Discussion 5

Ethical considerations of models once deployed

Recap of Case Study

- Your AI model has been deployed in the local hospital's emergency department
- Model automates the ordering of tests (this speeds up – and standardizes) how we deliver care
- Model was validated with a **92% accuracy**
- ***Each group has a different model architecture, with different development and validation processes***

Deployment Considerations

- **Data Drift**: Gradual change in data distribution over time in deployed settings
- **Model errors** – how are these identified and reported? Who is liable

Discussion 5 – Questions: Considerations

- How are models monitored overtime?
- What changes in performance are considered noteworthy?
- What responsibilities do developers have? Clinicians? Regulators?

TIME TO SHARE WITH THE LARGE GROUP

Share both overall thoughts – as well as those specific to your model's architecture and validation process

LARGE GROUP DISCUSSION

Consider your model type in communicating your thoughts to the group

Model A
Non-Generative Model

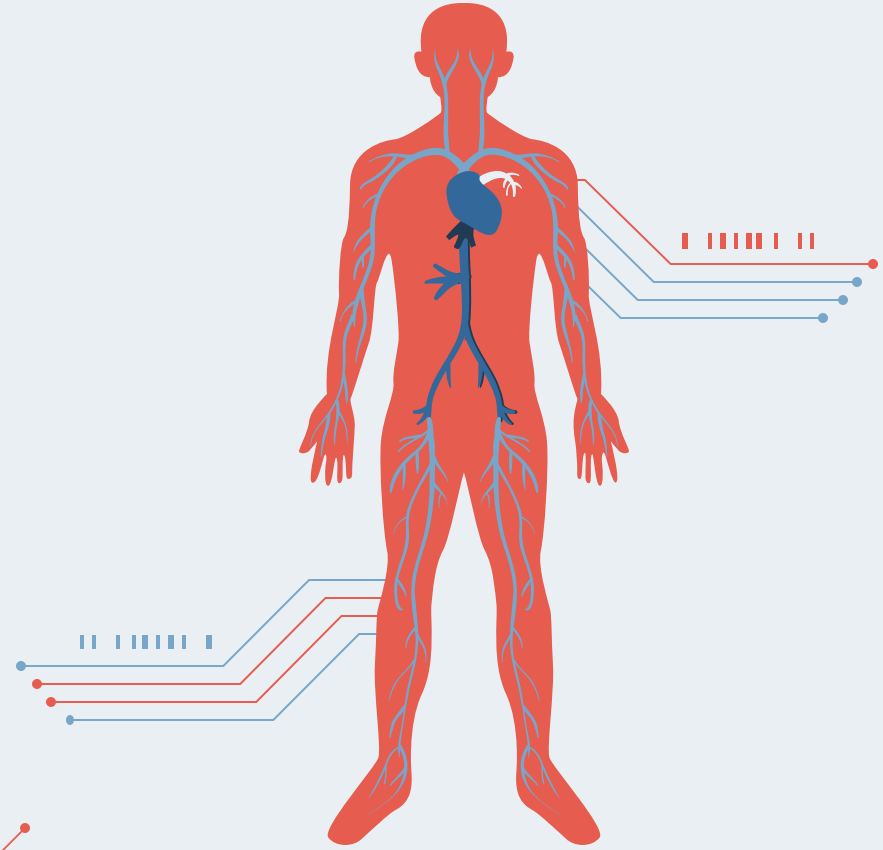
Model B
Non-generative, Image-based Model

Model C
Multimodal Model

Model D
Generative Model

EVIDENCE & RESOURCES

In the following slides, we recommend relevant evidence and resources related to this discussion



