International Journal of Communication 19(2025), Forum 1859–1865 1932–8036/2025FRM0002

When Faulty AI Falls Into the Wrong Hands: The Risks of Erroneous AI-Driven Healthcare Decisions

EUGENE JANG
University of Southern California, USA

This article investigates the consequences and implications of using artificial intelligence (AI) models in healthcare decision making. Specifically, it discusses a lawsuit in which a private healthcare company (UnitedHealthcare) allegedly used an erroneous AI model to deny coverage for its patients' medical services. While previous studies have found that AI can unintentionally produce biased outputs due to inadequate training data or statistical logics, this example is notable, as the insurer was accused of intentionally using a faulty AI model to their advantage. Furthermore, the victims in this case are elderly individuals who are particularly vulnerable to the negative impacts of AI biases. Although there has been a long history of health insurance companies rejecting medical payouts, the implementation of AI in decision-making processes has accelerated this trend and created loopholes for fraudulent practices. This article illustrates the detrimental consequences of erroneous AI-based healthcare decisions through a specific example, discusses how AI complicates the issue of liability and responsibility, and calls to the need for improved transparency and accountability in AI regulation.

Keywords: artificial intelligence (AI), algorithmic bias, error, healthcare, ageism

Efforts to mitigate the adverse effects of biases in algorithmic systems have been ongoing, yet significant gaps in research remain. This article examines a lawsuit where a private healthcare company was accused of using a flawed artificial intelligence (AI) model for medical claim assessments. Unlike most previous cases where AI biases were unintentionally or unknowingly brought out due to inadequate training data, the plaintiffs in this case allege that the insurance company knowingly used a defective algorithm to deny coverage for patients' medical services. By examining this case, this paper highlights the potential harms and societal implications of using AI in healthcare decision making, with a particular focus on the underexplored issue of ageism in AI bias (Chu et al., 2022).

A body of literature supports the notion that AI systems are more likely to work against the marginalized population, reproducing existent biases and discrimination. For instance, online advertisements for high-paying jobs are less likely to be shown to women (Datta, Tschantz, & Anupam, 2015), and job-hiring algorithms favor men over women (Ajunwa, 2020). Racist biases are also frequently observed in AI. African Americans are erroneously associated with crime records or negative ads (Noble, 2018), and facial recognition systems perform worse on people of color (Buolamwini & Gebru, 2018). These biases can root from diverse sources such as problematic training data, lack of data, or statistical logics, and these have a propensity to disadvantage already underserved minority groups, thus aggravating societal inequalities (Ananny, 2019).

Despite such risks, AI is increasingly applied to healthcare processes ranging from disease diagnosis, treatment, and risk evaluation to allocating resources (Chin et al., 2023). Studies have specifically explored how AI biases can lead to disproportionate or discriminatory healthcare access. Eubanks (2018) observed that the application of predictive models in welfare systems discriminates against the underserved community by denying access to medical aids and resources. Another study found that an algorithm widely used in the health industry to target patients for high-risk care management significantly disadvantaged Black patients seeking care and medical resources (Obermeyer, Powers, Vogeli, & Mullainathan, 2019). Many studies have found a similar pattern in the medical field (e.g., Ahmed et al., 2021; Straw & Wu, 2022; Vyas, Eisenstein, & Jones, 2020). The lawsuit introduced in the following section vividly illustrates how AI biases can inflict real-world harm to vulnerable populations when applied to healthcare decisions.

Class Action Lawsuit Filed Against UnitedHealthcare

In November 2023, a group of plaintiffs filed a class action lawsuit against UnitedHealthcare, one of the largest insurance companies in the United States. The plaintiffs in this lawsuit were elderly individuals with severe illnesses or injuries who were denied "post-acute care." The plaintiffs were entitled to this care through "Medicare Advantage Plans," an insurance plan provided by UnitedHealthcare, partnering with Medicare. However, UnitedHealthcare allegedly used an AI model, "nH Predict," instead of human medical experts to make determinations about insurance coverage. The plaintiffs argued that UnitedHealthcare knew that this AI model was blatantly inaccurate as "over 90 percent of patient claim denials are reversed through either an internal appeal process or through federal Administrative Law Judge (ALJ) proceedings" (Complaint, 2023, p. 14); however, the company continued to use it to override the determinations of the patient's physicians.

