Intelligence-Based Medicine 11 (2025) 100233

Contents lists available at ScienceDirect

Intelligence-Based Medicine

journal homepage: www.sciencedirect.com/journal/intelligence-based-medicine

Advancing drug discovery and development through GPT models: a review on challenges, innovations and future prospects

Zhinya Kawa Othman [a], Mohamed Mustaf Ahmed [b,*], Olalekan John Okesanya [c], Adamu Muhammad Ibrahim [d], Shuaibu Saidu Musa [e], Bryar A. Hassan [f,g], Lanja Ibrahim Saeed [h], Don Eliseo Lucero-Prisno III [i,j,k]

[a] Department of Pharmacy, Kurdistan Technical Institute, Sulaymaniyah, Kurdistan Region, Iraq
[b] Faculty of Medicine and Health Sciences, SIMAD University, Mogadishu, Somalia
[c] Faculty of Medicine, Department of Public Health and Maritime Transport, University of Thessaly, Volos, Greece
[d] Department of Immunology, School of Medical Laboratory Science, Usmanu Danfodiyo University Sokoto, Nigeria
[e] School of Global Health, Faculty of Medicine, Chulalongkorn University, Bangkok, Thailand
[f] Computer Science and Engineering Department, School of Science and Engineering, University of Kurdistan Hewler, Erbil, Iraq
[g] Department of Computer Science, College of Science, Charmo University, Chamchamal, Sulaimani, 44001, Iraq
[h] Bone Marrow Transplant Center, Hiwa Hospital, Sulaymaniyah, Kurdistan Region, Iraq
[i] Department of Global Health and Development, London School of Hygiene and Tropical Medicine, London, UK
[j] Center for Research and Development, Cebu Normal University, Cebu, Philippines
[k] Center for University Research, University of Makati, Makati City, Philippines

ARTICLE INFO

Keywords:
Drug discovery
GPT models
Drug development
Patient outcome

ABSTRACT

Advanced AI algorithms, notably generative pre-trained transformer (GPT) models, are revolutionizing healthcare and drug discovery and development by efficiently processing and interpreting large volumes of medical data. Specialized models, such as ProtGPT2 and BioGPT, extend their capabilities to protein engineering and biomedical text mining. Our study will contribute to ongoing discussions to revolutionize drug development, leading to a faster and more reliable validation of new therapeutic agents that are crucial for healthcare advancement and patient outcomes. GPT models, such as MTMol-GPT, are robust, generalizable, and provide important information for developing treatments for complicated disorders. SynerGPT utilizes a genetic algorithm to optimize prompts and select drug combinations for testing based on individual patient characteristics. Ligand generation for specific target proteins with potential drug activity is a significant stage in the drug design process, which enhances the quality of the synthesized compounds and augments the precision of capturing chemical structures and their activity correlations, highlighting the model's creativity and capability for innovative ligand design. Despite these advancements, there are still problems with the data volume, scalability, interpretability, and validation. Ethical considerations, robust methods, and omics data must be successfully integrated to develop AI for drug discovery and ensure successful deployment. In summary, these models significantly influence drug research and development, specifically in the earlier stages from initial target selection to post-marketing surveillance for medication safety monitoring.

1. Introduction

The emergence of new diseases and unmet healthcare requirements necessitate a rigorous drug discovery and development (DDD) process to create innovative drugs within the pharmaceutical industry [1]. Drug discovery aims to identify synthetic and biomolecular candidates for potential drug development. The process involves selecting and validating a druggable target, developing in vitro assays, and screening compound libraries to aid in finding hits that show promising activity, optimizing hits into leads with adequate potency, and testing lead

* Corresponding author.
E-mail addresses: zhinya.kawa@kti.edu.iq (Z.K. Othman), momustafahmed@simad.edu.so (M.M. Ahmed), ookesanya@uth.gr (O.J. Okesanya), 1411222099@udusok.edu.ng (A.M. Ibrahim), 6778006030@student.chula.ac.th (S.S. Musa), bryar.ahmad@ukh.edu.krd (B.A. Hassan), Lanja_farmacy@yahoo.com (L.I. Saeed), don-eliseo.lucero-prisno@lshtm.ac.uk (D.E. Lucero-Prisno).

https://doi.org/10.1016/j.ibmed.2025.100233
Received 9 February 2025; Received in revised form 25 February 2025; Accepted 2 March 2025
Available online 3 March 2025

efficacy in animal models. Subsequently, leads are further refined to enhance efficacy and pharmacokinetics before entering the drug development phase, which includes the preclinical and clinical stages. The discovery and development of innovative drugs is time- and cost-intensive, and approximately 12–15 years are needed to launch a new drug, with an estimated cost of approximately $2.5 billion to bring a drug to the market [1–3].

The utilization of machine learning (ML) and artificial intelligence (AI) to facilitate drug discovery and development makes the process more cost-effective and eliminates the need for clinical trials owing to the ability to conduct simulations using these technologies [4]. Large Language Models (LLMs), including generative pre-trained transformer (GPT) models, are a key branch of AI and focus on the interaction between computers and human language by leveraging mathematical and computational modeling to process and interpret large volumes of medical data [5]. These models have demonstrated remarkable efficacy in enhancing the precision and efficiency of healthcare [6]. GPT models employ transformer-based models that learn statistical patterns from natural languages, allowing them to synthesize human-like languages. The series began with GPT-1 in June 2018 and subsequently expanded to include GPT-2, GPT-3, and GPT-3.5 [7,8].

The most recent upgrade, GPT-4, was released in March 2023 and has made substantial progress in the production of logical and intelligible texts. GPT models are trained on large volumes of unstructured text data, allowing them to produce languages that are nearly identical to those of human-generated materials [9,10]. The knowledge gained by language models can be transferred to other activities, as humans routinely learn from one task and transfer the knowledge to another. GPT models combine transformer architecture, feedforward neural networks, and attention processes, which are pretrained on large volumes of data to learn broad language patterns, often via unsupervised learning. After pre-training, GPTs are fine-tuned for specific tasks, such as language modeling, sentiment analysis, text categorization, and question answering. Additionally, these models use the 'in-context learning' concept, in which models develop a broad set of skills and pattern recognition abilities at training time and then use those abilities at inference time to adapt to or recognize the desired task rapidly [11]. GPT models can learn chemical structures from large molecular datasets in a similar manner. In DDD, the core concept involves unsupervised pre-training to predict the next word (or token) based on prior input, enabling the model to acquire drug-like knowledge and generate valid simplified molecular input line-entry system (SMILES) strings. This approach supports conditional generation, allowing fine tuning with small supervised datasets for specific tasks, making GPT models versatile tools for both text and molecular generation [12]. Specialized GPT models such as ProtGPT2 and BioGPT have enhanced the application of GPT technology to areas such as protein engineering and biomedical text analysis [13]. ProtGPT2 is a language model based on the GPT2 Transformer architecture, designed to understand the protein language for protein design and engineering. With 36 layers, 1280-dimensionality, and 738 million parameters, it generates protein sequences retaining natural features, such as amino acid propensities and secondary structure, while exploring novel protein spaces. It causally predicts sequence tokens, enabling it to grasp protein language effectively, which is significant for advancements in protein engineering [14].

