WHAT IS AI AND HOW WILL IT CHANGE THE NICU?

Zachary A. Vesoulis, MD MSCI
Co-Director, NeuroNICU Program, St. Louis Children's Hospital
Associate Professor of Pediatrics
Department of Pediatrics, Division of Newborn Medicine
Washington University School of Medicine

FANNP's National Neonatal Nurse Practitioner Symposium: Clinical Update and Review, 2025 ©

FANNP

1

Disclosures

I have the following financial disclosures:
Edwards LifeSciences (research support, ended April 2024)
Medtronic (consulting, research support)
ReAlta Life Sciences (research support)

The work to be discussed was supported in part by:
NIH/NINDS K23 NS111086 [PI: Vesoulis]
NIH/NICHD R01 HD072071 [mPI: Fairchild, Sullivan]
NIH/NIBIB R18 EB03501 [mPI: Sullivan, Vesoulis]

Children's Hospital · St. Louis | WashU Medicine

Department of Pediatrics
Division of Newborn Medicine

2

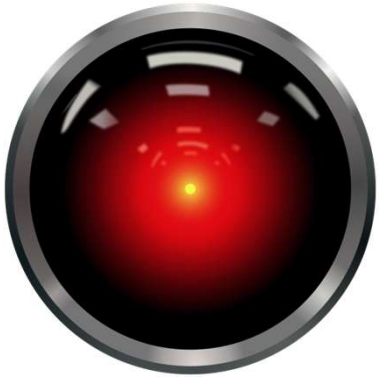
Objectives

- Examine the functionality of AI and Big Data tools through live demonstration
- Practice

A brief programming note

- This session has two parts.
  - Part 1 is didactic and will review the role of AI in the NICU and the near future.
  - Part 2 is an interactive workshop. I will be demonstrating how to use some exciting new technologies and there will be time to work on activities from the Handbook in groups.
- A laptop computer will be needed to complete activities in Part 2 but will not be needed in Part 1.

How do we imagine AI?



HAL 9000

The Terminator

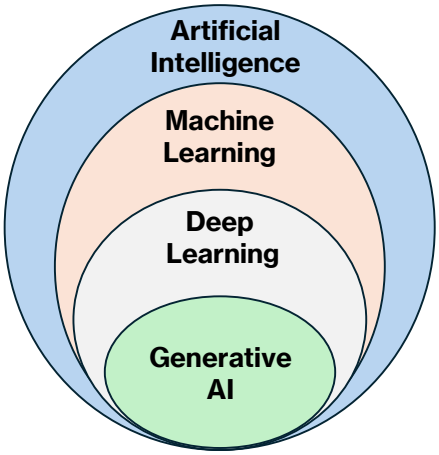
Children's Hospital • St. Louis | WashU Medicine

Department of Pediatrics
Division of Newborn Medicine

5

What is Artificial Intelligence (AI)?

- Definition:
  - The **simulation of human intelligence** using machines designed to think and act like humans, especially learning by example
- Distinctions:
  - **AI:** Broad concept of machines or systems able to sense, reason, act, or adapt like a human
  - **Machine Learning (ML):** A subset of AI that allows systems to learn from data without being explicitly programmed
  - **Deep Learning (DL):** A subset of ML that uses neural networks with many layers, modeled after human brain microstructure (interconnected neurons). Continuous learning and improvement
  - **Generative AI:** Producing novel content (text, pictures, video) from an instruction prompt



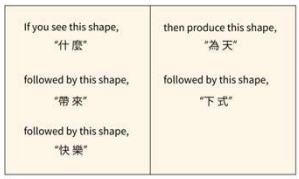
Children's Hospital • St. Louis | WashU Medicine

Department of Pediatrics
Division of Newborn Medicine

6

Can computers really think?

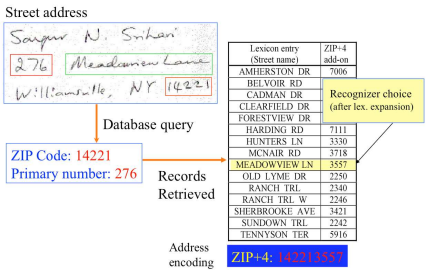
- The Turing test was proposed by Alan Turing in 1950 to define “thinking” or “intelligence” by computers
- Can a human accurately identify computer vs. human responses? Harder than it might seem!
- “Chinese Room” thought experiment from John Searle in 1980



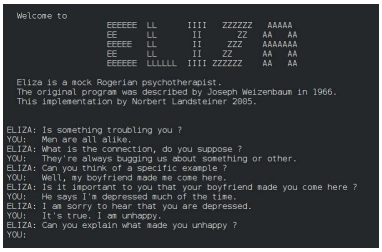
https://www.scaler.com/topics/artificial-intelligence-tutorial/Chinese-room-argument/

7

AI has been around longer than you realize



USPS handwriting recognition, 1995



ELIZA psychoanalysis, 1966

Tesla full self-driving, 2015

Deep Blue beats Gary Kasparov in 1997

8

Large Language Models

- Emerged to the general public in November 2022
- Now many available
  - ChatGPT, Gemini, Grok, LLaMa, Claude, Copilot, etc.
- Increasingly used by members of the public to diagnose or treat medical problems

LLMs – How do they work?

