AI in biomedical and health sciences – applications and challenges

Medicine, public health, pharmacy, and medical sociology deal with health, illness, and death - can AI be trusted?

The inroads of artificial intelligence (AI) into sociology and politics, so far, seem to be problematic, as discussed in the foregoing entries (1, 2). In health science disciplines are combined, which has enormous importance for society, as it is true for the legal system. AI models might pass law exams, such as multiple-choice or essay questions, but profoundly fail in real-world legal tasks. There, an AI model has to react to different interpretations of a given law or a former ruling, and to contradictory arguments between layers and judges or two legally battling individuals, not just to correctly answer multiple questions and write an essay (3). The Medical state exams might require multiple-choice answers throughout many different specialties, but will a suffering patient, anxiously waiting for the efficient help of an experienced physician, trust AI? In combining disciplines caring for health and healthcare, covered in the academic field as medicine, pharmacy, nursing, and public health. All four topics are intertwined and, to some extent, connected to important aspects of social science. (4).

AI technology in psychological research

A telling example of troublesome AI technology occurred in psychological research. The idea that LLMs simulate human psychology and could replace human participants in psychological studies was troublesome. Slight changes in wording by ‘prompting’ resulted in large changes and discrepancies between LLMs and human responses. Even specific ‘fine-tuned’ models designed for psychological AI research did not produce reliable results (5).

Silicon samples in sociology

The attempt by social scientists to use LLMs to ‘create’ fictitious samples, referred to as ‘silicon samples,’ for their research project is debated. A systematic investigation into this research technique revealed that there is practically no accuracy in the ‘silicon sample’ compared to human data. During the selection of silicon substitutes, very small changes resulted in significant discrepancies among the silicon samples. At best, one sample, out of an overall total of 252 samples, corresponded well to one category, but poorly to the others. Finally, it was not possible to come up with even ‘one-size-fits-all’ configuration (6).

The imbalance of popular LLMs could be due to leaning heavily during ‘training’ towards the European culture, and disregarding minorities and older people. It is possible that two researchers running the same study with silicon samples could reach opposite conclusions (7).

Synthetic samples for medical research

Not only in psychology but also in medicine, researchers started to use AI-generated medical data to avoid the need for ethics reviews (8). It is argued that synthetic data could be used in medical research, not originating from real populations, but created through mathematical models and algorithms, which could benefit hypothesis generation or preliminary testing for research approaches before collecting data from real populations (9).

An AI example directly related to clinical purposes is the use of AI models for the interpretation of patients' X-ray scans to generate reference scans. Radiology is a rare profession in medicine, so real data are limited, and AI models will help the radiologists work faster (10). Ethical and legal considerations prohibit replacing the medical doctor with AI. Radiology is only one example hinting towards the possible future use of synthetic data for healthcare.

Synthetic data in the research of various clinical specialties

A 'scoping' review of domains, motivations, and future applications covered a broad spectrum of the medical specialties, including oncology, neurology, cardiology, endocrinology, and gynecology (11).

Of 42 relevant publications, the main specialties were oncology, neurology, and cardiology. The primary motivation for testing the artificial data was to explore its future use, the types of data it generated, and its limitations and opportunities. The reasons to use such data rested on the difficulty of coming across enough real patients' data, the limitations of using real data in making sure that the privacy of the patients is secured, and the aspect that such an approach is considerably new and unexplored. The expectation is that it could help to improve diagnosis, treatment, and monitoring of the diseases, which is especially advantageous for rare diseases or those highly dangerous to health.

Concerns about the methodology were voiced regarding data security, quality, the way of AI development, and the direct medical and clinical applications. Uncertainty evolved in the identification of risk, which might not reflect the real data, which might be of particular concern in rare diseases. Images in the health field pose specific problems because they may not reflect reality.

Examples for data generator for study types

There are data generators for specific study types. For single-cell RNA sequencing data, the SPLATTER generator is mentioned as useful for oncology (12). The Splatter Bioconductor hints towards multiple simulation methods of relevance for AI technology. So far, most publications refer to a single domain, without mentioning subdomains. An exception is endocrinology, dealing with COVID and diabetes (11). The focus was given to diabetes type one. A suitable synthetic data generator includes the UVA/PADOVA Type I Diabetes simulator (13).

