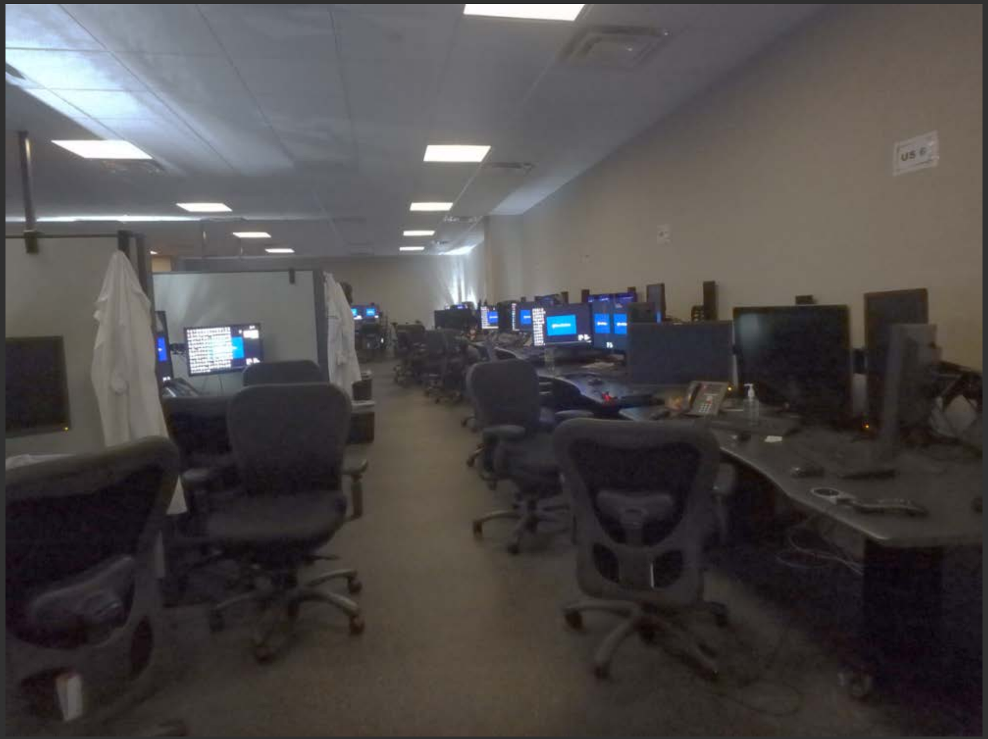
Challenges for artificial/machine intelligence for medical imaging

Will radiology ever disappear as a specialty?

Steven Horii, MD, FACR, FSIIM, FSPIE

Department of Radiology

Hospital of the University of Pennsylvania



Disclosures

- None relevant for this talk

Introduction

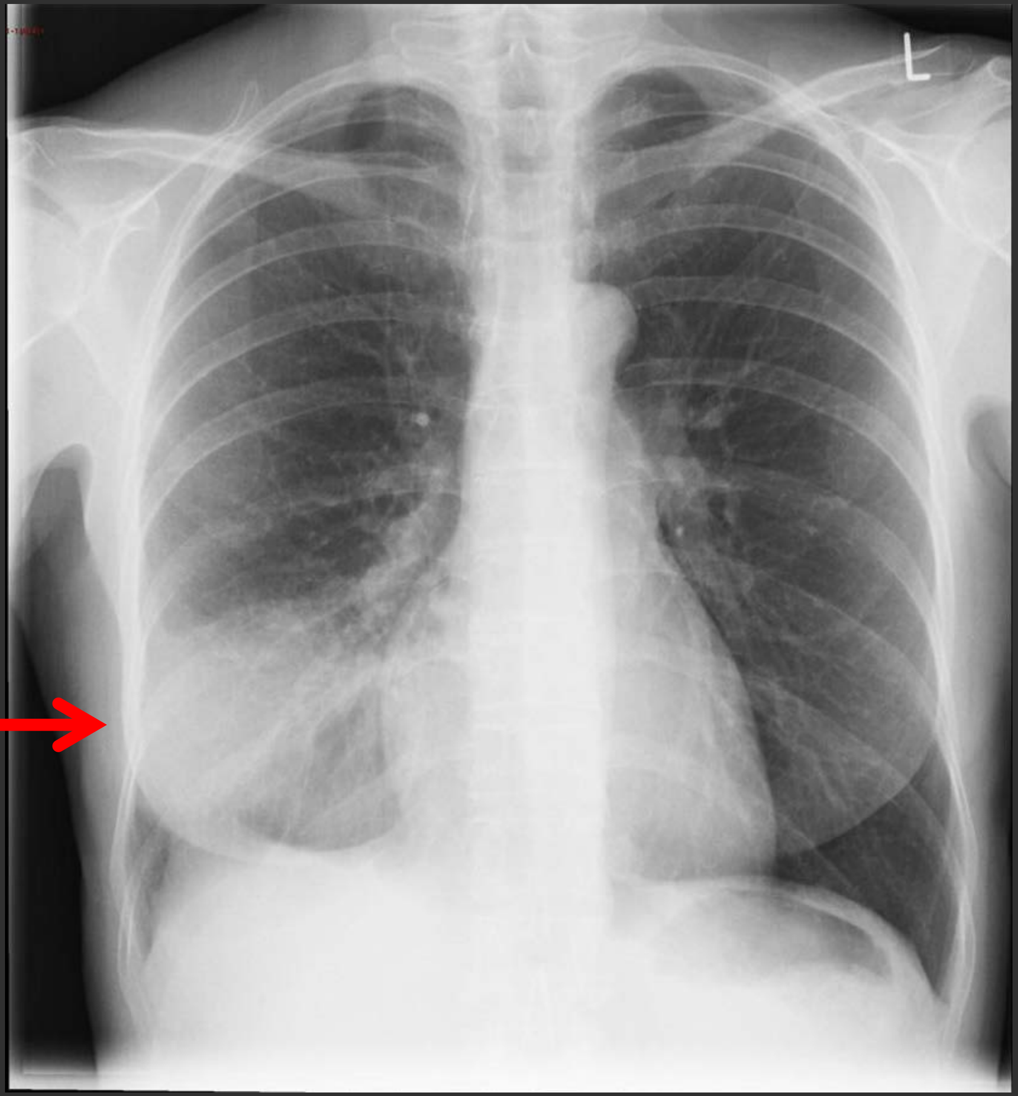
- Radiologists as detectors – good ones
- Challenges facing radiology (and medicine in general)
- What we do well, and don’t (or can’t)
- Where we need help
  - Radiologists’ pain points

