



# Understanding decision-making in diagnostic AI and its implications for clinician skills

Allison Gardner



WOMEN  
LEADING  
IN AI



# We need to consider HOW and WHY we use AI systems and long term consequences



WHAT IS DESKILLING?



LEARNING FROM  
OTHER SECTORS



IS IT A CONCERN FOR  
HEALTH PROFESSIONS?



WHAT CAN WE DO?

# Could we place a cap on developing and current expertise?

**Table 2 | Illustration of diagnostic performance of our AI model and physicians**

Disease conditions	Our model	Physicians				
		Physician group 1	Physician group 2	Physician group 3	Physician group 4	Physician group 5
Asthma	0.920	0.801	0.837	0.904	0.890	0.935
Encephalitis	0.837	0.947	0.961	0.950	0.959	0.965
Gastrointestinal disease	0.865	0.818	0.872	0.854	0.896	0.893
Group: 'Acute laryngitis'	0.786	0.808	0.730	0.879	0.940	0.943
Group: 'Pneumonia'	0.888	0.829	0.767	0.946	0.952	0.972
Group: 'Sinusitis'	0.932	0.839	0.797	0.896	0.873	0.870
Lower respiratory	0.803	0.803	0.815	0.910	0.903	0.935
Mouth-related diseases	0.897	0.818	0.872	0.854	0.896	0.893
Neuropsychiatric disease	0.895	0.925	0.963	0.960	0.962	0.906
Respiratory	0.935	0.808	0.769	0.89	0.907	0.917
Systemic or generalized	0.925	0.879	0.907	0.952	0.907	0.944
Upper respiratory	0.929	0.817	0.754	0.884	0.916	0.916
Root	0.888	0.847	0.862	0.908	0.902	0.912
<b>Average F1 score</b>	<b>0.885</b>	<b>0.841</b>	<b>0.839</b>	<b>0.907</b>	<b>0.915</b>	<b>0.923</b>

- model
- junior physicians
- senior physicians

If the junior physicians are trained using AI will they ever achieve the skill of the experienced physician?

Are they acquiring own knowledge while using the system?

We used the F1score to evaluate the diagnosis performance of our model and five different physician groups (group 1, group 2, group 3, group 4, and group 5) (see Methods section for description). We observed that our model performed better than junior physician groups but slightly worse than three experienced physician groups. Root is the first level of diagnosis classification.

# Could AI make clinicians lose or never gain expertise?

## What is automation-induced skills fade?

- This is the attenuation of skills and expertise over time due to reliance on automation or the failure to attain expertise due to automation dependent learning.
- This can relate to knowledge acquisition and manual skill.
- It is an unintended consequence that occurs slowly over time and often only becomes apparent in unusual or crisis situations where there is failure of the AI system or failure in human interaction.

## Principle 7: Data Transparency

- Consider how the introduction of AI will change relationships in health and care provision, and the **implications of these changes for responsibility and liability**. Use current best practice on how to explain algorithms to those taking actions based on their outputs.



# Examples are more overt in aviation



January 4, 2013 : Federal Aviation Administration published a Safety Alert for Operators #13002 (FAA):

“ This SAFO encourages operators to promote manual flight operations when appropriate”

