Philosophy of Medicine
Analysis

Privilege Under Review: The Evidence Challenge for Epistemic Injustice in Health AI

Patrik Hummel[1]

1 Department of Philosophy, University of Lucerne, Lucerne, Switzerland; Zug Institute for Blockchain Research, University of Lucerne, Lucerne, Switzerland. Email: patrik.hummel@unilu.ch

Abstract

Increasingly, scholars warn that artificial intelligence (AI) may replicate or intensify epistemic injustice (EI) in biomedical contexts. This paper argues that such concerns require demonstrable evidence on (i) the actual presence of EI, and (ii) whether EI, if present, makes system effects normatively undesirable. Without evidence on (i), the EI framework is, by its own lights, inapplicable. Evidence on (ii) is needed because reflective equilibrium can, in principle, prioritize competing considerations over epistemic justice. A key challenge is that discourses on (i) and (ii) can themselves be sites of EI. Two steps are proposed: distinguishing AI's effects on problem constructions from its effects on abilities to address constructed problems, and foregrounding relative accuracy and real-world impact on the quality of care in assessments of AI-induced EI.

1. Introduction

For a number of years, cross-disciplinary efforts have shown how artificial intelligence (AI) can replicate, perpetuate, and exacerbate injustice in society (Eubanks 2018; Noble 2018). Increasingly, it is argued that this includes epistemic injustice (EI) (El Kassar 2022; Milano and Prunkl 2025). As the AI buzz has grown in healthcare, in particular, there are concerns that AI could lead to new forms of EI in health (Pozzi 2023a, 2023b; Ho 2023; Hardalupas 2024; Herzog and Branford 2025). For example, Giorgia Pozzi (2023a) argues that NarxCare and similar prescription drug monitoring programs (PDMPs) for risk assessment in pain medication administration could suppress and preempt perspectives in patient–clinician encounters.

The following commentary raises two issues for applying EI to health AI. Both test the notion's analytical and normative weight. First, charges of EI, when not based on demonstrable evidence, can be self-defeating. Without such support for certain epistemic assumptions around AI-affected health contexts, there is reason to think that no EI occurs in the first place. Cautioning against EI is important. It thus should be done in a way that facilitates demonstrable support. Second, extant discussions of EI in health assume that EI

ISSN 2692-3963 | Date Received 06-03-2026 | Date Accepted 23-07-2026 | Date First Published 10-08-2026
Philosophy of Medicine | DOI 10.5195/pom.2026.289 | Volume 7 | Issue 1 | pp. 1–15

is a dealbreaker and in tension with AI frameworks such as meaningful human control or value alignment. This assumption, however, is problematic. Because of both issues, we should characterize more systematically how exactly EI can be ascertained and how it should be rectified.

2. Epistemic Injustice in Health

Discussions of EI typically refer to Miranda Fricker's understanding as "a wrong done to someone specifically in their capacity as a knower" (Fricker 2007, 1). Rachel McKinnon calls this tendency a "deep irony" (2016, 439) in view of similar, long-standing themes in Black feminist thought (Crenshaw 1991; Hill Collins 1990; hooks 1992; Alcoff 1999). In Fricker's framework, EI can be testimonial or hermeneutical. A distinctive wrong of testimonial EI is *epistemic objectification* (Fricker 2007, chapter 6): individuals are treated not as informants (Craig 1991); that is, epistemic agents conveying information, but as mere sources of information, mere "states of affairs from which the inquirer may be in a position to glean information" (Fricker 2007, 132). This is an injustice insofar as it involves *prejudices:* "judgements, which may have a positive or a negative valence, and which display some (typically, epistemically culpable) resistance to counterevidence owing to some affective investment on the part of the subject" (2007, 35; italics removed). Most relevant to Fricker's discussion are *identity* prejudices: those "held against people *qua* social type." In hermeneutical EI, a distinctive wrong is hermeneutic marginalization: "unequal hermeneutical participation with respect to some significant area(s) of social experience" (2007, 153).

EI has been highlighted as a challenge in health contexts (Carel and Kidd 2014; Kidd and Carel 2017). For example, prejudices can lead to dismissals of patient testimony because of

> stereotypical description of ill persons as cognitively impaired or emotionally compromised … or as existentially unstable, gripped by anxieties about mortality and morbidity such that they "cannot think straight"; or that they will be psychologically dominated by their illness in a way that warps their capacity to accurately describe and report their experiences. (Kidd and Carel 2017, 177–178)

EI need not result solely from individual decision-making but can also be driven by "structural and hierarchical features of the healthcare system" (Carel and Kidd 2014, 535), such as emphases on cost minimization and efficiency and/or more general societal inequities.

3. AI in Health: What's the Point?

The promise or vision around health AI is that it can enhance efficiency, safety, access to health, personalized care, and the time and attention clinicians can devote to patients. Reality checks of such claims are constantly needed (Chen and Asch 2017; Topol 2019; Goldenberg and Wiens 2026) as most current health AI might be deeply ridden with hype (Emanuel and Wachter 2019; Kulkarni and Singh 2023; Strange 2024). In the foreseeable future, AI's potential could be limited to specific, narrowly conceived tasks (Rampton 2020;

Nagendran et al. 2020), though some speculate that agentic AI and multi-agent systems are on the verge of altering this paradigm (Taylor 2025). Without begging too many questions, one can characterize the intended purposes of AI as pertaining in some way to augmentation and enhancement or, in the limiting case, replacement (Goldhahn et al. 2018) of human capacities and resources. The point is to uncover information one did not have before, to draw previously inaccessible inferences, to highlight trajectories human decision-makers might not have spotted or foreseen, and to fulfill tasks faster or otherwise more efficiently than human decision-makers could do on their own.

