientific literature and biological databases, large language models also improve biomedical knowledge graphs. This makes it possible to identify possible therapeutic options for a variety of illnesses, such as neurodegenerative, infectious, and autoimmune disorders. [8, 10]

3.6 Clinical Trial Optimisation

LLMs are being utilised to enhance clinical trial design by assisting with patient recruitment, eligibility evaluation, outcome prediction, and clinical decision-making. Systems such as TrialGPT, SEETrials, HINT, NYUTron, and Med-PaLM have proved AI's ability to increase trial efficiency and precision medicine. [7, 9, 10, 41, 42]

3.7 Analytical Chemistry and Autonomous Experimental Workflows

LLMs also help to advance analytical chemistry by improving spectral interpretation, compound identification, reaction analysis, and the automated extraction of chemical information from scientific data. These features improve the efficiency and accuracy of pharmaceutical research procedures. [8]

Beyond data analysis, LLMs are rapidly being integrated into self-contained laboratory systems to aid in experimental design, reaction optimisation, and automated analysis of toxicological and pharmacological data. These advancements have the potential to boost laboratory productivity while minimising manual labour. [9, 43, 44, 45]

4. Clinical Translation: From AI Predictions to Patient Benefit

The effective application of LLM-based drug discovery in clinical practice illustrates their expanding importance in pharmaceutical research. Rentosertib (INS018_055), produced utilising Insilico Medicine's AI platform, was the first AI-generated small-molecule medication to enter Phase 2a clinical trials and showed good clinical outcomes in patients with idiopathic pulmonary fibrosis. Furthermore, AI-assisted drug discovery has helped to develop therapeutic prospects for infectious, neurological, metabolic, and immunological illnesses, demonstrating the broad clinical potential of big language models. [8, 9, 10]

5. Limitations and Challenges

5.1 Hallucination and Factual Unreliability

The problem with Large Language Models is that they can make things up. This is called hallucination. This is an issue when it comes to using them in medicine. Hallucination can happen in ways like when Large Language Models design molecules that do not work, pair drugs with the wrong targets or say a drug is safe when it is not. Large Language Models like GPT-4 can also make mistakes with the types of atoms in a molecule and how they are connected, which means they can not produce molecules correctly. Because Large Language Models are trained on information from one place, they can be biased, and this makes hallucination more likely. [7, 8, 9]

People who use Large Language Models also tend to trust them a lot. This is because the things they write sound very good and certain.

This is called automation bias. It is a problem. Large Language Models are not always right. They can make things up. This can be very misleading. Large Language Models produce information. This is a problem. There are ways to make it better. One way is called Retrieval-augmented generation (RAG). It helps lower the rate of making things up. It does this by using information that is already checked. Using data from pharmaceuticals also helps. This is done by using knowledge graphs, special decoding algorithms and a step-by-step thought process. These ensure that the information is based on facts. Some researchers, like Gwon et al., have found that Large Language Models, such as ChatGPT, can find studies. They also make up fake references. This shows that people need to check all information from Large Language Models. [9, 46]

5.2 Model Interpretability and the Black-Box Problem

In the field of science, it is really important to understand how things work in order to get approval and be taken seriously. The problem with Large Language Models is that they are too complicated. It is hard to see how they make decisions. Researchers have a hard time figuring out how Large Language Models come up with their predictions, especially when it comes to finding new drugs. This is a problem because it makes it hard to trust Large Language Models and use them in a practical way. There are some techniques, like SHAP, that try to explain how Large Language Models think. They do not do a very good job, especially when Large Language Models are dealing with different types of information. [7, 8]

Large language models may sometimes focus more on things that look related on the surface than what is really going on with molecules. For example, they might keep parts of a molecule that do not actually help it work. In the future, we need to make systems that can explain things in a way that makes sense to regulators. These systems should include knowledge about biology. Show things in a way that experts can

understand. Deng did a thorough study and found out that how well a model can predict what a molecule will do depends on how the model is built and how it looks at things. This means we need to come up with ways to understand how these models are working. [8, 9, 47]