AI models in epidemiology and statistics

Epidemiology is one of the key methods for population health research. Even before AI models came into the limelight, epidemiological research drew on numerous mathematical models, for instance, in the description of the spread of infectious diseases (14). The most recent push in this direction initiated the COVID-19 calamity, and considered AI models to simulate networks, combined with additional established models, based on real data, estimated the distribution of infectious diseases over time (15, 16).

Among public health experts, the question of whether statistics come first and epidemiology is a sideline of statistics or the other way around is an unsolved controversy among professionals. Anyway, applied statistics nowadays focuses on AI as well, studying AI models themselves their reliability, and quality. This involves AI in applied statistics in the field of engineering and technological advances. In the medical field, ChatGPT was tested as an AI support tool for straightforward medical questions to verify suspected diagnoses. ChatGPT correctly answered over 90% of trivia questions in coronary artery diseases, pulmonary and venous embolism, fibrillation, heart failure, and risk management (17). Well-known statistical software such as SAS, JMP, Stata, and Python, were improved with the help of AI technology. Attention was paid to AI-generated coding, which worked well with easygoing problems, but still needs improvement for more difficult undertakings, where bugs and incorrect logic did occur (18).

Genomics is the main beneficiary of AI

From a biomedical perspective, the usefulness of AI applications in genomics is more convincing than in other related fields. Advances in research techniques emerged with a swiftness rarely seen before. Genomics, which has even modernized historical investigations, has explored the genomics of Neanderthals in Europe and Denisovans in Asia (19, 20). Among others, metabolomics, epigenomics, lipidomics, and proteomics, down to artificial and single-function proteins, digging deep into various fields of key parts of biomedical science (21).

The tens of thousands of proteins in the human body, encoded by genes, are made from a string of amino acids of various sizes. Previously, in a very tedious manner, the structures of proteins were assessed in the laboratory using X-ray crystallography and cryo-electron microscopy. The results of the investigations in that field were entered into databases. As a forerunner of an AI model, University of Washington scientist David Baker worked on software called ‘Rosetta’ to combine amino acid sequences of protein parts and insert the DNA coding for those sequences into bacteria, which produced the desired protein. These artificial proteins, however, were still non-functional (22).

The Nobel prize for physics in 2024

With AI protein language models (PLMs), categorized as ‘end-to-end-to-text sequence generation’, the task became easier, but only non-functional molecules could still be designed from short amino acid sequences. The proteins remained inert because their proper 3D structure, which controls their function, was not considered. Using the relevant 3D databases, the scientists Demis Hassabis and John Jumper trained a large language model (LLM) known as AlphaFold2. Both approaches —Baker's approach together with the models from Hassabis and Jumper — enhanced ‘structural biology’.

The model's training began with standard language instructions for designing sequences and eventually combined structure and language. Baker’s first functional protein was named Top7, an exceptional structure ‘not yet been seen in nature.’ All three scientists got the Nobel Prize for Physics in 2024 (22).

Functional proteins in the future?

The model can produce functional proteins similar to those of existing molecules, like enzymes and antibodies (23, 24). Intentionally introducing an artificial protein into human metabolism may not be to everyone's liking. There are enthusiastic hints of proteins with functions that we’ve never imagined.’ Envisaged are peptides against various strains of influenza, molecules that deliver drugs to the body, enzymes to clean water, and help recycle plastics and composite materials, as well as microbes sucking carbon dioxide out of the atmosphere (22).

Protein sequencing and the example of breast cancer

Yet, for many scientific applications, protein sequencing using laboratory technology served research for very pragmatic purposes in medicine and pharmacy, and with AI now protein sequencing is revolutionized (25). This could lead to a new area of personalized risk assessment, for instance, in oncology in the clinical setting, illustrated by the example of breast cancer.

