Biotechnology

SECTION EDITOR
Jennifer Bell
JenBellWS@outlook.com

Editorial

Ahmad Nazzal defines biotechnology as, "a field of science that involves using living organisms and organic substances to create or modify environments, produce goods, or improve human health."

Biotechnology is a vast topic and can be defined by using biological systems and living organisms in production processes. Biotechnology is commonly employed to make medical treatments using biomolecules and living cells.

It is vital that the massive amount of raw data generated during omics medical research is understood for medical treatments to be successful. Unfortunately, Big Data has resulted in a data bottleneck where lots of data are not processed because omics data processing technology needs to catch up. For some perspective, a 2015 estimate says about 40 petabytes of genomic data are produced at a population scale yearly.[1] One petabyte equals 1000 terabytes, and one terabyte equals one million megabytes. Almost 900 megabytes of data are needed to store information from one human genome.

Ahmad explains that bioinformatics is where biology, computer science, and information technology meet to help process large data sets. He glances over artificial intelligence (AI) regulation as a medical device because how to regulate it remains unclear. He writes about genomics and proteomics and how artificial intelligence is developing to process Big Data. Ahmad highlights the importance of public perceptions concerning how AI and biotechnology are embraced in the world.

I want to thank Ahmad for helping me to understand the developing role of AI in bioinformatics and biotechnology.

Jennifer Bell

1. Eisenstien M. Big data: The power of petabytes. Nature. 2015;527:S2–S4. doi:https://doi.org/10.1038/527S2a

Artificial intelligence, biotechnology, and trust

Ahmad M. Nazzal M.D., Ph.D.
https://scitribe.blog/
a-nazzal@outlook.com

Artificial intelligence (AI) is changing biotechnology. In this article, I discuss recent breakthroughs of AI in biotechnology, obstacles preventing faster progress, the future of the field, and the major role medical writers can play in building trust in the public about the future of AI in biotechnology.

First, let's look at the history of AI. In the 1950s, scientists seriously started to debate inventing machines that mimic human intelligence, and when breakthroughs in our understanding of DNA occurred. The field of artificial intelligence (AI) was born.[1] Today, AI can perform tasks such as visual perception, speech recognition, decision-making, translation between languages, and analysing data.[2]

However, AI is the new kid on the block when it comes to biotechnology.[3] Biotechnology is a field of science that involves using living organisms and organic substances to create or modify environments, produce goods, or improve human health.[4] AI is making it easier and faster to find new drug targets for diseases, detect diseases and harmful mutations, programme synthetic DNA, and analyse DNA sequence data.[3]

AI regulations in healthcare

The volume of data that AI can process is increasing. This makes it difficult to assess the quality and safety of AI-driven healthcare solutions. Therefore, in 2019, the US published a discussion paper that explains the approach to premarket review for AI-driven healthcare software when used in conjunction with a medical device.[5] Then, in 2021, the International Coalition of Medicines Regulatory Authorities published a report that provided recommendations for regulating therapies using AI. The report recommended adopting a risk-based approach, establishing governance structures, and fostering data reliability. Furthermore, it supported transparency, understanding, and real-world monitoring of patient functioning.[6] However, it is unclear how to keep up with the pace of innovation while regulating algorithms as medical devices.[7]

Breakthroughs

Bioinformatics, also known as computational biology, merges biology, computer science, and information technology. Bioinformatics accelerated areas of biotechnology that includes gene and protein sequencing, identification, prediction of function, understanding of complexity, structure and folding, and drug design and development.[8] AI has helped to make significant advances in biotechnology and bioinformatics.

Advances in genomics

Genomics is the study of the sequence and function of genes. The human genome is a set of approximately 3 billion base pairs on 23 pairs of chromosomes,[9] with differences between individuals called genetic variations.[10] There is a vast amount of data for each human being that standard statistical tools cannot analyse.[11] However, AI-based applications promise to help in this area. For example, DeepVariant,[12] a convolutional neural network, outperformed standard tools on variant-calling tasks.[13]

Furthermore, AI algorithms can improve variant classification and predict the impact of those variations. For example, PrimateAI,[14]

a convolutional neural network, outperformed previous methods in variation detection. It was trained on data from 120,000 human samples and showed superior performance compared to other variant pathogenicity prediction tools.[15]

However, not all genes code for proteins. Understanding non-coding genes remains an open challenge for the field.[16] It is estimated that up to 11% of rare genetic disorder causes could be traced to non-coding genes.[17] AI is expected to improve our understanding of non-coding genetic variations. For example, a deep layer neural network called SpliceAI[18] was able to predict non-coding genetic variants.[17]