5.3 Data Quality and Domain Knowledge

The training data for Language Learning Models is really important. If the data is not complete or is biased, it can affect how well the model works. This is especially true for datasets. If these datasets are not good, it can be hard to make predictions. This can be a problem for pharmaceutical applications. We need to make sure we have data that is specific to the area we are working in. [7, 10]

5.5 Computational Resource Requirements and Accessibility

Smaller research institutes and developing nations may find it more difficult to acquire the enormous computational resources needed for training and deploying LLMs, which could exacerbate already-existing disparities in drug development capacities. [7] Extremely high storage and computing resources are needed for global fine-tuning at the model level, usually requiring clusters of high-performance GPUs that are not available in most academic labs.

Rather than retraining the entire model, Parameter-Efficient Fine-Tuning (PEFT) methods, especially low-rank adaptation (LoRA) [48], modify only a small subset of model parameters, providing a workable solution by lowering computational costs and facilitating quick adaptation to several specific tasks. [7] Another useful method for controlling computational costs at the pharmaceutical scale is the use of cascaded pipeline techniques, which combine LLM re-ranking of smaller candidate sets with quick physics-based filters. [9]

6. Ethical Considerations

6.1 Patient Data Privacy and Confidentiality

When we use Large Language Models for pharmaceutical research, the privacy of patient data is an important issue. Large Language Models need a lot of data, such as profiles that show kinds of information about patients, their clinical records and the sequences of their genes. This data often has health information in it. Because Large Language Models can remember the data they are trained on, they might accidentally reveal parts of patient profiles. Other people might be able to get this information if they do the searches. [9] The General Data Protection Regulation is something we have to think about when we use Large Language Models for pharmaceutical research. If we train Large Language Models using patient omics data, we might break the General Data Protection Regulation rules. This could happen if patient genomic information is shared without permission. [8]

A key requirement for using Large Language Models or LLMs in medical research is ensuring that private data is properly anonymised. This is crucial for protecting information. Currently, researchers are focusing on creating frameworks that protect privacy. These frameworks combine privacy with federated learning. Federated learning trains models on data spread across locations without gathering sensitive information in one place. Building rules for managing data, controlling access and tracking changes is vital. This is important for following guidelines and maintaining public trust in AI systems used in medicine, as these systems continue to expand. Murdoch stresses that these steps are necessary. He believes they are essential

for both compliance and preserving public confidence in pharmaceutical AI systems as AI health systems grow. [10, 49]

6.2 Intellectual Property Ambiguity

The creation of molecular structures by Large Language Models raises serious questions about who owns what. When a chemical made by a computer program turns out to be good for people, it is not clear how to divide up the rights among the people who gave the computer the information it used, the people who made the computer program and the people who use it. [8, 7] When we think about Large Language Models and if we can patent the things they create, it gets really complicated. Most places say that a person has to be the one who comes up with an idea to patent it. When Large Language Models help make new drugs, people and computers work together in a way that makes it hard to say who really came up with the idea. [10]

The rules for using Artificial Intelligence in making medicines are not really clear yet and are all over the place. This makes it really hard to make sure Artificial Intelligence algorithms are good enough for people to use. Artificial Intelligence has to work for lots of people. In the future, we need laws that say how to figure out who made something with the help of Artificial Intelligence. We also need laws that say what it takes for something made by Artificial Intelligence to be considered an invention that can be patented. We need to find a way to make sure that people who come up with ideas and companies that build and maintain Large Language Model systems get paid fairly. It is also hard to say who really owns something and to make sure that science experiments can be repeated when we do not know how some Large Language Model systems work. The people who make these systems keep lots of things, like the information they use to train the systems and how they update them. [7, 8, 9]

6.3 Algorithmic Bias and Fairness

The fact that some diseases and groups of patients are not well represented in the data used to train Large Language Models is a problem. It is not fair. When these models are used in situations that affect people who are already marginalised, they may not work well. This is because they were trained on data from wealthy countries and well-studied diseases. For example, models that are trained on tissue that's easy to get to may not be accurate for diseases that affect tissue that is hard to reach. This means that some people will not benefit from medicines that are discovered with the help of Artificial Intelligence. The diseases and groups of people that do benefit are the ones that're already well studied. This is a problem that needs to be fixed. [9]

