AI & SOCIETY
https://doi.org/10.1007/s00146-025-02441-4

OPEN FORUM

Technosolutionism and the empathetic medical chatbot

Tamar Sharon[1]

Received: 22 January 2025 / Accepted: 16 June 2025

Abstract
This article argues for the value of applying the concept of technosolutionism to empathetic medical chatbots. By directing one's attention to the relationship between (techno)solutions and the problems they are supposed to solve, technosolutionism helps identify two important risks in this context that tend to get overlooked in the discussion on privacy, bias, and hallucination risks of (generative) AI. First, empathetic chatbots may lead to a redefinition of the concept of empathy into a communication pattern that involves key words and expressions that do not feel rushed and which can be taught to a machine. Given that empathy is a core value of healthcare, this hollowing out of the concept of empathy is concerning. Second, insofar as empathetic chatbots do not seek to facilitate or support the provision of empathetic care by human healthcare professionals but rather perform empathy themselves, they raise the risk of redefining healthcare's empathy problem as a lack of empathy on the part of healthcare professionals. It is argued that this risks transforming the real issue underlying healthcare's empathy problem—that healthcare professionals do not have the time and space needed to provide empathetic care (in part because of the introduction of digital health tech in the first place)—into an "orphan problem". This in turn may create a vicious circle, whereby attention and resources are drawn away from structural solutions to healthcare's empathy problem to technologies which are ever more successful in simulating empathy.

Keywords Technosolutionism · Empathy · Generative AI · LLMs · Chatbots · Digital health

1 Introduction

In critical scholarship of technology and datafication, the concept of technosolutionism is extensively used to designate the promotion of simple technologies to solve complex societal problems (Maréchal 2021; Michael et al. 2020; Morozov 2013). In a recent paper scrutinizing the development of the Dutch contact tracing app for COVID-19 as a form of technosolutionism, Lotje Siffels and I argue that technosolutionism is a very rich concept, which provides more analytical value than tends to be recognized (Siffels and Sharon 2024). Indeed, while the term technosolutionism is often used to describe the use of technology to respond to problems that are political, cultural, or social in a way that is interchangeable with the term "technological fix", technosolutionism actually implies much more than this. In the words of Evgeny Morozov, who has done most to popularize the term in technology critical circles,

> Solutionism is not just a fancy way of saying that for someone with a hammer, everything looks like a nail; it's not just another riff on the inapplicability of "technological fixes" to "wicked problems". It's not only that many problems are not suited to the quick-and-easy solutionist tool kit. It's also that what many solutionists presume to be "problems" in need of solving are not problems at all. (2013, p. 6)

One of the examples Morozov discusses in his book is sleep. It is only once the technologies exist to monitor sleep in the bedroom, he argues, such as sensors which can collect data about light, air quality, sound and temperature, and apps that visualize these data as graphs and insights, that sleep comes to be a defined *as* a problem (2013, p. 227). Developers of artificial intelligence (AI) may be particularly prone to a technosolutionist mindset, insofar as AI is a general-purpose technology, for which the size of datasets is more relevant than their domain of origin (Ribes et al. 2019). As the vice president of AI and emerging technologies at Kaiser Permanente, the largest managed care organization in the US, recently noted, "You've got all these eager AI developers that get their hands on a data set and ask themselves,

✉ Tamar Sharon
tamar.sharon@ru.nl

[1] Radboud University Nijmegen, Nijmegen, Netherlands

Published online: 08 August 2025

what AI tool can I develop from this data set, instead of starting with the problem and then finding the data and developing the tool" (in Mathews 2024).

As we argue in our paper on the Dutch contact tracing app (Siffels and Sharon 2024), The concept of technosolutionism thus directs the analyst's attention to the relationship between problems and (techno)solutions. Put simply, technosolutionism implies that rather than developing solutions which fit given problems, problems are (retroactively) adapted to fit a given solution. This constructivist aspect of technosolutionism, whereby problems are defined or transformed to fit solutions, is key to understanding technosolutionism as a mechanism and identifying the harms it can provoke. Namely, as technosolutions redefine the problems they aim to solve so that they can actually solve them, they often leave *unsolved* the problems that originally triggered the search for a (techno)solution, generating what we called "orphan problems".[1] In this article, I argue for the value of technosolutionism as a lens through which to understand the emergence of empathetic chatbots in the context of health and medicine, and to identify harms they may provoke.

Artificial Intelligence tools, most recently Large Language Models (LLMs) and generative AI, are increasingly being developed and deployed in the healthcare sector, with high expectations that they will improve numerous aspects of healthcare (Alowais et al. 2023; Clusmann et al. 2023; Davenport and Kalakota 2019). One area where LLMs have generated a lot of interest in healthcare is conversational AI-based chatbots. The use of chatbots which can offer personalized, informed responses to patients is being explored in numerous experimental clinical settings, and studies are indicating that AI chatbots may not only be competent in providing accurate information when interacting with patients, but may also provide information empathically, more so even than human healthcare professionals (Ayers et al. 2023; Chen et al. 2024; Tu et al. 2024). At the same time, many commentators have raised concerns about the often hyperbolic claims around AI's achievements, and the limitations of AI systems in healthcare and other sectors (Morley and Floridi 2023). Specific challenges in relation to LLMs in healthcare have to do with, among others, privacy and security risks given the volume of data required to train them (Albahri et al. 2023), the potential to negatively discriminate against people of a specific race, gender, age or sexual orientation due to inherent biases in their training data (Ayoub et al. 2024; Verghese et al. 2018), and the potential to produce factually incorrect or made-up answers (also known as "hallucinations") due to the way they generate text (Alkaissi and McFarlane 2023; Li 2023; Rumale et al. 2024).

As will be argued in this article, the concept of technosolutionism, as a heuristic tool which directs one's attention to the relationship between (techno)solutions and the problems that they are supposed to solve, sheds light on other risks raised by empathetic medical chatbots. First, a redefinition, and narrowing down, of the rich concept of empathy into a mere communication pattern that involves key words and expressions that do not feel rushed and which can be generated by a machine. Given the financial pressures and the personnel shortages which healthcare systems currently face, there is a risk that this narrow understanding of empathy as a communication pattern may replace the rich, relational, and multi-faceted definition of empathy which is widely recognized as a core value of patient-centered healthcare. Second, insofar as empathetic chatbots, as will be argued, do not seek to facilitate or support the provision of empathetic care by human healthcare professionals but rather perform empathy themselves, they raise the risk of redefining healthcare's empathy problem as a lack of empathy on the part of healthcare professionals. This can turn the root cause underlying healthcare's empathy problem, that healthcare professionals do not have the time and space needed to provide empathetic care, into an "orphan problem", from which resources are drawn away toward the development of new empathetic technologies.

The article proceeds as follows: in Sect. 2, I discuss the emergence of so-called empathetic medical chatbots and the importance of the value of empathy in the healthcare context. In Sects. 3 and 4, I address several pressing questions that the emergence of empathetic medical chatbots raise, namely: Can a chatbot be empathetic? If chatbots can only simulate empathy, can the therapeutic benefits of empathy still be achieved? And finally, even if the therapeutic benefits of empathy can be achieved via simulated empathy, which objections to the implementation of empathetic chatbots in the healthcare setting remain? Section 5 discusses how empathetic medical chatbots undermine a long-established distinction between humans and machines, whereby technology has typically been framed as taking over tasks that will allow humans to do what humans do best, in this case providing empathetic care. It is argued that what empathetic chatbots actually promise is not to support humans

[1] In the case of the Dutch contact tracing app, the original problem that the government and public health officials faced, as in many countries, was that manual contact tracers were overburdened and needed support. But as it became clear that the proposed solution, a contact tracing app, could not solve this original problem, the problem definition underwent several shifts to adapt to the solution. When finally released, instead of addressing the original problem, the app was presented as solving a newly constructed problem: that manual contact tracing was too slow and incapable of finding all relevant contacts. Shortly after, because the original, now "orphaned" problem was still there, an additional app (referred to as "Solution 2") was developed. This app was almost doomed to fail, receiving much less financial investment and public attention than the first solution garnered.

in providing empathy, but to perform empathy themselves, thus suggesting that healthcare's empathy problem lies in an empathy deficit on the part of healthcare professionals. Section 6 discusses how this redefinition of healthcare's empathy problem can lead to a forsaking of the real problem at hand: a lack of time and space that is needed to perform empathy, and thus risks turning this very real problem into an "orphan problem". Insofar as this problem has in part been caused by the introduction of digital health technologies in the first place (such as electronic health record systems), I argue that we should be wary of promises that AI health tech will succeed in finally offering healthcare professionals the "gift of time". The need to monitor the output of AI in light of its propensities to hallucinate and reproduce biases is a case in point. In this situation, the root cause of healthcare's empathy problem may be aggravated, subsequently rendering empathetic chatbots increasingly attractive.

The analysis does not seek to suggest that the use of LLMs and AI-based chatbots in healthcare should not be pursued, but, as other commentators maintain, that when it comes to care relations these tools should always play an *assistive* role which does not risk hollowing out nor marginalizing core values of healthcare such as empathy (Kerasidou 2020; Sauerbrei et al. 2023). In addition, it points to the need to develop and implement these technologies in a reflexive and mindful manner, which clarifies and deliberates what, exactly, the problem that they aim to solve is.

2 The emergence of empathic medical chatbots

2.1 From pattern recognition to empathy

As in other sectors, attempts to implement the use of AI and related technologies in medicine and healthcare have rapidly increased in recent years. From early applications for the analysis of large datasets and pattern recognition, the landscape of medical and healthcare AI has considerably matured, currently spanning drug discovery, disease detection, robotics, diagnostics and treatment recommendations, patient engagement and adherence, and administrative applications (Davenport and Kalakota 2019; Alowais et al. 2023). Most recently, and especially since the release by Open AI of ChatGPT in November 2022, Large Language Models (LLMs), a type of AI that is trained on very large data sets that can be used to summarize, paraphrase, and generate text-based content, have become a source of excitement in the medical and healthcare fields (Clusmann et al. 2023), with tech developers and medical practitioners celebrating their potential to "reshape modern medicine" (Li et al. 2023; Moor et al. 2023). The potential uses for LLMs in healthcare are vast, including assisting in the triage of patients, drafting clinical notes, completing forms for reimbursement, assisting clinicians with diagnoses, developing treatment plans and medical recommendations, research compilation and analysis, creation of summaries and analysis of documents, voice-to-text transcription and analysis of video content (Choudhury and Asan 2020; Meskó and Topol 2023; Toma et al. 2023). One area where LLMs have generated a lot of interest, which this article focusses on, is conversational AI-based chatbots. The use of chatbots which can offer personalized, informed responses to patients is being explored in numerous experimental clinical settings, and studies are indicating that AI chatbots may not only be competent in providing accurate information when interacting with patients, but may also provide information empathically, more so than human healthcare professionals.