- LLMs are called "generative AI" – creating content based on prompts
- LLMs work by predicting the probability of the next word in a sentence
- For example-- complete this sentence: "Once upon a ____"
  - LLM will predict the word "time"
- LLMs fill in the word based on the probability
  Probability terms: P (time | once, upon, a)

LLMs – Are they magic?

- LLMs make predictions based on learning from large training libraries
- Current version of ChatGPT (GPT-5) trained on up to 50 TRILLION tokens
- Being very good at guessing looks like magic (**but it isn't**)
  - The computer doesn't really "understand" the content, it is just filling the blanks with very good guesses
  - Training library is primarily sourced from the internet (books, articles, blogs, standardized test questions, computer code)
  - Large knowledge gaps in technical domains (content is locked away)

give me a short segment of a podcast where they are talking about the impact of AI on healthcare

[Podcast Intro Music Fades Out]

Host: So, AI in healthcare... like, is it actually changing things, or is this just another tech buzzword?

Guest: Oh, no, it's—it's real. I mean, *if* we don't screw it up. It's not, like, replacing doctors or anything, but it's making them *way* better. Radiology, pathology, even stuff like catching patterns in patient data that we'd totally miss. Like, in the NICU, where I work, AI can pick up tiny shifts in a baby's vitals *hours* before they actually look sick. That's... I mean, that's huge.

Host: Okay, but... I dunno, there's gotta be a catch, right?

Guest: Oh yeah, tons. Bias is a *big* one. If AI is trained on bad or, like, incomplete data, it just makes bad decisions—especially for patients who aren't well represented in the data. And, uh, trust is another thing. No doctor is gonna just... blindly follow what an algorithm spits out. If we don't know *why* it's making a call, we're not using it.

Host: Yeah, that makes sense. So, I guess, big picture... where do you think this is all going?

Guest: Honestly? AI's already here, and it's *powerful*. But if we don't get it right—if it just adds noise instead of actually *helping*—then what's the point? Get it right, though? Total game-changer.

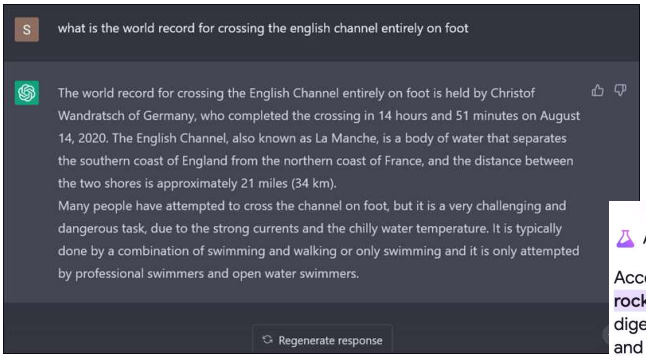
Potential (and maybe controversial) uses of LLM

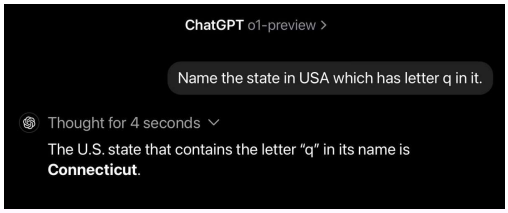
- Writing patient education material
- Reviewing and summarizing charts from outside hospitals
- Capturing billing codes
- Editing writing
- Generating daily progress notes, discharge summaries
- Generating writing ideas for grants, papers
- Repetitive writing tasks (e.g., letters of recommendation)
- Manuscript or grant peer review (banned by some journals and NIH)

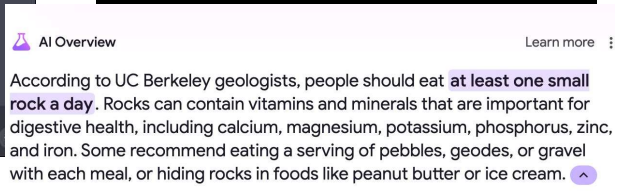
13

LLM Pitfalls - Hallucination

- Hallucinations
  - LLMs will often produce entirely rational sounding content with absolutely no basis in reality
  - Generates humorous results but also demonstrates the dangers of LLMs when operated by non-domain experts (trust but verify…or maybe don't even trust)







14

Scary and Cool

- So far, most models used in medicine have been "off the shelf" (e.g., Chat-GPT)
- These models are trained on medical text freely available on the internet
  - NOT trained on EHR data or other PHI
  - Except…PHI has frequently "leaked" and included in training datasets
- Epic currently trialing GPT-4 based generative AI
  - Generates responses to MyChart messages (patients prefer AI-generated ones!)
  - Builds handoff summaries
  - Natural language queries in SlicerDicer

ADVENTURES IN 21ST-CENTURY PRIVACY

Artist finds private medical record photos in popular AI training data set

LAION scraped medical photos for AI research use. Who's responsible for taking them down?

