Global Health Equity

Commentary

Guarding against ‘bias-accumulation’ in knowledge systems in the AI-era

Shiva Raj Mishra[1,2,3*], Ferdinand C. Mukumbang[4], Bipin Adhikari[5,6]

Received: July 12,2025 Accepted: November 23,2025

Introduction

The exponential growth of artificial intelligence (AI) has revolutionized numerous fields, offering unprecedented capabilities in data analysis, prediction, and automation. These large language models (LLMs) can process, understand, and generate human language based on enormous amounts of data that are fed into it. AI is dependent on the types and nature of data that they are trained to access and process. LLMs are trained on large amounts of text, predominantly obtained from the internet to respond to questions, provide summaries, translate and create stories and other forms of textual outputs [1]. Nonetheless, AI forms much of its foundational operating system based on the existing reserve of knowledge on the internet. But is the existing resource of knowledge in the internet adequate, saturated and not to question, (fairly) representative of the population that are either underserved by the internet or are systematically deprived from the modern databases? This question already poses an input bias to these LLIMs and are indeed likely to generate bias in processing and output.

Based on their dependence on the data they can access, a further question to consider is: how sustainable and reliable is AI's output if its foundational data are still developing? Data, as we know it, cannot grow meaningfully without continuous investment in primary research especially in global health, a field driven by evidence-based practice that continues to evolve by the empirical research.

Evidence-based practice and policymaking are the bedrock of public health. Primary and indeed secondary research (including systematic reviews which are also based on primary research) are used to generate the evidence that guide global health practice including policymaking. Consequently, evidence or information obtained through data retrieval and those considered as secondary sources (social media, internet searches, wearables) all form part of the data ecosystem, which is accessed by LLMs to answer questions, provide summaries and guide disease diagnosis for the general public.

Dr Ilya Sutskever, co-founder and former chief scientist at OpenAI which runs ChatGPT, speaking at the NeurIPS 2024 in Vancouver, Canada described data as the main limitation for AI models going forward [2]. In Ilya’s words, “We have but one internet”. Hidden in the pretext, a heavy reliance on AI models that scrape and recycle AI-generated internet content compounds the risk of amplifying biases and inaccuracies in AI outputs. Referring to deep learning revolution, Ilya said, “... more it [AI systems] reasons, more unpredictable it becomes”. Mimicking human ‘reasoning’ is complex and prone to biases; like humans fail to unfathom[ed] our own biases based on where we were born, live, work and our environment around [1,3].

Historical lessons from natural world examples offer a stark warning. The infamous Minamata disease, a neurological disorder caused by methylmercury poisoning, emerged from bioaccumulation—where contaminants progressively concentrated through the food chain [3]. Fish and shellfish in Japan's Minamata Bay were poisoned by methylmercury compounds discharged from a chemical plant, devastating human and ecological health. Just as toxic compounds accumulate in ecosystems, unchecked biases, imperfections, developing facts can propagate and intensify across AI systems, perpetuating systemic flaws [3,4]. This analogy underscores the urgency of implementing stronger safeguards against the accumulation of bias in AI technologies [4].

The Risk of Data Recycling

AI models are increasingly trained on vast quantities of online data. With the growth in AI models, it is likely that AI generated outputs are recycled into training large clusters of AI models those which are currently in operation, or forthcoming. While this approach accelerates model training, it also creates a closed feedback loop. Unlike mathematical models which perform comparatively better, linguistical models bear higher risks of reduction and generalization underestimating the enormity of linguistic complexity, meanings, interpretations, construction, cues, and nuances. In addition, the real-life scenario of how a particular intervention works, for who and how is even more complex, just as complex as human behavioral patterns are. Following a set of trends, patterns, and processes available so far risks simplifying the intricacies of interaction between science and the society.

Assuming an existing sets of data on cultural practice in a remote community in Laos for example, inadvertently adds to the narrowed understanding and insights that can inevitably lead to a futile model. Both cultural practices, and interventions evolve, including their interactions, and are thus dynamic; any assumption of 'saturation' and premature interpretation can lead to flawed outcomes. All primary research are prone to errors, biases, and gaps, thus these inputs can lead to a biased knowledge models including their recycling and reproduction. For example, studies have shown that large language models reflect and sometimes amplify societal stereotypes present in their training data. This can lead to discriminatory outcomes in critical areas such as hiring algorithms, loan approvals, and healthcare recommendations [3,4]. Indeed such models can become parochial and circumscribed, also aligning with the discourses on how current global health knowledge suffers from epistemic injustice including the unfair practices in global health research.