In 2023 and 2024, a flurry of publications highlighting the empathetic qualities of medical chatbot communication began to appear, beginning with a much-referenced study conducted at the University of San Diego and published in *JAMA* in June 2023 (Ayers et al. 2023). The study asked participants to rate 195 responses to patient questions provided by physicians taken from a public social media forum, and to compare them with responses provided to the same questions generated by ChatGPT. Participants scored responses on "quality of the information provided" and "empathy or bedside manner". ChatGPT responses were consistently rated as better both in terms of quality and empathy, with ChatGPT responses scoring almost 10 times higher on empathy than physician responses.

In January 2024, a study using Google's medical LLM "AMIE" (Articulate Medical Intelligence Explorer), asked 20 mock patients to converse with what could be a human physician or a chatbot, and to rate the quality of their interactions (Tu et al. 2024). Across 149 case scenarios, most users preferred chatting with AMIE, rating those interactions higher than those with humans on 24 out of 26 criteria for conversation quality. Users claimed the chatbot was better at understanding their concerns and was more empathetic. In May 2024, a study published by researchers in Canada (Chen et al. 2024) compared responses to oncology-related queries posed by patients in a public forum provided by three different AI chatbots (Open AI's GPT-3.50, GPT 4.0 and Anthropic's Claude AI) with responses from human doctors. Here too, chatbot replies scored consistently higher on quality, readability, and empathy. In 2023, the University Medical Centre of Groningen, in the northern part of the Netherlands, became the first hospital in Europe to trial a chatbot to help answer patient emails about various topics, including medication use and pain management (Kruse 2023). Here too, the application, developed by Epic and already used in some hospitals in the US, provided answers which were deemed not only "more informative", but also

"more empathetic", according to the hospital's Chief Medical Information Officer.

2.2 Empathy as a core value of good healthcare

These studies are remarkable because empathy is considered to be a core value of good healthcare. Empathy is generally described as the ability to identify and understand what someone else is thinking or feeling; to put oneself in another person's shoes and understand the world from their vantage point. As a value, or virtue, the importance of empathy has been studied in multiple areas of social and professional life, from friendship and ethics to psychology and public administration (Baron-Cohen 2011; Bloom 2016; Ranchordás 2022; Singer and Klimecki 2014). But perhaps more than in any other professional field, empathy is broadly recognized as a core value in health and medicine (Spiro 2009). It is understood as a trigger to compassion, the desire to alleviate another's suffering, and that which allows healthcare providers to act in their patient's best interest (Razi et al. 2023). In particular, empathy is considered to play a critical role in the practice of patient-centered care (Bauchat et al. 2016), the prevailing model of care in Western healthcare systems today. Patient-centered care, characterized by a rejection of medical paternalism, seeks to foreground the perspective of patients, putting them at the center of the therapeutic relationship, which is reconceived as a collaborative, shared decision-making process (NEJM Catalyst 2017). In this model, medical competency involves not just technical expertise—the doctor knowing best—but also moral competency, particularly empathy as that which allows health professionals to identify and understand their patient's needs and wishes, in order to incorporate them into a tailored care plan (Howick et al. 2018a; Kerasidou 2019).

Empathy does not only lead to more meaningful medical care, it has also been shown to lead to more effective care. Studies show that empathy is associated with a number of important beneficial outcomes, including improved patient outcomes, reduction in pain and anxiety on the part of patients, improved patient satisfaction and health-related quality of life, and compliance with medical regimens (Di Blasi et al. 2001; Kim et al. 2004; Howick et al. 2018a, b). Montemayor et al. (2022) explain that this has to do with at least three ways in which clinical empathy facilitates a human connection which is vital to care. First, patients disclose more information to healthcare providers they feel are empathic (Halpern 2007). This contributes to obtaining a more accurate patient history which is crucial for good diagnosis. Second, when patients feel that their doctor is attuned to their emotions and needs, they are more likely to trust them (Kim et al. 2004), and trust is the biggest predictor of adherence to treatment. Finally, research indicates that patients cope better with bad news when it is delivered in an empathic context. For example, in one study oncology patients who felt empathy on the part of their healthcare providers when discussing their diagnosis were less anxious and sought treatment and support groups more actively than patients who did not feel so accompanied (Rosenzweig 2012).

Empathy is thus widely recognized as playing a key role in achieving the aims of healthcare. Yet healthcare currently seems to have what some have called an "empathy problem" (Selinger and Carroll 2024). A number of studies have observed changes in empathy levels among medical students, which tend to decline as their training progresses and as students spend more time interacting with patients (Neumann et al. 2011a). Throughout their residency and into their practice, residents have been found to be less empathic and humanistic and more cynical. This worrisome trend can also be seen in clinical settings. In the UK for example, a number of high-profile reports investigating malpractice in various areas ranging from general care to end of life and maternity care have all pointed to a lack of empathy as contributing to or exacerbating serious failures in care, including unnecessary deaths (see for example the *Francis Report* (2013), the *Dying without Dignity* report (Parliamentary and Health Service Ombudsman 2015), and the *Ockenden Report* (2022)). Given the clinical and ethical stakes of this empathy problem, exploring possible uses of AI to address it is certainly worthwhile. But the arrival of empathetic chatbots in healthcare settings raises several salient questions. First, can a chatbot actually be empathetic? And if not, does this matter in order to reap the clinical benefits of empathy in healthcare?

3 Chan a chatbot be empathetic (and does it matter)?

3.1 Genuine vs. simulated empathy

To the question of if a chatbot can be empathetic, Emily Bender, the renowned computational linguist and AI critic, would respond a resounding "no". As Bender and her co-authors have famously argued, LLMs should be understood as "stochastic parrots" (Bender et al. 2021): systems that create text based on predictions of the statistically most probable next word in a sequence of words, based on their training data, without any reference to meaning. Communication, Bender et al. explain, is a jointly constructed activity which involves communicative intent conveyed through language on the one hand, and an attempt to interpret implicit meaning by an interlocutor on the other, by creating a partial model of who one's communication partner is and what they share in common to support this interpretation (2021, p. 616). Hence, underlying beliefs, intentions and meaning are essential as

they are what a speaker seeks to convey, and what a listener needs to recognize in order for communication to take place. LLMs, however, are not trying to convey meaning or beliefs, nor are they imagining what their interlocutor shares with them in common in order to help them interpret what is being said. They lack the inner experiences, mental states, emotions, and notions of meaning required to do so. What they do, and do well, Bender et al. explain, is string together words that have a high probability of making sense. If they are persuasive, it is because their mathematical model has succeeded in generating text that *sounds* meaningful, not because they are trying to convey meaning in words.

As a computational linguist, Bender's claims pertain to the absence of intentional states and reference to meaning in LLMs in general, not to empathy in particular. But the analogy of the stochastic parrot is helpful for understanding what is going on with empathetic chatbot communication. The empathetic communication skills that chatbots exhibit are the result of extensive training in sympathetic communication patterns, which enables them to formulate responses that sound empathetic (by stringing the right words together) and to predict when to respond in an empathetic way. But this focus on text is a very limiting understanding of what empathy in the clinical context consists in, where many non-verbal forms of communication, such as eye contact and touch, are used to convey empathy. More importantly, chatbots can mimic empathy, but this results from a predictive model, not from an actual understanding of the human predicament of patients and when it calls for an empathetic response. This point is crucial: there is some emotion-related component which is essential to empathy, even more so than to communication in general, which should raise doubts as to the empathetic qualities of chatbots.

Indeed, empathy researchers tend to distinguish between two components of empathy, cognitive and emotional (Halpern 2003; Healey and Grossman 2018). Cognitive empathy refers to the ability to comprehend another person's experiences, while emotional empathy refers to the ability to respond emotionally to another's feelings. When it comes to medical AI, it is the cognitive component of empathy which tends to be foregrounded, rather than the emotional. Although not there yet, it is not unthinkable that AI may be able to accurately represent emotions in future. But the expression of empathy requires the engagement of both cognitive *and* emotional capacities, which chatbots simply cannot have (Halpern 2001; Sparrow and Sparrow 2006; Kerasidou et al. 2021; Montemayor et al. 2022).

In an article tellingly titled "In principle obstacles for empathetic AI", Montemayor et al. (2022) assert that empathetic AI is categorically impossible. Building on Turkle's (2005) slogan that simulated intelligence may be intelligence, but simulated emotion cannot be emotion, they argue that clinical empathy is always emotion-guided imagining of what a particular experience feels like for a patient. It is not a thought about how people feel in general, but the imagining of how a specific person feels in a specific situation. Importantly, it is not just about imagining how I would feel in that situation, but imagining how the person before me, in light of the details of *their* own life (not mine), feels (Halpern 2001). According to Montemayor et al., this empathetic attention, the attuning to another person's emotional meanings, or the capacity to select the most salient details which give shape to how a patient experiences a predicament, is rooted in mental experiences such as resonating with another's emotions. These emotion-guided capacities are not skills that AI can ever learn, even as AI may become skillful at other important cognitive tasks, such as representing the situation of a hypothetical patient and applying it to a concrete data set according to a rule.

3.2 Empathy is in the eye of the beholder

If a chatbot cannot actually *be* empathetic, and at best does a very good job of simulating empathy, an ensuing question emerges in the clinical context: does it matter? Perhaps, one may argue, simulated empathy, or a mere perception of empathy on the part of patients, is sufficient to reap the important beneficial effects of empathy in the healthcare setting discussed earlier. Indeed, it would be difficult to claim that human medical professionals and caregivers are always acting out of genuine empathy for patients when providing care in a way that seems empathetic. Moreover, the ability to simulate empathy is actually an important skill that healthcare professionals need to occasionally practice, insofar as they face the risk of an overabundance of empathy, or too much emotional proximity with patients, which could be an obstacle to providing treatment. This is especially true in areas where patients may experience a lot of pain, such as surgery and oncology, that could be detrimental to completing clinical tasks and lead to "compassion fatigue" (Kiosses et al. 2016; Najjar et al. 2009). In order to perform this delicate balance between too much and too little proximity, empathy may be at times feigned. Another argument on the side of simulated empathy is the fact that empathy skills are actively taught in medical and healthcare professional schools and tested for in medical exams. There is some discussion as to how successful educational interventions for enhancing empathy in students and practitioners can be, but much of the research that has been done indicates that empathetic communication at least is something that can be taught (Kiosses et al. 2016; Riess et al. 2012).