These are broad brushstrokes for several reasons. A socio-technical system perspective cautions against simplistic, reductive understanding of how AI models function across diverse healthcare settings (Wiesenfeld et al. 2022). The broad brushstrokes are agnostic on how regularly the aspiration of augmentation is actually achieved. Moreover, they appear to presume the existence of some sort of ground truth that AI helps humans to discern more effectively (for example, the presence of disease, risk levels to progress to disease, or most effective clinical course of action). Uncritical notions of ground truth can neglect the human activity involved in conceptualizing disease, risk thresholds, endpoints, available options, and other key issues impacted by interests, convention, and pragmatics (Starke et al. 2021). However, reminders about the significance of (co-)construction are compatible with a further point sometimes lost in reactions against AI hype: *once* ground truths and categories are constructed, there is nothing particularly mysterious or necessarily unwarranted about using an AI tool to try to track them.

4. The Issue of Evidence for Epistemic Injustice

If EI affects pre-AI healthcare, for example, by stakeholders being "too often prejudicial" (Kidd and Carel 2017, 175) against ill persons, health AI could replicate or even aggravate EI. This supposition very much makes sense and is articulated forcefully by a number of authors (Pozzi 2023a, 2023b; Ho 2023; Hardalupas 2024; Herzog and Branford 2025). Clearly, biases exist in societies against members of certain groups, who are also patients. There is no reason to think that medical stakeholders or institutions are immune to such biases. Sex, gender, race, and other features affect biomedical research, care, and public health. AI could further contribute to this (Cirillo et al. 2020; Haider et al. 2026).

This suggests that EI likely happens in medicine. However, it does not speak to the important questions of when and how exactly it happens. It is incumbent on us to assess this. An uncomfortable truth is that EI's de facto pervasiveness, frequency, and mechanisms, as well as its impact in health remain understudied. For the sake of simplicity, let us focus on testimonial EI. Health AI could, in principle, cause clinicians to discredit patient testimony. Likewise, it could lead other stakeholders to discredit clinician perspectives vis-à-vis AI outputs. Conceptually, Fricker-style testimonial EI is based on prejudice. Part of determining whether there is EI is thus to determine what people used as a basis for credibility attributions, belief formation, and subsequent action. Prejudice must be shown to be what caused them. Here, a first observation is that, somewhat ironically, there exists a tendency to just presume that prejudice was or will be at work; systematic investigation of the extent to which prejudice is really what leads to assessments of testimony and subsequent action is lacking (Kious et al. 2023; Nielsen et al. 2025). Analogous remarks apply to EI approaches operating with constitutive features other than

prejudice, such as Elizabeth Anderson's institutional EI (2012) or Kristie Dotson's patterns of exclusion (2014).

In contrast, a more well-researched tendency is automation bias in clinicians using AI decision-support systems (for example, Khera et al. 2023). Note, however, that for such results to indicate EI, biases would have to result from prejudice; that is, attitudes that are identity-based, affective, and resistant to counterevidence, or other constitutive features of EI. Without such demonstration, discrediting patient testimony based on automation bias does not clearly constitute EI (though of course it might still be epistemically problematic for different, prejudice-independent reasons).

This highlights a predicament of navigating terrain rife with complexity and uncertainty: hypotheses can sound plausible, and yet whether and at which frequency they obtain can be an open question. Sensible suppositions are the subject of debate and contestation. Some argue that most EI concerns in health are "unfounded" and "too much emphasis on it could entail serious harm to patients, providers, and their relationships" (Kious et al. 2023, 1). This is not to take a stance in this debate, only to note that progress requires a shared understanding and ascertainment of what would count as sound evidence on these matters. A lack of such evidence introduces uncertainty about the supposition of EI itself.

5. The Possibility of Evidence-Attuned Credibility Deficits

The foregoing highlights a more general challenge that is overlooked surprisingly often: a mere tendency to discount the testimony of certain stakeholders is not sufficient for EI. One reason just mentioned is that discounting can have drivers other than prejudice. A special case of this is when such discounting actually signifies attunement, rather than prejudicial resistance to evidence.

Prejudice-driven discrediting has profound epistemic defects. The discredited perspective is epistemically relevant yet ignored. Such deficiency is obvious in exemplary instances of EI, such as when "simply stating the case [is] not enough" (hooks 1992, 80) for the uptake of credible testimony, including when the discrediting of pain reports (Hoffman et al. 2016; Loh 2026) or witness statements (Fricker 2013, 1325) is implausibly identity-based. But it is not the case that discrediting *must* be resistant to counterevidence. In the absence of such resistance, the case for discrediting being unjust is much less clear. This point can be made in several ways from within the EI framework itself.

As one example, David Coady (2010) distinguishes two concepts of EI: first, *distributive* injustice in the allocation of epistemic goods, such as education; second, *discriminatory* injustice as unjust deficits in credibility or intelligibility. He observes what he calls a "tension" between these concepts:

> To the extent that the former kind … is a problem, the latter kind … seems to be less of a problem, and perhaps less of an injustice. After all, if the relatively powerless group in question is ignorant or in error about the question at hand (however unjust it may be that they are), then it is surely just as well if they are not given much credibility, since they are likely to be wrong if indeed they have an opinion. (Coady 2010, 110–111)

Coady sketches out ways of handling this tension. His preferred strategy is to separate and elucidate propositions relative to which victims are unjustly ignorant or in error, on the one hand, from propositions relative to which these victims are unjustly disbelieved or misunderstood, on the other. What matters for our purposes is that asymmetries in the distribution of epistemic goods can undermine the case for reductions in credibility being deficient and/or unjust.

