Dimensions of Algorithmic Injustice in Medicine and Healthcare:

Issues and Future Perspectives

The increased introduction and use of AI systems in medicine and healthcare span various domains and contexts of applications. While the use of these systems is often met with optimism connected to their opportunity to improve healthcare delivery and optimize cost-efficiency and accuracy (Davenport & Kalakota, 2019), introducing these technologies in health environments is accompanied by concerns at the intersection between ethics and epistemology.

DR. GIORGIA POZZI
Researcher
Technology, Policy and Management
Delft University of Technology
E-mail: G.Pozzi@tudelft.nl

In this contribution, I will focus on problems raised by algorithmic applications in medicine and healthcare in terms of *justice*. Two clarifications are in order. First, regarding the context of the application, when referring to justice issues connected to AI in medicine, I refer to both clinical applications and AI-based applications that support individuals in health-related matters. Related to the former, many systems are being developed to enter clinical practices, and there are promising uses of AI systems, for example, in radiology and dermatology (Dembrower et al., 2023; Esteva et al., 2017). As for the latter, AI systems are also used to mediate sensitive decisions about pain management (Szalavitz, 2021; 2024). Other non-clinical applications are, for instance, related to the use of Large Language Models or other AI-assisted chatbots for mental health support (Tekin, 2023).

Second, I consider justice questions that pertain to the broader context of the application of an AI system. These issues do not exclusively pertain to the technological artifact per se but are also more fundamentally related to the social context in which they are implemented.

I will proceed as follows. In the section entitled 'Dimensions of injustice in medical AI' I will briefly review some of the dimensions of injustice available in the literature. Here, I highlight how the debate often focuses on issues connected to two related main domains: distributive justice issues (e.g., how AI systems lead to an unequal allocation of healthcare resources) and justice issues that emerge due to algorithmic unfairness and discriminatory patterns. However, as pointed out by Hull (2023), fairness-based critique often overfocuses on the technology itself but does not adequately address the social conditions under which an AI system operates. For this reason, in the next section, entitled 'Mechanisms of unrecognition: an epistemological-ethical issue' I complement with a discussion on a further domain of injustice pertaining to mechanisms of unrecognition, silencing, and marginalization that relate to an *epistemic* dimension of algorithmic injustice. This has received comparatively less attention but is nonetheless relevant in tackling forms of social injustice further perpetuated by AI systems in medicine and healthcare. As I will argue, the focus on epistemic injustice allows us to take a broader perspective and contextualize issues of unfairness and discrimination against the background of underlying power dynamics and social inequalities. In the section entitled 'Future perspectives and final remarks' I discuss future perspectives that need our attention to counter epistemic injustices that AI systems bring about.

Dimensions of injustice in medical AI

The use of AI systems in medicine and healthcare often leads to forms of injustice that have received extensive attention in the literature. Let us consider an often discussed case showing a paradigmatic form of injustice that can be found in Obermeyer et al. (2019).

These authors consider the use of an algorithmic system implemented in the United States with the aim of evaluating which patients qualify for additional healthcare support. Due to its goals, it is evident that the system is involved in crucial decisions that will impact patients' ability to access healthcare resources and its appropriate functioning is central to ensure fair and just outcomes. However, the system turned out to have several shortcomings. Crucially, it assesses patients' eligibility based on their past healthcare costs. Implicitly encoded in the system functioning is, thus, a definition of health that amounts to having had low previous medical expenses. Conversely, patients who have higher medical expenses are considered to need additional health support. While past healthcare costs seem to be, at first, a neutral and easily quantifiable metric, using this information as a proxy for eligibility for further medical assistance can lead to misguided and unjust decision-making. In fact, it erroneously remains unconsidered that due to inequalities in access to healthcare as a consequence of systemic racism, not every social group has the same opportunity to access healthcare services in the first place. As a matter of fact, Obermeyer and colleagues show that the system turned out to be heavily biased against Black people, as it erroneously systematically miscategorized these patients as being less in need of additional care compared to White people (Benjamin, 2019). By implicitly encoding a biased definition of health, equating to having low medical expenses, the system reproduces health inequalities and discriminatory patterns of care. Above and beyond the fact that this system brings about discriminatory outcomes, this case also highlights a further dimension of injustice pertaining to an unfair distribution of healthcare resources. And this is, unfortunately, not the only example. For instance, Simonite (2020) reports of an AI system exacerbating health disparity in kidney transplant allocation.