BioGPT is a domain-specific generative pretrained transformer model designed for biomedical text processing. Trained from scratch on 14 million PubMed abstracts, it leverages transformer architecture for tasks in the biomedical domain [15]. These algorithms have the ability to examine the scientific literature, extract relevant knowledge, and aid in the development of new treatment options. In de novo drug design, representing molecules with the SMILES notation enables the application of next-token prediction techniques. Using attention mechanisms, GPT captures the dependencies between tokens and focuses on the most important information in the input. Relying on its next-token prediction and attention mechanisms, GPT can generate novel drug molecules by learning from existing chemical data and creating structures with enhanced properties [16]. This review explores the advancements in drug discovery and development (DDD) by utilizing GPT models, which lead to the rapid and precise detection of new therapeutic compounds that are critical for enhancing patient outcomes and advancing medical science.

We further highlight the challenges associated with using GPTs in the DDD process and suggest addressing these problems by focusing on approaches that leverage GPT models in drug development to increase efficacy and success rates.

2. Application of GPT models in drug discovery and development

Although less than 500 human druggable targets have been identified as of 2022 [17], the development of successful treatments depends on the identification of new molecules for these targets. Significant efforts are being made to refine, revise, and reform the de novo drug discovery and development process with a focus on data-driven approaches to novel treatments. According to one report, AI technology can save at least 25–50 % of the time and money spent on drug development throughout the discovery and preclinical phases [18].

Consequently, AI-powered language models, such as GPTs, have the potential to open new avenues in drug development, paving the way for a faster and more cost-effective DDD [11].

Despite trained AI systems designed to target specific molecules require extensive and hand-labeled data, which are often unavailable owing to companies' reluctance to share proprietary information [19], a method known as federated analysis addresses the previous difficulties by examining various data sets independently and sharing just the resulting insights from each analysis. Federated learning, employed by corporations such as Google, allows users to exchange insights from their devices while safeguarding sensitive data from transfer. A more comprehensive system integrates differential privacy with secure multi-party computation (MPC), enabling many parties to jointly compute a function while maintaining the confidentiality of individual inputs. MPC guarantees that only the end outcome is disclosed, while the private inputs remain confidential [20]. Furthermore, the process of fine-tuning leverages a pre-trained model, such as OpenAI's GPT, by training it on a smaller, domain-specific dataset to improve performance on specialized tasks. This methodology leverages the model's pre-existing knowledge, minimizing data and computing demands while enhancing task-specific accuracy employed by pharmaceutical companies [21]. Hence, language models such as ChatGPT offer another potential solution by utilizing abundant unlabeled data, such as the 250 million protein sequences in the UniProt database, to autonomously learn relationships between molecular building blocks [19]. By utilizing its advanced language processing capabilities, ChatGPT can rapidly validate novel drug targets because of its ability to analyze literature and patent databases to identify disease-specific agents, compounds, and genes [19]. In drug research and development, ChatGPT functions as an effective instrument for virtual screening by identifying potential compounds with favorable interactions with biological targets. It can support the design of molecules with tailored properties, such as enhanced binding affinity and selectivity, or optimized ADMET (absorption, distribution, metabolism, excretion, and toxicity) profiles [22]. One study investigated how OpenAI's ChatGPT can help automate biomedical literature reviews for rapid drug discovery during health emergencies. Researchers have developed a pipeline to identify drug targets for viruses using PubMed articles, testing it for SARS-CoV-2 (with extensive literature) and Nipah (with limited literature). Articles were classified as relevant or irrelevant by human specialists, and ChatGPT results were evaluated against these expert assessments. Using GPT-4 and prompt engineering, the pipeline achieved high accuracy of 92.87 % and 87.40 % for SARS-CoV-2 and Nipah, respectively. The results showed an automated framework that used ChatGPT for biomedical literature

reviews to support drug development, incorporating six prompt templates to enhance its ability to extract relevant information from the vast biomedical literature. The goal was to reduce the time and resources needed for manual reviews and provide a tool for identifying drug targets, ultimately accelerating drug discovery and enabling more efficient responses to pandemics [23].

De novo drug design aims to create drugs with specific pharmacological properties; however, traditional methods often focus on single-target drugs, overlooking the complexity of diseases influenced by multiple factors. To address this, a novel model, MTMol-GPT, incorporates four key modules: a pre-trained generator, replay buffer, dual contrastive discriminator, and generative adversarial imitation learning (GAIL) to develop multi-target molecules against these receptors, such as dopamine type 2 (DRD2) receptor [24,25], epidermal growth factor receptor (EGFR), and thenon-receptor tyrosine kinase (c-Src) [26]. The research team initially pretrained transformer neural networks using the ChEMBL database to ensure that the resulting sequences conform to genuine drug-like structures in accordance with SMILES and Self-referencing Embedded Strings (SELFIES) [24]. They developed a dual-target molecular sequence creation system that utilized the GAIL algorithm. A replay buffer was employed to hold molecules from various target types, and a dual contrastive discriminator was implemented to assess the relative authenticity of the created and real molecules, serving as an incentive for training the generator. These findings indicate that MTMol-GPT can generate novel chemical entities that specifically target DRD2, EGFR, and c-Src receptors [27]. Molecular docking and pharmacophore mapping validated the drug-likeness and efficacy of these compounds, particularly in neuropsychiatric therapies and breast cancer treatment. Consequently, the MTMol-GPT exhibited robustness and generalizability and offered significant insights for the development of treatments aimed at complicated disorders [27]. Drug discovery and development entail the design of compounds that engage with biological targets within the body to express a specific therapeutic effect [28]. Recent advancements in GPT architectures for in-context learning have introduced a pre-trained model from scratch on drug synergy data without relying on textual corpora known as SynerGPT. The team evaluated the ability of this model to generalize in few-shot scenarios to drugs and patient cell lines not encountered during training, and found that it delivered strong competitive performance without utilizing external knowledge sources [29].

SynerGPT further introduced In-Context Learning for Drug Synergy (ICL-DS), a paradigm in which task adaptation occurs dynamically through contextual examples without parameter updates, making it particularly suited for personalized treatment and training on a limited dataset of 10–20 drug synergy relationships pertinent to cancer cell targets, thus enabling it to forecast further synergies within that framework. It incorporates a genetic algorithm to improve prompts and choose drug combinations for testing according to specific patient profiles. The model effectively retrieved relevant drugs or structurally similar alternatives from synergistic relationships, performing better than random even for unknown drugs. This capability enhances explainability and aids in identifying drugs with the desired synergies for drug candidate discovery [29]. Another model, called the Conditional Generative Pre-Trained Transformer (cMolGPT), was created for target-specific, de novo molecular design. By examining the data for drug-like structures, we created novel and diverse virtual chemical libraries. Target-specific molecules were created according to predetermined standards in the early phases of drug research to produce drug-like compounds with the required characteristics. The ability of cMolGPT to generate diverse chemical libraries and target-specific compounds was evaluated by comparing the results with actual data distributions. The efficient production of unique compounds with particular characteristics can streamline the drug design process, aid in the discovery of new therapeutic candidates, and improve molecular optimization [30].

To detect previously unexplored regions of the chemical space, a pre-trained ChemGPT model was developed. It is an autoregressive language model of small molecules that employs a transformer architecture with a self-attention mechanism to estimate conditional probabilities and generate new molecules. It was pre-trained on PubChem molecules using a causal language modeling task to predict the next token in a sequence based on previous tokens. The ChemGPT model is a robust and self-supervised representation learner that enables embedding-based nearest-neighbor search [31,32]. It is trained on datasets containing up to 10 million unique molecules and models with over one billion parameters, showcasing its capability to handle large-scale molecular data which can learn SELFIES molecule representations, while this the methodology could be applied to drug design SELFIES molecules. The study showed that the model size increased its performance by improving a pattern known as neural scaling, which is commonly seen in natural language processing (NLP) and is now proven to be applicable to molecular modeling as well [31]. Hence, the scaling strategies and results enable immediate and significant improvements in model performance, as well as computational and data efficiency for deep chemical models.

