

Conflict of Interest Disclosures: The author has completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest and none were reported.

1. Mez J, Daneshvar DH, Kiernan PT, et al. Clinicopathological evaluation of chronic traumatic encephalopathy in players of American football. *JAMA*. 2017; 318(4):360-370.
2. Corrigan JD, Bogner J. Initial reliability and validity of the Ohio State University TBI identification method. *J Head Trauma Rehabil*. 2007;22(6): 318-329.

In Reply Dr Zuckerman and colleagues are concerned that the definition of CTE used in our study was overly sensitive. The definitive diagnosis of Alzheimer disease, Lewy body disease, or CTE is based on strictly defined neuropathologic criteria.¹⁻³ In CTE and Lewy body disease, there is a pathognomonic lesion or a unique pathologic feature that is specific to the disorder and not found in healthy control patients. In Lewy body disease, the pathognomonic feature is the Lewy body, and a diagnosis is made even if only a single Lewy body is found.^{2,4} In CTE, the pathognomonic lesion is a cluster of tau-immunostained neurofibrillary tangles and neurites around a small vessel,³ and again, 1 lesion is sufficient for a pathologic diagnosis. The situation is analogous to finding a single microscopic focus of tumor—even a single small focus of tumor is sufficient for a diagnosis. Of the 177 CTE cases reported in our study, only 11 (6%) had 1 or 2 isolated CTE lesions and were considered stage I CTE; most cases (133 [75%]) showed advanced pathology and were stage III and IV. Unlike CTE and Lewy body disease, the diagnosis of Alzheimer disease requires the presence of 2 pathologic features, amyloid- β plaques and neurofibrillary tangles, distributed throughout the brain. Small amounts of amyloid- β plaques or neurofibrillary tangles are not sufficient for the diagnosis of Alzheimer disease.

Criteria exist for the clinical diagnoses of probable and possible Alzheimer disease and Lewy body disease, and criteria for traumatic encephalopathy syndrome have been proposed,⁵ but for all 3 disorders, the clinical diagnosis is not definitive and does not contribute to the pathologic diagnosis.

Zuckerman and colleagues and Dr Yari suggest that other diagnoses may have been responsible for the participants' symptoms. All the clinical evaluations for CTE in our study were conducted retrospectively after death with an informant. Of the CTE cases, 91% had behavior or mood symptoms in the year before death and 43% initially presented with mood or behavior symptoms. There is no claim in the article that the reported symptoms were due to CTE pathology alone. Moreover, participants who initially presented with mood or behavior symptoms commonly had a family history of psychiatric illness, suggesting that "CTE τ pathology may lower the threshold for psychiatric manifestations in susceptible individuals."

Although football data were collected via an online questionnaire beginning in 2014, all data gathered about TBI and concussions were only obtained via postmortem telephone interviews with informants by a clinician. These data were collected in an unstructured format prior to 2014 and via a structured telephone questionnaire after 2014. In the

structured telephone questionnaire, the clinician asked about concussion number after reading a lay definition of concussion to the informant.⁶ The median number of concussions reported was substantially higher after 2014, suggesting that reading the definition led to inclusion of milder concussions. We acknowledged the possibility of recall bias due to the retrospective nature of data collection in the article.

Ann C. McKee, MD

Jesse Mez, MD, MS

Bobak Abdolmohammadi, BA

Author Affiliations: Boston University Alzheimer's Disease and CTE Center, Boston University School of Medicine, Boston, Massachusetts.

Corresponding Author: Ann C. McKee, MD, Neuropathology Service, Veterans Affairs Boston Healthcare System, 150 S Huntington Ave, Boston, MA 02118 (amckee@bu.edu).

Conflict of Interest Disclosures: The authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Dr McKee reported that she has received funding from the National Football League (NFL) and World Wrestling Entertainment, is a member of the Mackey-White Committee of the NFL Players' Association, and has received honoraria for speaking engagements. No other disclosures were reported.