These cases show that AI systems in healthcare might lead to issues of injustice understood in *distributive* terms: when it comes to valuable social goods, like healthcare support and resources, AI systems can contribute to an unjust distribution that reinforces patterns of inequality (Gabriel, 2022). Other scholars have argued that the focus on justice in distributive terms does not leave enough space for considerations of justice that are *relational* in nature (Wegner et al., 2023). The latter refers to structural social mechanisms of oppression and power asymmetries that bring about issues beyond questions of fairness related to equity in distributing relevant goods. According to these authors, relevant questions about how power relations between decision makers and decision subjects play out in AI-mediated settings, who is disproportionately harmed, and how AI technologies shape the decision-making power require an approach beyond analysing these issues through a distributive injustice lens.

A similar call for a more structural approach to algorithmic (un)fairness and discrimination is also called for by Binns (2018). This author points out, for example, how certain forms of algorithmic discrimination can emerge because predictions about individuals are made given their membership in a social group that is considered relevant for the final prediction. Here, the harm often would not (exclusively) pertain to an unfair distribution of socially relevant goods but to the very fact that the individual is not considered in their particularities, but rather due to belonging to what is considered a relevant group. So, the harm amounts to people not being adequately acknowledged in their singularity.

There is a further dimension to injustice that has received comparatively less attention and is also related to structural mechanisms of inequality and power dynamics exacerbated by AI systems, specifically connected to their possibility to play a role as epistemic subjects. In the following, I address how AI systems as epistemic technologies bring about forms of epistemic injustice, a further dimension connected but not equal to the relational injustice just briefly addressed. I turn to this more specifically in the next section.

Mechanisms of unrecognition: an epistemological-ethical issue

As Alvarado (2023) argues, AI systems are epistemic technologies: they are used to manipulate information and produce relevant knowledge about individuals, among others. So understood, it is central to critically question to what extent AI systems can lead to forms of injustice that are specifically epistemic in nature. The concept of epistemic injustice has been coined by Miranda Fricker (2007) to refer to injustices related to unjustified credibility deficits connected to identity prejudices about individuals belonging to under-represented social groups (*testimonial injustice*) or to the inability to express one's situated lived experience and knowledge due to a lack of collectively available hermeneutical resources (*hermeneutical injustice*).

In previous work (Pozzi, 2023a, 2023b) I argued in more detail that both forms of epistemic injustice can emerge in AI in healthcare settings. However, due to the limited scope of this contribution, I focus exclusively on forms of testimonial injustice. These can emerge, for instance, using AI-based mental

health chatbots (increasingly Large Language Models) in socially relevant domains, which can lead to individuals' inability to have meaningful conversations with these systems and remain unrecognized in their testimonial offering. For example, chatbots are often used as mental health support but have shown considerable limitation when applied to under-served populations because their experiences do not match the syntax of the system that likely mirrors and codifies the lived experiences of dominant social groups (De Proost & Pozzi, 2023; Pozzi & De Proost, 2023). The difficulty in properly expressing oneself and being recognized as a conversation partner can lead to patterns of unrecognition in which the epistemic subject's role is unjustifiably degraded.

Issues of testimonial injustice might also pertain more generally to the unrecognition of the lived experience of patients that do not exclusively have to do with technical limitations intrinsic to the system, but rather to the role that these systems are allowed to play, attributing to them excessive decision-making power. Consider, for example, NarxCare, an AI-based platform used in the US to predict the risk of patients becoming addicted to opioid medication (Szalavitz, 2021; 2024). This case is paradigmatic of this particular form of injustice because pain management is often intrinsically difficult to objectively verify and bound to patients' subjective experience and perception in their singularity. In turn, this means that if the system is a-critically given more credibility than the patient's report of their experience of pain and disease, more generally, it might lead to forms of AI-induced testimonial injustice. Elsewhere, I argued that this happens because the system might decrease the value of a patient's testimony and report of pain since it is erroneously taken to objectively estimate patients' eligibility for pain medication (Pozzi, 2023). In these cases, the value of patients' testimony risks remaining unacknowledged, and they risk losing epistemic agency, thus being excluded from meaningful informational exchange. The latter is, however, central to the practice of shared decision-making, so AI systems risk bringing about a fall back to paternalistic modes of care (McDougall, 2019).

