Invited article

Outbreak, Surveillance, Investigation & Response (OSIR) Journal
ISSN 2651-1061 (Online)
https://he02.tci-thaijo.org/index.php/OSIR
Email: osir@ddc.mail.go.th

June 2026
Volume 19, Issue 2
Article no. 282987

The Grammar of Science: B&I Before AI—Bias and Inference Still Matter

Jaranit Kaewkungwal*

Faculty of Tropical Medicine, Mahidol University, Thailand

*Corresponding author, email address: jaranitk@biophics.org

Received: 23 May 2026; Revised: 6 Jun 2026; Accepted: 12 Jun 2026
https://doi.org/10.59096/osir.v19i2.282987

Artificial Intelligence (AI) is increasingly used in healthcare and research, from disease detection, chronic disease management, service delivery, and drug discovery to proposal development, data analysis, and scientific writing. Its influence is clear: in 2021, the World Health Organization issued its first global report on AI in health, and many universities now offer courses on AI and research.[1] At the same time, some quick training short-courses, both onsite and online, now promote using AI to conduct research without enough attention to scientifically and ethically sound research principles. AI can expand human capability, but its value still depends on data quality, representativeness, context, and human judgment. When data are biased, incomplete, or disconnected from lived experience, AI may produce inaccurate, unfair, or inequitable results.[2,3] Concerns also remain about hallucination, misuse, privacy, black-box decisions, unclear accountability, reinforced bias, and overreliance on pattern recognition without true understanding.[2,3]

Before the AI era, researchers had to consider bias in study design, data collection, analysis, interpretation, and generalization. They also had to judge whether findings supported association, inference, or causal claims. These issues have not disappeared; AI may even make them more complex. As AI becomes more common in research, we must return to the basic principles of bias and inference. AI can assist research, but researchers must still understand, question, interpret, and take responsibility for the meaning of the results. Let's look into these issues.

AI in Biomedical and Health Research

Let's start with what AI means. There is no single agreed definition, but AI generally refers to computer systems or programs that perform tasks commonly linked to human intelligence, such as reasoning, learning, adaptation, sensory understanding, interaction, decision-making, creativity, pattern recognition, and prediction.[1,3,4] In health and research, AI can analyze large datasets, text, and images; produce tailored recommendations; support decisions; and improve efficiency.[5] AI is best understood as a broad group of technologies, including machine learning, deep learning, natural language processing, robotics, computer vision, and neural networks.[4,6] More recently, AI has moved beyond conventional data processing tasks into areas such as content generation and causal reasoning. Generative AI can produce written content, visual materials, computer code, audio, and synthetic datasets, whereas causal AI focuses on understanding possible causal pathways and strengthening inference from data.[7–10]

Historically, AI has evolved over many decades, beginning in the 1950s with early questions about machine intelligence. Alan Turing, a British mathematician and computer scientist, raised the famous question of whether machines could "think," which later influenced the development of AI. The term "artificial intelligence" was formally introduced at the Dartmouth workshop in 1956. Early AI mainly used symbolic rules and expert systems, but progress slowed during the "AI winters" because early systems could not meet high expectations and were limited by computing power, data availability, and real-world performance. From the 1980s, machine learning became more useful as computers began learning patterns from data instead of relying only on programmed rules. Later advances in neural networks, backpropagation, big data, graphics processing units,

All material is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC-BY-NC-ND 4.0) License unless otherwise stated.

and deep learning drove major progress in the 2000s and 2010s, leading to breakthroughs in image recognition, speech recognition, translation, games, and decision-support systems. By the 2020s, foundation models and generative AI had brought AI into daily life, enabling systems to produce text, images, audio, and code. At the same time, growing interest in causal inference and causal AI has emphasized that AI should not only find patterns, but also help researchers think more carefully about cause-and-effect relationships.[6,7,11]

Recent literature from 2024 to early 2026 shows that AI is increasingly used across healthcare and public health.[12,13] In drug discovery, tools such as DeepChem MLP, DeepTox, DeepNeuralNetQSAR, ORGANIC, and PotentialNet help identify drug candidates, predict toxicity, assess molecular activity, generate new molecules, and estimate ligand–target binding.[12] AI is also used in neuroimaging to predict task-related information and possible clinical biomarkers, and AI-enabled medical devices are increasingly being used in radiology, cardiology, and neurology.[14,15] In public health, AI can support surveillance, epidemiological research, communication, resource allocation, recommendations, and decision-making.[5]

Numerous AI tools are now used throughout the research process, from idea development and proposal writing to evidence synthesis, analysis, manuscript preparation, and accessibility. General AI assistants such as ChatGPT, Claude, Microsoft Copilot, Gemini, Perplexity, Stanford AI Playground, and Stanford Secure GPT can support brainstorming, topic refinement, outlining, writing, summarizing, coding, and interpretation. Literature-review and evidence-synthesis tools, including SciSpace, Elicit, Rayyan, Abstrackr, Consensus, Scite AI Assistant, and Inciteful, can assist with paper reading, citation screening, evidence appraisal, systematic reviews, and study mapping. Medical search and clinical reference tools, such as PubMed AI, ClinicalKey AI, UpToDate Advanced AI Search, Cochrane AI, and OpenEvidence, can support natural-language searching, evidence retrieval, clinical summaries, and review production. Speechify can further support accessibility through text-to-speech functions.[16,17]