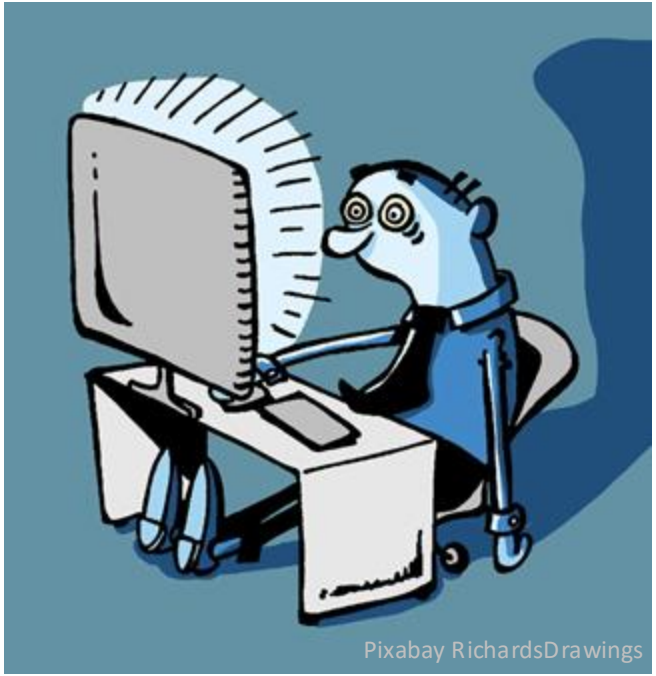
Challenges of automation & “human in the loop”

Humans are not good monitors

- If nothing happens, we stop paying attention
- Loss of situational awareness

Loss of critical skills from lack of practice

- Deskilling
- Never-skilling



Bainbridge. Ironies of automation. Automatica, Volume 19, Issue 6,1983,Pages 775-779

Ethical Duties of Various Roles

AMA Framework

Source: https://www.ama-assn.org/practice-management/digital-health/advancing-health-care-ai-through-ethics-evidence-and-equity

RESPONSIBILITY	DEVELOPER	DEPLOYER	PHYSICIAN
PLANNING AND DEVELOPMENT			
Ensure the AI system addresses a meaningful clinical goal	○		○
Ensure the AI system works as intended	○		○
Explore and resolve legal implications of the AI system[1] prior to implementation and agree upon appropriate professional and/or governmental oversight for safe, effective, and equitable use of and access to health care AI	○	○	○
Develop a clear protocol to identify and correct for potential bias	○	○	○
Ensure appropriate patient safeguards are in place for direct-to-consumer tools that lack physician oversight	○		
IMPLEMENTATION AND MONITORING			
Make clinical decisions such as diagnosis and treatment			○
Have the authority and ability to override the AI system			○
Ensure meaningful oversight is in place for ongoing monitoring		○	○
Ensure the AI system continues to perform as intended through performance monitoring & maintenance	○	○	
Ensure ethical issues identified at the time of purchase and during use have been addressed[2]		○	
Ensure clear protocols exist for enforcement and accountability, including a clear protocol to ensure equitable implementation	○	○	○

1. Such as issues of liability or intellectual property
2. Including but not limited to safeguarding patients' and other individuals' privacy interests and preserving the security and integrity of personal information; securing patient consent; and providing patients' access to records"

AI Alone Won’t Transform Healthcare

AI can exacerbate existing problems in the healthcare system UNLESS – at the local level – we:

1. **Revamp incentives –** compensate AI activities, including those focused on prevention
2. **Embed AI into medical education and training**
3. **Engage doctors and patients in AI development and use**

https://hbr.org/2025/05/ai-alone-wont-transform-u-s-healthcare

Avenues to Actualize AI's potential

1. Ensuring safe, effective and trustworthy use of AI
2. Promoting and development of an AI-competent health care workforce
3. Investing in AI research to support the science, practice, and delivery of health and health care
4. Promotion of policies and procedures to clarify AI liability and responsibility

Matheny et al. 2025. Health Affairs
(doi: 10.1377/hlthaff.2024.01003)

Relevant Readings

1. Char, D. S., Shah, N. H., & Magnus, D. (2018). *Implementing machine learning in health care—addressing ethical challenges*. New England Journal of Medicine, 378(11), 981–983.
2. Leslie, D. (2019). *Understanding artificial intelligence ethics and safety*. The Alan Turing Institute.
3. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). *Dissecting racial bias in an algorithm used to manage the health of populations*. Science, 366(6464), 447–453.

THANK YOU FOR PARTICIPATING

Time for another discussion!