

Artificial narrative medicine

David Sundemo, Charlotte Blease

David Sundemo, MD, PhD, Sahlgrenska Academy at the University of Gothenburg, Närhälsan Lerum Primary Healthcare Center. E-mail: david.sundemo@outlook.com
Charlotte Blease, Ph.D. Department of Women's and Children's Health Uppsala University and Digital Psychiatry, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, USA. E-mail: charlotte.blease@uu.se

Recently, artificial intelligence (AI) has generated impressive results in the medical domain, particularly in terms of diagnostic expertise and empathic demeanor. Currently, studies are limited to text-based interaction, but considering the current performance and rapid development of AI-models, questions arise whether there are any limits when it comes to technology displacing medical tasks. According to the philosophical concept of functionalism, intelligence may be realizable using different physical materials. If the theory is valid, it could imply that all functions normally performed by doctors could potentially be replaced by AI. But if AI is listening to the patient, is anyone truly listening? This article explores what happens to the patient's story and the role of the doctor in the AI-age, and discusses the possible future of artificial narrative medicine.

What makes a sound a sound

What is a story? Partly it's the component parts, the captivating first paragraphs, the development of characters and a paradoxical ending or twist in the tale. It can also be the pace of the language, the intonation of the spoken word or the gestures of an animated storyteller. It can even be as simple as the characters building words, building sentences and chapters. Then again, the story may not be an isolated thing existing alone in the world. It can be dependent on the listener. To paraphrase the slightly worn out phrase - if a tree falls in a desolate forest and no one is around to hear it, does it really make a sound?

The 18th century bishop and philosopher Berkley may have been falsely attributed the phrase of the falling tree, but he developed the idea of perception being a vital part of the realness of an object (Berkeley, 1710) . So, what about the story? In narrative medicine, the story is the essence. It is the story that carries the weight, that creates the bond between patient and provider, that contains the keys to diagnosis and treatment of disease. But is a story dependent on a conscious perception to be a story, or is it a story by merit of its physical existence in a book or by residing within a human mind? The questions are abstract but with the advent of artificial intelligence (AI), deciphering this conundrum should be an imminent part of the future of narrative medicine. The intuition for many of us may be to disregard computer programs as a meaningless or insignificant part of a conversation, or in the telling of a story. We will try to challenge that idea. We will also try to give context to what might be technically possible in order for us to seriously ponder the future place of the patient's story within medicine, and what it might mean *to honor the stories of illness* (Charon, 2006) in the era of AI.

The current state of AI in medicine

So far, artificially intelligent systems have been very narrow in their usage; there are algorithms with superhuman ability to perform certain very specific tasks like playing chess, but an algorithm that plays chess cannot assess X-rays or ECGs - it can only perform within the domain it is trained for.

However, there are computer scientists and philosophers who believe that AI can become generally intelligent on the same level as humans. Some, like Sébastien Bubeck, a mathematics professor from Princeton, claim that sparks of artificial general intelligence have emerged as a result of one of the most recent AI-models, in this case, GPT-4 (Bubeck et al., 2023). This model is a type of AIs known as large language models (LLMs). The first LLM that became publicly well-known was Chat-GPT, which was released in November 2022. We won't delve into the function and structure of language models here; instead, we'll focus on how they perform in healthcare contexts - and it is sometimes

alarmingly good. First, let's look at a slide showing the performance progression of language models on the American medical licensing exam, USMLE (figure 1).

