Revolutionizing Pharmaceuticals: Generative Artificial Intelligence as a bibliographic assistant

Ángel Canal-Alonso1, Noelia Egido1, Pedro Jiménez1, Juan Manuel Corchado

[1]Department of Bioinformatics and Computational Biology, AIR Institute, Carbajosa de la Sagrada, Spain

Email: acanal@air-institute.com

Abstract

Artificial Generative Intelligence (AGI) has exploded into biomedical and pharmaceutical research, fundamentally transforming the way scientists approach literature review, experiment design, and reagent and antibody selection. This article explores how IAG, supported by advanced machine learning and natural language processing models, has revolutionized these processes.

The IAG streamlines literature review, extracting relevant information, identifying emerging patterns and trends in the scientific literature, and generating innovative hypotheses. It also acts as an advanced search tool, allowing researchers to quickly access accurate information in an ocean of data.

A prominent example of this application is BenchSci, a platform that uses the IAG to recommend reagents and antibodies based on real experimental data and scientific literature. This integration of IAG into experimental design promises to accelerate research, reduce costs, and improve the precision of experiments.

Together, the IAG is presented as a catalyst for discoveries in pharmaceutical and biomedical research, offering unprecedented potential to advance the understanding and treatment of diseases, and improve decision-making in the industry.

Keywords: Generative Artificial Intelligence, Drug Design, Literature Review

1. Introduction

The Artificial Intelligence (AI) revolution has transformed countless fields of study and applications in recent decades. Among its many branches, Artificial Generative Intelligence (AGI) has emerged as one of the most fascinating and promising. IAG refers to AI systems that can produce content that is new, original and, in many cases, indistinguishable from that created by humans. This ability to generate content manifests itself in multiple ways: from the creation of images and music, to the generation of text and the design of molecules for drugs.

The birth of IAG can be traced to early efforts in machine learning, where algorithms began to identify and replicate patterns from data. However, it was with the advent of neural networks, and more specifically generative adversarial networks (GANs) in 2014, that the field really began to take shape. GANs, proposed by Ian Goodfellow and his colleagues, operate with two neural networks that work together: a generating network that produces images and a discriminator network that evaluates those images. These networks "compete" with each other in an iterative process until the generating network creates images that the discriminator cannot distinguish from real images.

Since then, the capacity and applicability of the IAG has grown exponentially. Its applications are not limited to the generation of multimedia content. In the field of biology and medicine, IAG has proven to be a valuable tool in the design and discovery of new drugs. Faced with the challenge of analyzing huge libraries of chemical compounds and discerning which ones might have desirable therapeutic properties, researchers have found in the IAG a powerful ally to optimize and accelerate this process.

In this article, we will explore how Generative Artificial Intelligence has positioned itself as an essential tool in literature reviews for drug design, outlining its impact and potential in this crucial field of research.

What is Generative Artificial Intelligence and how does it work?

Generative Artificial Intelligence (GAI) is a subfield of machine learning that focuses on the development of algorithms capable of generating new and original data that reflect similar characteristics to the input data with which they were trained. Unlike other AI models that are designed to make predictions or classifications, IAG seeks to create.

Below, we describe the main characteristics and operating mechanisms of the IAG:

1. Generative Models: These models are trained to understand how the data is distributed. Once they have "learned" this distribution, they can generate new examples. A common generative model is the Autoencoder, which compresses the input into a more compact representation and then reconstructs the input from that representation.

2. Generative Adversarial Networks (GANs): GANs consist of two neural networks: the generator and the discriminator. The first tries to create data, while the second tries to distinguish between real data and generated data. During training, these networks "compete" in a game of cat and mouse until the generator produces data so convincing that the discriminator can barely differentiate it from real data.

3. Latent Space Models: These models represent data in a reduced dimensional space, called "latent space." The idea is that the essential variations and characteristics of the data can be represented in this reduced space. When new data needs to be generated, the model takes points in this latent space and "translates" them back to the original data space.

4. Training: The key to obtaining effective generative models lies in training. Using large amounts of data and constant feedback, models adjust their parameters to produce increasingly realistic results. With sufficient training and appropriate data, a generative model can generate high-quality content that is very similar to the real thing.

5. Applications in various domains: Although image generation is one of the most popular applications of the IAG, its usefulness spans domains as diverse as text generation, music creation, video game design and, as we mentioned above, design and discovery of new drugs.

In short, Generative Artificial Intelligence uses advanced machine learning models to “learn” intrinsic patterns and distributions from data sets and subsequently uses that learning to create new data that reflects the characteristics of the original data. It is this ability to create that gives it vast potential in multiple fields and applications.