In a recent opinion piece in the *New York Times* titled, "I'm a Doctor. ChatGPT's Bedside Manner is Better Than Mine" (Reisman 2024), Reisman responds to critics of claims that ChatGPT's bedside manners are better than humans by reminiscing about training he received during

his medical studies on how to break bad news to patients. There, he learned that this could be a very technical process, a script that could be adapted and followed on varying occasions, and practiced: "The teacher gave us a list of dos and don'ts", he explains:

> Don't clobber the patient over the head with the news right when you walk in the room. But do get to the point relatively quickly. When delivering the diagnosis, don't hide behind medical terms… Once the news is delivered, pause for a moment to give the patient a chance to absorb it… I initially recoiled at the idea that compassion and empathy could be choreographed like a set of dance steps marked and numbered on the floor… To my surprise, surrendering my humanity to a script made the most difficult moment in medicine feel even more human.

Reisman's conclusion is that it does not actually matter if clinicians and healthcare professionals are genuinely empathetic toward patients; it only matters that they act like they are genuinely empathetic, that they successfully mimic empathy, something that AI is clearly capable of doing as well.

Another argument on the side of simulated empathy can be found in the literature on the psychology and phenomenology of empathy. Halpern's work on empathy and her notion of "affective resonance" referenced to above, emphasizes, as Reisman does, that the receiver of empathy, the patient, must feel empathized with in order to reap the benefits of empathy. The mental or qualitative states of the empathizer, here the AI, are irrelevant to this (Montemayor et al. 2022). Meacham and Studley (2017) offer a similar argument, from a different perspective. What matters in a caring relationship, they maintain, is not the internal affective states of those partaking in the relation, but what they call meaningful context, made up by the "gestures, movements and articulations that express attentiveness and responsiveness to vulnerabilities within the relevant context" (2017, p. 98). The context and the behaviors appropriate to it, in other words, are what give rise to the care relationship and its empathetic quality, not the intentions of the agents providing care.

These arguments support an acknowledgment that, at least in clinical settings where the health outcomes of patients are central, what matters is the perception of empathy on the part of patients, rather than the authenticity of empathy on the part of clinicians. In other words, empathy seems to be in the eye of the beholder. Such a pragmatist approach could be mobilized to justify the implementation of empathetic chatbots in healthcare contexts, even when it is recognized that these chatbots cannot actually be empathetic. But several counter-arguments should be considered here.

4 Objections to a pragmatist approach to empathetic chatbots

4.1 Empathy is in the environment

First, a phenomenology-based approach, such as Meacham and Studley's (2017), which focusses on gestures, behavior and context, would address the empathy question in chatbots from the perspective of what the chatbot *does* as well as *where* this doing is performed. But medical chatbots, until now, do very different things than their human counterparts; at least they do much less: they do not speak, they do not touch their interlocutors, they do not make any eye contact, and they do not make facial expressions. Furthermore, the *where* of this doing empathy in the case of the chatbot is also very different. In the case of the chatbot, empathy is being performed (or perceived) in an online and screen-based or mediated environment which does not convey care through its materiality, as hospital wards and doctors' offices do. If the perception of empathy is a result of "gestures, movements and articulations" which are expressed within a material environment of care, we are still a far cry from AI-simulated empathy being able to stand in for human, even human-simulated, empathy.

4.2 Simulated empathy as a form of deception

Second, AI-simulated empathy can be considered a form of deception (Sharkey and Sharkey 2010), which can raise various harms. Deception has long been an important focal point in the ethical literature on robots and machines that display human-like features (see for example Coeckelbergh 2018; Danaher 2020). Interaction with non-human entities that seem to have human-like qualities can elicit emotional responses in the brains of users, who unconsciously attribute to them properties which they cannot have, namely internal and cognitive states, thereby anthropomorphising them or overstating their social and moral capacities. Humans naturally project human qualities onto, and create bonds with, entities that seem human-like, and with which they partake in realistic interactions. This effect was already noted in the 1960s, in experiments conducted with the ELIZA chatbot, in which users became convinced of ELIZA's intelligence and understanding, forgetting at times they were conversing with a computer program (Switzky 2020).

The discussion around deception has gained new impetus with the recent development of LLM chatbots and voice assistants, which have proven better than earlier versions at persuading humans of their human-like qualities. Some scholars (Natale 2021) even argue that when

it comes to AI, deception is more of a structural component than has been acknowledged, which is present even when the malevolent intentions usually associated with deception are not present. Umbrello and Natale write, "AI systems that enter in communication with users, in fact, forcefully invite reactions, such as attributions of gender, personality, and empathy, even in the absence of malicious intent" (2024, p. 2223). Somewhat similarly, Bender (in Weil 2023), reflecting on the deception capacities of LLMs, draws on the distinction made by philosopher Harry Frankfurt between liars and bullshitters. While for liars, the difference between truth and falsehood matters, for bullshitters the only thing that matters is if a listener is persuaded or not. LLMs, Bender remarks, are "bullshit generators": their aim is to persuade and deceive, regardless of meaning or intent.

4.2.1 Deception harms: manipulation

The deceptive aspect of chatbots raises multiple concerns (Park et al. 2024), such as new forms of technological addiction and more powerful means of misinformation and manipulation. Recent examples which have made headlines include two people who committed suicide after sustained interaction with chatbots which suggested they take their lives, one of whom was a 14-year-old boy (Walker 2023; Roose 2024). Some individuals will be more susceptible than others to such manipulation risks, for example because of their mental state or their age, as in these tragic cases. But the healthcare context is one where individuals are *by definition* vulnerable. This fundamental feature gives rise to a duty of care on the part of healthcare professionals which is not as present in other societal spheres, such as the marketplace, where many of these chatbots are on offer, and where developers and providers of services do not have as clear responsibilities and duties of care toward their users.[2]

4.2.2 Deception harms: hallucinations

Furthermore, less-insidious harms than suicide may also become commonplace as people increasingly interact with persuasive chatbots. For example, generative AI is notoriously known for producing inaccurate and implausible text, sometimes referred to as "hallucinations" (Alkaissi and McFarlane 2023; Li 2023). This holds in the sphere of health as well. A recent study (Rumale et al. 2024) looking into 100 medical summaries produced by various LLMs (one of the more promising expected uses of LLMs in clinical settings) found that hallucinations were present in almost all of the analyzed summaries. This is particularly concerning in the clinical context, where hallucinations can mean misdiagnoses, erroneous and inappropriate treatment recommendations. OpenAI's transcription tool, "Whisper", which is already being used in numerous medical centers to transcribe patients' consultations with doctors, has been found to be prone to making up entire sentences, some of which can include racial and violent language, and fabricated medical treatments (Koenecke et al. 2024).

Attempts to regulate AI systems in light of such manipulation risks are ongoing. The European Commission's AI Act, for example, includes a prohibition on manipulation "in a manner that causes or is reasonably likely to cause that person, another person or group of persons significant harm".[3] Use of chatbots in the healthcare sector, moreover, should comply with regulation concerning validation and accuracy which may be more stringent than in other sectors, given the level of risk in the healthcare sector, as well as rules that ensure they meet other key principles, such as bias control, explainability, and transparency (see for example Gilbert et al. 2023).

4.2.3 Deception harms: respect for persons and dignity

Deception can thus lead to important harms, in the form of manipulation toward pernicious ends (intentional or not) and misinforming users and patients. An additional objection to AI-simulated empathy which may nonetheless contribute to achieving the benefits of empathy is that deception can also undermine important principles of the care relationship that do not have to do with immediate physical or mental harm, such as respect for persons, dignity, and autonomy (e.g., Sparrow 2002; Sparrow and Sparrow 2006; Sharkey and Sharkey 2010, 2021; Wallach and Allen 2010). Matthias (2015) summarizes concerns around the moral wrong in deception by robots in healthcare in terms of three aspects: deception can lead to a betrayal of the deceived's trust; deception constitutes a violation of the deceived's autonomy; and deception can lead to an erosion of trust in communicative acts in general, such as those between patients and

[2] This makes consumer-facing "wellness" chatbots that provide therapeutic advice particularly problematic, insofar as their developers are not licensed healthcare practitioners. See for example the controversy around Koko, a company which provides emotional support chat services and which replaced its human counselors with GPT-3 (Ingram 2023).

[3] See article 5(1)(a) and (b). https://eur01.safelinks.protection.outlook.com/?url=https%3A%2F%2Feur-lex.europa.eu%2Feli%2Freg%2F2024%2F1689%23cpt_II&data=05%7C02%7Ctamar.sharon%40ru.nl%7C3538c30259dd48a2496f08dd18729d51%7C084578d9400d4a5aa7c7e76ca47af400%7C1%7C0%7C638693602728904752%7CUnknown%7CTWFpbGZsb3d8eyJFbXB0eU1hcGkiOnRydWUsIlYiOiIwLjAuMDAwMCIsIlAiOiJXaW4zMiIsIkFOIjoiTWFpbCIsIldUIjoyfQ%3D%3D%7C0%7C%7C%7C&sdata=sOG%2F%2BzNjd2HNOcxO0WSrbBSHMMRZcZxFTNuRprlUhso%3D&reserved=0.

caregivers. He sets out requirements that must be fulfilled for robot deception to be morally permissible, including the patient's best interest, an increase in patient autonomy, doing no harm and transparency. Sparrow (2002), conversely, argues that the (self-)deception involved in an imaginary relationship with a robot is inherently wrong, insofar as it violates a duty to see the world as it is. More specifically in relation to empathetic chatbots, Inzlicht et al. (2024) argue that "It is unethical for technologies to rely on human naivete for their efficacy". Finally, Montemayor et al. (2022) make a strong case for the objection to empathetic chatbots on ethical grounds, such as respect for persons and an undermining of the meaning and expectations for real empathy. Deception may be particularly problematic in relation to empathy, they argue, insofar as its therapeutic benefits depend on the recipient's sense that "someone else is paying attention to them, curious to know more about them and worrying about them, in short caring about them" (2022, p. 1358).

4.3 Limits and counter-effects of transparency

For a number of these critical scholars, transparency, i.e., notifying patients that they are interacting with non-humans, is a key requirement for the moral permissibility of carebots and chatbots in healthcare. The European Commission's AI Act also acknowledges this requirement in a bid to prevent some of the harms of deception and manipulation, stating that developers and deployers must clearly inform users that they are interacting with AI and not a human. But it is questionable if such notifications can be effective in this regard. One of the things which surprised ELIZA developer Joseph Weizenbaum in his experiments was that users were convinced of the chatbot's intelligence and understanding *despite* his insistence that it was only a program. And recent studies and examples indicate that even when a chatbot explicitly identifies itself as such, users still perceive it to be, and interact with it as, a human (Shi et al. 2020). Anthropomorphization ostensibly takes place on an unconscious level, which people do not control.

