# Biotechnology, Big Data and Artificial Intelligence

A Review Report

COMP 472: Artificial Intelligence

***Submitted To:***
Sanjog Sigdel
COMP 472 Course Instructor
Kathmandu University


***Submitted By:***
Aayush Shrestha
CM IV-I
Roll no- 21


***Submitted On:***
November 2024

## Abstract

The combination of biotechnology, big data, and artificial intelligence (AI) is a significant step in modern science, influencing the course of the Fourth Industrial Revolution. This report studies how artificial intelligence and big data are transforming the sector of biotechnology at different fields, such as agriculture, medicine, animal research, industrial applications, and bioinformatics. AI's pattern recognition, predictive modeling, and process optimization abilities have enabled advancements in drug discovery, genome annotation, and the creation of precision tools such as AlphaFold for protein structure prediction. Key accomplishments highlighted in the paper include AI-driven molecular docking approaches and the use of digital twins for virtual modeling of biological systems. Ethical problems and constraints, such as data privacy, centralization of AI resources, and algorithmic biases, are addressed, highlighting the importance of balanced governance. Overall, this study provides an overview of current advancements and future directions at the intersection of biotechnology and AI.



***Keywords:*** Protein Folding, AlphaFold, Molecular Docking, Digital Twin, Genome

# Introduction

Biotechnology is defined as the use of biological systems to develop or create technologies with the aim to improve our life and condition. This field integrates principles from various scientific domains, including molecular biology, genetics, biochemistry, and engineering, to effectively utilize cellular and biomolecular processes for practical applications [1,2]. In simpler words, Biotechnology is the technology based on and around biology.

Historically, biotechnology has been utilized for thousands of years. The beginning is credited with traditional practices such as fermentation in food production i.e. traditional production of alcohal and the selective breedimg of plants and animals. However, modern biotechnology has evolved significantly with advancements in genetic engineering and molecular techniques that allow for a more precise modifications of organisms[3].

At present, with the growth and commodization of AI and Big Data, the field of Biotechnology has achieved an accelerated pace of advancements. Mainly, the integration of biotechnology, big data, and artificial intelligence (AI) represents a significant aspect of the 4th Industrial Revolution.

The application of big data has enabled the collection and analysis of vast amounts of biological information. With the paper by Researchers at the European Bioinformatics Institute (EMBL-EBI), having monitored the growth of the different types of data stored in their servers, concluding that the volume of data is doubling roughly every year.[3]. AI complements these advancements by providing sophisticated algorithms capable of processing these complex datasets, identifying patterns, and making predictions that were previously computationally off limits.

As these fields continue to evolve, the assmiliation between biotechnology, big data, and AI raises important questions regarding ethical considerations, data privacy, and the implications of automated decision-making in critical areas such as health and food security.

This review aims to explore the current landscape of biotechnology in conjunction with big data analytics and AI applications, highlighting key developments, ongoing research efforts, and potential future directions discussed at present.

# Applications In Biotechnology

The integration of AI and Big Data in Biotechnology can be observed in multiple subsectors of the field, types like agricultural biotechnology, medical biotechnology, animal biotechnology, industrial biotechnology, and bioinformatics[4].

- Agricultural Biotechnology

Agricultural biotechnology develops genetically modified plants to increase crop yields or introduce new characteristics to the existing plants. It involves conventional plant breeding, tissue culture, micropropagation, molecular breeding, and genetic engineering of plants.

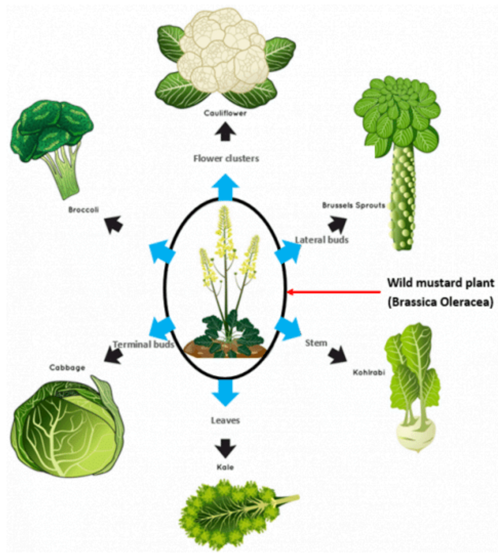


*Fig 1: Selective breeding of mustard plant and its current aftereffects*.

