

# Predictive Analytics in Health Care

Neil Sarkar, PhD, MLIS, FACMI, ACHIP

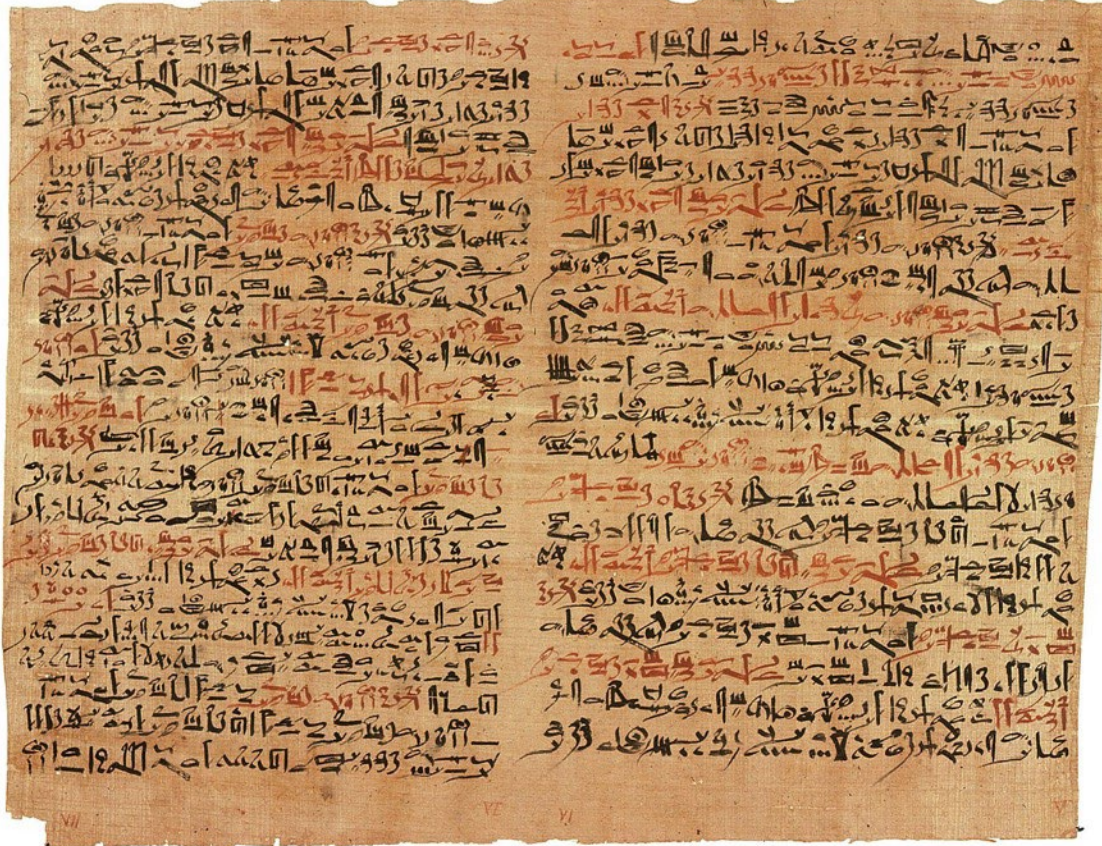
President & CEO

Rhode Island Quality Institute

Associate Professor of Medical Science

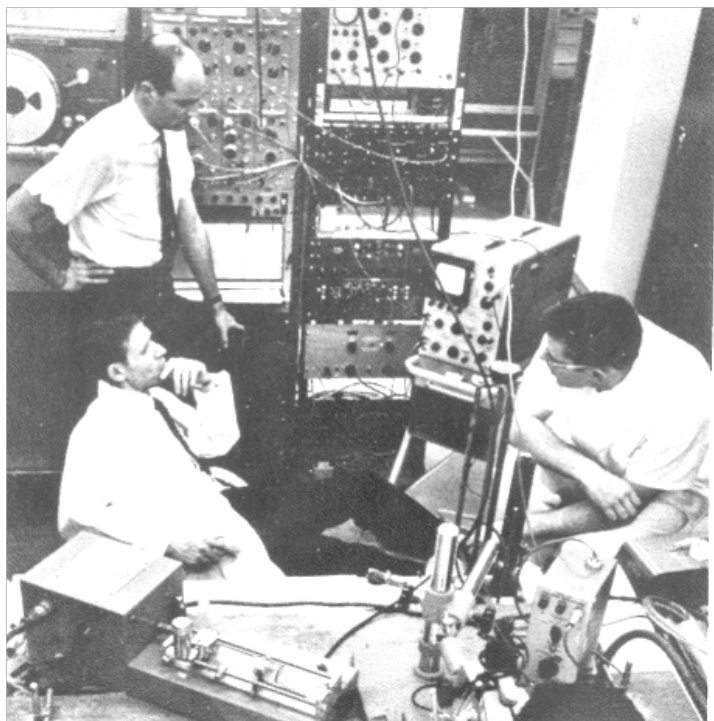
Associate Professor of Health Services, Policy and Practice

