Tidsskriftet Tidsskrift for Den norske legeforening

Artificial intelligence – one brain for all?

PERSPECTIVES

ERIK LANG

erik.lang@nhh.no

Erik Lang is a PhD student at the Department of Strategy and Management, Norwegian School of Economics (NHH) and is affiliated with the research centre DIG – Digital Innovation for Sustainable Growth. His research examines artificial intelligence and how it shapes professional practices, decision-making processes and organisation.

The author has completed the ICMJE form and declares no conflicts of interest.

Whether artificial intelligence helps or not depends less on whether it truly understands and more on how it is put to use.

Let us reflect on what artificial intelligence (AI) is good at today. A large language model – the technology behind tools such as ChatGPT and Claude – can read a patient's entire medical record together with a vast amount of relevant literature and produce a complex conclusion within seconds. It can identify patterns, interpret signals and arrive at a diagnosis, thereby assisting the human clinician in a meaningful way (1). In this article, 'artificial intelligence' refers to such language-generating models. AI is developing rapidly, and for some tasks these large language models already outperform what a single person, or even a team, can achieve (2). Yet despite this potential, AI models, like humans, can make mistakes and carry their own bias.

Meet your new colleague

Two types of bias are key to the argument that follows. Idiosyncratic bias affects individual clinicians: the clinician may be wrong, but in their own personal way. Systematic bias on the other hand, extends beyond the individual: in a large group of clinicians, many may make the same mistake, thus affecting the average outcome; or a

treatment or tool may repeatedly produce the same unusual or unfortunate result thus similarly affecting the average outcome. These two types of bias impact different tasks in different ways.

Let us imagine that an additional colleague is recruited to our team. Like the rest of us, this doctor is not infallible, but is more widely read than anyone else on the team. The colleague works extremely fast and, for the most part, makes excellent decisions. This colleague will generally improve outcomes, freeing up time for patient care and other duties. Despite their idiosyncratic bias, this doctor is so good that we want them on the team.

Next, imagine that we replicate this colleague across every hospital in the country. Each hospital could allocate part of its workload to this additional brain. The individual decisions will still improve, but all patients are now treated by the same person. The idiosyncratic bias that was previously confined to a single clinician would now be propagated throughout the health system. If this doctor were particularly weak in one area, the resulting errors would be replicated in all hospitals. At that point, idiosyncratic bias becomes systematic bias. We would lose the variation between independent clinicians who otherwise – collectively – might have identified the problem as a team.

A single human being cannot, of course, be replicated in this way, but artificial intelligence can be, and already has been. The use of AI in medical practice raises two immediate questions. First, is this artificial brain sufficiently similar to a human brain that it lends itself to the above type of reasoning, comparing biases and strengths? Second, does this artificial brain genuinely understand, or does it merely simulate understanding? Both questions merit attention, but ultimately this article argues that the key difference does not lie in how AI works or whether it truly understands, but in what happens when many independent human brains are replaced by a single artificial brain that is used by everyone.

> ***«The key difference does not lie in how AI works or whether it truly understands, but in what happens when many independent human brains are replaced by a single artificial brain that is used by everyone»***

What kind of brain are we dealing with?

A large language model and a human brain may look different, but are surprisingly similar on the levels relevant to the argument.

Current neuroscience accounts describe the human brain as a prediction machine. It is not a passive receiver of information but continuously generates expectations about incoming signals and revises them when reality deviates. These coordinated guesses (3) can in essence be understood as Bayesian updating (4): the brain adjusts its predictions as new evidence becomes available. As we read this sentence, the brain adjusts its expectations about its meaning, word by word, as the sentence unfolds.

A large language model is trained on vast amounts of text to predict the next word in a given context. It does not store or retrieve ready-made sentences; it carries a compressed mathematical representation of how language works, and uses this to reconstruct responses. There are two kinds of updating involved. During training, the model's internal parameters –referred to as 'weights' – are adjusted in each cycle. Once training is complete and the model is deployed, the weights are no longer updated. However, a different type of updating takes place within a single conversation. Each new word that is received or written changes the probability distribution of the next word. The predictions are updated word by word as new information becomes available.