Variants for particular genes are known to cause breast cancer. For quite some time, it has been known that carriers of the BRCA1 variant have a 57% risk (95% CI, 47% to 66%) and carriers of the BRCA2 variant have a 49% risk (95% CI, 13% to 23%). BRAC2 also has a relatively high risk of 40% for ovarian cancer, while the risk for this malignancy for BRAC1 is lower, with 18% (26). Some females won’t want to know about the possibility of being a carrier, because of the uncertainty, whether to really become a victim, or others might even decide to amputate their breasts. To know such a risk, therefore, is a double-edged sword.

Penetrance of variants of disease-related phenotypes

The uncertainty whether the variant is potentially dangerous depends on a protein function and its ‘possible dynamic mechanism’, called the ‘conformational status,’ which is different from its ‘compositional status’, of which the latter relates to various conditions, such as temperature, solvation, concentration, and other conditions that determine how the protein works (27). For the clinic, not only are the harmful consequences of a particular genomic variant important, but also the likelihood that the clinically relevant condition will evolve.

Formerly, the penetrance of breast cancer was estimated by epidemiological means, investigating disease occurrence frequently in particular families or measuring the incidence of carriers of the relevant variants within the total population (28). Very time-consuming and complex biophysical experiments and molecular dynamics simulations in the laboratory were also used to assess the penetrance with a ‘top-down’ approach. Recently, an innovative AI model was created and published. Using machine learning, the results of oligonucleotide and sequencing techniques generated a large volume of data, enabling AI to approximate disease risk by testing individual variants and their penetrance (29, 30). So far, machine-learning models for 10 selected diseases have been developed using patient data from more than one million individuals, incorporating 1600 variants in 31 genes. The models were validated on clinical data from the Mount Sinai

Health System. The system enables a ‘bottom-up’ methodology based on the individual patient and the disease she or he is suffering from. The AI technique in this example shifts diagnosis and treatment closer to ‘precision medicine’ (29). This innovative development in medicine obviously had the opportunity to assess data and software, partly available through sources of a well-known medical establishment.

The commercial aspect of AI in genomics

It is not known whether the above-described program will be free. The question arises because patient data are used to train the model. In the EU, key components of AI and training data are subject to scrutiny. Once a model claims to be free, it is allowed to be studied, modified, and the underlying model can be shared and scrutinized. It is known that firms might claim to offer open-source AI to avoid scrutiny, even though they do not actually have the ‘open-source’ label. There is the threat that the scientific community is in danger of finally facing ‘closed corporate systems and unverifiable models’ (31).

The AlphaFold model starts to lack data

The AlphaFold model, mentioned above as the one that won the Nobel prize, starts to have problems incorporating more data. The latest version, AlphaFold3, models the interaction with other molecules, including drugs. It seems that data useful for AlfaFold3 are also of great interest to drug companies, and relevant research results are being withheld and not published. Leading pharmaceutical companies are reportedly planning to develop their own AlphaFold3 AI model. Besides the 200.000 freely available protein structures in the Protein Data Bank (PDB), there are thousands of protein structures hidden in pharmaceutical company safes, neither shared by the pubic nor with other drug companies (32). There is the threat that AI in the future will not make full use of the methodology, unless drug companies allow independent organizations to manage the nonproprietary high-quality datasets to train and validate models for the overall benefit (33). It doesn’t look like this will happen soon.

Patents on AI-derived drugs not yet been tested in vivo

Another unfavorable aspect of AI-derived drugs is that companies that rely particularly on AI drug development have managed to patent more molecules without prior in vivo testing, compared to those firms that work more along a conservative path, actually testing their products more often in vivo before obtaining a patent. Speculative molecule structures, then, are blocked by the patent in that other companies are reluctant to test the remedy because of legal considerations. The incentive for research is then minimal, and many promising molecules will never become a final product. Although obtaining a patent for a molecule or invention requires proof that the item is novel and has not been previously disclosed, if it is plausible that the intervention will work as expected, the firm's lawyers might succeed in obtaining the patent. To avoid blogging promising developments in the future, evidence that the molecule really works as promised must be more strictly enforced (34).

