The AI Future of Emergency Medicine

Robert J. Petrella, MD*

*Corresponding Author. E-mail: robertjpetrella@yahoo.com.

In the coming years, artificial intelligence (AI) and machine learning will likely give rise to profound changes in the field of emergency medicine, and medicine more broadly. This article discusses these anticipated changes in terms of 3 overlapping yet distinct stages of AI development. It reviews some fundamental concepts in AI and explores their relation to clinical practice, with a focus on emergency medicine. In addition, it describes some of the applications of AI in disease diagnosis, prognosis, and treatment, as well as some of the practical issues that they raise, the barriers to their implementation, and some of the legal and regulatory challenges they create. [Ann Emerg Med. 2024;84:139-153.]

A **podcast** for this article is available at www.annemergmed.com.

0196-0644/$-see front matter

https://doi.org/10.1016/j.annemergmed.2024.01.031

SEE EDITORIALS, P. 154, P. 157.

INTRODUCTION

The terms artificial intelligence (AI) and machine learning are often used interchangeably, but machine learning is just one type of AI (Figure 1). AI pertains to computers performing intelligent, human-like tasks,[1] whereas machine learning, more specifically, has to do with "programming computers to learn from experience."[2] From its inception, computing has generally applied preformulated rules to inputs or data, but machine learning models make predictions by tuning their internal parameters to the data, in a sense creating their own rules. Deep learning models are, in turn, a subtype of machine learning models that are structured in multilayered networks of parameters (see also the Glossary [Appendix E1, available at http://www.annemergmed.com]), whereas large language models[3-8] such as Generative Pretrained Transformer (GPT),[9] Google Bard,[10] and Bidirectional Encoder Representations from Transformers (BERT)[4] are a subtype of deep learning models that have a very large number of parameters and generate responses to verbal prompts by processing them as a whole.[11-13]

Because of the rapid progress in the area of machine learning, the result of AI on our working lives in the emergency department (ED) is likely to increase in the near future, and the technology promises to transform both emergency medicine and medicine more broadly. The changes will likely occur in 3 stages: Map, Measure, and Manage (Figure 2). This article discusses the development of AI in emergency medicine in terms of these overlapping yet differing stages. At each stage, it describes important, related ideas, such as key properties of the models, and recognizes barriers to their development and implementation.

MAPPING STAGE

The first stage involves identifying pertinent clinical problems and exploring AI methods for solving them. The medical field, as a whole, has clearly progressed at least this far. Computer programs, often rule-based, have served as "clinical decision support systems" in the ED for years, providing preliminary ECG readings, voice recognition for dictation, and checks for drug interactions and allergies. In the near term, machine learning-based AI algorithms will surely contribute more to this kind of decision support by, for example, helping emergency medicine clinicians interpret radiology studies when the radiologists are not available.[14-18] In certain reading tasks, the accuracy of AI systems is already close to that of radiologists.[15,19-24] At some institutions, AI algorithms are currently being used for prioritizing abnormal studies.[25,26] For example, the ICH tool (Aidoc) helps rapidly detect intracranial hemorrhage.[27]

Diagnosis and Management Using AI Models

AI systems will no doubt help clinicians with many kinds of diagnoses, as well as the prediction of clinical outcomes. Multiple AI products have already been approved for assisting with the diagnosis and management of acute ischemic stroke. They appear to have fair sensitivity and good specificity for large vessel occlusion[28] and reportedly help reduce the time to mechanical

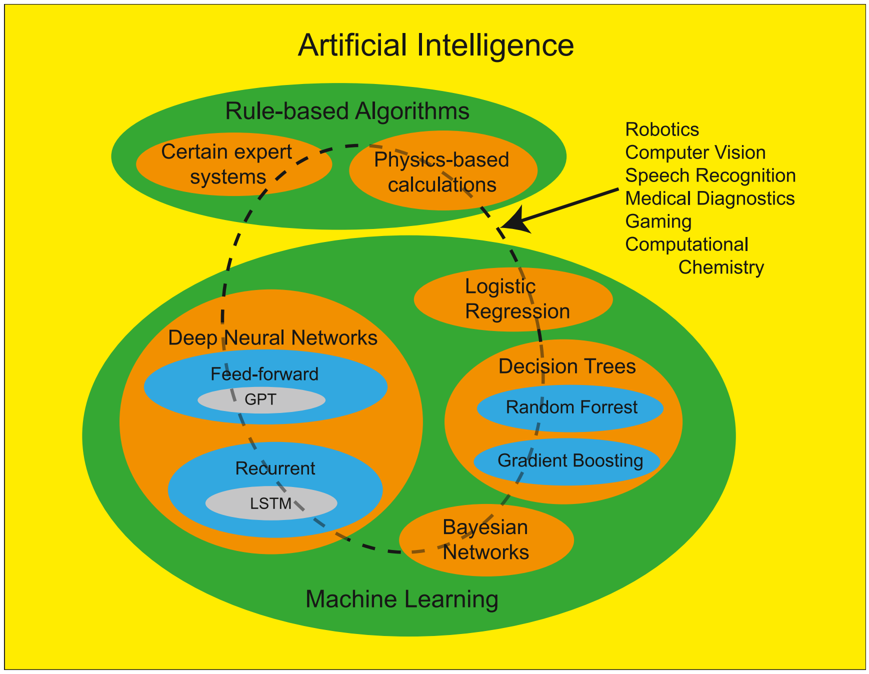

Figure 1. A general schema for artificial intelligence programs. Programs can roughly be divided into rule-based algorithms, which calculate or predict based on hard-coded rules, and machine learning, which are "trained" based on data sets. Rule-based programs comprised most of computing until fairly recently. Many fields, including those that traditionally used "hard-coded" programs such as robotics, computational chemistry, and mathematics, now incorporate elements of both rule-based computation and machine learning. Neural networks, of which there are several subtypes, are themselves a type of machine learning, of which there are also many subtypes. Feed-forward neural networks process data unidirectionally from input to output, ie, without any backward flow or loops, whereas recurrent neural networks include such loops. The lists of categories and subtypes shown here are representative, not exhaustive. *GPT*, Generative Pretrained Transformer, *LSTM*, long short-term memory (model).

thrombectomy.[29] In addition, EDs have started incorporating more radiology-related AI technology into their practices. For example, Critical Care Suite (GE Healthcare) and HealthPNX (Nanox) look for pneumothorax on chest radiographs. Many other diagnostic or predictive models have also been developed, some of which are listed in the Table. For example, machine learning algorithms have shown promise in identifying which patients with febrile neutropenia will develop multidrug-resistant infections[30] or which patients are likely to develop sepsis.[31-33] A number of machine learning sepsis predictors have been implemented in US EDs, including Sepsis Watch[34] and the Epic Sepsis Prediction Model[35] (but see below). A good deal of machine learning research has been done on ED triage.[36-46] At least one triage algorithm is being used in EDs as of this writing,[47] although validation studies are pending.

In addition, AI models are being built to make case-specific treatment recommendations, eg, antibiotic suggestions based on the clinical scenario, patient-specific factors, and local antibiograms.[48-51]

Types of AI Assistance in Clinical Decisionmaking

The choice of model for solving a particular clinical problem is important. There are at least 2 fundamentally different ways that AI programs can help clinicians make decisions: (1) solve a clinical problem with a "black box" model[52]—ie, in which data pertaining to the clinical scenario are put in, and the recommended decision is generated, without any accompanying description of the underlying rationale; or (2) help clinicians formulate better clinical decision rules, prediction rules[53] or clinical practice guidelines.

Model interpretability. The latter is a decades-old idea[54-56] that has never come to fruition, but modern AI may help.[57] Clinical decision rules are based on clinical patterns, and although human beings and deep learning models are both good at pattern recognition, computers can be trained on enormous amounts of data that would not be feasible for humans to sift through. However, obtaining fundamental insights from machine learning programs can be difficult because it requires model interpretability.[58] This is the capacity of a model to reveal relationships between the features in clinical data—ie, the

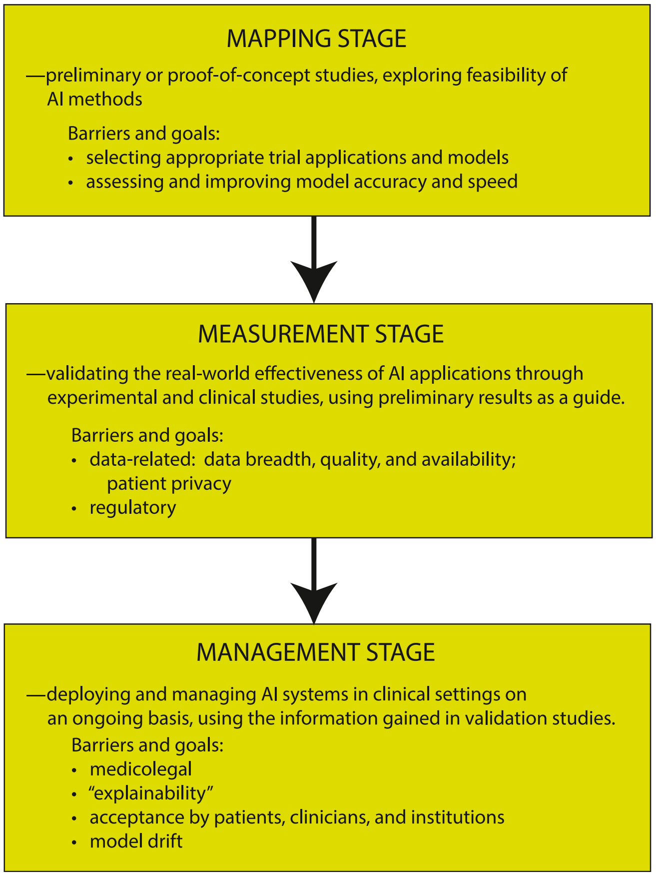

Figure 2. The 3 stages of medical AI development. This framework reflects that put forward in 2023 by The National Institute for Standards and Technology[160] for risk management in AI development. It is used here in a broader context. Each stage is associated with developmental goals and barriers. *AI*, artificial intelligence.

input variables—and the clinical decisions that produce the best results and to do it in a way that humans can understand.