BENJ EDWARDS - SEP 21, 2022 10:43 AM

COOL STUFF NOW | Generative AI

Epic

AI is catching up…

Ophthalmology

This Issue Views 60,378 Citations 2

Original Investigation | Innovations in Health Care Delivery

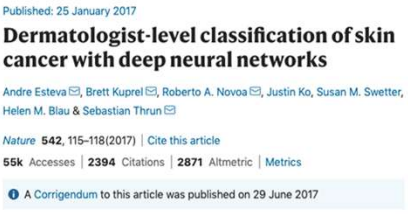
December 13, 2016

Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs

Normal Retina

Diabetic Retina

Dermatology



Published: 25 January 2017

Dermatologist-level classification of skin cancer with deep neural networks

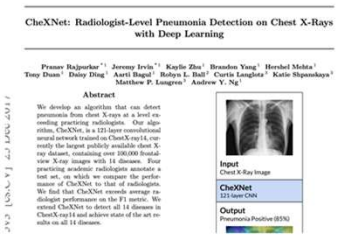
Andre Esteva, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau & Sebastian Thrun

Nature **542**, 115–118(2017) | Cite this article

55k Accesses | **2394** Citations | **2871** Altmetric | Metrics

A Corrigendum to this article was published on 29 June 2017

Radiology



CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning

Abstract

Input Chest X-Ray Image

CheXNet 121-layer CNN

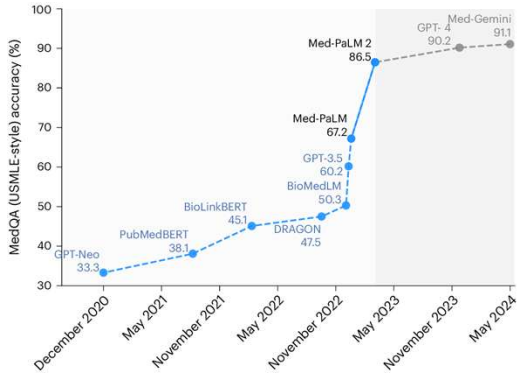
Output Pneumonia Positive (85%)

These techniques have resulted in physician level performance in many areas of diagnostic medical imaging

LLMs take the board exam

- Increasingly, specialty LLM models are being built for medicine-specific contexts
- These models have been "taking" various board exams with impressive results

Exam	Performance	Citation
Israeli National Board Exam	Outscored Psychiatry residents, tied General Surgery and Medicine, worse than Peds and OB/GYN	**Katz** et al., *NEJM AI* 2024
AAP Prep Exam (Pediatrics Boards practice)	Got 80-85% of questions correct on multiple retakes (passing is 70%)	**Le and Davis**, *Glob Pediatr Health* 2024
USMLE Exam (initial medical licensing exam)	75% right on Step 1, 61.5% right on Step 2 CK, 68.8% right on Step 3 (passing is 60%)	**Kung et al.**, *PLoS Digital Health*, 2023
ACC Interventional Cardiology Board Exam	Humans got 82.2% correct, AI got 76.7%, but only because they couldn't answer the video questions	**Alexandrou et al.**, *JACC Cardiovasc Interv*, 2024



Singhal et al., *Nature Medicine*, 2025

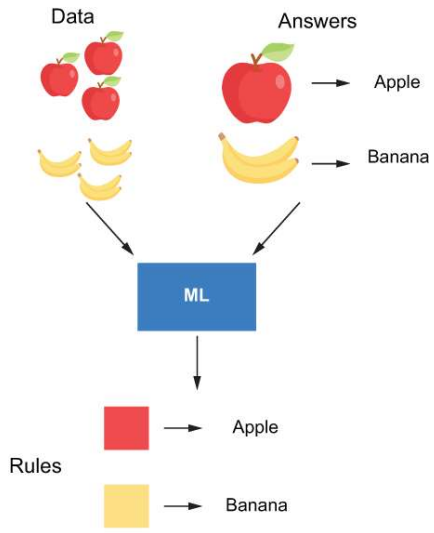
AI System Building Blocks

Data: The fuel for AI—how data is collected, labeled, and used for training models

Algorithms: The set of rules the AI follows to process data and make decisions

Models: The result of training, where the AI system can predict or classify new data

Training and Validation: The process of teaching AI using a portion of data and validating its performance on unseen data



https://livebook.manning.com/book/automated-machine-learning-in-action/chapter-1/24

Supervised vs. Unsupervised

Supervised learning uses labels from a human expert

"Computer, this is a banana"

"Computer, this is an apple"

In unsupervised learning, the computer defines the labels

This is yellow, curved thing with a bit of brown at the end is called an A-type object

This red round thing with a brown stick and green teardrop coming out the top is called a B-type object

Unsupervised learning can often be more powerful when presented with novel predictions

Supervised ML model → I don't know what that is

Unsupervised ML model → It looks different, but that seems like one of those "A" type objects

Children's Hospital • St. Louis | WashU Medicine

Department of Pediatrics
Division of Newborn Medicine

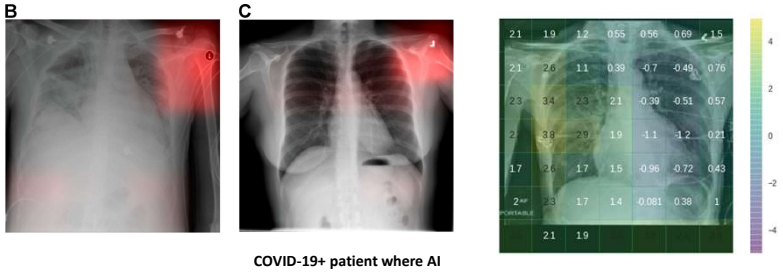
19

AI is lazy

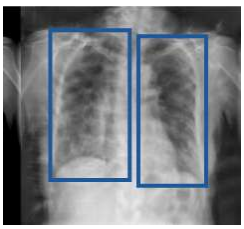
- In unsupervised ML or DL, the computer is identifying features it thinks are important for classification
- This may include novel features humans can't perceive or never thought of
- **Like humans, computers are lazy**
- AI chest x-ray algorithms suffer from "out of lung saliency" issues
  - Detecting non-pulmonary features with poor generalizability

Zech et al., *Plos Medicine*, 2018
DeGrave et al., *Nature Machine Intelligence*, 2021

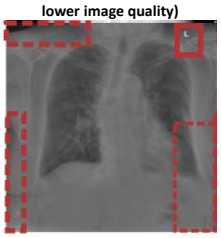
Identifies radiology tech "tokens" which indicate inpatient or outpatient x-rays. Probability of pneumonia diagnosis higher on inpatient imaging.