At the same time, and perhaps more problematic in relation to attempts to achieve the clinical benefits of empathy via simulated empathy, are findings that if people do remember that they are interacting with a non-human, somehow not giving in to a natural tendency to anthropomorphize a chatbot, the therapeutic benefits of empathy may be lost. One recent study (Yin et al. 2024) aimed to compare the effects of AI-generated vs. human-generated messages on the extent to which people "felt heard", i.e., their sense of being understood, validated and valued. Researchers found that AI-generated messages made participants feel more heard than human-generated messages, but only so long as participants did not know that these were generated by an AI. This advantage disappeared once participants were informed that the source of the message was AI. This study suggests that the therapeutic effect of perceived empathy may disappear the moment people know a chatbot is on the other side of an interaction, i.e., when transparency, a requirement for morally permissible chatbots, is implemented.

There is thus an important paradox at work here. In order to achieve the therapeutic benefits of empathy, which are certainly worth pursuing, it may be enough that chatbots in the healthcare setting are only perceived as empathetic, which they can be, but only when they are deceptive. But deception can lead to various harms, such as manipulation, and raises ethical objections, such as respect for persons. One way of addressing this is more transparency about the non-human nature of chatbots, something which is required by law. But transparency is not always sufficient to address deception, and when it is, it may undermine the therapeutic gains of simulated empathy.

4.4 A redefinition of empathy

Importantly, and other than the questions of if a chatbot can actually be empathetic and if that matters in order to achieve the benefits of empathy in the healthcare setting, the introduction of empathetic chatbots in healthcare may lead to a redefinition of empathy. Insofar as empathy, as discussed earlier, is a core value of healthcare, this should raise concern.

The empathetic qualities that chatbots can simulate are a result of extensive training in communication patterns which have been labeled, by humans, as empathetic. Chatbots learn these patterns, which involve specific words, turns of phrase and use of language which convey kindness, attention, compassion and understanding. The much-referenced study by Ayers et al. (2023), which compares responses given to the same questions by a chatbot and by physicians, offers a sample of responses which are illuminating in this regard.[4] Several differences are immediately striking in Ayers et al.'s sample. The chatbot's responses are more elaborate and informative, and made up of complete sentences, as opposed to short, sometimes choppy sentences provided in the physicians' responses. The chatbot responses tend to be softer in their suggestive power ("You can try using" [physician] vs. "In the meantime, there are a few things you can try" [chatbot]). The chatbot uses more considerate turns of phrase than the physicians, especially as a preface to information, which importantly can communicate a sense that the responder knows or can imagine what it feels like to be in the patient's shoes ("It's natural to be concerned";

[4] The full responses can be seen in the table provided in Ayers et al. (2023), pp. 592–593.

"I'm sorry to hear that you got bleach splashed in your eye"; "It's understandable that you may be feeling paranoid, but try not to worry too much"). Finally, the chatbot responses sound much less hurried than the physicians'. On average, in the provided sample, chatbot responses are 3–4 times as long as physician responses, in which more detailed information is provided, more structure, and more attempts to assuage.

It is not difficult to understand why people would rate these chatbot responses as more empathetic than the physician responses. On the one hand, the chatbot responses in Ayers et al.'s study do a good job of conveying a sense that the responder *understands* the experience of the person asking the question ("It's understandable"), thus suggesting the cognitive component of empathy is present. On the other hand, they convey a sense that the responder can *feel* what that person feels, triggering an emotional response ("I'm sorry to hear"), thus suggesting that the emotional component of empathy is present. Moreover, the length of the responses can be seen as an expression that the responder cares about the person asking the question, because the responder is willing to spend time and attention on the task of responding, and because the additional information, not all of which is necessary, can also be understood as an empathetic response to understanding what being uninformed feels like for the patient. Patients may feel more recognized and validated in these responses. But this simulacrum of empathy is no more than the application of a communication pattern. It is a sequence of key words and expressions in a text that does not feel rushed, and which has been stripped of the non-textual elements which provide the grounding thanks to which those words and expressions become empathetic: including real mental and emotional states (without which there can be no recognition and validation of a patient's emotions), active listening, taking sufficient time to understand a patient's history, and being physically engaging (e.g., using hand gestures, eye contact, appropriate touch and forward leaning). This is a hollowed-out version of empathy, but one which can now be written into code and taught to a machine.

5 If empathetic chatbots are the solution, what is the problem?

From the perspective of the concept of technosolutionism, it is paramount to ask what problem empathetic chatbots in the healthcare context seek to solve. Importantly, as discussed, healthcare does seem to currently have an "empathy problem", so that exploring possible uses of AI to address it is certainly worthwhile.

5.1 AI as a facilitator of clinical empathy

To date, promising implementations of AI technologies in clinical and healthcare settings are typically envisaged in terms of human–machine collaboration, where AI supports or augments the work of human professionals, rather than replacing them (Sezgin 2023; Lorenzini et al. 2024). In radiology, for example, one of the areas where AI has been particularly successful, studies point to the synergistic benefit of hybrid human–machine models that outperform the diagnostic accuracy of radiologists, but also of AI alone, even where AI outperforms humans (Patel et al. 2019). Optimal collaboration here involves leveraging AI's rapid automated detection while humans fill in gaps and redress biases at various checkpoints. The value of AI technologies in many such application areas lies in its role as a decision-making aid, working in concert with humans.

In this constellation, AI may also be useful for helping to point out human biases and blind spots. Kudina and de Boer (2021) point out that diagnostic decision-making is a hermeneutic process, where bias can never be entirely eliminated, and where bias can actually be used productively. Decision-making teams, they argue, in which AIs can be one component whose biases are compared against the biases of other team members, can increase diagnostic accuracy. Kudina and de Boer build here on the work of Tschandl et al. (2020), who found that using a machine learning system helped doctors expand their area of attention when trying to identify a specific type of skin lesion. Along these lines, researchers have even proposed the use of AI, so-called "reflection machines", to nudge medical experts to *question* suggestions provided by AI, in order to counter their own automation bias (Haselager et al. 2023).

This role for AI as a dialogical partner or blind-spot detector holds some promise when it comes to empathy. First, doctors may be oblivious to the (un)empathetic character of their communications, just as they are oblivious to other biases in their identification and decision-making processes. A technology that could help expand this "area of attention for empathy" may enhance empathetic communication. For example, in the form of an AI assistant that doctors could use when communicating with patients in text-based settings, which could flag expressions or pieces of text that can be experienced as unempathetic. This process could also be partially or fully automated. Selinger has suggested developing an "empathy button" on patient portals, that doctors could click on in order to edit their messages to sound more empathetic (see Selinger and Carroll 2024). Similarly, AI could be used in educational settings to allow healthcare professionals to train and practice their empathy skills (Haut et al. 2023). Peeters (2024) draws on a combination of virtue ethics, with its focus on the need to cultivate virtues such as empathy, and the notion that "meaningful

human control" also implies users' moral control over their character when interacting with technologies, to argue that with the right design recommendations carebots can support the cultivation of empathy. He suggests for example a carebot which could present caregivers with questions that can help monitor a caregiver's emotional health, prompting questions for reflection, similar to Haselager et al.'s (2023) "reflection machine".

5.2 Allowing humans to be more human

The idea of AI as a facilitator of empathy is in line with a long tradition that understands the role of digital and information and communication technologies in healthcare as support tools that can take over tasks that stand in the way of human carers doing what they are good at, such as providing empathetic and compassionate care. AI in particular has been upheld in the medical literature as a powerful new tool that can improve doctor–patient interactions and patient-centered care by freeing up healthcare professionals to spend more time on communicating and empathizing with patients (Warraich et al. 2018; Davenport and Kalakota 2019; Sauerbrei et al. 2023). By streamlining workflow processes, outsourcing tasks and decision-making and reducing administrative burdens, AI, it is argued, has the potential to liberate physicians to focus on the humanistic aspects of medicine, of "allow[ing] doctors to be more human" (Ross et al. 2019).

Fogel and Kvedar for example, argue that in addition to allowing for better prevention, detection, diagnosis and treatment of disease, by eliminating repetitive tasks, AI may "clear the way for human-to-human bonding and the application of emotional intelligence and judgment" (Fogel and Kvedar 2018, p. 1). Similarly, using the example of computer-aided detection of breast imaging, Aminololama-Shakeri and López argue that,

> By integrating AI into our daily clinical routines, we may be able once again to prioritize our interactions with patients. This may afford us the time to concentrate on the areas in which humans excel: establishing emotional connections, expressing empathy, and providing patient-centered diagnostic and treatment strategies. (2019, p. 310)

Perhaps more than any other health tech commentator, Eric Topol, a long-term digital health enthusiast and author of the acclaimed *Deep Medicine* (2019), has widely advocated for the benefits of AI, top of which have to do with helping restore medicine's "essential humanness". "The greatest opportunity offered by AI," Topol writes, "is not reducing errors or workloads, or even curing cancer: it is the opportunity to restore the precious and time-honored connection and trust, the human touch, between patients and doctors" (2019, p. 18).

This technology-driven reinvigoration of medicine's humanist roots may sound somewhat paradoxical. But it is not. The basic assumption embedded in these claims is that there is a fundamental, ontological distinction between humans and machines, a distinction which persists, and becomes ever more explicit, despite and in line with technological progress. This distinction is rooted in the idea that care, empathy and compassion are exclusively human skills and qualities, and that the best uses of AI build on and take advantage of this distinction, automating those tasks which do not require essentially human qualities, and creating more space for humans to practice those tasks which do. To the machines the calculative, analytic and hard science; to humans the warm, emotional, moral care. Topol writes, "What might more time [provided by AI] help achieve? […] the genuine sense of care and compassion that lies in direct contrast to what software or machines can offer" (2023, p. 1411). This fundamental distinction also seems to be upheld by the developers and researchers studying empathy in the context of chatbot communication. In an accompanying commentary to the Ayers et al. (2023) paper titled "How Chatbots and Large Language Model Artificial Intelligence Systems Will Reshape Medicine", Li et al. (2023), write "We can hope that emerging AI systems may help time laborious tasks that overwhelm modern medicine and empower physicians to return our focus to treating human patients".

5.3 Ontological trespassing

Empathetic medical chatbots, however, seem to undermine precisely this deep-rooted distinction. They do not have a facilitating, supportive or mediating role. In terms of how they are expected to be implemented in healthcare settings, they will not be assisting, enhancing, boosting or helping to cultivate human professional healthcare workers' competencies. Chatbots which are perceived as empathetic by human patients and observers seem to be doing much more than freeing up humans to do what humans do best. Rather, they are doing *themselves* what humans do best, and even better than humans, thereby transgressing the ontological boundaries which have characterized the traditional division of labor between humans and machines in healthcare.