Some machine learning models have *intrinsic explainability*, meaning explanatory metadata naturally falls out of them. Examples are logistic regression,[59] simple decision-tree algorithms,[60-62] and some Bayesian models.[63] So-called *explainable AI* has grown to be an active area of research over the last few years.[64-67] For harder problems, larger, more flexible models tend to be more accurate but more complex and "opaque" or less easily interpretable.[68] Deep learning models, such as convolutional neural networks[69] and large language models,[3-10] are very powerful,[70] but they tend to be opaque[71]—ie, they function as black boxes. In rough terms, deep learning models' predictions are akin to what has been called System 1 thinking in humans—quick, approximate, and unexplained.[72,73] Rationales have to be drawn from these models externally, or *post hoc*, after they are trained, normally with the help of some analysis tools (eg, see McKenzie et al,[74] 2022). Extracting a *buried clinical decision rule* from a deep learning model *post hoc* is, therefore, very difficult, and some data scientists believe it will become infeasible as the models grow larger and more complicated.[75]

Relation to background knowledge. Because they are reductions of complex systems, human-comprehensible clinical decision rules derived from machine learning models—like any clinical decision rule—are approximate and, in a sense, "contrived." They cannot possibly reflect all the relationships present in the model,[75] just as commonly used clinical decision rules like the History, Electrocardiogram, Age, Risk factors, and Troponin (HEART) score[76] do not capture all the complexity of clinical scenarios. Some argue that what matters most in machine learning is a correct prediction, not an approximate underlying rationale.[77,78]

However, medical providers and patients want to understand the thinking that underpins medical decisions,[79] especially for problems that have not been well studied or situations where critical decisions are at stake, like risk stratification in serious illness, treatments with significant side effects, or distribution of scarce resources.[80] For example, a higher age raises the HEART score, and that corresponds well to our knowledge of coronary disease, but if, instead, it lowered the score, we would want to understand why and might suspect the model is wrong. An AI model using its own internally generated version of the HEART score rule without an accompanying explanation would not allow us to check the rule against our background knowledge in that way. From this perspective, the capacity to be explained or interpreted will be essential to the usefulness of clinical decisionmaking models in emergency medicine.

Researchers in Seoul used opaque machine learning methods to predict one minute in advance whether a child might become hypoxic while under general anesthesia during a surgical procedure.[81] The inputs—ie, the features the models used—were a few demographic and physiologic parameters, such as age and end-tidal CO_2. The models made accurate predictions, but they did not identify any risk factors for hypoxia. That would have been useful but would have required more computation, eg, rerunning the model predictions many times, leaving out some of the features each time (eg, age or fraction of inspired oxygen), and checking how those changes affected the model's accuracy.[82,83]

In that study, the model used about 10 features, so the authors could conceivably have done that kind of thorough analysis. However, some deep learning models use hundreds

Table. Studies of some emergency department–related clinical artificial intelligence models.

Study Objective	Authors, References	Year	Type of Model	Validation
Diagnosis of pulmonary embolism by ECG	Valente Silva et al[212]	2023	Deep learning	Internal
Identifying appendicitis in undifferentiated abdominal pain	Su et al[213]	2022	Logistic regression, random forest	Internal
Risk stratification for				
TBI	Molaei et al[214]	2016	Random forest	Internal
Pediatric appendicitis	Reismann et al[215]	2019	Logistic regression	Internal
CHF	Sax et al[216]	2021	Various machine learning	Internal
Pulmonary embolism	Villacorta et al[217]	2021	Logistic regression	Internal
Chest pain	Liu et al[218]	2021	Various machine learning (dimensionality reduction)	Internal
Predicting				
Unscheduled 72-hour returns in patients with abdominal pain	Hsu et al[219]	2021	Various machine learning	Internal
Positive urine cultures	Taylor et al[220]	2018	Various machine learning	Internal
Need for admission in pediatric asthma	Patel et al[38]	2018	Various machine learning	Internal
multi-drug-resistant infections in febrile neutropenia	Garcia-Vidal et al[30]	2021	Various machine learning	Internal
Sepsis	Nemati et al[31]	2018	Artificial Intelligence Sepsis Expert*	Internal and external
	Delahanty et al[32]	2019	Gradient boosting	Internal
	Lin et al[22]	2021	Gradient boosting	Internal and external
	Zhang et al[33]	2021	LSTM (RNN)	Internal
Clinical deterioration	Choi et al[221]	2023	Gradient boosting	Internal and external
Adverse events in hyperglycemia	Hsu et al[222]	2023	Various machine learning	Internal
Frequent use of the ED	Chiu et al[223]	2023	Various machine learning	Internal
Risk of ED visits in older adults	Park et al[224]	2023	Various machine learning	Internal and external
In out-of-hospital patients				
Short-term mortality, need for ICU	Spangler et al[225]	2019	Gradient boosting	Internal
Cardiac arrest	Blomberg et al[226]	2019	Not specified	Internal
In patients with trauma				
Acute coagulopathy	Li et al[227]	2020	Random forest, logistic regression	Internal, prospective
Need for massive transfusion	Lammers et al[228]	2022	Various machine learning	Internal
ED volume	Jilani et al[229]	2019	FFNN, ARIMA, fuzzy time series	Internal
ED patient flow	Pak et al[230]	2021	Random forest and various linear regression models	Internal
ED triage				
Estimating ESI score in abdominal pain	Farahmand et al[36]	2017	Various machine learning	Internal, prospective
Predicting:				
Hospital admission	Hong et al[37]	2018	Logistic regression, deep learning, gradient boosting	Internal
Disposition	Raita et al[39]	2019	Various machine learning	Internal
	Chen et al[43]	2020	Deep learning model	Internal

Table. Continued.

Study Objective	Authors, References	Year	Type of Model	Validation
ESI levels	Ivanov et al[44]	2021	Gradient boosting (Not specified)	Internal
Hospitalization	Lee et al[45]	2021	Deep learning	Internal
Screening for septic shock	Kim et al[40]	2020	Various machine learning	Internal
Identifying critical illness	Joseph et al[42]	2020	Deep learning	Internal
Identifying patients with low-severity	Chang et al[46]	2022	Various machine learning	Internal and external

Study objective—the clinical goal of the model; Authors, References—authors of the study and study citation; Year—the year of study publication; Type of model—the type of AI model used in the study; Validation—how the model was validated. "Internal" means the data set collected was used for both training and validation, usually by partitioning it into separate sections. "External" means the model was also validated with other, unrelated data (from sources external to the training data institution). "Prospective" indicates testing was done on new data from the same institution.
ARIMA, autoregressive moving average integrated model; *CHF,* congestive heart failure; *ED,* emergency department; *ESI,* emergency severity index; *FFNN,* feed-forward neural network; *ICU,* intensive care unit; *LSTM,* long short-term memory (model); *RNN,* recurrent neural network; *TBI,* traumatic brain injury.
*Based on a Weibull-COX proportional hazard model.

or thousands of distinct features. Large language models, which have been used to try to predict the ED disposition based on triage notes[3] and to triage patients based on history of present illness alone,[84,85] rate words on several hundred[4,5] to thousands[86,87] of characteristics or dimensions, which is what they use as features of language. Boiling down a black box model with that many features to show which ones, and which interactions between them, are important in a given problem remains a daunting analytical challenge.

The emerging role of large language models. Because the more advanced large language models such as GPT-4 often articulate rationales along with their answers to queries, it is hoped that they may eventually circumvent the interpretability problem through self-explanation.[88,89] This is an area of active research,[67,90-92] but so far, the results have been mixed, as large language model-generated explanations are often implausible or inconsistent across related questions.[91,92]

It also points to a third way that AI could help clinicians make decisions: by summarizing and presenting the pertinent data that has already been established regarding a given clinical question,[93] possibly with an accompanying differential diagnosis[94,95] or list of treatment options.[96,97] This idea is far from being fully realized in publicly available large language models, however. A recent study by Berg et al[98] indicated that although Chat-GPT was fairly proficient at generating preliminary differential diagnoses in straightforward ED cases, it could be inconsistent in its answers. Moreover, although large language models show promise in their ability to answer medical questions,[99-102] they often "hallucinate" incorrect information[103-105] or otherwise fail to answer correctly[106-108] or give proper medical advice.[109] A 2023 Stanford study[110] found that although the responses provided by GPT-3.5 and -4 on clinical questions were, for the most part, unharmful, they correlated poorly with answers provided by an in-house informatics consultation service.[111] The development and use of large language models is currently in its infancy—as of May 2023, there were only 9 published articles relevant to emergency medicine[112]—but it will, no doubt, surge in the coming years.

MEASUREMENT STAGE

As described in the prior section, considerable research has been done in exploring and developing medical AI applications. However, these models need to be clinically validated to be of most use in real patient care settings. Many models perform better on internal tests—ie, on data related to their training data sets—than they do in real-world applications. An example is the underperformance in a 2021 study of the widely implemented Epic Sepsis Prediction Model.[113] As alluded to above, most clinical AI models have only been validated internally. In 2021, one review found more than 19,000 clinically related AI studies but only 41 randomized controlled trials,[114] and Yin et al[115] found only 51 studies in which AI made decisions in real-life clinical settings. This so-called "AI chasm"[116,117] argues that the medical field, on average, currently stands near the beginning of the Measurement stage.

During the coronavirus disease 2019 (COVID-19) pandemic, machine learning methods were used to rapidly identify the Janus kinase inhibitor baricitinib as a potential anti-SARS-CoV-2 drug,[118] and the drug's effectiveness was validated in subsequent experimental[119] and clinical[120-122] studies. To realize AI's potential, emergency medicine will need much more of that kind of validation of the basic research and clinical hypotheses generated by machine learning programs.

Data Set Quality

An array of barriers confronts the validation of clinical AI. A number of them involve patient data: (1) Data fragmentation. Many patients' medical information is scattered across a number of hospital systems. (2) Data locality. Algorithms trained on data from one hospital system or geographic area may not be applicable in another system or area.[123] (3) Data representation. Fewer data exist for certain populations, eg, pediatric patients.[124] (4) Data errors and ambiguities. For example, not all radiology readings (done by humans or otherwise) are correct,[125] and models trained on them inherit the errors.[126]

There are also technical issues with databases. For instance, clinical data exists in many different formats, and the reading/analysis software has to handle them all.[127,128] In addition, data sets are often biased. Current genetic databases, for example, contain data mainly from patients of European descent.[129] Machine learning algorithms are particularly vulnerable to misrepresentation bias because they will use whatever data features they are allowed to. For example, a COVID-19 prediction model based on radiographic studies failed because it was trained partly on ICU cases, so it learned that patients who were lying down were more acutely ill,[130] which is, of course, true but unhelpful.