The plaintiffs of the lawsuit also sued naviHealth—a technology company acquired by a subsidiary of the UnitedHealth Group in 2020—that developed the allegedly defective AI model involved in this lawsuit. On their website, naviHealth promotes the idea that their model can "reduce unnecessary post-acute care spending" (naviHealth, n.d., para. 5). The CEO of Optum and his coauthors also reported in a paper that naviHealth has been successful in "achieving an average of more than 20% cost savings per episode of care and reducing average length of stays in skilled nursing facilities (SNFs) by 15–25%" (Conway, Rosenblit, & Theisen, 2022, p. 5). As promoted by the developers, nH Predict helps reduce costs for insurers by predicting shorter average lengths of stays and fewer admissions in acute care facilities, which might not always guarantee better treatment for patients.

In December 2023, two other plaintiffs filed a lawsuit against the insurance company Humana for using nH Predict to wrongfully deny rehabilitation care for elderly people, which implies that this is not a singular incident caused by one bad-faith actor. Drawing from this example, I address the following research questions: How does the use of AI models in healthcare decision making impact the health outcomes and experiences of marginalized populations? What are the implications of AI errors or erroneous use of AI in relation to public health? How does the introduction of AI complicate the issue of liability when negative consequences occur?

AI Errors, Ageism, and Public Health

Ageism differs from other "-isms" (e.g., racism, sexism) in that everyone is susceptible to it during their life course, yet it tends to be more socially accepted and challenged less often compared to other types of discrimination (World Health Organization [WHO], 2021). Ageism against older individuals is prevalent in our current society, and it severely impacts health outcomes such as declining physical/mental health and shorter lifespans (WHO, 2021). Moreover, the use of AI in medicine can exacerbate or introduce new forms of ageism. Several factors, including the lack of representation of elderly experiences in AI training data, uneven knowledge/access levels between younger and older generations (i.e., lack of algorithmic awareness), and lacking input and oversight from the elderly population in AI design can amplify ageism (WHO, 2022).

Likewise, nH Predict reinforces inequalities and age-based discrimination embedded in our current system, as the algorithm does not account for factors beyond clinical conditions such as family support, access to housing, or financial affordability (Obermeyer, 2023, as cited in Ross & Herman, 2023). For instance, elderly patients who do not have a safe and sheltered home to return to or those who do not have family members to assist them after they are discharged from the hospital might require additional days at nursing facilities. However, AI models' determinations of optimal length of stay in SNFs are based on a database of millions of patient records that may not accurately represent lived elderly experiences and lack contextual information for each individual. Despite such shortcomings, case managers from naviHealth were instructed to adhere to the algorithm's projections rather than accounting for multiple factors, such as the patient's health conditions and their life situations (Ross & Herman, 2023).

Moreover, the use of AI models in medical claim assessments increases elderly patients' burden of denial appeals. First, appeal processes are made more challenging due to the proprietary nature of AI models and their opacity. In the lawsuit against UnitedHealthcare, the plaintiffs claimed that they were denied access to nH Predict reports when they requested them, as that information was considered proprietary by the company. Appealing for AI-assisted claim decisions can be particularly difficult for elderly patients as they are usually in "impaired conditions," have a "lack of knowledge," and "lack of resources to appeal the erroneous AI-powered decisions" (Complaint, 2023, p. 2). Even if the patients could get access to reports produced by an AI model, it is highly unlikely that they would be able to decipher the results and dispute to what extent the final decisions were affected by AI's predictions. Typically, only 0.2% of denied health insurance claims are appealed by patients (Pollitz, Lo, Wallace, & Mengistu, 2023), and the use of AI may drop the numbers even lower.

AI Complicates the Issue of Liability and Responsibility

Some might argue that bad-faith actors have always existed in the insurance industry. However, the introduction of AI into the healthcare system has further complicated the issue of liability by creating more incentives and loopholes for bad-faith actors and by shifting responsibilities. First, AI tools can be used to produce standardized guidelines that justify insurers cutting off care. Ross and Herman (2023) reported that UnitedHealthcare pressured employees to strictly adhere to nH Predict recommendations for length of stays regardless of patients' circumstances. They revealed an internal document showing that

the company's target goal for 2022 was to keep the patients' stay in nursing facilities within a 3% variance of the days predicted by the AI model. naviHealth even tightened their performance goal to a variance of under 1% in 2023, and their former employee testified that failing to meet the target could result in disciplinary actions, including potential termination (Ross & Herman, 2023). As such, predictions from AI models can function as guidelines to make quicker and stricter decisions within insurance companies, thus driving the number of claim denials.