As another example, Morten Fibieger Byskov argues that a necessary condition for EI is what he calls the *epistemic condition:* "In order for someone to be discriminated against as a knower they must possess relevant knowledge about the subject matter being discussed. If this were not the case, we would hesitate to classify it as an *epistemic* injustice" (Byskov 2021, 125). Having a stake in a decision is not the same as being entitled to contribute epistemically: "Passengers on a plane, for example, have a deep interest in the plane being flown safely, yet, that does not mean that they should have a say in how to fly the plane because most passengers simply have little qualification of how to fly a plane." It is not EI to discount testimony that does not represent relevant knowledge.

Likewise, it is built into the notion of *credibility deficit* that it applies "when we attribute less credibility than a speaker deserves" (McKinnon 2016, 438). This suggestion does not reject but endorses that there are proper levels of credibility for certain speakers. Even those who caution against EI in health suggest as much. Ian James Kidd and Havi Carel (2017, 175) clarify that the claim cannot be that "all ill persons are always epistemically reliable, for that is clearly not the case." Instead, "judgments about the epistemic credibility of ill persons are too often prejudicial and generated and sustained by negative stereotypes and structural features of healthcare practice." Paul Crichton et al. note that a person "may lack credibility for a good reason, for instance if she is a known liar" but highlight that in EI, "this credibility deficit is unjustified and hence constitutes an epistemic injustice" (Crichton et al. 2017, 66). Awais Aftab highlights that there is no tension between epistemic justice and the conviction that "beliefs should be attributed the credence that is merited" (Aftab 2023, 7978). Pozzi specifically considers "unjustified" (2023b, 537) withdrawal of credibility, implicitly recognizing the possibility of justified withdrawal.

Extant commentaries do not consider that the challenge of determining proper levels of credibility is even more pronounced in connection with health AI and big data. As already mentioned, the purpose of such applications is to kick in where human abilities are exceeded, to leverage their epistemic edge over human reasoners when designed, implemented, and validated in the right ways for specific-use cases. Once the prospect of meeting such milestones of validation becomes a salient possibility, health AI could introduce epistemic and, ultimately, clinical benefits. If and when this is the case, it becomes less clear that instances of discounting the perspectives of particular clinical stakeholders constitute EI. At the very least, presumption alone would not suffice to establish this. Moreover, if such presumption leads to nonuse of AI, this could impede valuable and perfectly attainable improvements to service delivery and research. Presumption alone also does not empower anyone to intervene and redesign systems in targeted ways.

There is no straightforward formula for specifying key notions in the above, such as "relevant" or "justified," and applying "justice" and "injustice" on that basis. AI ethics has long been grappling with the fact that notions like these need context and negotiation for their application (Whittlestone et al. 2019; Hummel 2025). The discounting of certain testimony as well as the composition of (and even gaps in) hermeneutical resources are not

unjust per se but need to be assessed and classified as such. The framework of EI does not preempt this assessment and negotiation process.

These points are not criticisms of the EI framework, its content, or its importance. Conjectures that AI will propel EI yield useful prompts to investigate this possibility. The point is that staying true to the framework itself requires at least two things: first, that credibility is not assigned and withheld through mere supposition alone but approximates *proper* credibility assignments. Second, that misalignments with proper credibility assignments cannot merely be presumed to exist but need to be *shown* to be objectionable and unjustified.

6. Do Control and Alignment Require Epistemic Justice?

Various recent guidance documents on the ethics and governance of AI could inform how we navigate the risk of AI-induced EI in health. Paradigmatically, the World Health Organization (WHO) seeks that in the face of AI, "humans should remain in full control of health-care systems and medical decisions" (WHO 2021, vi).

Two ecumenical AI ethics frameworks for operationalizing this aspiration are meaningful human control (MHC) and value alignment (VA). MHC over automated systems requires that the system tracks reasons of human decision-makers, and that outcomes are traceable to human decisions and actions (Santoni de Sio and van den Hoven 2018; Santoni de Sio and Mecacci 2021; Santoni de Sio 2024). VA requires overlap between what humans value and what systems do (Gabriel 2020).

When applying these frameworks to health AI, it is tempting to think that EI is not among the set of human reasons (Santoni de Sio 2024, 179; Pozzi and Santoni de Sio 2026) or values (Kay et al. 2024). Thus, MHC and VA for health AI, when pursued properly, preclude EI. But the issue is more complicated.

The orthodox view is that relevant norms in biomedicine encompass autonomy, non-maleficence, beneficence, and justice (Beauchamp and Childress 2013). These and other inputs into moral deliberation need to be specified, balanced, and made coherent via reflective equilibrium: a state "in which equilibrium occurs after assessment of the strengths and weaknesses of the full body of all relevant and impartially formulated judgments, principles, theories, and facts" (2013, 406).

Crucially for the present purposes, justice is not the only applicable principle to be considered in reflective equilibrium, nor does it enjoy lexical priority over other principles of importance. If MHC and VA require approximation of human values, these frameworks do not necessarily disqualify EI as inappropriate. Instead, reflective equilibrium requires, among other things, context-specific judgments about how principles outweigh one another if and when they conflict. Under certain circumstances, reflective equilibrium might determine that beneficence outweighs respect for autonomy; for example, when administering an intervention with proven benefits even against the will of the patient. Other times, autonomy can outweigh non-maleficence; for example, in competent refusals of chemotherapy in end-of-life care. Likewise for considerations of justice, including EI.

Responding to Brent M. Kious and colleagues' concerns about the utility of the concept of EI in psychiatry (2023), Aftab clarifies that epistemic justice "is not something that is *outside* of good clinical care. Good clinical care is inclusive of our best ethical practices" (2023, 7978), including epistemic justice. Similarly, Lisa Bortolotti highlights EI as an

"obstacle to successful clinical interactions" and "bad for medicine" (2025, 2, 13). The above picture captures common ground: discussants are after good care. It also enables us to observe a twofold disagreement: first, about the extent to which EI de facto obstructs good care; second, how different normative goods within good care are to be weighed.