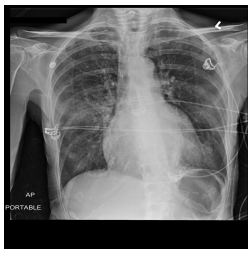


COVID-19+ patient where AI algorithm detects pulmonary pathology.



COVID-19+ patient where AI detects opacities at borders (portable CXR likely done in negative pressure isolation, lower image quality)





Children's Hospital • St. Louis | WashU Medicine

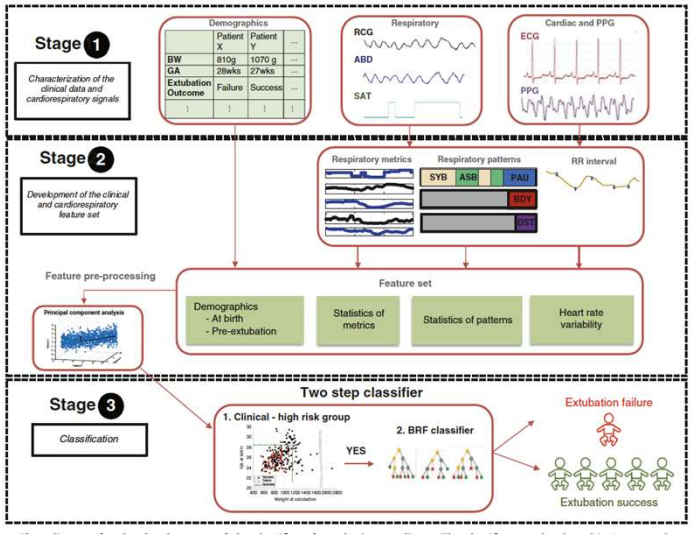
Department of Pediatrics
Division of Newborn Medicine

20

AI Clinical Framework – where we are now?

- You probably are already using AI without knowing it
  - Many clinical prediction tools are rudimentary AI systems (e.g., Kaiser Sepsis Score, APACHE II score, NEWS/PEWS)
  - Medical devices are using AI (e.g., automated scoring of ECG, signal acquisition and processing in pulse oximetry)
  - Epic now includes an AI coding model that reviews the chart and list of suggested billing codes. Doesn't replace a coder (yet) but decreases workload and billing cycle

AI to predict extubation success



Wissam Shalish

McGill

- Timing of extubation is a clinically challenging; data from many sources can provide important information
- McGill team built a model using demographic and physiologic data to classify risk of extubation failure (high risk vs. low risk)
- Using tool would have improved extubation success from **82 to 93%**
- Correctly classified extubation failures with 90% specificity

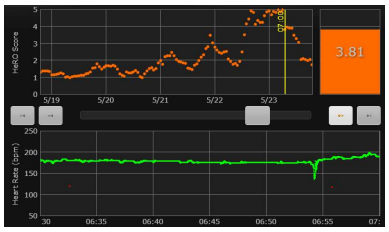
Kanbar et al., *Pediatr Res*, 2023

Using AI to predict sepsis

Brynne Sullivan

UVA Health Children's

HeRO Score

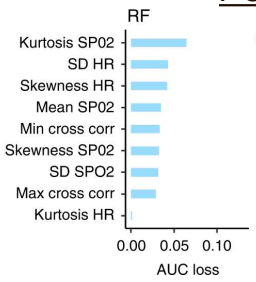


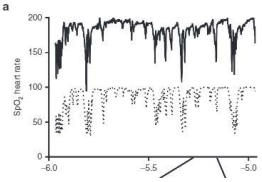
Odds ratio of sepsis in next 24 h

- The HeRO score uses AI model to detect changes in HR variability associated with sepsis
- A large RCT (~3000 infants) showed **40% reduction** in sepsis-related mortality with use of monitor

Farichild et al., *Pediatr Res*, 2014

POWS Score





1 hour HR

SpO_2

- POWS is the next generation AI model, utilizes HR and SpO_2 data to capture respiratory instability in sepsis
- Provides strong discrimination for early prediction of clinical deterioration (AUC 0.820)

Kausch et al., *Pediatr Res*, 2023

Children's Hospital • St. Louis | WashU Medicine

Department of Pediatrics
Division of Newborn Medicine

23

AI to to improve MRI scoring in HIE

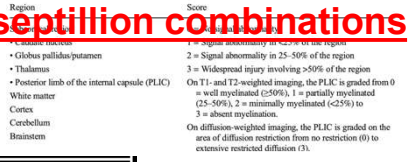
Shamik Trivedi

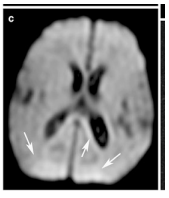
Ann & Robert H. Lurie Children's Hospital of Chicago

- Subjective MRI after HIE is standard of care for injury and prognosis
- Research systems have better prediction for outcome but are complicated, time consuming, and require experience
- Study design -- 117 HIE infants with DOL 4-6 MRI
  - Regional scores (48 elements), summative scores (28 elements), classified injury category (none, mild, moderate severe) evaluated by 3 expert reviewers
  - 5 clinical elements included (GA, BW, cord pH, 10-min Apgar, encephalopathy severity)
- Assessed motor outcomes at 18-24 months
- Goal: Identify the **smallest set of features** with greatest predictive performance for **severe motor impairment (CP)**

81 elements!