Brown University



Edwin Smith Papyrus [c. 1600 BCE]  
(Medical summaries of trauma case studies)

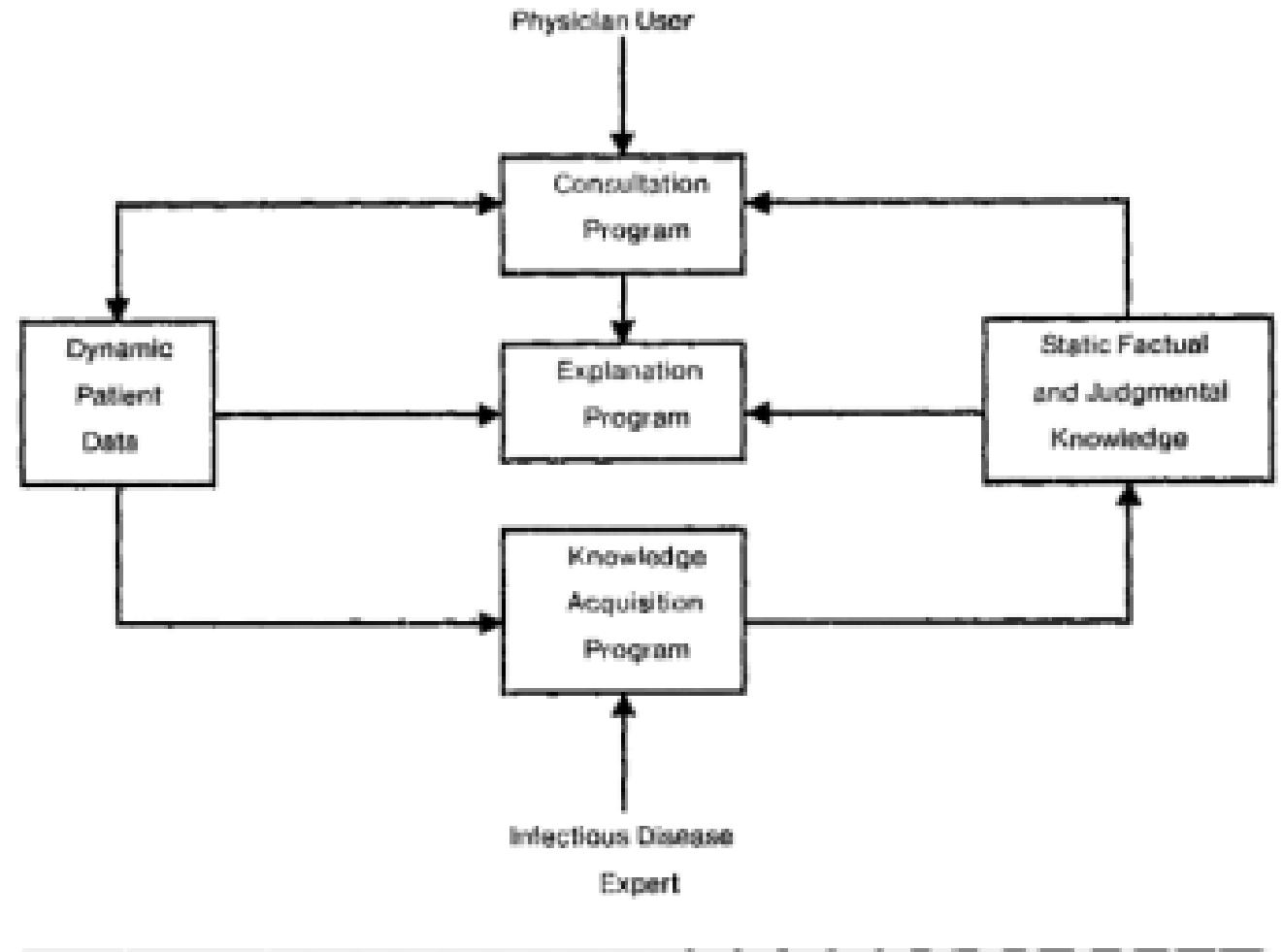
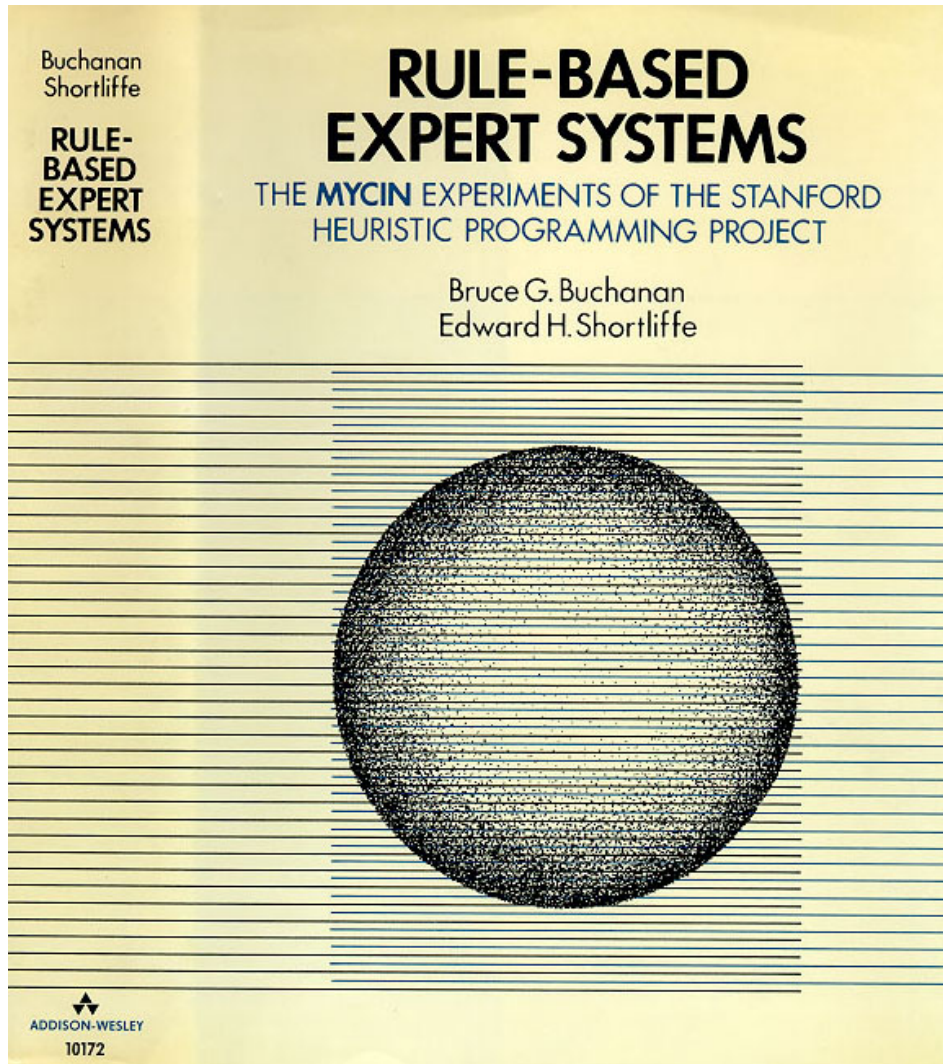




Octo Barnett, MD  
(Development of MUMPS)







# ARTIFICIAL INTELLIGENCE

Early artificial intelligence stirs excitement.



1950's

1960's

1970's

1980's

# MACHINE LEARNING

Machine learning begins to flourish.



1990's

2000's

2010's

# DEEP LEARNING

Deep learning breakthroughs drive AI boom.