The importance of literature reviews

Literature reviews in the pharmaceutical industry, especially in drug design, are of utmost importance and serve multiple essential purposes in the development and approval process of new drugs.

Firstly, the process of drug discovery and design is an intricate and complex one, involving the interaction and manipulation of molecules at microscopic levels. Before investing significant resources in the research and development (R&D) of a new drug, it is vital to ensure that the chosen direction is informed by previous scientific data and findings. Literature reviews provide a comprehensive compendium of these findings, helping researchers better understand the context and scientific basis of their efforts.

Furthermore, drug design is not only about the effectiveness of the proposed drug but also its safety. Knowing potential side effects, interactions with other drugs, and mechanisms of action and resistance are crucial. Literature reviews allow scientists to access prior information on similar or related compounds, providing valuable insight into potential risks and opportunities.

On the other hand, in a field that advances as quickly as pharmacology and biomedicine, it is essential to stay updated. Literature reviews provide a panoramic view of recent advances, emerging techniques, and key discoveries. This way, R&D teams can quickly incorporate innovations and adjust their strategies based on current trends and discoveries.

Finally, regulations in the pharmaceutical industry are rigorous and require that any claims about a drug's effectiveness and safety be supported by robust scientific evidence. Literature reviews help consolidate this evidence and facilitate the communication and justification process before regulatory bodies.

For all these reasons, literature reviews are a cornerstone in the pharmaceutical industry. They not only guide and inform research, but also ensure that the medicines developed are both effective and safe for patients.

2. Direct applications

Automated text analysis and summaries

Generative Artificial Intelligence (AI) has experienced impressive development in recent years, especially with the emergence of models based on Transformers. These models have revolutionized the field of natural language processing (NLP) due to their ability to manage and understand textual information in an unprecedented way.

Transformers, a specific architecture within NLP, are characterized by their ability to handle long sequences of data and establish relationships between different parts of a text, no matter how distant they are from each other. This ability makes them uniquely suited to analyzing the vast and complex biomedical literature, which often features dense, technical texts replete with conceptual interrelationships.

Given the avalanche of publications in the biomedical field, Transformers-based models can play a vital role in digesting and synthesizing this information. They are capable of automatically summarizing scientific articles, distilling essential content and eliminating redundancy. This automation not only facilitates access to key information, but also allows researchers to address a greater number of documents in less time.

Furthermore, these advanced models can extract specific and crucial details, such as experimental data, results and conclusions, identifying what is truly relevant to the research at hand. Its ability to perform deep text analysis allows you to identify key topics, detect emerging trends in research, and establish relationships between different concepts, molecules, or biological mechanisms.

As if that were not enough, by integrating these AI models with databases and other search tools, complex queries can be performed and specific answers obtained, making it easier to identify gaps in knowledge or areas of opportunity for future research.

In essence, generative AI, particularly Transformers-based models, is presented as a bridge between information overload and the need for deep understanding in the biomedical field. These tools not only streamline the literature review process, but also enhance researchers' ability to gain insights and advance the complex landscape of drug design and study.

Generation of new hypotheses

Generative Artificial Intelligence has gone beyond simply being a tool to analyze and summarize scientific literature. Its potential lies in its ability to merge and recombine information in unexpected ways, opening doors to new perspectives and understandings in biomedical research.

Diving into the vastness of scientific literature, generative AI models do not simply interpret content in a linear or direct manner. Instead, they "read" and process thousands of articles, understanding the intricate networks of information that span diverse fields of study. This vast knowledge allows them to detect patterns, analogies and correlations that, due to the vastness and complexity of the data, could go unnoticed by human researchers.

From this processing, generative AI has the ability to generate new hypotheses, many of which may be counterintuitive or innovative. For example, a model could identify indirect relationships between a specific gene and a rare disease, based on a series of seemingly unrelated studies. Or it could propose a potential interaction between a protein and a drug, suggesting a new therapeutic application for a previously approved drug.

These automatically generated hypotheses can be extremely valuable, challenging and expanding the boundaries of current knowledge. They can inspire researchers to explore previously unconsidered areas and validate AI proposals with laboratory experiments. This feedback loop between AI and experimental research can significantly accelerate the process of scientific discovery.

Furthermore, in a world where multidisciplinarity is becoming the norm, generative models can act as bridges between different fields of knowledge, fusing information from disciplines such as genomics, proteomics, pharmacology and epidemiology. This holistic integration can unleash a new era of interdisciplinary discovery, where solutions to complex problems emerge from the intersection of multiple areas of expertise.

