Artificial Intelligence in Perioperative Medicine: A Proposed Common Language With Applications to FDA-Approved Devices

Ryan L. Melvin*, Matthew G. Broyles¹, Elizabeth W. Duggan¹, Sonia John¹, Andrew D. Smith² and Dan E. Berkowitz¹

¹Department of Anesthesiology and Perioperative Medicine, University of Alabama at Birmingham, Birmingham, AL, United States, ²Department of Radiology, University of Alabama at Birmingham, Birmingham, AL, United States

As implementation of artificial intelligence grows more prevalent in perioperative medicine, a clinician’s ability to distinguish differentiating aspects of these algorithms is critical. There are currently numerous marketing and technical terms to describe these algorithms with little standardization. Additionally, the need to communicate with algorithm developers is paramount to actualize effective and practical implementation. Of particular interest in these discussions is the extent to which the output or predictions of algorithms and tools are understandable by medical practitioners. This work proposes a simple nomenclature that is intelligible to both clinicians and developers for quickly describing the interpretability of model results. There are three high-level categories: transparent, translucent, and opaque. To demonstrate the applicability and utility of this terminology, these terms were applied to the artificial intelligence and machine-learning-based products that have gained Food and Drug Administration approval. During this review and categorization process, 22 algorithms were found with perioperative utility (in a database of 70 total algorithms), and 12 of these had publicly available citations. The primary aim of this work is to establish a common nomenclature that will expedite and simplify descriptions of algorithm requirements from clinicians to developers and explanations of appropriate model use and limitations from developers to clinicians.

Keywords: artificial intelligence, AI, machine learning, algorithm, FDA approval

INTRODUCTION

The list of medical uses for Artificial Intelligence (AI) and Machine Learning (ML) is expanding rapidly (1). Recently, this trend has been particularly true for anesthesiology and perioperative medicine (2, 3). Deriving utility from these algorithms requires medical practitioners and their support staff to sift through a deluge of technical and marketing terms (3). This paper provides an aid for separating the signal of utility from the noise of jargon.

This work proposes a straightforward nomenclature for describing the interpretability and appropriate use of AI/ML products that will be intuitive to developers and clinicians alike. The applicability and utility of this terminology of these terms is then applied to Food and Drug Administration (FDA) approved AI/ML algorithms (1) with perioperative utility. Such a standardized language may speed discussion and understanding among technical developers and
clinical users. To this end, there are three standardized terms for conveying interpretability and indicating the appropriate use of systems and products based on AI/ML. The terms are transparent, translucent, and opaque.

Opaque describes a system that (a) estimates non-linearly applied parameters that require advanced analysis (external to the product itself) to be understood, (b) estimates such a large set of parameters that a human cannot interpret them unless aided by a tool external to the product, or (c) provides a prediction with no indication as to the reason for the prediction. Translucent describes a product that incorporates non-specific methods for assisting the end user in understanding possible reasons for a prediction that would otherwise be categorized as “opaque.” Transparent describes a product that estimates a linearly applied parameter or relatively small set of parameters and is implemented to indicate to the user how much each feature influences the output or indicates the rationale for a prediction (e.g., providing exact weights for features considered). Alternatively, a transparent system’s prediction is easily verifiable at the time the prediction is made using information the system provides.

For example, consider a clinician wanting to predict future glucose values for a patient (4). A careful developer for such a model would be concerned with several factors including whether the clinician desires simply a prediction (opaque) or a prediction with an explanation (translucent or transparent). If an explanation is needed, the developer may wonder how specific to each patient the explanation needs to be. Would a list of factors considered be sufficient (translucent), or does the clinician need to know exactly how much each factor contributed to the prediction for each patient (transparent)?

This imagined developer seeks to understand the level of interpretability required for this algorithm. However, there is also a tradeoff that the developer is considering. Requiring more interpretability limits the types of algorithms that can be used. The deep learning algorithms that currently automate driving, image recognition, and recently folded protein structure estimation (among many other things) tend to lack interpretability (opaque). On the other end of the spectrum, generalized linear models (such as logistic regression) tend toward high interpretability (transparent) but often have less accuracy than deep learning algorithms.