<https://www.datasciencecentral.com/profiles/blogs/artificial-intelligence-vs-machine-learning-vs-deep-learning>

# A “Fundamental Theorem” of Biomedical Informatics

---

CHARLES P. FRIEDMAN, PHD

**Abstract** This paper proposes, in words and pictures, a “fundamental theorem” to help clarify what informatics is and what it is not. In words, the theorem stipulates that a person working in partnership with an information resource is “better” than that same person unassisted. The theorem is applicable to health care, research, education, and administrative activities. Three corollaries to the theorem illustrate that informatics is more about people than technology; that in order for the theorem to hold, resources must be informative in addition to being correct; and that the theorem can fail to hold for reasons explained by understanding the interaction between the person and the resource.

■ J Am Med Inform Assoc. 2009;16:169–170. DOI 10.1197/jamia.M3092.



**Figure 1.** A “Fundamental Theorem” of informatics.

# The Right Data at the Right Time

## How can big data change science?

Here's how medical research traditionally works:



**1** Come up with a question or hypothesis.

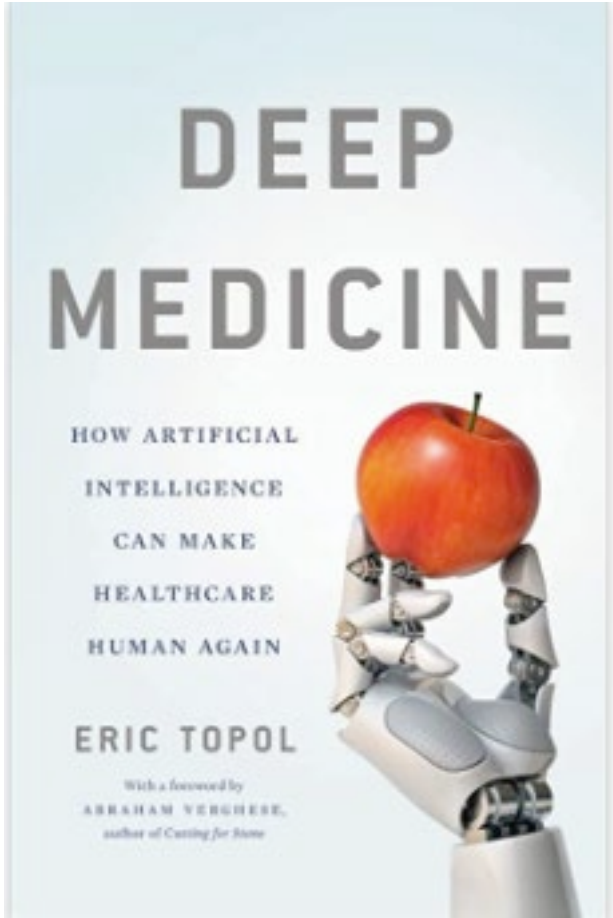


**2** Design an experiment to test it. Wait for new data to come in.



**3** Form your conclusion.

<http://ww2.kqed.org/futureofyou/2014/09/29/how-big-data-is-changing-medicine/>



**Step 1**  
**Collect demonstration data and train a supervised policy.**

A prompt is sampled from our prompt dataset.



A labeler demonstrates the desired output behavior.

This data is used to fine-tune GPT-3.5 with supervised learning.

**Step 2**  
**Collect comparison data and train a reward model.**

A prompt and several model outputs are sampled.

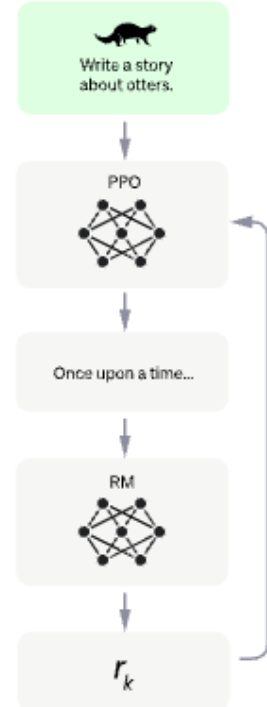


A labeler ranks the outputs from best to worst.

This data is used to train our reward model.