In short, generative AI is not only a tool for the digestion and understanding of scientific literature, but it is also a source of inspiration and a catalyst for new lines of research, which can lead to significant discoveries that transform the way we that we understand and address diseases and their treatment.

Pattern and trend detection

Generative Artificial Intelligence, equipped with advanced machine learning and natural language processing (NLP) techniques, represents one of the most revolutionary tools for the analysis and interpretation of scientific literature in the contemporary era. Its ability to delve into vast databases of publications and extract essential insights transcends mere automation, offering a window into the depths of accumulated knowledge.

Science, in its nature, is incremental and evolutionary, with findings that build on previous work. However, the colossal volume of information published daily can make it difficult to see subtle links between investigations or the emergence of patterns over time. Generative AI, with its computational power and sophisticated algorithms, can navigate this ocean of data, identifying trends and patterns that would perhaps remain hidden to the human eye.

A critical area is the identification of emerging therapies. While researchers can stay abreast of advances in their specific fields, AI can connect dots between disciplines, recognizing potential therapeutic applications of compounds or techniques that were not initially intended for that purpose. This ability to "think outside the box" is essential in a field where drug repurposing and multidisciplinarity are increasingly common.

Early detection of unexpected side effects is another invaluable benefit. Through analysis of scientific literature, case reports, clinical studies, and other resources, generative AI can alert you to potential adverse effects of a medication before they become bigger problems. These early warnings can lead to deeper reviews, dosage modifications, or even withdrawal of a drug from the market, thus protecting patients' health.

Finally, in the field of public health, the ability of AI to identify risk factors is crucial. It can recognize correlations between demographic, environmental, genetic or behavioral variables and the prevalence or severity of certain diseases. These insights can guide health policies, prevention campaigns and intervention strategies.

For decision makers in the pharmaceutical industry, generative AI is a compass in an ever-evolving landscape. It offers a clear and up-to-date view of trends, risks and opportunities, ensuring that decisions are based on the most complete and up-to-date knowledge available. In an industry where the stakes are high and the margin for error is small, the contribution of generative AI is invaluable.

Assistance in searching for information

The digital revolution has led to an explosion in the amount of information available to researchers, with an avalanche of articles, reports and studies published daily. In this vast sea of data, finding specific and relevant information can be comparable to searching for a needle in a haystack. This is where generative Artificial Intelligence presents itself as an invaluable ally for the scientific community.

Equipped with cutting-edge machine learning and natural language processing techniques, generative AI rises above traditional search tools. While conventional search engines rely heavily on keywords and simple algorithms, generative models have the ability to understand the context and semantics behind queries, allowing for a much richer and more nuanced interpretation of user requests.

This means that, rather than relying solely on specific terms, researchers can ask complex queries, asking questions or describing research scenarios and objectives. For example, instead of simply searching for "effects of drug X," a researcher could query "interactions between drug X and disease-associated proteins Y." The generative model, understanding the nature and scope of the query, will scour the literature to find documents that address that specific interaction.

Additionally, generative AI can rank and prioritize results based on relevance, impact, and recency, presenting researchers with a curated set of documents that closely aligns with their needs. This eliminates the tedious task of manually filtering large amounts of information, allowing scientists to focus on what really matters: data analysis and interpretation.

It is also worth mentioning that, over time, these models can adapt and learn from user preferences and behaviors, further improving the accuracy and relevance of their results.

In short, generative AI, by acting as an advanced search tool, not only democratizes access to scientific information, but also streamlines and optimizes the research process. In a world where time is a precious resource and speed of innovation is essential, this technology promises to be a key piece in the future of scientific research.

Application of generative AI to assist in the design of experiments and selection of reagents and antibodies

Experimental design and selection of appropriate reagents and antibodies are crucial steps in biomedical and pharmacological research. Choosing the correct reagents and antibodies can significantly influence the validity and reproducibility of an experiment. However, with the vast number of reagents and antibodies available on the market, identifying the most suitable one for a particular experiment can be a daunting task.

This is where generative Artificial Intelligence comes in, transforming and optimizing this selection process. Using advanced machine learning techniques and big data analysis, generative AI can help researchers identify reagents and antibodies that have been shown to be effective and specific in previous similar experiments.

BenchSci Example

A prominent example of the application of AI in this area is BenchSci. This platform uses Artificial Intelligence to decode and structure complex information from millions of antibody technical sheets and research papers. BenchSci's system allows users to enter parameters specific to their experiment, such as tissue type or application (e.g., immunohistochemistry or Western blot), and receive antibody recommendations based on actual experimental data.