The situation also works in reverse. Consider a developer attempting to explain a new model to a randomly selected clinician. The clinician may wonder how an algorithm works or why it makes certain predictions. For some products, such questions are easy to answer (transparent). For others, it is exceedingly difficult (opaque). And the distinction between such systems at times seems arbitrary. Rather than developers and clinicians continuously engaging in this discussion de novo for every project and collaboration, presented here are three standardized terms for conveying the interpretability and indicating the appropriate use of AI/ML models—transparent, translucent, and opaque.

These ideas behind the terms are not new (5–7). Phrases such as “glass box,” (8) “white box,” (9) “gray box,” (10) “interpretable,” (11) “explainable,” (6, 8, 11–15) and “black box” (8, 10–14) are often used to describe the complexity of algorithms from a somewhat technical perspective. Proposed here are less technically intended terms meant to describe the perspective of the end user rather than the developer—a distinction discussed later with several examples. In addition to being first and foremost clinician-friendly, these terms are intended to convey sufficient technical information to developers for understanding the types of algorithms appropriate for the desired use case. Note that these terms do not describe the underlying mathematics of an algorithm or even the technical details of a particular implementation; rather, they describe the experience of the end user.

For readers familiar with technical usages of “black box,” (8, 10–14) our usage of opaque is similar with the exception it focuses on the experience of the end user and what they reasonably know or are presented with by a specific algorithm implementation. For readers familiar with the concept of “Explainable AI,” the “explainable” piece often refers to a secondary technology applied to a trained AI/ML model that extracts information about why the model makes certain predictions (6, 8, 11–15). An algorithm that makes use of such a technology and shows the output to an end user would be classified under our nomenclature as translucent.

METHODS AND DEFINITIONS

The initial source of algorithms for consideration in this review is a constantly updated online database of FDA-approved algorithms. At the time of this writing, the database contained 70 such algorithms (1). As reported by the primary citation for the database (1), the majority of algorithms in this database were approved with 510(k) clearance. Other approval methods seen in the database are de novo pathway clearance and premarket approval clearance. The database makes broad categorizations of applicable fields for these algorithms. The fields most represented are Radiology, Cardiology, and Internal Medicine/General Practice.

Perioperative medicine is not explicitly mentioned in the database. Therefore, the categorization of “perioperative utility” in this review was made under the best judgment of the authors. The primary purpose of this work, though, is to establish a nomenclature, using the algorithms labeled as having perioperative utility in examples of applying this terminology. Of the 70 algorithms in the database, 22 were determined to have perioperative utility.

Each record in the database includes the name of the algorithm and the parent company. Using this information, along with the details in the corresponding FDA announcements themselves, journal reviewed articles describing the function of these algorithms were sought. This search included—but was not limited to—searching the parent company’s website for mentions of journal articles. From this search, citations for 12 of the 22 algorithms were found.

The categories applied to these 12 algorithms were opaque, translucent, and transparent. Table 1 summarizes these categories. Opaque describes a system that (a) estimates
linear functions of features, variable importance methods, and categorized as “opaque.” Examples include plotting non-linearly applied parameters that require advanced analysis to be understood

Opaque
Concrete example: A clinician desires a prediction of future glucose values for a patient with no need to understand how the prediction was made.

Translucent
Concrete example: A clinician desires a prediction of future glucose values for a patient with an explanation of what clinical factors were involved in making the prediction.

Transparent
Concrete example: A clinician desires a prediction of future glucose values for a patient with an explanation of exactly how much each involved factor contributed to the prediction for each patient.