It is very plausible to call for MHC and VA, but what the foregoing shows is that it is a substantive, potentially vexed further question to figure out *how* to control and *what* to align with (Hille et al. 2023). Pozzi is concerned that AI could lead to "automated hermeneutical appropriation" (2023a, 9): it can, in principle, create meanings, understandings, and valuations while disfiguring and superseding those of human agents. This important insight notwithstanding, we should also consider it hermeneutically appropriating to privilege EI over all other, potentially competing concerns.

This picture can become more complicated upon further inspection. For instance, one might assume there is a connection between EI and violations of non-maleficence, such that EI causes epistemic and/or non-epistemic harms (Della Croce 2023). This would mean EI is more costly than it might have initially appeared. By itself, however, this insight is silent on whether and how relevant aspects of justice and non-maleficence outweigh other considerations.

On a bold reading of this suggestion, while EI might be pro tanto wrong, all-things-considered judgments could still permit it. The idea is not only that MHC and VA are each compatible with EI; that is, that in a given situation, other "human reasons" or "values" judged to be weightier would be undermined by pursuits of epistemic justice. Instead, the idea is also that reflective equilibrium could, in principle, *justify* a degree of EI if and when the overall set of moral commitments judged most coherent in that situation support alternative priorities. Either way, the connection between MHC and VA, on the one hand, and epistemic injustice, on the other, is contingent, not inherent or conceptual. Neither framework privileges epistemic justice categorically. This should not, however, be considered a shortcoming, quite the opposite; it would be dubious if the frameworks would curtail rather than catalyze the process of arriving at reflective equilibrium.

7. AI-Induced Epistemic Injustice in Medical Problem Construction

How should we proceed? A gap in extant discussions of EI through the use of health AI is the systematic consideration of accuracy and real-world impact. We should distinguish more consistently between two different areas in which EI can affect healthcare. First, the process of constructing (Dewey 1938; Conrad 2007; Siffels and Sharon 2024) a given medical problem: deciding what matters, conceiving of criteria for disease, prioritizing between different targets, choosing endpoints we optimize for, and forming reflective equilibrium about how to harmonize case judgments and more principled considerations. In these processes, there is potential for the exclusion or non-consideration of views from certain individuals or groups, who are not taken seriously as knowers and have no way of contributing to the hermeneutical repertoire. Indeed, the availability of tools can impact how human agents conceptualize a given problem. In particular, AI can interfere with the problem construction and the formation of reflective equilibrium: a PDMP might have the effect of reducing prescription volumes and, suddenly, in line with Pozzi's concern, minimizing prescriptions becomes the target of the practice.

Shifts like this can happen, and it is important to be aware of such possibilities. At the same time, there is no reason for fatalism. Our problem constructions are not necessarily doomed if AI is used to tackle them. We are not oblivious to AI-induced shifts but can discern them and intervene. While challenges are real, there is no need to collapse from critical, reflective distance into "negativity bias" (Königs 2025) that focuses exclusively on potentially exaggerated AI-induced threats to how we understand and approach medical challenges.

Authors cautioning against EI worry about the neglect of patients' lived experiences as legitimate sources of knowledge (Kidd and Carel 2017; Pozzi 2023b). However, it is not objectionable per se that particular problems are constructed in ways that set aside lived experiences. Rather, the trouble is that not all healthcare challenges are of this kind: sometimes lived experiences are indispensable to reconfigure the problem construction for the better. The key question we need to ask when taking the threat of EI seriously is: in a given instance, is the problem construction still contestable and revisable, or does AI calcify (Hardalupas 2024) that process? If we think it does, how does it have that effect, and how can such dynamics be made transparent in order to facilitate responsive action?

One complication is that the mentioned shifts can be subtle, influenced by institutional and commercial actors with differential power (Pozzi and Santoni de Sio 2026) over development and deployment decisions, and thus somewhat removed from the access of individual decision-makers whose reflective and critical intervention is key. This concern can (and should) be considered in any process of medical problem construction. Vigilance is needed to recognize when and how AI interacts with differential power and introduces additional, distinctive challenges to these processes.

Contrast this with a second area of interest: with the problem construction in place, is the tool under consideration helpful for addressing the problem? This is an important juncture in the assessment. Does AI use yield accuracy gains or losses? Does it positively impact real-world care (Goldenberg and Wiens 2026)? To assess whether the use of health AI is epistemically problematic and/or unjust, one relevant consideration is how its outputs and their uptake compare with the perspective they would replace. For instance, when assessing addiction risk, one possibility is that unaided clinician and patient (self-) assessments are less reliable guides than AI-driven risk scores. The key question is whether the latter are accurate and lead to better care.

This is, in fact, the fundamental concern with NarxCare. Commentators note that it "lacks sufficient evidence for its wide clinical implementation" (Buonora et al. 2024, 859), produces "inflated scores for marginalized patients" (Oliva 2022, 52), and is in need of independent external validation (McElfresh et al. 2023). In a validation study assessing a NarxCare component risk metric for opioid overdose against a misuse-risk reference standard (WHO ASSIST, the Alcohol, Smoking and Substance Involvement Screening Test), patients over-flagged by the metric were more likely to report disability, poor general health, and greater pain severity (Cochran et al. 2021). In view of such accuracy issues, especially if distributed inequitably, it is quite clear that the epistemic condition of EI is satisfied: given how PDMPs are de facto mandatory to use, clinical stakeholders are discredited despite possessing knowledge that, relative to NarxCare's outputs, must count as epistemically relevant. There is EI, resulting from issues with relative accuracy. As suggested above, the ascertainment of a deviation of proper levels of credibility attributions is what drives this judgment.