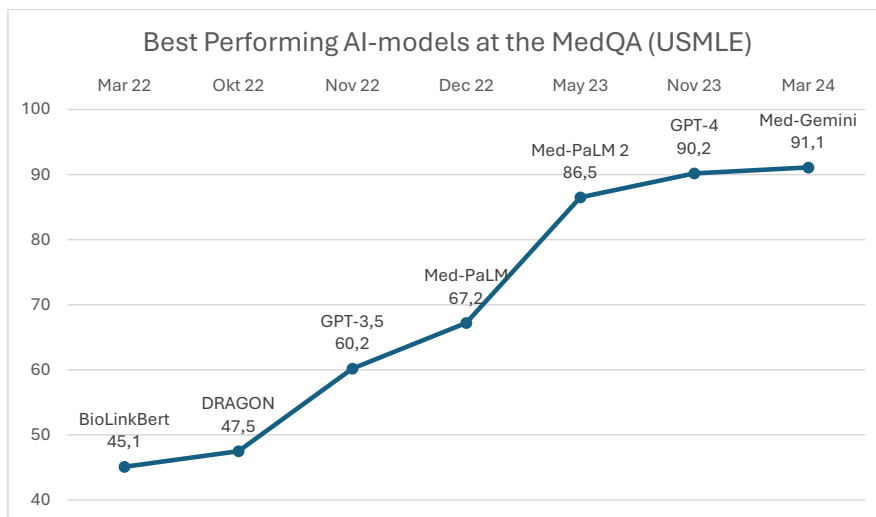


Figure 1. The diagram illustrates the progress of the best-performing AI-models from 2022 to 2024, using the benchmark of the American medical licensing exam, MedQA/USMLE. AI: artificiell intelligens

In March 2022, the best model got 45% correct answers on the exam. Two years later, the Med-Gemini model was launched with more than double the result - 91% (Saab et al., 2024). As a reference, around 60% correct answers are required for a passing grade, necessary for medical licensing. The medical exam - restricted though it may be - is no longer a problem for AI models. Can this AI handle real medical cases in the real world? And regardless, does it not require a human, a real doctor, to convey empathy and compassion?

As a response to this question, a study was published in April 2023 (Ayers et al., 2023). This was a study conducted by researchers in the USA with the aim to compare responses from real doctors with responses from ChatGPT, with regards to both medical quality and empathetic demeanor. The researchers collected questions from an anonymous online forum (Reddit/AskDocs). This specific section of the forum has verified doctors answering questions from patients and the public, openly on the web. In the study, the researchers collected both the patients' questions as well as the doctors' answers from the website. Subsequently, the researchers posed the same questions to ChatGPT, which generated its own answers. Three doctors were then asked to review the questions and the blinded responses from both ChatGPT and the doctors, and then rate, on a Likert scale (a 5-step scale), how they performed.

The results were commanding. The proportion of responses rated as good or very good was 3.6 times higher for the chatbot, and the proportion of responses rated as empathetic or very empathetic was 9.8 times higher for the chatbot than for the doctors.

The circumstances in the study were far from optimal, and quite different from natural clinical situations. It may also seem obvious that doctors express empathy and compassion more frequently in physical situations. But it is nonetheless thought-provoking when these systems not only have the ability to achieve accurate diagnostics and medical precision but can also simulate something we consider so deeply human, like empathy and compassion.

Philosophy, consciousness and true intelligence

Considering the current performance of AI-models in medicine, the question about the limits of automation arise. Can all of the functions of a healthcare worker be automated using AI? Functionalism is a theory of mind, suggesting that the question above may be a future possibility. The nascent phase of functionalism dates back to Aristotle and Thomas Hobbes, but the more modern form started to take shape in the late 60s (Putnam, 1975). In short, the theory states that it is the function of an organism, and the information processing occurring, that determines whether the organism can be intelligent or conscious. The theory is linked to multiple realizability, meaning that intelligence can emerge independent of the component parts of an organism. Biological carbon atoms, silicon chips or in fact any stable materia can provide the building blocks. If the flow of information - the function of an artificial system - is identical to the patterns of a human brain, intelligence would emerge. The theory of functionalism is a theory, and has been contested. For example, some argue that brain structures may not be as multiply realized as philosophers have argued. Nonetheless, and although far from a decisive argument in its favour, functionalism remains a dominant mind-theory according to a recent global survey of philosophers (Bourget and Chalmers, 2023).

Many philosophers, including those who are not advocates of functionalism, also consider intelligent and conscious AI-agents possible. But at present, are the current state-of-the art generative AI-models conscious or even intelligent? It is evident that (LLMs) operate in a fundamentally different way compared to the human brain and whether the models are truly intelligent is an open question. One contributor to the debate is Noam Chomsky, philosopher and professor of linguistics at MIT. In a recent article from the NY Times, he describes that LLMs differ significantly from the human way of reasoning and using language, and that these differences give them an inherent defect that cannot be fixed (Chomsky et al., 2023). He expresses it as follows:

The human mind is not, like ChatGPT and its ilk, a lumbering statistical engine for pattern matching, gorging on hundreds of terabytes of data and extrapolating the most likely conversational response or most probable answer to a scientific question. On the contrary, the human mind is a surprisingly efficient and even elegant system that operates with small amounts of information; it seeks not to infer brute correlations among data points but to create explanations.