We need to deal with bias in two ways. First, we have to make sure that the data we use to train Large Language Models includes people from ethnic groups, places and types of illnesses. We have to do this on purpose. Before we start using these models, pharmaceutical companies and the people who regulate them have to come up with a way to check for bias. They have to make sure the models work well for all kinds of patients. To make sure we are being fair when we compare how well these models work and to see how we are doing at finding drugs in a fair way, to everyone, we need to have some standard datasets that include all kinds of human diseases. These datasets have to be consistent, include a lot of types of data and be representative of all people. [2, 9]

6.4 Dual-Use Risks and Misuse Prevention

Large language models or LLMs can be used to make good molecules, for people. They can also be used to make bad things. The same power that helps us find new medicines can be used to make compounds. For example, there is a model called MegaSyn that can make molecules. This model was used to make chemicals. This shows that these models can be used for things. We also have to think about models that make proteins or peptides that can fight off germs. These models can be used to make things that can hurt people. So, we need to make rules to stop people from using these models for things. We also do not want to stop people from using them to make new medicines. We have to be careful when we let people use these models. We have to track how they are used and who can use them. This is a job, and we have to do it just right. LLMs are tools, and we have to use them in a good way. [9, 10]

Ethical review procedures, safety evaluations of produced compounds for toxicity and synthesizability, and regulatory compliance frameworks should be essential components of LLM-enabled pharmaceutical research rather than afterthoughts.[10] Before pursuing any experimental validation, research should examine the synthesizability and toxicity predictions of molecules produced using chemical language models, making sure that drug development rules are followed. [8]

6.5 Accountability and Graduated Autonomy

Assigning blame for success or failure becomes more difficult as LLMs have a greater impact on drug research decisions. Clear ethical principles and updated regulatory frameworks are necessary due to their rapid evolution and opaque reasoning. Zheng supports a graduated autonomy framework that is similar to the EU AI Act: high-stakes or clinical decisions require formal uncertainty estimates, counterfactual explanations, and regulatory oversight with required human-in-the-loop checkpoints; moderate-stakes tasks like hit triage require human verification and audit logs; and low-stakes tasks like literature triage may be fully automated. [9]

A three-tier paradigm for human-LLM collaboration is outlined by Lu as a tool, helper, or partner. [50] Each tier represents increasing levels of AI autonomy and correspondingly rising requirements for validation and oversight. To guarantee scientific reproducibility and ease regulatory scrutiny, future research should incorporate "model provenance" statements that are comparable to data availability sections, recording training data sources, model versions, and validation procedures. [9, 50] An institutional need for acceptable LLM deployment in pharmaceutical sciences is the creation of thorough ethical norms for interdisciplinary collaboration, involving chemists, physicians, AI specialists, ethicists, and patient representatives. [10]

7. Future Directions

7.1 Multimodal-Hybrid LLM Ecosystems

Multimodal architectures that natively process the diverse evidence base of drug discovery—chemical graphs, protein structures, microscopy images, electronic health records, and scientific literature—within unified computational frameworks will produce the most revolutionary advancements in the future. [9] Researchers will be able to query intricate cross-domain inquiries in natural language using next-generation Multimodal Large Language Models (MLLMs), receiving responses that concurrently synthesise chemical, biological, and clinical information. In order to enable general LLMs to more precisely adapt to particular task requirements in the biomedical field, Lin et al. [7] highlight the systematic

integration of general LLMs (like GPT-4) with specialised biochemical tools (like protein language models and reaction engines) as a priority research direction. [7]

7.2 Retrieval-Augmented Generation and Knowledge Integration

Retrieval-Augmented Generation (RAG) can improve LLM reliability by grounding model responses on reliable biomedical databases. This method helps to reduce the number of hallucinations and helps with more accurate drug-target prediction and clinical decision-making. [9, 46]