1. Montine TJ, Phelps CH, Beach TG, et al; National Institute on Aging; Alzheimer's Association. National Institute on Aging-Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease: a practical approach. *Acta Neuropathol*. 2012;123(1):1-11.
2. McKeith IG, Galasko D, Kosaka K, et al. Consensus guidelines for the clinical and pathologic diagnosis of dementia with Lewy bodies (DLB): report of the Consortium on DLB International Workshop. *Neurology*. 1996;47(5):1113-1124.
3. McKee AC, Cairns NJ, Dickson DW, et al; TBI/CTE group. The first NINDS/NIBIB consensus meeting to define neuropathological criteria for the diagnosis of chronic traumatic encephalopathy. *Acta Neuropathol*. 2016;131(1):75-86.
4. Crane PK, Gibbons LE, Dams-O'Connor K, et al. Association of traumatic brain injury with late-life neurodegenerative conditions and neuropathologic findings. *JAMA Neurol*. 2016;73(9):1062-1069.
5. Montenegro PH, Baugh CM, Daneshvar DH, et al. Clinical subtypes of chronic traumatic encephalopathy: literature review and proposed research diagnostic criteria for traumatic encephalopathy syndrome. *Alzheimers Res Ther*. 2014;6(5):68.
6. Robbins CA, Daneshvar DH, Picano JD, et al. Self-reported concussion history: impact of providing a definition of concussion. *Open Access J Sports Med*. 2014;5:99-103.

Benefits and Risks of Machine Learning Decision Support Systems

To the Editor Machine learning decision support systems (ML-DSS) have been touted as a key driver of major changes in the way medicine is practiced.¹ Although most articles describe the benefits of incorporating these techniques, Dr Cabitza and colleagues² rightly cautioned that there are potential risks as well as benefits. They pointed to (1) the risk of "deskilling," in which physicians will become unable to recognize when the algorithms are incorrect, (2) errors that the systems themselves make because of inability to address the whole context of care, (3) the intrinsic uncertainty of medical data, and (4) the inscrutability of the black box of many algorithms.

These cautions are well taken but they are similar to those found with the introduction of other innovations. As

described by Crenner,³ the blood pressure cuff was initially resisted because physicians thought that it could not possibly be as accurate as they were in palpating patients' pulses. More than 50 years passed before blood pressure machines were routinely used. Galletta et al⁴ showed that when spell-checkers are incorrect, even highly verbal students may do worse than without the spell-checker. Most people would agree that these tools can be helpful when used properly.

Machine learning also requires proper use to avoid some of the pitfalls described by Cabitza and colleagues.² Strategies to ensure proper use include embedding warning mechanisms to identify unreliable data; creating ancillary systems that incorporate the context; modifying physician training to include best practices in the use of ML-DSS (eg, only using them for the specific purpose for which they were designed); using the systems for decision support, not decision making; and constantly improving ML-DSS to tackle unexpected consequences.

When new tools are introduced, they are often dismissed, as the blood pressure cuff was. As they improve and there is advocacy for their use in clinical practice, the scrutiny increases, as it should. Machine learning algorithms, clinical decision support systems, and bioinformatics approaches for genomic medicine are probably at the stage of increased examination of strategies and consequences. As more research accumulates and the systems themselves improve (similar to the automated blood pressure cuff and the spell-checker), most likely their routine use will be accepted.

Although the cautions raised in the Viewpoint by Cabitza and colleagues² should not be ignored, they also should not be used to inhibit the development of potentially innovative systems that can improve clinical care.

Eta S. Berner, EdD
Bunyamin Ozaydin, PhD

Author Affiliations: Department of Health Services Administration, University of Alabama, Birmingham.

Corresponding Author: Eta S. Berner, EdD, Graduate Programs in Health Informatics, Department of Health Services Administration, University of Alabama, 1716 Ninth Ave S, Room 590J, Birmingham, AL 35294 (eberner@uab.edu).

Conflict of Interest Disclosures: The authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest and none were reported.