Biotechnology firms are now leveraging Artificial Intelligence and Machine Learning techniques to develop and program autonomous robots that handle important agricultural tasks like harvesting crops at a much faster pace than humans. Computer Vision and Deep Learning algorithms are leveraged to process and analyze the data captured by drones. This helps in monitoring crop and soil health. Machine Learning algorithms help in tracking and predicting various environmental changes like the weather changes that impact the crop yield[5].

- Medical biotechnology

Medical biotechnology uses living cells for the betterment of human health by producing drugs and antibiotics. It also involves the study of DNA and genetically manipulates the cells to increase the production of important and beneficial characteristics.

Drug discovery makes substantial use of machine learning and artificial intelligence. Depending on known target structures, machine learning aids in the discovery of tiny compounds that may have therapeutic advantages. Since machine learning uses actual results to improve diagnostic testing—that is, the more diagnostic tests performed, the more accurate the results—it is frequently utilized in illness diagnosis. AI is also assisting in the reduction of the planning process for radiation therapy, which saves time and enhances patient care. Enhancing EHRs with evidence-based medications and clinical decision support systems is another area where AI and machine learning are showing promise. In addition to the aforementioned uses, these technologies are extensively employed in radiology, customized medicine, gene editing, medication management, and other fields.

- Animal biotechnology

Using molecular biology techniques, this branch genetically engineers or modifies animals to increase their sustainability for use in agriculture, industry, or pharmaceuticals.

One field where machine learning models and artificial intelligence offer useful insights is animal breeding. Selective breeding is a widely used technique in which animals with the best attributes are bred with one another to produce offspring that share those traits. In order to choose and breed certain animals based on their genetic traits, this technique is also carried out at the molecular level. Large genomic data sets are being interpreted and a wide range of genomic sequence elements are being annotated using machine learning.

- Industrial Biotechnology

Industrial biotechnology is all about biopolymers substitutes, the invention in various areas like vehicle parts, fuels, fibers, new chemicals, and the production process.

Internet of Things (IoT), Machine Learning, and Artificial Intelligence analyze the machines, predict outages, optimize equipment, etc to provide efficient production and better product quality. Computer-aided designs and Artificial Intelligence are coming up with the desired molecule design. Robotics and Machine Learning cultivate the strains and test to what extent the desired molecule was reached.


*Fig 2: AI co-created Coke Flavor*

- Bioinformatics

In order to comprehend the biological relevance of a range of data, bioinformatics uses methods from mathematics, computer science, and biology to assist in the collection, storing, processing, distribution, analysis, and interpretation of biochemical and biological information. Large data pools are used to arrange this information.

To obtain remarkable insights, this data must be utilized. Machine learning and artificial intelligence are used in DNA sequencing because of the massive amount of data involved, protein classification and its biological and catalytic roles, gene expression analysis, genome annotation, where some automation is needed to locate genes, computer-aided drug design, etc.

## Key Developments\ Achievements

1. Protien Folding Problem

The protein folding problem involves predicting a protein's three-dimensional structure from its amino acid sequence. The protein-folding problem was first posed about one half-century ago. The term refers to three broad questions:

(i) What is the physical code by which an amino acid sequence dictates a protein's native structure?

(ii) How can proteins fold so fast?

(iii) Can we devise a computer algorithm to predict protein structures from their sequences?

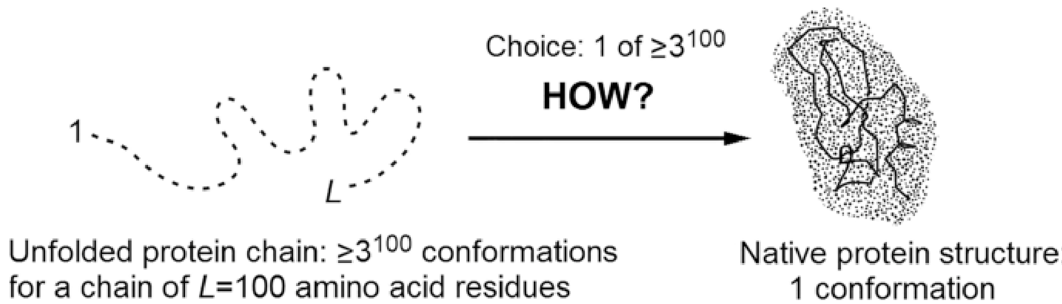


*Fig 3: Protein Folding problem*

This challenge has significant implications for understanding diseases and developing new therapies. Recent advancements in computational methods, particularly through artificial intelligence (AI), have made strides in addressing this problem[6].