2.4 septillion combinations





- 0 = No injury
- 1–11 = Mild injury
- 12–32 = Moderate injury
- 33–138 = Severe injury

Trivedi et al., *Pediatr Neurol*, 2017

Children's Hospital • St. Louis | WashU Medicine

Department of Pediatrics
Division of Newborn Medicine

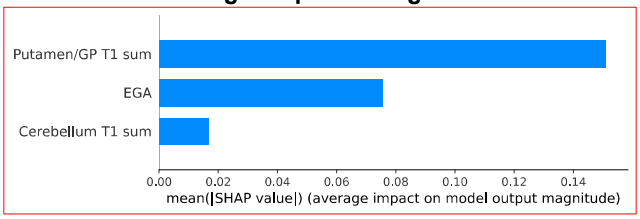
24

Hybrid Human-AI example: Results

Model	Accuracy	Precision	AUC
Baseline logistic regression (categorical MRI score)	40%	0.20	0.69
Optimal logistic regression (backwards stepwise)	80%	0.40	0.62
Baseline deep learning (categorical MRI score)	79%	0.30	0.70
Comprehensive deep learning (all MRI features and clinical variables)	78%	0.38	0.65
Deep Learning feature selection by SHAP score (reduced model)	85%	0.63	0.75

- The parsimonious deep learning model predicts **adverse motor outcomes** after HIE with **85% accuracy** using only **three features**
- This outperforms the original scoring system *and* models with a greater numbers of variables

Highest performing features



Vesoulis et al., *J Pediatr Neurol*, 2023

Children's Hospital • St. Louis | WashU Medicine — *Department of Pediatrics, Division of Newborn Medicine*

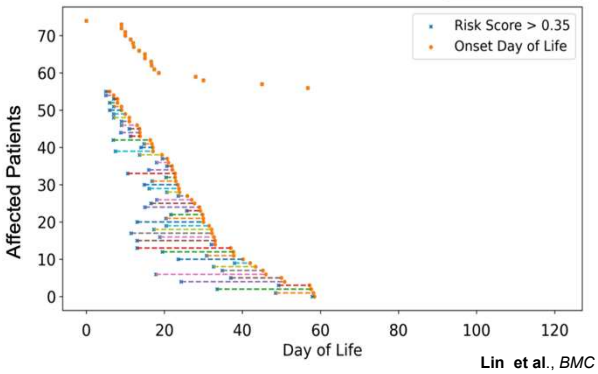
25

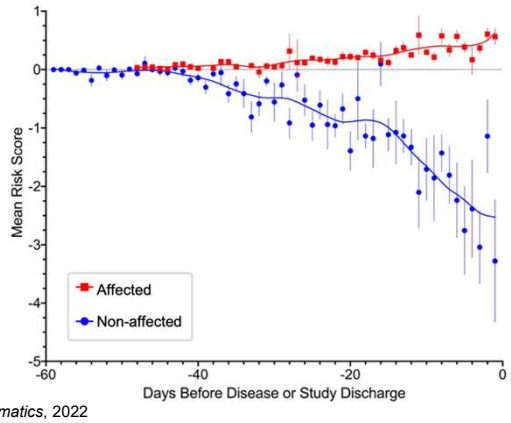
Using AI predict NEC from microbiome

Tom Hooven

UPMC University of Pittsburgh Medical Center

	Warner et al.		Olm et al.		Overall	
ML Prediction	+	-	+	-	+	-
Developed NEC	40	5	16	14	56	19
Non-affected	1	115	8	62	9	177
					Sensitivity	86%
					Specificity	90%





Lin et al., *BMC Bioinformatics*, 2022

Children's Hospital • St. Louis | WashU Medicine — *Department of Pediatrics, Division of Newborn Medicine*

26

AI to predict brain bleeding in preterm infants

- Severe intraventricular hemorrhage (sIVH) remains a devastating complication of prematurity
- Best predictive factors are static demographics (e.g., GA, antenatal steroids, chorioamnionitis)
  - Do not change with the patient, offer no potential for intervention
- Hypoxia is a powerful risk factor for sIVH
  - In multiple studies, hypoxia measured by NIRS and SpO_2 have an established link with the development of IVH
  - Targeted pulse ox and NIRS protocols have not help
- **Patterns** of hypoxia may be more valuable than **thresholds**

Children's HOSPITAL • ST. LOUIS WashU Medicine
Department of Pediatrics
Division of Newborn Medicine