Outlook for public health, statistics, and publications

This third reflection on AI focuses on biomedical science, including medicine, public health, pharmacy, and genomics. The field with the most potential to benefit from AI appears to be genomics. Statistics, among all the remaining specialties, is the most twisted between the established way of operating and a rather challenging future. The ‘two’ cultures in statistical modeling belong to ‘data modeling’ on one side and ‘algorithmic modeling’ on the other. The conventional attitude is to estimate the truth by operating on the data with linear or logistic regression, Cox model, etc., starting from the independent (input) variables “x” to the response variable “Y”, and validating the outcome with “yes” or “no”, goodness-of-fit, and residual examination. This conservative approach is now being challenged by the “algorithmic modeling culture” (35). Those very much in favor of the algorithmic modeling might overlook the difference between statistics used for engineering as opposed to natural science, here especially related to human health and survival (36).

The question should be allowed while using AI methods in statistics, what happened to the need to calculate a sample size, and how to deal with confounding, Berkson’s Bias, Simpson’s Paradox, and other significant biases.

AI and scientific publications

A real threat to science publications is the misuse of AI methodology. Recently, journals related to PLOS, Frontiers announced that submissions on open health data from surveys of the National Health and Nutrition Examination (NHANES) based on more than 130.000 individuals will be automatically rejected because ‘of unscrupulous actors use these data sets to churn out nonsense scientific papers’ (37). This drastic step is a response to an overwhelming increase in low-quality papers generated by public data and AI, which turned out to be an industrial-scale publishing fraud, including ‘unethical’ AI research (38-40).

Conclusion

The usefulness of AI in future in health science varies across specialties. Medical expertise, particularly in clinical medicine, will benefit from genomics in precision medicine, rare genetic diseases, and the interplay of genetics with the environment through epigenetics. Innovative tools might be developed to assist in patient care and support clinicians in making informed decisions, but they cannot relieve clinicians of their responsibility. Public health in epidemiology and statistics could contribute to improving AI technology, but be careful not to disregard biases and erroneous conclusions that have long been known to be associated with epidemiological research. The pharmaceutical economic sector should be carefully kept within the legal boundaries and resist towards financial gain against fairness for the whole society.

Medical sociology, psychiatry, and related fields using AI models designed for the general public should be more conscious of the social impact of the tools made available and assess their impacts effectively. Digital technologies can be easily deployed to billions of people, while ‘corporate actors resist public knowledge creation or data sharing’ (41). Scientists' expectations for the future of AI are quite diverse.

A balanced view is expressed by about 4.000 researchers who were asked how they think AI will improve life (42). Those from China are the most optimistic, seeing more benefits than risks in about 70%. From the listed countries, the UK has the lowest optimism, well below 50%, while the EU, India, and low-income countries hover around 50%. One main expectation is that AI will increase access to learning and education. Additionally, positive changes are foreseen for easier jobs, improved health care, and reduced working hours at the same pay as before. The often-voiced adverse effect is the unreliability of the information, distinguished from fakes (42).

Others do not share this somewhat moderate, positive impression and expectation of AI, and note that an overreliance on AI-driven modeling is harmful to science and that its growing role could do more harm than good (36). An error called ‘leakage’ is supposed to occur frequently in machine learning modeling, enabling AI to make predictions and test hypotheses. Evaluation data influence the training process by focusing on the ‘patterns’ of the evaluation data, but fail to capture the real meaning behind the phenomenon observed (43). A systematic review of papers on the COVID-19 calamity revealed that most studies used poor evaluation methods and were uncertain about the disease's reality in adult patients, while the negatives were in young children. The conclusion was that a simple chest X-ray or CT scan is sufficient for the diagnosis. Often overlooked are individual findings and their meaning for the progress of scientific understanding (36).