AI has been used in data analysis to process large and complex datasets more quickly. It can identify patterns, predict disease risks, and support more personalized treatment and prevention strategies.[18] A paper on statistical analysis in internal medicine research noted that AI can help write analysis code, perform data analyses, suggest preliminary statistical tests, detect reporting errors, check guidelines, and assist with data extraction. AI tools mentioned include ChatGPT, Bard, Llama, ChatGPT-4o, Gemini, and Elicit. These tools produced results comparable to those obtained using SPSS, R, and SAS for data analyses.[8]

Bias and Confounding in Research

Bias refers to systematic error in study design, participant selection, data collection, measurement, analysis, interpretation, or reporting, which can weaken validity and lead to misleading conclusions.[19–21] Because bias can occur at any stage, careful design and implementation are needed to protect study validity. Bias should always be considered when interpreting findings because it can hide a true association, create a false one, or overestimate or underestimate effect size.[19,22] Understanding bias is critical because it can distort scientific interpretation, damage public trust, and influence health decisions and policy. As an example, the discredited measles, mumps, and rubella vaccine–autism study had serious methodological problems and contributed to public concern, lower vaccination uptake, and later measles increases.[19] Importantly, increasing sample size does not remove bias.[21]

Common types of bias in epidemiological studies are often grouped into two broad categories: selection bias and information bias, also called misclassification bias. Although many specific types of bias have been described in research over the years, they can generally be classified under these two broader groups.[20,23] Selection bias occurs when study participants differ systematically from those not included, or when comparison groups are not truly comparable, which may limit generalizability or distort group comparisons; examples include sampling bias, allocation bias, and loss to follow-up. Information bias occurs when exposure or outcome data are measured, recorded, or collected incorrectly or differently across groups, leading to misclassification and distorted associations; examples include measurement error, observer bias, interviewer bias, recall bias, and related data collection problems.[20,21]

The types of bias shown in Table 1 are drawn from early epidemiologic work in the 1970s and are listed according to the stage of research in which they occur. This listing is based on the “Catalogue of Bias” in epidemiological and clinical research compiled by M. Tevfik Dorak which was adapted from David Sackett’s work.[21,24]

Table 1. Types of biases in research.

Stages of research	Major categories/examples of bias
Literature review	**Foreign language exclusion bias, literature search bias, one-sided reference bias, rhetoric bias**
Study design	**Selection bias** **Sampling frame bias** (Berkson (admission rate) bias, centripetal bias, diagnostic access bias, diagnostic purity bias, hospital access bias, migrator bias, prevalence-incidence (Neyman/ selective survival; attrition) bias, telephone sampling bias) **Nonrandom sampling bias** (autopsy series bias, detection bias, diagnostic work-up bias, door-to-door solicitation bias, previous opinion bias, referral filter bias, sampling bias, self-selection bias, unmasking bias) **Noncoverage bias** (early-comer bias, illegal immigrant bias, loss to follow-up (attrition) bias, response bias, withdrawal bias) **Noncomparability bias** (ecological (aggregation) bias, healthy worker effect (HWE), lead-time bias, length bias, membership bias, mimicry bias, nonsimultaneous comparison bias, sample size bias)
Study execution	**Bogus control bias, contamination bias, compliance bias**
Data collection	**Instrument bias** (case definition bias, diagnostic vogue bias, forced choice bias, framing bias, insensitive measure bias, juxtaposed scale bias, laboratory data bias, questionnaire bias, scale format bias, sensitive question bias, stage bias, unacceptability bias, underlying/contributing cause of death bias, voluntary reporting bias **Data source bias** (competing death bias, family history bias, hospital discharge bias, spatial bias) **Observer bias** (diagnostic suspicion bias, exposure suspicion bias, expectation bias, interviewer bias, therapeutic personality bias) **Subject bias** (apprehension bias, attention bias (Hawthorne effect), culture bias, end-aversion bias (end-of-scale or central tendency bias), faking bad bias, faking good bias, family information bias, interview setting bias, obsequiousness bias, positive satisfaction bias, proxy respondent bias) **Recall bias** (reporting bias, response fatigue bias, unacceptable disease bias, unacceptable exposure bias, underlying cause (rumination bias), yes-saying bias) **Data handling bias** (data capture error, data entry bias, data merging error, digit preference bias, record linkage bias)
Analysis	**Confounding bias** (latency bias, multiple exposure bias, nonrandom sampling bias, standard population bias, spectrum bias) **Analysis strategy bias** (distribution assumption bias, enquiry unit bias, estimator bias, missing data handling bias, outlier handling bias, overmatching bias, scale degradation bias) **Post hoc analysis bias** (data dredging bias, post hoc significance bias, repeated peeks bias)
Interpretation of results	**Assumption bias, cognitive dissonance bias, correlation bias, generalization bias, magnitude bias, significance bias, under-exhaustion bias**
Publication	**All's well literature bias, positive result bias, hot topic bias**

As shown in Table 1, bias can occur at any stage of research, from planning and conduct to analysis, interpretation, and reporting. Therefore, uncertainty should be reported honestly to avoid overstating findings. Bias can also be classified using other tools such as the Cochrane Risk of Bias tool for randomized trials, the Quality of Cohort Studies tool (Q-Coh) for cohort studies, and Risk of Bias in Non-randomized Studies of Interventions (ROBINS-I) for non-randomized intervention studies.[25] Readers who are interested in these approaches can find more details and examples in the literature.