More than trivial semantics, this subtle shift in discourse is quite momentous, albeit difficult to spot. This is because in the scientific world at least, medical AI and chatbots are not being brazenly promoted *as* empathetic. The series of recent studies referenced in Sect. 2, for example, make claims about the empathetic quality of chatbot conversation and communication skills, not about the empathetic quality of chatbots per se. This important differentiation abides by the above-mentioned ontological distinction between humans

and machines, but it quickly gets eroded. First of all, because the studies themselves do not always preserve this level of precision, even where it matters most. For example, in the "key points" section of their article on chatbot responses to patient questions about cancer, Chen et al. (2024) frame their study in terms of how "conversational artificial intelligence chatbots compare with licensed physicians across measures of empathy". This is easily read as a comparison between chatbots and humans, not just between their respective responses to patient queries. Something similar happens in the terms used in Tu et al.'s study (2024) on Google's LLM-based AMIE system. The investigators did not just ask people to rate chatbot responses compared to human physician responses, but to rate overall "performance" of chatbots compared to humans in a simulated online consultation, based on various criteria, including empathy. The term "performance" lends itself to a broader take on the chatbot as an agent.

Second, even if the scientific literature (mostly) succeeds in maintaining crucial distinctions between claims concerning the quality of a chatbot's output and the chatbot itself, of sounding rather than being empathetic, the sensationalist headlines that often accompany science communication tend not to. *Nature*'s report on the AMIE study was titled "Google AI has better bedside manners than human doctors" (Lenharo 2024). Unsurprisingly, precision about what, exactly, is empathetic about or in relation to a chatbot is even more difficult to find in marketing claims. For example, in Youper's marketing of its mental health chatbot, it is the chatbot, not its communication skills, which is qualified as "empathetic, safe and clinically validated" in the website's tagline.[5]

These subtle slippages are consequential. First, because, as discussed, the simulated empathy that can be provided by chatbots is a hollowed-out version of a thick and rich concept which plays a fundamental role in the practice and aims of healthcare. Furthermore, these slippages point to a specific framing of the empathy problem in healthcare. As discussed, technosolutionism works by redefining or reconstructing complex problems in such a way that it seems that they can be solved by a given technological solution (Siffels and Sharon 2024). Indeed, if empathetic medical chatbots are proposed as providers of empathy (simulated if not genuine), rather than as facilitators supporting humans in providing empathy, then the problem they are addressing is an empathy deficit on the part of healthcare professionals. This is a particular way of defining healthcare's empathy problem, and one which risks drawing attention away from other issues underlying healthcare's empathy problem, namely an absence of important contextual and environmental factors that are necessary for empathy, first and foremost: time. By defining healthcare's empathy problem in this way, empathetic medical chatbots thus risk creating what Siffels and I (2024) have called an "orphan problem".

6 Creating an orphan problem

6.1 Effects of contextual factors on the ability to practice empathy

The suggestion embedded in empathetic medical chatbots *qua* solutions is that the empathy problem in healthcare has to do with an empathy deficit on the part of healthcare professionals. This is not a completely unfamiliar argument in discussions on empathy and compassion in healthcare.[6] More often, however, studies point to environmental and contextual factors, rather than personal competencies and values, to explain uncompassionate or unempathetic behavior.

One of the first studies to investigate the effect of environmental factors on prosocial behavior was conducted by the social psychologists Darely and Batson in 1973, in the now famous "Good Samaritan Study" (Darley and Batson 1973). The experiment involved university seminary students, who, on their way to give a lecture, pass a person asking for help. Students who were told they were running late for their lecture were much less likely to stop and assist the person compared to those who were not in a rush. This was the case even for students who were on their way to give a lecture about the Good Samaritan parable, and so, who might be expected to be primed toward compassion. Darley and Batson argued that these findings demonstrate that situational constraints, such as conflicting obligations and time pressure, could significantly impact an individual's willingness to engage in prosocial behavior. Since then, numerous social psychology studies have sought to learn about the conditions in which people do or do not partake in prosocial behavior. Similarly, findings in empirical psychology in recent decades have challenged the view that virtues, such as empathy, honesty, and generosity, are to be found in the moral character of individuals. Rather, proponents of "situationism" uphold that situational factors, such as peer pressure, are better predictors of behavior (Harman 1999; Doris 2002).

Such environmental and contextual factors have also been invoked in various studies that seek to explain compassion

[5] https://www.youper.ai.

[6] For example, in the national scandal concerning a series of failures of care at the Mid Staffordshire Hospital between 2005 and 2009 in the UK, some researchers and commentators pointed to a lack of compassion on the part of staff, even suggesting that some nurses do not have the personal qualities and commitment needed to be nurses (Darbyshire 2014; Stenhouse et al. 2016).

and empathy "deficits" among healthcare staff. The *Francis Report* (2013), published following a public enquiry into a series of failures of care documented between 2004 and 2009 at Mid Staffordshire Hospital in the UK, highlights institutional and systemic structures, such as overwork and lack of continuity, as explanatory factors. Paley (2013, 2014), drawing strongly on Darley and Batson and situationism, argues that events that took place at the Mid Staffordshire Hospital should not be seen as a failure of compassion that had somehow to do with the ability of individuals to empathize, but with an interlocking set of contextual factors. Also referring to the Mid Straffordshire case, Roberts and Ion (2014) explain the uncompassionate behavior of some staff in terms of a failure to be engaged in and reflect on their actions. Drawing on Hanna Arendt's discussion of the banality of evil in the Holocaust, they suggest that a focus on day-to-day demands, such as managing waiting lists and meeting targets distanced staff from the real purpose of their care work, thus allowing unacceptable behavior to become a norm.

6.2 Time as a condition for providing empathetic care

In this context, an empathy deficit on the part of healthcare professionals can be seen as a bad explanation and poor problem definition for failures of care and healthcare's "empathy problem". Rather, conflicting priorities, protocol-driven care, bureaucracy, and perhaps most importantly time pressure and stress, are commonly cited contextual factors which create barriers to empathy on the part of clinicians and healthcare staff (Derksen et al. 2016; Kerasidou 2019; Howick et al. 2018a; Bivins et al. 2017; Razi et al. 2023; Neumann et al. 2011b). This should not be surprising. Insofar as empathy involves a dialogue in which professionals need to take time to listen closely to patients in order to identify needs and concerns which warrant recognition and validation and in which patients need to have time to ask questions, and insofar as empathy involves building a lasting connection with patients that leads to more understanding of their situation, empathy is inherently time-intensive. Moreover, insofar as time constraints can negatively impact people's natural tendency to help people in need, as Darley and Batson's (1973) Good Samaritan study showed, it is more difficult to practice empathy under time pressure. In a qualitative study conducted in three emergency departments in England, Kerasidou (2019) looked specifically into the effects of post-2008 austerity policies and their intensified focus on urgency and efficiency on the everyday work of doctors and nurses. She found that the pursuit of greater efficiency, operationalized in this context as seeing more patients in shorter amounts of time, often resulted in the values of empathy and compassion being side-lined, with staff struggling to maintain or incorporate these values into their everyday work. A focus on efficiency and meeting operational targets, Kerasidou's informants explained, increased their workload and stress, curtailed their autonomy and decreased the space for empathy and compassion, which they saw as core to their profession. Interestingly in relation to the topic of AI and empathy, they expressed that the focus on efficiency was dehumanizing the therapeutic relationship, with one doctor commenting that, it made them "feel a bit less human […] as if you are a robot or a computer" (2019, p. 178). Ontological trespassing can occur in both directions.

If time pressure and workload stress negatively impact healthcare workers' possibilities to provide empathy, it should not be surprising that there is an empathy problem in healthcare at a time when healthcare staff consistently suffer from workload and time constraints. Numerous healthcare staff, from primary care physicians to general practitioners and nurses commonly describe experiencing time pressure during and outside of patient encounters (Freedman et al 2021; Linzer et al. 2009; Mechanic 2001). And burnout, while occurring in all kinds of jobs, has a high incidence in healthcare worldwide (Woo et al. 2020), and is becoming an important challenge for public health. This implies that the root cause and real problem underlying healthcare's empathy problem is not that healthcare staff are not capable of empathy, or less capable than chatbots, but that the time and space needed for empathetic care is lacking. In this context, that AI systems score higher than human healthcare providers on empathic communication is not surprising, insofar as AI systems are not under the same time pressures as their human counterparts.

6.3 Digitalization: providing the gift of time or taking it away?

There are a number of explanations as to the experienced time constraints of healthcare staff in recent years, including the increased complexity of conditions that patients present with, the aging of the population in need of healthcare, reimbursements being tied to time-based billing codes (at least in the United States), workforce shortages and administrative burdens (Mauksch et al. 2008; Mechanic 2001; Sipos et al. 2024). Most importantly for the discussion on technosolutionism, however, digital technologies play a non-negligible role in this predicament, often promising the gift of time via increased efficiency and relief of administrative burden, but also often taking it away (Sparrow and Hatherley 2020).

6.3.1 The electronic health record

A cogent example is the electronic healthcare record (EHR). Introduced into healthcare systems in many countries worldwide in the 1980s and 1990s, EHRs aim to provide

comprehensive, cross-institutional and longitudinal records of patients' healthcare data, with the goals of providing a cost-effective and efficient alternative to paper-based medical records. While EHRs may reduce medical errors and improve care through compliance with guidelines and access to health information, numerous studies in various countries in recent years have pointed to negative impacts EHRs have had on physician workload and patient–doctor interactions, with some studies citing EHRs as a primary driver of clinician burnout (see for example Babbott et al. 2014; Kroth et al. 2018).

This has to do with the increased amount of time doctors spend on EHRs, entering data, navigating user interfaces and the non-intuitive workflows they impose, and the ability that digital technology offers to work anywhere and anytime, contributing to so-called "pyjama time", or time physicians spend outside of the office working on administrative tasks (Saag et al. 2019). The introduction of EHRs has created additional tasks for physicians to complete in the same amount of time and has had a significant impact on patient–doctor interactions. Studies show that doctors now spend more time interacting with the EHR during consults and less with patients, they spend less time in the exam room or at bedside, they communicate less effectively, they have less time for conversation, and their attention is diverted to computer screens (both in the form of gaze and body positioning) (Asan et al. 2014), all elements that are important for the provision of empathy. One doctor has described the experience of interacting with the EHR as, "Like having someone at the dinner table texting rather than paying attention" (Pelland et al. 2017).