The Human Genome Project highlights many genomics field advances. On October 1, 1990, the Human Genome Project was officially launched after planning in the late 1980s. It took scientists 13 years of work and almost US $3 billion in funds to map the human genome, a mosaic of sequences from 13 individuals.[9,19] However, even today, amendments are being made to the human genome reference sequence. Advances in the genome sequencing field have led to considerable reductions in the cost of genome sequencing. In 2014, the US$ 1000 genome was announced, and 20,000 human genomes could be sequenced in 1 year.[20,21] Costs have continued to reduce, and a human genome can be sequenced for US$ 600 with the US$ 100 genome not far behind due to advances in AI and computing.[22]

Without data, there can be no usage of AI.

The protein folding problem and AI

Proteins are the building units and the working molecules of the cells. They are made up of chains of amino acids, which can be arranged in a variety of ways in 3D space. Therefore, studying protein folding is difficult. In fact, it is referred to as a "grand challenge" in biology.

In biology, it is crucial to understand the way a protein folds because it reveals its function.[23] Scientists study protein folding using x-ray crystallography – an expensive, time-consuming, and error-prone process.[24] In theory, it is possible to determine protein structure by reading its amino acid sequence – this was impossible until 2020.

In December 2020, Google DeepMind[25] – a division of Alphabet Inc.[26] responsible for developing AI – introduced a neural network-based model, AlphaFold,[27] to accurately predict how proteins fold. This was acclaimed as the solution to the 50-year-old protein folding problem.[28,29] Three months later, and in partnership with the European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI), DeepMind launched the AlphaFold Protein Structure Database. Now, scientists can access and download the shape of every single protein in the human body and proteins of 20 additional organisms. This is expected to help scientists find solutions to antibiotic resistance, microplastic pollution, and climate change – a One Health approach.[30]

Drug discovery and AI

Drug discovery is an expensive, complex, and uncertain process.[31] The process of drug development can take around 12 years and costs up to €3 billion.[32] Drug development can be divided into three stages: hypothesis generation, candidate development, and commercialisation. Hypothesis generation involves target identification and validation, assay development, and lead generation. Once a couple of candidate leads are generated, the drug goes into animal studies for optimisation. Optimisation is an elaborate, time-intensive, and costly process. Once optimisation is done, first-in-human testing starts with a Phase IA clinical trial, in which the drug is tested on a small number of healthy volunteers. This is followed by Phase IB trials to establish safety, steady-state pharmacokinetics, and maximum tolerated dose. If the drug provides a proof of concept, it can move forward to larger randomised Phase II and Phase III trials to establish safety and efficacy. This is followed by marketing and continuous global optimisation.[33]

AI will enhance the process substantially.[34] Machine learning or deep learning – both sub-topics in the field of AI – can be used to discover and optimise therapeutic candidates faster.[35] In addition, AI is expected to cut down the costs of drug discovery. According to a one market size analysis, AI has the potential to save US$ 70 billion by the year 2028 in the drug discovery field.[36] Therefore, leading pharma companies are forging partnerships with AI start-ups and companies.[34] As such, faster drug discovery will require more medical writers to help get approvals from regulatory agencies.

Challenges

New technologies from AI have made significant advances in data-driven fields like biotechnology and medicine.[37] This AI-driven innovation in biotechnology and medicine will continue to increase.[38] However, this progress faces challenges.[39]

The data set challenge

One of the biggest challenges in applying AI to biotechnology, medicine, and healthcare is the lack of properly annotated, standardised, and non-biased data sets. Without data, there can be no usage of AI.[40]

AI specialists train AI algorithms on real-world data.[41] In AI there are two approaches to train an algorithm: supervised learning and unsupervised learning. In supervised learning we need to input a set of input data and to have the desired output labelled by human experts. The input data is made up of a set of things or events which can be repeated in a predictable pattern. The desired output is a prediction of the next event in the sequence. For example, in the prediction of an alphabet, the desired output is a letter. The input is a sequence of letters. It is a lot like teaching a small child the alphabet. In teaching a small child the alphabet, a person can tell the child what the alphabet is and what sound each letter makes. This would be the input data, and the desired output would be the child recognising the letter once presented. Thus, learning the alphabet.

As such, biased data sets will create biased AI. Many AI algorithms are trained with data sets that consist of data that reflect the biases of the culture that created it. In 2019, researchers found that an algorithm – used in US hospitals to predict which patients need extra medical care – favoured white patients over black patients. This algorithm affected 200 million patients.[42] To create an ethical non-biased benevolent AI, we need to assess data sets for biases, inequities, and discriminations.[43] Nevertheless, credible data sets remain a challenge for AI progress in data-driven fields including biotechnology.