DrugGPT is an autoregressive ligand design method that integrates computational chemistry and deep learning to accurately generate ligands for specific target proteins. This GPT model was trained on extensive protein-ligand interaction data, in which the model captured chemical structural features and activity relationships to generate compounds with potential drug activity. Additionally, by tokenizing proteins and ligands using the Byte Pair Encoding (BPE) algorithm, DrugGPT reduces the representation of the vast chemical space to a finite vocabulary, effectively exploring it through learned combinations and arrangements of these tokens [33]. Therefore, it enhances the quality of synthesized compounds and augments the precision of capturing chemical structures and activity relationships by employing its own output as an input for predictions through an autoregressive generation approach [34]. A previous study allowed the DrugGPT tokenizer to combine vocabularies from ligands and proteins, removing duplicate tokens and conflicts. This process significantly reduced the average text length, making sequences easier to manage and train. With a finalized vocabulary size of 53,080 tokens, the tokenizer improved the efficiency of DrugGPT training, ensuring a better performance in drug discovery tasks. Three inference methods were employed, including designing ligands from given protein sequences, where the model was provided with the amino acid sequence of a protein (e.g., Bcl-2) to create corresponding ligands. The second method adds ligand-specific prompt information to the given protein sequence, where the researchers provide a starting SMILES representation of a ligand (e.g., "COc1ccc(cc1)C(=O)") to generate specific ligands for the protein. The third method allows the model to autonomously generate ligand designs without any input information, testing its ability to create potential ligands for proteins without prior instructions. Each of these methods demonstrates how DrugGPT can be applied to targeted ligand design. Consequently, the 112 matching ligands demonstrated that the DrugGPT model effectively learned protein-ligand relationships, as it was able to regenerate these known pairs [33]. Additionally, the 55 newly designed ligands showed that, despite being trained on a large number of existing Ectonucleotide pyrophosphatase 2 (ENPP2) ligands (known as autotaxin) are phosphodiesterase enzymes that have potential pharmaceutical value in many disease processes, especially in cancer, inflammatory diseases, and fibrosis, the model can still generate entirely novel ligands, highlighting the model's creativity and capability for innovative ligand design [33,35].

In the clinical trial phase, which is essential to medical research and provides critical evidence regarding the safety and efficacy of new treatments prior to public use, TWIN-GPT is an innovative methodology that employs large language models to generate personalized digital twins for efficient virtual clinical trial simulations. This approach tackles the difficulties of clinical trials, including significant time demands and participant engagement, particularly in contexts with inadequate data.

This model was trained and assessed using two primary datasets to improve its efficacy in virtual clinical trials. The initial dataset came from a Phase III breast cancer trial (NCT00174655) that evaluated Disease-Free Survival (DFS) results among 2887 individuals [36]. Participants were randomly assigned to groups to evaluate the efficacy of docetaxel, doxorubicin, cyclophosphamide, methotrexate, and fluorouracil (CMF) regimens. This openly available dataset offers a dependable basis for assessing the efficacy of TWIN-GPT in actual clinical cases. The Nearest Neighbor Adversarial Accuracy (NNAA) quantifies a model's overfitting and the corresponding risk of privacy leakage. This study utilized 71 individuals, 500 visit records, and 500 clinical records from actual data for training, resulting in the TWIN-GPT attaining a NNAA score of 0.271. A score near 0.5 indicates that TWIN-GPT successfully avoided excessive memory of the original data, instead of acquiring generalizable patterns. As a result, TWIN-GPT excels in fidelity, utility, and privacy, effectively addressing the privacy issues associated with conventional clinical trials [36]. Large language models like GPT-3.5 and GPT-4 are used to automate literature screening in pharmacovigilance, aiding in the identification of relevant articles for drug safety reviews [37]. Pharmacovigilance involves the science and practice of monitoring, evaluating, and preventing adverse effects of medications after they have been marketed, which significantly contributes to public health by improving drug safety [38]. These models improve the efficiency of the screening process by categorizing the articles as relevant or irrelevant, thereby enhancing the detection of safety issues. GPT-3.5 and GPT-4 demonstrated high accuracy, with reproducibility and sensitivity rates of 93 % and 97 %, respectively. Insights from their use help in developing automated systems for drug safety monitoring, ensuring more efficient literature screening, and improved safety evaluations [37].

3. Challenges of Applying GPT Models in Drug Discovery and Development

Although GPT models offer significant advantages in the drug discovery and development process, various challenges hinder their successful implementation of GPT models. Relying exclusively on pretrained LLM decoding for drug discovery leads to inconsistencies in producing valid, high-quality molecules, and lacks the flexibility to adapt to different objectives [39].

3.1. Lack of chemical and biological insights

Current GPT models face challenges in various tasks and benchmarks for molecular and protein generation. For example, although models trained on the basic QM9 dataset may produce a high frequency of legitimate and stable compounds, they encounter difficulties with the more intricate GEOMDrugs dataset. Moreover, tasks involving protein-molecule binding present significant challenges, as existing models are inadequate for producing compounds with high binding affinities for targets [40]. Additionally, tokenization is a common problem in the drug development process because chemical representations, such as SMILES or SELFIES, may not effectively capture or enhance chemical information, potentially limiting prediction accuracy and data efficiency [41]. GPT-3 lacks an inherent understanding of the chemical principles driving molecular properties and relies solely on pattern recognition in data, without deep comprehension [41].

Although molecule generation has been well documented with standardized benchmarking procedures, generative tasks in protein design have not been established, making it difficult to objectively evaluate designed proteins [40]. Furthermore, current chemical LLMs struggle to fully understand the complexity and breadth of chemical knowledge, particularly in retrosynthesis, and need to integrate more advanced knowledge from quantum chemistry [42]. Although ChatGPT can generate human-like text, it does not comprehend the underlying biology of the systems it simulates, resulting in predictions that may not accurately reflect their true complexities [43]. Similarly, AI, including GPTs, may improve early drug candidate selection, leading to safer and more viable drugs for human trials. However, in Phase II trials, where drug efficacy and side effects were evaluated in larger patient groups, the success rate dropped to 40 % (4 out of 10 completed trials), which aligns with historical industry averages of 30–40 % [44,45]. Thus, careful assessment of their impact in preclinical and in silico models is needed to balance treatment efficacy and tolerable toxicity in diverse patient populations [46].

3.2. Resource limitations

The availability of labeled data, which is crucial for training effective models with peptides or proteins, poses several challenges during the use of GPT models in DDD [47]. These models require large datasets and significant computational power, which can be demanding. Another challenge is the lack of information about certain regions of the protein/chemical landscape, which restricts current methodologies to exploiting existing data rather than fully exploring new data [48]. Limited training data can also mean that the models lack recent or specialized information. Furthermore, the massive computational resources required to handle and train extensive datasets pose significant challenges, such as identifying useful molecules without adding noise [49]. Fine-tuning GPT-3 for chemistry tasks demands considerable computational resources and financial investment, which can be prohibitive for broader adoption [41], especially with large datasets.