According to Chomsky, language models are good at generating descriptions and predictions but they cannot say with certainty what could never happen. For instance, describing an apple falling and predicting its movement before it is dropped is something language models can do - description and prediction. But coming up with something like gravity governing the motion - that is an unlikely explanation and something language models would never be able to intuit. Being able to say what could never happen and categorically being able to say what is possible and impossible is an example of an explanation, according to Chomsky, a watermark of what he calls true intelligence. However, Chomsky does not say that AI systems can never become intelligent in the way he describes; he merely asserts that today's modern language models do not exhibit general intelligence. However, he concedes they are intelligent in a narrow sense.

Other scholars like Richard and Daniel Susskind argue that AI need not be functionally analogous to human cognition (functionalism) or mimic the same kinds of cognitive development as humans (as Chomsky suggests) in order to automate expert human tasks. They argue that commentators too often commit what they dub the "AI fallacy" - the presumption that AI systems must function anything like human experts to succeed or even surpass what we've currently got when it comes to performance on professional tasks.

Clearly, the contours of the AI debate are complicated with no received wisdom. AI could potentially be both intelligent and conscious. The models are efficient and well performing in terms of complicated medical tasks, but at present it is unclear whether they actually express real intelligence. We will now see how this may affect the role of the listener, the role of the doctor.

The role of the listener

As doctors, one learns to listen to the patient. The patient's story holds the key to diagnosis, and is the most important part of the medical investigation (Roshan and Rao, 2000). Undoubtedly, there are instrumental reasons for listening to the patient. However, if we ponder the possibility that the theory of functionalism is valid, AI could gather the story from the patient as efficiently as a doctor, and with as much nuance and empathy. Then through the AI-doctor the

correct diagnosis is reached and the patient gets the information presented in an seemingly natural and emphatic manner. But is something lost in the process? If AI is listening, is someone really listening?

We have concluded that the listening doctor has an instrumental role to fill, listening is an important part of the traditional process of reaching diagnosis and prescribing treatment. However, the instrumental role does not end with practical medical performance. Making the patient feel attended to, feel comforted and cherished are also instrumental tasks that could potentially be automated using technology. Thus, the practical medical and first order psychological tasks of the doctor could be fully automated using AI in this context. Still, it is unclear if someone or something is listening, and whether that matters beyond the functions we have stated so far. Proponents of narrative medicine may be prejudiced in believing that a humanistic worldview can only be performed by humans as opposed to the same acts, rendering the same results, if performed by artificial agents. Discerning what is pure bias from what is difference in value is important here, as to not become dogmatic.

If the patient has a pro-human disposition, favoring the interaction with a human doctor, the humanness in itself may provide instrumental value for the patient. Irrespective of medical quality, irrespective of empathic demeanor, the fact that the doctor is a human may for some people provide second order instrumental value. This will most likely not be the case for everyone, there are numerous reports of humans quitting normal human-to-human dating in favor of courting AI-avatars for romantic purposes (Singh-Kurtz, 2023). If AI is deemed enough to replace a romantic partner, it will for some people most probably also be sufficient for medical purposes.

However, for some of us, this may not be enough. Knowing we interact with *someone* will probably be essential to some. But does this someone have to be a human? Perhaps, but then again it may not be the essential humanness that we value but something more ethereal. Perhaps this more abstract, second order instrumental value that we attribute to the listener of the story is instead instantiated in the existence of a conscious mind. That it is the knowledge of being listened to by a conscious being that we value, that the species listening is not what is most important. There are already serious attempts by renowned philosophers and data scientists trying to build frameworks to evaluate the probability of emergence of consciousness in artificial systems (Butlin et al., 2023). According to these researchers, and authorities like Noam Chomsky and John Searle, present LLM-driven AI systems are far from reaching consciousness. However, for patients it might not even be the actual existence of consciousness that is of importance, but rather our perception of the existence of consciousness that matters. If we believe the agent is conscious, the aim may be met irrespective of the factual truth.

Artificial narrative medicine

Can narrative medicine survive the impending anticipated revolution of AI? In this piece, we have discussed the potential of AI in medicine, highlighting the possible validity of functionalism leading to artificial agents performing on par with us humans in each and every possible task. We evaluated the role of the listener, its instrumental functions and the psychological aspects of listening. The value that AI can and will provide in medicine is obvious. Many tasks today performed by clinicians will be replaced, in line with our discussion perhaps everything clinicians do could be replaced. This will probably create professional tension, technical problems and much turmoil but if we leave our biases aside, much of the changes can be beneficial to patients and the common good.