The mindset challenge and building trust

AI is a powerful tool that could help but also could magnify flaws in the system at a damaging scale. Troncoso suggested that the greatest challenge to the usage of AI in healthcare is to change mindsets towards AI.[44]

Troncoso proposed that this challenge is representative of each society's mindset. And to achieve a change in attitudes, we need to educate and increase awareness of AI, to make AI more explainable, and to build trust between parties involved in the process. By the same token, it is recommended to implement ethical frameworks, encouraging positive behavioural intentions behind using AI and strike a balance between the exchange of individual data and the public for the greater good.[44]

In 2019, researchers found that an algorithm – used in US hospitals to predict which patients need extra medical care – favoured white patients over black patients. This algorithm affected 200 million patients.

The role of medical writers in building trust

AI is becoming more present in the world of biotechnology, where it will work with human scientists to solve real-world problems. Unfortunately, biotechnology often suffers from a bad reputation. And negative stories in the media, such as those related to controversies involving biotech companies, do not help.[45] On the other hand, AI has potentially dangerous implications; many scientists believe it could lead to a dystopian future.[46] Both are likely to be difficult for the public to accept without effective communication efforts.

The advancements of biotechnology research are dependent on the use of AI. However, due to the complexity of this field, it requires a significant amount of trust from the public.[47] In order to possess a certain level of trust, the public needs to be able to understand the benefits and risks of biotechnology advancements. Here is where medical communicators come in. Medical writers are experts in writing complex medical information and conveying it to others in easy-to-understand language.

The main challenge with AI is that it is increasingly difficult to understand – the future of technology is a mystery even to the best scientists. Despite this, medical writing will be a considerable part of the future. Therefore, medical writers would do well to learn about how AI works. They can easily do so by taking free online courses on the topic and grasping the subject.

Future

In today's world, we produce a massive amount of biomedical data. AI can digest massive datasets. Thus, in the future, AI will increase the accuracy of diagnosis, reduce the human error rate, and help healthcare professionals deal with a growing workload.[48]

One particularly exciting area is personalised medicine. Future personalised medicine will consider patient's genome to design tailored therapies. AI will make genetic testing more accurate, cheaper, and accessible.[11] Researchers are already using the technology to identify genes that are responsible for rare diseases in individual patients,[49] to understand cancer genomics,[50] and to create new therapies.[34]

Conclusion

The future of AI is ours to create. It offers tremendous potential to the fields of medicine and biotechnology. It can help us to accelerate screening and diagnosing diseases, as well as provide better patient-centred healthcare. Nevertheless, we need to be aware of AI's risks, such as aggravating existing social biases. Today, there is more hype and less reality around AI in the biotech industry and start-ups – executives believe there is a gap between what AI can do and what people think it can do.[51] I think that with continuous education we can recognise the hype from reality in the field of AI, determine its risks, and overcome obstacles facing it.

Disclaimers

The opinions expressed in this article are the author's own and not necessarily shared by his employer or EMWA.

Disclosures and conflicts of interest

The author declares no conflicts of interest.