Confounding is commonly viewed as another major source of bias because it can distort the interpretation of an exposure–outcome association. While bias often involves error in how a variable is measured or selected, confounding involves error in what the association means. The observed effect may be partly or entirely due to a third factor related to both the exposure and the outcome. Confounding can exaggerate, reduce, create, or hide an association, even when the collected data are correct.[20,21,26] Like bias, confounding is not solved by increasing sample size. Larger samples reduce random error, but they do not

remove systematic distortion. Confounding can be assessed by comparing estimates before and after adjustment. A substantial change suggests that the factor may be acting as a confounder and should be controlled. Confounding can be controlled during study design through randomization, restriction, or matching, and during analysis through stratification, multivariable analysis, or matched analysis.[21]

Bias remains a major concern in healthcare AI because biased data and confounding can lead to unreliable and unfair results. A 2023 systematic assessment of 48 healthcare AI studies found that 50% had a high risk of bias, often because of missing data, imbalanced or incomplete datasets, or weak algorithm design; only about 20% were judged to have a low risk of bias.[15] It has also been noted in studies using large datasets for precision medicine that AI and machine learning can support prediction and variable selection, but confounders must still be carefully addressed because they can distort results and reduce clinical validity.[27,28] Beyond these technical and statistical concerns, health AI also raises ethical concerns about privacy, confidentiality, fairness, and accountability, especially when sensitive data are used or when datasets do not represent all patient groups well.[18]

Some define bias in healthcare AI as systematic or unfair differences in predictions or study results. The phrase "bias in, bias out" highlights how biased training data can lead to unfair real-world performance.[29] As shown in Table 2, based on the paper by Fereshteh Hasanzadeh et al, bias may occur throughout the AI life cycle. Therefore, comprehensive bias detection and mitigation are needed to support fair, equitable, and reliable healthcare AI.[15]

Table 2. Biases in AI life cycle.

AI model life cycle	major categories/examples of bias
Conception (human introduced biases)	**Implicit bias:** unconscious stereotypes shape decisions. **Systemic bias:** institutional practices reinforce inequities. **Confirmation bias:** evidence is selected or interpreted to fit prior beliefs. **Sensitive attribute bias or concept shift:** changing definitions, coding, or disease patterns alter data meaning over time.
Data collection (data biases)	**Representation bias:** some groups are underrepresented. **Sampling/selection bias:** data or participants do not reflect the target population. **Participation bias:** volunteers differ from non-participants. **Measurement bias:** data are measured differently across groups or sites. **Historical bias:** past inequities are learned and reproduced by AI.
Pre-processing	**Aggregation bias:** distinct groups are inappropriately combined. **Missing data bias:** incomplete data misrepresent some groups. **Feature selection bias:** important variables are omitted. **Representation bias:** processing may worsen under representation.
In-processing (algorithm development and validation)	**Algorithmic bias:** model design or training introduces unfair patterns. **Validation bias:** testing fails to detect poor performance or bias. **Representation bias:** validation data lack enough subgroup diversity.
Post-processing (clinical deployment)	**Evaluation bias:** black-box models limit explainability and oversight. **Predictive bias:** predictions differ unfairly across target groups, affecting care.
Post-deployment surveillance	**Concept drift:** model accuracy declines as data or practice changes. **Automation bias:** clinicians overtrust AI outputs. **Feedback loop bias:** wrong recommendations are reinforced over time. **Dismissal bias/alarm fatigue:** alerts are ignored because of overload or false alarms.

Representation bias can affect several stages of AI research, including data collection, pre-processing, and model validation. Imbalanced data, a common challenge in big data and AI studies, may also be viewed as a form of representation bias when certain groups or outcomes are insufficiently represented.

Inferences in Research: Association and Causation

A central aim of epidemiology is to determine whether an observed association reflects a true cause of disease. This distinction is important because a statistical association does not prove causation, and the absence of an association does not always rule out a causal relationship.[30,31] Causal inference therefore requires careful interpretation of the observed relationship,

including whether it is direct, influenced by confounders, or supported by appropriate assumptions and methods.[32,33] Random error and bias must also be considered. Quantitative bias analysis can further help estimate how findings might change under assumed biases, allowing researchers to judge how strongly the results support a causal conclusion.[22,32]

Evidence of causality may also be weakened by small sample size, lack of randomization, or weak study design.[34] Study design plays an important role in causal inference. Randomized controlled trials generally provide stronger causal evidence because randomization can reduce confounding, whereas observational studies are more vulnerable to bias and confounding and therefore require more cautious interpretation.[34] However, many public health questions cannot be answered through experiments alone. In public health, causal inference is often pragmatic: it can guide action and policy, but it should be applied carefully when data are incomplete, systems are complex, and experiments are not feasible.[35]

There are several models and frameworks that help explain causal inference in epidemiology. The sufficient-component cause model or "causal pies" explains that disease usually develops from a combination of contributing factors, rather than from one cause alone. For example, smoking can increase the risk of lung cancer, but other factors may also be needed, which helps explain why not every smoker develops lung cancer.[25] Early thinking about causal inference in epidemiology was shaped by Bradford Hill's viewpoints, shown in Table 3. These viewpoints became a well-known and commonly used framework for considering whether an observed association might be causal. It should be noted that Hill proposed them as practical guides rather than strict criteria or a rigid checklist.[25,31,36]

Table 3. Hill's Viewpoints for Causal Inference.