2. AlphaFold

Google's AlphaFold has made advancements in the field of structural biology since its introduction in 2020. Developed by DeepMind, AlphaFold utilizes artificial intelligence to predict protein structures with remarkable accuracy. This addresses the "protein folding problem." This breakthrough has not only accelerated research in various biological fields but has also been recognized with prestigious awards, including the 2024 Nobel Prize in Chemistry for its contributions to molecular understanding and drug discovery[7,8,9].

Launched in 2020, AlphaFold 2 demonstrated accuracy in predicting protein structures based on their amino acid sequences. It could forecast protein configurations to near-experimental precision, reducing the time and cost of traditional methods like X-ray crystallography. By making predictions for over 200 million proteins, AlphaFold 2 has become an invaluable resource for researchers globally. [7]


*Fig 4: Google Alphafold*

Recently introduced, AlphaFold 3 expands upon this capability by predicting not only proteins but also other essential biomolecules such as DNA, RNA, and ligands. This model incorporates advanced deep learning techniques and new modules which enhance its predictive power. Early analyses indicate that AlphaFold 3 achieves up to a 50% improvement in accuracy for protein-ligand interactions compared to previous methods[7,8].

3. Docking

Molecular docking is a computational technique used to predict the preferred orientation of a small molecule (ligand) when bound to a protein target[10], which is crucial for understanding biological processes and designing new medicines. The traditional docking process involves searching conformational space to find optimal binding poses, often requiring high computational resources. However, with the help of AI and Big Data, the landscape of molecular docking is rapidly evolving.

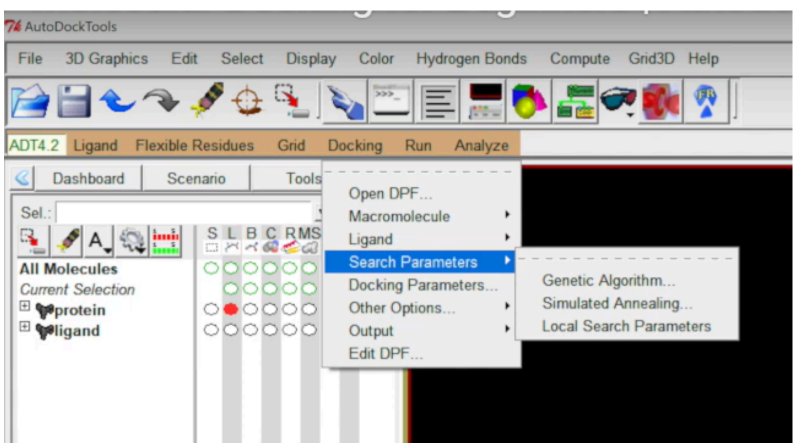


*Fig 5: Snapshot of Autodock, docking search parameters*

One such implementation is the AutoDock Vina software, which employs a variety of stochastic global optimization approache, including genetic algorithms,particle swarm optimization, simulated annealing as part of its optimization strategy for molecular docking. GAs are adaptive heuristic search algorithms inspired by the principles of natural selection and genetics. In AutoDock Vina, these algorithms are utilized to optimize ligand conformations and orientations relative to target proteins through a population-based approach.[11]

There are also paper discussing the application of deep learning in docking, i.e. Deep Docking, an AI-driven approach that accelerates structure-based virtual screening[12]. It allows researchers to dock only a subset of a chemical library while iteratively synchronizing with ligand-based predictions for the remaining scores. This method can achieve up to 100-fold acceleration in screening processes, making it feasible to evaluate billion-sized libraries without excessive computational costs.