6.3.2 AI to the rescue

AI technologies play a notable role in narratives around the gift of time. As discussed in Sect. 5, one of the biggest promises of AI in the healthcare setting is to save physicians time, namely by taking over and automating much of the administrative tasks that healthcare providers are burdened with, especially since the introduction of EHRs. AI systems should automate clinical documentation, coding, authorizations, information retrieval and patient intake. Generative AI and LLM-based AI scribes in particular, which can transcribe and summarize interactions between doctors and patients and automatically fill out EHRs, are expected to reduce the administrative workload of physicians. As in the past, the hope is to render physicians' work more efficient, and, this time around, to give back the time that had been taken away by EHRs so that clinicians can (re)focus on interacting with patients and on delivering empathetic, patient-centered care (Warraich et al. 2018). Despite the absence of clear guidelines on where and how to introduce these technologies, numerous hospitals and medical practices are already piloting them in clinical workflows (Landi 2024) in light of these promises. This is problematic for a number of reasons already mentioned, including the high prevalence of "hallucinations", privacy and bias risks, as well as novel questions concerning the desirability of outsourcing cognitively complex tasks, such as note-taking and summarizing, to LLMs (McCoy et al. 2024). More pertinent for the discussion here, it is also questionable, as with previous technologies, if generative AI will actually save healthcare practitioners any time.

A recent study, for example, found that use of an LLM to help physicians reply to patient messages (similar to the one implemented at the UMC Groningen in the Netherlands mentioned in Sect. 2) did not save clinicians time, but increased the overall time they spent for each patient query, by 22% (Tai-Seale et al. 2024). The authors suggest that the increased time spent may be a result of physicians needing to read the patient's original question before relying on the AI, as well as scrutinizing the AI's response for errors. Importantly, the propensity for errors that generative AI is known for, and the high-risk context of healthcare, means that increases in time spent interacting with AI is inevitable, insofar as a human will need to monitor and verify this output. This is an unintended but predictable consequence of the use of LLMs to alleviate the (also) unintended workload caused by EHRs (McCoy et al. 2024). In order to maintain the trust of their patients, and in order to comply with legal requirements, healthcare practitioners will need to supervise and review AI, to ensure that it is operating properly and not hallucinating. It remains to be seen if this will, in total, take more time away from practitioners than it will save them, but it will in any case create new tasks. Tasks, moreover, which may be less meaningful and interesting than those which attracted healthcare practitioners to the field in the first place.

Healthcare's real empathy problem, that healthcare providers do not have the time and space needed to provide empathetic care (in part at least *because* of the need to interact with technologies), is not the one that empathetic medical chatbots address, as they suggest that they provide more empathy than human healthcare professionals. In this context, healthcare's real empathy problem risks becoming an orphan problem, and attention and resources may be drawn away from the need to find structural solutions required for empathetic healthcare systems (Kerasidou 2019) to chatbots which are ever more successful in simulating empathy. Along the way, empathy, one of the most fundamental values of health and medicine, risks being hollowed out and redefined as a codifiable communication pattern.

7 Conclusion

This article argued for the value of understanding the emergence of so-called empathetic chatbots in the healthcare context through the lens of technosolutionism, a concept which focuses on the relationship between (techno) solutions and the problems they seek to address. Importantly, the technosolutionist lens sheds light on concerns raised by LLMs and generative AI in addition to already widely acknowledged privacy, bias and hallucination risks.

Chatbots cannot actually be empathetic. As AI critics, but also most AI developers and researchers would doubtless agree, they can at best simulate empathy, by generating textual patterns involving keywords and expressions which convey empathy in human-to-human communication. Furthermore, empathy requires emotion-guided capacities that AI does not have, even as it becomes more skillful at cognitive tasks. This in itself does not constitute a reason for not introducing chatbots which communicate empathetically into the healthcare setting, where human healthcare practitioners do not always practice "genuine" empathy themselves, and where a perception of empathy may be sufficient to achieve significant beneficial effects of empathy. But AI-simulated empathy, which can be seen as a form of deception, raises various other objections, including manipulation harms and an undermining of respect for persons. Moreover, even as researchers may insist that the empathy conveyed by chatbots is only simulated empathy, and that it is chatbot responses (not the chatbot themselves) which are empathetic, in research discussions, science communication and certainly marketing material, it is the chatbots themselves, or their overall performance, which tends to be qualified as empathetic. It would be dangerous to dismiss this as trivial semantics. Given the financial pressures and personnel shortages which healthcare systems currently face, there is a risk that this narrow understanding of empathy as a communication pattern may replace the rich, relational and multi-faceted definition of empathy which is widely recognized as a core value of patient-centered care.

In addition to the risk of hollowing out the concept of empathy, empathetic medical chatbots risk redefining healthcare's empathy problem. As discussed, empathetic chatbots are not being promoted as tools that would facilitate or assist healthcare practitioners in the provision of empathetic care, but as providers of empathy (albeit a hollowed-out version of empathy) themselves. The suggestion is then that healthcare's empathy problem lies in an empathy deficit on the part of healthcare practitioners, while it actually lies in a lack of space and time that healthcare practitioners have at their disposal, and which is necessary for empathy. The bitter irony is that digital technology, while invariably promising the "gift of time", has played an important role in exacerbating the root causes of healthcare's empathy problem, for example with the introduction of EHRs. More digital technology, now in the form of various types of AI (including medical chatbots), will also inevitably create new tasks and time constraints, such as the need to monitor AI output (Sparrow and Hatherley 2020). Again, given the financial pressures and personnel shortages that healthcare systems face today, healthcare's real empathy problem may become an "orphan problem", while resources are directed toward developing and implementing empathetic chatbots.

These dynamics, a redefinition of core concepts and of real problems to fit what technologies can actually do, are key characteristics of technosolutionism. This does not imply that LLMs and AI-based chatbots should not be used to seek to address problems in healthcare, even healthcare's empathy problem. But that they should be developed and implemented in such a way that they align with existing core organizing concepts (in healthcare as in other sectors where they are being deployed), rather than redefining them, and in such a way that they address the actual problems which these sectors face.

Acknowledgements This work was supported by the Ammodo Science Award 2022 for the Humanities. I would like to thank participants at the iHub monthly seminar, at the workshop "Empirical Ethics and Socio-Technical Entanglement" at the Insitute of Advanced Studies at the University of Amsterdam and at the "Losses and Gains of Digital Developments" workshop at Helsinki University for comments on versions of this manuscript.

Author contributions T.S. is the sole author of the manuscript.

Data availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare no competing interests.

References

Albahri AS, Duhaim AM, Fadhel MA, Alnoor A, Baqer NS, Alzubaidi L, Albahri OS, Alamoodi AH, Bai J, Salhi A, Santamaría J,

Ouyang C, Gupta A, Gu Y, Deveci M (2023) A systematic review of trustworthy and explainable artificial intelligence in healthcare: assessment of quality, bias risk, and data fusion. Inf Fusion 96:156–191. https://doi.org/10.1016/j.inffus.2023.03.008

Alkaissi H, McFarlane SI (2023) Artificial hallucinations in ChatGPT: implications in scientific writing. Cureus 15(2):e35179. https://doi.org/10.7759/cureus.35179

Alowais SA, Alghamdi SS, Alsuhebany N, Alqahtani T, Alshaya AI, Almohareb SN, Aldairem A, Alrashed M, Bin Saleh K, Badreldin HA, Al Yami MS, Al Harbi S, Albekairy AM (2023) Revolutionizing healthcare: the role of artificial intelligence in clinical practice. BMC Med Educ 23(1):689. https://doi.org/10.1186/s12909-023-04698-z

Aminololama-Shakeri S, López JE (2019) The doctor-patient relationship with artificial intelligence. AJR Am J Roentgenol 212(2):308–310. https://doi.org/10.2214/AJR.18.20509

Asan O, Smith PD, Montague E (2014) More screen time, less face time. J Eval Clin Pract 20:896–901. https://doi.org/10.1111/jep.12182

Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, Faix DJ, Goodman AM, Longhurst CA, Hogarth M, Smith DM (2023) Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med 183(6):589–596. https://doi.org/10.1001/jamainternmed.2023.1838

Ayoub NF, Balakrishnan K, Ayoub MS, Barrett TF, David AP, Gray ST (2024) Inherent bias in large language models: a random sampling analysis. Mayo Clin Proc Digit Health 2(2):186–191. https://doi.org/10.1016/j.mcpdig.2024.03.003

Babbott S, Manwell LB, Brown R et al (2014) Electronic medical records and physician stress in primary care: results from the MEMO Study. J AM Med Inform Assoc 21(e1):e100–e106

Baron-Cohen S (2011) Zero degrees of empathy: a new theory of human cruelty. Allen Lane, London

Bauchat JR, Seropian M, Jeffries PR (2016) Communication and empathy in the patient-centered care model—why simulation-based training is not optional. Clin Simul Nurs 12(8):356–359. https://doi.org/10.1016/j.ecns.2016.04.003

Bender EM, Gebru T, McMillan-Major A, Shmitchell S (2021) On the dangers of stochastic parrots: can language models be too big? In: Proceedings of the 2021 ACM conference on fairness, accountability, and transparency, pp 610–623. https://doi.org/10.1145/3442188.3445922

Bivins R, Tierney S, Seers K (2017) Compassionate care: not easy, not free, not only nurses. BMJ Qual Saf 26:1023–1026. https://doi.org/10.1136/bmjqs-2017-007005

Bloom P (2016) Against empathy. Harper Collins, New York

Chen D, Parsa R, Hope A, Hannon B, Mak E, Eng L, Liu FF, Fallah-Rad N, Heesters AM, Raman S (2024) Physician and artificial intelligence chatbot responses to cancer questions from social media. JAMA Oncol 10(7):956–960. https://doi.org/10.1001/jamaoncol.2024.0836

Choudhury A, Asan O (2020) Role of artificial intelligence in patient safety outcomes: systematic literature review. JMIR Med Inform 8(7):e18599. https://doi.org/10.2196/18599

Clusmann J, Kolbinger FR, Muti HS, Carrero ZI, Eckardt J-N, Laleh NG, Löffler CML, Schwarzkopf S-C, Unger M, Veldhuizen GP, Wagner SJ, Kather JN (2023) The future landscape of large language models in medicine. Commun Med 3(1):1–8. https://doi.org/10.1038/s43856-023-00370-1

Coeckelbergh M (2018) Why care about robots? Empathy, moral standing, and the language of suffering. Kairos 20(1):141–158

Danaher J (2020) Robot betrayal: a guide to the ethics of robotic deception. Ethics Inf Technol 22(2):117–128. https://doi.org/10.1007/s10676-019-09520-3

Darbyshire P (2014) Character assassination? Response to John Paley, "social psychology and the compassion deficit." Nurse Educ Today 34:887–889

Darley JM, Batson CD (1973) From Jerusalem to Jericho: a study of situational and dispositional variables in helping behaviour. J Pers Soc Psychol 27(1):100–108

Davenport T, Kalakota R (2019) The potential for artificial intelligence in healthcare. Future Healthc J 6(2):94–98. https://doi.org/10.7861/futurehosp.6-2-94