Criterion	Brief Definition
Strength of association	A stronger link between exposure and outcome makes causality more likely.
Consistency	Similar findings are seen across different populations, places, times, or study designs.
Specificity	One exposure is linked to one specific outcome. This is rare in real life.
Temporality	The exposure must happen before the outcome.
Biological gradient	Greater exposure leads to greater risk or severity, showing a dose-response pattern.
Biological plausibility	A reasonable biological mechanism can explain the association.
Coherence	The association fits with existing scientific knowledge about the disease.
Experiment	Evidence can be obtained from experimental or semi-experimental studies (e.g., stopping the exposure leads to a reduction in the disease).
Analogy	Similar known causal relationships support the possibility of causation.

Several critiques note that Bradford Hill's viewpoints do not always clearly separate causal from non-causal associations. For example, a true cause may show only a weak association, specificity is often unrealistic because one cause can lead to many outcomes, and dose-response patterns may still result from confounding or bias.[31,37] Because of these limitations, modified approaches based on Hill's viewpoints have been proposed.[35,37,38] Readers interested in these approaches can find more detailed discussions and examples in the literature.

This distinction between association and causation is also important when considering the role of AI in healthcare and health research. Traditional AI applications are mainly association- or correlation-based rather than causal. They are designed to detect patterns, make predictions, classify information, or generate outputs, but they do not directly determine whether one factor causes another. Examples include machine learning for predicting outcomes from patterns, neural networks and deep learning for learning complex relationships in large datasets, large language models (LLMs) for generating text, and computer vision AI for detecting image patterns.[10,39]

Increasingly, AI has expanded beyond traditional "pattern-seeking" systems toward causal AI, which focuses more on "why" and "what-if" questions. Traditional AI asks, "given that I see X, how likely is Y?" and mainly detects correlations or predicts outcomes. Causal AI asks, "if I change X, what will happen to Y?" and "why did Y happen?" It aims to identify cause-and-effect relationships, explain outcomes, and support better intervention-focused decisions. This is why causal AI is considered an important advance: it moves beyond prediction to help decision-makers understand causes, reduce risks, and influence outcomes more

effectively.[9,40–43] In healthcare, traditional AI may show that high cholesterol is linked with cardiovascular disease, while causal AI asks whether high cholesterol is truly a cause or only a related risk marker. Causal AI can also support diagnosis, personalized treatment, drug discovery, clinical trial design, and patient monitoring by estimating which interventions may produce better outcomes.[44]

Because most traditional AI systems are prediction- or pattern-based, explainable AI (XAI) is often used to make these systems more interpretable by clarifying how their outputs are produced. Unlike causal AI, which focuses on estimating cause-and-effect relationships, XAI mainly explains prediction-based AI or machine-learning models. Its purpose is to clarify how or why a model produces a prediction, classification, recommendation, or decision, such as disease-risk prediction, cancer image classification, or treatment recommendation. One important approach in evaluating XAI and LLM-generated explanations is model faithfulness testing, which assesses whether an explanation accurately reflects the model's reasoning process rather than merely appearing plausible. Faithful explanations are essential because imperfect or misleading explanations can have serious consequences, especially in healthcare.[45] Faithfulness differs from factuality: faithfulness refers to whether an answer is consistent with the given input or retrieved context, whereas factuality refers to whether it is objectively true according to real-world or accepted knowledge. Thus, an answer may be faithful to a retrieved source but still factually wrong if the source itself is incorrect.[46] Faithfulness testing is particularly relevant to Retrieval-Augmented Generation (RAG), a technique that integrates external sources, such as medical literature or clinical guidelines, into LLM-generated answers. In RAG systems, faithfulness testing usually involves checking whether claims in the generated answer are supported by the retrieved context. This is especially important in healthcare, where generated explanations or answers should be grounded in reliable source information to reduce inaccurate or hallucinated responses.[47,48] Common tools and systems used for this purpose include RAGAS, MedRAG, OpenEvidence, and others.[49,50]

It should be noted that XAI does not automatically prove causality. It may show which variables are important for prediction, but this does not mean that they caused the outcome. While XAI explains model predictions or decisions, causal AI focuses on cause-and-effect relationships, interventions, and counterfactuals. Therefore, XAI and LLM explanations are often evaluated using model faithfulness testing to assess whether the explanation faithfully reflects the model, the generated answer, or the supporting source evidence. In contrast, causal AI requires causal reasoning frameworks and appropriate assumptions to assess whether one factor is likely to cause another.