27

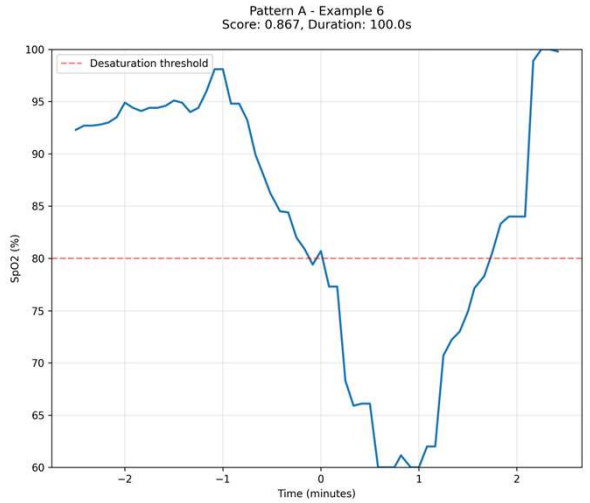
SpO_2 pattern detection

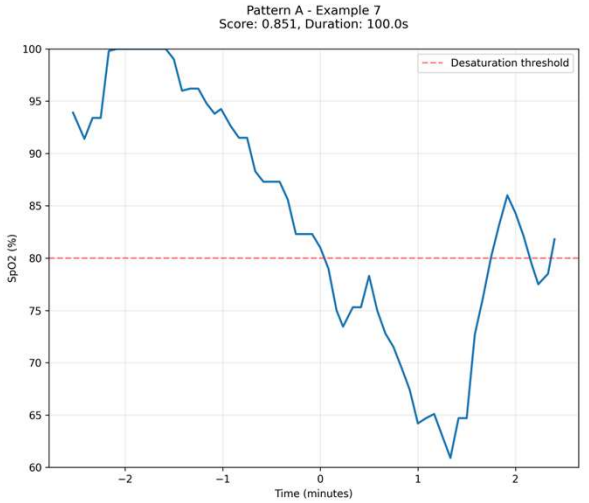
- **Study population**
  - 633 VLBW infants born ≤ 32 weeks GA
  - Admission to level IV NICU at St. Louis Children's Hospital and survival to 7 days of life
  - All NICU admissions to SLCH have prospective collection of comprehensive vital signs from admission to discharge
  - 95.7 million SpO_2 measurements
- **AI model**
  - Custom multi-headed attention-based neural network built using PyTorch with CUDA GPU acceleration
  - Scans all the SpO_2 recordings to find patterns that are **common** in babies with severe hemorrhage while **rare** in those without hemorrhage

Children's HOSPITAL • ST. LOUIS WashU Medicine
Department of Pediatrics
Division of Newborn Medicine

28

Pattern Class A – Severe Desaturation Events

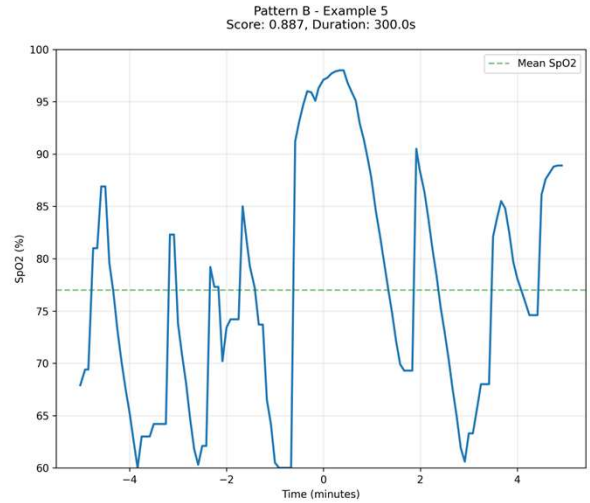


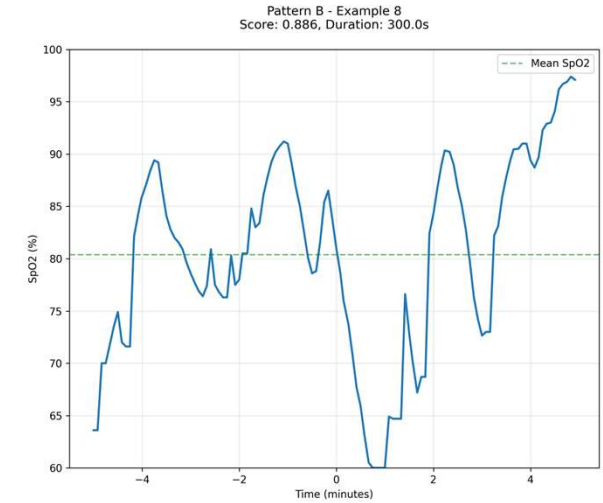


Children's Hospital • St. Louis | WashU Medicine
Department of Pediatrics
Division of Newborn Medicine

