# An awakening in medicine: the partnership of humanity and intelligent machines

In concurrence with the introduction of the internet, widely networked computers, and the collection of large amounts of digital data, the medical profession as a whole has become more self-aware and self-critical. It is increasingly apparent that suboptimal decisions are made at times and, on other occasions, are fatally flawed. Most clinical decisions rest largely on what is referred to as the art of medicine: that is, decision-making that is based on inconsistent and incomplete provider knowledge; variable skills, training, and experience; and last but not the least, an array of biases. Unsurprisingly, the result is an unacceptable degree of care variation that is not explained by patient factors or the clinical context.[1] Every minute, a medical decision is being made somewhere that could be more informed, more objective, more precise, and more safe. How does medicine move on to adapt to an era of big data and a need to make consistent, data driven, evidence and value-based clinical decisions?

Artificial intelligence (AI) refers to the ability of computers to learn the associations within troves of data to assist with classification (eg, diagnosis), prediction (eg, triaging and prognostication), and optimisation (eg, precision treatment). Studies have reviewed current applications of AI, as well as the opportunities and challenges it poses in the field of health care.[2] However, it is important to reiterate the vast potential for AI beyond the use in medical imaging, in a wide array of disparate clinical situations. For example, AI has long been used in the development of severity scoring systems, but could AI assist the psychiatrist in assessing suicide risk[3] or assist the pediatrician in the recognition of rare genetic syndromes?[4] AI is already being built into the development of physiological monitors and has been proposed as an aid to decisions such as the treatment of sepsis, assessment of readmission risk, and recognition of consciousness in unresponsive patients.[5–8] One can envision the possibilities of AI guidance and support in the care of patients with complex conditions, such as those with multiple, chronic medical problems, or the decision to proceed to major surgery in fragile, complicated patients. The question is not whether computers can outperform humans in specific tasks, but how humanity will embrace and adopt these capabilities into the practice of medicine.

For AI to achieve adoption in medicine, hospitals and health-care systems must be willing to buy it, and patients and providers must accept it. The AI task must be deemed useful to patients, providers, and payors. If the gains are trivial, the unsystematic and uncontrolled mushrooming of minimal value-added medical AI algorithms will only add to the over-referral, over-diagnosis, over-treatment and, ultimately, to overall health-care costs, as well as a distrust of the technology.

With compelling use cases defined, one can get to the challenge of building trusted models. The development of AI algorithms requires data extraction and integration that is engineering intensive, as well as data standardisation and curation that is clinical domain expertise intensive. Like the scientist to clinician partnership in translational medicine, a new partnership of data scientist, engineers, and clinicians needs to be developed. However, this process is wrought with challenges, both technical and non-technical. The most fundamental challenge to the development and implementation of machine learning in health-care is access to reliable, well curated data. Clinical data usually resides across different systems, often locked in a proprietary format requiring additional costly software for extraction. And once data has been freed from proprietary servers, health data standards and common data models for research purposes are far from ideal, creating unnecessary work to unlock the value in the data. The task of establishing these standards requires not only expertise in medical informatics but time-consuming input from domain experts and researchers who might have little interest in providing this important housekeeping task. It would be worthwhile to learn from the experience of the Laboratory for Computational Physiology at the Massachusetts Institute of Technology who have created the publicly available and greatly used Medical Information Mart for Intensive Care (MIMIC) database and the eICU Collaborative Research Database, and have crowd-sourced the data curation across MIMIC's 12,000 users.[9]

With trusted models established, AI adoption will then require addressing the practical, but too often overlooked, matter of designing AI applications that are intelligently integrated with clinical workflows. Following the example of the current generation of electronic medical records, burnout has often followed as providers struggled with usability. Poorly designed AI could analogously worsen information overload and cognitive fatigue. The controversial IBM Watson-MD Anderson partnership is an example of the difficulties encountered in applying AI to complex clinical issues. Will it require decades of trial and (unfortunately, sometime disastrous) errors in human-machine interfaces? To achieve adoption, AI will need to be an invisible, seamless, and unbiased aid, helping patients and physicians make better decisions in an efficient, effective, and acceptable manner.[9]

It is clear that humans can trust machines for particular decisions, such as in airline travel today. But software has wrestled with pilots and won – with catastrophic results.[10] The Boeing 737 control issues showed that any changes made to a functional system must be well communicated, completely tested, and accompanied by a thorough education of all involved in the use of these systems. Clinical practice should evolve as a hybrid enterprise with clinicians who know what to expect from, and how to work with, what is fundamentally a very sophisticated clinical support tool.