4. GMO

GMO refers to the Genetically Modified Organisms and, the integration of big data and artificial intelligence (AI) into the field of genetic modification is a major shift in the development and enhancement of genetically modified organisms. These impact the food and ecosystem preservation at the globa scale.

One of the most significant applications of AI in GMO development is its ability to identify genes associated with desirable traits efficiently. Traditionally, this process involved extensive experimental trials that could take years to yield results. However, AI algorithms can analyze vast datasets containing plant genomes and their corresponding phenotypic traits to quickly pinpoint genes linked to characteristics such as enhanced nutritional content, disease resistance, and stress tolerance[13].

Also, AI technologies complement advanced gene-editing tools like CRISPR-Cas9[14] by enhancing precision in genetic modifications. CRISPR-Cas9 functions as molecular scissors that enable precise insertions, deletions, or modifications of DNA sequences. AI algorithms assist researchers in predicting potential off-target effects and optimizing the design of CRISPR strategies.

5. Digital Twin

A digital twin (DT) is a theoretical representation or virtual simulation of an object or system composed of a computer model and real-time data[15].In biotechnology, digital twins can model biological processes or systems, allowing researchers to predict outcomes and optimize experiments without the need for extensive physical trials.  For instance, AI-enabled digital twins can run concurrent simulations of multiple processes, allowing manufacturers to conduct "what-if" analyses that inform decision-making and improve operational efficiency. These are currently in development stages in various developed nations already in their trail or experimentation phases[16].

6. Perfect Beer

Artificial intelligence (AI) is revolutionizing the brewing industry by enabling brewers to innovate and refine their beer recipes with high precision. By employing algorithms that evaluate consumer feedback to continuously modify recipes, breweries such as IntelligentX have produced the first AI-brewed beer in history. Similar to this, Carlsberg uses AI in their Beer Fingerprinting Project to examine the flavor profiles of hundreds of samples every day. While Champion Brewing Company works with Metis Machine to employ machine learning to design successful IPAs based on market trends, other breweries, like St Austell Brewery with its Hand Brewed by Robots program, show how AI can inspire creativity in recipe development. Furthermore, initiatives like RoboBEER train AI models to forecast customer preferences in order to optimize particular sensory aspects, such foam quality[17,18,19].