Literature:

1. Does it require a Carrington Event* to save humanity? Thailand: Faculty of Public Health, Khon Kaen University; 2025 [Available from: https://ph.kku.ac.th/eng/index.php/research/journal-club-phkku/236-180868.
2. AI in society and politics - Chatbots stimulate suicide, and the US president’s “assault on science”: Khon Kaen, Thailand: Faculty of Public Health, Khon Kaen University; 2025 [Available from: https://ph.kku.ac.th/eng/index.php/research/journal-club-phkku/237-300968.
3. Chaudhri VK. We need a new Turing test to assess AI's real-world knowledge. Nature. 2025;646(8087):1027.
4. Dunleavy P, Bastow, S. Tinkler, J. The contemporary social sciences are now converging strongly with STEM disciplines in the study of ‘human-dominated systems’ and ‘human-influences systems’ Great Britain: London School of Economy; 2014 [Blog]. Available from: https://blogs.lse.ac.uk/impactofsocialsciences/2014/01/20/social-sciences-converging-with-stem-disciplines/.
5. Schröder S, Morgenroth, T., Kuhl, U., Vaquet, V. Paasßen, B. . Large language models do not simulate human psychology. aeXiv:250806950v3 [csAI]. 2025.
6. Cummins J. The threat of analytical flexibility in using large language models to simulate human data: A call to attention. arXiv: 250913397v2[csCY]. 2025.
7. O’’Grady C. AI-generated ‘participants’ can lead social science experiments astray, study finds. NewsSceinceInsider. 2025.
8. Extance A. AI-generated medical data can sidestep usual ethics review, universities say. Nature. 2025.

9. Synthetic data can benefit medical research - but risks must be recognized. Nature. 2025;645(8080):283.
10. Achour N, Zapata T, Saleh Y, Pierscionek B, Azzopardi-Muscat N, Novillo-Ortiz D, et al. The role of AI in mitigating the impact of radiologist shortages: a systematised review. Health Technol (Berl). 2025;15(3):489-501.
11. Rujas M, Martin Gomez Del Moral Herranz R, Fico G, Merino-Barbancho B. Synthetic data generation in healthcare: A scoping review of reviews on domains, motivations, and future applications. Int J Med Inform. 2025;195:105763.
12. Zappia L, Phipson B, Oshlack A. Splatter: simulation of single-cell RNA sequencing data. Genome Biol. 2017;18(1):174.
13. Man CD, Micheletto F, Lv D, Breton M, Kovatchev B, Cobelli C. The UVA/PADOVA Type 1 Diabetes Simulator: New Features. J Diabetes Sci Technol. 2014;8(1):26-34.
14. Pujante-Otalora L, Canovas-Segura B, Campos M, Juarez JM. The use of networks in spatial and temporal computational models for outbreak spread in epidemiology: A systematic review. J Biomed Inform. 2023;143:104422.
15. Bernick L. Modelhuman networks using random graphs 2018 [Available from: https://www.google.com/search?client=firefox-b-d&q=what+is+the+Erd%C3%B6s-R%C3%A9nyi+random+model.
16. Ford E. Barabási-Albert random graphs, scale-free distribution and bounds for approximation through Stein's method University Dissertation: Oxford; 2009 [Available from: https://www.google.com/search?client=firefox-b-d&q=what+is+the+Barab%C3%A1si-Albert+scale-free+model#vhid=o4qX5-rmQRSenM&vssid=l.
17. Harskamp RE, De Clercq L. Performance of ChatGPT as an AI-assisted decision support tool in medicine: a proof-of-concept study for interpreting symptoms and management of common cardiac conditions (AMSTELHEART-2). Acta Cardiol. 2024;79(3):358-66.
18. Min J, Song, X., Zheng, S., King, C.B., Deng, X., Hong, Y. Applied statistics in the era of artificial intelligence: A review and vision arXiv: 2412.10331v1; 2024 [Available from: https://arxiv.org/html/2412.10331v1.
19. Genomics - Part1: The Basis - an Obligation for Public Health Khon Kaen Thailand: Faculty of Public Health; 2023 [Available from: https://ph.kku.ac.th/eng/index.php/research/journal-club-phkku/208-260766.
20. Denisovans in Asia join the Eurocentric Neanderthal in the awareness of our evolutinary past Thailand: Faculty of Public Health, KHon Kaen University; 2023 [Available from: https://ph.kku.ac.th/eng/index.php/research/journal-club-phkku/205-170366.
21. Higgs M. The state of multi-omics England Wales: Frontline Genomics; 2024 [Available from: https://frontlinegenomics.com/the-state-of-multi-omics/.
22. Offord C. Protein designer and structure solvers win chemistry Nobel USA: AAAS; 2024 [Available from: https://www.science.org/content/article/protein-designer-and-structure-solvers-win-chemistry-nobel.
23. Callaway E. I told AI to make me a protein. Here's what it came up with. Nature. 2025;641(8065):1079-80.
24. Dai F, Fan, Y., Su, J. Wang, C., Han, C., Zhou, X., Liu, J., Qian, H., Wang, S., Zeng, A., Wang, Y., Yuan, F. Toward de novo protein design from natural language. Reprint bioRxiv. 2025.
25. Service RF. The AI revolution comes to protein sequencing. Science. 2025;388(6743):142.

