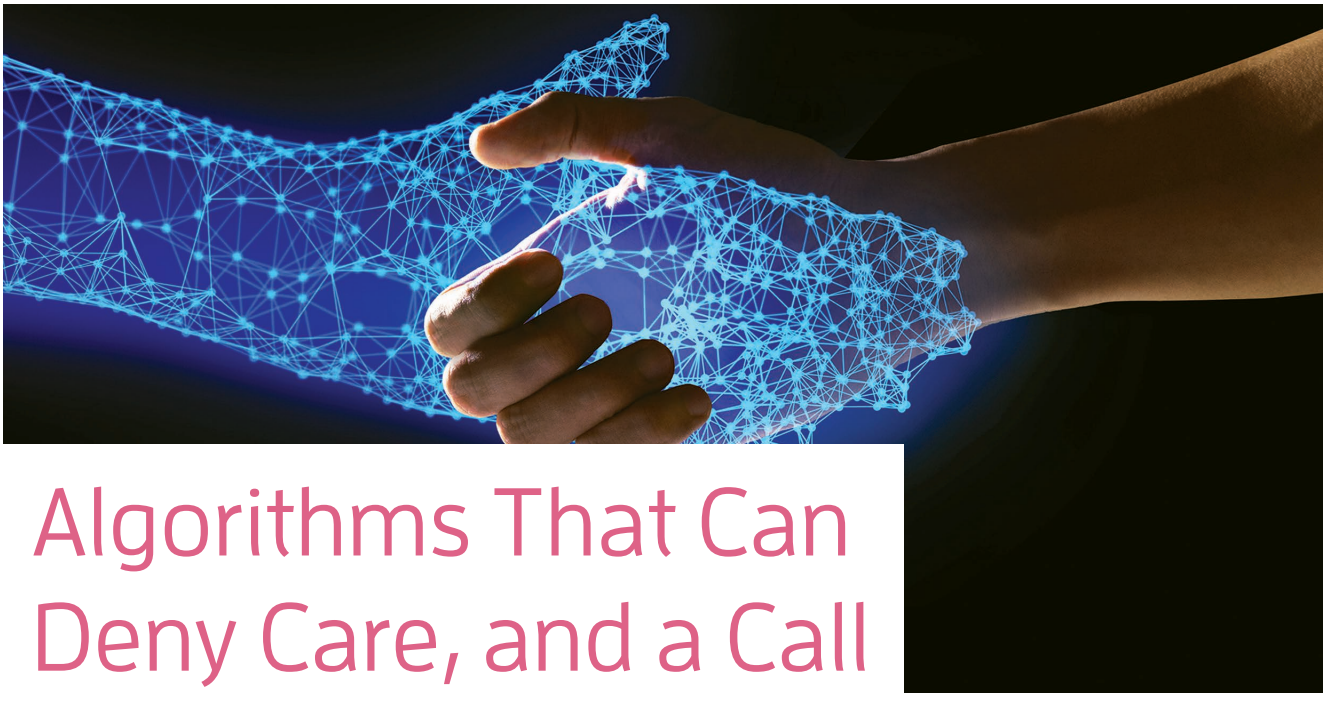
EDITOR DOMENICO TALIA
University of Calabria; talia@dimes.unical.it



Algorithms That Can Deny Care, and a Call for AI Explainability

Domenico Talia, University of Calabria

Recent examples of negative use of big data and machine learning come from their use in health care decisions involving a large number of patients. These cases highlight the need for algorithm explainability to help us better understand how artificial intelligence works in solving problems so we can then evaluate its accuracy and effectiveness.

In the first article published in this column,[1] we discussed several success stories about artificial intelligence (AI) and machine intelligence solutions effectively used in health care for detecting diseases, supporting doctors, helping patients, improving health-care processes, and saving lives. Together with many cases where big data, machine learning algorithms, and AI systems are improving people's diagnoses and treatments, recently raised examples of negative use of these technologies in real scenarios involving a large number of patients. The use of faulty data analysis and machine learning is becoming a problem for patients and the health-care industry because of problematic uses due to profit goals, algorithmic racial bias in patient care,[2] and ethical issues.