Working together, humans and machines can address many of the decisional fragilities intrinsic to current practice. The human-driven scientific method can be powerfully augmented by computational methods sifting through the necessarily large amounts of longitudinal patient- and provider-generated data. But the guidance of these methods also requires more precise definitions of some of the foundational principles in medicine: eg, what is normal, what abnormalities require clinical intervention, what outcomes are we trying to achieve, and what costs are acceptable to do so? For example, defining normal is fundamentally important when considering the use of AI for evaluating chest x-rays. Today, we use a radiologist's opinion as the gold standard. But it is also crucial to know which radiological abnormalities are significant and require intervention versus those which represent clear-cut overdiagnoses with the potential for overtreatment generating unnecessary costs, complications, and suffering. Similarly, for diagnosis of diabetic retinopathy from fundus photographs, the use of a consensus agreement standard with ophthalmologists might not be the gold standard. One opportunity is to mine huge, longitudinal, linked population datasets (including imaging results, treatments, and clinical and patient related outcomes) to fully define what constitutes normal in various contexts. A single conclusion might not apply in all circumstances. In the intensive care unit, for example, in each individual case we need to establish the value that the patient and family place on simply extending survival versus attempting to provide quality of remaining life. In obstetrics, what is the optimal outcome we are trying to achieve in terms of the decision to employ a caesarian section or not? Is it a minimal rate of surgery, a minimal rate of fetal mortality, or a minimal rate of maternal mortality? These are complex, non-trivial issues in terms of providing such decision support to the obstetrician.

One of the indirect benefits of AI might be in forcing us to clearly define the major challenges in health care in a way that we have not been forced to do before. Anyone who has attempted to develop clinical software understands that coding an application requires so-called black and white solutions (eg, symptoms yes or no, disease present or absent) that are often difficult to obtain in medicine. We can only design AI for those issues in which we possess a precise, contextual, and optimally complete level of understanding of health and disease. And after we do develop AI solutions, we will still need to continue to provide supervision using human intelligence with all its quirks, inconsistencies, and potential for deficits in situational awareness. AI is not going to produce a perfect medical system, but if thoughtfully designed and implemented, it has the potential to produce a better one.

*Leo Anthony Celi, Benjamin Fine, David J Stone*
Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA, USA (LAC); Division of Pulmonary, Critical Care and Sleep Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA (LAC); Department of Diagnostic Imaging and Operational Analytics Lab, Trillium Health Partners, Mississauga, ON, Canada (BF); Department of Medical Imaging, University of Toronto, ON, Canada (BF); Departments of Anesthesiology and Neurosurgery and the Center for Advanced Data Analytics, University of Virginia, Charlottesville, USA (DJS)

We declare no competing interests. We thank Dr Mazyeh Ghassemi for inspiring this work during a debate at the Machine Learning for Health Unconference in Toronto in May 28, 2019.

1 Wennberg J. Unwarranted variation in healthcare delivery: implications for academic medical centres. *BMJ* 2001; **325:** 961–64.

2 The Lancet Digital Health. Walking the tightrope of artificial intelligence guidelines in clinical practice. *Lancet Digit Health* 2019; **1:** e100.

3 Simon GE, Johnson E, Lawrence JM, et al. Predicting suicide attempts and suicide deaths following outpatient visits using electronic health records. *Am J Psychiatry* 2018; **175:** 951–60.

4 Hsieh T-C, Mensah MA, Pantel JT, et al. PEDIA: prioritization of exome data by image analysis. *Genet Med* 2019; published online June 5. DOI:10.1038/s41436-019-0566-2.

5 Rush B, Celi LA, Stone DJ. Applying machine learning to continuously monitored physiological data. *J Clin Monit Comput* 2018; published online Nov 11. DOI:10.1007/s10877-018-0219-z.

6 Komorowski M, Celi LA, Badawi O, Gordon AC, Faisal AA. The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. *Nat Med* 2018; **24:** 1716.

7 Badawi O, Breslow MJ. Readmissions and death after ICU discharge: development and validation of two predictive models. *PloS one* 2012; **7:** e48758.

8 Claassen J, Doyle K, Matory A, et al. Detection of brain activation in unresponsive patients with acute brain injury. *New Engl J Med* 2019; **380:** 2497–505.

9 Johnson AEW, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. *Sci data* 2016; **3:** 160035.

10 Hatton L, Rutkowski A. "Lessons must be learned"—but are they? IEEE software. 2019; **36:** 91–95.