Psychiatry Research 328 (2023) 115466

Contents lists available at ScienceDirect

Psychiatry Research

journal homepage: www.elsevier.com/locate/psychres

Ethical considerations for the use of artificial intelligence in medical decision-making capacity assessments

Michael R. MacIntyre [a,*], Richard G. Cockerill [b], Omar F. Mirza [c], Jacob M. Appel [d]

[a] Department of Psychiatry and Biobehavioral Sciences at the David Geffen School of Medicine, University of California, Los Angeles, 760 Westwood Plaza, Los Angeles, CA, U.S.A.
[b] Department of Psychiatry and Behavioral Neuroscience, University of Chicago Pritzker School of Medicine, 5815 S. Maryland Ave., Chicago, Illinois, USA
[c] Department of Psychiatry, NYC Health+Hospitals/Harlem, 506 Lenox Ave., New York, New York, USA
[d] Department of Psychiatry and Medical Education, Mount Sinai's Icahn School of Medicine, 1 Gustave L. Levy Pl., New York, New York, USA

ARTICLE INFO

Keywords:
Capacity
Artificial intelligence
Machine learning
Ethics
Informed consent

ABSTRACT

The rapid advancement of artificial intelligence (AI) and machine learning are providing new tools to clinicians. AI tools have the potential to process vast amounts of data in a short amount of time, providing new insights and changing how we approach complicated healthcare problems. AI has the potential to assist clinicians in medical decision-making capacity assessments by providing additional insights to an evaluation process that currently lacks universal objective standards. However, despite the promise of AI in this setting, there remain significant concerns making it unlikely to replace human evaluators anytime soon. AI remains highly susceptible to biased inputs and thus biased decisions, raises questions about autonomy, and creates uncertainty for who is accountable for the ultimate decision of capacity. In this paper we explore these ethical considerations of using AI for capacity assessments. While we acknowledge AI may not be ready to replace physicians in determining patient medical-decision making capacity, these new technologies have significant near-term potential as a tool to screen patients, uncover physician biases, and guide next steps after a capacity determination has been made.

1. Introduction

Artificial intelligence (AI) and machine learning are rapidly changing numerous fields, with medicine being no exception. The development of technology capable of absorbing and processing large amounts of data in seconds has the potential to transform the practice of medicine. Several fields may be especially primed for significant changes due to AI. In radiology, AI tools have already been show to outperform human radiologists - both in speed and in accuracy - on specific tasks, like identifying malignancies on breast imaging (McKinney et al., 2020).

While much research has focused on AI's utility for providing objective answers to complex questions, as AI progresses, it will be used for highly nuanced, complex assessments involving both emotional states and human behavior. For example, a recent survey found an online chatbot provided not only higher quality responses to patient questions than physicians, but was also seen as more empathetic (Ayers et al., 2023). In another study, a deep learning algorithm was trained to interpret facial expressions to categorize pain among a cohort of post-operative surgical patients on a scale of 0-10. The algorithm dramatically outperformed human nurses performing the same tasks, giving an exactly correct pain score in 53% of cases, while the nurses did so in only 14.9% of cases (Fontaine et al., 2022). In psychiatry, AI offers potential for improving suicide and violence risk assessments, each of which currently defy reliable prediction by mental health professionals (Bernert et al., 2020; Cockerill, 2020; Zheng et al., 2020).

AI also has potential in the assessment of decision-making capacity. The assessment of capacity is an essential task of any medical team because a patient must be able to understand, appreciate, and weigh the risks and benefits of any treatment to which they consent (Appelbaum and Grisso, 1988). Assuring concordance between the patient's underlying values and current preferences is also a valuable part of the assessment process (Appel, 2022). Without an appropriate assessment of capacity, medicine becomes paternalistic with those most vulnerable at the greatest risk of losing autonomy in their own healthcare.

The lack of universal objective standards and the highly personal nature of decision-making may initially seem poorly suited to computers. However, the large language models (LLMs) in use today are massive, with hundreds of layers of artificial neurons. They can consider billions of variables simultaneously and are constantly improving. It is

* Corresponding author at: Box 3342, Santa Monica, California, 90408.
E-mail address: michaelmacintyremd@gmail.com (M.R. MacIntyre).

https://doi.org/10.1016/j.psychres.2023.115466
Received 1 August 2023; Received in revised form 3 September 2023; Accepted 4 September 2023
Available online 7 September 2023

Abbreviations

AI artificial intelligence

also worth considering that "messy" problems like assessing capacity are notoriously difficult for even the best trained human practitioners. In such areas, the bar for an AI tool to be "good enough" is much lower than in more straightforward tasks like interpreting chest radiographs.