A CLASS ACTION LAWSUIT THAT INVOLVES AN ALGORITHM

Recently, a class action lawsuit was filed against UnitedHealth and a subsidiary based on the hypothesis that they are illegally using an algorithm to deny or limit rehabilitation care to seriously ill patients, even though the companies know the algorithm has a high error rate. After one

Digital Object Identifier 10.1109/MC.2024.3387012
Date of current version: 26 June 2024

month, a similar class action lawsuit was filed against Humana for the same algorithm. We must consider that UnitedHealth and Humana are among the top providers of the popular plans for seniors and account for nearly half of all medical assistance enrollees in the United States.

The software system used by UnitedHealth Group and Humana is the nH Predict AI tool that uses a database of a few million patients compiled over years.

According to the filed lawsuits, the United States's largest health insurance companies pressured their medical staff to reduce payments for very ill patients at large scale, exploiting an algorithm that calculates the needed rehabilitation days and limits rehabilitation care for older and disabled Americans while profits increased.[3] On the other hand, UnitedHealth and Humana state that the algorithm they used, which predicts how long patients will need to stay in rehabilitation, is just a tool used to estimate recoveries. They say the decision is made by (human) managers and employees. Some newspapers in the United States wrote that managers must follow the algorithm predictions precisely so payment could be discontinued by the date it predicted.

The software system used by United Health Group and Humana is the nH Predict AI tool that uses a database of a few million patients compiled over years. The algorithm of nH Predict uses this large database to analyze a patient's diagnosis, living situation, age, gender, physical function, admission date, and other information for predicting how much postacute care a patient "should" need. According to the lawsuits, the algorithm settles on the day when UnitedHealthcare or Humana will cut payment for patient care. nH Predict also considers the patient's usual living setting, that is, if they are at home alone or in an assisted living facility.

The outcome report provided by nH Predict provides a sort of profile of each patient that includes a score for a few of the patient's functions also based on the data of similar patients analyzed in the past. The profile includes scores on the patient's basic mobility, such as wheelchair skills or ability to take the stairs, cognitive abilities, such as memory and communication, and daily activity (for example, dressing and bathing). The profile report produced by nH Predict includes a total average score for the patient that is based on a combination of single scores.

When the nH Predict patient report is ready, physicians and company managers use it for making evidence-based decisions on a rehabilitation plan for the patient based on personal medical condition and functional necessities. It is evident that the role of the prediction algorithm in this scenario is very important and significantly contributes to the final decision about the rehabilitation duration.

Unfortunately, the machine learning algorithm used in the nH Predict system is proprietary and not disclosed in detail by the owner NaviHealth. Thus it is unclear how the nH Predict system works exactly, that is how the patient data are used together with the information of other similar patients to estimate medical needs, length of stay, and discharge date. In a few words, the nH Predict algorithm suggests a decision to the health insurance company manager to process patient data in a way that is not known to patients and their families. In fact, when patients or their physicians have requested to know how the nH Predict's reports are generated, as expected, UnitedHealth has denied their requests, telling them the algorithm is proprietary. When prescribing physicians disagree with UnitedHealth's algorithm-based determination of how much postacute care their patients need, their judgments are ignored.

Doctors and patients in those cases are neither aware of the algorithm procedures, nor able to question their data processing strategies and final decisions. In other words, insurers use opaque (for the public) predictive algorithms that are becoming progressively influential in decisions about patient care and coverage. These cases call for algorithm explainability, an issue that in the area of AI is becoming urgent.

AN APPEAL TO EXPLAINABILITY

As discussed by de Franco et al.,[4] a problem may occur when algorithms make important decisions that affect people or are coded to influence humans providing suggestions and information to the public. This is not only the case of the nH Predict system, but it may occur with many algorithms, such as social media profiling systems, financial intelligence products, customer profiling, and other algorithms that are designed to increase profits for companies rather than do the best for the common good. In all these cases, individuals and/or society need to know how the algorithms work and question about their real goals. This calls for algorithm explainability that describes what happens in the code from input to output. AI algorithm explainability makes learning models transparent and solves the "black box" problem. To reach this goal, auditing of algorithms can be effective to guarantee that they are legal, ethical, and safe.[5,6] Algorithm auditors can look at the inputs and outputs of a decision system from the outside.