3.3. Ethical concern

Numerous legal and ethical issues may arise from the utilization of GPT models, notably the ChatGPT. Inappropriate utilization may result in breach of copyright statutes, health restrictions, and other legal frameworks. This may stem from biases in training data, insufficient information, inadequate comprehension of reality, or constraints of the algorithm [50]. Structure-based de novo design has restricted practical implications, likely attributable to biases and constraints in current protein-ligand affinity datasets [51]. Concerns about privacy, equity, and the ethical application of models, especially skewed data, exacerbate this dilemma [52]. The development of new drugs is plagued by challenges and expenses due to limitations and ethical issues, including potential copyright violations, medicolegal complications, occasional inaccuracies, generation of harmful or biased outcomes, imprecision in assessing source reliability, creation of irrelevant references, confidentiality concerns, and plagiarism [53].

3.4. Expense and training time

A significant limitation of artificial intelligence in small molecule drug development is the training of ML models that encompass chemical properties, commonly known as quantitative structure-activity relationships (QSAR), based on chemical structures. The timing and methodology for implementing machine-learning models vary based on the distinct requirements of each drug development initiative. Projects may concentrate on elements such as target selectivity, delivery, toxicity, or patent issues, necessitating distinct computational strategies for each [54]. Developing and maintaining these models require significant financial resources [52]. GPT models require extensive data and prolonged training times, which can be impractical for small research entities. Storing model weights, intermediate representations, and training data can also exhaust a significant storage capacity [55]. The rapid pace of scientific advancement poses a challenge to LLMs. New discoveries and methodologies have emerged, making models trained on older datasets obsolete. Hence, access to advanced AI tools requires substantial computing power, financial investment, education, and impeding progress in certain regions or communities [56].

3.5. Data quality & security

LLMs depend on vast amounts of data and a deep understanding of scientific terminology, including sensitive information such as patient details, genetic sequences, and proprietary research data. Ensuring the privacy and security of these data is crucial, as unauthorized access or breaches can have severe consequences, requiring robust measures to protect sensitive information [56]. The performance of the ChatGPT is only as good as the quality and availability of its training data. If the data are incomplete, biased, or inaccurate, the predictions of the model may be unreliable. The available data often have significant biases stemming from the limited sampling of the chemical space and issues with data quality and reproducibility [48]. Additionally, bioactivity assay data are biased towards the platform and contain intrinsic experimental errors, which disrupt accurate predictions [57].

3.6. Interpretability and transparency challenges

Interpretability issues arise from the challenge of understanding and explaining how a model makes decisions [52]. Many studies have used engineered molecular descriptors to train conventional machine learning models, but these models often suffer from intrinsic feature biases [57]. For example, GPT-3 functions as a black box, indicating that its internal workings are not easily understood or modified based on chemical knowledge [41]. The black-box characteristic of ChatGPT complicates the comprehension and elucidation of its internal mechanisms, resulting in difficulties in trusting its predictions [43]. Expert systems must disclose their predictive rules and uncertainty assessments, and provide explicit descriptions of the optimization goals and quantitative performance metrics. This transparency assists users in determining the appropriate timing and application of algorithms to achieve legitimate outcomes and enhance decision making [46]. The absence of a completely open-source GPT-3 model constrains its capacity to comprehend the rationale behind certain forecasts, which is essential in situations demanding transparency [41].

4. Strategies to overcome the challenges

A comprehensive strategy that tackles the challenges related to their implementation is required to maximize the use of GPT models in drug discovery and development. Enhancing the caliber and variety of the training data is essential. Increasing the amount and improving the quality of training data should be the main goals of future studies [50]. By compiling large, high-quality datasets that cover a wide range of chemical and biological variables, researchers can significantly increase the effectiveness and versatility of GPT models [58]. This approach not only minimizes the limitations caused by incomplete or skewed data but also increases the model's ability to generate novel and useful insights in a variety of drug development domains [33]. According to AI, removing false information necessitates cooperative community efforts [54]. Furthermore, by using inclusive and diverse training data, putting bias detection techniques into practice, and establishing strict evaluation standards, researchers and developers must actively identify and correct biases in their models [52]. A more thorough basis for models can be established by including data from various sources such as literature, clinical trials, and experimental results, enabling them to produce accurate and well-informed predictions [59]. Therefore, experimental validation is required to validate the model predictions [43]. Although general-purpose language models show promise, their usefulness in drug discovery can be greatly increased by creating models customized to domain-specific data and terminology [60]. This specialization can address the challenges associated with insufficient chemical and biological understanding, as well as tokenization problems that frequently hinder general-purpose models when managing complicated chemical structures [61]. Recent discussions have explored ways to enhance SELFIES to address the shortcomings of the current molecular string representations. These discussions can influence updates for both SMILES and DeepSMILES [62]. Further progress in property data mining and aggregation is expected to mitigate the challenges resulting from insufficient engineering knowledge of structure-property relationships among diverse molecular types, such as crystal structures, alloys, proteins, nucleic acids, polymers, ionic liquids, and biologics. Enhancing open-source datasets and repositories of accessible property models is essential for advancing the development of this domain [63]. By integrating chemical representation learning methodologies and refining the tokenization process to effectively manage SMILES or SELFIES notations, these specialized models can more precisely reflect the complexities and characteristics of the molecular structures [64].

Resolving the interpretability and transparency issues of GPT models is essential for their effective incorporation into drug development. To increase model clarity, researchers have investigated a variety of approaches, such as layer-wise relevance propagation and attention visualization [65]. By clarifying the reasoning behind a model's predictions, these techniques aim to make it easier for scientists to validate and build confidence in their results [66]. In addition, a compromise between efficacy and transparency can be achieved by developing hybrid approaches that combine the generative capabilities of GPT models with more interpretable machine-learning techniques [67]. This method satisfies the regulatory requirements for transparency in drug development processes and promotes confidence among researchers [68]. One study suggested using LLMs to extract features, after which interpretable machine learning models, such as random forests or linear classifiers, are applied to the features [69]. By encouraging LLMs to communicate more intermediate thinking processes, the chain-of-thought (CoT) prompting idea improves their interpretability while preserving or even improving their performance [70].

To overcome the resource limitations and reduce the expenses associated with training and implementing large GPT models, researchers have investigated effective model architectures and training techniques. Methods such as model pruning, quantization, and knowledge distillation facilitate the development of more compact models that preserve a substantial portion of the performance of their larger equivalents while demanding considerably fewer computational resources [71]. Utilizing cloud computing and distributed training methodologies might democratize access to these advanced models, allowing smaller research institutions and corporations to use their capabilities without requiring extensive in-house computing resources [72].

Collaboration among AI researchers, domain specialists, and regulatory authorities is crucial to address ethical issues and the absence of established norms regarding the application of GPT models in drug development [73]. Formulating explicit protocols for the ethical application of AI in pharmaceutical research, tackling data privacy concerns, and guaranteeing equity in model outcomes are essential measures as well [74]. Establishing defined benchmarks and evaluation metrics tailored to drug discovery tasks may facilitate the objective assessment of various models and methodologies [75]. This collaborative endeavor may result in the development of shared resources and pretrained models, promoting a more transparent and expedited rate of innovation in the domain [76]. Incorporating GPT models into the current DDD process requires meticulous attention to workflow efficiency and user interface design [19]. Creating intuitive interfaces that enable researchers to engage with these models efficiently, without requiring extensive knowledge of AI, can substantially improve their uptake and influence [77]. Furthermore, establishing automated pipelines that effectively integrate GPT model outputs into the subsequent phases of drug development, including virtual screening and ADMET prediction, might enhance the overall efficiency of the process [78]. Consequently, access to extensive datasets is essential. By concentrating on these integration components, researchers can guarantee that the capabilities of GPT models are completely actualized within the practical limitations of real-world drug discovery initiatives.