Table 1
| Category and example usage | Defining features |
|---------------------------|-------------------|
| Opaque                    | - Estimates non-linearly applied parameters that require advanced analysis to be understood |
|                           | - OR estimates such a large set of parameters that a human cannot interpret them unaided |
|                           | - OR provides a prediction with no indication as to the reason for the prediction |
|                           | - Includes techniques for explaining the predictions from an otherwise opaque algorithm |
|                           | - Examples: |
|                           |   - Plotting non-linear functions of features |
|                           |   - Variable importance methods |
|                           |   - Providing a list of factors considered |
|                           | - Estimates a linearly applied parameter or relatively small set of parameters that indicate how much each feature influences the output |
|                           | - OR indicates the specific rationale for a prediction |
|                           | - Examples providing the exact features and weights responsible for a given prediction |
|                           | - OR provides a prediction that is easily verifiable at the time of the prediction |

Table 2

Examples of FDA-Approved Algorithms

Through the process described above, three of the 12 products with a located citations and clear perioperative utility were categorized as opaque (Table 2). These are RhythmAnalytics from Biofourmis Singapore Pte. Ltd., and the Guardian Connect System from Medtronic. RhythmAnalytics along with its underlying Biovitals Analytics Engine monitors, which categorizes cardiac arrhythmias via a convolutional neural network—a deep learning technique—that consumes wavelet transforms and short-time Fourier transforms of electrocardiogram (ECG) signals (19). Such a deep learning technique is inherently opaque, as the number of weights and their non-linear combination makes unaided understanding of the reasons for a prediction not feasible. The Guardian Connect System alerts users to interstitial glucose levels outside of a specified range. The system offers two alert types, “threshold” and “predictive.” The “threshold” alerts are transparent in that they indicate if the sensor glucose reading is above or below a threshold. The “predictive” alerts indicate whether glucose is predicted to be outside the specified range within the next 10–60 min (4). While the manufacturer’s user guide (20) explains that these predictions are formula-based, prediction-based alerts are provided using the name of the prediction, not the reasons for the prediction. The categorizations presented here are based on the information provided to a user when a prediction is provided. Therefore, this system is opaque.

In the translucent category are products that provide a prediction alongside an upfront list of features considered—but no specific weights—or a non-specific visualization of signal.
used in making the prediction. Of the 12 systems, three fit the criteria for this category (Table 2). For example, the physIQ Heart Rhythm and Respiratory Module from physIQ Inc. collapses improving or declining factors related heart failure into a single index. The signals related to this index are viewable by patients and providers along with the corresponding calculated index (21), which led to categorizing this algorithm as translucent. Another cardiac monitoring product, eMurmer ID from CSD Labs GmbH, predicts whether murmurs are pathologic. Along with the prediction, the algorithm shows the systolic and diastolic phases considered; however, the specifics of the predictions “are protected under proprietary regulations,” (22) leading to inclusion of this algorithm in the translucent category. Likewise, the DreaMed algorithm from DreaMed Diabetes, Ltd., produces reports with important features related to its insulin dose recommendations (23, 24), placing it in the translucent category.

The remaining six (of 12) products fall in the transparent category (Table 2). Of these, three are implementations of atrial fibrillation detection that provide either annotated signals along with predictions in the case of ECG App from Apple Inc. (25) and FibriCheck from Qompium NV (26, 27) or use explicit “if...then...” rules in the case of the Irregular Rhythm Notification Feature from Apple Inc. (28). Similarly, the Wave Clinical Platform from Excel Medical Electronics LLC provides vital sign alerts based on a set of rules and provides the reason for the alert when triggered (29, 30). Also, in the transparent category is the EchoMD Automated Ejection Fraction Software which integrates with the Caption Guidance system from Caption Health Inc. to indicate when an echocardiogram transducer is correctly placed for the automated calculation of ejection fraction (31). Since the reasons for the provided feedback are intuitively obvious (the transducer is or is not physically placed correctly), these systems are transparent.