Another difficulty lies in the nature of clinical diagnosis itself. Details like the general impression (eg, the patient "looks terrible") are often crucial to proper ED management, but medical providers often do not record them in the electronic medical record for capture in patient data sets. Conceivably, we could video record all patients at presentations to teach our computers to encode general impressions, but this, of course, raises patient privacy issues. Video recording or photography of ED patients, usually for educational or documentation purposes, is sometimes done,[131] as is video monitoring of inpatients,[132,133] but images and recordings are currently subject to Health Insurance Portability and Accountability Act (HIPAA) restrictions and generally require patient authorization for use in research.[134]

Privacy

More generally, whose data will be made accessible to researchers? If not everyone's, then should access be made voluntary? That may introduce self-selection biases. Moreover, offering citizens financial incentives for their data raises both ethical and political concerns.

Many institutions and hospital systems already have their own patient data warehouses.[135-137] Moreover, there are large patient data networks, like the Patient-Centered Outcomes Research Network (PCORnet)[138] and the Accrual to Clinical Trials (ACT) Network,[139] which is National Institutes of Health-funded, that contain data on tens of millions of Americans. These data systems make inroads toward solving the problems of locality because they cover large areas and fragmentation because they aggregate data, but they have varying levels of patient privacy[140] and security.

The United States has traditionally had a more *laissez-faire* approach to corporate use of personal data than Europe,[141,142] although large data security breaches and the rise of AI seem to be spurring changes. Data privacy in the United States is covered under HIPAA, the Federal Information Security Management Act, the Gramm-Leach-Bliley Act, and several other laws. There are also various disconnected laws at the state level.[143] Four states (California, Connecticut, Colorado, and Virginia) did pass general data privacy laws in 2022-2023 that contain provisions governing automated decisionmaking,[144,145] a regulatory term that includes AI algorithms.

The great research benefits of feeding large swaths of clinical data into machine learning models may ultimately overwhelm people's privacy concerns and motivate solutions to other problems. Organizations have already developed methods for "anonymizing" patients' personal data,[146] and the reading abilities of large language models may circumvent the problems posed by differing patient data formats.[147]

Regulatory

Despite the United States Food and Drug Administration's (FDA's) long-time involvement with computer-related medical devices,[148-150] the overall regulatory regime for AI products is still in its formative stages, with regulators scrambling to keep pace with rapid changes in the field. In 2017, the FDA launched a precertification program meant to streamline the approval of digital health applications (called Software As a Medical Device), such as mobile phone-based symptom checkers, but the program failed because of the speed of innovation in this area.[151] Smartphone apps are becoming increasingly popular with consumers and prescribed by clinicians. They obviously pose risks if they give faulty results or advice, which is why early AI-related symptom checkers tended to be risk-averse.[61] A

2022 study suggested that the apps have become more specific but less sensitive, and their overall accuracy remains similar to that of laypersons.[152] Some studies have suggested improved outcomes with these apps,[153-156] and although they are not yet commonly prescribed from the ED, that is likely to change.

The Federal Trade Commission and The National Institute for Standards and Technology (NIST) have laid down some ground rules for commercial AI products, including ones ensuring that the results of AI algorithms are explainable to consumers,[157,158] but as of this writing, both bodies are still formulating their reliability, accuracy, and safety guidelines/regulations[159,160] and NIST's AI Risk Management Framework is nonbinding.[161,162]

MANAGEMENT STAGE

Once AI-related medical tools have been developed and validated, they need to be implemented or deployed, ie, integrated into daily clinical workflows and managed. If and when this is achieved for enough applications, it is conceivable that AI will drive, or at least guide, the entire clinical process.[163] Clinicians may eventually come to rely on AI algorithms,[164] even if they have become so internally complex as to make the rationales for their recommendations practically impossible to decipher.

No one really knows what medical diagnosis and treatment will look like in this management stage of medical AI development, but medicine in general will likely be more preventative, detecting more conditions before they are clinically apparent.[165,166] Still, there will no doubt be accidents and injuries, as well as unexpected infections, thromboembolic events, and inflammatory conditions, so emergency medicine services will likely continue to fill a need. One can reasonably surmise emergency medicine will involve more *embodied AI*, including robotics,[167] computer vision, natural language processing, and speech recognition—ie, computers interacting with their environment, as in embodied conversational agents.[168]

However, as described below, there are many obstacles to the successful deployment of AI systems in medical settings, including medicolegal, regulatory, technical, and social/philosophical ones.

Medicolegal

If an AI's reading of a radiograph is in error and causes a bad clinical outcome, is the software developer liable? The treating clinician? The hospital that purchased the software? If not, then who is? Incorrect suggestions by AI systems can degrade the quality of clinical decisions.[169] Although the FDA issued its first approval of a software-enabled medical device in 1995,[148] the case law in this area has not yet been well developed.[170] In related areas, the "learned intermediary" doctrine[171] has generally held that manufacturers and pharmaceutical companies relieve themselves of liability by disclosing to the physician the risks of the device or drug. By this principle, software developers would not be liable for incorrect diagnoses or recommendations made by an AI that led to poor outcomes, provided the risks were disclosed beforehand. Courts have been reluctant to subject software developers to product liability law, especially in health care.[170,172] As pointed out by Price et al,[173] because current tort law rewards following the standard of care, liability concerns encourage physicians to avoid using AI to make patient-specific decisions that might deviate from a broader standard, thereby negating the potential advantages of AI-enabled personalization of care. Some advocate for more AI developer liability, arguing that sophisticated AI systems possessing some level of autonomous thinking and giving medical advice should be held to a higher legal standard than, say, a heart valve.[174,175]

Technical/Logistical

The compressed time frames in emergency medicine place special demands on AI diagnostics and treatment recommendations. For example, a tension pneumothorax is a clinical diagnosis made rapidly by history and physical examination because there's often no time for anything else. Timely AI diagnosis and advice in cases like this would require sensors (eg, biomechanical,[176] electrochemical,[177] and ultrasonographic[178]) to gather and process the initial data within the first minute or two. Other examples of situations requiring AI timeliness in the ED would include deciding whether or not to treat with thrombolytics, intubate a patient, perform an emergency cricothyrotomy, administer uncross-matched blood, or transfer a hypotensive patient to the radiology suite. Further, ED time constraints make accuracy a necessity in AI recommendations, as emergency medicine clinicians would need to trust them on the spot. This distinguishes the implementation of AI-based clinical decision support systems in the ED from that in, say, oncology, where virtual tumor boards have the time to meet and discuss the validity of AI recommendations for individual patients.[97]

Model Drift

Even if a machine learning model is very accurate when it is first deployed, its performance will often deteriorate over time because of changes in real-world data—eg, patient demographics or disease pathology—relative to its training data[179] or changes in the data's context of use.[180] In an example of this "AI aging"[181] phenomenon, machine learning models developed before the COVID-19 pandemic to predict admissions from the ED[182] or trigger sepsis alerts[183] saw a large increase in false positives during the pandemic. Thus, many models need periodic updating or validation. However, because they are usually proprietary commercial products (refer to FDA list[150] and Alaskar et al[184]), EDs and hospitals have limited control over them. Open-source development[117,185] might allow institutions to update models more frequently as well as train them on local data. In addition, models that are more adaptable to, or tolerant of, data set shifts are in development.[180,186-188]

Reasoning Ability

A key determinant of how far AI can advance as a medical research and clinical decision-making tool will be how well it can reason. Much of medical decision making involves common sense, a form of reasoning. A patient with bilateral upper extremity injuries will likely be incapable of using crutches (eg, for his/her sprained ankle), and an AI engine should be able to recognize that, despite the paucity of published clinical trials on the question. Formulating sensible mechanisms of disease involves multiple forms of reasoning, as does interpreting opaque machine learning models.

Although deep learning models are based on something like System 1 abilities—a knack for recognizing patterns that is critical for hypothesis generation[189]—reasoning in AI algorithms would be more akin to System 2 thinking in humans, which is analytic. Scientific inquiry,[190,191] diagnosis,[192] and theory building[193,194] require both intuition and analysis acting in concert.

As of this writing, publicly available deep learning AI models can do some reasoning—GPT-4 apparently scored in the 88th percentile on the Law School Admission Test[195]—but it is currently limited and error-prone,[196,197] especially in areas like common sense.[198] These models currently mimic the reasoning process by recognizing the statistical features of a particular problem and constructing answers from that rather than using transferable, explicit rules of logic.[199] Because the construction of accurate, cogent, and human-comprehensible explanations of a model's results involves reasoning, advances in this area will likely have a direct bearing on the ability of large language models to self-explain, which in turn could increase their utility in the ED and other areas of medicine.

Whether "Black Box" Will Ultimately Be Enough

Some believe that the current "intuitive" capabilities of AI, if taken to the extreme, will eliminate the need to design System 2-like analytical abilities into the models. They point to the Universal Approximation Theorem,[200,201] which says that a neural network if it is large enough, can approximate any function to any arbitrary degree of accuracy. In principle, this means that if one had enough relevant data, a large enough model, and enough time and computational resources, one could arrive at a sufficient answer to just about any clinical question. So, if deep learning algorithms could predict the optimal course of action in every possible clinical situation, would it matter that the programs didn't really "understand"—in the sense of being able to articulate reliable rationales—how they arrived at their decisions?

As of now, patients and clinicians would clearly reject the notion of machines making all the life-and-death treatment decisions without any human involvement. The more critical the decisions and the more serious the conditions for which we are using AI, the more important interpretability would seem to be. However, one supposes that with enough time, familiarity, and technical advancement, humans would someday learn to trust AI recommendations[164] even if we did not understand them or could not verify accompanying explanations, as odd as that may now seem.

In that event, the opacity of the models would make it more challenging for clinicians to stay in the decisionmaking loop. An AI's estimate of patients with trauma chances of survival as 78% with a chest tube and 33% without it would likely be very accurate because it would be tailored to that specific clinical scenario and patient. However, if the clinician expected the opposite advice, the absence of a verified explanation would likely prove confusing.

Who and What Will Drive

Presumably, in such a future world, patients would still make the decisions, supported by professional advice and discussion, but would be very heavily influenced by AI recommendations. The American Medical Association prefers the term *augmented intelligence* over artificial intelligence to stress that computers are assisting human beings rather than replacing us.[202] But if AI engines are better at a certain task than clinicians are, there will probably be a tradeoff between our level of involvement and the medical system's overall performance on that task. This is a general AI problem. If self-driving automobile technology advances to the point where an integrated network of self-driving cars dramatically reduces the number of car accidents,[203,204] human driving may be restricted[205] (see Summary Table).