Key features of causal AI usually include three related ideas: (1) causal inference, which helps distinguish cause-and-effect relationships from simple associations; (2) structural causal models, which describe how one variable may influence another; and (3) counterfactual reasoning, which asks what might happen under different conditions. Together, these tools help researchers estimate the likely effects of interventions or key exposures on outcomes and support better-informed decisions.[9,39] In practice, causal inference analysis usually begins with a clear causal question and measurable variables. Researchers then develop a map of assumed relationships to identify possible confounders. Next, they choose appropriate methods, such as matching, weighting, or regression, and use sensitivity analysis to assess whether the findings remain credible if some assumptions are violated. Finally, causal effects, such as the average treatment effect (ATE) or conditional average treatment effect (CATE), are interpreted and visualized so the findings can be understood and applied in practice.[51,52]

One key framework in causal AI is Judea Pearl's Causal Hierarchy, also called the Ladder of Causation. It explains causal reasoning in three levels: (1) association, (2) intervention, and (3) counterfactuals. The first level, association or "seeing," focuses on observed patterns and correlations, such as "what is associated with what?". The second level, intervention or "doing," asks what would happen if an action were taken, such as giving a treatment, applying an intervention, or changing the main exposure of interest. The third level, counterfactuals or "imagining," asks what would have happened under a different scenario. This framework helps AI move beyond simple prediction toward deeper causal reasoning about interventions, exposures, and alternative outcomes.[10,42]

Another important framework is the Potential Outcomes Model (POM), also known as the Rubin Causal Model (RCM). This framework estimates the effect of an intervention or main exposure by comparing possible outcomes. For example, it asks what would happen to the same person or group if they were exposed compared with if they were not exposed. In reality, only one of these outcomes can be observed. The observed outcome is called the factual outcome, while the outcome that did not happen is called the

counterfactual outcome. Because this counterfactual outcome is missing, researchers need strong assumptions, careful study design, randomization when possible, and appropriate statistical adjustment. Causal effects can then be estimated using measures such as the individual treatment effect (ITE).[10,33,42]

While Pearl's hierarchy provides a roadmap for causal reasoning, causal AI commonly uses complementary tools such as Directed Acyclic Graphs (DAGs), Structural Causal Models (SCMs), and the Potential Outcomes Model (POM). DAGs provide a causal map by showing assumed relationships among variables and helping identify pathways, confounders, and whether causal effects may be estimated from observational data. SCMs add mathematical structure to this map through equations, allowing researchers to examine interventions and counterfactual reasoning. POM focuses on estimating causal effects by comparing potential outcomes under different exposures or treatments. Together, these approaches make causal assumptions clearer and help causal AI move from observed associations toward intervention-based and counterfactual inference.[10,33,40,42,53]

To judge whether causal inference is strong enough, researchers need to examine the assumptions behind causal reasoning. In causal discovery, important principles include: conditional dependence, temporal precedence, and the common cause principle.[54] In simple terms, a possible cause should be related to the outcome, happen before the outcome, and be evaluated together with other possible hidden causes. These assumptions help prevent researchers from mistaking patterns in the data for true causal effects. Another common approach focuses on estimating the average causal effect, using three important assumptions: exchangeability, positivity, and consistency.[55] Exchangeability means the comparison groups are sufficiently similar, with no unmeasured confounding or informative censoring. Positivity means each relevant subgroup has a real chance of receiving each exposure level. Consistency means the exposure is clearly defined, so the observed outcome matches the potential outcome under that same exposure. These assumptions are commonly used in causal machine-learning research because they help define when causal effects can be estimated from data. They guide how researchers select variables, adjust for confounding, and judge whether machine-learning-based causal estimates are credible.[56,57]

Recently, Yu Yunguo discussed a causal explanation framework for clinical AI. The paper proposed a set of core metrics to evaluate whether AI-generated explanations contain the intended causal structure, as shown in Table 4. The paper also noted that these metrics mainly assess the structural and evidence-based organization of the explanations, rather than proving that the proposed causal pathways are truly correct or that they reflect the model's actual internal reasoning. Therefore, as in other forms of XAI, model faithfulness testing is still needed to assess model fidelity, while expert clinician review is required to assess the clinical and causal correctness of the proposed pathways.[58]

Table 4. Causal explanation framework for clinical AI.

Core metric	Definition
Fidelity	Degree to which the causal explanation is consistent with the AI prediction and relevant clinical knowledge.
Completeness	Extent to which the explanation covers important causal pathways and mechanistic reasoning
Clinical alignment	Degree to which the explanation follows accepted clinical reasoning patterns and guideline-based logic.
Evidence quality	Strength, relevance, and appropriateness of evidence supporting the causal relationships.
Clarity	Extent to which the explanation is understandable, organized, and easy to follow.
Structural reliability	Consistency, verifiability, and reliability of the causal explanation structure.

It should be noted that evaluating causality in AI is difficult because real-world data are often complex, incomplete, and biased. Confounders, limited causal labels, missing data, observational bias, and unclear causal assumptions can make simple correlations appear to be true causal effects. Causal AI also faces challenges in model validation, explainability, computational complexity, and proving that its causal explanations reflect real causal mechanisms. Therefore, causal AI still requires good data quality, careful assumptions, appropriate methods, and expert judgment.[51,59] The future of causal AI will likely be a hybrid field that brings together causal inference, deep learning, automated causal discovery, and explainable AI. It will integrate data-driven learning with causal reasoning methods to make AI systems less opaque and more transparent, interpretable, and grounded in causal logic.[42,51]

Final Words

Responsible AI should not end with faster analysis or more automated outputs. The real challenge is whether AI can support research that is valid, fair, explainable, and useful in practice. This requires more than technical performance; it requires attention to confounding, bias, transparency, human oversight, and regulatory accountability.[3,14,44] Causal AI has shown promise in helping researchers move beyond prediction toward clearer explanation, better decision support, and more trustworthy use of AI in health and biomedical research. I do not know what the future will bring in this fast-moving AI world. Perhaps AI will become as imperfectly perfect, or as flawlessly flawed, as humans. Maybe the real promise of AI is not that it thinks like us, but that it may eventually reason with better logic than us, helping us see beyond human biases and draw more accurate inferences.