Additional Examples and Future Work
Here we have applied our proposed common language to FDA-approved algorithms, as initial examples of how these terms might be used in clinical contexts. We recognize there are many tools and devices in the world of perioperative medicine used for patient care, education, research, quality improvement, and operations. Covering all of these would be at least a book-length task, well beyond the scope of this project. However, we consider the next step in developing this nomenclature to be a review article addressing the (additional) most common algorithms for patient care. As a small sample, such a review article might include products such as BIS (Bispectral Index) (32, 33), Sedline (34), Datex-Ohmeda Entropy (35), and Edwards Hemosphere (36).

Beyond these additional examples in perioperative medicine, these terms are immediately extensible to other medical fields. While all examples provided herein dealt with algorithms surrounding surgery, note that the definitions themselves (Table 1) are agnostic to any medical subfield.

Additionally, some commentary seems warranted with respect to proprietary algorithms. Indeed, in this review, some products were given opaque or translucent classifications, which may change if the algorithmic details and/or source code for such products were ever released by the intellectual property owners. This dimension of the nomenclature whereby a product categorization could be changed by a public release of information further emphasizes that these terms are from the perspective of the end user rather than technical details.

CONCLUSIONS
This work presents a nomenclature for describing algorithm implementations and applies it to several examples in the
literature. This terminology is composed of three high-level categories: transparent, translucent, and opaque. These terms are applied the point of view of the clinician. To indicate how these terms can be used to categorize AI systems, AI/ML systems with FDA are presented as examples. A database of these examples that will be updated as new systems gain FDA approval is available at https://sites.uab.edu/periopt-datasience/algo-database.

This nomenclature aids in understanding the appropriate use of models. In high-risk situation, the requirement for accuracy may be paramount. Alternatively, in high-profile situations, predictions may need to be explainable to stakeholders. For example, the FDIC in the United States requires financial institutions to develop “conceptually sound” (37) models. An assessment of conceptual soundness would be easiest for transparent models and most difficult for opaque models.

The primary values of common nomenclature are expediting and simplifying descriptions of model requirements and appropriate use between clinicians and developers.

AUTHOR CONTRIBUTIONS

RM performed the initial algorithm search and filtering and prepared the first manuscript draft. MB, ED, and SJ helped assess the perioperative utility of each algorithm and revised the manuscript. AS suggested the framework for the project, helped assess the appropriate category of each algorithm, and revised the manuscript. DB helped organize the project team, helped assess the perioperative utility of each algorithm, and revised the manuscript. All authors contributed to the article and approved the submitted version.