8. Epistemic Authority in Tackling Problems Once Constructed

Suppose this was different. Imagine a case in which at this second stage (where the problem is constructed, and we now want to address it), AI has high relative accuracy compared to human agents' judgment; for example, about subjective risk for addiction, so accurate that it leads us to ignore patients and clinicians in their capacity as knowers. What this illustrates is that the concern that AI enjoys unwarranted epistemic privilege (McCradden et al. 2023; Pozzi 2023a) cuts both ways: unwarranted epistemic privileges can also be conferred upon human decision-makers, who could be better off if health AI was leveraged successfully to augment care. Ignoring testimony could mean two things. Either, we avoid EI because the epistemic condition is undercut: in light of the high accuracy, clinical stakeholders do not possess relevant knowledge on the particular issue under consideration (among other things, *not* discrediting in these cases could even give rise to rather than counteract EI if it perpetuates stereotypes about that stakeholder's [un]reliability on relevant matters). Or there is EI, but as highlighted above, a debate still needs to be had about whether human reasons are tracked, nevertheless, and whether values are sufficiently aligned.

There are risks with labeling a practice an instance of EI. Application of the label could be misleading, contentious, false, or unwarranted. This can damage the practical utility of the label. It complicates the handling of cases that clearly are EI. Precisely because healthcare tackles highly complex subject matter and medical problem construction is a social endeavor, assertions about epistemic privilege cannot always be taken for granted. Often, it must be articulated what that privilege consists of. In order to properly flag EI in connection with AI, the proactive generation of evidence about relative epistemic authority is indispensable. Only then can we understand and characterize the epistemic privilege of EI-discredited perspectives for the relevant problem construction and also recognize any epistemic authority the AI might deserve upon rigorous validation.

Again, such authority likely pertains to relatively specific issues and narrowly confined tasks; for example, accurate and validated predictions of addiction risk operationalized in a particular way. Moreover, authority can be withdrawn if stakeholder views and circumstances change in ways that shift the problem construction. As such, countenancing AI epistemic authority need not lead to a clinical culture that privileges AI per se, that neglects stakeholders as active informants, or inhibits epistemic agency in other ways. Deference to AI epistemic authority must, of course, be based on good evidence, in the same way as deference to non-AI health tools such as medical guidelines, diagnostic tests, therapeutic interventions, and so on. If so, such deference need not have a detrimental effect on the epistemic agency of clinical stakeholders, but quite the opposite: it would result from enactments of their epistemic vigilance.

Several deep challenges that health AI poses anew in the quest for evidence about EI in health must be acknowledged. It is not set in stone what counts as evidence and for which hypotheses we should even search for evidence. An important risk is that a given conceptualization of evidence is insensitive to distinctive standpoints (Alcoff 1999) to which different clinical stakeholders have privileged access and that matter to a given problem construction. For example, for diseases like fibromyalgia, it is partly constitutive that complaints persist in the absence of signals from established testing methods, inviting an "antagonism between science and … individual experience" (Heggen and Berg 2021). Cases

like this illustrate that discourses on best evidence, its sources, and whether we have good evidence for EI, can themselves be sites of EI. It could be an instance of EI to insist on evidence for EI when the notion of evidence is not flexible and multidimensional enough to capture diverse, potentially marginalized standpoints. Moreover, gathering evidence for EI requires awareness of the phenomenon to test for. Such awareness can be suppressed by EI itself. A current notion of evidence can capture diverse, external, or hybrid phenomena only in familiar terms that are accessible to established bodies of theorizing, which can lead to distortions and imperfections in access and understanding.

The problem with these challenges is not just their depth but that there is no way around tackling them. It is possible, even if difficult, to purposefully bend established understandings of evidence, but it is impossible to operate without evidence. Evidence is needed, but as users we ought to be aware of its provisionality and handle it with care: evidence is continuously shaped by the translational work of rendering unexplored perspectives intelligible for consideration while maintaining openness to even radical otherness.

Finally, one should not lose sight of opportunities of health AI for addressing rather than perpetuating EI. Accurate, validated AI could, in principle, counteract prejudices of clinical stakeholders (Lalumera 2025, 147–151). It could trace patient well-being in ways that are resistant to preemptive silencing (that is, structural obstacles to the articulation of perspectives) or to testimonial smothering (that is, patients' self-suppression of their testimony due to EI). For rare or stigmatized diseases for which evidence generation is challenging and/or hermeneutical resources are lacking, AI could be instrumental in gathering data and inferring correlations. For testimony whose epistemic status appears initially unclear to human observers, AI could confer credibility and reinforce rather than undermine the epistemic standing of voices offering distinctive, underrepresented perspectives. Policies generated by reflective equilibrium could be encoded into "algorithmic bioethics" (Savulescu 2025), yielding consistent, codified application of stable value judgments. In each of these examples, the representative, surrogate role taken on by AI is not unproblematic. It could indeed be a symptom of EI that AI is needed to make certain voices heard. But this is not a reason to refrain from initiating improvements, and to try to search for and leverage the tools we deem most useful to advance that process.