Streaming memory transformers allow LLMs to retrieve arbitrary data slices without catastrophic forgetting. [9]

7.3 Parameter-Efficient Fine-Tuning and Democratisation

Methods such as LoRA and adapter-based approaches for Parameter-Efficient Fine-Tuning (PEFT) allow adapting large language models to pharmaceutical applications at lower computational costs. These techniques reduce training costs and still offer good performance, thus making sophisticated AI tools accessible to academic and clinical researchers. [7, 48]

The open-source pharmaceutical LLMs and the standard benchmark datasets development will enable the AI models' accessibility, transparency and fair evaluation across research institutions. [2, 9]

7.4 Strengthening Prediction Reliability and Validation Frameworks

LLM-generated predictions need to be incorporated into clinical practice only after robust validation frameworks. Standardised benchmark datasets, disease-specific evaluation strategies and experimental validation are needed for the accuracy, reliability and clinical relevance of AI-assisted drug discovery. [7, 8]

Improving model interpretability and developing validation frameworks will make us more confident in predictions made by Large Language Models. In the future, AI systems that use agents may speed up biomedical research even more. These systems can combine literature analysis, generating hypotheses, designing experiments and interpreting data into one workflow. This unified approach can help researchers work efficiently. [9, 51]

8. Conclusion

Large language models and generative AI are revolutionising pharmaceutical sciences, supporting drug target discovery, molecular design, virtual screening, ADMET prediction, drug repurposing, clinical trial optimisation, and analytical chemistry. The evidence reviewed from the studies shows their increasing contribution to the improvement of the efficiency and accuracy of drug discovery.

The clinical success of rentosertib shows the potential of AI-assisted drug discovery in the real world and the ability of LLM-enabled platforms to speed up drug research and cut down on development time.

However, challenges such as hallucinations, lack of interpretability, data quality issues, computational requirements, and ethical issues related to privacy, fairness, intellectual property, and accountability need to be addressed to facilitate the wide adoption of LLMs in clinical practice.

Future progress will depend on AI systems that can work with things like retrieval-augmented generation and fine-tuning that use fewer parameters. We also need to test these systems carefully and have AI researchers, pharmaceutical scientists, clinicians and regulatory authorities work together. With

development and cooperation among different fields, Large Language Models (LLMs) can help make finding new medicines more efficient and reliable.
This can also make it more accessible to people who need it.

REFERENCES:

1. Guan S, Wang G. Drug discovery and development in the era of artificial intelligence: From machine learning to large language models. Artif Intell Chem. 2024;2:100070. doi:10.1016/j.aichem.2024.100070
2. Liu X, Zhang J, Wang X, Teng M, Wang G, Zhou X. Application of artificial intelligence large language models in drug target discovery. Front Pharmacol. 2025;16:1597351. doi:10.3389/fphar.2025.1597351
3. Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. arXiv. 2017. arxiv.org/abs/1706.03762
4. Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv. 2018. arxiv.org/abs/1810.04805
5. OpenAI, Openai AR, Openai KN, Openai TS. Improving language understanding by generative pre-training (GPT). 2018. gluebenchmark.com/leaderboard
6. Brown TB, Mann B, Ryder N, et al. Language models are few-shot learners. arXiv. 2020. arxiv.org/abs/2005.14165
7. Lin A, Fang X, Jiang A, et al. Large language models in drug development: Current progress and future directions. Curr Mol Pharmacol. 2025;18:1–5. doi:10.1016/j.cmp.2025.06.003
8. Ma J, Liu J, Xu D, Song X, Zhang Z. Application and prospects of large language models in small-molecule drug discovery. Anal Chem. 2025;97:27453–27477. doi:10.1021/acs.analchem.5c04083
9. Zheng Y, Koh HY, Ju J, et al. Large language models for drug discovery and development. Patterns. 2025;6:101346. doi:10.1016/j.patter.2025.101346
10. Chakraborty C, Bhattacharya M, Pal S, et al. AI-enabled language models to large language models and multimodal large language models in drug discovery and development. J Adv Res. 2025;78:377–389. doi:10.1016/j.jare.2025.02.011
11. Raschka S. Build a Large Language Model (From Scratch). Manning Publications; 2024.
12. Ouyang L, Wu J, Jiang X, et al. Training language models to follow instructions with human feedback. arXiv. 2022. arxiv.org/abs/2203.02155
13. Lee Y, Yoon W, Kim S, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics. 2020;36:1234–1240. doi:10.1093/bioinformatics/btz682
14. Gu Y, Tinn R, Cheng H, et al. Domain-specific language model pretraining for biomedical natural language processing. ACM Trans Comput Healthc. 2021;3:1–23. doi:10.1145/3458754
15. Luo R, Sun L, Xia Y, et al. BioGPT: generative pre-trained transformer for biomedical text generation and mining. Brief Bioinform. 2022;23:bbac409. doi:10.1093/bib/bbac409.
16. Wang J, et al. 3DSMILES-GPT: 3D molecular pocket-based generation with token-only large language model. Chem Sci. 2025;16(2):637–648. doi:10.1039/d4sc06174d