REFERENCES

1. Benjamens S, Dhunnoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. NPJ Digit Med. (2020) 3:1–8. doi: 10.1038/s41746-020-00324-0
2. Mathis MR, Khetarpal S, Najar A. Artificial intelligence for anesthesia: what the practicing clinician needs to know. Anesthesiology. (2018) 129:619–22. doi: 10.1097/ALN.0000000000002384
3. Maheshwari K, Ruetzler K, Saugel B. Perioperative intelligence: applications of artificial intelligence in perioperative medicine. J Clin Monit Comput. (2020) 34:625–8. doi: 10.1007/s10877-019-00379-9
4. Abraham SB, Arunanachalam S, Zhong A, Agrawal P, Cohen O, McMahon CM. Improved real-world glycemic control with continuous glucose monitoring system predictive alerts. J Diabetes Sci Technol. (2019) 15:91–7. doi: 10.1177/193296811959334
5. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R, editors. Advances in Neural Information Processing Systems 30. Long Beach, CA: Curran Associates, Inc. (2017). p. 4765–74. Available online at: http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf
6. Gunning D, Stefik M, Choi J, Miller T, Stumpf S, Yang G-Z. XAI-Explainable artificial intelligence. Sci Robot. (2019) 4:eaay7120.
7. Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. BMC Med Res Methodol. (2019) 2:749–60. doi: 10.1038/s41551-018-0304-0
8. Gunning D, Stefik M, Choi J, Miller T, Stumpf S, Yang G-Z. XAI-Explainable artificial intelligence. Sci Robot. (2019) 4:eaay7120.
9. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R, editors. Advances in Neural Information Processing Systems 30. Long Beach, CA: Curran Associates, Inc. (2017). p. 4765–74. Available online at: http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf
10. Rajput KS, Wibowo S, Hao C, Majmudar M. On Arrhythmia Detection by Deep Learning and Multidimensional Representation. (2019). Available online at: http://arxiv.org/abs/1904.00138
11. Medtronic. Guardian Connect. Medtronic (2017). Available online at: https://guardianconnect.medtronic-diabetes.co.uk
12. Stehlik J, Schmalfuss C, Bozkurt B, Nativi-Nicolaus J, Wohlfahrt P, Wegerich S, et al. Continuous wearable monitoring analytics predict heart failure hospitalization: the link-hf multicenter study. Circ Heart Fail. (2020) 13: e006513. doi: 10.1161/CIRCHEARTFAILURE.119.006513
13. Lai LSW, Redington AN, Reinisch Al, Unterberger MJ, Schriefl MJ. Computerized automatic diagnosis of innocent and pathologic murmurs in pediatrics: a pilot study. Congenit Heart Dis. (2016) 11:386–95. doi: 10.1111/chd.12328
14. Nimri R, Dassau E, Segall T, Muller I, Bratina N, Kordonouri O, et al. Adjusting insulin doses in patients with type 1 diabetes who use insulin pump and continuous glucose monitoring: variations among countries and physicians. Diabetes Obes Metab. (2018) 20:2458–66. doi: 10.1111/dom.13408
15. Nimri R, Oron T, Muller I, Kraljevic I, Alonso MM, Keskinen P, et al. Adjustment of insulin pump settings in type 1 diabetes management: a user-pro device compared to physicians’ recommendations. J Diabetes Sci Technol. (2022) 16:364–72. doi: 10.1177/1932296820965561
16. Holzinger A. From machine learning to explainable AI. In: DISA 2018 - IEEE World Symp Digit Intell Syst Mach Proc. Košice, Slovakia (2018) 55–66. doi: 10.1109/DISA.2018.8490530
17. Bhatt U, Xiang A, Sharma S, Weller A, Taly A, Jia Y, et al. Explainable machine learning in deployment. In: FAT* 2020 - Proc 2020 Conf Fairness, Accountability, Transpar. Barcelona, Spain (2020) 648–57. doi: 10.1145/3351095.3375624
18. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R, editors. Advances in Neural Information Processing Systems 30. Long Beach, CA: Curran Associates, Inc. (2017). p. 4765–74. Available online at: http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf
19. Rajput KS, Wibowo S, Hao C, Majmudar M. On Arrhythmia Detection by Deep Learning and Multidimensional Representation. (2019). Available online at: http://arxiv.org/abs/1904.00138
20. Medtronic. Guardian Connect. Medtronic (2017). Available online at: https://guardianconnect.medtronic-diabetes.co.uk
21. Stehlik J, Schmalfuss C, Bozkurt B, Nativi-Nicolaus J, Wohlfahrt P, Wegerich S, et al. Continuous wearable monitoring analytics predict heart failure hospitalization: the link-hf multicenter study. Circ Heart Fail. (2020) 13: e006513. doi: 10.1161/CIRCHEARTFAILURE.119.006513
22. Lai LSW, Redington AN, Reinisch Al, Unterberger MJ, Schriefl MJ. Computerized automatic diagnosis of innocent and pathologic murmurs in pediatrics: a pilot study. Congenit Heart Dis. (2016) 11:386–95. doi: 10.1111/chd.12328
23. Nimri R, Dassau E, Segall T, Muller I, Bratina N, Kordonouri O, et al. Adjusting insulin doses in patients with type 1 diabetes who use insulin pump and continuous glucose monitoring: variations among countries and physicians. Diabetes Obes Metab. (2018) 20:2458–66. doi: 10.1111/dom.13408
24. Nimri R, Oron T, Muller I, Kraljevic I, Alonso MM, Keskinen P, et al. Adjustment of insulin pump settings in type 1 diabetes management: a user-pro device compared to physicians’ recommendations. J Diabetes Sci Technol. (2022) 16:364–72. doi: 10.1177/1932296820965561
25. Turakhia MP, Desai M, Hedlin H, Rajmane A, Talati N, Ferris T, et al. Rationale and design of a large-scale, app-based study to identify cardiac arrhythmias using a smartphones: the Apple Heart Study. Am Heart J. (2019) 207:66–75. doi: 10.1016/j.ahj.2018.09.002
26. Selder JL, Proemsans T, Breukel L, Dur O, Gielen W, van Rossum AC, et al. Assessment of a standalone photoplethysmography (PPG) algorithm for detection of atrial fibrillation on wristband-derived data. Comput Methods Programs Biomed. (2020) 197:105753. doi: 10.1016/j.cmpb.2020.105753
27. Verbrugge FH, Proemsans T, Vigen J, Mullens W, Rivero-Ayerza M, Van Herendael H, et al. Atrial fibrillation screening with photoplethysmography through a smartphone camera. Europace. (2019) 21:1167–75. doi: 10.1093/europace/euz119
28. Ip JE. Evaluation of cardiac rhythm abnormalities from wearable devices. *JAMA*. (2019) 321:1098. doi: 10.1001/jama.2019.1681