Second, the lack of regulations for AI use in health industries or laissez-faire regulations creates a potential for profit-driven commercial companies to abuse these technologies. The Centers for Medicare & Medicaid Services (CMS) established a federal register stating:

> MA [Medicare Advantage] organizations must ensure that they are making medical necessity determinations based on the circumstances of the specific individual, as outlined at § 422.101(c)[1], as opposed to using an algorithm or software that doesn't account for an individual's circumstances. (Medicare Program, 2023, p. 22195)

In a memorandum that the CMS sent out after a series of lawsuits were filed for wrongful use of AI, they stated that "MA organizations should, prior to implementing an algorithm or software tool, ensure that the tool is not perpetuating or exacerbating existing bias, or introducing new biases" (The Centers for Medicare & Medicaid Services, 2024, p. 3). As is evident in these rules, the regulators have shifted the responsibility of monitoring the algorithms to the MA organizations (i.e., insurers). However, letting those who will profit most from using AI models supervise its biases or potential misuse is like a fox guarding the henhouse.

Furthermore, the liability of those who developed the AI models is ambiguous when the algorithm results in negative health outcomes. As mentioned earlier, the plaintiffs sued the developer of nH Predict, arguing that they "intended for it to save insurance companies money" (Complaint, 2023, p. 8). In this specific case, the company (naviHealth) that developed the algorithm is a subsidiary of the insurer (UnitedHealthcare), which makes it easier to draw the connections between their motives and the resultant consequences. However, it is still questionable whether we can punish developers for building a model that predicts lower costs over better treatment. How can we assume liability if the developers allegedly did not know that the use of their AI model would lead to negative outcomes? More fundamentally, are we just being more sensitive to AI's decision fairness (Jones-Jang & Park, 2022) and shifting the blame to AI when, in actuality, humans are more prone to error (Shen et al., 2019)?

Conclusion

The lawsuit against UnitedHealthcare serves as a critical example for scrutinizing the societal implications of applying AI tools in healthcare, especially concerning the elderly population. Central to the

[1] §422.101(c) states that MA organizations must make medical necessity determinations based on multiple factors including the patient's medical history, physician's recommendations, and clinical notes (Requirements relating to basic benefits, 2021).

lawsuit was the allegation that erroneous AI models were used to deny elderly patients necessary care without accounting for contextual factors that might vary by individual. Such practice not only endangers patients' health by delaying essential care but also increases their burden to appeal for claim denials. As such, this case study underscores the potential risks of using AI models for health-related decisions in that it can reinforce and exacerbate inequalities, which underserved minority groups are more susceptible to.

However, in this example, it is not accurate to say that the AI was erroneous; it was rather the humans who "erroneously used" the AI model. If the end goal of the AI was to predict lower costs, the algorithm was functional or even accurate on the insurer's terms. Furthermore, AI-assisted decisions make it challenging to pinpoint liability when negative consequences arise—whether it rests with insurers, AI developers, or health professionals who make the final claim decisions. The proprietary and black-box nature of algorithms hinders both the patient's ability to appeal for wrongful denials as well as adequate regulatory oversight. However, the series of similar lawsuits suggests that this is a systematic problem requiring more robust regulatory frameworks instead of a laissez-faire operation, especially when it comes to healthcare, which is directly linked to basic human rights. This article calls for the need for improved transparency and accountability in AI regulation, in alignment with ethical healthcare principles, and a reevaluation of how AI tools can be integrated without harming vulnerable populations.

References

Ahmed, S., Nutt, C. T., Eneanya, N. D., Reese, P. P., Sivashanker, K., Morse, M., . . . Mendu, M. L. (2021). Examining the potential impact of race multiplier utilization in estimated glomerular filtration rate calculation on African-American care outcomes. *Journal of General Internal Medicine*, *36*(2), 464–471. doi:10.1007/s11606-020-06280-5

Ajunwa, I. (2020). The paradox of automation as anti-bias intervention. *Cardozo Law Review*, *41*(5), 1671–1742. Retrieved from https://cardozolawreview.com/the-paradox-of-automation-as-anti-bias-intervention/

Ananny, M. (2019). Probably speech, maybe free: Toward a probabilistic understanding of online expression and platform governance. *Knight First Amendment Institute at Columbia University*. Retrieved from https://knightcolumbia.org/content/probably-speech-maybe-free-toward-a-probabilistic-understanding-of-online-expression-and-platform-governance

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Conference on fairness, accountability and transparency*. *Proceedings of Machine Learning Research*, *81*, 77–91. Retrieved from https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