17. Yue J, et al. Unlocking comprehensive molecular design across all scenarios with large language model. Chem Sci. 2024;15(34):13727–13740. doi:10.1039/d4sc03744h
18. Bagal V, Aggarwal R, Vinod PK, Priyakumar UD. MolGPT: molecular generation using a transformer-decoder model. J Chem Inf Model. 2022;62:2064–2076. doi:10.1021/acs.jcim.1c00600
19. Ye G, et al. DrugAssist: a large language model for molecule optimization. Brief Bioinform. 2024;26(1):bbae693. doi:10.1093/bib/bbae693
20. Zhao WX, Zhou K, Li J, et al. A survey of large language models. arXiv. 2023. arxiv.org/abs/2303.18223
21. Ji Y, Zhou Z, Liu H, Davuluri RV. DNABERT: Pre-trained bidirectional encoder representations from transformers model for DNA-language in genome. Bioinformatics. 2021;37:2112–2120. doi:10.1093/bioinformatics/btab083
22. Theodoris CV, Xiao L, Chopra A, et al. Transfer learning enables predictions in network biology. Nature. 2023;618:616–624. doi:10.1038/s41586-023-06139-9
23. Cui H, Wang C, Maan H, et al. scGPT: toward building a foundation model for single-cell multi-omics using generative AI. Nat Methods. 2024;21:1470–1480. doi:10.1038/s41592-024-02201-0
24. Yang F, Wang W, Wang F, et al. scBERT as a large-scale pretrained deep language model for cell type annotation of single-cell RNA-seq data. Nat Mach Intell. 2022;4:852–866. doi:10.1038/s42256-022-00534-z
25. Yang X, Liu G, Feng G, et al. GeneCompass: deciphering universal gene regulatory mechanisms with a knowledge-informed cross-species foundation model. Cell Res. 2024;34:830–845. doi:10.1038/s41422-024-01034-y
26. Lin Z, Akin H, Rao R, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. Science. 2023;379:1123–1130. doi:10.1126/science.ade2574
27. Abramson J, Adler J, Dunger J, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature. 2024;630:493–500. doi:10.1038/s41586-024-07487-w
28. Ren F, Ding X, Zheng M, et al. AlphaFold accelerates artificial intelligence powered drug discovery: efficient discovery of a novel CDK20 small molecule inhibitor. Chem Sci. 2023;14:1443–1452. doi:10.1039/d2sc05709c
29. Chen J, Gu Z, Xu Y, et al. QuoteTarget: a sequence-based transformer protein language model to identify potentially druggable protein targets. Protein Sci. 2023;32:e4555. doi:10.1002/pro.4555
30. Leem J, Mitchell LS, Farmery JHR, Barton J, Galson JD. Deciphering the language of antibodies using self-supervised learning. Patterns. 2022;3:100513. doi:10.1016/j.patter.2022.100513
31. Ferruz N, Schmidt S, Höcker B. ProtGPT2 is a deep unsupervised language model for protein design. Nat Commun. 2022;13:4348. doi:10.1038/s41467-022-32007-7
32. Nijkamp E, Ruffolo JA, Weinstein EN, Naik N, Madani A. ProGen2: exploring the boundaries of protein language models. Cell Syst. 2023;14:968–978. doi:10.1016/j.cels.2023.10.002
33. Chen S, Zhong F. GPCRSPACE: a new GPCR real expanded library based on large language models. J Med Chem. 2024;67(18):16912–16922. doi:10.1021/acs.jmedchem.4c01257
34. Born J, Manica M. Regression transformer enables concurrent sequence regression and generation for molecular language modelling. Nat Mach Intell. 2023;5:432–444. doi:10.1038/s42256-023-00639-z

