[18-185] How AIs “Read” Sonograms: Mechanisms of Machine Interpretation in Fetal Imaging

Christian Macedonia, M.D.

University of Michigan

USA

CEO Lancaster Maternal Fetal Medicine LLC

https://thektproject.org/

IFMSS 2026
KYOTO, JAPAN

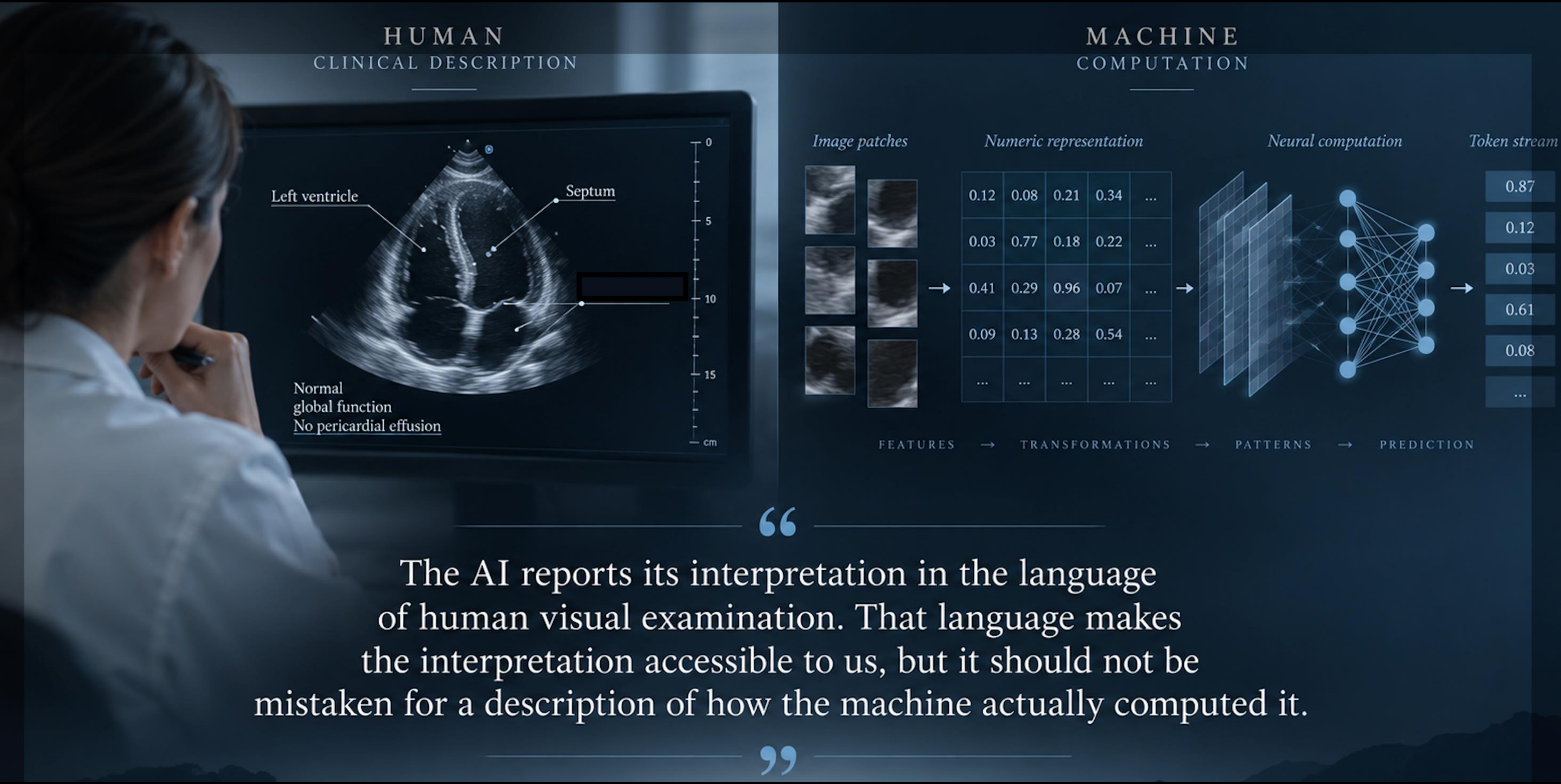
How do AIs "read" sonograms?

And can we <u>trust them</u>?