Summary Table. A few take-home points from this discussion.

Key Ideas

- Artificial intelligence (AI) has the potential to transform many aspects of emergency medicine, from diagnosis and treatment of conditions to management of patient flow.
- Although many AI algorithms have been developed and internally tested, there has been a paucity of clinical validation to date.
- The most powerful AI methods, such as deep learning models, tend to be "opaque," so the best answers AI can give us may not be accompanied by verified explanations.
- Large language models, such as the GPT series, promise to be highly useful in emergency medicine, but their development is still in its early stages, and they are currently error-prone.
- Many important questions surrounding medical AI have not yet been answered, such as who is liable for errors caused by AI systems and how the technology will be regulated.
- When medical AI applications have become sufficiently advanced, there will likely be a tradeoff between the level of clinician involvement and the efficiency of the medical system, at least with regard to certain tasks.

GPT, Generative Pretrained Transformer.

Efficiencies created by AI systems could have a material result on ED and other hospital staffing needs within the next decade or two. A Goldman Sachs analysis estimated that 28% of tasks in health care are exposed to AI automation.[206] McKinsey & Co estimate that about a third of medical provider tasks are automatable[207] and that, by 2030, AI systems could free up about 12% of physicians' and surgeons' hours and 8% of nurse practitioners' hours.[208]

Still, it is likely that clinicians will retain a role in decisionmaking long after that. The application of clinical evidence to a particular case usually involves some degree of expert opinion to handle factors that lie outside of the prescribed guidelines. Even with advanced AI generating the recommendations, that will likely remain the case. Some variables probably will not be reflected in training databases, such as an individual patient's feelings and preferences, as well as practical, idiosyncratic, or situational factors. For instance, the patient may be on a novel or experimental medication that makes the recommended treatment unnecessary or contraindicated. Sometimes, patients refuse chest tube placements, even if they may prove crucial to their care and outcome, because they involve an invasive procedure that can be painful.

As to scientific advancement, if more powerful machine learning methods make the prediction problem easier to solve than the insight problem, in the future, basic medical understanding may lag clinical decisionmaking. That is not to say that we now thoroughly understand everything we do in clinical medicine—we do not. For example, we commonly use many drugs, including lithium,[209] acetaminophen,[210] and levetiracetam,[211] for which the mechanism of action is poorly understood. However, for the most part, theory and concepts drive science. Our findings in the clinic and the laboratory either validate or disprove not only the study hypothesis but also the underlying theory. In the future of machine learning, one might anticipate there will be a paradigm shift, in which the latter part of that will be missing. Scientific knowledge will no doubt run much deeper than it does today, and the new, effective treatments produced for patients will still be good for them, but those advances may be sparked by pattern matching that is not initially understood.

In conclusion, AI systems and machine learning models, in particular, have made great strides in their predictive abilities and demonstrated enormous potential as research and clinical tools. AI algorithms being developed to make diagnostic, prognostic, or therapeutic predictions in the ED or other clinical settings need to be validated in those settings before they can be helpfully integrated into daily workflows. This process, currently in its early stages, faces many technical, legal, logistical, social, and regulatory hurdles but promises to be transformative.

The author thanks Victor Ovchinnikov, PhD, Marc Hoffman, MD, Dave King, Paul Hershenson, and several Annals of Emergency Medicine *reviewers for their helpful comments on the manuscript. The author also thanks Martin Karplus, PhD and the Department of Chemistry and Chemical Biology at Harvard University, as well as Harvard Medical School and the Boston VA Medical Center, for support of this work.*

Supervising editor: Stephen Schenkel, MD, MPP. Specific detailed information about possible conflict of interest for individual editors is available at https://www.annemergmed.com/editors.

Author affiliations: From the Emergency Departments, CharterCARE Health Partners, Providence and North Providence, RI; Emergency Department, Boston VA Medical Center, Boston, MA; Emergency Departments, Steward Health Care System, Boston and Methuen, MA; Harvard Medical School, Boston, MA; the Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA; and the Department of Medicine, Brigham and Women's Hospital, Boston, MA.

Authorship: All authors attest to meeting the four ICMJE.org authorship criteria: (1) Substantial contributions to the conception

or design of the work; or the acquisition, analysis, or interpretation of data for the work; AND (2) Drafting the work or revising it critically for important intellectual content; AND (3) Final approval of the version to be published; AND (4) Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding and support: By *Annals'* policy, all authors are required to disclose any and all commercial, financial, and other relationships in any way related to the subject of this article as per ICMJE conflict of interest guidelines (see www.icmje.org). The authors have stated that no such relationships exist. The authors report this article did not receive any outside funding or support.

Publication dates: Received for publication November 3, 2023. Revision received January 23, 2024. Accepted for publication January 24, 2024.

REFERENCES

1. Andresen SL. John McCarthy: Father of AI. *IEEE Intelligent Systems*. 2002;17:84-85.
2. Samuel AL. Some studies in machine learning using the game of checkers. *IBM J Res Dev*. 1959;3:210-229.
3. Tahayori B, Chini-Foroush N, Akhlaghi H. Advanced natural language processing technique to predict patient disposition based on emergency triage notes. *Emerg Med Australas*. 2021;33:480-484.
4. Kravets A. A Deep Dive into the Code of the BERT Model. *Towards Data Science*. 2021. Accessed April 30, 2024. https://towardsdatascience.com/deep-dive-into-the-code-of-bert-model-9f618472353e
5. Zanella-Beguelin S, Tople S, Paverd A, et al. Grey-box Extraction of Natural Language Models. Proceedings of the 38th International Conference on Machine Learning. Proceedings of Machine Learning Research; 2021.
6. Montii R. Google Bard: Everything You Need To Know. *Search Engine Journal*. 2023. Accessed April 30, 2024. https://www.searchenginejournal.com/google-bard/482860/#close
7. Touvron H, Lavril T, Izacard G, et al. LLaMA: Open and Efficient Foundation Language Models. *arXiv* 2023; 2302.13971. Accessed April 30, 2024. https://arxiv.org/abs/2302.13971
8. Henshall W. What to Know About Claude 2, Anthropic's Rival to ChatGPT. *TIME |Tech| Artificial Intelligence* 2023. Accessed April 30, 2024. https://time.com/6295523/claude-2-anthropic-chatgpt/
9. OpenAI. GPT-4 Technical Report. *arXiv* 2023; 2303.08774. Accessed April 30, 2024. https://arxiv.org/abs/2303.08774
10. Manyika J, Hsiao S. An overview of Bard: an early experiment with generative AI. 2023. Accessed April 30, 2024. https://ai.google/static/documents/google-about-bard.pdf
11. Vaswani A, Shazeer N, Parmar N, et al. Attention Is All You Need. *arXiv* 2023; 1706.03762. Accessed April 30, 2024. https://arxiv.org/abs/1706.03762
12. Yenduri G, Ramalingam M, Chemmalar SG, et al. Generative Pre-trained Transformer: A Comprehensive Review on Enabling Technologies, Potential Applications, Emerging Challenges, and Future Directions. *arXiv* 2023; 2305.10435. Accessed January, 2024. https://arxiv.org/abs/2305.10435
13. Sengupta S. A Deep Dive into GPT's Transformer Architecture: Understanding Self-Attention Mechanisms. *GPTFrontier*. 2023. Accessed January, 2024. https://www.gptfrontier.com/a-deep-dive-into-gpts-transformer-architecture-understanding-self-attention-mechanisms/#:~:text=Instead%2C%20GPT%20leverages%20the%20decoder%27s,coherent%20and%20contextually%20relevant%20text
14. Seah JCY, Tang CHM, Buchlak QD, et al. Effect of a comprehensive deep-learning model on the accuracy of chest x-ray interpretation by radiologists: a retrospective, multireader multicase study. *Lancet Digit Health*. 2021;3:e496-e506.
15. Lee JH, Kim KH, Lee EH, et al. Improving the Performance of Radiologists Using Artificial Intelligence-Based Detection Support Software for Mammography: A Multi-Reader Study. *Korean J Radiol*. 2022;23:505-516.
16. Rudolph J, Huemmer C, Ghesu FC, et al. Artificial Intelligence in Chest Radiography Reporting Accuracy: Added Clinical Value in the Emergency Unit Setting Without 24/7 Radiology Coverage. *Invest Radiol*. 2022;57:90-98.
17. Abadia AF, Yacoub B, Stringer N, et al. Diagnostic Accuracy and Performance of Artificial Intelligence in Detecting Lung Nodules in Patients With Complex Lung Disease: A Noninferiority Study. *J Thorac Imaging*. 2022;37:154-161.
18. Lindsey R, Daluiski A, Chopra S, et al. Deep neural network improves fracture detection by clinicians. *Proc Natl Acad Sci U S A*. 2018;115:11591-11596.
19. Li D, Li S. An artificial intelligence deep learning platform achieves high diagnostic accuracy for Covid-19 pneumonia by reading chest X-ray images. *iScience*. 2022;25:104031.
20. Homayounieh F, Digumarthy S, Ebrahimian S, et al. An Artificial Intelligence-Based Chest X-ray Model on Human Nodule Detection Accuracy From a Multicenter Study. *JAMA Netw Open*. 2021;4:e2141096.
21. Sun J, Peng L, Li T, et al. Performance of a Chest Radiograph AI Diagnostic Tool for COVID-19: A Prospective Observational Study. *Radiol Artif Intell*. 2022;4:e210217.
22. Liu X, Faes L, Kale AU, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digit Health*. 2019;1:e271-e297.
23. Rajpurkar P, Irvin J, Ball RL, et al. Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. *PLoS Med*. 2018;15:e1002686.
24. Nam Y, Choi Y, Kang J, et al. Diagnosis of nasal bone fractures on plain radiographs via convolutional neural networks. *Sci Rep*. 2022;12:21510.
25. O'Neill TJ, Xi Y, Stehel E, et al. Active Reprioritization of the Reading Worklist Using Artificial Intelligence Has a Beneficial Effect on the Turnaround Time for Interpretation of Head CT with Intracranial Hemorrhage. *Radiol Artif Intell*. 2021;3:e200024.
26. Park A. How AI Is Changing Medical Imaging to Improve Patient Care. *Time |Health*. 2022. Accessed April, 2023. https://time.com/6227623/ai-medical-imaging-radiology/
27. Kundisch A, Hönning A, Mutze S, et al. Deep learning algorithm in detecting intracranial hemorrhages on emergency computed tomographies. *PLoS One*. 2021;16:e0260560.
28. Yahav-Dovrat A, Saban M, Merhav G, et al. Evaluation of Artificial Intelligence-Powered Identification of Large-Vessel Occlusions in a Comprehensive Stroke Center. *AJNR Am J Neuroradiol*. 2021;42:247-254.
29. Elijovich L, Dornbos D III, Nickele C, et al. Automated emergent large vessel occlusion detection by artificial intelligence improves stroke workflow in a hub and spoke stroke system of care. *J NeuroInterv Surg*. 2022;14:704-708.
30. Garcia-Vidal C, Puerta-Alcalde P, Cardozo C, et al. Machine Learning to Assess the Risk of Multidrug-Resistant Gram-Negative Bacilli Infections in Febrile Neutropenic Hematological Patients. *Infect Dis Ther*. 2021;10:971-983.
31. Nemati S, Holder A, Razmi F, Stanley MD, et al. An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU. *Crit Care Med*. 2018;46:547-553.
32. Delahanty RJ, Alvarez J, Flynn LM, et al. Development and Evaluation of a Machine Learning Model for the Early Identification of Patients at Risk for Sepsis. *Ann Emerg Med*. 2019;73:334-344.