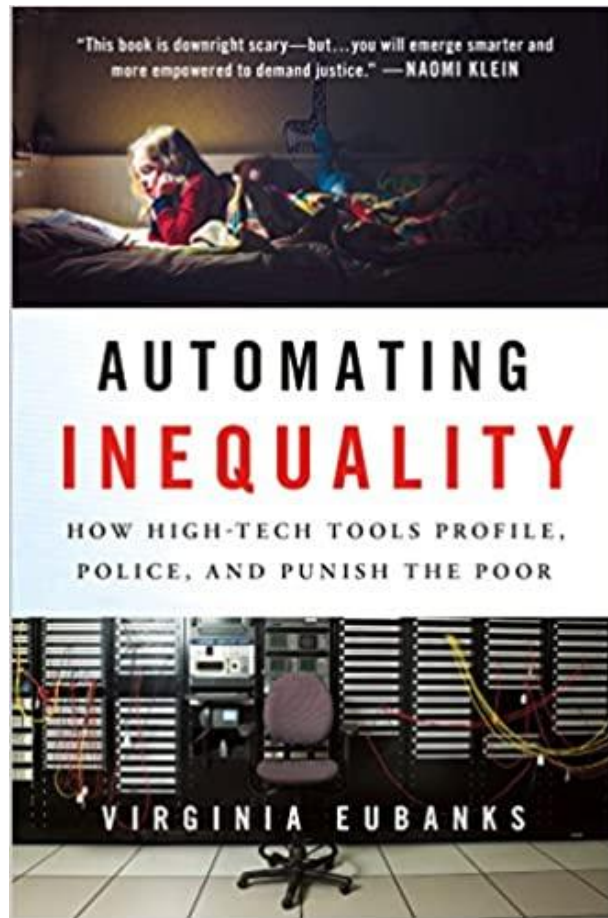
The report stated that overuse of automation led to “a degradation of the pilot’s ability to quickly recover the aircraft from an undesired state”

Colgan airflight 3407 crash on February 12, 2009  
“probable cause to be the pilots' inappropriate response to the stall warnings.” – killed 50 people

Similar incidence have occurred e.g. Air France A330 May 31 2009 – killed 228 people due to “loss of situational awareness”

Glass Cage, Nicolas Carr, 2015

# Examples in social care:



VIRGINIA EUBANKS BUSINESS 01.15.18 08:00 AM

## A CHILD ABUSE PREDICTION MODEL FAILS POOR FAMILIES



Also examples in accountancy and auditing



# Do we need to consider possibility of deskilling in healthcare?

https://twitter.com/Eric\_Topol/status/943559317055094764/photo/2

## Opinion Viewpoint

accurate differential diagnoses, might reshape the traditional CPC (clinical problem solving) exercise, just as the development of imaging modalities and sophisticated laboratory testing made the autopsy less relevant.

As with the EMR, there are legitimate concerns that artificial intelligence applications might jeopardize critical social interactions between colleagues and with the patient, affecting the lived experiences of both groups. But concerns about physician "unemployment" and "de-skilling" are overblown.<sup>9</sup> In the same manner that automated blood pressure measurement and automated blood cell counts freed clinicians from some tasks, artificial intelligence could

bring back meaning and purpose in the practice of medicine while providing new levels of efficiency and accuracy. Physicians must proactively guide, oversee, and monitor the adoption of artificial intelligence as a partner in patient care.

In the care of the sick, there is a key function played by physicians, referred to by Tinsley Harrison as the "priestly function of the physician." Human intelligence working with artificial intelligence—a well-informed, empathetic clinician armed with good predictive tools and unburdened from clerical drudgery—can bring physicians closer to fulfilling Peabody's maxim that the secret of care is in "caring for the patient."

### ARTICLE INFORMATION

**Published Online:** December 20, 2017.  
doi:10.1001/jama.2017.19198

**Conflict of Interest Disclosures:** All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Dr Verghese reported receiving royalties from Random House and Scribner; receiving speakers fees from Leigh Bureau; and receiving fees for serving on the Gilead Global Access Advisory Board. Dr Harrington reported receiving research grants outside the topic area from Merck, CSL Behring, GlaxoSmithKline, Regado, Sanofi-Aventis, AstraZeneca, Portola, Janissen, Bristol Myers Squibb, Novartis, and The Medicines Company; serving as a consultant for Amgen, Adverse Events, Bayer, Element Science, Gilead, MyoKardia, Merck, The Medicines Company, and WebMD; and serving on the boards of directors of Signal Path, Scanadu, the American Heart Association, and Stanford Healthcare. Dr Shah reported no disclosures.