**Step 3**  
**Optimize a policy against the reward model using the PPO reinforcement learning algorithm.**

A new prompt is sampled from the dataset.



The PPO model is initialized from the supervised policy.

The policy generates an output.

The reward model calculates a reward for the output.

The reward is used to update the policy using PPO.

**Limitations**

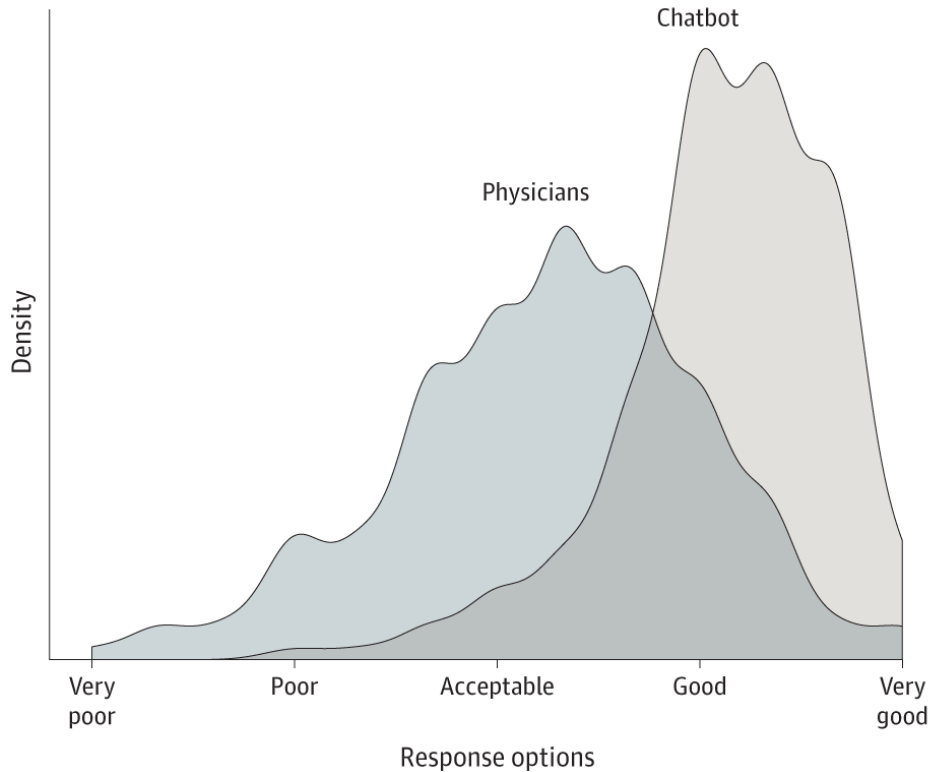
- ChatGPT sometimes writes plausible sounding but incorrect or nonsensical answers. Fixing this issue is challenging as (1) prompt engineering, (2) training the model to be more cautious, (3) better prompts that can avoid ambiguity, and (4) better model training.
- ChatGPT sometimes fails to answer questions that it can answer correctly, and it sometimes writes answers that are not in our prompt dataset. (1) Better prompts, (2) better model training, (3) better model training, (4) better model training.
- ChatGPT sometimes fails to answer questions that it can answer correctly, but given a slight rephrasing, it can answer correctly.
- The model is often excessively verbose and contains certain phrases, such as stating that it is a language model trained by OpenAI. These issues arise from biases in the training data (e.g., longer answers that look more comprehensive and well-tuned can optimize better).
- In some cases, the model will not answer questions unless the user provides an explicit query context, but current models usually provide what the user intended.
- While we do make efforts to make the model follow responsible requests, it will sometimes respond to harmful instructions or exhibit biased behavior. We're using the latest GPT-3.5 to serve in this version. Types of harmful content that we expect to have some bias against are: (1) hate speech, (2) sexual content, (3) child sexual abuse material, (4) self-harm, (5) suicide, (6) terrorism, (7) violence, (8) illegal activities, (9) disinformation, (10) harassment, (11) discrimination, (12) harassment, (13) harassment, (14) harassment, (15) harassment.