Let me further note that while the epistemic injustices I briefly sketched are often connected to forms of bias and discrimination, they should not be reduced to those. Epistemic injustice encompasses the broader idea that AI systems might support mechanisms of healthcare delivery that lead to the systematic unrecognition of the role patients can and should play as relevant epistemic subjects. This relates to the misguided idea that hard-to-quantify information from patients' testimonial reports can be substituted by referring to allegedly objective information *about* patients generated by AI systems. However, it is paramount to insist that the information coming directly *from* patients can be, on occasion, an irreplaceable source of medical information that needs to play a role in medical discourse. This is particularly the case in AI-supported pain management, as previously mentioned. Recognizing the value of patients' reports of symptoms and other circumstantial knowledge about their experience of illness as a way to access important clinical information requires rethinking the role of AI systems in healthcare to avoid further shifting epistemic power away from patients.

Future perspectives and final remarks

While there is no one-size-fits-all solution to epistemic injustice in healthcare arising through the deployment of AI systems, different approaches have been recently advanced in the literature to mitigate it. For instance, to alleviate credibility issues ensuing from using AI technologies in pain management, Katz et al. (2025) argue for the need for epistemic humility, particularly from agents in power. Epistemic humility requires that agents "internally examine and critique their own claims to cognitive authority in order to ensure that other perspectives are not only heard but incorporated into furthering understanding" (ibid., p. 7). According to these authors, this shift in AI-supported medical practice needs to take place by, among others, appropriately educating medical personnel to critically question and reflect upon their practice. This reflective practice should ideally become part of medical curricula, thus rendering sufficient information available to students regarding their work practices supported by new technologies. This is crucial to enabling critical reflection on how AI outcomes (should) flow into medical decision-making as but one piece of relevant information that needs to be further contextualized and not necessarily directly acted upon.

Along similar lines, there is a need to steer public discourse toward a more realistic description of what AI systems effectively can and cannot achieve. Often, there is an over-inflated and unjustified hype around AI systems that suggests a level of accuracy that often lacks appropriate contextualization. As Drogt et al. (2024) point out, claims of high accuracy and AI systems outperforming clinicians *tout court* are often misguided and not specific enough. A possible overestimation of the knowledge AI systems can effectively generate might lead to attributing to them a higher epistemic status than other relevant sources of information (e.g., patients' testimony). To counter

these issues, efforts to provide a realistic assessment of AI systems' limitations seem paramount to avoid attributing to them an unwarranted epistemic privilege.

Other authors have highlighted the potential of AI systems to alleviate forms of epistemic injustice. For example, Kay et al. (2024) propose harnessing the potential of Generative AI to create new hermeneutical resources, thus possibly ameliorating forms of hermeneutical ignorance. They claim that "(g)enerative AI can enable the creative exploration of new experiences by simulating them, and help articulate experiences which are otherwise ineffable." (p. 10) A similar proposal might be envisaged in medicine when it comes to patients that have difficulties to verbalize and render accessible to physicians their symptoms and possible associated suffering. Using Generative AI as these authors suggest might be a helpful tool to bring across what they would like to express, thus, if used responsibly, giving voice to experiences that would otherwise remain unacknowledged. However, we must still be mindful of not misrepresenting their experiences by giving more legitimacy to the AI's representation. Hence, this needs to be done in awareness of possible power dynamics that attribute to the AI too much power and credibility, and might again shift the balance away from patients.

Finally, it is worth pointing out the importance of appropriately designing the epistemic environment where clinicians operate when AI systems support medical decision-making. Particularly, it is paramount to ensure that clinicians do not overly rely on these systems to the extent that their critical scrutiny becomes compromised. Elsewhere, it has been argued that an unjustified over-reliance on AI systems can compromise clinicians' ability to be receptive to reasons for medical action that go beyond the AI system's output (Pozzi et al., 2025). The design of their epistemic environment by, for instance, securing a division of cognitive labour in which clinicians not only monitor the system but also come, in parallel, to independent conclusions seems one possible way to counter these issues (ibid.). All in all, more needs to be said about which epistemic conditions need to be in place to mitigate forms of epistemic injustice resulting in the illegitimate dismissal of patients' testimony and unrecognition of their lived experience. Thus, these issues ought to remain the object of future research in the context of epistemically and ethically responsible medical AI.