References

1. History of Artificial Intelligence. 2021 [cited 2022 Jan 2]. Available from: https://en.wikipedia.org/w/index.php?title=History_of_artificial_intelligence&oldid=1062216803
2. Korteling JEH, van de Boer-Visschedijk GC, Blankendaal RAM, et al. Human – versus artificial intelligence. Front Artif Intell. 2021;4:14. Available from: https://www.frontiersin.org/articles/10.3389/frai.2021.622364/full. doi:10.3389/frai.2021.622364
3. Oliveira AL. Biotechnology, big data and artificial intelligence. Biotechnol J. 2019;14(8):1800613. doi:10.1002/biot.201800613
4. Dahms AS. Biotechnology: What it is, what it is not, and the challenges in reaching a national or global consensus. Biochem Mol Biol Educ. 2004;32(4):271–8. doi:10.1002/bmb.2004.494032040375
5. Artificial Intelligence and Machine Learning in Software as a Medical Device. 2021 [cited 2022 Jan 9]. Available from: https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device
6. International Coalition of Medicines Regulatory Authorities (ICMRA). Horizon Scanning Assessment Report – Artificial Intelligence. 2021. p. 37. Available from: https://www.icmra.info/drupal/sites/default/files/2021-08/horizon_scanning_report_artificial_intelligence.pdf
7. Medical Algorithms Need Better Regulation. 2022 [cited 2022 Mar 27]. Available from: https://www.scientificamerican.com/article/the-fda-should-better-regulate-medical-algorithms/
8. Kumar A and Chordia N. Role of bioinformatics in biotechnology. Res Rev Biosci. 2017;12(1):116.
9. Human Genome Project FAQ. 2020 [cited 2022 Jan 10]. Available from: https://www.genome.gov/human-genome-project/Completion-FAQ
10. What is genetic variation | Human genetic variation. [Cited 2022 Mar 4]. Available from: https://www.ebi.ac.uk/training/online/courses/human-genetic-variation-introduction/what-is-genetic-variation/
11. Dias R, Torkamani A. Artificial intelligence in clinical and genomic diagnostics. Genome Med. 2019;11(1):70. doi:10.1186/s13073-019-0689-8
12. Run DeepVariant | Cloud Life Sciences Documentation | Google Cloud. 2022 [cited 2022 Mar 9]. Available from: https://cloud.google.com/life-sciences/docs/tutorials/deepvariant
13. Poplin R, Chang PC, Alexander D, et al. A universal SNP and small-indel variant caller using deep neural networks. Nat Biotechnol. 2018;36(10):983-987. https://www.nature.com/articles/nbt.4235
14. PrimateAI: deep residual neural network for classifying the pathogenicity of missense mutations. Illumina; 2022 [cited 2022 Mar 9]. Available from: https://github.com/Illumina/PrimateAI
15. Sundaram L, Gao H, Padigepati SR, et al. Predicting the clinical impact of human mutation with deep neural networks. Nat Genet. 2018;50(8):1161–70. doi:10.1038/s41588-018-0167-z
16. Chatterjee S, Ahituv N. Gene regulatory elements, major drivers of human disease. Annu Rev Genomics Hum Genet. 2017;18:45–63. doi:10.1146/annurev-genom-091416-035537
17. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF,et al. Predicting splicing from primary sequence with deep learning. Cell. 2019;176(3):535–48.e24. doi:10.1016/j.cell.2018.12.015
18. Illumina/SpliceAI. Illumina; 2022 [cited 2022 Mar 9]. Available from: https://github.com/Illumina/SpliceAI
19. Bromberg Y. Building a genome analysis pipeline to predict disease risk and prevent disease. J Mol Biol. 2013;425(21):3993–4005. doi:10.1016/j.jmb.2013.07.038
20. Illumina's DNA Supercomputer Ushers in the $1,000 Human Genome. 2014 [cited 2022 Mar 15]. Available from: https://www.bloomberg.com/news/articles/2014-01-14/illumina-unveils-its-hiseq-x-dna-supercomputer
21. The Cost of Sequencing a Human Genome. 2021 [cited 2022 Jan 10]. Available from: https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost
22. Innovation at Illumina: The road to the $600 human genome. 2022 [cited 2022 Mar 15]. Available from: https://www.nature.com/articles/d42473-021-00030-9
23. Alberts B. A Grand Challenge in Biology. Science. 2011;333 (6047);1200. doi:10.1126/science.1213238
24. Alberts B, Johnson A, Lewis J, et al. The shape and structure of proteins. In: Molecular Biology of the Cell. 4th ed. New York: Garland Science; 2002.
25. DeepMind Homepage. [Cited 2022 Mar 9]. Available from: https://deepmind.com/
26. Alphabet [cited 2022 Mar 9]. Available from: https://abc.xyz/
27. AlphaFold Protein Structure Database. 2021 [cited 2022 Mar 9]. Available from: https://alphafold.ebi.ac.uk/
28. DeepMind. AlphaFold: a solution to a 50-year-old grand challenge in biology. 2020 [cited 2022 Jan 2]. Available from: https://deepmind.com/blog/article/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology
29. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583–9. doi:10.1038/s41586-021-03819-2
30. DeepMind. Putting the power of AlphaFold into the world's hands. 2021 [cited 2022 Jan 2]. Available from: https://deepmind.com/blog/article/putting-the-power-of-alphafold-into-the-worlds-hands
31. DiMasi JA, Feldman L, Seckler A, Wilson A. Trends in risks associated with new drug development: Success rates for investigational drugs. Clin Pharmacol Ther. 2010;87(3):272–7. doi:10.1038/clpt.2009.295
32. DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: New estimates of R&D costs. J Health Econ. 2016;47:20–33. doi:10.1016/j.jhealeco.2016.01.012
33. Posey Norris SM, Pankevich De, Davis M, et al. Improving and accelerating therapeutic development for nervous system disorders: Workshop summary. National Academies Press (US); 2014. . Available from: https://www.ncbi.nlm.nih.gov/books/NBK195047/ doi:https://doi.org/10.17226/18494
34. Fleming N. How artificial intelligence is changing drug discovery. Nature. 2018;557(7707):S55–7. doi:10.1038/d41586-018-05267-x
35. Preuer K, Klambauer G, Rippmann F, et al. Interpretable Deep Learning in Drug Discovery. In: Samek W, Montavon G, Vedaldi A, Hansen LK, Müller K-R, editors. Explainable AI: Interpreting, Explaining