### REFERENCES

1. Hill RG Jr, Sears LM, Melanson SW. 4000 clicks: a productivity analysis of electronic medical records in a community hospital ED. *Am J Emerg Med.* 2013;31(11):1591-1594.
2. Angwin J, Larson J, Mattu S, Kirchner L. Machine bias: there's software used across the country to predict future criminals—and it's biased against blacks. ProPublica website. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>. May 23, 2016. Accessed November 30, 2017.
3. Caruana R, Lou Y, Gehrke J, Koch P, Sturm M, Elhadad N. Intelligible models for healthcare: predicting pneumonia risk and hospital 30-day readmission. In: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY: Association for Computing Machinery; 2015:1721-1730.
4. Breiman L. Statistical modeling: the two cultures. *Stat Sci.* 2001;16(3):199-123.
5. Calif RM, Rosati RA. The doctor and the computer. *West J Med.* 1981;135(4):321-323.
6. Avati A, Jung K, Harman S, Downing L, Ng A, Shah NH. Improving palliative care with deep learning. Presented at: 2017 IEEE International Conference on Bioinformatics and Biomedicine; Kansas City, MO; November 13-16, 2017.
7. Verghese A. Culture shock—patient as icon, icon as patient. *N Engl J Med.* 2008;359(26):2748-2751.
8. Halpern Y, Horng S, Choi Y, Sontag D. Electronic medical record phenotyping using the anchor and learn framework. *J Am Med Inform Assoc.* 2016;23(4):731-740.
9. Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. *JAMA.* 2017;318(6):517-518.

6

122

142

6

https://twitter.com/Eric\_Topol/status/943559317055094764/photo/2

Eric Topol  
@Eric\_Topol

Medicine and #AI: spot on, by @cuttingforstone [jamanetwork.com/journals/jama/...](http://jamanetwork.com/journals/jama/...) @JAMA\_current @dmigam @HeartBobH @StanfordMed

7:10 PM · Dec 20, 2017 · Twitter Web Client

106 Retweets 16 Quote Tweets

142 Likes

Reply Retweet Like Share

Tweet your reply Reply

Federi... @c... · Dec 20, 2017

Replying to @Eric\_Topol @cuttingforstone and 4 others  
Thank you for noticing a blunder. We [?] never raised concerns of physician (quote) "unemployment". As regards (quote) "de-skilling", who "overblew" what? We just wrote that overreliance can induce deskilling. Disputable?

1 1 3 6

Federi... @c... · Dec 20, 2017

However, I am wondering whether thinking that AI "could bring back" meaning and purpose in the practice of medicine" is more pessimistic about the current state of medical practice than it is optimistic about the redeeming

# In healthcare: different deskilling outcomes for novices, non-specialists and experts

ARTIFICIAL INTELLIGENCE

## Artificial Intelligence Makes Bad Medicine Even Worse

A new study out from Google seems to show the promise of AI-assisted health care. Actually, it shows the threat.



PHOTOGRAPH BY GUY LAWRENCE

[Artificial Intelligence Makes Bad Medicine Even Worse | WIRED](#)  
2020

- “The Google study found that AI performed better than radiologists who were not specifically trained in examining mammograms.

**Would it come out on top against a team of more specialized experts? It’s hard to say without a trial.”**

- Also provided interesting insight into role of second reader

[International evaluation of an AI system for breast cancer screening | Nature 2019](#)