1. Obermeyer Z, Emanuel EJ. Predicting the future: big data, machine learning, and clinical medicine. *N Engl J Med*. 2016;375(13):1216-1219.
2. Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. *JAMA*. 2017;318(6):517-518.
3. Crenner CW. Introduction of the blood pressure cuff into US medical practice: technology and skilled practice. *Ann Intern Med*. 1998;128(6):488-493.
4. Galletta DF, Durcikova A, Everard A, Jones BM. Does spell-checking software need a warning label? *Communications*. 2005;48(7):82-86.

To the Editor We would like to contribute to the discussion on the unintended consequences of machine learning in medicine¹ by suggesting that the unintended consequences can be viewed as opportunities to drive methodological changes in ML-DSS to improve health care.

The risks related to overreliance on ML-DSS owe primarily to the lack of methodological transparency of deep-learning algorithms. Why they work or do not work in certain cases is not clear. For example, the Deep Patient program at Mount Sinai Hospital (New York, New York) proved accurate at predicting many diseases but the research group was unable to explain how it worked.² New machine learning-based training is needed to provide physicians with the conceptual skills to interpret the output of algorithmic decision support systems in light of the known challenges.

Big data analytics typically provide only the final output. Careful modeling assumptions, as in the case of weather forecasting,³ serve primarily to discard data that may hinder the accuracy of the prediction with respect to a given context. New methodological research is needed to determine whether this approach applies to medicine as well. If this aspect of modeling does not apply to medicine, and given that all algorithms are biased,^{4,5} ways to open the machine learning black boxes must be found by focusing on basic research principles to increase output accuracy and reliability while minimizing uncertainty.

Thus, ML-DSS must be paired with robust methods to enlighten the relevant aspects of the black boxes. This is particularly relevant for rare pathologies in which spurious correlations with uncommon conditions may lead to inaccurate classifications.

We agree with Dr Cabitza and colleagues that much needs to be done before ML-DSS can be safely used; however, we do not consider the methodological difficulties as reasons to shy away from the theoretical challenges created by this technology. To the contrary, we think that taking up the challenge will improve diagnostic models and contribute to a deeper understanding of the role of uncertainty in medicine.

Lisa Licitra, MD
Annalisa Trama, MD, PhD
Hykel Hosni, PhD

Author Affiliations: Fondazione IRCCS Istituto Nazionale dei Tumori, University of Milan, Milan, Italy (Licitra); Evaluative Epidemiology Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy (Trama); Department of Philosophy, University of Milan, Milan, Italy (Hosni).

Corresponding Author: Lisa Licitra, MD, Medical Oncology Unit 3, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Via Venezia 1, 20133 Milan, Italy (lisa.licitra@istitutotumori.mi.it).

Conflict of Interest Disclosures: The authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Dr Licitra reported receiving grants and personal fees from EISAI Co, Merck Sharp and Dohme, Merck Serono, Boehringer Ingelheim, Novartis, AstraZeneca, and Roche; personal fees from Bristol-Myers Squibb, Debiopharm, and Sobi; and travel expenses from Merck Sharp and Dohme, Merck Serono, Bayer, Debiopharm, and Sobi. No other disclosures were reported.

1. Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. *JAMA*. 2017;318(6):517-518.
2. Knight W. The dark secret at the heart of AI. <https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/>. Accessed August 10, 2017.
3. Hosni, H, Vulpiani, A. Forecasting in light of big data [published online May 26, 2017]. *Philosophy Technol*. doi:10.1007/s13347-017-0265-3
4. O'Neil K. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York, NY: Penguin Random House; 2016.
5. Pasquale F. *The Black Box Society*. Cambridge, Massachusetts; London, England: Harvard University Press; 2015.

To the Editor We offer a counterpoint to the views of Dr Cabitza and colleagues¹ on the adverse effects of machine learning in medicine. We argue that the negative consequences described by the authors are more often a product of misuse of machine learning, rather than anything intrinsic to its methods. Moreover, new approaches developed in the past decade make it easy to avoid some of the discussed problems.