The Centers for Medicare & Medicaid Services. (2024, February 6). *Frequently asked questions related to coverage criteria and utilization management requirements in CMS final rule (CMS-4201-F)*. American Hospital Association. Retrieved from https://www.aha.org/system/files/media/file/2024/02/faqs-related-to-coverage-criteria-and-utilization-management-requirements-in-cms-final-rule-cms-4201-f.pdf

Chin, M. H., Afsar-Manesh, N., Bierman, A. S., Chang, C., Colón-Rodríguez, C. J., Dullabh, P., . . . Ohno-Machado, L. (2023). Guiding principles to address the impact of algorithm bias on racial and ethnic disparities in health and health care. *JAMA Network Open*, *6*(12), 1–13. doi:10.1001/jamanetworkopen.2023.45050

Chu, C. H., Leslie, K., Shi, J., Nyrup, R., Bianchi, A., Khan, S. S., . . . Grenier, A. (2022). Ageism and artificial intelligence: Protocol for a scoping review. *JMIR Research Protocols*, *11*(6), 1–9. doi:10.2196/33211

Complaint at Estate of Gene B. Lokken et al. v. UnitedHealth Group, Inc. et al. (2023, November 14). Retrieved from https://litigationtracker.law.georgetown.edu/wp-content/uploads/2023/11/Estate-of-Gene-B.-Lokken-et-al_20231114_COMPLAINT.pdf

Conway, P. H., Rosenblit, A., & Theisen, S. (2022). The future of home and community care. *NEJM Catalyst Innovations in Care Delivery*, *3*(4), 1–11. doi:10.1056/CAT.22.0141

Datta, A., Tschantz, M. C., & Anupam, D. (2015, March). Automated experiments on ad privacy settings. *Proceedings on Privacy Enhancing Technologies*, *2015*(1), 92–112. doi:10.48550/arXiv.1408.6491

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. New York, NY: St. Martin's.

Jones-Jang, S. M., & Park, Y. J. (2022). How do people react to AI failure? Automation bias, algorithmic aversion, and perceived controllability. *Journal of Computer-Mediated Communication*, *28*(1), 1–8. doi:10.1093/jcmc/zmac029

Medicare Program; Contract Year 2024 Policy and Technical Changes to the Medicare Advantage Program, Medicare Prescription Drug Benefit Program, Medicare Cost Plan Program, and Programs of All-Inclusive Care for the Elderly, 88 F.R. 22120 (to be codified at 42 C.F.R § 417, § 422, § 423, § 455, § 460). (2023). Retrieved from https://www.federalregister.gov/d/2023-07115

naviHealth (n.d.). *Post-acute care management*. Retrieved from https://navihealth.com/solutions/post-acute-care-management/

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York: New York University Press.

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

Pollitz, K., Lo, J., Wallace, R., & Mengistu, S. (2023, February 9). *Claims denials and appeals in ACA marketplace plans in 2021*. Retrieved from https://www.kff.org/private-insurance/issue-brief/claims-denials-and-appeals-in-aca-marketplace-plans/

Requirements relating to basic benefits, 45 C.F.R. § 422.101. (2021). Retrieved from https://www.ecfr.gov/current/title-42/section-422.101

Ross, C., & Herman, B. (2023, November 14). UnitedHealth pushed employees to follow an algorithm to cut off Medicare patients' rehab care. *STATnews*. Retrieved from https://www.statnews.com/2023/11/14/unitedhealth-algorithm-medicare-advantage-investigation/

Shen, J., Zhang, C. J. P., Jiang, B., Chen, J., Song, J., Liu, Z., . . . Ming, W. K. (2019). Artificial intelligence versus clinicians in disease diagnosis: Systematic review. *JMIR medical informatics*, *7*(3), 1–15. doi:10.2196/10010

Straw, I., & Wu, H. (2022). Investigating for bias in healthcare algorithms: A sex-stratified analysis of supervised machine learning models in liver disease prediction. *BMJ health & care informatics*, *29*(1), 1–8. doi:10.1136/bmjhci-2021-100457

Vyas, D. A., Eisenstein, L. G., & Jones, D. S. (2020). Hidden in plain sight—reconsidering the use of race correction in clinical algorithms. *New England Journal of Medicine*, *383*(9), 874–882. doi:10.1056/NEJMms2004740

World Health Organization. (2021, March 18). *Global report on ageism*. Retrieved from https://www.who.int/publications/i/item/9789240016866

World Health Organization. (2022, February 9). *Ageism in artificial intelligence in health*. Retrieved from https://www.who.int/publications/i/item/9789240040793