A Glimpse of the Problem

Fact:

- Up to 70% of CHDs are undetected on routine prenatal ultrasound
- ... and roughly ***<u>half of all fetal anomalies</u>*** missed!
- *...and we cannot say why, for any individual case.*
- *...but the question remains; can we trust AI to fill the gaps? Can they be relied upon to help us read images properly? To diagnose? To navigate?*

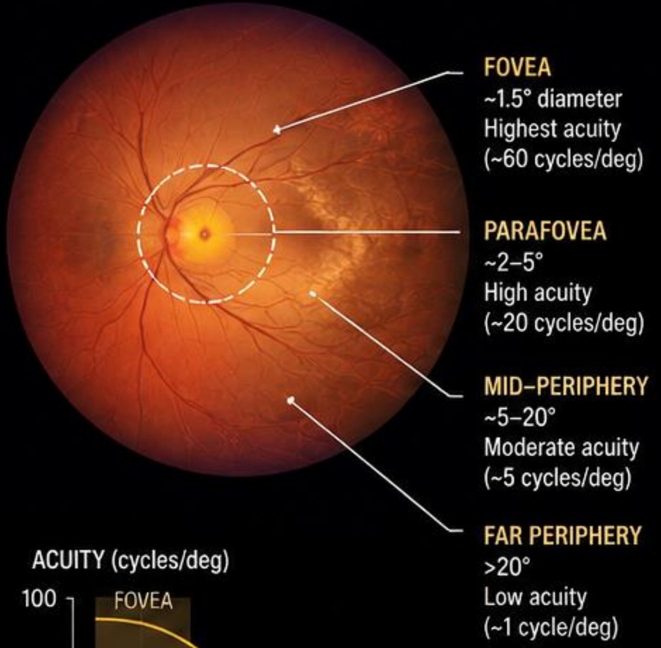

HUMAN
CLINICAL DESCRIPTION
Left ventricle
Septum
[redacted]
0
5
10
15
cm
Normal
global function
No pericardial effusion
MACHINE
COMPUTATION
Image patches
Numeric representation
Neural computation
Token stream
0.12 0.08 0.21 0.34 ...
0.03 0.77 0.18 0.22 ...
0.41 0.29 0.96 0.07 ...
0.09 0.13 0.28 0.54 ...
... ...
0.87
0.12
0.03
0.61
0.08
...
FEATURES → TRANSFORMATIONS → PATTERNS → PREDICTION
“The AI reports its interpretation in the language
of human visual examination. That language makes
the interpretation accessible to us, but it should not be
mistaken for a description of how the machine actually computed it.”

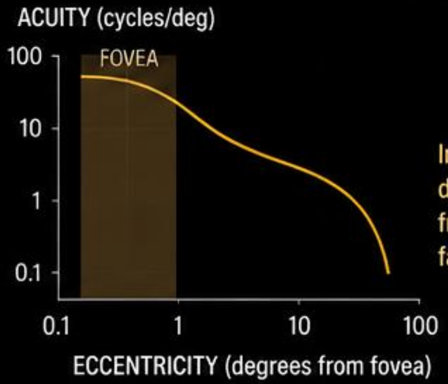
HUMAN VISION IS LOCAL. MACHINE VISION IS UNIFORM.

Ultrasound should match the observer, not the machine.

1. HUMAN RETINA: FOVEAL ACUITY FALLOFF

We see the world with a high-resolution center and rapidly decreasing resolution in the periphery.

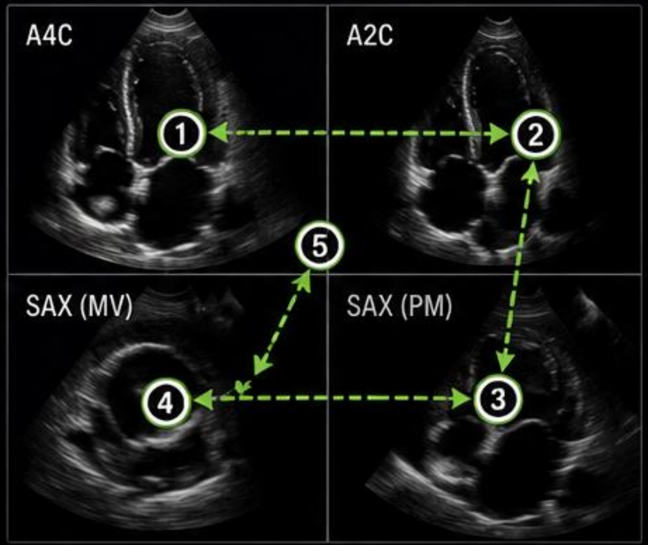




Information density drops ~100–1000× from fovea to far periphery.