Acknowledgements

An artificial intelligence tool, ChatGPT (OpenAI, 2025), was used for language support during the preparation of this manuscript.[60] All generated content was carefully reviewed for accuracy and appropriateness. The author revised and edited the material and assumes full responsibility for the final content.

References

1. World Health Organization. WHO issues first global report on artificial intelligence (AI) in health and six guiding principles for its design and use [Internet]. Geneva: World Health Organization; 2021 Jun 28 [cited 2026 May 1]. <https://www.who.int/news/item/28-06-2021-who-issues-first-global-report-on-ai-in-health-and-six-guiding-principles-for-its-design-and-use>
2. Nuffield Council on Bioethics. Bioethics briefing note: artificial intelligence (AI) in healthcare and research [Internet]. London: Nuffield Council on Bioethics; 2018 May [cited 2026 May 1]. 8 p. <https://www.nuffieldbioethics.org/wp-content/uploads/Artificial-Intelligence-AI-in-healthcare-and-research-1.pdf>
3. Krubiner C, Studman A, Singh S. How is AI reshaping health research? [Internet]. London: Wellcome; 2023 Nov 1 [cited 2026 May 1]. <https://wellcome.org/insights/articles/how-ai-reshaping-health-research>
4. National Aeronautics and Space Administration. What is artificial intelligence? [Internet]. Washington: NASA; 2024 May 13 [cited 2026 May 1]. <https://www.nasa.gov/what-is-artificial-intelligence/>
5. Panteli D, Adib K, Buttigieg S, Goiana-da-Silva F, Ladewig K, Azzopardi-Muscat N, et al. Artificial intelligence in public health: promises, challenges, and an agenda for policy makers and public health institutions. Lancet Public Health. 2025 May;10(5):e428–32. doi:10.1016/S2468-2667(25)00036-2.
6. Chumakov E. History and definitions of AI. Eur Psychiatry. 2025 Aug 26;68(Suppl 1):S61. doi:10.1192/j.eurpsy.2025.238.
7. Foote KD. A brief history of generative AI [Internet]. Studio City (CA): Dataversity; 2025 Sep 10 [cited 2026 May 1]. <https://www.dataversity.net/articles/a-brief-history-of-generative-ai>
8. Ordak M. Artificial intelligence-assisted statistical analysis and statistical review: evidence (2023-2025) and implications for internal medicine. Pol Arch Intern Med. 2026 Mar 24;136(3):17243. doi:10.20452/pamw.17243.
9. Takyer A. Causal AI: importance, use cases, benefits, challenges and implementation strategies [Internet]. Gurugram: LeewayHertz; [cited 2026 May 1]. <https://www.leewayhertz.com/causal-ai/>
10. Jiao L, Wang Y, Liu X, Li L, Liu F, Ma W, et al. Causal inference meets deep learning: a comprehensive survey. Research (Wash D C). 2024;7:0467. doi:10.34133/research.0467.
11. Karimov A. History of artificial intelligence [Internet]. Zurich: Swiss Cyber Institute; 2026 Jan 21 [cited 2026 May 1]. <https://swisscyberinstitute.com/blog/history-artificial-intelligence/>
12. Fahim YA, Hasani IW, Kabba S, Ragab WM. Artificial intelligence in healthcare and medicine: clinical applications, therapeutic advances, and future perspectives. Eur J Med Res. 2025 Sep 23;30(1):848. doi:10.1186/s40001-025-03196-w.
13. Deonandan R. Role of artificial intelligence in health research: opportunities, challenges, and implications for medical education. Cureus. 2026;18(4):e106579. doi:10.7759/cureus.106579.
14. Dinga R, Schmaal L, Penninx B, Veltman DJ, Marquand AF. Controlling for effects of confounding variables on machine learning

predictions. bioRxiv [preprint]. 2020 Aug 18 [cited 2026 May 1]. <https://doi.org/10.1101/2020.08.17.255034>