33. Zhang D, Yin C, Hunold KM, et al. An interpretable deep-learning model for early prediction of sepsis in the emergency department. *Patterns (N Y)*. 2021;2:100196.
34. Sendak MP, Ratliff W, Sarro D, et al. Real-World Integration of a Sepsis Deep Learning Technology Into Routine Clinical Care: Implementation Study. *JMIR Med Inform*. 2020;8:e15182.
35. Bennett TD, Russell S, King J, et al. Accuracy of the Epic Sepsis Prediction Model in a Regional Health System. *arXiv* 2019; 1902.07276. https://arxiv.org/abs/1902.07276
36. Farahmand S, Shabestari O, Pakrah M, et al. Artificial Intelligence-Based Triage for Patients with Acute Abdominal Pain in Emergency Department; a Diagnostic Accuracy Study. *Adv J Emerg Med*. 2017;1:e5.
37. Hong WS, Haimovich AD, Taylor RA. Predicting hospital admission at emergency department triage using machine learning. *PLoS One*. 2018;13:e0201016.
38. Patel SJ, Chamberlain DB, Chamberlain JM. A Machine Learning Approach to Predicting Need for Hospitalization for Pediatric Asthma Exacerbation at the Time of Emergency Department Triage. *Acad Emerg Med*. 2018;25:1463-1470.
39. Raita Y, Goto T, Faridi MK, et al. Emergency department triage prediction of clinical outcomes using machine learning models. *Crit Care*. 2019;23:64.
40. Kim J, Chang H, Kim D, et al. Machine learning for prediction of septic shock at initial triage in emergency department. *J Crit Care*. 2020;55:163-170.
41. Miles J, Turner J, Jacques R, et al. Using machine-learning risk prediction models to triage the acuity of undifferentiated patients entering the emergency care system: a systematic review. *Diagn Progn Res*. 2020;4:16.
42. Joseph JW, Leventhal EL, Grossestreuer AV, et al. Deep-learning approaches to identify critically Ill patients at emergency department triage using limited information. *J Am Coll Emerg Physicians Open*. 2020;1:773-781.
43. Chen CH, Hsieh JG, Cheng SL, et al. Emergency department disposition prediction using a deep neural network with integrated clinical narratives and structured data. *Int J Med Inform*. 2020;139:104146.
44. Ivanov O, Wolf L, Brecher D, et al. Improving ED Emergency Severity Index Acuity Assignment Using Machine Learning and Clinical Natural Language Processing. *J Emerg Nurs*. 2021;47:265-278.e7.
45. Lee JT, Hsieh CC, Lin CH, et al. Prediction of hospitalization using artificial intelligence for urgent patients in the emergency department. *Sci Rep*. 2021;11:19472.
46. Chang Y-H, Shih H-M, Wu J-E, et al. Machine learning–based triage to identify low-severity patients with a short discharge length of stay in emergency department. *BMC Emerg Med*. 2022;22:88.
47. Johns Hopkins Technology Ventures. Digital Health Startup That Assists Emergency Department Decision Making Acquired. 2022. Accessed March, 2023. https://ventures.jhu.edu/news/stocastic-beckman-coulter-acquisition-digital-health/
48. Shen Y, Yuan K, Chen D, et al. An ontology-driven clinical decision support system (IDDAP) for infectious disease diagnosis and antibiotic prescription. *Artif Intell Med*. 2018;86:20-32.
49. Ben Souissi S, Abed M, El Hiki L, et al. PARS, a system combining semantic technologies with multiple criteria decision aiding for supporting antibiotic prescriptions. *J Biomed Inform*. 2019;99:103304.
50. Cai T, Anceschi U, Prata F, et al. Artificial Intelligence Can Guide Antibiotic Choice in Recurrent UTIs and Become an Important Aid to Improve Antimicrobial Stewardship. *Antibiotics (Basel)*. 2023;12:375.
51. Corbin CK, Sung L, Chattopadhyay A, et al. Personalized antibiograms for machine learning driven antibiotic selection. *Communications Medicine*. 2022;2:38.
52. Price WN. Big data and black-box medical algorithms. *Sci Transl Med*. 2018;10:eaao5333.
53. Eastwood KW, May R, Andreou P, et al. Needs and expectations for artificial intelligence in emergency medicine according to Canadian physicians. *BMC Health Serv Res*. 2023;23:798.
54. Musen MA, Tu SW, Das AK, et al. EON: a component-based approach to automation of protocol-directed therapy. *J Am Med Inform Assoc*. 1996;3:367-388.
55. Shahar Y, Miksch S, Johnson P. The Asgaard project: a task-specific framework for the application and critiquing of time-oriented clinical guidelines. *Artif Intell Med*. 1998;14:29-51.
56. Terenziani P, Molino G, Torchio M. A modular approach for representing and executing clinical guidelines. *Artif Intell Med*. 2001;23:249-276.
57. The Lancet Digital H. Walking the tightrope of artificial intelligence guidelines in clinical practice. *Lancet Digit Health*. 2019;1:e100.
58. Murdoch WJ, Singh C, Kumbier K, et al. Definitions, methods, and applications in interpretable machine learning. *Proc Natl Acad Sci U S A*. 2019;116:22071-22080.
59. Schober P, Vetter TR. Logistic Regression in Medical Research. *Anesth Analg*. 2021;132:365-366.
60. Ali MA, Hickman PJ, Clementson AT. The Application of Automatic Interaction Detection (AID) in Operational Research. *Operational Research Quarterly (1970-1977)*. Palgrave Macmillan Journals; 1975;26:243-252.
61. Semigran HL, Linder JA, Gidengil C, et al. Evaluation of symptom checkers for self diagnosis and triage: audit study. *BMJ*. 2015;351:h3480.
62. Jijo BT, Mohsin Abdulazeez A. Classification Based on Decision Tree Algorithm for Machine Learning. *JASTT*. 2021;2:20-28.
63. Madhukar NS, Khade PK, Huang L, et al. A Bayesian machine learning approach for drug target identification using diverse data types. *Nat Commun*. 2019;10:5221.
64. Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*. 2020;58:82-115.
65. Angelov PP, Soares EA, Jiang R, et al. Explainable artificial intelligence: an analytical review. *WIREs Data Mining and Knowl Discov*. 2021;11:e1424.
66. Grigoryan G, Collins AJ. Is Explainability Always Necessary? Discussion on Explainable AI. Modeling, Simulation and Visualization Student Capstone Conference. 2. DOI:10.25776/ 2ta8-8058. 2022. Accessed April 30, 2024. https://digitalcommons.odu.edu/msvcapstone/2022/scienceengineering/2
67. Jain S, Wiegreffe S, Pinter Y, et al. Learning to Faithfully Rationalize by Construction. *arXiv* 2020; 2005.00115. Accessed January, 2024. https://arxiv.org/abs/2005.00115
68. Myrianthous G. Understanding The Accuracy-Interpretability Trade-Off. Towards Data Science. 2021. Accessed April 30, 2024. https://towardsdatascience.com/accuracy-interpretability-trade-off-8d055ed2e445
69. Park JJ, Kim KA, Nam Y, et al. Convolutional-neural-network-based diagnosis of appendicitis via CT scans in patients with acute abdominal pain presenting in the emergency department. *Sci Rep*. 2020;10:9556.
70. Janiesch C, Zschech P, Heinrich K. Machine learning and deep learning. *Electronic Markets*. 2021;31:685-695.
71. Belle V, Papantonis I. Principles and Practice of Explainable Machine Learning. *Front Big Data*. 2021;4:688969.
72. Epstein S. Integration Of The Cognitive And The Psychodynamic Unconscious. *Am Psychol*. 1994;49:709-724.
73. Kahneman D. *Thinking Fast and Slow*. New York: Farrar, Straus and Giroux; 2011.
74. McKenzie AT, Marx GA, Koenigsberg D, et al. Interpretable deep learning of myelin histopathology in age-related cognitive impairment. *Acta Neuropathol Commun*. 2022;10:131.

75. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell*. 2019;1:206-215.
76. Six AJ, Backus BE, Kelder JC. Chest pain in the emergency room: value of the HEART score. *Neth Heart J*. 2008;16:191-196.
77. Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. *Lancet Digit Health*. 2021;3:e745-e750.
78. Castelvecchi D. Can we open the black box of AI? *Nature*. 2016;538:20-23.
79. Tonekaboni S, Joshi S, McCradden MD, et al. What Clinicians Want: Contextualizing Explainable Machine Learning for Clinical End Use. *arXiv* 2019; 1905.05134. Accessed April 30, 2024. https://arxiv.org/abs/1905.05134
80. Wang F, Kaushal R, Khullar D. Should Health Care Demand Interpretable Artificial Intelligence or Accept "Black Box" Medicine? *Ann Intern Med*. 2020;172:59-60.
81. Park JB, Lee HJ, Yang HL, et al. Machine learning-based prediction of intraoperative hypoxemia for pediatric patients. *PLoS One*. 2023;18:e0282303.
82. Shapley LS. *Notes on the n-Person Game—II: the Value of an n-Person Game*. Rand Corporation; 1951.
83. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017.
84. Levine DM, Tuwani R, Kompa B, et al. The Diagnostic and Triage Accuracy of the GPT-3 Artificial Intelligence Model. *medRxiv*. 2023. Preprint posted online February 1, 2023. Accessed April 30, 2024. https://doi.org/10.1101/2023.01.30.23285067
85. Sarbay İ, Berikol GB, Özturan İU. Performance of emergency triage prediction of an open access natural language processing based chatbot application (ChatGPT): A preliminary, scenario-based cross-sectional study. *Turk J Emerg Med*. 2023;23:156-161.
86. OpenAI. Guides/Embeddings. *Docs* 2023. Accessed March, 2023. https://platform.openai.com/docs/guides/embeddings
87. Lee TB, Trott S. A jargon-free explanation of how AI large language models work. *Ars Technica*. 2023. Accessed April 30, 2024. https://arstechnica.com/science/2023/07/a-jargon-free-explanation-of-how-ai-large-language-models-work/8/#:~:text=In%202020%2C%20OpenAI%20released%20GPT,significantly%20larger%20than%20GPT%2D3
88. Huang S, Mamidanna S, Jangam S, et al. Can Large Language Models Explain Themselves? A Study of LLM-Generated Self-Explanations. *arXiv* 2023; 2310.11207. Accessed January, 2024. https://arxiv.org/abs/2310.11207
89. Zhao H, Chen H, Yang F, et al. Explainability for Large Language Models: A Survey. *arXiv* 2023; 2309.01029. Accessed January, 2024. https://arxiv.org/abs/2309.01029
90. Zhao J, Yao Z, Yang Z, et al. SELF-EXPLAIN: Teaching Large Language Models to Reason Complex Questions by Themselves. *arXiv* 2023; 2311.06985. Accessed January, 2024. https://arxiv.org/abs/2311.06985
91. Chen Y, Zhong R, Ri N, et al. Do Models Explain Themselves? Counterfactual Simulatability of Natural Language Explanations. *arXiv* 2023; 2307.08678. Accessed January, 2024. https://arxiv.org/abs/2307.08678
92. Rancourt F, Vondrlik P, Maupomé D, et al. Investigating Self-Rationalizing Models for Commonsense Reasoning. *Stats*. 2023;6:907-919.
93. Singhal K, Tu T, Gottweis J, et al. Towards Expert-Level Medical Question Answering with Large Language Models. *arXiv* 2023; 2305.09617. Accessed January, 2024. https://arxiv.org/abs/2305.09617
94. Gao Y, Li R, Caskey J, et al. Leveraging A Medical Knowledge Graph into Large Language Models for Diagnosis Prediction. *arXiv* 2023; 2308.14321. Accessed January, 2024. https://arxiv.org/abs/2308.14321
95. Hirosawa T, Harada Y, Yokose M, et al. Diagnostic Accuracy of Differential-Diagnosis Lists Generated by Generative Pretrained Transformer 3 Chatbot for Clinical Vignettes with Common Chief Complaints: A Pilot Study. *Int J Environ Res Public Health*. 2023;20:3378.
96. Suwanvecho S, Suwanrusme H, Jirakulaporn T, et al. Comparison of an oncology clinical decision-support system's recommendations with actual treatment decisions. *J Am Med Inform Assoc*. 2021;28:832-838.
97. Shapiro MA, Stuhlmiller TJ, Wasserman A, et al. AI-Augmented Clinical Decision Support in a Patient-Centric Precision Oncology Registry. *AI in Precision Oncology*. 2023;1:58-68.
98. Berg Ht, van Bakel B, van de Wouw L, et al. ChatGPT and Generating a Differential Diagnosis Early in an Emergency Department Presentation. *Ann Emerg Med*. 2024;83:83-86.
99. Ayers JW, Poliak A, Dredze M, et al. Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum. *JAMA Intern Med*. 2023;183:589-596.
100. Liévin V, Hother CE, Motzfeldt AG, et al. Can large language models reason about medical questions? *arXiv* 2023; 2207.08143. Accessed January, 2024. https://arxiv.org/abs/2207.08143
101. Johnson D, Goodman R, Patrinely J, et al. Assessing the Accuracy and Reliability of AI-Generated Medical Responses: An Evaluation of the Chat-GPT Model. Research Square Platform LLC. 2023. Accessed April 30, 2024. https://doi.org/10.21203/rs.3.rs-2566942/v1
102. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digit Health*. 2023;2:e0000198.
103. Ji Z, Lee N, Frieske R, et al. Survey of Hallucination in Natural Language Generation. *ACM Comput Surv*. 2023;55:248.
104. Emsley R. ChatGPT: These are not hallucinations – they're fabrications and falsifications. *Schizophrenia*. 2023;9:52.
105. OpenAI. GPT-4. *OpenAI | Research* 2023. Accessed January, 2024. https://openai.com/research/gpt-4
106. Tamayo-Sarver J. I'm an ER doctor: Here's what I found when I asked ChatGPT to diagnose my patients. Fast Company| Future of Health. 2023. Accessed January, 2024. https://www.fastcompany.com/90863983/chatgpt-medical-diagnosis-emergency-room.
107. Grossman S, Zerilli T, Nathan J. ChatGPT: Evaluation of Its Ability to Respond to Drug Information Questions. American Society of Health-System Pharmacists Midyear Clinical Meeting & Exhibition; 2023.
108. Liyan T, Zhaoyi S, Betina I, et al. Evaluating Large Language Models on Medical Evidence Summarization. medRxiv. 2023. 2023.04.22.23288967. Accessed April 30, 2024. https://www.medrxiv.org/content/10.1101/2023.04.22.23288967v1
109. Nastasi AJ, Courtright KR, Halpern SD, et al. A vignette-based evaluation of ChatGPT's ability to provide appropriate and equitable medical advice across care contexts. *Scientific Reports*. 2023;13:17885.
110. Dash D, Thapa R, Banda JM, et al. Evaluation of GPT-3.5 and GPT-4 for supporting real-world information needs in healthcare delivery. *arXiv* 2023; 2304.13714. Accessed April 30, 2024. https://doi.org/10.48550/arXiv.2304.13714
111. Callahan A, Gombar S, Cahan EM, et al. Using Aggregate Patient Data at the Bedside via an On-Demand Consultation Service. *NEJM Catalyst*. 2021;2.
112. Ashenburg N, Preiksaitis C, Dayton J, et al. 312 When AI Meets the Emergency Department: Realizing the Benefits of Large Language Models in Emergency Medicine. *Ann Emerg Med*. 2023;82(4, Supplement):S136-S137.
113. Wong A, Otles E, Donnelly JP, et al. External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients. *JAMA Intern Med*. 2021;181:1-6.
114. Plana D, Shung DL, Grimshaw AA, et al. Randomized Clinical Trials of Machine Learning Interventions in Health Care: A Systematic Review. *JAMA Network Open*. 2022;5:e2233946.