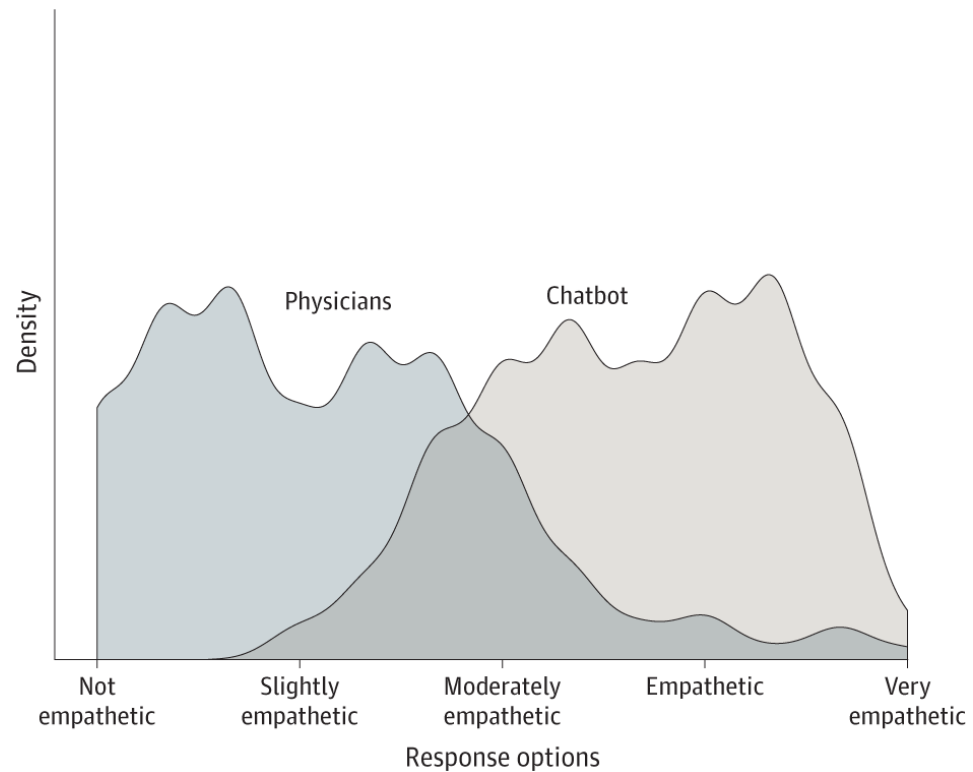
April 28, 2023

# Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum

A Quality ratings



B Empathy ratings



**Limitations**

- ChatGPT sometimes writes plausible sounding but incorrect or nonsensical answers. Finding this issue to be challenging as it depends on training. There's currently no source of truth. (2) Training the model to be more accurate requires a better question that is not answer specific. (3) Improving training requires that the model see more of the user answer. (4) ChatGPT will still have biases, such as when it will not respond, give and phrasing of a question, the model can claim to not know the answer, but give a slight response, and answer correctly.
- The model is often excessively verbose and contains certain phrases, such as stating that it's a language model trained by OpenAI. These issues arise from biases in the training data (people come longer answers that look more comprehensive and well thought out - optimization issues).<sup>17</sup>
- Ideally, the model should ask clarifying questions when the user provides an ambiguous query. Instead, our current model usually gives what the user intended.
- While we've made efforts to make the model utilize responsive requests, it will sometimes require helpful instructions or exhibit biased behavior. We're using the Microsoft API to view or block certain types of user content, but we expect to have some bias requests and problems for now. We're eager to collect user feedback to aid our ongoing work to improve this system.

# Limitations

- ChatGPT sometimes writes plausible-sounding but incorrect or nonsensical answers. Fixing this issue is challenging, as: (1) during RL training, there's currently no source of truth; (2) training the model to be more cautious causes it to decline questions that it can answer correctly; and (3) supervised training misleads the model because the ideal answer depends on what the model knows, rather than what the human demonstrator knows.
- ChatGPT is sensitive to tweaks to the input phrasing or attempting the same prompt multiple times. For example, given one phrasing of a question, the model can claim to not know the answer, but given a slight rephrase, can answer correctly.
- The model is often excessively verbose and overuses certain phrases, such as restating that it's a language model trained by OpenAI. These issues arise from biases in the training data (trainers prefer longer answers that look more comprehensive) and well-known over-optimization issues.<sup>1, 2</sup>
- Ideally, the model would ask clarifying questions when the user provided an ambiguous query. Instead, our current models usually guess what the user intended.
- While we've made efforts to make the model refuse inappropriate requests, it will sometimes respond to harmful instructions or exhibit biased behavior. We're using the Moderation API to warn or block certain types of unsafe content, but we expect it to have some false negatives and positives for now. We're eager to collect user feedback to aid our ongoing work to improve this system.