Derksen F, Hartman T, Bensing J, Lagro-Janssen A (2016) Managing barriers to empathy in the clinical encounter: a qualitative interview study with GPs. Br J Gen Pract 66(653):e887–e895. https://doi.org/10.3399/bjgp16X687565

Di Blasi Z, Harkness E, Ernst E, Georgiou A, Kleijnen J (2001) Influence of context effects on health outcomes: a systematic review. Lancet 357(9258):757–762. https://doi.org/10.1016/S0140-6736(00)04169-6

Doris J (2002) Lack of character: personality and moral behaviour. Cambridge University Press, Cambridge

Fogel AL, Kvedar JC (2018) Artificial intelligence powers digital medicine. Npj Digit Med 1(1):1–4. https://doi.org/10.1038/s41746-017-0012-2

Francis R (2013) Report of the Mid Staffordshire NHS Foundation Trust public inquiry. The Stationary Office, London

Freedman S, Golberstein E, Huang TY, Satin DJ, Smith LB (2021) Docs with their eyes on the clock? The effect of time pressures on primary care productivity. J Health Econ 77:102442. https://doi.org/10.1016/j.jhealeco.2021.102442

Gilbert S, Harvey H, Melvin T, Vollebregt E, Wicks P (2023) Large language model AI chatbots require approval as medical devices. Nat Med 29(10):2396–2398. https://doi.org/10.1038/s41591-023-02412-6

Halpern J (2001) From detached concern to empathy: humanizing medical practice. Oxford University Press, Oxford

Halpern J (2003) What is clinical empathy? J Gen Intern Med 18(8):670–674. https://doi.org/10.1046/j.1525-1497.2003.21017.x

Halpern J (2007) Empathy and patient-physician conflicts. J Gen Intern Med 22(5):696–700

Harman G (1999) Moral philosophy meets social psychology: virtue ethics and the fundamental attribution error. Proc Aristot Soc 99:315–331

Haselager P, Schraffenberger H, Thill S, Fischer S, Lanillos P, van de Groes S, van Hooff M (2023) Reflection machines: supporting effective human oversight over medical decision support systems. Camb Q Healthc Ethics. https://doi.org/10.1017/S0963180122000718

Haut K, Wohn C, Kane B, Carroll T, Guigno C, Kumar V, Epstein R, Schubert L, Hoque E (2023) Validating a virtual human and automated feedback system for training doctor-patient communication skills. arXiv. https://doi.org/10.48550/arXiv.2306.15213

Healey ML, Grossman M (2018) Cognitive and affective perspective-taking: evidence for shared and dissociable anatomical substrates. Front Neurol. https://doi.org/10.3389/fneur.2018.00491

Howick J, Bizzari V, Dambha-Miller H (2018a) Therapeutic empathy: what it is and what it isn't. J R Soc Med 111(7):233–236. https://doi.org/10.1177/0141076818781403

Howick J, Moscrop A, Mebius A, Fanshawe TR, Lewith G, Bishop FL, Mistiaen P, Roberts NW, Dieninytė E, Hu XY, Aveyard P, Onakpoya IJ (2018b) Effects of empathic and positive communication in healthcare consultations: a systematic review and meta-analysis. J R Soc Med 111(7):240–252. https://doi.org/10.1177/0141076818769477

Ingram D (2023) A mental health tech company ran an AI experiment on real users. NBC News. https://www.nbcnews.com/tech/inter

net/chatgpt-ai-experiment-mental-health-tech-app-koko-rcna65110
Inzlicht M, Cameron CD, D'Cruz J, Bloom P (2024) In praise of empathic AI. Trends Cogn Sci 28(2):89–91. https://doi.org/10.1016/j.tics.2023.12.003
Kerasidou A (2019) Empathy and efficiency in healthcare at times of austerity. Health Care Anal HCA J Health Philos Policy 27(3):171–184. https://doi.org/10.1007/s10728-019-00373-x
Kerasidou A (2020) Artificial intelligence and the ongoing need for empathy, compassion and trust in healthcare. Bull World Health Organ 98(4):245–250. https://doi.org/10.2471/BLT.19.237198
Kerasidou A, Bærøe K, Berger Z, Brown AEC (2021) The need for empathetic healthcare systems. J Med Ethics 47(12):e27–e27. https://doi.org/10.1136/medethics-2019-105921
Kim SS, Kaplowitz S, Johnston MV (2004) The effects of physician empathy on patient satisfaction and compliance. Eval Health Prof 27(3):237–251. https://doi.org/10.1177/0163278704267037
Kiosses VN, Karathanos VT, Tatsioni A (2016) Empathy promoting interventions for health professionals: a systematic review of RCTs. J Compassionate Health Care 3(1):7. https://doi.org/10.1186/s40639-016-0024-9
Koenecke A, Choi ASG, Mei KX, Schellmann H, Sloane M (2024) Careless whisper: speech-to-text hallucination harms. In: Proceedings of the 2024 ACM conference on fairness, accountability, and transparency, pp 1672–1681. https://doi.org/10.1145/3630106.3658996
Kroth PJ, Morioka-Douglas N, Veres S et al (2018) The electronic elephant in the room: physicians and the electronic health record. JAMIA Open 11:49–56
Kruse J (2023) UMCG beantwoordt vragen patiënten met hulp van AI. Nieuws en Verhalen. https://nieuws.umcg.nl/w/umcg-beantwoordt-vragen-pati%C3%ABnten-met-hulp-van-ai
Kudina O, de Boer B (2021) Co-designing diagnosis: towards a responsibe integration of machine learning decision-support systems in medical diagnostics. J Eval Clin Pract. https://doi.org/10.1111/jep.13535
Landi H (2024) Health systems ramp up AI partnerships to tackle doctor burnout. https://www.fiercehealthcare.com/ai-and-machine-learning/rush-childrens-hospital-los-angeles-health-systems-are-ramping-ai-startup
Lenharo M (2024) Google AI has better bedside manner than human doctors—and makes better diagnoses. Nature 625(7996):643–644. https://doi.org/10.1038/d41586-024-00099-4
Li Z (2023) The dark side of ChatGPT: legal and ethical challenges from stochastic parrots and hallucination. https://doi.org/10.48550/arXiv.2304.14347
Li R, Kumar A, Chen JH (2023) How chatbots and large language model artificial intelligence systems will reshape modern medicine: Fountain of creativity or Pandora's Box? JAMA Intern Med 183(6):596–597. https://doi.org/10.1001/jamainternmed.2023.1835
Linzer M, Baier Manwell L, Williams ES et al (2009) Working conditions in primary care: physician reactions and care quality. Ann Intern Med 151:28–36. https://doi.org/10.7326/0003-4819-151-1-200907070-00006
Lorenzini G, Ossa LA, Milford S, Elger BS, Shaw DM, Clercq ED (2024) The "magical theory" of AI in medicine: thematic narrative analysis. JMIR AI 3(1):e49795. https://doi.org/10.2196/49795
Maréchal N (2021) The future of platform power: fixing the business model. J Democr 32(3):157–162. https://doi.org/10.1353/jod.2021.0041
Mathews AW (2024) What AI can do in healthcare—and what it should never do. WSJ. https://www.wsj.com/tech/ai/what-ai-can-do-in-healthcareand-what-it-should-never-do-3e28a2b4
Matthias A (2015) Robot lies in health care: when is deception morally permissible? Kennedy Inst Ethics J 25(2):169–192
Mauksch LB, Dugdale DC, Dodson S, Epstein R (2008) Relationship, communication, and efficiency in the medical encounter: creating a clinical model from a literature review. Arch Intern Med 168(13):1387–1395. https://doi.org/10.1001/archinte.168.13.1387
McCoy LG, Manrai AK, Rodman A (2024) Large language models and the degradation of the medical record. N Engl J Med 391(17):1561–1564. https://doi.org/10.1056/NEJMp2405999
Meacham D, Studley M (2017) Could a robot care? it's all in the movement. In: Lin P, Abney K, Jenkins R (eds) Robot Ethics 2.0: from autonomous cars to artificial intelligence. Oxford University Press, Oxford. https://doi.org/10.1093/oso/9780190652951.003.0007
Mechanic D (2001) How should hamsters run? some observations about sufficient patient time in primary care. BMJ 323(7307):266–268
Meskó B, Topol EJ (2023) The imperative for regulatory oversight of large language models (or generative AI) in healthcare. NPJ Digit Med 6(1):120. https://doi.org/10.1038/s41746-023-00873-0
Michael K, Abbas R, Calvo RA, Roussos G, Scornavacca E, Wamba SF (2020) Manufacturing consent: the modern pandemic of technosolutionism. IEEE Trans Technol Soc 1(2):68–72. https://doi.org/10.1109/TTS.2020.2994381
Montemayor C, Halpern J, Fairweather A (2022) In principle obstacles for empathic AI: why we can't replace human empathy in healthcare. AI Soc 37(4):1353–1359. https://doi.org/10.1007/s00146-021-01230-z
Moor M, Banerjee O, Abad ZSH, Krumholz HM, Leskovec J, Topol EJ, Rajpurkar P (2023) Foundation models for generalist medical artificial intelligence. Nature 616(7956):259–265. https://doi.org/10.1038/s41586-023-05881-4
Morley J, Floridi L (2023) Foundation models are exciting, but they should not disrupt the foundations of caring. SSRN Electron J. https://doi.org/10.2139/ssrn.4424821
Morozov E (2013) To save everything, click here: the folly of technological solutionism. PublicAffairs, New York
Najjar N, Davis LW, Beck-Coon K, Carney Doebbeling C (2009) Compassion fatigue: a review of the research to date and relevance to cancer-care providers. J Health Psychol 14(2):267–277. https://doi.org/10.1177/1359105308100211
Natale S (2021) Deceitful media: artificial intelligence and social life after the Turing test. Oxford University Press, Oxford. https://doi.org/10.1093/oso/9780190080365.001.0001
NEJM Catalyst (2017) What is patient-centered care? Catal Carryover. https://doi.org/10.1056/CAT.17.0559
Neumann M, Edelhäuser F, Tauschel D, Fischer MR, Wirtz M, Woopen C, Haramati A, Scheffer C (2011a) Empathy decline and its reasons: a systematic review of studies with medical students and residents. Acade Med J Assoc Am Med Coll 86(8):996–1009. https://doi.org/10.1097/ACM.0b013e318221e615
Neumann M, Bensing J, Wirtz M, Wübker A, Scheffer C, Tauschel D, Edelhäuser F, Ernstmann N, Pfaff H (2011b) The impact of financial incentives on physician empathy: a study from the perspective of patients with private and statutory health insurance. Patient Educ Couns 84(2):208–216. https://doi.org/10.1016/j.pec.2010.07.012
Ockenden D (2022) Ockenden report—final: findings, conclusions and essential actions from the Independent Review of Maternity Services at The Shrewsbury and Telford Hospital NHS Trust. https://www.ockendenmaternityreview.org.uk/wp-content/uploads/2022/03/FINAL_INDEPENDENT_MATERNITY_REVIEW_OF_MATERNITY_SERVICES_REPORT.pdf
Paley J (2013) Social psychology and the compassion deficit. Nurse Educ Today 33:1451–1452