2. HUMAN OBSERVATION: SCANPATH ON FOUR-CHAMBER VIEW

Humans sample selectively. We fixate, process, then move our gaze. The periphery guides, the fovea decides.



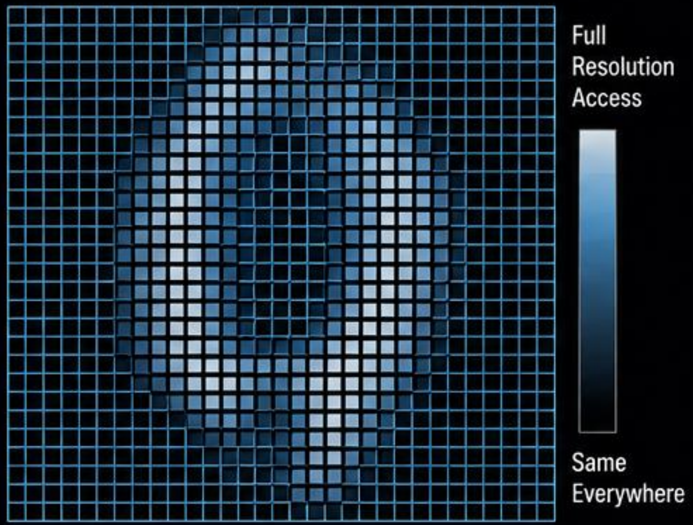
1. Global assessment (A4C)
2. Volume/Function (A2C)
3. Ventricular geometry (SAX PM)
4. Valve assessment (SAX MV)
5. Integration / Decision

- ~200–300 ms per fixation
- ~3–6 fixations per view
- Peripheral preview
- Foveal decision
- Sparse sampling, not uniform reading

HUMAN VISION IS ADAPTIVE AND PURPOSE-DRIVEN. WE DON'T READ EVERY PIXEL.

3. MACHINE ACCESS: UNIFORM AND COMPLETE

Machines have no fovea. Every pixel is available simultaneously at full resolution.



No attention bottleneck.
No acuity falloff.
No information is inherently discarded.

MACHINES CAN SEE EVERYTHING. LET AI USE THE WHOLE FIELD.

KEY IMPLICATION

Traditional display and reading paradigms match human limitations, not machine potential. To unlock AI performance, ultrasound systems must deliver uniform, high-fidelity data and let intelligence decide where to look.

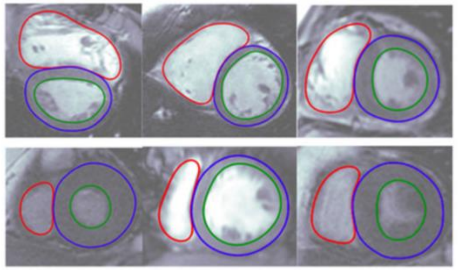
THE GOAL IS NOT PRETTIER IMAGES. THE GOAL IS TO LET INTELLIGENCE SEE THE SIGNAL BEFORE WE THROW IT AWAY.

A Condensed History of AI as Image Reader

- 1960s–1980s — Computer-aided image analysis begins: early rule-based systems quantify edges, shapes, densities, and lesions in X-ray, CT, and NUCMED.
- 1990s — Classical CAD enters clinical imaging: handcrafted features and statistical classifiers are applied to mammography, chest radiography, and CT; AI remains largely a second-reader aid.
- 2000s — Machine learning broadens: support-vector machines, random forests, atlas methods, and early neural networks improve segmentation, detection, and organ classification.
- **2012 — Deep learning inflection point: convolutional neural networks transform computer vision, and medical imaging rapidly adopts end-to-end learned feature extraction. ImageNet 2012: AlexNet Won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) by a massive margin, slashing the top-5 error rate to 15.3%.**
- 2015–2018 — Deep learning reaches radiologist-level tasks: CNNs become highly effective for diabetic retinopathy, CXR, CT, MRI, pathology, and dermatologic image classification; U-Net and related architectures make segmentation practical at scale. Medical-imaging deep-learning publications accelerate sharply after 2017. “Attention is All You Need” is published.