While it is unclear whether AI can replace physician-based clinical determinations of capacity, AI has near-term potential to serve as a screening or adjunct tool to increase accuracy and validity, minimize subjectivity and biases, enhance efficiency in health systems, and increase standardization between clinicians. AI may be utilized as a check against implicit biases of physicians requesting capacity consultations, which have been demonstrated to be subject racial disparities (Garrett et al., 2023). Additionally, it has been suggested that utilizing AI for predictive modeling would be useful in employing substituted judgment for a patient without a surrogate decision maker, after it has been determined the patient does not have medical decision making capacity (Rid and Wendler, 2014). Some value may also exist for the use of AI in assessing capacity in patients who retain "conscious awareness" while otherwise meeting the clinical criteria for a "diagnosis of a vegetative state" (Owen et al., 2006).

The adoption of any AI tool for decision-making assessments will have challenges. AI must be trained, and output is only as good as the data used to train the model. As a result, AI is highly susceptible to perpetuating or even magnifying biases already present in capacity assessments (Garrett et al., 2023). AI algorithms are neither transparent nor explainable, and clinicians using these tools might be unaware of how specific biased or incorrect information affected outputs, preventing an opportunity for a physician to intervene. Beyond that, ethical challenges may arise regarding if and when a physician should intervene and "override" an AI model if that model was previously demonstrated to be more consistently accurate than typical care clinical assessments.

Finally, concerns arise about AI removing an essential interpersonal element from capacity assessments due to an overreliance on technology. Ideally, the informed consent process should be used as a tool to facilitate patient autonomy and self-respect. Evaluating patients' understanding of treatment and their true capacities to consent may foster placebo effects while minimizing less desirable nocebo effects (Weimer et al., 2020). A capacity evaluation is more than a benign diagnostic tool for which another mechanism may easily substitute. Rather, a capacity evaluation is itself an intervention affecting the overall care of a patient, endowed with its own set of risks, side effects, and consequences (Mirza and Appel, 2023). A patient's values are inherent to the capacity assessment itself and challenges may arise regarding whether patients will even consent to the use of AI models in their care (Wendler, 2022).

While the future of AI in improving patient care remains promising, caution must be exerted, particularly to minimize harm from using this new technology in ways for which it was not intended or capable of doing accurately. In this paper, we discuss the ethics issues clinicians and researchers must consider when considering the use of AI as a medical decision-making tool. We also review the strong potential of AI in assisting with complicated capacity assessments.

2. Challenges in decision making capacity and assessment

Medical decision-making capacity refers to an individual's ability to understand relevant medical information, appreciate the implications and consequences of the proposed treatment, and make a reasoned choice based upon their own values and preferences. It is essential to allow patients to make their own decisions to respect their rights to autonomy and self-determination. Conversely, healthcare providers must ensure that a patient has the capacity to make informed decisions to protect vulnerable individuals and to provide appropriate care.

When a patient's capacity to render a decision is uncertain, providers must complete a thorough assessment to determine the presence of capacity or the causes of any impairment to capacity. This assessment is a complex process requiring careful evaluation of multiple factors. It often demands a collaborative approach, involving the primary treatment team, medical specialists, psychiatric consultants, and family or trusted allies of a patient to gather information, provide support, and make informed judgements. When assessments are performed, they focus on the specific decision applying to the patient at that precise time. If decisional capacity is impaired, clinicians must attempt to determine why and whether anything can be done to restore capacity so that patients may once again engage in their own healthcare decisions. Numerous factors may affect one's ability to render decisions about one's own medical care including cognitive impairment, psychiatric conditions, developmental disabilities, medical syndromes, and intoxication. A clinician must not conflate a patient's disagreement with the care team's recommendations with lack of capacity. Additionally, willful refusal to engage meaningfully with the care team or the capacity evaluation process should not automatically be assumed to indicate a patient's lack of decisional capacity (Appel, 2023). Clinicians must assess decision making in the context of the patient's own values and not those of the treatment team.