## Desiderata for Ideal Algorithms in Health Care



### *Explainable*

conveys the relative importance of features in determining outputs



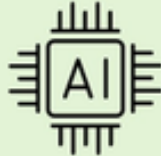
### *Dynamic*

captures temporal changes in physiologic signals and clinical events



### *Precise*

data frequency matches physiology, architecture is aptly complex



### *Autonomous*

executes without time-consuming, manual data entry



### *Fair*

evaluates and mitigates implicit bias and social inequity



### *Reproducible*

validated externally and prospectively, shared with research communities

Loftus TJ, Tighe PJ, Ozrazgat-Baslanti T, Davis JP, Ruppert MM, et al. (2022) Ideal algorithms in healthcare: Explainable, dynamic, precise, autonomous, fair, and reproducible. PLOS Digital Health 1(1): e0000006. <https://doi.org/10.1371/journal.pdig.0000006>  
<https://journals.plos.org/digitalhealth/article?id=10.1371/journal.pdig.0000006>

Table 1. Checklist for ideal algorithms in healthcare.

Desiderata	Criteria	Yes	Location	No	N/A
<b>Explainable</b> Yes <sup>a</sup> Partially <sup>b</sup> No <sup>c</sup> N/A <sup>d</sup>	<i>Feature importance</i> : conveys the relative importance of features in determining algorithm outputs				
	<i>Descriptive accuracy</i> : describes what the algorithm has learned (e.g., weights in a neural network)				
	<i>Simulatability</i> : clinicians can understand and mentally simulate the model's process for generating predictions				
	<i>Relevance</i> : describes relevancy as judged by the algorithm's target human audience				
<b>Dynamic</b> Yes <sup>a</sup> Partially <sup>b</sup> No <sup>c</sup> N/A <sup>d</sup>	<i>Temporality</i> : captures temporal changes in physiologic signals and clinical events				
	<i>Continuous monitoring</i> : performance is reassessed at several time points, including the point at which performance is expected to plateau				
<b>Precise</b> Yes <sup>a</sup> Partially <sup>b</sup> No <sup>c</sup> N/A <sup>d</sup>	<i>Data frequency</i> : rate of data collection matches the rate of physiologic changes				
	<i>Complexity</i> : algorithm complexity matches the complexity of the prediction or classification task				
<b>Autonomous</b> Yes <sup>a</sup> Partially <sup>b</sup> No <sup>c</sup> N/A <sup>d</sup>	<i>Efficiency</i> : the algorithm executes without the need for time-consuming, manual data entry by the end user (i.e., patient, provider, or investigator)				
<b>Fair</b> Yes <sup>a</sup> Partially <sup>b</sup> No <sup>c</sup> N/A <sup>d</sup>	<i>Generalizability</i> : algorithm is developed and validated across diverse patient demographics and practice settings				
	<i>Selectivity</i> : excludes features that lack pathophysiologic or linguistic association with outcomes, but may introduce bias				
	<i>Objectivity</i> : includes variables that are minimally influenced by clinician judgments (e.g., vital signs)				
<b>Reproducible</b> Yes <sup>a</sup> Partially <sup>b</sup> No <sup>c</sup> N/A <sup>d</sup>	<i>Generalizability</i> : validated externally, prospectively				
	<i>Collaboration</i> : algorithm is shared with the research community				
	<i>Compliance</i> : fulfills SPIRIT-AI extension guidelines (if trial) and fulfills CONSORT-AI guidelines				

<sup>a</sup>Overall adjudication is "Yes" when all criteria are either met or not applicable.

<sup>b</sup>Overall adjudication is "Partially" when some but not all criteria are either met or not applicable.

<sup>c</sup>Overall adjudication is "No" when no criteria are met.

<sup>d</sup>Overall adjudication is "N/A" when all criteria are not applicable.