1998- R2 Technology’s ImageChecker M1000 system

2012- AlexNet Creators:
Sutskever, Krizhevsky, Hinton
The power of Deep Learning



2017- Arterys Cardio DL™

Generally regarded as the most important scientific paper of the 21st century.

Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com

Noam Shazeer* Google Brain noam@google.com

Niki Parmar* Google Research nikip@google.com

Jakob Uszkoreit* Google Research usz@google.com

Llion Jones* Google Research llion@google.com

Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu

Łukasz Kaiser* Google Brain lukaszkaiser@google.com

Illia Polosukhin* ‡ illia.polosukhin@gmail.com

arXiv:1706.03762v7 [cs.CL] 2 Aug 2023

Abstract

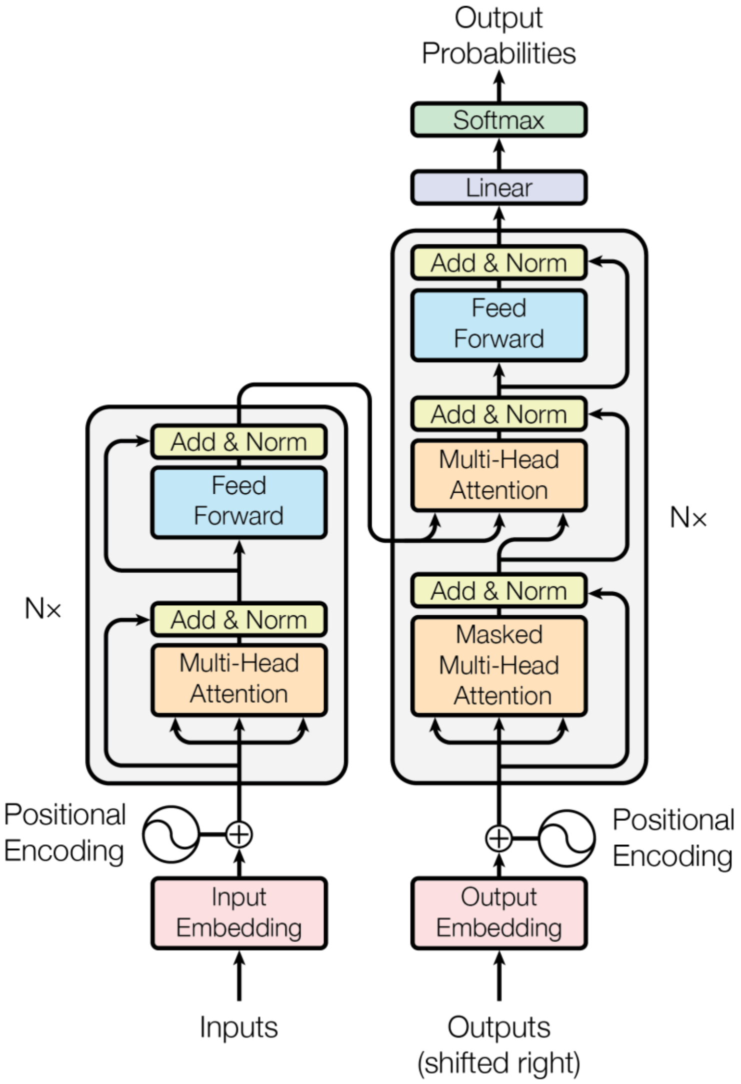
The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.

*Equal contribution. Listing order is random. Jakob proposed replacing RNNs with self-attention and started the effort to evaluate this idea. Ashish, with Illia, designed and implemented the first Transformer models and has been crucially involved in every aspect of this work. Noam proposed scaled dot-product attention, multi-head attention and the parameter-free position representation and became the other person involved in nearly every detail. Niki designed, implemented, tuned and evaluated countless model variants in our original codebase and tensor2tensor. Llion also experimented with novel model variants, was responsible for our initial codebase, and efficient inference and visualizations. Lukasz and Aidan spent countless long days designing various parts of and implementing tensor2tensor, replacing our earlier codebase, greatly improving results and massively accelerating our research.

†Work performed while at Google Brain.