26. Chen S, Parmigiani G. Meta-analysis of BRCA1 and BRCA2 penetrance. J Clin Oncol. 2007;25(11):1329-33.
27. Lewis S, Hempel T, Jimenez-Luna J, Gastegger M, Xie Y, Foong AYK, et al. Scalable emulation of protein equilibrium ensembles with generative deep learning. Science. 2025;389(6761):eadv9817.
28. Anglian Breast Cancer Study G. Prevalence and penetrance of BRCA1 and BRCA2 mutations in a population-based series of breast cancer cases. Anglian Breast Cancer Study Group. Br J Cancer. 2000;83(10):1301-8.
29. Forrest IS, Vy HMT, Rocheleau G, Jordan DM, Petrazzini BO, Nadkarni GN, et al. Machine learning-based penetrance of genetic variants. Science. 2025;389(6763):eadm7066.
30. Raiken H, Stein A. Penetrance and variant consequences-Two sides of the same coin? Science. 2025;389(6763):880-1.
31. Maffulli S. 'Open source' AI isn't truly open - here's how researchers can reclaim the term. Nature. 2025;640(8057):9.
32. Callaway E. AlphaFold is running out of data - so drug firms are building their own version. Nature. 2025;640(8058):297-8.
33. Gold ER, Cook-Deegan R. AI drug development's data problem. Science. 2025;388(6743):131.
34. Freilich J, Rai AK. What patents on AI-derived drugs reveal. Science. 2025;388(6750):924-6.
35. Breiman L. Statistical modelling: The two cultures. Statistical Science 2001;16(3):33.
36. Narayanan A, Kapoor S. Why an overreliance on AI-driven modelling is bad for science. Nature. 2025;640(8058):312-4.
37. O'Grady C. Papers based on open health data face bans. Science. 2025;390(6770):222-3.
38. O'Grady C. Low-quality papers surge thanks to public data and AI. Science. 2025;388(6749):807-8.
39. O'Grady C. 'Unethical' AI research on Reddit under fire. Science. 2025;388(6747):570-1.
40. O'Grady C. Study reveals industrial-scale publishing fraud. Science. 2025;389(6760):557-8.
41. Orben A, Matias JN. Fixing the science of digital technology harms. Science. 2025;388(6743):152-5.
42. Schwaller F. Will AI improve your life? Here's what 4,000 researchers think. Nature. 2025;640(8059):577-8.
43. Kapoor S, Narayanan, A. Leakage and the reproduibility crisis in machine - learning - based science: Patterns (N Y), AUG 4, 4(9): 100804; 2023 [Available from: https://pubmed.ncbi.nlm.nih.gov/37720327/.

Frank P. Schelp is responsible for the manuscript's content, and the points of view expressed might not reflect the stance and policy of the Faculty of Public Health, Khon Kaen University, Thailand.
For comments and questions, please contact <awuso11@gmail.com.

Grammarly software was used to improve English, but the AI function was disabled.