First, the original publication of the pneumonia mortality model² cited by Cabitza and colleagues makes it clear that the errors arose from misapplication rather than lack of treatment context. The model was trained using mortality data of patients who received usual care, but it was applied to predict mortality in patients conditioned on being sent home. It is not surprising that this mismatch resulted in errors. In fact, the mismatch was carefully identified and discussed by the original investigators, who chose to make an explicit assumption that the difference would be negligible in patients with low predicted mortality. The assumption turned out to be wrong, but the errors it caused were not unforeseen. It can be easy to apply a model in conditions mismatched to its training scenario, which is why skillful design and thorough testing are essential.

Second, we agree that discretizing continuous variables is often an analytic mistake. However, it is common in medical research³ and not unique to machine learning. As Cabitza and colleagues stated, the historically popular practice of modeling diseases as present vs absent, and using clinical experts to add this discrete label to potentially ambiguous training examples, can limit a model's accuracy. But now there is increasing literature on using the disagreement between experts as a meaningful signal rather than a source of error,⁴ and some new learning algorithms do not require labels at all.

Third, new techniques can make any arbitrary model interpretable,⁵ allowing avoidance of the historic trade-off between interpretability and accuracy. However, a model's interpretability is only truly an issue when its accuracy is questionable. If a clinician has experienced high and consistent accuracy with a clinically useful and efficacious model, the interpretability of the model will be less important.

To us, the salient message of this conversation is the need for highly trained medical data scientists who have a deep understanding of both clinical medicine and computational methods.

Thomas A. Lasko, MD, PhD
Colin G. Walsh, MD
Bradley Malin, PhD

Author Affiliations: Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, Tennessee.

Corresponding Author: Thomas A. Lasko, MD, PhD, Department of Biomedical Informatics, Vanderbilt University School of Medicine, 2525 W End Ave, Ste 1475, Nashville, TN 37203 (tom.lasko@vanderbilt.edu).

Conflict of Interest Disclosures: The authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Dr Lasko reported receiving honoraria and travel expenses for lecturing on machine learning from the Harvard T. H. Chan School of Public Health; and a grant from Pfizer for a machine-learning project. Dr Walsh reported serving as a consultant

on machine learning for Florida State University, Johns Hopkins University, and Raiven Healthcare. Dr Malin reported serving as a consultant on privacy and security issues to Google; and receiving honoraria and travel expenses for lecturing on privacy and security from the Harvard T. H. Chan School of Public Health.

1. Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. *JAMA*. 2017;318(6):517-518.
2. Cooper GF, Aliferis CF, Ambrosino R, et al. An evaluation of machine-learning methods for predicting pneumonia mortality. *Artif Intell Med*. 1997;9(2):107-138.
3. Fedorov V, Mannino F, Zhang R. Consequences of dichotomization. *Pharm Stat*. 2009;8(1):50-61.
4. Sharmanska V, Hernández-Lobato D, Hernández-Lobato JM, Quadrianto N. Ambiguity helps: classification with disagreements in crowdsourced annotations. Presented at: 29th IEEE Conference on Computer Vision and Pattern Recognition (CVPR); June 27-30, 2016; Las Vegas, Nevada.
5. Adler P, Falk C, Friedler SA, et al. Auditing black-box models for indirect influence. Presented at: 16th IEEE International Conference on Data Mining (ICDM); December 12-15, 2016; Barcelona, Spain.

To the Editor I disagree with the premises and conclusions of the Viewpoint by Dr Cabitza and colleagues on the unintended consequences of machine learning in medicine.¹

The authors' conclusion raises the bar for artificial intelligence higher than that for new pharmaceuticals, medical devices, or changes in care delivery. Cabitza and colleagues espoused a standard that artificial intelligence must provably affect clinically important and relevant outcomes, as well as satisfy patients and physicians. Yet imaging artificial intelligence tools have already received regulatory approval this year in Europe, Australia, and New Zealand. Well-designed artificial intelligence algorithms exist today for the detection of pulmonary tuberculosis,² breast malignancy,³ and serious brain findings such as stroke, hemorrhage, and mass effects.⁴

The authors' premises unfairly pointed to problems with artificial intelligence that are identical to problems faced by humans. For example, it is true that a poorly designed artificial intelligence study that misses context will fail or produce counterintuitive output. However, the design problem underlying the pneumonia mortality example in the Viewpoint¹ would lead to failure even for a human running a simple logistic regression. Blame should not be extended from one poor design to a class of tools.