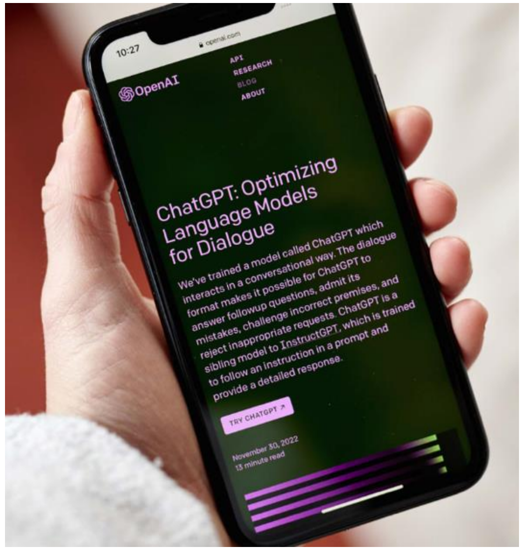
‡Work performed while at Google Research.

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.



A Short Hx of AI as Image Reader…cont.

- 2020–2023 — Ultrasound catches up: AI begins automating view classification, measurements, quality control, fetal biometry, echocardiographic analysis, and acquisition guidance.
- 2022 — **GPT-3.5** introduced the concept of utilizing **Large Language Models (LLMs).** Vision Transformers (ViTs) are now challenging CNNs in medicine.
- 2024 — Sonio commercializes prenatal-ultrasound AI in the U.S.: Sonio launches commercially at SMFM 2024, and Sonio Detect receives FDA clearance for automatic fetal-view recognition, anatomy detection, and exam-quality verification.
- 2025 — AI moves from ultrasound workflow assistance toward diagnostic interpretation: Sonio Suspect receives FDA clearance for AI-driven fetal-anomaly detection, while BrightHeart expands FDA-cleared fetal-heart software for standard-view recognition and congenital-heart-disease assessment.
- 2025–2026 — BrightHeart and Sonio establish commercial fetal-ultrasound AI platforms: BrightHeart's FDA-cleared fetal-heart tools and Sonio's Detect/Suspect platform mark the transition from "AI helps acquire the image" toward AI actively reading and interpreting prenatal ultrasound examinations. BrightHeart uses Meta's open DINOv2 (ViT backbone). Sonio uses a proprietary CNN …SSD (Single Shot MultiBox Detector).



GPT 3.5 November 30, 2022

What matters to doctors	Convolutional Neural Networks — [CNNs]	Vision Transformers — [ViTs]
How they analyze an image	Build from local patterns—edges, textures, and shapes—combining them through successive layers.	Represent image regions as numerical "tokens" and use attention to learn relationships among them.
Principal strength	**Efficient extraction of visual detail.** Strong for tasks such as identifying standard views, outlining structures, and locating measurement targets.	**Flexible integration of context.** Directly connects information from separated image regions, helping model relationships among structures.
Principal limitation	Relationships between distant regions are built more indirectly through the network. Broader context may require additional architectural components.	High-resolution images and long sequences can demand substantial memory and computation. Fine detail still requires adequate resolution and an appropriate design.
Training requirements	Built-in assumptions about image structure often help when training with limited data. Can also start from pretrained models.	Particularly powerful when adapting a broadly pretrained model. **The local clinic does not need to repeat the original large-scale training.**
Role in a broader imaging assistant	Can supply visual features to systems that also process video, clinical information, and language.	Token-based representations fit naturally into systems combining images, multiple views, video, and language. **It is multimodal by nature.**
Running beside the scanner	Already attractive for fast, low-power analysis. Modern chips make them faster still.	Optimized and compact models **increasingly run locally** on Jetson and Apple silicon, reducing dependence on cloud processing.
Likely direction — a forecast	Remain valuable as efficient specialists and components within larger systems.	Gain a larger role in broadly pretrained, multimodal imaging systems—often **combined with CNNs rather than replacing every convolution**.

Feature	CNN Deployment	ViT Deployment
NVIDIA Jetson	Native out-of-the-box; extremely fast on GPU & DLA.	Requires TensorRT compilation for viable edge speeds.
Apple Silicon	Native via PyTorch `mps` or Core ML on the ANE.	Best implemented using Apple ML's optimized transformer guide or hybrid models.

Other things that “AI Vision” Enables

Sex differences in prenatal oral-motor function and development

J L Miller* PhD, Department of Health and Human Services/National Institutes of Health, National Institutes of Child Health and Human Development-Clinical Center/Rehabilitation Medicine Department-Physical Disabilities Branch;**
C Macedonia LTCOL MC, Women’s and Children’s Health, National Naval Medical Center, Bethesda;
B C Sonies PhD, Department of Health and Human Services/National Institutes of Health, National Institutes of Child Health and Human Development-Clinical Center/Rehabilitation Medicine Department-Physical Disabilities Branch,** Bethesda, Maryland, USA.