115. Yin J, Ngiam KY, Teo HH. Role of Artificial Intelligence Applications in Real-Life Clinical Practice: Systematic Review. *J Med Internet Res*. 2021;23:e25759.
116. Keane PA, Topol EJ. With an eye to AI and autonomous diagnosis. *NPJ Digit Med*. 2018;1:40.
117. Aristidou A, Jena R, Topol EJ. Bridging the chasm between AI and clinical implementation. *Lancet*. 2022;399:620.
118. Richardson P, Griffin I, Tucker C, et al. Baricitinib as potential treatment for 2019-nCoV acute respiratory disease. *Lancet*. 2020;395:e30-e31.
119. Stebbing J, Krishnan V, de Bono S, et al. Mechanism of baricitinib supports artificial intelligence-predicted testing in COVID-19 patients. *EMBO Mol Med*. 2020;12:e12697.
120. Stebbing J, Sánchez Nievas G, Falcone M, et al. JAK inhibition reduces SARS-CoV-2 liver infectivity and modulates inflammatory responses to reduce morbidity and mortality. *Science Advances*. 2021;7:eabe4724.
121. Cantini F, Niccoli L, Nannini C, et al. Beneficial impact of Baricitinib in COVID-19 moderate pneumonia; multicentre study. *J Infect*. 2020;81:647-679.
122. Kalil AC, Patterson TF, Mehta AK, et al. Baricitinib plus Remdesivir for Hospitalized Adults with Covid-19. *N Engl J Med*. 2021;384:795-807.
123. Zech JR, Badgeley MA, Liu M, et al. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. *PLoS Med*. 2018;15:e1002683.
124. Ramgopal S, Sanchez-Pinto LN, Horvat CM, et al. Artificial intelligence-based clinical decision support in pediatrics. *Pediatr Res*. 2023;93:334-341.
125. Berlin L. Radiologic Errors and Malpractice: A Blurry Distinction. *AJR Am J Roentgenol*. 2007;189:517-522.
126. Northcutt CG, Athalye A, Mueller J. Pervasive Label Errors in Test Sets Destabilize Machine Learning Benchmarks. *arXiv* 2021; 2103.14749. Accessed June, 2023. https://arxiv.org/abs/2103.14749
127. Schmidt C. M. D. Anderson Breaks With IBM Watson, Raising Questions About Artificial Intelligence in Oncology. *J Natl Cancer Inst*. 2017;109. https://doi.org/10.1093/jnci/djx113
128. Herper M. MD Anderson Benches IBM Watson In Setback For Artificial Intelligence In Medicine. *Forbes* 2017. Accessed April 30, 2024. https://www.forbes.com/sites/matthewherper/2017/02/19/md-anderson-benches-ibm-watson-in-setback-for-artificial-intelligence-in-medicine/?sh=3b9522be3774
129. Sirugo G, Williams SM, Tishkoff SA. The Missing Diversity in Human Genetic Studies. *Cell*. 2019;177:26-31.
130. Heaven WD. Hundreds of AI tools have been built to catch covid. None of them helped. *MIT Technology Review* 2021. Accessed April, 2023. https://www.technologyreview.com/2021/07/30/1030329/machine-learning-ai-failed-covid-hospital-diagnosis-pandemic/
131. Iserson KV, Allan NG, Geiderman JM, et al. Audiovisual recording in the emergency department: Ethical and legal issues. *Am J Emerg Med*. 2019;37:2248-2252.
132. Davoudi A, Malhotra KR, Shickel B, et al. Intelligent ICU for Autonomous Patient Monitoring Using Pervasive Sensing and Deep Learning. *Sci Rep*. 2019;9:8020.
133. Lennon C. Protecting patients with live video monitoring and analytics. Modern Healthcare. 2021. Accessed April 30, 2024. https://www.modernhealthcare.com/technology/protecting-patients-live-video-monitoring-and-analytics
134. HHS. Research. *HIPAA|Special Topics* 2003. Accessed April, 2023. https://www.hhs.gov/hipaa/for-professionals/special-topics/research/index.html
135. Eschrich SA, Teer JK, Reisman P, et al. Enabling Precision Medicine in Cancer Care Through a Molecular Data Warehouse: The Moffitt Experience. *JCO Clin Cancer Inform*. 2021;5:561-569.
136. Boston Medical Center. BMC Clinical Data Warehouse (CDW) for Research. *Research at Boston Medical Center* 2023. Accessed April, 2023. https://www.bmc.org/research/clinical-data-warehouse-cdw
137. Stanford Medicine. Clinical Data Warehouse Reimagined. *Research IT|News* 2023. Accessed April, 2023. https://med.stanford.edu/researchit/news/CDW-reimagined.html
138. Timbie JW, Rudin RS, Towe VL, et al. *National Patient-Centered Clinical Research Network (PCORnet) Phase I: Final Evaluation Report*. Santa Monica, CA: RAND Corporation; 2015.
139. Visweswaran S, Becich MJ, D'Itri VS, et al. Accrual to Clinical Trials (ACT): A Clinical and Translational Science Award Consortium Network. *JAMIA Open*. 2018;1:147-152.
140. Yu YW, Weber GM. Balancing Accuracy and Privacy in Federated Queries of Clinical Data Repositories: Algorithm Development and Validation. *J Med Internet Res*. 2020;22:e18735.
141. FRA. Article 8 - Protection of personal data. *EU Charter of Fundamental Rights* 2007. Accessed April, 2023. http://fra.europa.eu/en/eu-charter/article/8-protection-personal-data
142. Casey B, Farhangi A, Vogl R. Rethinking Explainable Machines: The GDPR's "Right to Explanation" Debate and the Rise of Algorithmic Audits in Enterprise. *Berkeley Technology Law Journal*. 2019;34:143-188.
143. Klosowski T. The State of Consumer Data Privacy Laws in the US (And Why It Matters). *NY Times|Wirecutter*; 2021. Accessed April 30, 2024. https://www.nytimes.com/wirecutter/blog/state-of-privacy-laws-in-us/
144. Zhu K. The State of State AI Laws: 2023. 2023. Accessed April 30, 2024. https://epic.org/the-state-of-state-ai-laws-2023/
145. Bonta R. California Consumer Privacy Act (CCPA). *State of California Department of Justice*; 2023. Accessed April, 2023. https://oag.ca.gov/privacy/ccpa#sectionb
146. Zuo Z, Watson M, Budgen D, et alAl Moubayed N. Data Anonymization for Pervasive Health Care: Systematic Literature Mapping Study. *JMIR Med Inform*. 2021;9:e29871.
147. Agrawal M, Hegselmann S, Lang H, et al. Large language models are few-shot clinical information extractors. *arXiv* 2022; 2205.12689. Accessed April 30, 2024. https://arxiv.org/abs/2205.12689
148. FDA. PAPNET (R) TESTING SYSTEM. *FDA|Premarket Approval* 1995. Accessed April, 2023. https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpma/pma.cfm?id=P940029
149. FDA. Center for Devices and Radiological Health. FDA Organization 2023. Accessed April, 2023. https://www.fda.gov/about-fda/fda-organization/center-devices-and-radiological-health
150. FDA. Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices. *Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices* 2022. Accessed April, 2023. https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices
151. FDA. The Software Precertification (Pre-Cert) Pilot Program: Tailored Total Product Lifecycle Approaches and Key Findings. *U.S. Food and Drug Administration|Media* 2022. Accessed April 30, 2024. https://www.fda.gov/medical-devices/digital-health-center-excellence/digital-health-software-precertification-pre-cert-pilot-program
152. Schmieding ML, Kopka M, Schmidt K, et al. Triage Accuracy of Symptom Checker Apps: 5-Year Follow-up Evaluation. *J Med Internet Res*. 2022;24:e31810.
153. Nundy S, Razi RR, Dick JJ, et al. A text messaging intervention to improve heart failure self-management after hospital discharge in a largely African-American population: before-after study. *J Med Internet Res*. 2013;15:e53.
154. Hägglund E, Lyngå P, Frie F, et al. Patient-centred home-based management of heart failure. Findings from a randomised clinical trial evaluating a tablet computer for self-care, quality of life and effects on knowledge. *Scand Cardiovasc J*. 2015;49:193-199.
155. Flores Mateo G, Granado-Font E, Ferré-Grau C, et al. Mobile Phone Apps to Promote Weight Loss and Increase Physical Activity: A

Systematic Review and Meta-Analysis. *J Med Internet Res*. 2015;17:e253.
156. Wu Y, Yao X, Vespasiani G, et al. Mobile App-Based Interventions to Support Diabetes Self-Management: A Systematic Review of Randomized Controlled Trials to Identify Functions Associated with Glycemic Efficacy. *JMIR Mhealth Uhealth*. 2017;5:e35.
157. Smith A. Using Artificial Intelligence and Algorithms. *FTC|Business Blog* 2020. Accessed April, 2023. https://www.ftc.gov/business-guidance/blog/2020/04/using-artificial-intelligence-and-algorithms
158. Phillips PJ, Hahn CA, Fontana PC, et al. Four Principles of Explainable Artificial Intelligence. *NIST Interagency/Internal Report (NISTIR 8312)*. Gaithersburg, MD: National Institute of Standards and Technology; 2021. Accessed April 30, 2024. https://nvlpubs.nist.gov/nistpubs/ir/2021/NIST.IR.8312.pdf
159. FTC. Commercial Surveillance and Data Security Rulemaking. *FTC|Federal Register Notices* 2022. Accessed April, 2023. https://www.ftc.gov/legal-library/browse/federal-register-notices/commercial-surveillance-data-security-rulemaking
160. National Institute of Standards and Technology. AI RMF Core. *Trustworthy & Responsible AI Resource Center* 2023. Accessed April, 2023. https://airc.nist.gov/AI_RMF_Knowledge_Base/AI_RMF/Core_And_Profiles/5-sec-core#tab:govlongtblr
161. Sisto A, Halm KC, Seiver JD. NIST Releases Final Risk Management Framework for Developing Trustworthy AI. *Artificial Intelligence Law Advisor* 2023. Accessed August, 2023. https://www.dwt.com/blogs/artificial-intelligence-law-advisor/2023/01/ai-risk-management-framework-nist#:∼:text=The%20RMF%20is%20a%20non,the%20development%20of%20trustworthy%20AI
162. Kerry CF. NIST's AI Risk Management Framework plants a flag in the AI debate. *Brookings*. 2023. Accessed August, 2023. https://www.brookings.edu/articles/nists-ai-risk-management-framework-plants-a-flag-in-the-ai-debate/
163. Rao A, Pang M, Kim J, et al. Assessing the Utility of ChatGPT Throughout the Entire Clinical Workflow: Development and Usability Study. *J Med Internet Res*. 2023;25:e48659.
164. Fuhrman JD, Gorre N, Hu Q, et al. A review of explainable and interpretable AI with applications in COVID-19 imaging. *Med Phys*. 2022;49:1-14.
165. Sánchez-Cabo F, Rossello X, Fuster V, et al. Machine Learning Improves Cardiovascular Risk Definition for Young, Asymptomatic Individuals. *J Am Coll Cardiol*. 2020;76:1674-1685.
166. Palm V, Norajitra T, von Stackelberg O, et al. AI-Supported Comprehensive Detection and Quantification of Biomarkers of Subclinical Widespread Diseases at Chest CT for Preventive Medicine. *Healthcare (Basel)*. 2022;10:2166.
167. Nathan J. Four Ways Artificial Intelligence Can Benefit Robotic Surgery. *Forbes|Innovation*. 2023. Accessed April, 2023. https://www.forbes.com/sites/forbestechcouncil/2023/02/15/four-ways-artificial-intelligence-can-benefit-robotic-surgery/?sh=2128bdd859f8
168. Laranjo L, Dunn AG, Tong HL, et al. Conversational agents in healthcare: a systematic review. *J Am Med Inform Assoc*. 2018;25:1248-1258.
169. Jacobs M, Pradier MF, McCoy TH, et al. How machine-learning recommendations influence clinician treatment selections: the example of antidepressant selection. *Transl Psychiatry*. 2021;11:108.
170. Maliha G, Gerke S, Cohen IG, et al. Artificial Intelligence and Liability in Medicine: Balancing Safety and Innovation. *Milbank Q*. 2021;99:629-647.
171. Husgen J. Product liability suits involving drug or device manufacturers and physicians: the learned intermediary doctrine and the physician's duty to warn. *Mo Med*. 2014;111:478-481.
172. Price WN. Artificial Intelligence in Health Care: Applications and Legal Implications. *The SciTech Lawyer*. 2017;14:10-13.
173. Price WN, Gerke S, Cohen IG. Potential Liability for Physicians Using Artificial Intelligence. *JAMA*. 2019;322:1765-1766.
174. Allain JS. From Jeopardy! to Jaundice: The Medical Liability Implications of Dr. Watson and Other Artificial Intelligence Systems. *La Law Rev*. 2013;73:1049-1079.
175. Chagal-Feferkorn KA. Am I an algorithm or a product: when products liability should apply to algorithmic decision-makers. *Stanford Law Pol Rev*. 2019;30:61-114.
176. Gurchiek RD, Cheney N, McGinnis RS. Estimating Biomechanical Time-Series with Wearable Sensors: A Systematic Review of Machine Learning Techniques. *Sensors (Basel)*. 2019;19:5227.
177. Ronkainen NJ, Halsall HB, Heineman WR. Electrochemical biosensors. *Chem Soc Rev*. 2010;39:1747-1763.
178. Wang C, Chen X, Wang L, et al. Bioadhesive ultrasound for long-term continuous imaging of diverse organs. *Science*. 2022;377:517-523.
179. Varshney KR. Trustworthy machine learning and artificial intelligence. *XRDS: Crossroads, The ACM Magazine for Students*. 2019;25:26-29.
180. Sahiner B, Chen W, Samala RK, et al. Data drift in medical machine learning: implications and potential remedies. *Br J Radiol*. 2023;96:20220878.
181. Vela D, Sharp A, Zhang R, et al. Temporal quality degradation in AI models. *Sci Rep*. 2022;12:11654.
182. Duckworth C, Chmiel FP, Burns DK, et al. Using explainable machine learning to characterise data drift and detect emergent health risks for emergency department admissions during COVID-19. *Sci Rep*. 2021;11:23017.
183. Finlayson SG, Subbaswamy A, Singh K, et al. The Clinician and Dataset Shift in Artificial Intelligence. *N Engl J Med*. 2021;385:283-286.
184. Alaskar K, Tamboli FA, Memon S, et al. Artificial Intelligence (AI) in Healthcare Management. *J Pharm Neg Res*. 2022;13:1011-1020.
185. Harish KB, Price WN, Aphinyanaphongs Y. Open-Source Clinical Machine Learning Models: Critical Appraisal of Feasibility, Advantages, and Challenges. *JMIR Form Res*. 2022;6:e33970.
186. Celik B, Vanschoren J. Adaptation Strategies for Automated Machine Learning on Evolving Data. *IEEE Trans Pattern Anal Mach Intell*. 2021;43:3067-3078.
187. Piratla V. Robustness, Evaluation and Adaptation of Machine Learning Models in the Wild. *arXiv* 2023; 2303.02781. Accessed July, 2023. https://arxiv.org/abs/2303.02781
188. Thakur A, Armstrong J, Youssef A, et al. Self-Aware SGD: Reliable Incremental Adaptation Framework For Clinical AI Models. *IEEE J Biomed Health Inform*. 2023:1624-1634.
189. Bowers KS, Regehr G, Balthazard C, Parker K. Intuition in the context of discovery. *Cognitive Psychology*. 1990;22:72-110.
190. Strathern P. *Mendeleyev's Dream: The Quest for the Elements*. Thomas Dunne Books; 2001.
191. Haffner J. Scientific reasoning requires the irrationality of intuition. *EPFL* 2023. Accessed April, 2023. https://actu.epfl.ch/news/scientific-reasoning-requires-the-irrationality-of/
192. Van den Brink N, Holbrechts B, Brand PLP, et al. Role of intuitive knowledge in the diagnostic reasoning of hospital specialists: a focus group study. *BMJ Open*. 2019;9:e022724.
193. Davis PJ, Hersh R, Hersch R. *Descartes' Dream: The World According to Mathematics*. Harcourt Brace Jovanovich; 1986.
194. Ellenberg J. *How Not to Be Wrong: The Power of Mathematical Thinking*. Penguin Publishing Group; 2014.
195. Fowler GA. ChatGPT can ace logic tests now. But don't ask it to be creative. *The Washington Post* 2023. Accessed April, 2023. https://www.washingtonpost.com/technology/2023/03/18/gpt4-review/
196. Arkoudas K. GPT-4 Can't Reason. *Medium* 2023. Accessed January, 2024. https://medium.com/@konstantine_45825/gpt-4-cant-reason-2eab795e2523
197. Wan Y, Wang W, Yang Y, et al. Triggering Logical Reasoning Failures in Large Language Models. *arXiv* 2024; 2401.00757. Accessed January, 2024. https://arxiv.org/abs/2401.00757