9. Conclusion

This paper raises two issues for considering EI in health AI. First, credibility deficits and hermeneutical gaps are not unjust per se. Credibility deficits can, in principle, indicate attunement rather than resistance to evidence. This possibility needs to be examined more systematically in complex domains such as health, particularly for AI tools designed to augment and enhance human capacities. Taking the EI framework seriously involves attending to relevant empirical evidence for its application. In the absence of such evidence, the framework's own application conditions are undermined. Yet, the pervasiveness of EI in health and its relation to AI remain understudied. Second, frameworks like MHC and VA do not privilege epistemic justice categorically. Reflective equilibrium can, in principle, prioritize competing considerations. To make these issues tractable, this paper proposes a distinction between AI's effects on the construction of medical problems and its effects on our abilities to address problems once constructed, with relative accuracy and real-world

impact on care as key considerations at the second stage. A key challenge at each stage is that discourses about evidence for EI can themselves become sites of EI. Future work should explore where evidential standards for these stages remain least developed, operationalize what evidence for EI could look like, and measure how AI affects it.

Acknowledgments
For feedback on earlier versions of this material, I am grateful to two anonymous reviewers for *Philosophy of Medicine*; to participants of the Bridging Justice and Meaningful Human Control in Medical AI symposium at IACAP-AISB 2025 and the Justice, Power, and Control in Medical AI symposium at DMM&L 2025; and to my colleagues in the Department of Philosophy at the University of Lucerne.

Disclosure Statement
No competing interest was reported by the author.

References
Aftab, Awais. 2023. "Epistemic Justice Is an Essential Component of Good Psychiatric Care." *Psychological Medicine* 53, no. 16: 7978–7979. https://doi.org/10.1017/s0033291723001113.

Alcoff, Linda Martin. 1999. "On Judging Epistemic Credibility: Is Social Identity Relevant?" *Philosophic Exchange* 29, no. 1: 73–93. http://hdl.handle.net/20.500.12648/3169.

Anderson, Elizabeth. 2012. "Epistemic Justice as a Virtue of Social Institutions." *Social Epistemology* 26, no. 2: 163–173. https://doi.org/10.1080/02691728.2011.652211.

Beauchamp, Tom L., and James F. Childress. 2013. *Principles of Biomedical Ethics*. 7th edition. Oxford University Press.

Bortolotti, Lisa. 2025. "Agential Epistemic Injustice in Clinical Interactions Is Bad for Medicine." *Philosophy of Medicine* 6, no. 1: 1–19. https://doi.org/10.5195/pom.2025.222.

Buonora, Michele J., Sydney A. Axson, Shawn M. Cohen, and William C. Becker. 2024. "Paths Forward for Clinicians Amidst the Rise of Unregulated Clinical Decision Support Software: Our Perspective on NarxCare." *Journal of General Internal Medicine* 39, no. 5: 858–862. https://doi.org/10.1007/s11606-023-08528-2.

Byskov, Morten Fibieger. 2021. "What Makes Epistemic Injustice an 'Injustice'?" *Journal of Social Philosophy* 52, no. 1: 114–131. https://doi.org/10.1111/josp.12348.

Carel, Havi, and Ian James Kidd. 2014. "Epistemic Injustice in Healthcare: A Philosophical Analysis." *Medicine, Health Care and Philosophy* 17, no. 4: 529–540. https://doi.org/10.1007/s11019-014-9560-2.

Chen, Jonathan H., and Steven M. Asch. 2017. "Machine Learning and Prediction in Medicine—Beyond the Peak of Inflated Expectations." *New England Journal of Medicine* 376, no. 26: 2507–2509. https://doi.org/10.1056/NEJMp1702071.

Cirillo, Davide, Silvina Catuara-Solarz, Czuee Morey, et al. 2020. "Sex and Gender Differences and Biases in Artificial Intelligence for Biomedicine and Healthcare." *NPJ Digital Medicine* 3, no. 1, 81. https://doi.org/10.1038/s41746-020-0288-5.

Coady, David. 2010. "Two Concepts of Epistemic Injustice." *Episteme* 7, no. 2: 101–113. https://doi.org/10.3366/epi.2010.0001.

Cochran, Gerald, Jennifer Brown, Ziji Yu, et al. 2021. "Validation and Threshold Identification of a Prescription Drug Monitoring Program Clinical Opioid Risk Metric with the WHO Alcohol, Smoking, and Substance Involvement Screening Test." *Drug and Alcohol Dependence* 228 (November), 109067. https://doi.org/10.1016/j.drugalcdep.2021.109067.

Conrad, Peter. 2007. *The Medicalization of Society: On the Transformation of Human Conditions into Treatable Disorders*. Johns Hopkins University Press.

Craig, Edward. 1991. *Knowledge and the State of Nature: An Essay in Conceptual Synthesis*. Clarendon Press.

Crenshaw, Kimberle. 1991. "Mapping the Margins: Intersectionality, Identity Politics, and Violence Against Women of Color." *Stanford Law Review* 43, no. 6: 1241–1299. https://doi.org/10.2307/1229039.

Crichton, Paul, Havi Carel, and Ian James Kidd. 2017. "Epistemic Injustice in Psychiatry." *BJPsych Bulletin* 41, no. 2: 65–70. https://doi.org/10.1192/pb.bp.115.050682.

Della Croce, Yoann. 2023. "Epistemic Injustice and Nonmaleficence." *Journal of Bioethical Inquiry* 20, no. 3: 447–456. https://doi.org/10.1007/s11673-023-10273-4.

Dewey, John. 1938. *Logic: The Theory of Inquiry*. Henry Holt and Co.

Dotson, Kristie. 2014. "Conceptualizing Epistemic Oppression." *Social Epistemology* 28, no. 2: 115–138. https://doi.org/10.1080/02691728.2013.782585.

El Kassar, Nadja. 2022. "Epistemische Ungerechtigkeiten in und durch Algorithmen – ein Überblick." *Zeitschrift für Praktische Philosophie* 9, no. 1: 279–304. https://doi.org/10.22613/zfpp/9.1.11.