**Correspondence to first author at* National Institutes of Health, Rehabilitation Medicine Department, Building 10

An important factor in assessing the development of young children is the delineation of sex differences in patterns of physical growth and function. These differences occur very early in human development, most likely before birth, where prenatal patterns of morphologic development form antecedents of subsequent postnatal behaviors.[1] The identification of potential sex-related dimorphisms has been of particular interest in the study of prenatal motor development, primarily in the emergence of general body or leg movements of male and female fetuses.[2,3] Progressive, simple to complex, differentiated patterns of movement represent a developmental trajectory of neurobehavioral maturation and serve as markers of fetal well-being;[4] however, evidence to support sex-specific patterns of emergent motor activity has been controversial. Early investigations reported no sex differences in leg movements throughout various stages of maturation[5,6] while other studies reported data favoring males in the frequency of late gestation leg movements.[7,8]

An awareness of sex variations in motor activity is particularly important in evaluating behaviors that need to reach maturation by birth, such as the oral, pharyngeal, and laryngeal movements required for neonate feeding, respiration, vocalization, and ingestion.[9] Few data exist as to whether developmental differences between males and females occur in these regions, morphologically and physiologically. While male fetuses are biometrically more advanced in general body weight,

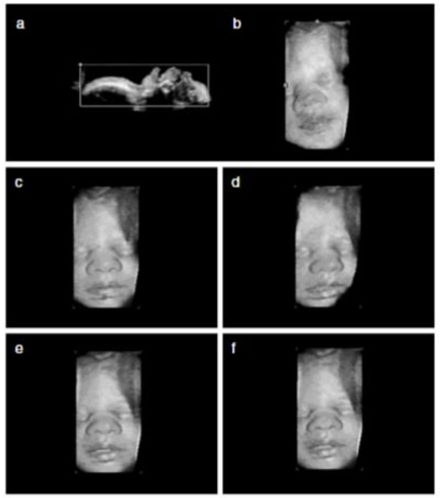


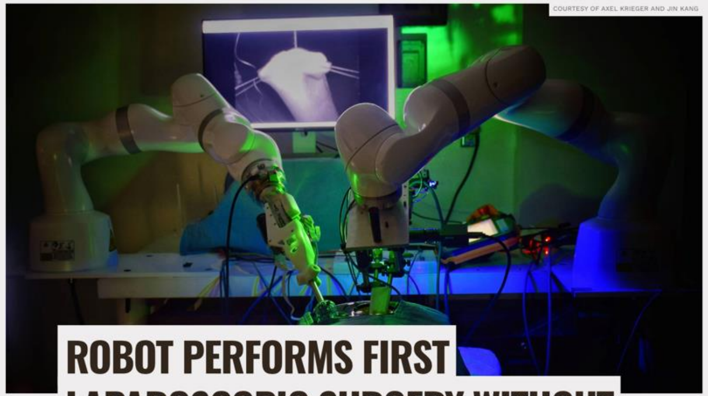
Figure 1: *Fetal ultrasound images to observe motor activity: (a) Two-dimensional sagittal plane view of fetal face, head, and body; (b) same fetal scan rendered in a three-dimensional (3D) frontal view; (c and d) sequence of movements of oral mouthing; and (e and f) anterior lingual thrust patterns derived from 3D scan.*

Table III: Percentage of observed fetal movements across sexes and trimesters

Trimester	*Percentage observed*										
	Lingual			*Oral*				*Pharyngeal*		*Laryngeal*	
	Thrust	*Cup*	*Suck*	*Jaw*	*Mouth*	*Lip*	*Yawn*	*Periodic*	*Complete*	*Periodic*	*Complete*
2nd											
Male	100	50	63	71	100	0	42	66	66	17	40
Female	100	88	87	78	88	71	88	85	100	44	53
3rd											
Male	100	90	76	100	100	66	66	83	66	55	55
Female	100	88	77	100	100	50	50	80	100	44	45

Thrust, anterior thrust of tongue to lower labial margin; Cup, tongue cupping (elevation of lateral lingual margins and midline depression along dorsum of tongue); Suck, sucking; Jaw, jaw excursion; Mouth, small mouthing movements of lips with mouth slightly open; Lip, lip pursing; Yawn, yawning; Periodic, periodic contractions; Complete, complete contractions.