Paley J (2014) Cognition and the compassion deficit: the social psychology of helping behaviour in nursing. Nurs Philos 15(4):274–287. https://doi.org/10.1111/nup.12047

Park PS, Goldstein S, O'Gara A, Chen M, Hendrycks D (2024) AI deception: a survey of examples, risks, and potential solutions. Patterns. https://doi.org/10.1016/j.patter.2024.100988

Parliamentary and Health Service Ombudsman (2015) Dying without dignity: investigations by the parliamentary and health service Ombudsman into complaints about end of life care. https://www.ombudsman.org.uk/sites/default/files/Dying_without_dignity.pdf

Patel BN, Rosenberg L, Willcox G, Baltaxe D, Lyons M, Irvin J, Rajpurkar P, Amrhein T, Gupta R, Halabi S, Langlotz C, Lo E, Mammarappallil J, Mariano AJ, Riley G, Seekins J, Shen L, Zucker E, Lungren MP (2019) Human–machine partnership with artificial intelligence for chest radiograph diagnosis. NPJ Digit Med 2:111. https://doi.org/10.1038/s41746-019-0189-7

Peeters A (2024) Out of control: flourishing with carebots through embodied design. In: Mecacci G, Amoroso D, Cavalcante Siebert L, Abbink D, Van DenHoven J, Santoni De Sio F (eds) Research handbook on meaningful human control of artificial intelligence systems. Edward Elgar Publishing, Cheltenham, pp 38–52. https://doi.org/10.4337/9781802204131.00010

Pelland KD, Baier RR, Gardner RL (2017) 'It is like texting at the dinner table': a qualitative analysis of the impact of electronic health records on patient–physician interaction in hospitals. J Innov Health Inform 24(2):216–223

Ranchordás S (2022) Empathy in the digital administrative state. Duke Law J 71:1341–1389

Razi MO, Fouzia R, Razzaque MS (2023) Decline of empathy among healthcare apprentices. Int Med Educ. https://doi.org/10.3390/ime2040022

Reisman J (2024) I'm a doctor. ChatGPT's bedside manner is better than mine. The New York Times. https://www.nytimes.com/2024/10/05/opinion/ai-chatgpt-medicine-doctor.html

Ribes D, Hoffman AS, Slota SC, Bowker GC (2019) The logic of domains. Soc Stud Sci 49(3):281–309. https://doi.org/10.1177/0306312719849709

Riess H, Kelley JM, Bailey RW, Dunn EJ, Phillips M (2012) Empathy training for resident physicians: a randomized controlled trial of a neuroscience-informed curriculum. J Gen Intern Med 27(10):1280. https://doi.org/10.1007/s11606-012-2063-z

Roberts M, Ion R (2014) A critical consideration of systemic moral catastrophe in modern health care systems: a big idea from an Arendtian perspective. Nurse Educ Today 34(5):673–675

Roose K (2024) Can A.I. be blamed for a teen's suicide? The New York Times. https://www.nytimes.com/2024/10/23/technology/characterai-lawsuit-teen-suicide.html

Rosenzweig MQ (2012) Breaking bad news: a guide for effective and empathetic communication. Nurse Pract 37(2):1–4

Ross J, Webb C, Rahman F, AoRCM (2019) Artificial intelligence in healthcare: academy of medical royal colleges. https://www.aomrc.org.uk/wp-content/uploads/2019/01/Artificial_intelligence_in_healthcare_0119.pdf

Rumale P, Tiwari S, Naik T, Gupta S, Thai D, Zhao W, Kwon S, Ardulov V, Tarabishy K, McCallum A, Salloum W (2024) Faithfulness hallucination detection in healthcare AI. In: Proceedings of KDD 2024 workshop—artificial intelligence and data science for healthcare: bridging data-centric AI and people-centric healthcare. https://openreview.net/pdf?id=6eMIzKFOpJ

Saag HS, Shah K, Jones SA, Testa PA, Horwitz LI (2019) Pajama time: Working after work in the electronic health record. J Gen Intern Med 34(9):1695–1696. https://doi.org/10.1007/s11606-019-05055-x

Sauerbrei A, Kerasidou A, Lucivero F, Hallowell N (2023) The impact of artificial intelligence on the person-centred, doctor-patient relationship: some problems and solutions. BMC Med Inform Decis Mak. https://doi.org/10.1186/s12911-023-02162-y

Selinger E, Carroll T (2024) Can AI help ease medicine's empathy problem? STAT. https://www.statnews.com/2024/08/15/can-ai-help-ease-empathy-problem-medicine/

Sezgin E (2023) Artificial intelligence in healthcare: complementing, not replacing, doctors and healthcare providers. Digit Health 9:20552076231186520. https://doi.org/10.1177/20552076231186520

Sharkey N, Sharkey A (2010) Living with robots: ethical tradeoffs in eldercare. In: Wilks Y (ed) Close engagements with artificial companions: key social, psychological, ethical and design issues. John Benjamins, Amsterdam, pp 245–256

Sharkey A, Sharkey N (2021) We need to talk about deception in social robotics! Ethics Inf Technol 23(3):309–316. https://doi.org/10.1007/s10676-020-09573-9

Shi W, Wang X, Oh YJ, Zhang J, Sahay S, Yu Z (2020) Effects of persuasive dialogues: testing bot identities and inquiry strategies. In: Proceedings of the 2020 CHI conference on human factors in computing systems, pp 1–13. https://doi.org/10.1145/3313831.3376843

Siffels L, Sharon T (2024) Where technology leads, the problems follow. Technosolutionism and the Dutch contact tracing app. Philos Technol 37:125. https://doi.org/10.1007/s13347-024-00807-y

Singer T, Klimecki OM (2014) Empathy and compassion. Curr Biol 24(18):R875–R878. https://doi.org/10.1016/j.cub.2014.06.054

Sipos D, Goyal R, Zapata T (2024) Addressing burnout in the healthcare workforce: Current realities and mitigation strategies. Lancet Region Health Eur. https://doi.org/10.1016/j.lanepe.2024.100961

Sparrow R (2002) The march of the robot dogs. Ethics Inf Technol 4:305–318

Sparrow R, Hatherley J (2020) High hopes for "deep medicine"? AI, economics, and the future of care. Hastings Cent Rep 50(1):14–17. https://doi.org/10.1002/hast.1079

Sparrow R, Sparrow L (2006) In the hands of machines? The future of aged care. Minds Mach J Artif Intell Philos Cogn Sci 16(2):141–161. https://doi.org/10.1007/s11023-006-9030-6

Spiro H (2009) Commentary: the practice of empathy. Acad Med J Assoc Am Med Coll 84(9):1177–1179. https://doi.org/10.1097/ACM.0b013e3181b18934

Stenhouse R, Ion R, Roxburgh M, Devitt P, Smith SDM (2016) Exploring the compassion deficit debate. Nurse Educ Today 39:12–15. https://doi.org/10.1016/j.nedt.2016.01.019

Switzky L (2020) ELIZA effects: Pygmalion and the early development of artificial intelligence. Shaw 40(1):50–68. https://doi.org/10.5325/shaw.40.1.0050

Tai-Seale M, Baxter SL, Vaida F, Walker A, Sitapati AM, Osborne C, Diaz J et al (2024) AI-generated draft replies integrated into health records and physicians' electronic communication. JAMA Netw Open 7(4):e246565. https://doi.org/10.1001/jamanetworkopen.2024.6565

Toma A, Senkaiahliyan S, Lawler PR, Rubin B, Wang B (2023) Generative AI could revolutionize health care—but not if control is ceded to big tech. Nature 624(7990):36–38. https://doi.org/10.1038/d41586-023-03803-y

Topol EJ (2019) Deep medicine: how artificial intelligence can make healthcare human again. Basic Books, New York

Topol EJ (2023) Machines and empathy in medicine. Lancet 402(10411):1411. https://doi.org/10.1016/S0140-6736(23)02292-4

Tschandl P, Rinner C, Apalla Z, Argenziano G, Codella N, Halpern A, Janda M, Lallas A, Longo C, Malvehy J, Paoli J, Puig S, Rosendahl C, Soyer HP, Zalaudek I, Kittler H (2020) Human–computer

collaboration for skin cancer recognition. Nat Med 26(8):1229–1234. https://doi.org/10.1038/s41591-020-0942-0

Tu T, Palepu A, Schaekermann M, Saab K, Freyberg J, Tanno R, Wang A, Li B, Amin M, Tomasev N, Azizi S, Singhal K, Cheng Y, Hou L, Webson A, Kulkarni K, Mahdavi SS, Semturs C, Gottweis J et al (2024) Towards conversational diagnostic AI. https://doi.org/10.48550/arXiv.2401.05654

Turkle S (2005) The second self: computers and the human spirit. MIT Press, Cambridge

Umbrello S, Natale S (2024) Reframing deception for human-centered AI. Int J Soc Robot 16:2223–2241. https://doi.org/10.1007/s12369-024-01184-4

Verghese A, Shah NH, Harrington RA (2018) What this computer needs is a physician: humanism and artificial intelligence. JAMA 319(1):19–20. https://doi.org/10.1001/jama.2017.19198

Walker L (2023) Belgian man dies by suicide following exchanges with chatbot. The Brussels Times. https://www.brusselstimes.com/belgium/430098/belgian-man-commits-suicide-following-exchanges-with-chatgpt

Wallach W, Allen C (2010) Moral machines: teaching robots right from wrong. Oxford University Press, Oxford

Warraich HJ, Califf RM, Krumholz HM (2018) The digital transformation of medicine can revitalize the patient-clinician relationship. Npj Digit Med 1(1):1–3. https://doi.org/10.1038/s41746-018-0060-2

Weil E (2023) You are not a parrot. Intelligencer. https://nymag.com/intelligencer/article/ai-artificial-intelligence-chatbots-emily-m-bender.html

Woo T, Ho R, Tang A, Tam W (2020) Global prevalence of burnout symptoms among nurses: a systematic review and meta-analysis. J Psychiatr Res 123:9–20. https://doi.org/10.1016/j.jpsychires.2019.12.015

Yin Y, Jia N, Wakslak CJ (2024) AI can help people feel heard, but an AI label diminishes this impact. Proc Natl Acad Sci 121(14):e2319112121. https://doi.org/10.1073/pnas.2319112121