2.1. Lack of standardization in capacity assessments

While Applebaum proposed the most widely utilized criteria for determining capacity (Appelbaum and Grisso, 1988), there have been numerous proposed variations and alternative frameworks. Beyond the core components of determining capacity, clinicians employ a variety of approaches to the assessment ranging from the clinical interview to the use of formal structured tools (Tunzi, 2001). While very little standardization exists with clinical interviews more generally, the inherent contextual nature of the construct of decisional capacity in healthcare renders standardization even more challenging. For example, regarding capacity assessments, psychometric instruments cannot be fully standardized (Palmer and Harmell, 2016). It has been argued that effective capacity assessments by their very nature cannot be reduced to operationalized criteria, but rather are complex encounters in which a clinician must actively engage with patients through a distinctive decision-making process (Banner, 2012). While this may be accurate, health sciences still strive for reliability between clinicians and the lack of a gold-standard prevents any conclusions about the validity of any singular capacity assessment (Cairns et al., 2005).

2.2. Biases in capacity assessments

Derived from legal standards, the prevailing decision making models have never been calibrated to an empiric truth originating in the patient (Appelbaum and Grisso, 1988). Following, it is no surprise that decentering patients from their decision making via capacity challenges opens the door for external bias. In the most overt instance, this bias manifests in the disproportionate requests for capacity assessments in marginalized identities (Garrett et al., 2023). In other cases the bias is disguised in identifying low education level, acknowledged fear of or discomfort with institutional health care setting, and significant cultural or language barriers as risk factors for impaired medical decision making capacity rather than an output of a flawed model (Barstow et al., 2018). More concerning is the fact that physician personal values shape capacity judgements (Hermann et al., 2015). Upstream of the actual capacity determination, instances of bias described as "value impregnation" influence the physician's judgment of medical indication for a treatment so much so as to conflict with the established evidence base (Björk et al., 2016). In models that are heavily based in understanding, such bias is even more nefarious and may even undermine the

medical decision-making process (Lynøe et al., 2018). Even though the potential for bias from the physician is plenty, even surrogates can be perpetrators of influence that does not align with the patient's preferences and principals (Devnani et al., 2017).

2.3. When patients do not have capacity

Following a finding of incapacity, additional challenges arise. Treatment teams must then decide upon a course of future action. Sometimes, a patient has an advance directive, such as a living will or designated health care agent or has clearly made their wishes known prior to losing capacity allowing clinicians to proceed comfortably, knowing that they are honoring the patient's treatment goals. Sometimes a surrogate decision maker is available. Depending on jurisdiction, a spouse, family member, or close friend may speak on behalf of the incapacitated patient, making decisions as the patient would have wanted if that individual retained decisional capacity.

Frequently the prior wishes of an incapacitated patient are not clear. Often, loss of capacity occurs suddenly in high stakes situations, such as in a medical emergency or during an inpatient hospitalization. The patient may not have had the opportunity to discuss such a situation in advance, the treatment team may have no prior history with the patient, and family or friends may not be readily available. Complications may arise if divergent wishes are presented by different third parties, such as two close family members arguing that the patient would have wanted different treatment paths, or a spouse stating that the patient had planned to alter the contents of a living will but had not yet had the chance to do so. While many states and hospital systems have processes in place for rendering medical decisions in these situations, no certainty exists that a treatment team is deciding in accordance with the patient's wishes.

3. Utilizing artificial intelligence in medical decision-making capacity assessments

To appreciate how AI might affect capacity assessments, it is essential to understand AI basics and machine learning. AI is a broad term describing the use of automated processes to perform cognitive tasks. Machine learning, a subset of AI, describes the processes by which machine processes (or algorithms) improve their predictive power over time, as they are exposed to more data. A subset of machine learning, deep learning, refers to the process by which a computer can mimic the functions of the human brain by using artificial neural networks (Ahuja, 2019). Deep learning involves minimal human supervision as it trains itself based upon whether its output is correct or false. The most powerful models of today rely upon deep learning.

Today's best AI models, like OpenAI's GPT4 (the engine behind chatGPT), are incredibly complex, with hundreds of layers of artificial neurons weighing billions of parameters. The artificial neural network structure utilized by deep learning models is inherently nontransparent, even with few layers, and human operators are unable to understand how an AI model reaches the conclusions it does. This has been described as the "black box problem" (Bathaee, 2018). The progress made in recent years has exceeded almost all expectations. For example, in 2010, when asked to identify an unknown object from an image, the average AI algorithm was wrong over half the time, while the best was incorrect one quarter of the time. But by 2017, after the wide adoption of deep learning, the error rate for the average AI algorithm was less than 5% (Gershgorn, 2017)! Rapid progress has also been made in the use of predictive algorithms in search and social media, computer vision in self-driving cars, and most recently in Generative AI (apparently creative work done by computer algorithms). AI has even made a convincing entry into the world of art. The DALL-E model, also developed by OpenAI, has also made apparently creative graphic designs available at low cost to any interested customers. This transformation has led some to predict that AI is soon to "challenge our notions of beauty, creativity, and the nature of art" (Chatterjee, 2022).