Literature

Alvarado R. AI as an Epistemic Technology. Science and Engineering Ethics 2023; 29(5). https://doi.org/10.1007/s11948-023-00451-3

Benjamin R. Assessing risk, automating racism. Science 2019; 366(6464), 421–422. https://doi.org/10.1126/science.aaz3873

Binns, R. Fairness in Machine Learning: Lessons from Political Philosophy. Proceedings of Machine Learning Research 2018; (Vol. 81). https://ssrn.com/abstract=3086546

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

De Proost M & Pozzi G. Conversational Artificial Intelligence and the Potential for Epistemic Injustice. The American Journal of Bioethics 2023; 23(5), 51–53. https://doi.org/10.1080/15265161.2023.2191020

Dembrower K, Crippa A, Colón E, Eklund M & Strand F. Artificial intelligence for breast cancer detection in screening mammography in Sweden: a prospective, population-based, paired-reader, non-inferiority study. The Lancet Digital Health 2023; 5(10), e703–e711. https://doi.org/10.1016/S2589-7500(23)00153-X

Drogt J et al. Ethical guidance for reporting and evaluating claims of AI outperforming human doctors. npj Digital Medicine 2024; (Vol. 7, Issue 1). Nature Research. https://doi.org/10.1038/s41746-024-01255-w

Durán JM & Pozzi G. Trust and Trustworthiness in AI. Philosophy & Technology 2025; 38(1), 16. https://doi.org/10.1007/s13347-025-00843-2

Esteva A et at. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017; 542(7639), 115–118. https://doi.org/10.1038/nature21056

Fricker M. Epistemic Injustice. Power & the Ethics of Knowing. Oxford University Press, 2007.

Gabriel I. Toward a Theory of Justice for Artificial Intelligence. Daedalus 2022; 151(2), 218–231. https://doi.org/10.1162/daed_a_01911

Hull G. Dirty data labeled dirt cheap: epistemic injustice in machine learning systems. Ethics and Information Technology 2023; 25(3). https://doi.org/10.1007/s10676-023-09712-y

Katz RA, Graham SS & Buchman DZ. The need for epistemic humility in AI-assisted pain assessment. Medicine, Health Care and Philosophy 2025. https://doi.org/10.1007/s11019-025-10264-9

Kay J, Kasirzadeh A & Mohamed S. Epistemic Injustice in Generative AI, 2024. http://arxiv.org/abs/2408.11441

McDougall RJ. Computer knows best? The need for value-flexibility in medical AI. Journal of Medical Ethics 2019; 45(3), 156–160. https://doi.org/10.1136/medethics-2018-105118

Obermeyer Z, Powers B, Vogeli C & Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019; 366(6464), 447–453. https://doi.org/10.1126/science.aax2342

Pozzi G. Automated opioid risk scores: a case for machine learning-induced epistemic injustice in healthcare. Ethics and Information Technology 2023a; 25(1), 3.

Pozzi G. Testimonial injustice in medical machine learning. Journal of Medical Ethics 2023b; 49, 536–540. https://doi.org/10.1136/jme-2022-108630

Pozzi G, Buijsman S & van den Hoven J. Harmful epistemic dependence on medical machine learning and its moral implications. Journal of Medical Ethics 2025. https://doi.org/10.1136/jme-2024-110552

Pozzi G & De Proost M. Keeping an AI on the mental health of vulnerable populations: reflections on the potential for participatory injustice. AI and Ethics 2024. https://doi.org/10.1007/s43681-024-00523-5

Simonite T. How an Algorithm Blocked Kidney Transplants to Black Patients. Wired 2020. https://www.wired.com/story/how-algorithm-blocked-kidney-transplants-black-patients/