5. Future prospects

The pharmaceutical sector is progressively employing machine-learning and deep-learning algorithms to expedite drug research and development. This provides new opportunities for drug repurpose and reutilization by uncovering novel therapeutic indications [79]. A research team analyzed AI-discovered molecules that entered various stages of DDD and suggested that AI, including GPT models, improves drug development success rates, particularly in early stage trials. Currently, there are no FDA-approved AI-discovered drugs; however, artificial intelligence algorithms, including GPT models, have the potential to advance the pharmaceutical industry [80]. By December 2023, 24 AI-discovered drugs had completed Phase I trials of drug development process, with 21 succeeding, indicating an 80–90 % success rate which significantly higher than the historical industry average of 40–65 % [44,81,82]. Therefore, AI has the potential to optimize later stages of drug development in the future.

By early 2024, the top 20 pharmaceutical companies integrated AI into Research and Development (R&D), with many collaborations involving AI-native biotech firms. This has led to a significant increase in the number and scale of AI partnership deals over the past five years [44]. For example, Sanofi (a multinational pharmaceutical company) embraces AI through partnership with OpenAI and Formation Bio to develop AI-powered drug discovery software [80]. The head of R&D at Sanofi (“Houman Ashrafian ”) stated that AI and machine learning enhance R&D efficiency by improving drug modeling, optimizing dosing, and refining clinical trial recruitment, which has doubled since AI’s implementation. Expanding AI applications will further accelerate drug development and bring transformative treatments to patients more quickly [83]. Therefore, future initiatives must concentrate on creating resilient machine learning models capable of integrating and analyzing multimodal omics data while addressing issues such as data integration, feature selection, and the explanation of intricate interactions across several omics’ layers [84]. To transform drug discovery and development utilizing GPT models, researchers must create frameworks for their integration with current databases and experimental data, foster collaboration between data scientists and domain experts, and conduct pilot studies to validate the predictive insights offered by these models. Standardizing data formats and protocols among research institutes and databases, implementing advanced data curation methods, and establishing open-access data repositories are crucial for enhancing the efficacy of the GPT model [85]. Improving the interpretability and transparency of GPT models is essential for effective implementation in drug development. Investments must be allocated to research for the development of interpretability tools and visualization approaches, and guidelines should be formulated to interpret model outputs and incorporate them into decision-making processes [86].

Facilitation of multidisciplinary collaboration, including cross-disciplinary workshops and conferences, the establishment of interdisciplinary research teams, and the support of joint research projects, are essential to tackle specific challenges in DDD associated with technologies [87]. The implementation of rigorous validation and testing processes is essential to assess the precision and dependability of GPT models in drug discovery applications. Investing in training and capacity development is essential for efficient application of GPT models in drug discovery. Educational programs and workshops aimed at GPT model applications should be established; resources and support for ongoing learning should be provided; and information exchange through seminars and online courses should be encouraged. The integration of deep learning with structural biology may enhance the accuracy of predictions regarding protein-ligand interactions and protein dynamics. It is advisable to tackle the difficulties of incorporating protein flexibility and conformational alterations into deep learning models as well as comprehending complicated chemical relationships [88,89]. Addressing these obstacles will improve the efficiency and efficacy of drug discovery, perhaps resulting in expedited and successful therapeutic advancements [90]. The National AI Research Resource (NAIRR) is a pivotal program designed to furnish infrastructure for responsible and extensive AI research, facilitating access to computational resources and datasets, necessitating substantial federal investment and collaborative assistance [91]. Ensuring secure access to federal databases is essential for drug discovery and development. Expanding pilot projects that provide secure data access and formulating standards that include modern privacy technology can fulfill these requirements. Collaboration among academia, industry, national laboratories, and federal agencies is essential for progressing AI research in drug discovery [92]. Funding agencies must expand their support to encompass many research methodologies and collaborations while equilibrating contributions from private sector organizations with public and scientific objectives. Specific agreements on intellectual property rights and licensing must be formulated to synchronize commercial incentives with research goals [93].

Embracing the fundamentals of responsible, transparent, and reliable AI during the research process is crucial for ethical and efficient application of AI in drug discovery. Organizations such as the National Science Foundation (NSF) and National Institute of Standards and Technology (NIST) ought to spearhead initiatives to formulate regulations and develop methods for bias identification and data validation [93]. Continuous discourse among AI developers, policymakers, and specialists from many domains will enhance the formulation of responsible AI utilization policies and adjust them to changing technological and ethical dilemmas in drug discovery and development [94]. Data security and safety are essential to safeguard patient privacy and confidentiality. Comprehensive data-handling processes are crucial for protecting privacy and maintaining ethical norms [95]. In accordance with these principles, solutions must be created to recognize and alleviate biases in AI systems. Regulatory authorities must implement rigorous standards, processes, and guidelines to address ethical issues, ensure patient safety, and promote animal welfare. FDA’s discussion paper on artificial intelligence in drug research and clinical trials indicates optimal procedures for the implementation of AI in these domains [96]. Attaining ethical AI utilization in the pharmaceutical industry entails thorough research and development, robust legislation, and ongoing assessment [97].

6. Conclusion

GPT models significantly enhance the multiple stages of drug discovery and development, particularly in the early stages. Initially, these models facilitated the validation of prospective therapeutic targets by evaluating massive datasets pertaining to diseases or proteins and suggested innovative chemical structures. In the drug discovery phase, GPT models aid in the design of new compounds and prediction of their pharmacokinetic and pharmacodynamic features, thus expediting the identification of potential candidates. In clinical trials, they may assist in assessing trial data, forecasting patient reactions, and enhancing trial designs through virtual simulations. In the later stages of post-marketing evaluation, their role in drug safety monitoring is important for preventing adverse effects and other medication-related issues. Therefore, GPT models can offer substantial benefits in accelerating drug discovery and development. Addressing current challenges and implementing robust solutions are key to maximizing their potential and ensuring their effective use in the pharmaceutical industry.

CRediT authorship contribution statement

Zhinya Kawa Othman: Writing – review & editing, Writing – original draft, Conceptualization. **Mohamed Mustaf Ahmed:** Writing – review & editing, Conceptualization. **Olalekan John Okesanya:** Writing – review & editing, Writing – original draft, Data curation. **Adamu Muhammad Ibrahim:** Writing – review & editing, Writing – original draft, Data curation. **Shuaibu Saidu Musa:** Writing – review &

editing. **Bryar A. Hassan:** Writing – review & editing. **Lanja Ibrahim Saeed:** Writing – review & editing. **Don Eliseo Lucero-Prisno:** Supervision.

Ethics statement

Approval from the ethics committee was not required.

Data availability statement

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Funding

The authors have not received any funding for this study.

Declaration of competing interest

The authors declare that they have no conflicts of interest.

Acknowledgement

None.