Regarding capacity assessments, one feature that makes AI such a powerful tool is its ability to analyze massive data sets in a fraction of the time it would take a human. These data sets can be extremely broad, including but not limited to medical records for an individual patient cross referenced with the records of thousands of comparable patients, observations of the patient's facial expressions, tone of voice, and vital signs, and personal information like social media posts, texts, or e-mails. As data is obtained, whether through patient interview, result of diagnostic tests, or change in the treatment plan, an AI algorithm could quickly incorporate this new information into its analysis. Recalling that the threshold for decision-making for capacity exists on a sliding scale (Atluru, 2016), AI will likely incorporate the changing threshold more accurately than a human provider in any given case. Finally, an AI algorithm is likely to be able to identify specific causes of incapacity more aptly than a human practitioner, such as a knowledge deficit, poor communication, or a cognitive deficit. This may allow reversible causes of incapacity to be addressed such that capacity is restored.

AI has the potential to fill in gaps in clinician abilities as well. While all physicians are responsible for assessing capacity, many clinicians lack formal training in this assessment, particularly with more nuanced situations. Many physicians, despite training, demonstrate poor patient-doctor communication skills that may include struggling to present the right amount and quality of information needed for patient decision-making. Physicians commonly over-estimate the health literacy of their patients (Kelly and Haidet, 2007), which can have disastrous consequences when assessing decisional capacity. One of the most common reasons patients are found to lack decision-making capacity is failures of communication (Ubel et al., 2017).

AI has potential to address these concerns. While physicians overestimate health literacy across all patient demographics, this error is most common with minority patients and particularly Black patients (Kelly and Haidet, 2007). These overestimates might partially explain research showing that minority patients are more likely to be subject to formal requests for psychiatric evaluation of capacity (Garrett et al., 2023). Furthermore, many physicians lack formal knowledge or training in cultural competence and cultural humility, which can raise significant doubts regarding the validity of a capacity assessment in cases in which cultural or linguistic factors are more significant (Stubbe, 2020; Tervalon and Murray-García, 1998). Socioeconomic status and cultural background shape views on autonomy and decision-making, and AI may help inform providers of issues they are not aware of or have not yet encountered. If deficits in communication can be identified, AI algorithms may be able to provide additional materials, specifically targeted to a patient, which may improve comprehension of key material. Computers themselves do not have inherent biases beyond those of their programmers and training data and have the potential to promote fairness and objectivity if used appropriately.

Presently, rather than replacing humans, AI has considerable potential to serve as a tool to guide clinical decision-making, helping to indicate areas of bias, screen for appropriate cases for further assessment, and alert physicians to nuances that might otherwise go undetected. Soon AI may be used as a standalone screening tool or stratification tool (i.e., high, medium, or low risk of not having capacity) to guide more thorough evaluations. Additionally, AI might provide insights that a doctor might miss, such as subtle deficits in a patient's health literacy, communication skills, or cultural barriers between provider and patient. AI may be particularly suited to tackle the most challenging situations in capacity assessments, such as ascertaining the wishes of an incapacitated patient who lost the ability to communicate and there is disagreement between surrogate decision-makers. Though technological, logistical, and privacy concerns make this highly unlikely in the near future, in such a case an algorithm could theoretically rapidly review reams of data, including every social media post, text, or email the patient had ever sent, to predict what that patient most likely wants in any given clinical situation. While it is nearly impossible to predict

exactly how AI will be used in the assessment of decision-making capacity, it is nearly certain that it will play an important, and growing role in this arena moving forward.

4. Challenges inherent to using AI in this setting

4.1. Addressing bias

While all doctors utilizing the same AI tool might create a specious sense of standardization, one of the major drawbacks is the lack of understanding of *how* AI comes to its conclusions (the "black box problem"). That is, though the AI trains itself and appears to continuously improve as the outputs continue to improve, the person utilizing the AI does not know the intervening steps or data that the AI actually harnessed in formulating its final opinions. The human user of the algorithm does not know which factors were weighed heavily, undervalued or even omitted. Even if this information could be known, it would be nearly impossible to interpret, as the best AI models in use today weigh billions of parameters simultaneously (Ananthaswamy, 2023; Google AI, 2022). This creates a uniquely challenging problem when attempting to determine whether human biases have infected AI algorithms.