468 *Developmental Medicine & Child Neurology* 2006, 48: 465–470



ROBOT PERFORMS FIRST LAPAROSCOPIC SURGERY WITHOUT HUMAN HELP

In four experiments on pig tissues, the robot excelled at suturing two ends of intestine—one of the most intricate and delicate tasks in abdominal surgery

Catherine Graham / Jan 26, 2022

Automating Long-Horizon Surgical Procedures

(Gallbladder Removal Surgery, **Fully Autonomous**)

8x speed

1) Detach gallbladder 2) Clip/cut ducts 3) Remove

https://youtu.be/1wCwl_E7ZgM?si=OPUBOURg0LYTnPm5

Autonomous Robots in Surgery: How Does AI Fit In?

FOCAL, Workshop on AI
Washington, DC, October 9th, 2025

Axel Krieger, PhD
Associate Professor and Carol Croft Linde Faculty Scholar
Intelligent MEdical Robotic Systems & Equipment (IMERSE) Lab
Department of Mechanical Engineering
Laboratory for Computational Sensing and Robotics (LCSR)
Malone Center for Engineering in Healthcare
Data Science and AI Institute
Johns Hopkins University, 116 Hackerman Hall, Baltimore, MD

axel@jhu.edu

Now, the Future of AI is at the *Edge*

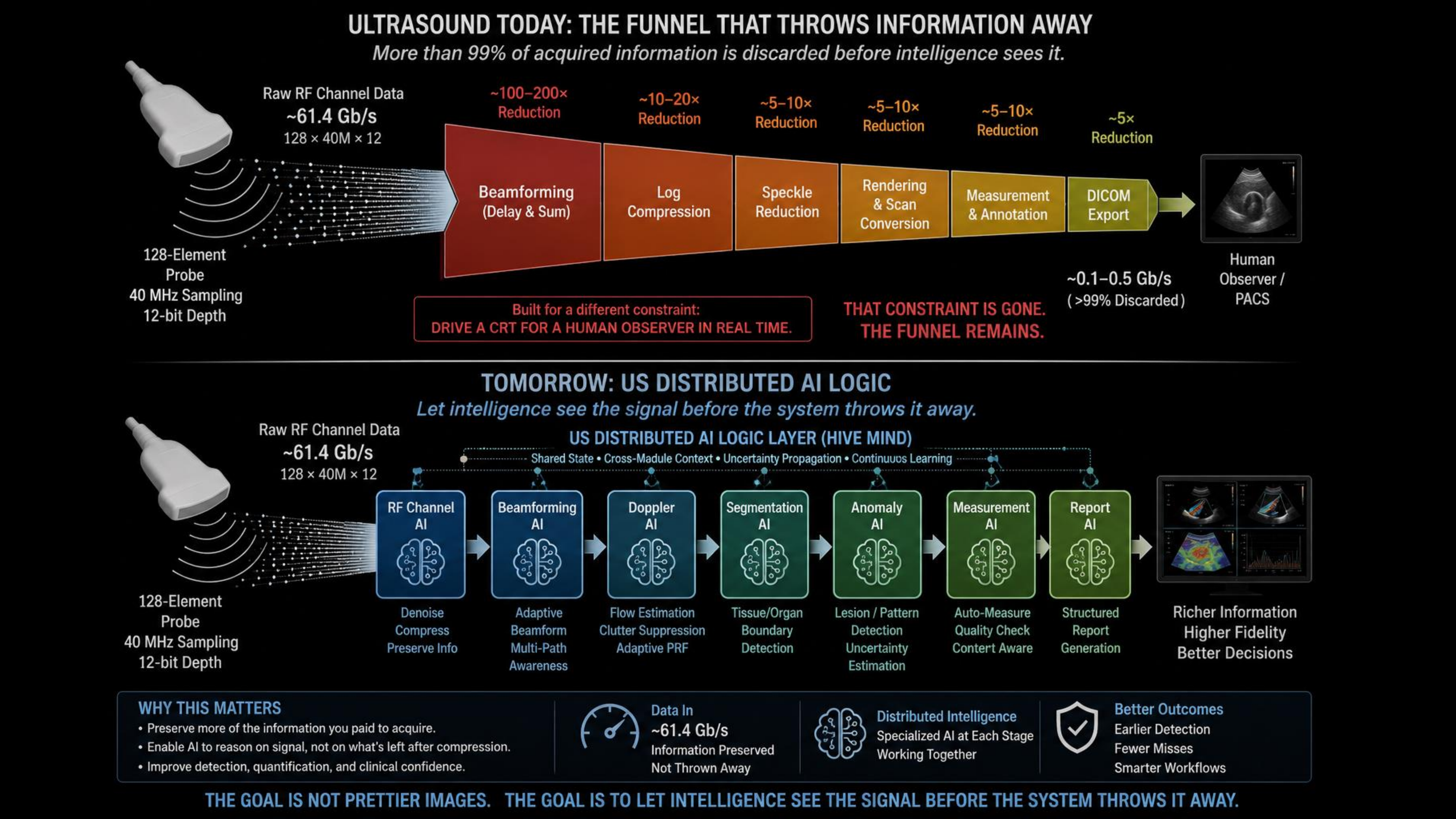
«Это система-на-кристалле (SOC), которая была компонентом российского ударного беспилотника, применённого против Украины. Мы должны использовать искусственный интеллект только в мирных целях. Война — это глупый и совершенно постыдный способ поведения для любого разумного вида. Искусственный интеллект должен служить мирным целям: помогать лечить больных и безопасно приводить детей в этот мир.»