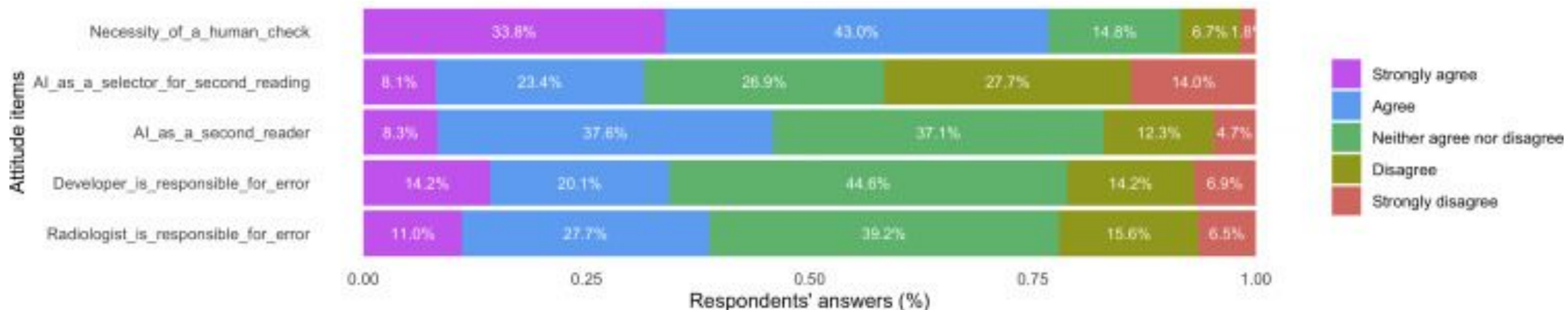


# Example of automation bias and impact on novices versus experts:

- 2013 article showed automation bias in radiologists looking at mammograms was greater than thought
- For novice readers computer-aided detection improved the reliability for easy cases.
- However, it degrades the experts ability to evaluate difficult cases and thus overlooks certain cancers.

[How to Discriminate between Computer-Aided and Computer-Hindered Decisions: A Case Study in Mammography - Andrey A. Povyakalo, Eugenio Alberdi, Lorenzo Strigini, Peter Ayton, 2013 \(sage pub.com\)](#)

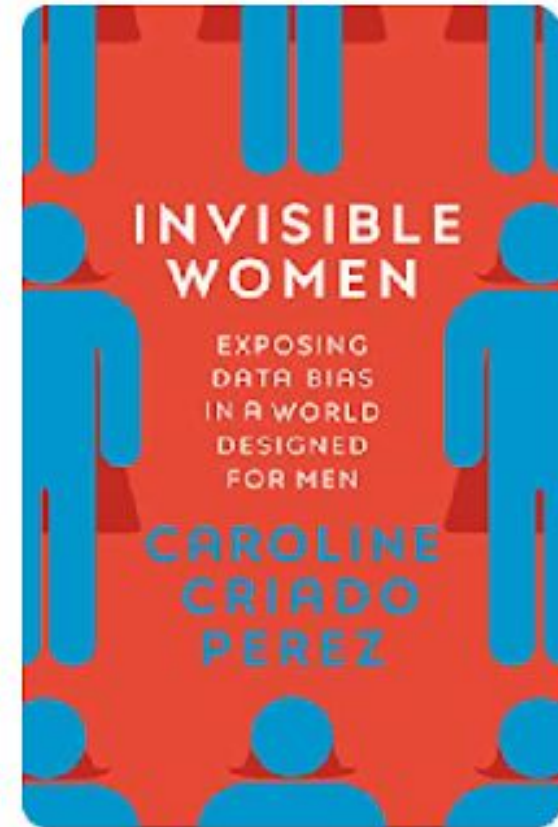
## Raises questions of accountability



# Just a thought regarding algorithmic bias and deskilling



- Could deskilling exhibit biases in terms of one group more impacted than another?



# A guide to good practice for digital and data-driven health technologies



- Undergraduate training and professional development on AI assisted diagnostics and predictions
- Consider training regarding manual skills and unassisted decision making, particularly in crisis situations.
- Decision-making autonomy for clinical staff and diagnosing FIRST
- **Utilise expertise in HCI and knowledge acquisition to help design interfaces and working practices.**
- Real world monitoring of working protocols and AI system performance
- Monitoring for deskilling and implementation of guidance and operational processes to minimise impact.



“The knowledge is not ‘lost’—it lives on in the few remaining specialists and researchers and in a database—but it is no longer being added to”

(Froomkin, Kerr and Pineau, 2018)

Thank you for listening

Dr Allison Gardner

@AllisonCGardner a.gardner@keele.ac.uk