15. Hasanzadeh F, Josephson CB, Waters G, Adedinsewo D, Azizi Z, White JA. Bias recognition and mitigation strategies in artificial intelligence healthcare applications. NPJ Digit Med. 2025 Mar 11;8(1):154. doi:10.1038/s41746-025-01503-7.
16. American University of Beirut. Artificial intelligence in healthcare: AI in health research [Internet]. Beirut: American University of Beirut; [cited 2026 May 1]. <https://aub.edu.lb.libguides.com/c.php?g=1486495&p=11088096>
17. Lane Medical Library, Stanford Medicine. AI tools and resources for biomedical research [Internet]. Stanford (CA): Stanford Medicine; [cited 2026 May 1]. <https://laneguides.stanford.edu/AI/tools>
18. Williams B. Use of AI in Health Data Analysis for Research [Internet]. Gainesville: Insight7; [cited 2026 May 1]. <https://insight7.io/artificial-intelligence-in-data-analysis-for-research/>
19. Smith J, Noble H. Understanding sources of bias in research. Evid Based Nurs. 2025 Jun 19;28(3):137–9. doi:10.1136/ebnurs-2024-104231.
20. HealthKnowledge, Faculty of Public Health. Biases and confounding [Inernet]. London: Faculty of Public Health; [cited 2026 May 1]. <https://www.healthknowledge.org.uk/public-health-textbook/research-methods/1a-epidemiology/biases>
21. Steingart KR. Systematic reviews of diagnostic test accuracy [Internet]. Montreal: TeachEpi; [cited 2026 May 1]. 85 p. <https://www.teachepi.org/wp-content/uploads/OldTE/documents/courses/tbdiagrx2/day4/Steingart%20SRs%20of%20DTA_Montreal_FINAL_r1.pdf>
22. Walker AR, Venetis CA, Opdahl S, Chambers GM, Jorm LR, Vajdic CM. Estimating the impact of bias in causal epidemiological studies: the case of health outcomes following assisted reproduction. Hum Reprod. 2024 May 2;39(5):869–75. doi:10.1093/humrep/deae053.
23. Althubaiti A. Information bias in health research: definition, pitfalls, and adjustment methods. J Multidiscip Healthc. 2016 May 4;9:211–7. doi:10.2147/JMDH.S104807.
24. Sackett DL. Bias in analytic research. J Chronic Dis. 1979;32(1–2):51–63. doi:10.1016/0021-9681(79)90012-2.
25. Watkins TR. Understanding uncertainty and bias to improve causal inference in health intervention research [dissertation]. Sydney: School of Public Health, Faculty of Medicine and Health, University of Sydney; 2019 Jul [cited 2026 May 1]. 373 p. <https://ses.library.usyd.edu.au/handle/2123/20772>
26. Greenland S, Morgenstern H. Confounding in health research. Annu Rev Public Health. 2001;22:189–212. doi:10.1146/annurev.publhealth.22.1.189.
27. Benasseur I, Talbot D, Durand M, Holbrook A, Matteau A, Potter BJ, et al. A comparison of confounder selection and adjustment methods for estimating causal effects using large healthcare databases. Pharmacoepidemiol Drug Saf. 2022 Apr;31(4):424–33. doi:10.1002/pds.5403.
28. Komeyer V, Eickhoff SB, Grefkes C, Patil KR, Raimondo F. A framework for confounder considerations in AI-driven precision medicine. medRxiv [Preprint]. 2024 [cited 2026 May 1]. <https://doi.org/10.1101/2024.02.02.2430219>
29. Ott MA. Bias in, bias out: ethical considerations for the application of machine learning in pediatrics. J Pediatr. 2022 Aug;247:124. doi:10.1016/j.jpeds.2022.01.035.
30. Harvard Graduate School of Education. Unit 4 - causal and non-causal associations [Internet]. Cambridge (MA); Harvard Graduate School of Education; [cited 2026 May 1]. <https://s040.gse.harvard.edu/unit4.html>
31. HealthKnowledge, Faculty of Public Health. Association and causation [Internet]. London: Faculty of Public Health; [cited 2026 May 1]. <https://www.healthknowledge.org.uk/public-health-textbook/research-methods/1a-epidemiology/association-causation>
32. Karamitros G, Grant MP, Lamaris GA. Associations in medical research can be misleading: a clinician's guide to causal inference. J Surg Res. 2025 Jun;310:145–54. doi:10.1016/j.jss.2025.03.043.
33. Raden N. How causal analysis and AI intersect-methods of causal inference [Internet]. Wallington: Diginomica; 2023 Oct 24 [cited 2026 May 1]. <https://diginomica.com/how-

causal-analysis-and-ai-intersect-methods-causal-inference>

34. Osborne V, Shakir SAW. What is the difference between observed association and causal association, signals and evidence? examples related to COVID-19. Front Pharmacol. 2021;11:569189. doi:10.3389/fphar.2020.569189.

35. Glass TA, Goodman SN, Hernan MA, Samet JM. Causal inference in public health. Annu Rev Public Health. 2013;34:61–75. doi:10.1146/annurev-publhealth-031811-124606.

36. Hill AB. The environment and disease: association or causation? 1965. J R Soc Med. 2015 Jan;108(1):32–7. doi:10.1177/0141076814562718.

37. Rothman KJ, Greenland S. Causation and causal inference in epidemiology. Am J Public Health. 2005;95 Suppl 1:S144–50. doi:10.2105/AJPH.2004.059204.

38. Shimonovich M, Pearce A, Thomson H, Keyes K, Katikireddi SV. Assessing causality in epidemiology: revisiting Bradford Hill to incorporate developments in causal thinking. Eur J Epidemiol. 2021 Sep;36(9):873–87. doi:10.1007/s10654-020-00703-7.

39. Matthay EC, Neill DB, Titus AR, Desai S, Troxel A, Cerda M, et al. Integrating artificial intelligence into causal research in epidemiology. Curr Epidemiol Rep. 2025;12(6). doi:10.1007/s40471-025-00359-5.