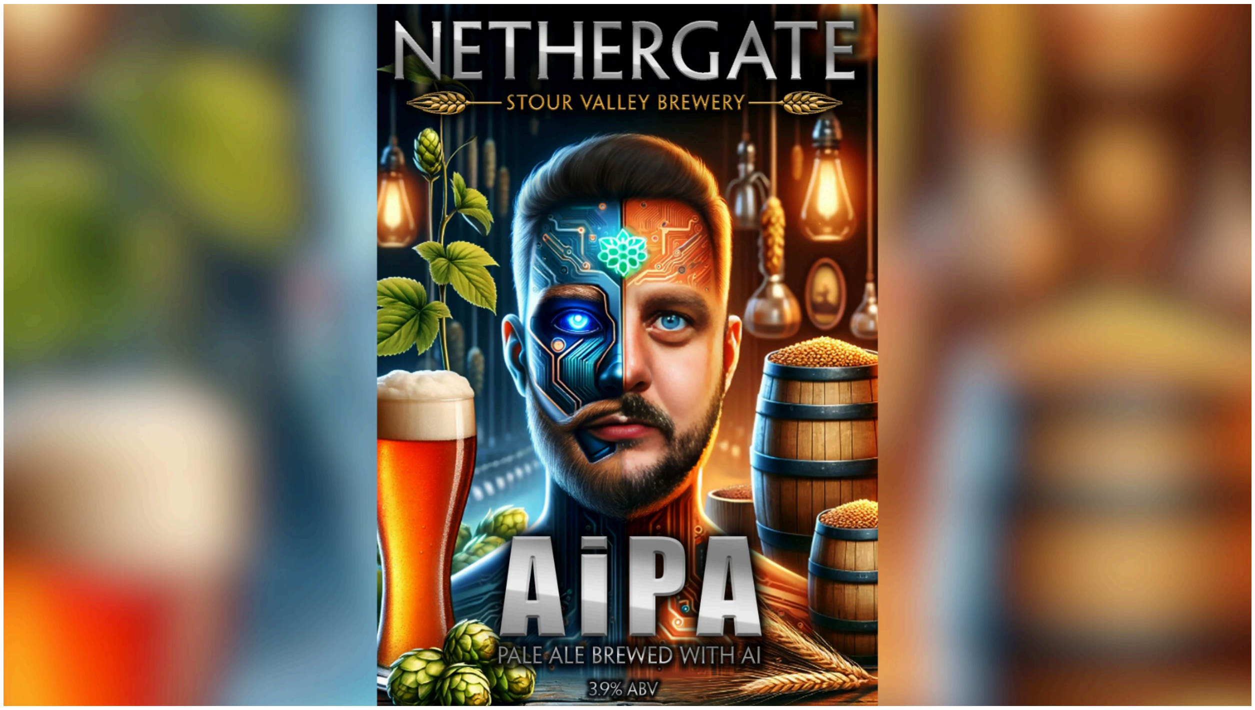


*Fig 6: AI generated poster for AI brewed Beer*

7. Nobel Prize Winner
   "In a second Nobel win for AI, the Royal Swedish Academy of Sciences has awarded half the 2024 prize in chemistry to Demis Hassabis, the cofounder and CEO of Google DeepMind, and John M. Jumper, a director at the same company, for their work on using artificial intelligence to predict the structures of proteins. The other half goes to David Baker, a professor of biochemistry at the University of Washington, for his work on computational protein design. The winners will share a prize pot of 11 million Swedish kronor ($1 million). " was the exact wording used in the MIT Technology Review[20]

This achievement marks a significant milestone, as it is one of the first instances where an AI-driven solution has led to a Nobel Prize for their on AlphaFold.

## Challenges\ Limitation

Integration of Artificial Intelligence in Biotechnology opens many doors of innovation and efficiency, but equally presents as many challenges and limitations.
One of these will be the centralization of computational power which normally would mean unequal access to AI technologies among many parties, especially among large corporations and well-endowed research institutions, leaving smaller players and under-endowed organizations without steam.

Another significant challenge is the misuse of information, as a lot of data is a prerequisite for effective working of AI systems, the possibility of data misappropriation or bad management becomes more likely.

Given the nature of biological and health data, ethical concerns regarding privacy and informed consent as well as discrimination based on genetic information arise.

In addition, many important ethical issues arising from the deployment of new AI applications in biotechnology can't be captured easily because of their complexity and multi-dimensionality. Some of those include algorithmic bias, transparency, and accountability. Thus, to make fair and avoid operations of AI systems that perpetuate already-existing inequalities, as much as possible, all these have to be given consideration. This is because technological advancement always outpaces the development of appropriate laws.

## Discussions

AI has a lot of potential to transform biotechnology because it can improve research, optimize processes, and personalize medicine and it is evident to the progress made in current time; however, its limitations or challenges need to become well understood. The ethics and misuse of these advancements are a challenge where effort needs to be poured by the lot to discuss the ethics of data use and establish good governance .

## References

[01] *What is Biotechnology? | BIO*. (2024). Bio.org.
https://www.bio.org/what-biotechnology

[02] Barney, N., & Lewis, S. (2022). biotechnology (biotech). WhatIs; TechTarget.
https://www.techtarget.com/whatis/definition/biotechnology

[03] Cook, C. E., Bergman, M. T., Cochrane, G., Apweiler, R., & Birney, E. (2017). The European Bioinformatics Institute in 2017: data coordination and integration. Nucleic Acids Research, 46(D1), D21–D29.
https://doi.org/10.1093/nar/gkx1154

[04] Vaish, S., Agarwal, A., & Raheja, R. (n.d.). ROLE OF ARTIFICIAL INTELLIGENCE IN BIOTECHNOLOGY. In *International Research Journal of Modernization in Engineering Technology and Science* (pp. 2582–5208).
https://www.irjmets.com/uploadedfiles/paper/issue_4_april_2022/20852/final/fin_irjmets1650098311.pdf

[05] Shi, J.-P., Xie, R.-H., Yang, F., Zhou, W., Jiang, X.-Q., Hu, H.-F., Yang, C., Xie, Y.-Y., & Xiao, S. (2024). When agriculture meets biotechnology: a route for future agricultural innovation. Advanced Biotechnology, 2(4).
https://doi.org/10.1007/s44307-024-00047-3