References

[1] Sinha S, Vohora D. Drug discovery and development. Pharmaceutical Medicine and Translational Clinical Research; 2018. p. 19–32.
[2] Berdigaliyev N, Aljofan M. An overview of drug discovery and development. Future Med Chem 2020 Apr 9;12(10).
[3] Chan HCS, Shan H, Dahoun T, Vogel H, Yuan S. Advancing drug discovery via artificial intelligence. Trends Pharmacol Sci 2019 Aug;40(8):592–604.
[4] Patel V, Shah M. A comprehensive study on artificial intelligence and machine learning in drug discovery and drug development. Intelligent Med 2021 Nov;2(3). Available from: https://www.sciencedirect.com/science/article/pii/S2667102621001066.
[5] Mohammad AF, Clark B, Hegde R. Large Language model (LLM) & GPT, A monolithic study in generative AI. 2023 Jul 24.
[6] Yenduri Gokul, Ramalingam M, Chemmalar Selvi G, Supriya Y, Srivastava G, Kumar P, et al. GPT (generative pre-trained transformer) – a comprehensive review on enabling technologies, potential applications, emerging challenges, and future directions. IEEE Access 2024 Jan 1;12. 1–1.
[7] OpenAI licenses GPT-3 technology to microsoft. OpenAI 2020. https://openai.com/blog/openai-licenses-gpt-3-technology-to-microsoft/. [Accessed 3 March 2025].
[8] Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I. Language models are unsupervised multitask learners [internet]. Available from: https://insightcivic.s3.us-east-1.amazonaws.com/language-models.pdf; 2019.
[9] OpenAI. GPT-4 technical report. arXiv (cornell university) [internet]. Available from: https://arxiv.org/abs/2303.08774; 2023 Mar 15.
[10] aka A, Taiwo R, Saka N, Salami BA, Ajayi S, Akande K, et al. GPT models in construction industry: opportunities, limitations, and a use case validation. Developments Built Environ 2024 Mar 1;17:100300. Available from: https://www.sciencedirect.com/science/article/pii/S2666165923001825.
[11] Liu Z, Roberts RA, Lal-Nag M, Chen X, Huang R, Tong W. AI-based language models powering drug discovery and development. Drug Discov Today 2021 Nov;26(11):2593–607.
[12] Wang Y, Zhao H, Sciabola S, Wang W. cMolGPT: a conditional generative pre-trained transformer for target-specific de novo molecular generation. Molecules 2023 May 30;28(11). 4430–0.
[13] Ray PP. ChatGPT: a comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. Internet of Things and Cyber-Physical Systems 2023 Apr;3(1):121–54. Available from: https://www.sciencedirect.com/science/article/pii/S266734522300024X.
[14] Ferruz N, Schmidt S, Höcker B. ProtGPT2 is a deep unsupervised language model for protein design. Nat Commun 2022 Jul 27;13(1).
[15] Luo R, Sun L, Xia Y, Qin T, Zhang S, Poon H, et al. BioGPT: generative pre-trained transformer for biomedical text generation and mining. Briefings Bioinf 2022 Sep 24;23(6).
[16] Haroon S. Generative Pre-trained Transformer (GPT) based model with relative attention for de novo drug design. Comput Biology and Chem 2023 Oct 1;106:107911. CA Hafsath. https://www.sciencedirect.com/science/article/pii/S1476927123001020 [cited 2023 Aug 19];Available from:.
[17] Dhudum Rushikesh, Ankit Ganeshpurkar, Pawar A. Revolutionizing drug discovery: a comprehensive review of AI applications. Drugs and Drug Candidates 2024 Feb 13;3(1):148–71.
[18] How 2025 can be a pivotal year of progress for Biopharma [Internet]. World Economic Forum; 2025. Available from: https://www.weforum.org/stories/2025/01/2025-can-be-a-pivotal-year-of-progress-for-pharma/.
[19] Savage N. Drug discovery companies are customizing ChatGPT: here's how. Nat Biotechnol 2023;41:585–6. https://doi.org/10.1038/S41587-023-01788-7.
[20] Mugunthan Vaikkunth, Byrd David, Hybinette Balch Tucker, Morgan JP, Research AI. SMPAI: secure multi-party computation for federated learning. 2019.
[21] Parthasarathy Venkatesh Balavadhani, Zafar Ahtsham, Iqbal khan Aafaq, Shahid Arsalan. The ultimate guide to fine-tuning LLMs from basics to breakthroughs: an exhaustive review of technologies, research, best practices, applied research challenges and opportunities. 2024. ArXiv abs/2408.13296.
[22] Pradhan T, Gupta O, Chawla G. The future of ChatGPT in medicinal chemistry: harnessing AI for accelerated drug discovery. ChemistrySelect 2024 Mar 27;9(13).
[23] Yang J, Walker KC, Bekar-Cesaretli AA, Hao B, Bhadelia N, Joseph-McCarthy D, et al. Automating biomedical literature review for rapid drug discovery: leveraging GPT-4 to expedite pandemic response. Int J Med Inf 2024 Sep;189:105500.
[24] Krenn M, Häse F, Nigam A, Friederich P, Aspuru-Guzik A. Self-Referencing Embedded Strings (SELFIES): a 100% robust molecular string representation. Mach Learn: Sci Technol 2020 Nov 3;1(4):045024. Available from: https://arxiv.org/abs/1905.13741#:~:text=version%2C%20v2)%5D.
[25] Popova M, Isayev O, Tropsha A. Deep reinforcement learning for de novo drug design. Sci Adv 2018 Jul;4(7):eaap7885. Available from: https://advances.sciencemag.org/content/4/7/eaap7885.
[26] Larroque-Lombard Anne-Laure, Ning N, Rao S, Lauwagie S, Halaoui Ruba, Coudray Laëtitia, et al. Biological effects of AL622, a molecule rationally designed to release an EGFR and a c-src kinase inhibitor. Chem Biol Drug Des 2012 Sep 3;80(6):981–91.
[27] Ai C, Yang H, Liu X, Dong R, Ding Y, Guo F. MTMol-GPT: de novo multi-target molecular generation with transformer-based generative adversarial imitation learning. PLoS Comput Biol 2024 Jun 26;20(6):e1012229. 9.
[28] Bassani D, Moro S. Past, present, and future perspectives on computer-aided drug design methodologies. Molecules 2023 May 5;28(9). 3906-6.
[29] Edwards CK, Naik A, Khot Tushar, Burke MD, Ji H, Hope T. SynerGPT: in-context learning for personalized drug synergy prediction and drug design. 2023. https://doi.org/10.1101/2023.07.06.547759.
[30] Wang Y, Zhao H, Sciabola S, Wang W. cMolGPT: a conditional generative pre-trained transformer for target-specific de novo molecular generation. Molecules 2023 May 30;28(11). 4430–0.
[31] Ye G. De novo drug design as GPT language modeling: large chemistry models with supervised and reinforcement learning. J Comput Aided Mol Des 2024 Apr 22;38(1).
[32] Skinnider MA, Greg Stacey R, Wishart DS, Foster LJ. Chemical language models enable navigation in sparsely populated chemical space. Nat Mach Intell 2021 Jul 19;3(9):759–70.
[33] Li Y, Gao C, Song X, Wang X, Xu Y, Han S. DrugGPT: a GPT-based strategy for designing potential ligands targeting specific proteins. 2023. https://doi.org/10.1101/2023.06.29.543848.
[34] Atz K, Cotos L, Isert C, Håkansson M, Focht D, Hilleke M, et al. Prospective de novo drug design with deep interactome learning. Nature Commun 2024 Apr 22;15(1):3408 [cited 2024 May 14];Available from: https://www.nature.com/articles/s41467-024-47613-w.
[35] Linden J, Koch-Nolte F, Dahl G. Purine release, metabolism, and signaling in the inflammatory response. Annu Rev Immunol 2019 Apr 26;37(1):325–47.
[36] US20230034559A1 - automated prediction of clinical trial outcome - Google Patents [Internet]. Google.com. 2022 [cited 2025 Jan 30]. Available from: https://patents.google.com/patent/US20230034559A1/en.
[37] Li D, Wu L, Zhang M, Svitlana Shpyleva, Lin YC, Huang HY, et al. Assessing the performance of large language models in literature screening for pharmacovigilance: a comparative study. Frontiers in Drug Safety and Regulation 2024 Jun 27:4.
[38] Sai Prakash MV, Parab G, Veeramalla M, Reddy S, V V, Gopalakrishnan S, et al. Accelerating pharmacovigilance using large Language Models. In: Proceedings of the 17th ACM international conference on web search and data mining; 2024 Mar 4. p. 1182–3.
[39] Liu X, Tien C, Ding P, Jiang S, Stevens RL. Entropy-reinforced planning with large Language Models for drug discovery. Proc Mach Learn Res 2024;235:31917–32. https://doi.org/10.48550/arXiv.2406.07025.
[40] Tang X, Dai H, Knight E, Wu F, Li Y, Li T, et al. A survey of generative AI for *de novo* drug design: new frontiers in molecule and protein generation. Briefings Bioinf 2024 May 23;25(4):bbae338.
[41] Xie Z, Evangelopoulos Xenophon, Omar ÖH, Troisi A, Cooper AI, Chen L. Fine-tuning GPT-3 for machine learning electronic and functional properties of organic molecules. Chem Sci 2024 Jan 1;15(2):500–10.
[42] Liao C, Yu Y, Mei Y, Wei Y. From words to molecules: a survey of large Language Models in chemistry. 2024. https://doi.org/10.48550/arXiv.2402.01439.
[43] Sharma G, Thakur A. ChatGPT in Drug Discovery 2023. https://doi.org/10.26434/CHEMRXIV-2023-QGS3K.
[44] Jayatunga M KP, Ayers M, Bruens L, Jayanth D, Meier C. How successful are AI-discovered drugs in clinical trials? A first analysis and emerging lessons. Drug Discovery Today 2024 Jun 1;29(6):104009. Available from: https://www.sciencedirect.com/science/article/pii/S135964462400134X.
[45] Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J. Clinical development success rates for investigational drugs. Nature Biotechnol 2014 Jan;32(1):40–51. Available from:http://diyhpl.us/~nmz787/pdf/Clinical_development_success_rates_for_investigational_drugs.pdf.