AI tools are only as good as the data used to train them, and these data are inevitably produced by humans, who all have their own biases. This could lead inherently unbiased algorithms to produce bias-infected data which on the surface appears entirely objective. Given the significant power of these computing systems, a data set with even slight bias may be magnified extensively in the final output. At the same time, an AI system might be trained to identify and account for biases in its decision-making or to identify biases present in human decision making.

Since human beings are not privy to the processes through which AI reaches its conclusions, it may not be evident if or precisely where in the process an error was made, and at which point a human physician might intervene. One study showed how AI models can accurately predict self-reported race from corrupted, cropped, and noisy medical images, often when medical experts cannot (Gichoya et al., 2022). This finding is significant, as it suggests the possibility that an AI model may be using its knowledge of self-reported race to reach conclusions, without the clinician having access to the same data the model used to make its decision. As a result, the clinician may remain unaware of an underlying racial bias used in decision making.

Though AI might serve as a valuable tool in identifying implicit biases of physicians that may impact their clinical assessment of capacity, we caution against an overreliance on technology. Completely relying upon AI to account for detection of one's personal biases might tempt the physician to not challenge themselves, when our hope is that physicians constantly strive to improve their understanding of cultural and socioeconomic experiences to minimize biases in care.

Because of their transformative potential, the deployment of AI algorithms across disciplines is inevitable. AI certainly has the potential to improve objectivity and reduce bias. However, if more attention is not devoted to bias-mitigation in the development of AI, it is likely that, rather than reduce human biases, AI will magnify and conceal them (Ratajczak and Cockerill, 2023).

4.2. Autonomy

One of the core reasons for capacity assessments is to assure that the patient has capacity to exercise autonomy, as an impaired patient does not necessarily have autonomy to make complicated decisions. However, it has been argued that patients may neither trust AI nor want AI involved in this process, and supplementation of the capacity assessment with AI to predict a patient's preference may actually undermine the values of a patient who objects to it (Wendler, 2022). A capacity assessment is not a harmless endeavor (Mirza and Appel, 2023), with a consensus among western bioethicists that patients should not be coerced into participating in care against their will (Atluru, 2016). To the best of our knowledge, there currently are no methods to obtain informed consent from patients regarding the use of AI in their treatment, and it is critical that clear and concise educational tools be developed before utilizing AI in any clinical assessment, including those involving decisional capacity (Darby et al., 2023). This is a complex challenge as there may be some aspects of medical care when a patient will enthusiastically embrace AI involvement (such as reading their imaging) but others where the patient may not (such as making decisions about life prolonging care).

4.3. Moral responsibility

When high-stake decisions are made about an individual's life, such as whether to withdraw care or move forward with a high-risk procedure, traditionally a responsible party is accountable for the ultimate decision. In modern medical systems, when patient autonomy is overridden due to lack of capacity, the decision is made by a treating physician, often with consultation and input from psychiatry consultants specializing in capacity assessments, service chiefs, and hospital ethics committees.

However, if a decision were to be made by an AI algorithm, it is less clear how to improve the system should a bad outcome occur. If AI algorithms are ever to be involved, it is essential that guidelines be established regarding when and how a physician may "overrule" an algorithm.

While one might be tempted to afford the treating physician final authority to overrule an AI algorithm output, this approach carries its own challenges. If AI is demonstrated to be consistently valid, performing a task more efficiently and accurately than a human doctor, physician input runs the risk of being a disadvantage to patients. Given the known errors in human judgment and the limitations of purely clinical decision-making, to ignore a robust AI model's verdict may be akin to a primary physician ignoring the read of a radiologist or a resident dismissing the direction of the attending physician.

5. Clinical implications

The most significant promise for AI in capacity assessment is likely in the use of the technology for triage, stratification and limiting unnecessary and inappropriate assessments. Since capacity evaluation is not merely a diagnostic tool, but rather a clinical intervention with its own risks, side effects and implications for the physician-patient therapeutic relationship (Mirza and Appel, 2023), all efforts must be made to prevent unnecessary capacity assessments and consultations. Such caution is especially important because both the burden of these excess capacity consults and their negative consequences fall disproportionately upon racial minorities and patients with limited social capital (Garrett et al., 2023). By having AI review the medical records and consultation requests in advance, the technology might warn psychiatrists that in the overwhelming majority of similar cases, capacity assessments have had no impact upon patient care–either because patients possessed capacity or because their choices coincided fully with those in their advance directives. Consultants might bring this information back to the clinical team requesting the consultation and spark a discussion as to whether the evaluation is necessary and whether the potential value of such an assessment exceeds its negative consequences. Knowing in advance the specific biases that have affected clinicians in similar cases of capacity assessment may also help clinicians and consultants check their own biases when approaching the patient in front of them.