Emanuel, Ezekiel J., and Robert M. Wachter. 2019. "Artificial Intelligence in Health Care: Will the Value Match the Hype?" *JAMA* 321, no. 23: 2281–2282. https://doi.org/10.1001/jama.2019.4914.

Eubanks, Virginia. 2018. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press.

Fricker, Miranda. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford University Press.

———. 2013. "Epistemic Justice as a Condition of Political Freedom?" *Synthese* 190, no. 7: 1317–1332. https://doi.org/10.1007/s11229-012-0227-3.

Gabriel, Iason. 2020. "Artificial Intelligence, Values, and Alignment." *Minds and Machines* 30, no. 3: 411–437. https://doi.org/10.1007/s11023-020-09539-2.

Goldenberg, Anna, and Jenna Wiens. 2026. "Is AI Actually Improving Healthcare?" *Nature Medicine* 32, no. 4: 1182–1183. https://doi.org/10.1038/s41591-026-04329-2.

Goldhahn, Jörg, Vanessa Rampton, and Giatgen A. Spinas. 2018. "Could Artificial Intelligence Make Doctors Obsolete?" *BMJ,* November 7, k4563. https://doi.org/10.1136/bmj.k4563.

Haider, Syed Ali, Sahar Borna, Cesar A. Gomez-Cabello, Sophia M. Pressman, Clifton R. Haider, and Antonio Jorge Forte. 2026. "The Algorithmic Divide: A Systematic Review on AI-Driven Racial

Disparities in Healthcare." *Journal of Racial and Ethnic Health Disparities* 13, no. 1: 188–217. https://doi.org/10.1007/s40615-024-02237-0.

Hardalupas, Mahi. 2024. "Contributory Injustice, Epistemic Calcification and the Use of AI Systems in Healthcare." *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* 7, no. 1: 573–583. https://doi.org/10.1609/aies.v7i1.31659.

Heggen, Kristin Margrethe, and Henrik Berg. 2021. "Epistemic Injustice in the Age of Evidence-Based Practice: The Case of Fibromyalgia." *Humanities and Social Sciences Communications* 8, no. 1, 235. https://doi.org/10.1057/s41599-021-00918-3.

Herzog, Christian, and Jason Branford. 2025. "Relational Ethics and Structural Epistemic Injustice of AI in Medicine." *Philosophy & Technology* 38, no. 4, 160. https://doi.org/10.1007/s13347-025-00987-1.

Hill Collins, Patricia. 1990. *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*. *Perspectives on Gender*. Routledge.

Hille, Eva Maria, Patrik Hummel, and Matthias Braun. 2023. "Meaningful Human Control over AI for Health? A Review." *Journal of Medical Ethics* 52, no. e1: 50–58. https://doi.org/10.1136/jme-2023-109095.

Ho, Calvin Wai-Loon. 2023. "Generative AI and the Foregrounding of Epistemic Injustice in Bioethics." *American Journal of Bioethics* 23, no. 10: 99–102. https://doi.org/10.1080/15265161.2023.2250319.

Hoffman, Kelly M., Sophie Trawalter, Jordan R. Axt, and M. Norman Oliver. 2016. "Racial Bias in Pain Assessment and Treatment Recommendations, and False Beliefs About Biological Differences Between Blacks and Whites." *Proceedings of the National Academy of Sciences of the United States of America* 113, no. 16: 4296–4301. https://doi.org/10.1073/pnas.1516047113.

hooks, bell. 1992. *Black Looks: Race and Representation*. South End Press.

Hummel, Patrik. 2025. "Algorithmic Fairness as an Inconsistent Concept." *American Philosophical Quarterly* 62, no. 1: 53–68. https://doi.org/10.5406/21521123.62.1.04.

Kay, Jackie, Atoosa Kasirzadeh, and Shakir Mohamed. 2024. "Epistemic Injustice in Generative AI." *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* 7, no. 1: 684–697. https://doi.org/10.1609/aies.v7i1.31671.

Khera, Rohan, Melissa A. Simon, and Joseph S. Ross. 2023. "Automation Bias and Assistive AI: Risk of Harm From AI-Driven Clinical Decision Support." *JAMA* 330, no. 23: 2255–2257. https://doi.org/10.1001/jama.2023.22557.

Kidd, Ian James, and Havi Carel. 2017. "Epistemic Injustice and Illness." *Journal of Applied Philosophy* 34, no. 2: 172–190. https://doi.org/10.1111/japp.12172.

Kious, Brent M., Benjamin R. Lewis, and Scott Y.H. Kim. 2023. "Epistemic Injustice and the Psychiatrist." *Psychological Medicine* 53, no. 1: 1–5. https://doi.org/10.1017/S0033291722003804.

Königs, Peter. 2025. "The Negativity Crisis of AI Ethics." *Synthese* 206, no. 6, 277. https://doi.org/10.1007/s11229-025-05378-9.

Kulkarni, Prathit A., and Hardeep Singh. 2023. "Artificial Intelligence in Clinical Diagnosis: Opportunities, Challenges, and Hype." *JAMA* 330, no. 4: 317–318. https://doi.org/10.1001/jama.2023.11440.

Lalumera, Elisabetta. 2025. "Ameliorating Epistemic Injustice with Digital Health Technologies." In *Epistemic Justice in Mental Healthcare,* edited by Lisa Bortolotti. Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-68881-2_8.

Loh, Wulf. 2026. "Can You Feel My Pain Tonight? Automated Pain Assessment and Epistemic Injustice." *Philosophy & Technology* 39, no. 2, 72. https://doi.org/10.1007/s13347-026-01074-9.