The magic behind BenchSci is its ability to extract and understand information from scientific publications, identifying in which contexts a particular antibody has been used and with what degree of success. This eliminates much of the guess work in antibody selection, and reduces the likelihood of investing in antibodies that will not perform as expected.

In addition to providing literature-based recommendations, BenchSci also offers visual product comparisons, showing experimental images of actual antibody use. This allows researchers to make an informed assessment of the efficacy and specificity of the antibody in question.

In summary, the integration of generative Artificial Intelligence in the process of experimental design and selection of reagents and antibodies promises to accelerate research, reduce costs and improve the reproducibility and precision of experiments. Tools like BenchSci demonstrate how AI can act as a powerful ally for researchers, helping them make informed decisions and advance their work with confidence.

Conclusions

Artificial Generative Intelligence (AGI) has emerged as a transformative tool in the field of pharmaceutical and biomedical research. Its ability to analyze, understand and generate content based on the vast scientific literature offers significant benefits that are redefining the way drug design challenges and decision-making are addressed in the industry.

First, the IAG has demonstrated its ability to accelerate and optimize the literature review process. With the ability to summarize, extract relevant information, and detect patterns in large volumes of scientific literature, researchers can gain deeper and faster insight into previous research, allowing them to make informed decisions and reduce time spent manually searching. .

Furthermore, the IAG has revealed itself as a tireless generator of new hypotheses and perspectives in biomedical research. By analyzing multiple studies and correlating data in innovative ways, this technology can inspire entirely new lines of research and significant discoveries. Its ability to detect complex relationships between genes, proteins, diseases and drugs has immense potential in identifying new therapies and understanding the molecular bases of diseases.

Generative AI has also proven to be an invaluable ally in the early detection of unexpected side effects and in the identification of risk factors in public health. This ability to foresee potential problems can save lives and improve the safety of drug treatments.

Finally, in assisting experiment design and reagent and antibody selection, generative AI, as exemplified by BenchSci, demonstrates its ability to simplify critical processes, save time and resources, and ensure the reproducibility of experiments.

In summary, generative AI has established itself as a multidimensional and powerful tool that drives pharmaceutical and biomedical research. As technology continues to advance and become further integrated into decision-making and drug design, we will hopefully accelerate the pace of discoveries and advances in the field, improving people's quality of life and the effectiveness of medical treatments.

References

Garcia-Retuerta D, Canal-Alonso A, Casado-Vara R, Rey AM, Panuccio G, Corchado JM. Bidirectional-Pass Algorithm for Interictal Event Detection. In Practical Applications of Computational Biology & Bioinformatics, 14th International Conference (PACBB 2020). PACBB 2020. Advances in Intelligent Systems and Computing, vol 1240. Springer, Cham. https://doi.org/10.1007/978-3-030-54568-0_20

Castillo Ossa LF, Chamoso P, Arango-López J, Pinto-Santos F, Isaza GA, Santa-Cruz-González C, Ceballos-Marquez A, Hernández G, Corchado JM. A Hybrid Model for COVID-19 Monitoring and Prediction. Electronics. 2021; 10(7):799.
https://doi.org/10.3390/electronics10070799

Intelligent Platform Based on Smart PPE for Safety in Workplaces. Márquez-Sánchez S, Campero-Jurado I, Herrera-Santos J, Rodríguez S, Corchado JM. Sensors (Basel). 2021 Jul 7;21(14):4652
https://doi.org/10.3390/s21144652

A. Canal-Alonso, R. Casado-Vara and J. Manuel Corchado, "An affordable implantable VNS for use in animal research," 2020 27th IEEE International Conference on Electronics, Circuits and Systems (ICECS), 2020, pp. 1-4,
doi: 10.1109/ICECS49266.2020.9294958

An Agent-Based Clustering Approach for Gene Selection in Gene Expression Microarray. Ramos J, Castellanos-Garzón JA, González-Briones A, de Paz JF, Corchado JM. Interdiscip Sci. 2017 Mar;9(1):1-13
DOI 10.1007/s12539-017-0219-6

Acknowlegments

This study has been funded by the AIR Genomics project (with file number CCTT3/20/SA/0003), through the call 2020 R&D PROJECTS ORIENTED TO THE EXCELLENCE AND COMPETITIVE IMPROVEMENT OF THE CCTT by the Institute of Business Competitiveness of Castilla y León and FEDER funds.