Szalavitz M. The Pain Was Unbearable. So Why Did Doctors Turn Her Away? Wired 2021. https://www.wired.com/story/opioid-drug-addiction-algorithm-chronic-pain/
Szalavitz M. Say Hello to Your Addiction Risk Score-Courtesy of the Tech Industry. New York Times 2024. https://www.nytimes.com/2024/04/20/opinion/addiction-risk-score-avertd-narxcare.html
Tekin Ş. Ethical Issues Surrounding Artificial Intelligence Technologies in Mental Health: Psychotherapy Chatbots. In: Technology Ethics: A Philosophical Introduction and Readings (pp. 152–159). Routledge, 2023.
Wegner L, Decker MC & Leicht-Scholten C. Beyond Distributions: A Systematic Review on Relational Algorithmic Justice. 2023.

Because of medical AI, individuals can be wronged in their capacity to convey knowledge and receive appropriate consideration

Patients and vulnerable social groups risk being degraded in their status as epistemic subjects

ABSTRACT

In this contribution, I analyse different dimensions of algorithmic injustice. After considering forms of injustice understood in distributive and relational terms that often receive attention in the literature, I turn to considerations pertaining to forms of *epistemic* injustice arising in medical AI. Generally, the latter amounts to situations in which individuals are wronged in their capacity to convey knowledge and receive appropriate consideration in medical interactions. I then spell out how mechanisms of unrecognition materialize as an expression of this issue, in which patients and vulnerable social groups risk being degraded in their status as epistemic subjects. Finally, I reconstruct some approaches recently advanced in the literature to mitigate the previously considered issues and lay out some perspectives for future research.
Keywords: *Medical AI; epistemic injustice; ethics and epistemology of AI.*

Woltring N. De marktconforme verzorgingsstaat. Nederlands neoliberalisme in de lange jaren negentig. Amsterdam: Boom uitgevers, 2024. 376 blz. ISBN: 9789024468140. Prijs: € 29,90.

Historica Naomi Woltring heeft al een mooie staat van dienst. Ze was werkzaam bij de Wiardi Beckman Stichting en bij het centrum voor parlementaire geschiedenis onderzoekt ze het financieel-economisch beleid van de kabinetten-Lubbers. Ondertussen promoveerde ze op de geschiedenis van het neoliberalisme in Nederland, waar een handelseditie van is uitgegeven. In dit boek ontrafelt ze de interacties tussen politieke ideologieën, beleidsvorming en de dagelijkse praktijk van de verzorgingsstaat.
In de inleiding geeft Woltring aan wat ze onder neoliberalisme verstaat, namelijk een set van ideeën over de inrichtring van de samenleving met bijbehorende sturingsfilosofie over effectief en goed bestuur, hetgeen zichtbaar wordt in beleidspraktijken. De beleidsuitkomsten en de sociale, economische, culturele, institutionele gevolgen, zoals de verdeling van kennis, inkomen en macht, van levenskansen en de maatschappelijke gevolgen daar weer van op de langere termijn, worden dan een niet uit te vlakken factor binnen het neoliberalisme.
Ze werkt dit uit aan de hand van drie onderwerpen: marktconform beleid, volkshuisvesting en woningmarkt en sociale zekerheid. De inhoud is vrij taai, maar eens je mee bent zie je de hele evolutie van het neoliberaal beleid in Nederland aan je voorbij gaan. Het is ons niet overkomen, het is opgelegd en gestuurd door de overheid. Het lijkt misschien verbazingwekkend dat de uitwassen niet op tijd konden worden bijgestuurd, maar als je dit boek leest leer je dat veel radertjes in het neoliberalisme in elkaar grijpen. Lees; dezelfde mensen die andere petjes opzetten op de juiste momenten om het politiek idee van marktwerking te laten doorgaan. De marktwerking binnen de gezondheidszorg bespreekt ze niet expliciet, maar wat ze beschrijft is daar ook zeker van toepassing.
Nu we dankzij dit belangrijke werk op de hoogte zijn hoe marktwerking ons dagelijks bestaan beheerst, is het een grote uitdaging om een alternatieve weg te ontwikkelen.

Dr. Karin Janssen van Doorn, klinisch adviseur bij het Belgische Federale Agentschap voor Geneesmiddelen en Gezondheidsproducten, lid van de 'Scientific Advice Working Party' van het Agentschap voor geneesmiddelen van de Europese Unie.