McCradden, Melissa, Katrina Hui, and Daniel Z. Buchman. 2023. "Evidence, Ethics and the Promise of Artificial Intelligence in Psychiatry." *Journal of Medical Ethics* 49, no. 8: 573–579. https://doi.org/10.1136/jme-2022-108447.

McElfresh, Duncan C., Lucia Chen, Elizabeth Oliva, Vilija Joyce, Sherri Rose, and Suzanne Tamang. 2023. "A Call for Better Validation of Opioid Overdose Risk Algorithms." *Journal of the American Medical Informatics Association* 30, no. 10: 1741–1746. https://doi.org/10.1093/jamia/ocad110.

McKinnon, Rachel. 2016. "Epistemic Injustice." *Philosophy Compass* 11, no. 8: 437–446. https://doi.org/10.1111/phc3.12336.

Milano, Silvia, and Carina Prunkl. 2025. "Algorithmic Profiling as a Source of Hermeneutical Injustice." *Philosophical Studies* 182, no. 1: 185–203. https://doi.org/10.1007/s11098-023-02095-2.

Nagendran, Myura, Yang Chen, Christopher A. Lovejoy, et al. 2020. "Artificial Intelligence Versus Clinicians: Systematic Review of Design, Reporting Standards, and Claims of Deep Learning Studies." *BMJ* 368, m689. https://doi.org/10.1136/bmj.m689.

Nielsen, Kasper Møller, Julie Nordgaard, and Mads Gram Henriksen. 2025. "Fundamental Issues in Epistemic Injustice in Healthcare." *Medicine, Health Care and Philosophy* 28, no. 2: 291–301. https://doi.org/10.1007/s11019-025-10259-6.

Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism.* NYU Press.

Oliva, Jennifer D. 2022. "Dosing Discrimination: Regulating PDMP Risk Scores." *California Law Review*, ahead of print. https://doi.org/10.15779/Z38Z31NP8J.

Pozzi, Giorgia. 2023a. "Automated Opioid Risk Scores: A Case for Machine Learning-Induced Epistemic Injustice in Healthcare." *Ethics and Information Technology* 25, no. 1, 3. https://doi.org/10.1007/s10676-023-09676-z.

———. 2023b. "Testimonial Injustice in Medical Machine Learning." *Journal of Medical Ethics* 49, no. 8: 536–540. https://doi.org/10.1136/jme-2022-108630.

Pozzi, Giorgia, and Filippo Santoni de Sio. 2026. "Epistemic Justice as a Condition for Meaningful Human Control Over Medical AI." *Minds and Machines* 36, no. 1, 10. https://doi.org/10.1007/s11023-026-09762-3.

Rampton, Vanessa. 2020. "Artificial Intelligence versus Clinicians." *BMJ* 369, m1326. https://doi.org/10.1136/bmj.m1326.

Santoni de Sio, Filippo. 2024. *Human Freedom in the Age of AI*. 1st edition. Routledge. https://doi.org/10.4324/9781003303244.

Santoni de Sio, Filippo, and Giulio Mecacci. 2021. "Four Responsibility Gaps with Artificial Intelligence: Why They Matter and How to Address Them." *Philosophy & Technology* 34, no. 4: 1057–1084. https://doi.org/10.1007/s13347-021-00450-x.

Santoni de Sio, Filippo, and Jeroen van den Hoven. 2018. “Meaningful Human Control over Autonomous Systems: A Philosophical Account.” *Frontiers in Robotics and AI* 5, 15. https://doi.org/10.3389/frobt.2018.00015.

Savulescu, Julian. 2025. “Collective Reflective Equilibrium, Algorithmic Bioethics and Complex Ethics.” *Cambridge Quarterly of Healthcare Ethics* 34, no. 2: 204–219. https://doi.org/10.1017/S0963180124000719.

Siffels, Lotje E., and Tamar Sharon. 2024. “Where Technology Leads, the Problems Follow: Technosolutionism and the Dutch Contact Tracing App.” *Philosophy & Technology* 37, no. 4, 125. https://doi.org/10.1007/s13347-024-00807-y.

Starke, Georg, Eva de Clercq, and Bernice S. Elger. 2021. “Towards a Pragmatist Dealing with Algorithmic Bias in Medical Machine Learning.” *Medicine, Health Care and Philosophy* 24, no. 3: 341–349. https://doi.org/10.1007/s11019-021-10008-5.

Strange, Michael. 2024. “Three Different Types of AI Hype in Healthcare.” *AI and Ethics* 4, no. 3: 833–840. https://doi.org/10.1007/s43681-024-00465-y.

Taylor, R. Andrew. 2025. “AI Agents, Automaticity, and Value Alignment in Health Care.” *NEJM AI* 2, no. 8, AIp2401165. https://doi.org/10.1056/AIp2401165.

Topol, Eric J. 2019. “High-Performance Medicine: The Convergence of Human and Artificial Intelligence.” *Nature Medicine* 25, no. 1: 44–56. https://doi.org/10.1038/s41591-018-0300-7.

Whittlestone, Jess, Rune Nyrup, Anna Alexandrova, and Stephen Cave. 2019. “The Role and Limits of Principles in AI Ethics: Towards a Focus on Tensions.” *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, January 27, 195–200. https://doi.org/10.1145/3306618.3314289.

Wiesenfeld, Batia Mishan, Yin Aphinyanaphongs, and Oded Nov. 2022. “AI Model Transferability in Healthcare: A Sociotechnical Perspective.” *Nature Machine Intelligence* 4, no. 10: 807–809. https://doi.org/10.1038/s42256-022-00544-x.

WHO (World Health Organization). 2021. *Ethics and Governance of Artificial Intelligence for Health: WHO Guidance*. World Health Organization. https://apps.who.int/iris/handle/10665/341996.