Intrinsic uncertainty of predictors and outcomes is a problem, but these affect humans just as much as artificial intelligence. When there is no criterion standard referent, the US Food and Drug Administration simply requires, for example, the use of positive percent agreement instead of sensitivity as a performance measure. The US Food and Drug Administration does not insist on provable improvement over a human diagnosis when no criterion standard exists.

It is also true that good artificial intelligence is often a black box, but so is the physician gestalt that drives much human decision making. Although humans consciously think in 3 dimensions, functional neuronal structures may include 11 dimensional network representations.⁵ Extracting rich interactive visualization from either cognitive process is likely to be difficult to impossible, and may not add to human reassurance.

Deskilling is a longer-term concern when artificial intelligence is implemented in ways that collaborate with humans. But this will either lead to autonomous artificial intelligence or specialization by physicians in other higher value-added, patient-facing skills.

Marco D. Huesch, MBBD, PhD

Author Affiliation: Department of Radiology, Milton S. Hershey Medical Center, Hershey, Pennsylvania.

Corresponding Author: Marco D. Huesch, MBBD, PhD, Department of Radiology, Milton S. Hershey Medical Center, 500 University Dr, Mailcode H-066, Hershey, PA 17033 (mhuesch@pennstatehealth.psu.edu).

Conflict of Interest Disclosures: The author has completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest and reported receiving personal fees and nonfinancial support for consulting on machine learning and artificial intelligence from Ping An of China; and being paid for conducting research on machine learning and deep learning by Pennsylvania State University.

1. Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. *JAMA*. 2017;318(6):517-518.
2. Lakhani P, Sundaram B. Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology*. 2017;284(2):574-582.
3. Teare P, Fishman M, Benzaquen O, Toledano E, Elnekave E. Malignancy detection on mammography using dual deep convolutional neural networks and genetically discovered false color input enhancement. *J Digit Imaging*. 2017;30(4):499-505.
4. Prevedello LM, Erdal BS, Ryu JL, et al. Automated critical test findings identification and online notification system using artificial intelligence in imaging [published online July 3, 2017]. *Radiology*. doi:10.1148/radiol.2017162664
5. Reimann MW, Nolte M, Scolamiero M, et al. Cliques of neurons bound into cavities provide a missing link between structure and function. *Front Comput Neurosci*. 2017;11:48.

To the Editor A consortium of leading artificial intelligence researchers predict that artificial intelligence machines will surgically outperform humans by 2053;¹ therefore, it is essential to consider how these technologies will affect medical care. The unintended consequences of machine learning discussed by Dr Cabitza and colleagues² can be grouped into 2 categories: maladaptive changes in physician work and the forcing of information into machine-interpretable chunks. This analysis suffers from 2 central flaws.

First, the authors presupposed that most currently held medical skills are essential for physicians. The skills necessary for a physician are those that allow him or her to serve as a trusted advisor for patients and to affect the prevention, diagnosis, and treatment of disease. Cabitza and colleagues³ mentioned electrocardiogram reading as currently being a necessary human-performed step in the spectrum of prevention, diagnosis, and treatment; however, computers may soon outperform and replace humans at this task.

Consider the following thought experiment: imagine a world in which any single simple medical task can be replaced with a better, faster, cheaper alternative. Then consider whether ceasing to perform the original test constitutes deskilling. Electrocardiogram reading is unlikely to pass this test; however, a deep understanding of the pathophysiology of cardiovascular disease and quickly bonding with a patient pass easily.

Second, the authors worried that the human factors in medicine (eg, context and uncertainty) will be minimized in favor of information that can be readily converted to data. However, we believe that it is more likely that these factors, which already constitute much of medicine and cannot be easily performed by a machine, will rightly become the focus of physicians. Well-trained physicians were easily able to grasp the necessary contextual intangibles to reject the machine-predicted protective effect of asthma in pneumonia⁴ as cited in the Viewpoint by Cabitza and colleagues. Similar applications of reasoned judgment to machine predictions based on the pathophysiology of disease and the individual patient will become the essential role of the physician.