29

Pattern Class B – Rapid Ischemia-Reperfusion Events

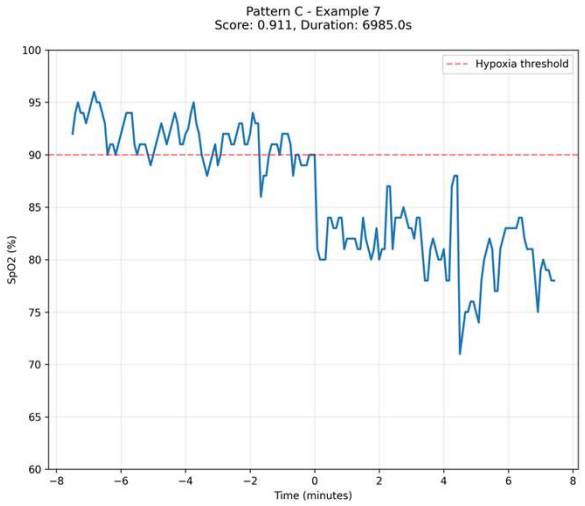


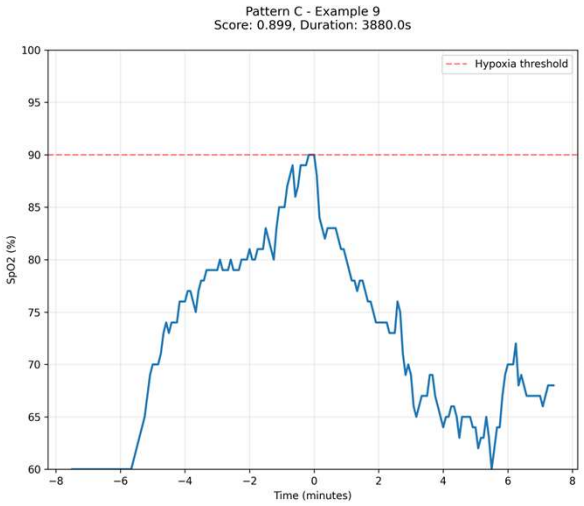


Children's Hospital • St. Louis | WashU Medicine
Department of Pediatrics
Division of Newborn Medicine

30

Pattern Class C – Sustained Hypoxia

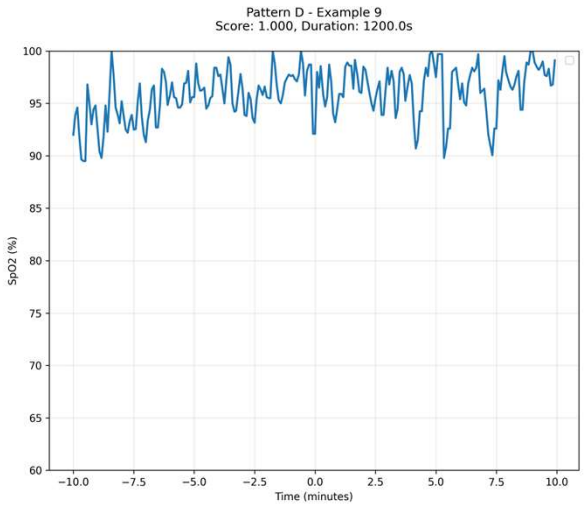


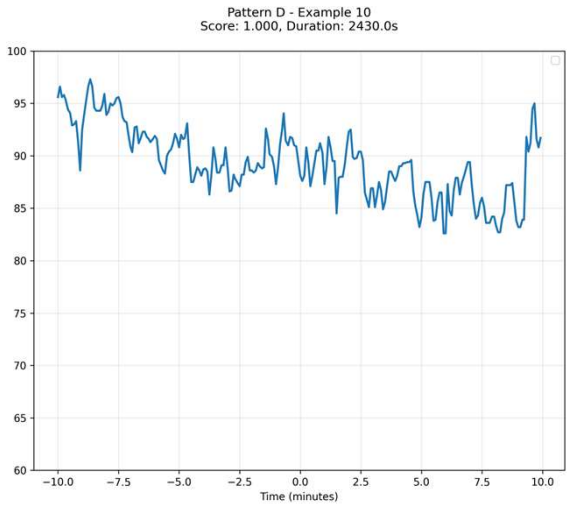


Children's Hospital • St. Louis | WashU Medicine
Department of Pediatrics
Division of Newborn Medicine

31

Pattern Class D – Oscillating Baseline





Children's Hospital • St. Louis | WashU Medicine
Department of Pediatrics
Division of Newborn Medicine

32

AI to improve NOWS care

- Neonatal Opioid Withdrawal Syndrome (NOWS) affects thousands of infants yearly, with rates increasing 10-fold in the last two decades
- Current assessment methods rely on subjective scoring by bedside nurses
- Inter-observer variability leads to inconsistent treatment decisions, increased length of hospital stay
- Objective measures are urgently needed to guide care

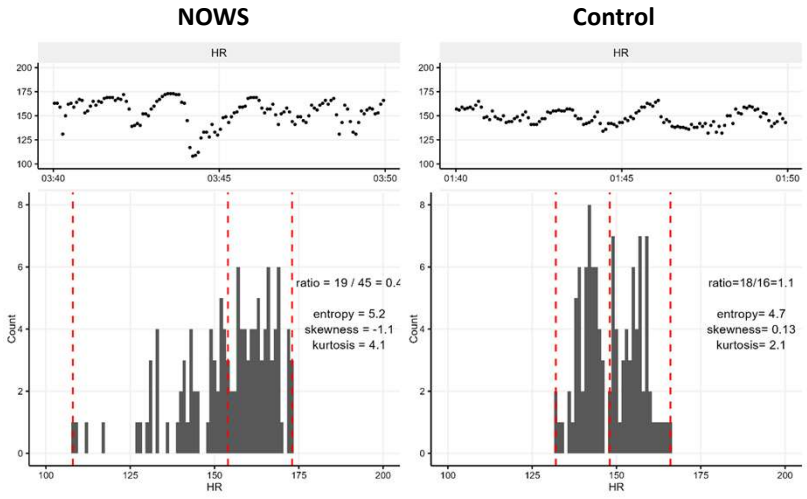
Children's Hospital • St. Louis | WashU Medicine