[06] Dill, K. A., & MacCallum, J. L. (2012). The Protein-Folding Problem, 50 Years On. Science, 338(6110), 1042–1046.
https://doi.org/10.1126/science.1219021

[07] Adams, H. S. (2024, October 19). Google DeepMind's AlphaFold Wins Nobel Prize, Ushering in a New Era of Healthcare and Drug Discovery. Healthcare-Digital.com; Bizclik Media Ltd.
https://healthcare-digital.com/ai-ml/alphafold-2-the-ai-system-that-won-google-a-nobel-prize

[08] Thomason, J. (2024, May 8). Google's AlphaFold 3 AI predicts the very building blocks of life. VentureBeat.
https://venturebeat.com/ai/googles-alphafold-3-ai-predicts-the-very-building-blocks-of-life/

[09] AlphaFold. (2024, November 20). Google DeepMind.
https://deepmind.google/technologies/alphafold/

[10] Stanzione, F., Giangreco, I., & Cole, J. C. (2021). Use of molecular docking computational tools in drug discovery. Progress in Medicinal Chemistry, 273–343.
https://doi.org/10.1016/bs.pmch.2021.01.004

[11] Trott, O., & Olson, A. J. (2009). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of Computational Chemistry, 31(2), 455–461.
https://doi.org/10.1002/jcc.21334

[12] Gentile, F., Yaacoub, J. C., Gleave, J., Fernandez, M., Ton, A.-T., Ban, F., Stern, A., & Cherkasov, A. (2022). Artificial intelligence–enabled virtual screening of ultra-large chemical libraries with deep docking. Nature Protocols, 17(3), 672–697.
https://doi.org/10.1038/s41596-021-00659-2

[13] Harnessing the Power of Artificial Intelligence: Transforming Produce for Enhanced Nutrition | Agritech Future. (2024, February 13). Agritech Future.
https://www.agritechfuture.com/big-data-ai/harnessing-the-power-of-artificial-intelligence-transforming-vegetables-for-enhanced-nutrition/

[14] Redman, M., King, A., Watson, C., & King, D. (2016). What is CRISPR/Cas9? Archives of Disease in Childhood - Education & Practice Edition, 101(4), 213–215.
https://doi.org/10.1136/archdischild-2016-310459

[15] Liu, X., Jiang, D., Tao, B., Xiang, F., Jiang, G., Sun, Y., Kong, J., & Li, G. (2023). A systematic review of digital twin about physical entities, virtual models, twin data, and applications. Advanced Engineering Informatics, 55, 101876–101876.
https://doi.org/10.1016/j.aei.2023.101876

[16] Mayer, K. (2024, September 9). Digital Twins of Biological Systems Inform Drug Development. GEN - Genetic Engineering and Biotechnology News.
https://www.genengnews.com/topics/omics/digital-twins-of-biological-systems-inform-drug-development/

[17] Williams, R. (2024). AI could make better beer. Here's how. MIT Technology Review.
https://doi.org/1085094/10-breakthrough-technologies-2024

[18] We're using AI to create a new beer! | St Austell Brewery. (2024). Staustellbrewery.co.uk.
https://staustellbrewery.co.uk/about-us/our-blog/were-using-ai-to-create-new-beer

[19] Marr, B. (2021, July 2). How Artificial Intelligence Is Used To Make Beer | Bernard Marr. Bernard Marr.
https://bernardmarr.com/how-artificial-intelligence-is-used-to-make-beer/

[20] Heikkilä, M. (2024). Google DeepMind leaders share Nobel Prize in chemistry for protein prediction AI. MIT Technology Review.
https://doi.org/1085094/10-breakthrough-technologies-2024

[21] How AI Revolutionized Protein Science, but Didn't End It | Quanta Magazine. (2024, June 26). Quanta Magazine.
https://www.quantamagazine.org/how-ai-revolutionized-protein-science-but-didnt-end-it-20240626/

[22] Hefferon, K. L. (2016). Food Security of Genetically Modified Foods. Elsevier EBooks.
https://doi.org/10.1016/b978-0-08-100596-5.03532-0

[23] Wiseyak. (2024). Wiseyak.com.
https://www.wiseyak.com/aboutus