While the "four skills" and value concordance elements of capacity, associated with Appelbaum & Grisso and Mirza & Appel respectively (Grisso et al., 1995; Mirza and Appel, 2023), are crucial for decisional capacity evaluation, the level of scrutiny applied in each case is also important and likely to vary (Buchanan and Brock, 1990). The level of investigation required may vary on a "sliding scale" based upon such factors as the complexity of the clinical decision at hand (Mellgard and

Gligorov, 2022) and its potential dangerousness (Drane, 1984). While the same mechanism of assessment is theoretically applied to determine whether a patient can consent to have vital signs taken as to refuse to consent to an emergent, life-saving appendectomy, the degree of investigation involved necessarily differs. AI may prove extremely helpful in advance of evaluation in determining the degree of investigation necessary in a particular situation and the types of information that must be gathered to render a consistently accurate assessment.

As a clinician prepares to conduct an assessment, AI may be able to pinpoint the likely underlying factors affecting capacity or even propose specific questions that might target subtle deficiencies otherwise not easily detectable. For example, if AI suggests that a patient lacks capacity due to cognitive limitations, the assessor might adopt one approach to investigation; if the AI suggests a higher likelihood that lack of healthcare knowledge is the factor raising concerns, the assessor might target the evaluation toward determining whether the patient has been sufficiently informed about the decision in manner that he or she can be expected to understand.

AI is unlikely to replace decision making by human physicians entirely any time in the near future, and this is certainly true regarding the assessment of decisional capacity in the clinical setting. However, AI holds great potential as a mechanism for screening and stratifying such cases. At the outset, one might easily identify the cases in which a patient clearly does or does not have capacity, allowing the primary team to spend only limited efforts confirming these predictions, rather than expending significant resources investigating or requesting a formal capacity consultation from a psychiatrist.

6. Conclusion

AI holds tremendous potential for psychiatry as the technology rapidly improves and develops the ability to understand human cognitive and emotional states. AI may prove to be a vital tool in the essential task of ensuring that patients have medical decision-making capacity. It offers the possibility of increasing the reliability of a capacity assessment, which are currently subject to the nuances of any given evaluator's personal experiences and style. AI can consider orders of magnitude more data much more quickly than any human practitioner. This allows AI systems to utilize far more information in making these important clinical decisions, with the potential to minimize the impact of human biases.

However, medical systems remain far from being ready to safely incorporate AI tools at this time. Numerous issues must be addressed including how this novel tool might be effectively implemented, what checks and balances a treatment provider may be able to employ with such a computer model, and assurances that the AI model is not providing biased results derived from biased data.

It is unclear when, if ever, AI tools will operate independently in assessing decision-making capacity. The more pressing concern is addressing how they will be deployed as a tool for human practitioners. We discuss the strong potential for AI to assist in screening for capacity and in aiding clinicians with identifying areas that may need more formal clinical assessment. If used cautiously and with careful attention paid to its limitations and biases, AI will be a powerful mechanism for enhancing our ability to assess capacity efficiently and objectively, thus improving our protection of patient autonomy.

AI declaration

During the preparation of this work the authors did not utilize any generative AI or AI-assisted technologies.

Declaration of Competing Interest

The authors report no proprietary or commercial interest in any product mentioned or concept discussed in this article.