The dependence on such self-referential data pipelines risks diminishing the diversity, depth and evolutionary properties of empirical datasets that AI models rely upon. Diversity in training data is crucial for AI systems to approach effective generalization and equitable representation across varied contexts. Without continuous empirical research, AI systems may continue to propagate and exacerbate systemic flaws in data—recycling the errors and biases.

The Need for Guardrails

Guarding against bias in AI systems requires a multifaceted approach. First and foremost, investment in primary research must be prioritized. Diverse types of empirical research are critical to produce multi-dimensional aspects of a disease or an intervention including its interaction with the communities. Such investment can integrate diverse datasets, from qualitative to quantitative, to inform the design of community-acceptable interventions, strategies, and policies. For example, in healthcare, the inclusion of underrepresented populations in clinical trials ensures that AI models trained on such data provide equitable recommendations across de-

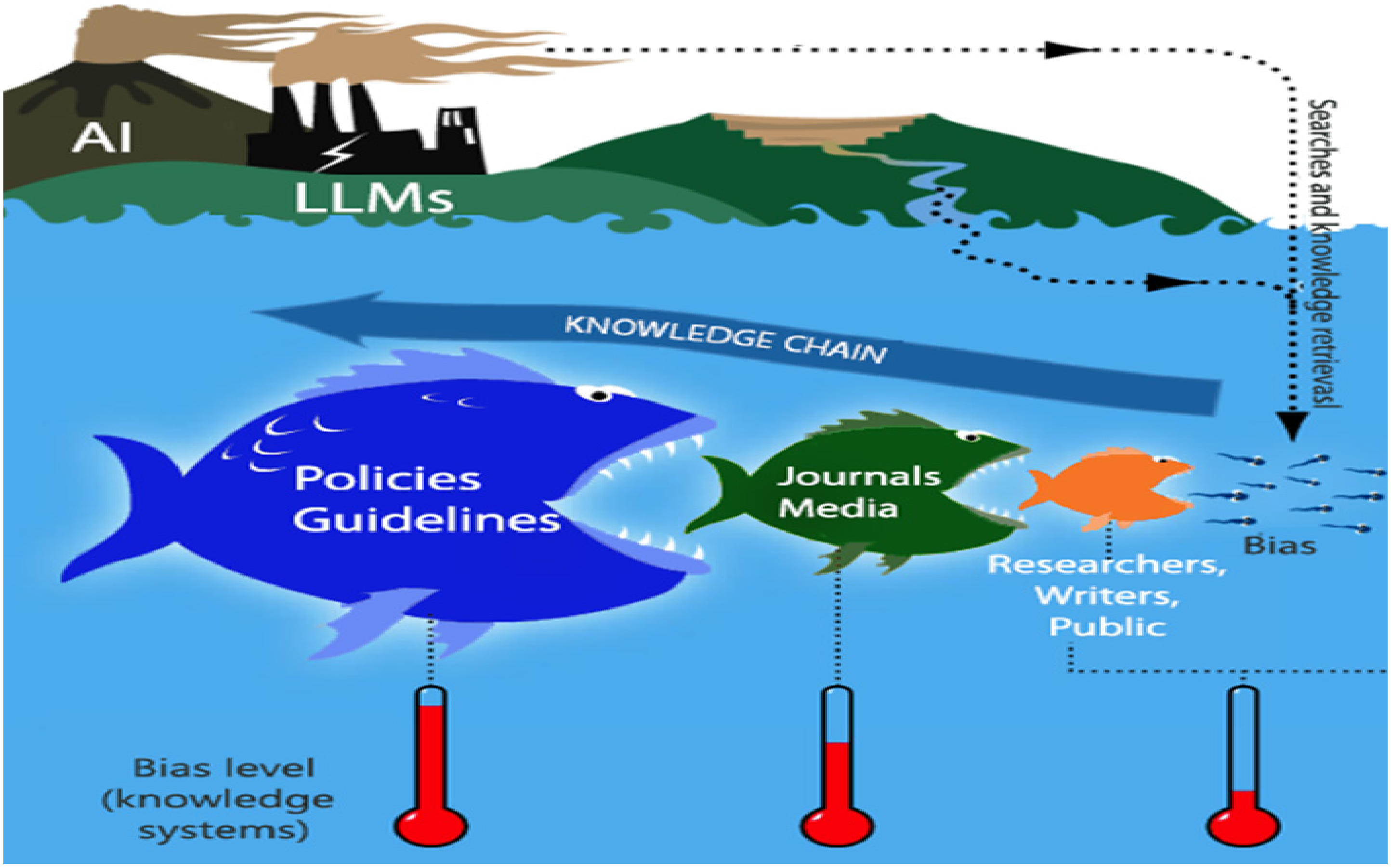

Figure 1. Bias accumulation in knowledge systems with advent of AI and LLMs. Adapted using historical example of Minamata Diaster[3,5]

mographics [6].