35. Wu KH, et al. TamGen for target-aware molecule generation. Nat Commun. 2024;15(1):9360. doi:10.1038/s41467-024-53443-3
36. Bran AM, Cox S, Schilter O, et al. Augmenting large language models with chemistry tools. Nat Mach Intell. 2024;6(5):525–535. doi:10.1038/s42256-024-00832-8
37. Niu Z, et al. PharmaBench: enhancing ADMET benchmarks with large language models. Sci Data. 2024;11(1):985. doi:10.1038/s41597-024-03793-0
38. Fang X, Liu L, Lei J, et al. Geometry-enhanced molecular representation learning for property prediction. Nat Mach Intell. 2022;4:127–134. doi:10.1038/s42256-021-00438-4
39. Fang Y, Zhang Q, Zhang N, et al. Knowledge graph-enhanced molecular contrastive learning with functional prompt. Nat Mach Intell. 2023;5:542–553. doi:10.1038/s42256-023-00654-0
40. Wei J, et al. DrugReAlign: a multisource prompt framework for drug repurposing based on large language models. BMC Biol. 2024;22(1):226. doi:10.1186/s12915-024-02028-3
41. Singhal K, Tu T, Gottweis J, et al. Towards expert-level medical question answering with large language models. arXiv. 2023. arxiv.org/abs/2305.09617
42. Makarov N, Bordukova M, Rodriguez-Esteban R, Schmich F, Menden MP. Large language models forecast patient health trajectories enabling digital twins. medRxiv. 2024. doi:10.1101/2024.07.05.24309957
43. Boiko DA, MacKnight R, Kline B, Gomes G. Autonomous chemical research with large language models. Nature. 2023;624:570–578. doi:10.1038/s41586-023-06792-0
44. Slattery A, Wen Z, Tenblad P, et al. Automated self-optimization, intensification, and scale-up of photocatalysis in flow. Science. 2024;383:eadj1817. doi:10.1126/science.adj1817
45. Silberg J, Swanson K, Simon E, et al. UniTox: leveraging LLMs to curate a unified dataset of drug-induced toxicity from FDA labels. medRxiv. 2024. doi:10.1101/2024.06.21.24309315
46. Lewis P, Perez E, Piktus A, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. arXiv. 2020. arxiv.org/abs/2005.11401
47. Deng J, Yang Z, Wang H, et al. A systematic study of key elements underlying molecular property prediction. Nat Commun. 2023;14:6395. doi:10.1038/s41467-023-41948-6
48. Hu EJ, Shen Y, Wallis P, et al. LoRA: low-rank adaptation of large language models. arXiv. 2021. arxiv.org/abs/2106.09685
49. Murdoch B. Privacy and artificial intelligence: challenges for protecting health information in a new era. BMC Med Ethics. 2021;22:122.
50. Anderson W, Braun I, Bhatnagar R, et al. Unlocking the capabilities of large language models for accelerating drug development. Clin Pharmacol Ther. 2024;116(1):38–41. doi:10.1002/cpt.3378
51. Yang J, Xu Z, Wu WKK, Chu Q, Zhang Q. GraphSynergy: a network-inspired deep learning model for anticancer drug combination prediction. J Am Med Inform Assoc. 2021;28:2336–2345. doi:10.1093/jamia/ocab162