> ***«It is worth noting that these mechanisms in large language models are remarkably similar to the continuous Bayesian updating process in the human brain»***

It is worth noting that these mechanisms in large language models are remarkably similar to the continuous Bayesian updating process in the human brain. In fact, the artificial systems that most accurately predict human brain activity during language comprehension are precisely these prediction-trained large language models (5–7).

Does the artificial brain actually understand?

In a recent article in this journal, Næss drew the opposite conclusion (3). He used the concept of generative reconstruction – the idea that each human brain builds understanding internally rather than by direct transfer from other brains – to argue that AI is fundamentally different from human cognition. The evidence outlined above, however, suggests that the framework points in the opposite direction. Prediction and reconstruction are precisely what human brains and large language models have in common. Seen through the Bayesian framework they become more similar, not less.

This similarity raises a deeper question: does AI genuinely understand what it is doing? And does the answer matter for the artificial colleague we are considering hiring? Searle's Chinese Room thought experiment provides a clear illustration of the problem (8). A person who does not speak Chinese sits in a closed room. The person receives Chinese characters through a slot in the wall and then follows an extensive set of rules that specifies which characters should be returned. To a Chinese speaker outside the room, the exchange looks like a fluent conversation, but the person inside the room has no understanding of any of the symbols. Searle would argue that a large language model does exactly the same thing: it manipulates symbols without understanding them.

No individual neuron in the human brain understands English or what a disease is. By the same logic: if we deny the Chinese Room understanding because its components do not understand, we must also deny the human brain understanding. The so-called systems reply summarises this argument: understanding can either arise in a system even though its components do not possess such understanding, or understanding does not exist anywhere. If behaviour is sufficient to attribute understanding to humans, then there must be an explicit justification for not attributing it to AI systems that exhibit the

same behaviour (9). Moreover, we only have access to our own subjective experience, never that of others (10). We attribute understanding to others on the basis of their behaviour, not on an inspection of their internal state of mind.

Philosophically and clinically, the question is a dead end. The more important difference lies not so much in mechanism or in understanding, but in distribution. This article reaches a different conclusion from Næss regarding how Bayesian updating distinguishes artificial intelligence from the human brain, yet the framework is a productive tool for seeing how the concept of 'coordinated guessing' can help us better understand the use of AI in the health sector.

The asymmetry that matters

Communication between two human brains is a process of coordinated guessing rather than direct transfer of information. This has important implications at population level. If every brain reconstructs understanding individually, then human cognition is fundamentally diverse. This implies billions of unique reconstructions, each shaped by a particular history, body, and set of priors. Even when two clinicians read the same case note, they may reconstruct it differently. Two clinicians can of course have the same bias, but the variation is not merely noise; it is part of what enables a group to identify errors or find solutions that a single person may overlook.

> ***«Even when two clinicians read the same case note, they may reconstruct it differently. Two clinicians can of course have the same bias, but the variation is not merely noise; it is part of what enables a group to identify errors»***

Multiple deployments of a single large language model is the opposite. One set of trained parameters is distributed to every user. Although today's models have very long context windows, in the order of millions of words, and although persistent memory across conversations is already in use, the fundamental underlying model remains the same. Personalisation is only layered on top of the one shared brain. When a new version is released, all users move simultaneously to the same new starting point, with the same blind spots, wherever the model is used. Empirical research demonstrates these effects. Writers using generative AI produced more creative stories individually, but the collective diversity across the group decreased (11). Technical measures at the model level are insufficient to prevent the resulting monoculture (12).

The source of this asymmetry is structural rather than philosophical. Whether or not the model genuinely understands does not change this. The concentration of a single shared AI model, trained on a single set of texts, is an intrinsic characteristic of how the technology is built and applied. This is what makes 'one artificial brain for all' fundamentally different from all human brains. An individual human brain has its own biases and its own limitations, but it represents one perspective among many. A single large language model deployed across all contexts is a single perspective, replicated everywhere.