Second, the development of robust regulatory frameworks to oversee the AI generated data is essential. Just as environmental regulations mitigate industrial pollution, stringent guidelines can prevent the accumulation of bias in AI systems. These frameworks must mandate transparency in data sourcing, ensure diverse representation in training datasets, and require continuous monitoring of AI outputs for signs of bias.

Lastly, interdisciplinary collaboration between AI developers, ethicists, sociologists, and content experts can help identify and mitigate potential biases early in the development process. This collaborative approach mirrors successful strategies in environmental management, where scientists, policymakers, and industry stakeholders work together to address complex challenges.

Lessons from the Natural World

The parallels between bioaccumulation and AI bias accumulation serve as a cautionary tale. In the case of Minamata disease, the absence of preventive measures allowed pollutants to devastate communities and ecosystems.1,3 Ironically, the pollution was not stopped even after 12 years of identification of methylmercury in 1956 [7]. Much like the powerful industry lobby and low community awareness in 1956, efforts to curb AI bias today suffer from a lack of advocacy and critical awareness among the public and policymakers. Likewise, if we fail to address the root causes of bias in AI systems, the consequences could be far-reaching, affecting societal trust, equity, and well-being.

The lessons from Minamata emphasize the importance of proactive intervention. Environmental disasters have taught us that investing in prevention is far more economical than remediation when considering the expenses related to hidden, opportunity and disease-related costs. In AI, this means investing in ethical and responsible data practices from the outset. Just as environmental impact assessments are required for industrial projects, ethical impact assessments should be mandatory for AI systems, particularly those deployed in high-stakes domains like healthcare, finance, and criminal justice.

Conclusion

As AI becomes increasingly integrated into the fabric of society, its potential to benefit humanity is immense. However, this potential can only be realized if we address the systemic flaws in its development and deployment. Investment in primary research, coupled with strong regulatory frameworks and interdisciplinary collaboration, can help mitigate the risks of bias accumulation in AI systems. By learning from historical and natural analogies, such as the bioaccumulation of pollutants, we can chart a more ethical and sustainable path forward for AI.

The call for action is clear: to build AI systems that are not only intelligent but also just and equitable, we must prioritize the quality and diversity of the data that underpin them. Mimicking biological systems without understanding inheritance of bias, propagation and accumulation will increase AI associated risk for human civilization. Without risk mitigation measures, we risk perpetuating and amplifying the very inequalities that AI has the potential to solve.

Author Affiliations

[1] *Nepal Development Society, Bharatpur-6, Chitwan, Nepal*

[2] *NHMRC Clinical Trials Centre, University of Sydney, Australia*

[3] *Westmead Applied Research Centre, University of Sydney, Australia*

[4] *Department of Global Health, School of Public Health, University of Washington, Seattle, Washington, USA*

[5] *Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand*

[6] *Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom*

Corresponding Author

Shiva Raj Mishra, MPH, PhD,

School of Medicine, Western Sydney University

Email: s.mishra5@westernsydney.edu.au

References

1. Clusmann J, Kolbinger FR, Muti HS, et al. The future landscape of large language models in medicine. Communications medicine 2023; 3(1): 141.
2. Ilya Sutskever. NeurIPS 2024 in Vancouver, Canada. . 2024. https://www.youtube.com/watch?v=1yvBqasHLZs (timestamp 8:48) (accessed December 12 2024).
3. McAlpine D, Araki S. Minamata disease. An unusual neurological disorder caused by contaminated fish. 1958.
4. Binns R, Van Kleek M, Veale M, Lyngs U, Zhao J, Shadbolt N. 'It's Reducing a Human Being to a Percentage' Perceptions of Justice in Algorithmic Decisions. Proceedings of the 2018 Chi conference on human factors in computing systems; 2018; 2018.

p. 1-14.

5. Ground Truth Alaska. How Does Mercury End Up in Your Food?Mercury in fish. https://groundtruthalaska.org/static/uploads/files/MercuryFoodChain.svg69md34/MercuryFoodChain.svg (accessed December 12 2024).

6. Food and Drug Administration. Enhancing the Diversity of Clinical Trial Populations — Eligibility Criteria, Enrollment Practices, and Trial Designs Guidance for Industry. 2024. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/enhancing-diversity-clinical-trial-populations-eligibility-criteria-enrollment-practices-and-trial (accessed December 12 2024).

7. Yorifuji T. Lessons from an early-stage epidemiological study of Minamata disease. Journal of epidemiology 2020; 30(1): 12-4.