198. Yu P, Wang T, Golovneva O, et al. ALERT: Adapting Language Models to Reasoning Tasks. *arXiv* 2022; 2212.08286. Accessed April 30, 2024. https://arxiv.org/abs/2212.08286
199. Zhang H, Li LH, Meng T, et al. On the paradox of learning to reason from data. *arXiv* 2022; 2205.11502. Accessed April 30, 2024. https://arxiv.org/abs/2205.11502
200. Hornik K, Stinchcombe M, White H. Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. *Neural Networks*. 1990;3:551-560.
201. Sahay M. Neural Networks and the Universal Approximation Theorem. Towards Data Science 2020. Accessed April 30, 2024. https://towardsdatascience.com/neural-networks-and-the-universal-approximation-theorem-8a389a33d30a
202. Crigger E, Khoury C. Making Policy on Augmented Intelligence in Health Care. *AMA J Ethics*. 2019;21:E188-E191.
203. Ramsey M. Self-Driving Cars Could Cut Down on Accidents, Study Says. *WSJ|Tech* 2015. Accessed October, 2023. https://www.wsj.com/articles/self-driving-cars-could-cut-down-on-accidents-study-says-1425567905
204. Metz J. Road Test: Driverless Cars Reduce Injuries To Zero. *Forbes|Advisor* 2023. Accessed October, 2023. https://www.forbes.com/advisor/car-insurance/autonomous-cars-reduce-insurance-claims/#:~:text=Nearly%204%20million%20miles%20driven,by%20Swiss%20Re%20and%20Waymo
205. Sahota N. The Real Question: When Do We Ban Human Drivers? *Forbes|Innovation* 2020. Accessed September, 2023. https://www.forbes.com/sites/neilsahota/2020/10/12/the-real-question-when-do-we-ban-human-drivers/?sh=764a3ace2baa
206. Hatzius J, Briggs J, Kodnani D, et al. The Potentially Large Effects of Artificial Intelligence on Economic Growth *Goldman Sachs|Economics Research* 2023. Accessed June, 2023. https://www.key4biz.it/wp-content/uploads/2023/03/Global-Economics-Analyst_-The-Potentially-Large-Effects-of-Artificial-Intelligence-on-Economic-Growth-Briggs_Kodnani.pdf
207. Ellingrud K, Gupta R, Salguero J. Building the vital skills for the future of work in operations. *McKinsey & Company| Operations* 2020. Accessed June, 2023. https://www.mckinsey.com/capabilities/operations/our-insights/building-the-vital-skills-for-the-future-of-work-in-operations#/
208. Spatharou A, Hieronimus S, Jenkins J. Transforming healthcare with AI: The impact on the workforce and organizations. *Healthcare| Our Insights* 2020. Accessed June, 2023. https://www.mckinsey.com/industries/healthcare/our-insights/transforming-healthcare-with-ai
209. Lewis T. Mystery Mechanisms. *TheScientist* 2016. Accessed April 30, 2024. https://www.the-scientist.com/mystery-mechanisms-33119
210. Ohashi N, Kohno T. Analgesic Effect of Acetaminophen: A Review of Known and Novel Mechanisms of Action. *Front Pharmacol*. 2020;11:580289.
211. Contreras-García IJ, Cárdenas-Rodríguez N, Romo-Mancillas A, et al. Levetiracetam Mechanisms of Action: From Molecules to Systems. *Pharmaceuticals (Basel)*. 2022;15:475.
212. Valente Silva B, Marques J, Nobre Menezes M, et al. Artificial intelligence-based diagnosis of acute pulmonary embolism: Development of a machine learning model using 12-lead electrocardiogram. *Rev Port Cardiol*. 2023;42:643-651.
213. Su D, Li Q, Zhang T, et al. Prediction of acute appendicitis among patients with undifferentiated abdominal pain at emergency department. *BMC Med Res Methodol*. 2022;22:18.
214. Molaei S, Korley FK, Reza Soroushmehr SM, et al. A machine learning based approach for identifying traumatic brain injury patients for whom a head CT scan can be avoided. *Annu Int Conf IEEE Eng Med Biol Soc*. 2016;2016:2258-2261.
215. Reismann J, Romualdi A, Kiss N, et al. Diagnosis and classification of pediatric acute appendicitis by artificial intelligence methods: An investigator-independent approach. *PLoS ONE*. 2019;14:e0222030.
216. Sax DR, Mark DG, Huang J, et al. Use of Machine Learning to Develop a Risk-Stratification Tool for Emergency Department Patients With Acute Heart Failure. *Ann Emerg Med*. 2021;77:237-248.
217. Villacorta H, Pickering JW, Horiuchi Y, et al. Machine learning with D-dimer in the risk stratification for pulmonary embolism: a derivation and internal validation study. *Eur Heart J Acute Cardiovasc Care*. 2022;11:13-19.
218. Liu N, Chee ML, Koh ZX, et al. Utilizing machine learning dimensionality reduction for risk stratification of chest pain patients in the emergency department. *BMC Med Res Methodol*. 2021;21:74.
219. Hsu C-C, Chu C-CJ, Lin C-H, et al. A machine learning model for predicting unscheduled 72 h return visits to the emergency department by patients with abdominal pain. *Diagnostics (Basel)*. 2021;12:82.
220. Taylor RA, Moore CL, Cheung KH, et al. Predicting urinary tract infections in the emergency department with machine learning. *PLoS One*. 2018;13:e0194085.
221. Choi A, Choi SY, Chung K, et al. Development of a machine learning-based clinical decision support system to predict clinical deterioration in patients visiting the emergency department. *Sci Rep*. 2023;13:8561.
222. Hsu C-C, Kao Y, Hsu C-C, et al. Using artificial intelligence to predict adverse outcomes in emergency department patients with hyperglycemic crises in real time. *BMC Endocr Disord*. 2023;23:234.
223. Chiu YM, Courteau J, Dufour I, et al. Machine learning to improve frequent emergency department use prediction: a retrospective cohort study. *Sci Rep*. 2023;13:1981.
224. Park S, Lee C, Lee S-B, et al. Machine learning-based prediction model for emergency department visits using prescription information in community-dwelling non-cancer older adults. *Sci Rep*. 2023;13:18887.
225. Spangler D, Hermansson T, Smekal D, et al. A validation of machine learning-based risk scores in the prehospital setting. *PLoS ONE*. 2019;14:e0226518.
226. Blomberg SN, Folke F, Ersbøll AK, et al. Machine learning as a supportive tool to recognize cardiac arrest in emergency calls. *Resuscitation*. 2019;138:322-329.
227. Li K, Wu H, Pan F, et al. A Machine Learning–Based Model to Predict Acute Traumatic Coagulopathy in Trauma Patients Upon Emergency Hospitalization. *Clin Appl Thromb Hemost*. 2020;26:1076029619897827.
228. Lammers D, Marenco C, Morte K, et al. Machine Learning for Military Trauma: Novel Massive Transfusion Predictive Models in Combat Zones. *J Surg Res*. 2022;270:369-375.
229. Jilani T, Housley G, Figueredo G, et al. Short and Long term predictions of Hospital emergency department attendances. *Int J Med Inform*. 2019;129:167-174.
230. Pak A, Gannon B, Staib A. Predicting waiting time to treatment for emergency department patients. *Int J Med Inform*. 2021;145:104303.