and Visualizing Deep Learning [Internet]. Cham: Springer International Publishing; 2019 [cited 2022 Jan 2] . p. 331–45. (Lecture Notes in Computer Science). https://link.springer.com/chapter/10.1007/978-3-030-28954-6_18
36 Artificial Intelligence in Drug Discovery Market Size Analysis and Opportunity, 2018-2028. Bekryl Market Analysts. 2018 [cited 2022 Jan 2]. Available from: https://bekryl.com/industry-trends/ai-artificial-intelligence-in-drug-discovery-market-size-analysis
37. He J, Baxter SL, Xu J, et al. The practical implementation of artificial intelligence technologies in medicine. Nat Med. 2019;25(1):30–6. doi:10.1038/s41591-018-0307-0
38. Emmert-Streib F. From the digital data revolution toward a digital society: Pervasiveness of Artificial Intelligence. Mach Learn Knowl Extr. 2021;3(1): 284–98. doi:10.3390/make3010014
39. Emmert-Streib F. Grand Challenges for Artificial Intelligence in Molecular Medicine. Front Mol Med. 2021;1:2. doi:10.3389/fmmed.2021.734659
40. Kelly CJ, Karthikesalingam A, Suleyman M, et al. Key challenges for delivering clinical impact with artificial intelligence. BMC Med. 2019;17(1):195. doi:10.1186/s12916-019-1426-2
41. How to Train Your AI [Internet]. Revain. 2019 [cited 2022 Jan 9]. Available from: https://medium.com/revain/how-to-train-your-ai-98113bdac101
42. Obermeyer Z, Powers B, Vogeli C, et al. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366(6464):447–53. Available from: https://www.science.org/doi/abs/10.1126/science.aax2342
43. Khoury MJ, Iademarco MF, Riley, WT. Precision public health for the era of precision medicine. Am J Prev Med. 2016;50(3):398-401. doi:10.1016/j.amepre.2015.08.031 . do
44. Troncoso EL. The Greatest Challenge to Using AI/ML for Primary Health Care: Mindset or Datasets? Front Artif Intell. 2020;3:53. doi:10.3389/frai.2020.00053
45. Human Nature. Box Office Mojo [cited 2022 Mar 29]. Available from: https://www.boxofficemojo.com/release/rl2948170241/
46. Cellan-Jones R. Stephen Hawking warns artificial intelligence could end mankind. BBC News. 2014 [cited 2022 Mar 27]. Available from: https://www.bbc.com/news/technology-30290540
47. Savadori L, Savio S, Nicotra E, et al. Expert and public perception of risk from bio-technology. Risk Anal. 2004;24(5):1289–99. doi:10.1111/j.0272-4332.2004.00526.x
48. Amisha, Malik P, Pathania M, Rathaur VK. Overview of artificial intelligence in medicine. J Fam Med Prim Care. 2019;8(7):2328–31. doi:10.4103/jfmpc.jfmpc_440_19doi
49. De La Vega FM, Chowdhury S, Moore B, et al. Artificial intelligence enables comprehensive genome interpretation and nomination of candidate diagnoses for rare genetic diseases. Genome Med. 2021;13(1):153. doi:10.1186/s13073-021-00965-0
50. Greatbatch O, Garrett A, Snape K. The impact of artificial intelligence on the current and future practice of clinical cancer genomics. Genet Res. 2019;101:e9. doi:10.1017/S0016672319000089
51. VentureBeat. KPMG survey finds execs worry about AI hype – but they can address it. Making sense of AI. 2021 [cited 2022 Mar 27]. Available from: https://venturebeat.com/2021/05/26/kpmg-survey-finds-execs-are-worried-about-ai-hype-but-they-can-overcome-it/

Author information

Ahmad M. Nazzal, M.D., Ph.D. is a medical writer, physician, and researcher. His experience in medicine, research, and computer engineering puts him in a position to write about advances of AI in healthcare. He is dedicated to using digital health to make healthcare better and more affordable.