[46] Aittokallio T. What are the current challenges for machine learning in drug discovery and repurposing? Expet Opin Drug Discov 2022 Mar 8:1–3.
[47] Gangwal Amit, Ansari A, Ahmad I, Kalam Azad Abul, Kumarasamy Vinoth, Subramaniyan Vetriselvan, et al. Generative artificial intelligence in drug discovery: basic framework, recent advances, challenges, and opportunities. Front Pharmacol 2024 Feb 7:15.
[48] Öztürk H, Özgür A, Schwaller P, Laino T, Ozkirimli E. Exploring chemical space using natural language processing methodologies for drug discovery. Drug Discovery Today 2020 Apr 1;25(4):689–705 [cited 2021 Nov 19];Available from: https://www.sciencedirect.com/science/article/abs/pii/S1359644620300465.
[49] Payne J, Srouji M, Yap DA, Kosaraju Vineet. BERT learns (and teaches) chemistry. 2020. https://doi.org/10.48550/arXiv.2007.16012.
[50] Mu Y, He D. The potential applications and challenges of ChatGPT in the medical field. Int J General Med 2024 Mar 5;17:817–26.
[51] Grisoni F. Chemical language models for de novo drug design: Challenges and opportunities. Curr Opin Struct Biol 2023 Apr;79:102527.
[52] Hadi MU, Tashi QA, Qureshi R, Shah A, Muneer A, Irfan M, et al. A survey on large Language Models: applications, challenges, limitations, and practical usage [internet]. www.techrxiv.org; 2023.
[53] Bitri R, Ali M. A comparative review of GPT-4's applications in medicine and high decision making, vol. 14; 2023 Dec. p. 61–8.
[54] Hasselgren Catrin, Oprea TI. Artificial intelligence for drug discovery: are we there yet? arXiv. Cornell University; 2023 Jul 12.
[55] Zhang Y, Liu C, Liu M, Liu T, Lin H, Huang CB, et al. Attention is all you need: utilizing attention in AI-enabled drug discovery. Briefings in Bioinformatics 2024 Jan 6;25(1). Available from: https://academic.oup.com/bib/article/25/1/bbad467/7512647?login=false#434882392.
[56] Rane NS. Contribution and challenges of ChatGPT and similar generative artificial intelligence in biochemistry, genetics and molecular biology. SSRN Electronic Journal 2023. https://doi.org/10.2139/SSRN.4603219.
[57] Kim H, Kim E, Lee I, Bae B, Park M, Nam H. Artificial intelligence in drug discovery: a comprehensive review of data-driven and machine learning approaches. Biotechnol Bioproc Eng 2020 Dec;25(6):895–930.
[58] Patel L, Shukla T, Huang X, Ussery DW, Wang S. Machine learning methods in drug discovery. Molecules 2020 Nov 12;25(22):5277.
[59] Güvenç Paltun Betül, Mamitsuka Hiroshi, Kaski S. Improving drug response prediction by integrating multiple data sources: matrix factorization, kernel and network-based approaches. Briefings Bioinf 2019 Nov 11;22(1):346–59.
[60] Wysocka M, Wysocki O, Delmas M, Mutel V, Freitas A. Large Language Models, scientific knowledge and factuality: a framework to streamline human expert evaluation. J Biomed Informatics 2024 Sep 1. 104724-4. [cited 2025 Jan 30]; Available from: http://arxiv.org/abs/2305.17819.
[61] Guo T, Guo K, Nan B, Liang Z, Guo Z, Chawla NV, et al. What can Large Language Models do in chemistry? A comprehensive benchmark on eight tasks. Adv Neural Inf Process Syst 2023;36. https://doi.org/10.48550/arXiv.2305.18365.
[62] Bender A, Cortés-Ciriano I. Artificial intelligence in drug discovery: what is realistic, what are illusions? Part 1: ways to make an impact, and why we are not there yet. Drug Discov Today 2020 Dec;26(2).
[63] Alshehri AS, Gani R, You F. Deep learning and knowledge-based methods for computer-aided molecular design—toward a unified approach: state-of-the-art and future directions. Comput Chem Eng 2020 Oct;141:107005.
[64] Volkan Ucak Umit, Ashyrmamatov Islambek, Lee J. Improving the quality of chemical language model outcomes with atom-in-SMILES tokenization. J Cheminf 2023 May 29;15(1).
[65] Katz S, Belinkov Y. VISIT: visualizing and interpreting the semantic information flow of transformers. 2023. https://doi.org/10.48550/arXiv.2305.13417.
[66] Bahnsen FH, Berger BJ, Fey G. GLRP: guided by layer-wise relevance propagation – selecting crucial neurons in artificial. Neural Netw 2023 Jun;28:1–6.
[67] Himanshu Giroh, Kumar V, Singh G. Improving the performance of hybrid models using machine learning and optimization techniques. Int J Membrane Sci Technol 2023 Aug 10;10(2):3396–409.
[68] Massella M, Dri DA, Gramaglia D. Regulatory considerations on the use of machine learning based tools in clinical trials. Health and Technol 2022;12(6):1085–96. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9638313/.
[69] Zheng Y, Koh HY, Ju J, Anh May LT, Webb GI, et al. Large Language models for scientific synthesis, inference and explanation. 2023. https://doi.org/10.48550/arXiv.2310.07984.
[70] Zhanshun Wei Jason, Wang X, Schuurmans D, Bosma M, Chi EH, Le QV, et al. Chain-of-Thought prompting elicits reasoning in large Language Models. 2022 Jan 27.
[71] Rohanian O, Nouriborji M, Kouchaki S, Clifton DA. On the effectiveness of compact biomedical transformers. Bioinformatics 2023;39. https://doi.org/10.1093/bioinformatics/btad103.
[72] Lopez Garcia A, De Lucas JM, Antonacci M, Zu Castell W, David M, Hardt M, et al. A cloud-based framework for machine learning workloads and applications. IEEE Access 2020;8:18681–92.
[73] Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. Frontiers in Artificial Intelligence 2023 May 4;6. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10192861/.