29. Hravnak M, Devita MA, Clontz A, Edwards L, Valenta C, Pinsky MR. Cardiorespiratory instability before and after implementing an integrated monitoring system. *Crit Care Med.* (2011) 39:65–72. doi: 10.1097/CCM.0b013e3181f7b1c

30. Tarassenko L, Hann A, Young D. Integrated monitoring and analysis for early warning of patient deterioration. *Br J Anaesth.* (2006) 97:64–8. doi: 10.1093/bja/ael113

31. Schneider M, Bartko P, Geller W, Dannenberg V, König A, Binder C, et al. A machine learning algorithm supports ultrasound-naive novices in the acquisition of diagnostic echocardiography loops and provides accurate estimation of LVEF. *Int J Cardiovasc Imaging.* (2021) 37:577–86. doi: 10.1007/s10554-020-02046-6

32. Morimoto Y, Hagihira S, Koizumi Y, Ishida K, Matsumoto M, Sakabe T. The relationship between bispectral index and electroencephalographic parameters during isoflurane anaesthesia. *Anesth Analg.* (2004) 98:1336–40. doi: 10.1213/0000105587.17108.B6

33. Connor CW. A Forensic disassembly of the BIS monitor. *Anesth Analg.* (2020) 131:1923–33. doi: 10.1213/ANE.0000000000005220

34. Drover D, Ortega HRR. Patient state index. *Best Pract Res Clin Anaesthesiol.* (2006) 20:121–8. doi: 10.1016/j.bpa.2005.07.008

35. Viertiö-Oja H, Maja V, Särkelä M, Talja P, Tenkanen N, Tolvanen-Laakso H, et al. Description of the entropy algorithm as applied in the datex-ohmeda S/5 entropy module. *Acta Anaesthesiol Scand.* (2004) 48:154–61. doi: 10.1111/j.0001-5172.2004.00322.x

36. Hatib F, Jian Z, Buddi S, Lee C, Settels J, Sibert K, et al. Machine-learning algorithm to predict hypotension based on high-fidelity arterial pressure waveform analysis. *Anesthesiology.* (2018) 129:663–74. doi: 10.1097/ALN.0000000000002300

37. FDIC. Supervisory Guidance on Model Risk Management. (2017). Available online at: https://www.fdic.gov/news/financial-institution-letters/2017/fil17022a.pdf

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher’s Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Melvin, Broyles, Duggan, John, Smith and Berkowitz. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.