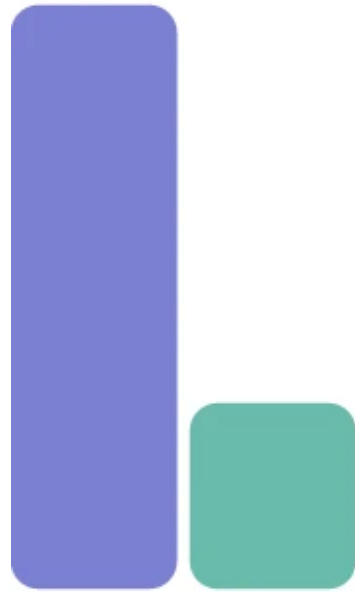
CONSORT-AI, Consolidated Standards of Reporting Trials-Artificial Intelligence; N/A, not applicable; SPIRIT-AI, Standard Protocol Items: Recommendations for Interventional Trials-Artificial Intelligence.

<https://doi.org/10.1371/journal.pdig.0000006.t001>

Loftus TJ, Tighe PJ, Ozrazgat-Baslanti T, Davis JP, Ruppert MM, et al. (2022) Ideal algorithms in healthcare: Explainable, dynamic, precise, autonomous, fair, and reproducible. PLOS Digital Health 1(1): e0000006. <https://doi.org/10.1371/journal.pdig.0000006>  
<https://journals.plos.org/digitalhealth/article?id=10.1371/journal.pdig.0000006>



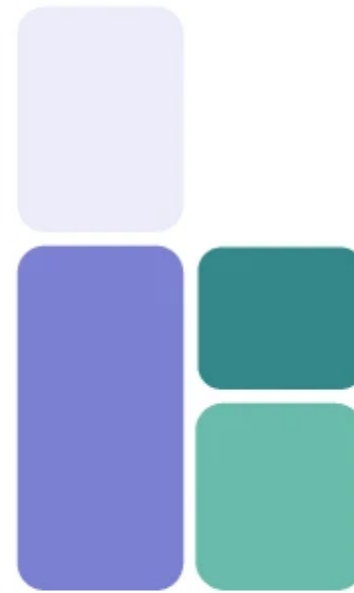
● Majority ● Minority ● Added ● Removed



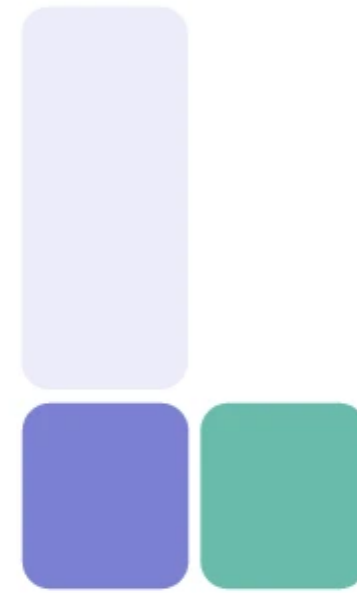
Original



Oversampling



Hybrid



Undersampling

Imbalanced data preprocessing techniques for machine learning: a systematic mapping study

Vitor Werner de Vargas, Jorge Arthur Schneider Aranda, Ricardo dos Santos Costa, Paulo Ricardo da Silva Pereira & Jorge Luis Victória Barbosa

Knowledge and Information Systems 65, 31–57 (2023) | Cite this article

April 15, 2021

# Algorithmovigilance—Advancing Methods to Analyze and Monitor Artificial Intelligence-Driven Health Care for Effectiveness and Equity

Peter J. Embi, MD, MS<sup>1</sup>

[» Author Affiliations](#) | [Article Information](#)

*JAMA Netw Open.* 2021;4(4):e214622. doi:10.1001/jamanetworkopen.2021.4622



**FDA** U.S. FOOD & DRUG  
ADMINISTRATION

## The Software Precertification (Pre-Cert) Pilot Program: Tailored Total Product Lifecycle Approaches and Key Findings

September 2022

# Thank You!

**Neil Sarkar, PhD, MLIS, FACMI, ACHIP**

President & CEO

Rhode Island Quality Institute

Associate Professor of Medical Science

Associate Professor of Health Services, Policy and Practice

Brown University

[nsarkar@riqi.org](mailto:nsarkar@riqi.org)