Explainable AI (XAI) defines a set of methods and techniques that refer to

learning algorithms that humans can comprehend and trust together with the results/decisions they generate. XAI is the opposite of black box systems where owners and users cannot (or don't want to) explain why an AI algorithm reached a certain decision or a specific result. Black box AI systems designed for decision-making map data features into a class by predicting events or the behavioral traits of citizens without providing the reasons why they did it. This opaque approach is elusive not only for the lack of transparency but also for possible biases the black box algorithms inherit from human discrimination and/or errors originating (often hidden) in the training data, which may lead to unjust or erroneous decisions.[7]

AI models used in the United States by physicians to detect diseases such as cancer or mental diseases, or suggest the most effective treatment, are assessed by the Food and Drug Administration. But tools used by insurers for determining whether hospitalization or treatments should be reimbursed are not subjected to the same inspection; however, they are equally important for patients. In Europe, the impact of algorithm decisions is regulated under the EU's General Data Protection Regulation (GDPR). A professional using personal data for automated processing must explain to the people concerned how the system makes decisions. The individual data subject (for instance, the person who was rejected for a payment or a patient who was denied a certain medical service) has the right to ask the company why it made the decision it did, and the company must then explain how the system came to its decision. If the company can't explain the decision in response to an individual's request, it would not be complaint with the GDPR and can be prosecuted by law.

For instance, Recital 71 of GDPR[8] states:

> "The data subject should have the right not to be subject to a decision, which may include a measure, evaluating personal aspects relating to him or her which is based solely on automated processing and which produces legal effects concerning him or her or similarly significantly affects him or her, such as automatic refusal of an online credit application or e-recruiting practices without any human intervention. ... In any case, such processing should be subject to suitable safeguards, which should include specific information to the data subject and the right to obtain human intervention, to express his or her point of view, to obtain an explanation of the decision reached after such assessment, and to challenge the decision."

In addition, Article 22 grants an individual a "right of human intervention." Under this right, an individual may ask for a human to review the AI algorithm's decision to determine whether or not the system made a mistake.

An approach similar to that one used in EU GDPR should be embodied in laws and regulations in other countries allowing people to have clear explanation of how algorithms work, how they use their data, and the rules they follow to make decisions.

NEW SKILLS AND ABILITIES TO FACE AI-BASED DECISIONS

Together with the increasing role of algorithms in our daily lives and with the increasingly widespread use of AI in decision making, we can witness that the scarcity of digital culture and computational thinking becomes an ever-greater handicap for individuals and for the entire society. The digital illiteracy widely spread in the population risks of being unaware users of "intelligent" systems that make decisions about us. The digital tsunami is advancing, and citizens must get skills and abilities to handle it, otherwise they will be overwhelmed.

> It is evident that the role of the prediction algorithm in this scenario is very important and significantly contributes to the final decision about the rehabilitation duration.

People skill and knowledge on AI and the most promising areas of digital technology show a delay that, if it will not be filled, will be another key element for the failure to progress. We must not work to train everyone to become a software designer, but we must understand that it is essential to provide basic knowledge to deal with apps, software tools, and platforms that use AI applications. We need concrete and widespread actions to increase expertise of young people in information technology and put the citizens in condition of living the present time with the necessary skills. Otherwise, we risk the rising of a community of people who will be passive entities instead of being active subjects in a world where AI is becoming pervasive.