Acknowledgements and disclosures

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Ahuja, A.S., 2019. The impact of artificial intelligence in medicine on the future role of the physician. PeerJ 2019. https://doi.org/10.7717/peerj.7702.
Ananthaswamy, A., 2023. In AI, is bigger always better? Nature. https://doi.org/10.1038/d41586-023-00641-w.
Appel, J.M., 2023. Anything you do not say can be used against you: volitional refusal to engage in decisional capacity assessment. J. Clin. Ethics 34, 204–210. https://doi.org/10.1086/725049.
Appel, J.M., 2022. A values-based approach to capacity assessment. J. Leg. Med. 42, 53–65. https://doi.org/10.1080/01947648.2022.2162171.
Appelbaum, P.S., Grisso, T., 1988. Assessing patients' capacities to consent to treatment. N. Engl. J. Med. 319 https://doi.org/10.1056/nejm198812223192504.
Atluru, A., 2016. Revisiting decision-making capacity. Am. J. Psychiatry Resid. J. 11 https://doi.org/10.1176/appi.ajp-rj.2016.111105.
Ayers, J.W., Poliak, A., Dredze, M., Leas, E.C., Zhu, Z., Kelley, J.B., Faix, D.J., Goodman, A.M., Longhurst, C.A., Hogarth, M., Smith, D.M., 2023. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern. Med. https://doi.org/10.1001/jamainternmed.2023.1838.
Banner, N.F., 2012. Unreasonable reasons: Normative judgements in the assessment of mental capacity. J. Eval. Clin. Pract. 18 https://doi.org/10.1111/j.1365-2753.2012.01914.x.
Barstow, C., Shahan, B., Roberts, M., 2018. Evaluating medical decision-making capacity in practice. Am. Fam. Physician 98.
Bathaee, Y., 2018. The artificial intelligence black box and the failure of intent and causation. Harv. J. Law Technol. 31.
Bernert, R.A., Hilberg, A.M., Melia, R., Kim, J.P., Shah, N.H., Abnousi, F., 2020. Artificial intelligence and suicide prevention: A systematic review of machine learning investigations. Int. J. Environ. Res. Public Health. https://doi.org/10.3390/ijerph17165929.
Björk, J., Lynöe, N., Juth, N., 2016. Empirical and philosophical analysis of physicians' judgments of medical indications. Clin. Ethic. 11 https://doi.org/10.1177/1477750916657666.
Buchanan, A.E., Brock, D.W., 1990. Deciding for Others: The Ethics of Surrogate Decision Making, 1st ed. Cambridge University Press, (CUP), Cambridge.
Cairns, R., Maddock, C., Buchanan, A., David, A.S., Hayward, P., Richardson, G., Szmukler, G., Hotopf, M., 2005. Reliability of mental capacity assessments in psychiatric in-patients. Br. J. Psychiatry 187. https://doi.org/10.1192/bjp.187.4.372.
Chatterjee, A., 2022. Art in an age of artificial intelligence. Front. Psychol. 13 https://doi.org/10.3389/fpsyg.2022.1024449.
Cockerill, R.G., 2020. Ethics implications of the use of artificial intelligence in violence risk assessment. J. Am. Acad. Psychiatry Law 48. https://doi.org/10.29158/JAAPL.003940-20.
Darby, W.C., MacIntyre, M., Cockerill, R.G., Stephens, D.B., Weinstock, R., Darby, R.R., 2023. In the courts: ethical and legal implications of emerging neuroscience technologies used for forensic purposes. Ethic. Clin. Neuroinnov. https://doi.org/10.1007/978-3-031-14339-7_10.
Devnani, R., Slaven, J.E., Bosslet, G.T., Montz, K., Inger, L., Burke, E.S., Torke, A.M., 2017. How surrogates decide: a secondary data analysis of decision-making principles used by the surrogates of hospitalized older adults. J. Gen. Intern. Med. 32 https://doi.org/10.1007/s11606-017-4158-z.
Drane, J.F., 1984. Competency to give an informed consent: a model for making clinical assessments. JAMA J. Am. Med. Assoc. 252 https://doi.org/10.1001/jama.1984.03350070043021.
Fontaine, D., Vielzeuf, V., Genestier, P., Limeux, P., Santucci-Sivilotto, S., Mory, E., Darmon, N., Lanteri-Minet, M., Mokhtar, M., Laine, M., Vistoli, D., 2022. Artificial intelligence to evaluate postoperative pain based on facial expression recognition. Eur. J. Pain (United Kingdom) 26. https://doi.org/10.1002/ejp.1948.
Garrett, W.S., Verma, A., Thomas, D., Appel, J.M., Mirza, O., 2023. Racial disparities in psychiatric decisional capacity consultations. Psychiatr. Serv. 74 https://doi.org/10.1176/appi.ps.202100685.
Gershgorn, D., 2017. The data that transformed AI research -and possibly the world [WWW Document]. Quartz.
Gichoya, J.W., Banerjee, I., Bhimireddy, A.R., Burns, J.L., Celi, L.A., Chen, L.C., Correa, R., Dullerud, N., Ghassemi, M., Huang, S.C., Kuo, P.C., Lungren, M.P., Palmer, L.J., Price, B.J., Purkayastha, S., Pyrros, A.T., Oakden-Rayner, L., Okechukwu, C., Seyyed-Kalantari, L., Trivedi, H., Wang, R., Zaiman, Z., Zhang, H., 2022. AI recognition of patient race in medical imaging: a modelling study. Lancet Digit. Heal. 4 https://doi.org/10.1016/S2589-7500(22)00063-2.
Google, AI, 2022. Pathways language model (PaLM): Scaling to 540 billion parameters for breakthrough performance [WWW Document]. Google AI Blog.
Grisso, T., Appelbaum, P.S., Mulvey, E.P., Fletcher, K., 1995. The MacArthur treatment competence study. II: measures of abilities related to competence to consent to treatment. Law Hum. Behav. 19, 127–148. https://doi.org/10.1007/BF01499322.
Hermann, H., Trachsel, M., Biller-Andorno, N., 2015. Physicians' personal values in determining medical decision-making capacity: a survey study. J. Med. Ethics 41. https://doi.org/10.1136/medethics-2014-102263.