Physicians generally do not enjoy the type of monotonous tasks devoid of human interaction that are likely to be replaced by machines. The unintended consequence of artificial intelligence in medicine may mean that physicians will be able to focus on the tasks that are uniquely human: building trust-based relationships and applying reason and judgment to complex problems to help individual patients.

Alexander L. Fogel, MBA

Joseph C. Kvedar, MD

Author Affiliations: Stanford University School of Medicine, Stanford, California (Fogel); Department of Dermatology, Massachusetts General Hospital, Boston (Kvedar).

Corresponding Author: Joseph C. Kvedar, MD, Partners Connected Health, Massachusetts General Hospital/Department of Dermatology, Harvard Medical School, 25 New Chardon St, Third Floor, Ste 300, Boston, MA 02114 (jkvedar@partners.org).

Conflict of Interest Disclosures: The authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Dr Kvedar reported being an advisor to Claritas Mindsciences, Mavericks Capital, PureTech, and MD Revolution; receiving grants from Samsung, Hitachi, and Phillips; and receiving personal fees from Qualcomm Life. No other disclosures were reported.

1. Grace K, Salvatier J, Dafoe A, Zhang B, Evans O. When will AI exceed human performance? evidence from AI experts. <https://arxiv.org/abs/1705.08807>. Accessed June 27, 2017.
2. Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. *JAMA*. 2017;318(6):517-518.
3. Tsai TL, Fridsma DB, Gatti G. Computer decision support as a source of interpretation error: the case of electrocardiograms. *J Am Med Inform Assoc*. 2003;10(5):478-483.
4. Caruana R, Lou Y, Gehrke J, et al. Intelligible models for healthcare: predicting pneumonia risk and hospital 30-day readmission. In: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Cham, Switzerland: Springer International Publishing AG; 2015:1721-1730.

In Reply We concur that the solutions mentioned by Drs Berner and Ozaydin and Dr Licitra and colleagues can help reduce the negative consequences of the adoption of oracular (ie, as much accurate as inscrutable) ML-DSS. Solutions to open the black box of ML-DSS can be considered technical, and we expect they will be refined in the future.¹ More important are sociotechnical solutions, which must be endorsed and promoted at the management and policy level. These solutions are partly methodological, as pointed out by Licitra and colleagues, and partly related to training physicians better, as suggested by Berner and Ozaydin.

The letter authors raised some objections to our article. The primary objection was that artificial intelligence applications are not different from other medical technological innovations and will eventually be accepted. The secondary objection was that artificial intelligence should not be blamed for the shortcomings reported, but rather the humans that misuse it, as suggested by Dr Lasko and colleagues and Dr Huesch.

We believe that ML-DSS are different from any other technology adopted in health care to date. It is not simply about augmenting the sensory and perceptual capabilities of physicians and providing them with more or better signs to interpret; rather, it is about giving them advice, as the result of an interpretative process, which physicians neither govern nor necessarily understand, that can condition medical decisions and affect outcomes. Thus, these systems affect learning and knowledge at the core of the medical profession. Although some medical innovations initially hindered by skeptical observers later became essential in practice, how many allegedly innovative medical interventions, which were implemented even when lacking strong evidence, were later withdrawn?²

The second argument has frail foundations because it entails a juxtaposition (artificial intelligence vs humans) that simply does not exist. No technology exists as an object per se, aside from its practical and situated use, and artificial intelligence is no exception. The case of misuse mentioned by Lasko and colleagues regarding mortality risk prediction in patients with pneumonia is common to predictive models, which are always biased to some extent, and possibly affected by confounding variables. However, in the case of ML-DSS, this risk is exacerbated by the capability to mine huge amounts of data and ground recommendations based on spurious associations.³

In regard to electrocardiogram reading, Mr Fogel and Dr Kvedar note that “computers may soon outperform and replace humans at this task.” Deskilling itself is rarely a problem; however, it becomes one if the technology fails or breaks down and if the sociotechnical system relies on the failed technology to complete the work. Similar concerns have been raised in regard to some accidents in aviation⁴ and also in medicine for electrocardiogram readings.⁵