40. Chauhan VK, Dhami DS, Gao B, Wang X, Clifton L, Clifton DA. Beyond correlations: the necessity and the challenges of causal AI. TechRxiv [Preprint]. 2025 Aug 18 [cited 2026 May 1]. <https://www.techrxiv.org/doi/full/10.36227/techrxiv.175554759.96327720/v1>

41. Pearl J, Mackenzie D. The book of why: the new science of cause and effect. New York: Basic books; 2018.

42. Uppili S. Causal AI in 2026: use cases, tools & how it works [Internet]. Austin (TX): Kanerika Inc; 2025 Oct 14 [cited 2026 May 1]. <https://kanerika.com/blogs/causal-ai/>

43. Lee H, Qiu S, Haupert S, Innes GK, Naumann T, Alemayehu D, van der Laan M. Methodological and regulatory considerations for causal AI in drug development. NPJ Digit Med. 2026 Feb 27;9(1):295. doi:10.1038/s41746-026-02477-w.

44. Lee AG. Causal AI: current state-of-the-art & future directions [Internet]. San Francisco; Medium; 2025 Mar 7 [cited 2026 May 1]. <https://medium.com/@alexglee/causal-ai-current-state-of-the-art-future-directions-c17ad57ff879>

45. Jacovi A, Goldberg Y. Towards faithfully interpretable NLP systems: How should we define and evaluate faithfulness? [Internet]. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics; 2020 Jul 5–10. Stroudsburg (PA): Association for Computational Liguistics; 2020. 8 p.

46. Fadeeva E, Rubashevskii A, Piatrashyn D, Vashurin R, Dhuliawala S, Shelmanov A, et al. Faithfulness-aware uncertainty quantification for fact-checking the output of retrieval augmented generation. arXiv:2505.21072 [Preprint]. 2025 [cited 2026 May 1]. <https://doi.org/10.48550/arXiv.2505.21072>

47. Meleka M. Retrieval-augmented AI assistants for healthcare: system design and evaluation [dissertation]. Oxford: Kellogg College, University of Oxford; [cited 2026 May 1]. <https://ora.ox.ac.uk/objects/uuid:9add0c17-f4fe-4051-9a2a-027f8818a5aa/files/d3197xm748>

48. Gargari OK, Habibi G. Enhancing medical AI with retrieval-augmented generation: A mini narrative review. Digit Health. 2025 Apr 21;11:20552076251337177. doi:10.1177/20552076251337177.

49. Zhao X, Liu S, Yang SY, Miao C. MedRAG: enhancing retrieval-augmented generation with knowledge graph-elicited reasoning for healthcare copilot. arXiv:2502.04413 [Preprint]. 2025 [cited 2026 May 1]. <https://doi.org/10.48550/arXiv.2502.04413>

50. Low YS, Jackson ML, Hyde RJ, Brown RE, Sanghavi NM, Baldwin JD, et al. Answering real-world clinical questions using large language model, retrieval-augmented generation, and agentic systems. Digit Health. 2025 Jun 9;11:20552076251348850. doi:10.1177/20552076251348850.

51. Stane D. Powering causal inference analysis with machine learning [Internet]. San Diego: Teradata; 2025 Nov 28 [cited 2026 May 1]. <https://www.teradata.com/insights/ai-and-machine-learning/powering-causal-inference-analysis>

52. Parikh H, Varjao C, Xu L, Tchetgen E. Validating causal inference models [Internet]. Seattle (WA): Amazon Science; 2022 [cited 2026 May 1]. <https://www.amazon.science/publications/validating-causal-inference-models#>

53. Boge F, Mosig A. Causality and scientific explanation of artificial intelligence systems in biomedicine. Pflugers Arch. 2025 Apr;477(4): 543–54. doi:10.1007/s00424-024-03033-9.

54. Reid A. Causal discovery: an introduction [Internet]. Tilburg (NL): Andrew Reid; 2024 Sep 23 [cited 2026 May 1]. <https://www.reid-lab.org/blog>

55. Cole SR, Frangakis CE. The consistency statement in causal inference: a definition or an assumption? Epidemiology. 2009 Jan;20(1):3–5. doi:10.1097/EDE.0b013e31818ef366.

56. Im DJ, Cho K. Active and passive causal inference learning. arXiv:2308.09248 [Preprint]. 2024 Aug 25 [cited 2026 May 1]. <https://arxiv.org/pdf/2308.09248>

57. Jeon N, Brand JE. Causal machine learning: a deductive–inductive framework for sociological research. Köln Z Soziol. 2026. doi:10.1007/s11577-026-01053-0.

58. Yunguo Yu. A causal explanation framework for clinical AI: methods and technical evaluation. hal-05484638 [Preprint]. 2026 Jan 30 [cited 2026 May 1]. <https://hal.science/hal-05484638>

59. Kartik NVJK. Evaluating causality in AI models in 2026: how to build AI that reasons beyond correlation [Internet]. San Francisco: Future AGI; 2025 Jan 30 [cited 2026 May 1]. <https://futureagi.com/blog/evaluating-causality-in-ai-models/>

60. OpenAI. ChatGPT (GPT-5.5 Thinking) [Internet]. San Francisco: OpenAI; 2026 [cited 2026 May 1]. <https://chat.openai.com/>