Kelly, P.A., Haidet, P., 2007. Physician overestimation of patient literacy: a potential source of health care disparities. Patient Educ. Couns. 66 https://doi.org/10.1016/j.pec.2006.10.007.

Lynøe, N., Helgesson, G., Juth, N., 2018. Value-impregnated factual claims may undermine medical decision-making. Clin. Ethic. 13 https://doi.org/10.1177/1477750918765283.

McKinney, S.M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., Back, T., Chesus, M., Corrado, G.C., Darzi, A., Etemadi, M., Garcia-Vicente, F., Gilbert, F.J., Halling-Brown, M., Hassabis, D., Jansen, S., Karthikesalingam, A., Kelly, C.J., King, D., Ledsam, J.R., Melnick, D., Mostofi, H., Peng, L., Reicher, J.J., Romera-Paredes, B., Sidebottom, R., Suleyman, M., Tse, D., Young, K.C., De Fauw, J., Shetty, S., 2020. International evaluation of an AI system for breast cancer screening. Nature 577, 89–94. https://doi.org/10.1038/s41586-019-1799-6.

Mellgard, G., Gligorov, N., 2022. Complexity, not severity: reinterpreting the sliding scale of capacity. Cambridge Q. Healthc. Ethics 31. https://doi.org/10.1017/S0963180122000111.

Mirza, O.F., Appel, J.M., 2023. Capacity reconceptualized: from assessment tool to clinical intervention. Cambridge Q. Healthc. Ethics. https://doi.org/10.1017/s0963180123000105.

Owen, A.M., Coleman, M.R., Boly, M., Davis, M.H., Laureys, S., Pickard, J.D., 2006. Detecting awareness in the vegetative state. Science (80-.) 313. https://doi.org/10.1126/science.1130197.

Palmer, B.W., Harmell, A.L., 2016. Assessment of healthcare decision-making capacity. Arch. Clin. Neuropsychol. 31 https://doi.org/10.1093/arclin/acw051.

Ratajczak, R., Cockerill, R.G., 2023. Artificial intelligence in violence risk assessment: addressing racial bias and inequity. J. Psychiatr. Pract. 29, 239–245. https://doi.org/10.1097/PRA.0000000000000713.

Rid, A., Wendler, D., 2014. Use of a patient preference predictor to help make medical decisions for incapacitated patients. J. Med. Philos. (United Kingdom) 39. https://doi.org/10.1093/jmp/jhu001.

Stubbe, D.E., 2020. Practicing cultural competence and cultural humility in the care of diverse patients. Focus (Madison) 18. https://doi.org/10.1176/appi.focus.20190041.

Tervalon, M., Murray-García, J., 1998. Cultural humility versus cultural competence: A critical distinction in defining physician training outcomes in multicultural education. J. Health Care Poor Underserv. 9 https://doi.org/10.1353/hpu.2010.0233.

Tunzi, M., 2001. Can the patient decide? Evaluating patient capacity in practice. Am. Fam. Phys. 64.

Ubel, P.A., Scherr, K.A., Fagerlin, A., 2017. Empowerment failure: how shortcomings in physician communication unwittingly undermine patient autonomy. Am. J. Bioeth. 17 https://doi.org/10.1080/15265161.2017.1378753.

Weimer, K., Enck, P., Dodd, S., Colloca, L., 2020. Editorial: placebo and nocebo effects in psychiatry and beyond. Front. Psychiatry. https://doi.org/10.3389/fpsyt.2020.00801.

Wendler, D., 2022. Promoting the values for surrogate decision-making. JAMA. https://doi.org/10.1001/jama.2022.10347.

Zheng, L., Wang, O., Hao, S., Ye, C., Liu, M., Xia, M., Sabo, A.N., Markovic, L., Stearns, F., Kanov, L., Sylvester, K.G., Widen, E., McElhinney, D.B., Zhang, W., Liao, J., Ling, X. B., 2020. Development of an early-warning system for high-risk patients for suicide attempt using deep learning and electronic health records. Transl. Psychiatry 10. https://doi.org/10.1038/s41398-020-0684-2.