Hugging Face

Qwen

MISTRAL AI_

deepseek Math-V2



ULTRASOUND TODAY: THE FUNNEL THAT THROWS INFORMATION AWAY
More than 99% of acquired information is discarded before intelligence sees it.
Raw RF Channel Data
~61.4 Gb/s
128 × 40M × 12
128-Element Probe
40 MHz Sampling
12-bit Depth
~100–200× Reduction
Beamforming (Delay & Sum)
~10–20× Reduction
Log Compression
~5–10× Reduction
Speckle Reduction
~5–10× Reduction
Rendering & Scan Conversion
~5–10× Reduction
Measurement & Annotation
~5× Reduction
DICOM Export
Human Observer / PACS
~0.1–0.5 Gb/s
(>99% Discarded)
Built for a different constraint:
DRIVE A CRT FOR A HUMAN OBSERVER IN REAL TIME.
THAT CONSTRAINT IS GONE.
THE FUNNEL REMAINS.
TOMORROW: US DISTRIBUTED AI LOGIC
Let intelligence see the signal before the system throws it away.
Raw RF Channel Data
~61.4 Gb/s
128 × 40M × 12
128-Element Probe
40 MHz Sampling
12-bit Depth
US DISTRIBUTED AI LOGIC LAYER (HIVE MIND)
Shared State • Cross-Madule Context • Uncertainty Propagation • Continuuos Learning
RF Channel AI
Denoise
Compress
Preserve Info
Beamforming AI
Adaptive
Beamform
Multi-Path
Awareness
Doppler AI
Flow Estimation
Clutter Suppression
Adaptive PRF
Segmentation AI
Tissue/Organ
Boundary
Detection
Anomaly AI
Lesion / Pattern
Detection
Uncertainty
Estimation
Measurement AI
Auto-Measure
Quality Check
Contert Aware
Report AI
Structured
Report
Generation
Richer Information
Higher Fidelity
Better Decisions
WHY THIS MATTERS
• Preserve more of the information you paid to acquire.
• Enable AI to reason on signal, not on what's left after compression.
• Improve detection, quantification, and clinical confidence.
Data In
~61.4 Gb/s
Information Preserved
Not Thrown Away
Distributed Intelligence
Specialized AI at Each Stage
Working Together
Better Outcomes
Earlier Detection
Fewer Misses
Smarter Workflows
THE GOAL IS NOT PRETTIER IMAGES. THE GOAL IS TO LET INTELLIGENCE SEE THE SIGNAL BEFORE THE SYSTEM THROWS IT AWAY.

Pro tip from a former DARPA Program Manager

https://thektproject.org/

When solving any problem with technology, consider using this three-step framework:

1. Define the problem in as much detail as possible and, in doing so, challenge everything- every assumption. Cartesian Doubt
2. Then question every assumption about why you do things using your new approach, and ruthlessly ask if your technology actually addresses the real pain points.
3. Put yourself on the receiving end. This applies to most tech, but particularly medicine. Imagine the person on the table is someone you love. Will your solution make their life better, or worse? What is the human benefit?

The Skeptics Rule, The Realist Rule…The Golden Rule