Regarding the conflict between accuracy and interpretability of machine learning algorithms, we do not agree with Lasko and colleagues that “a model’s interpretability is only truly an issue when its accuracy is questionable.” Accuracy measures are useful variables, but remain surrogate end points in the evaluation of these systems. This is why it is necessary to raise the bar for the approval of ML-DSS. Because we recognize the potential of ML-DSS to improve medical practice, these systems must be considered on par with any other medical intervention, and their benefits and risks must be appraised in high-quality trials: “Innovation without sufficient evidence is a disservice to all.”⁶

Our Viewpoint advocated a due appraisal of ML-DSS effectiveness but should not be considered disapproval. We refused an objectivistic view of technology and advocated for a consequentialist approach for its adoption, calling for more

phase 3 studies to identify best practices to embed artificial intelligence into real-world medical settings and demonstrate advantages that outweigh both costs and the unavoidable drawbacks. We hope that a more cognizant use of these powerful predictive tools “will improve diagnostic models and contribute to a deeper understanding of the role of uncertainty in medicine,” as Licitra and colleagues state. That said, physicians should not be unprepared for the opposite.

Federico Cabitza, PhD
Raffaele Rasoini, MD
Gian Franco Gensini, MD

Author Affiliations: Department of Informatics, University of Milano-Bicocca, Milan, Italy (Cabitza); Centro Studi Medicina Avanzata, Florence, Italy (Rasoini, Gensini).

Corresponding Author: Federico Cabitza, PhD, Università degli Studi di Milano-Bicocca, Viale Sarca 336, Milan, Italy 20126 (cabitza@disco.unimib.it).

Conflict of Interest Disclosures: The authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest and none were reported.

1. Ribeiro MT, Singh S, Guestrin C. Why should I trust you? explaining the predictions of any classifier. <http://www.kdd.org/kdd2016/papers/files/rfp0573-ribeiroA.pdf>. Accessed November 1, 2017.
2. Prasad V, Vandross A, Toomey C, et al. A decade of reversal: an analysis of 146 contradicted medical practices. *Mayo Clin Proc*. 2013;88(8):790-798.
3. Calude CS, Longo G. The deluge of spurious correlations in big data. *Found Sci*. 2016;(Mar):1-8.
4. Carr N. *The Glass Cage: Where Automation Is Taking Us*. New York, NY: Random House; 2015.
5. Bogun F, Anh D, Kalahasty G, et al. Misdiagnosis of atrial fibrillation and its clinical consequences. *Am J Med*. 2004;117(9):636-642.
6. McCartney M. Margaret McCartney: innovation without sufficient evidence is a disservice to all. *BMJ*. 2017;358:j3980.

Guidelines for Letters

Letters discussing a recent *JAMA* article should be submitted within 4 weeks of the article's publication in print. Letters received after 4 weeks will rarely be considered. Letters should not exceed 400 words of text and 5 references and may have no more than 3 authors. Letters reporting original research should not exceed 600 words of text and 6 references and may have no more than 7 authors. They may include up to 2 tables or figures but online supplementary material is not allowed. All letters should include a word count. Letters must not duplicate other material published or submitted for publication. Letters not meeting these specifications are generally not considered. Letters being considered for publication ordinarily will be sent to the authors of the *JAMA* article, who will be given the opportunity to reply. Letters will be published at the discretion of the editors and are subject to abridgement and editing. Further instructions can be found at <http://jamanetwork.com/journals/jama/pages/instructions-for-authors>. A signed statement for authorship criteria and responsibility, financial disclosure, copyright transfer, and acknowledgment and the ICMJE Form for Disclosure of Potential Conflicts of Interest are required before publication. Letters should be submitted via the *JAMA* online submission and review system at <https://manuscripts.jama.com>. For technical assistance, please contact jama-letters@jamanetwork.org.

Section Editor: Jody W. Zylke, MD, Deputy Editor.