[74] Li F, Ruijs N, Lu Y. Ethics & AI: a systematic review on ethical concerns and related strategies for designing with AI in healthcare. AI [Internet] 2022 Dec 31;4(1):28–53. Available from: https://www.mdpi.com/2673-2688/4/1/3.
[75] Tian T, Li S, Zhang Z, Chen L, Zou Z, Zhao D, et al. Benchmarking compound activity prediction for real-world drug discovery applications. Commun Chem 2024 Jun 4;7(1) [cited 2024 Oct 17];Available from: https://www.nature.com/articles/s42004-024-01204-4.
[76] Zhang Z, Bian Y, Xie A, Han P, Huang LK, Zhou S. Can pre-trained models really learn better molecular representations for AI-aided drug discovery? J Chem Inf Model 2022;64:2921–30. https://doi.org/10.1021/acs.jcim.3c01707.
[77] Liu S, Wang J, Yang Y, Wang C, Liu L, Guo H, et al. ChatGPT-powered conversational drug editing using retrieval and domain feedback. 2023. https://doi.org/10.48550/arXiv.2305.18090.
[78] Gupta A, Zhou HX. Machine learning-enabled pipeline for large-scale virtual drug screening. J Chem Inf Model 2021 Aug 17;61(9):4236–44.
[79] Vora LK, Gholap AD, Jetha K, Thakur RRS, Solanki HK, Chavda VP. Artificial intelligence in pharmaceutical technology and drug delivery design. Pharmaceutics 2023 Jul 10;15(7). 1916-6, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10385763/.
[80] Buntz B. Sanofi teams with OpenAI for AI-powered drug discovery drive [Internet]. Drug Discovery and Development; 2024. Available from: https://www.drugdiscoverytrends.com/sanofi-openai-partnership-accelerates-ai-drug-discover/.
[81] Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, et al. How to improve R&D productivity: the pharmaceutical industry's grand challenge. Nat Rev Drug Discov 2010 Feb 19;9(3):203–14.
[82] Wong CH, Siah KW, Lo AW. Estimation of clinical trial success rates and related parameters. Biostatistics 2018 Jan 31;20(2):273–86. Available from: https://academic.oup.com/biostatistics/article/20/2/273/4817524#131976077.
[83] Achieving R&D excellence to shape the medicines of tomorrow l Sanofi. https://www.sanofi.com/en/magazine/our-science/achieving-research-development-excellence-shape-medicines-tomorrow. [Accessed 3 March 2025].
[84] Mukherjee A, Abraham S, Singh A, Balaji S, Mukunthan KS. From data to cure: a comprehensive exploration of multi-omics data analysis for targeted therapies. Mol Biotechnol 2024 Apr;2.
[85] Chakraborty C, Bhattacharya M, Pal S, Lee S. From machine learning to deep learning: an advances of the recent data-driven paradigm shift in medicine and healthcare. Current Res Biotechnol 2023 Nov 1;7. 100164-4.
[86] Hayawi Kadhim, Shahriar Sakib, Alashwal Hany, Mohamed Adel Serhani. Generative AI and large language models: a new frontier in reverse vaccinology. Inform Med Unlocked 2024 Jan 1;48. 101533-3.
[87] Patel AU, Gu Q, Esper R, Maeser D, Maeser N. The crucial role of interdisciplinary conferences in advancing explainable AI in Healthcare. BioMedInformatics 2024;4:1363–83. https://doi.org/10.3390/biomedinformatics4020075.
[88] Wang H. Prediction of protein–ligand binding affinity via deep learning models. Briefings in Bioinformatics 2024 Jan 22;25(2). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10939342/.
[89] Dhakal A, McKay C, Tanner JJ, Cheng J. Artificial intelligence in the prediction of protein–ligand interactions: recent advances and future directions. Briefings Bioinf 2021 Nov 27;23(1).
[90] Tan RK, Liu Y, Xie L. Reinforcement learning for systems pharmacology-oriented and personalized drug design. Expet Opin Drug Discov 2022 Aug 3;17(8):849–63.
[91] National Artificial Intelligence Research Resource Pilot | NSF - National Science Foundation. 2024. https://new.nsf.gov/focus-areas/artificial-intelligence/nairr. [Accessed 3 March 2025].
[92] Paul M, Maglaras L, Ferrag MA, Almomani I. Digitization of healthcare sector: a study on privacy and security concerns. ICT Express [Internet] 2023;9(4):571–88. Available from: https://www.sciencedirect.com/science/article/pii/S2405959523000243.
[93] Alowais SA, Alghamdi SS, Alsuhebany N, Alqahtani T, Alshaya A, Almohareb SN, et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. BMC Med Education 2023 Sep 22;23(1). Available from: https://bmcmededuc.biomedcentral.com/articles/10.1186/s12909-023-04698-z.
[94] Sankar BS, Gilliland D, Rincon J, Henning Hermjakob, Yan Y, Adam I, et al. Building an ethical and trustworthy biomedical AI ecosystem for the translational and clinical integration of foundation models. Bioengineering 2024 Sep 29;11(10). 984-4. [cited 2024 Nov 3];Available from: https://www.mdpi.com/2306-5354/11/10/984.
[95] Dhudum Rushikesh, Ankit Ganeshpurkar, Pawar A. Revolutionizing drug discovery: a comprehensive review of AI applications. Drugs and Drug Candidates 2024 Feb 13;3(1):148–71.
[96] Jeyaraman M, Balaji S, Jeyaraman N, Yadav S. Unraveling the ethical enigma: artificial intelligence in healthcare. Cureus 2023 Aug 10;15(8). Available from: https://www.cureus.com/articles/178557-unraveling-the-ethical-enigma-artificial-intelligence-in-healthcare.pdf.
[97] Wang C, Liu S, Yang H, Guo J, Wu Y, Liu J. Ethical considerations of using ChatGPT in health care. J Med Internet Res 2023 Aug 11;25(1):e48009. Available from: https://www.jmir.org/2023/1/e48009/.