«An individual human brain has its own biases and its own limitations, but it represents one perspective among many. A single large language model deployed across all contexts is a single perspective, replicated everywhere»

What does this mean for clinical practice and research?

The distinction between the two types of bias provides a practical test. For many tasks, AI is faster, more thorough, and makes fewer mistakes than a single clinician. Replacing one doctor's idiosyncratic limitations with a broader, more systematic model is therefore an appropriate trade-off. The trade-off is reversed when the value lies in multiple clinicians arriving at different conclusions and each arguing their case: contested clinical reasoning, the framing of research questions, and the interpretation of ambiguous evidence.

The practical question for clinicians is therefore not only whether AI is good enough (its capabilities are developing rapidly), or whether it truly understands (a philosophical dead end), but whether its use strengthens our judgement or quietly replaces it. The answer varies case by case, between colleagues and between organisations, and depends on how AI is integrated into workflows and decision-making structures (13). Understanding this distinction can help us use AI where it strengthens clinical practice and research, and remain mindful of where it might weaken them.

«The practical question for clinicians is therefore not only whether AI is good enough, but whether its use strengthens our judgement or quietly replaces it»

When everyone consults the same model, the absence of dissent prevents the identification of errors that may otherwise have occurred. When uncertainty is genuinely novel, a single shared brain may produce convergence where many independent clinicians would have diverged.

The author would like to thank Susanne Albrechtsen for valuable comments on earlier versions of this article.

REFERENCES

1. Liu X, Liu H, Yang G et al. A generalist medical language model for disease diagnosis assistance. Nat Med 2025; 31: 932–42. [PubMed][CrossRef]

2. Dell'Acqua F, McFowland E, Mollick E et al. Navigating the jagged technological frontier: field experimental evidence of the effects of artificial intelligence on knowledge worker productivity and quality. Organ Sci 2026; 37: 403–23. [CrossRef]

3. Næss H. Kommunikasjon mellom mennesker er koordinerte gjetninger. Tidsskr Nor Legeforen 2025; 145. doi: 10.4045/tidsskr.25.0418. [PubMed][CrossRef]

4. Friston K. The free-energy principle: a unified brain theory? Nat Rev Neurosci 2010; 11: 127–38. [PubMed][CrossRef]

5. Schrimpf M, Blank IA, Tuckute G et al. The neural architecture of language: Integrative modeling converges on predictive processing. Proc Natl Acad Sci U S A 2021; 118. doi: 10.1073/pnas.2105646118. [PubMed][CrossRef]

6. Goldstein A, Zada Z, Buchnik E et al. Shared computational principles for language processing in humans and deep language models. Nat Neurosci 2022; 25: 369–80. [PubMed][CrossRef]

7. Caucheteux C, King JR. Brains and algorithms partially converge in natural language processing. Commun Biol 2022; 5: 134. [PubMed][CrossRef]

8. Searle JR. Minds, brains, and programs. Behav Brain Sci 1980; 3: 417–24. [CrossRef]

9. Dennett DC. The Intentional Stance. Cambridge, MA: MIT Press, 1987.

10. Chalmers DJ. Facing up to the problem of consciousness. J Conscious Stud 1995; 2: 200–19.

11. Doshi AR, Hauser OP. Generative AI enhances individual creativity but reduces the collective diversity of novel content. Sci Adv 2024; 10. doi: 10.1126/sciadv.adn5290. [PubMed][CrossRef]

12. Wu F, Black E, Chandrasekaran V. Generative monoculture in large language models. International Conference on Learning Representations (ICLR) 2025: 33068–107. https://proceedings.iclr.cc/paper_files/paper/2025/file/5178b2f2d7c44aa390c0777dc77b3f0c-Paper-Conference.pdf Accessed 20.3.2026.

13. Knudsen ES, Lang E, Lien LB et al. KI er klar, vi er ikke: hvorfor KI-suksess handler om strategi og ledelse. MAGMA 2026; 29: 126–30. [CrossRef]

Publisert: 3 July 2026. Tidsskr Nor Legeforen. DOI: 10.4045/tidsskr.26.0261