To address these issues, it is necessary to build a widespread basic training that makes people understand, for example, what AI is, how computers work in solving problems, being able to evaluate information online, its truthfulness, and its usefulness. It serves to provide young and old people the opportunity to understand the computational processes, the logic that governs them, and their potential that can be exploited in work and in everyday life. While everyone agrees on the obvious need to teach mathematics from elementary schools, it should be

equally normal to teach the principles of computer science in all schools. Kids need to know how to understand and compose simple algorithms, computational thinking, and thinking methodically to strategies useful for defining solutions. They need to know how AI works and how to cooperate with it as intelligent actors and not to be subjected to it. On the other hand, adults must be able to discuss and understand how AI algorithms are used when they concern decisions about them, their rights, their health, and their work.

The widespread use of AI solutions in daily life is creating a large social and cultural divide between the few experts and the many oblivious users. This separation, other than creating new inequalities, will cause much damage to our society. According to a recent study of the World Economic Forum, employers estimate that 44% of workers' skills will be disrupted in the next four years. It is necessary to take steps to manage the future that arrived too quickly. Politics, schools, and universities are called to face this main challenge. If this will not be done now, enormous problems will arise in the coming decades, when the intensive use of new intelligent technologies will create many new jobs that do not exist today and will abolish many of those being done today. At the same time, private companies using AI in their decision processes about clients, should realize that they must provide clear explanation of automated decisions. We are now also experiencing the new big data divide that refers to the split between those who have access and ownership to large and distributed data sources and those that do not. Andrejevic[9] worked on the concepts of access, ownership, and use of digital data, arguing that a big data divide is occurring today that deals with the asymmetric relationship between those who collect, store, and mine large quantities of data, and those whom data collection targets.

> Insurers use opaque (for the public) predictive algorithms that are becoming progressively influential in decisions about patient care and coverage.

The issues involved in this asymmetric relationship relate to a basic divide that occurs for people and organizations without the technical means to access and analyze data stored on the Internet or in large data repositories. To fill this gap is essential to work on new cognitive paradigms, to change education programs and practices, and to offer the opportunity to people to access data and learn how to use them successfully and how them are used in AI systems to profile them or to make decisions on their lives. This integrated strategy could result in new opportunities for those who wish to have an active role in the new digital society, offer a way to defend people rights, and can improve the global knowledge of people on intelligent technologies.

REFERENCES

1. D. Talia, "Machine intelligence for human health: A few noteworthy cases," *Computer*, vol. 55, no. 10, pp. 82–86, Oct. 2022, doi: 10.1109/MC.2022.3190786.
2. D. M. Christensen, J. Manley, and J. Resendez, "Medical algorithms are failing communities of color," Health Affairs, Washington, DC, USA, Sep. 2021. [Online]. Available: https://www.healthaffairs.org/content/forefront/medical-algorithms-failing-communities-color
3. C. Ross and B. Herman. "Denied by AI: How Medicare Advantage plans use algorithms to cut off care for seniors in need." STAT. Accessed: Apr. 15, 2024. [Online]. Available: https://www.statnews.com/2023/03/13/medicare-advantage-plans-denial-artificial-intelligence/
4. J. F. DeFranco, J. Voas, and N. Kshetri, "Algorithms: Society's invisible puppeteers," *Computer*, vol. 55, no. 4, pp. 12–14, Apr. 2022, doi: 10.1109/MC.2021.3128675.
5. D. Talia, *From Algorithms to Thinking Machines: The New Digital Power*. New York, NY, USA: ACM, 2023.
6. M. Jovanović and M. Schmitz, "Explainability as a user requirement for artificial intelligence systems," *Computer*, vol. 55, no. 2, pp. 90–94, Feb. 2022, doi: 10.1109/MC.2021.3127753.
7. N. Prentzas, A. Kakas, and C. S. Pattichis, "Explainable AI applications in the medical domain: A systematic review," 2023, *arXiv:2308.05411*.
8. "General data protection regulation." Intersoft Consulting. Accessed: Feb. 1, 2024. [Online]. Available: https://gdpr-info.eu/
9. M. Andrejevic, "Big data, big questions: The big data divide," *Int. J. Commun.*, vol. 8, no. 1, pp. 1673–1689, 2014.

DOMENICO TALIA is a full professor of computer engineering at the University of Calabria, 87036 Rende, Italy. Contact him at talia@ dimes.unical.it.