We should not forget that the importance of the narrative in a medical encounter varies. A patient with a sore throat, a sprained ankle or arachnophobia might of course carry a story of great significance, and sharing that story might be of value to both patient and listener. However, to be frank, these patients are probably more interested in getting their sore throat treated and their ankle and phobia rehabilitated. AI could without much loss automate these encounters. One should not generalize, but of course other maladies could with greater probability awaken thoughts and stories with existential and psychological weight. The experience of having a heart attack, of suffering from cancer or severe depression will, for many, sprout stories of existential proportions. How we treat these narratives says something about who we are. It might go against the premise, but perhaps artificial narrative medicine could be more selective, with active listeners honoring the story when it matters. Other times, the Hobbsian automata could do the job. Determining what patients perceive and want in these scenarios will be critical. Recall the study by Ayers on GPT outperforming doctors. It is unclear from this blinded study whether patients perceive the responses as empathic when they know AI has delivered it.

Narrative medicine emerged as a concept from dr Rita Charon and the university of Columbia. In narrative medicine, the listener becomes part of the story and helps to shape it. Comparatively, what if a tree falls in the woods, does it make a sound? Waves are pulsating through the medium of oxygen and nitrogen molecules but sounds do not emerge without an organism with a hearing apparatus. If current AI-systems interact with patients, they might produce medically valuable and empathically delivered messages, though it is improbable that consciousness, an equivalent of sound, emerges within the AI:s. It remains to be seen, when it comes to our illness stories, whether for some the perception of *feeling* listened to is paramount. Further, it could be a matter of perceiving a listener who is a conscious witness and clearly human that really counts. The narrative of AI in medicine is just beginning, and there's still a vast frontier waiting to be explored and understood.

References

- AYERS, J. W., POLIAK, A., DREDZE, M., LEAS, E. C., ZHU, Z., KELLEY, J. B., FAIX, D. J., GOODMAN, A. M., LONGHURST, C. A., HOGARTH, M. & SMITH, D. M. 2023. Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum. *JAMA Internal Medicine*, 183, 589-596.
- BERKELEY, G. 1710. *A treatise concerning the principles of human knowledge*, Indianapolis: Hackett Pub. Co., [1982] ©1982.
- BOURGET, D. & CHALMERS, D. J. 2023. Philosophers on Philosophy: The 2020 PhilPapers Survey. *Philosophers' Imprint*, 23.
- BUBECK, S., CHANDRASEKARAN, V., ELKAN, R., GEHRKE, J., HORVITZ, E., KAMAR, E., LEE, P., LEE, Y. T., LI, Y. & LUNDBERG, S. 2023. Sparks of artificial general intelligence: Early experiments with gpt-4. *arXiv preprint arXiv:2303.12712*.
- BUTLIN, P., LONG, R., ELMOZNINO, E., BENGIO, Y., BIRCH, J., CONSTANT, A., DEANE, G., FLEMING, S. M., FRITH, C. & JI, X. 2023. Consciousness in artificial intelligence: insights from the science of consciousness. *arXiv preprint arXiv:2308.08708*.
- CHARON, R. 2006. *Narrative Medicine: Honoring the Stories of Illness: Honoring the Stories of Illness*, Oxford University Press, USA.
- CHOMSKY, N., ROBERTS, I. & WATUMULL, J. 2023. The False Promise of ChatGPT. *The New York Times*, 2023-03-08.
- PUTNAM, H. 1975. The nature of mental states. In: PUTNAM, H. (ed.) *Philosophical Papers: Volume 2: Mind, Language and Reality*. Cambridge: Cambridge University Press.
- ROSHAN, M. & RAO, A. P. 2000. A Study on Relative Contributions of the History, Physical Examination and Investigations in Making Medical Diagnosis. *The Journal of the Association of Physicians of India*, 48, 771-5.
- SAAB, K., TU, T., WENG, W.-H., TANNO, R., STUTZ, D., WULCZYN, E., ZHANG, F., STROTHER, T., PARK, C. & VEDADI, E. 2024. Capabilities of gemini models in medicine. *arXiv preprint arXiv:2404.18416*.
- SINGH-KURTZ, S. 2023. The Man of Your Dreams. *New York Magazine*.