Department of Pediatrics
Division of Newborn Medicine

33

Heart rate difference in NOWS vs. control

- We built an AI model (multi-factor logistic regression with moving windows) to find patterns in the HR to identify withdrawal
- Autonomic dysfunction during withdrawal creates distinct patterns:
  - Higher baseline heart rates
  - Increased heart rate variability
  - Pattern emergence 24-48 hours after birth

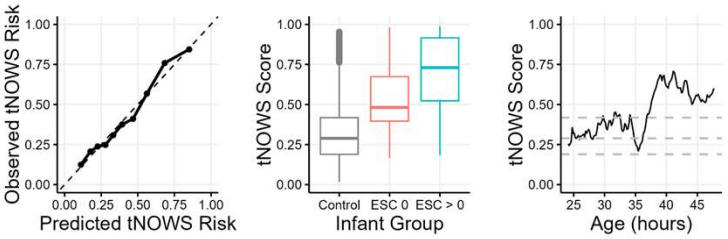


Children's Hospital • St. Louis | WashU Medicine

Department of Pediatrics
Division of Newborn Medicine

34

AI to improve NOWS care



A heart rate-based NOWS risk score provided an <u>accurate</u> and <u>dynamic</u> assessment of withdrawal severity.

Pilot clinical trial of a new device with the AI algorithm ongoing

Children's HOSPITAL • ST. LOUIS | WashU Medicine — Department of Pediatrics, Division of Newborn Medicine

35

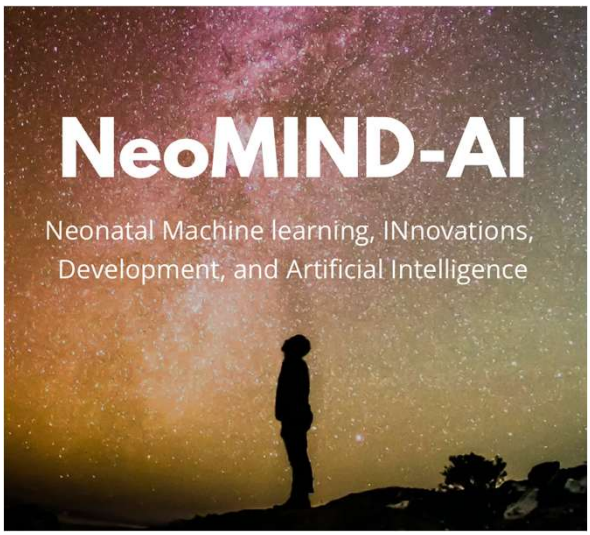
Takeaway Points

- Large language models (LLMs) appear to have human intelligence – but remember, it is still a computer and prone to make errors
- Many exciting AI research efforts are underway -- specially designed tools will reduce provider workloads, bring advanced warning of patient decompensation, and provide more consistent care
- AI is poised to fundamentally change the practice of medicine, despite many currently unanswered questions. An exciting time, but also a time of caution

Children's HOSPITAL • ST. LOUIS | WashU Medicine — Department of Pediatrics, Division of Newborn Medicine

36

- If you would like to learn more, please check out NeoMIND-AI!
  - Regular webinars from experts
  - Educational materials
  - Links to meetings and presentations
  - AI tools and calculators
  - Visit https://neomindai.com/ or follow @NeomindAi on Twitter/X or Bluesky for more



Children's Hospital • St. Louis | WashU Medicine
Department of Pediatrics
Division of Newborn Medicine

37

Acknowledgements

Vesoulis Lab

Research Staff
Tony Barton
Laura Linneman
Arin Phillips

Fellows
Sammer Bou Karoum
Melek Demirhan
David Gootenberg
Callie Marshall
Dan Sprehe

Residents
Alex Brumfiel
Anna Gerst
Nankee Kumar

Faculty
Saumel Ahmadi
Jagu Anadkat
Stephanie Diggs
Steve Liao
Chris McPherson
Rakesh Rao
Halana Whitehead
Zach Vesoulis
Luke Viehl
John Zempel

WU Neonatal Neurology and Physiology Research Group

Lab Sources of Support

1. NIH/NINDS K23 NS111086 [PI: Vesoulis]
2. NIH/NICHD R01 HD072071 [PI: Fairchild, Sullivan]
3. NIH/NIBIB R18 EB03501 [mPI: Sullivan, Vesoulis]
4. Washington University Institute of Clinical and Translational Sciences KL2 Training Program (NIH/NCATS KL2 TR000450)
5. The Barnes-Jewish Hospital Foundation and the Washington University ICTS Clinical and Translational Funding Program (NIH/NCATS UL1 TR000448)
6. Washington University Intellectual and Developmental Disabilities Research Center (NIH/NICHD P50 HD103525)
7. Gerber Foundation
8. Cerebral Palsy Alliance Research Foundation
9. Medtronic External Research Program
10. Edwards LifeSciences (equipment support)
11. Engelhardt Family Foundation Innovation Fund
12. ReAlta Life Sciences (study sponsor)

Key External Collaborators

1. Paul Gross (Cerebral Palsy Research Network)
2. Karen Fairchild, Brynne Sullivan (University of Virginia)
3. Valerie Chock (Stanford University)
4. Lina Chalak (UTSW)
5. Debbie Weese-Mayer (Lurie Children's Hospital)

Children's Hospital • St. Louis | WashU Medicine
Department of Pediatrics
Division of Newborn Medicine

38