A little perception test

- To illustrate what visual challenges radiologists have to meet
- Get ready!

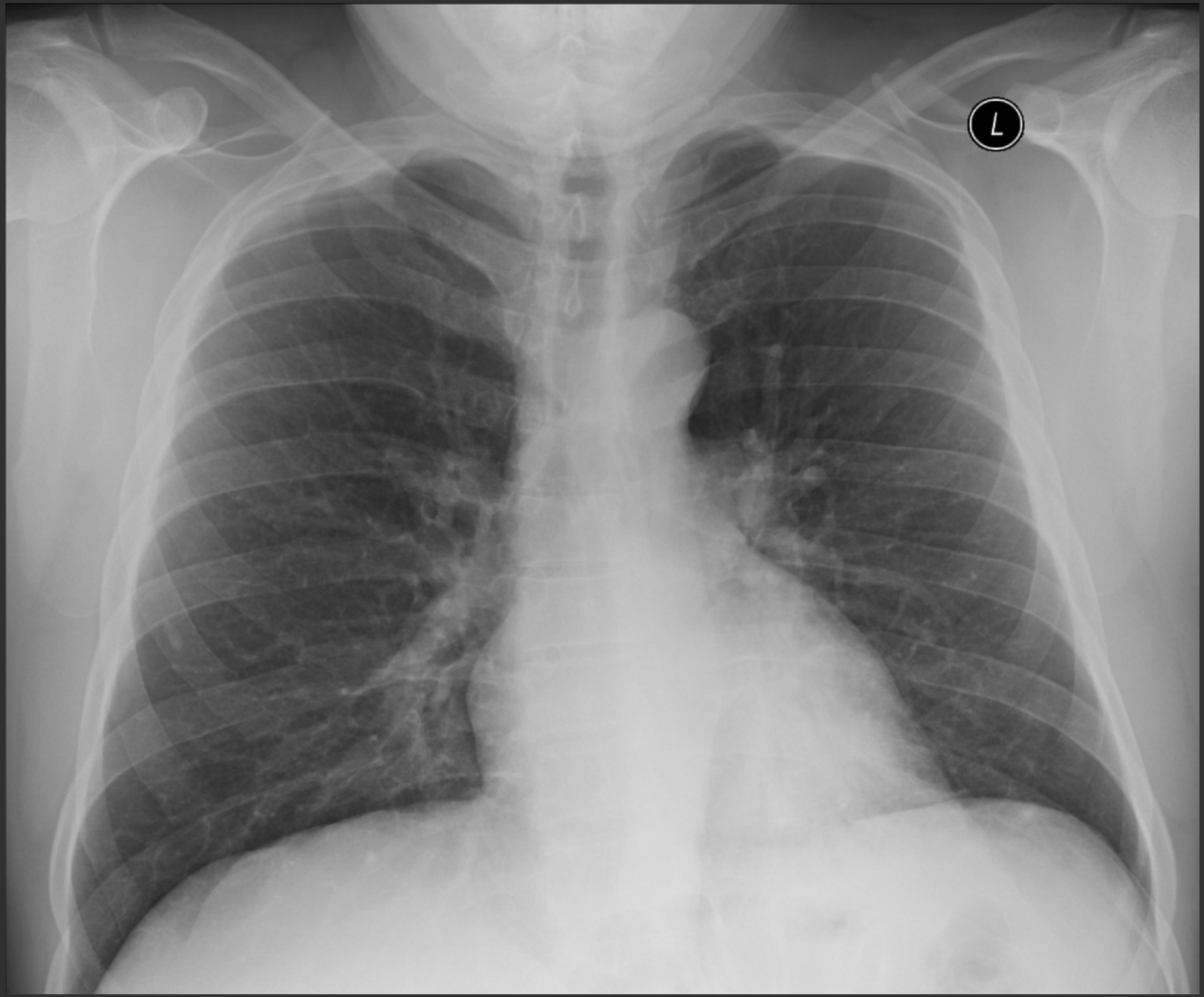
Quick look

Quick look

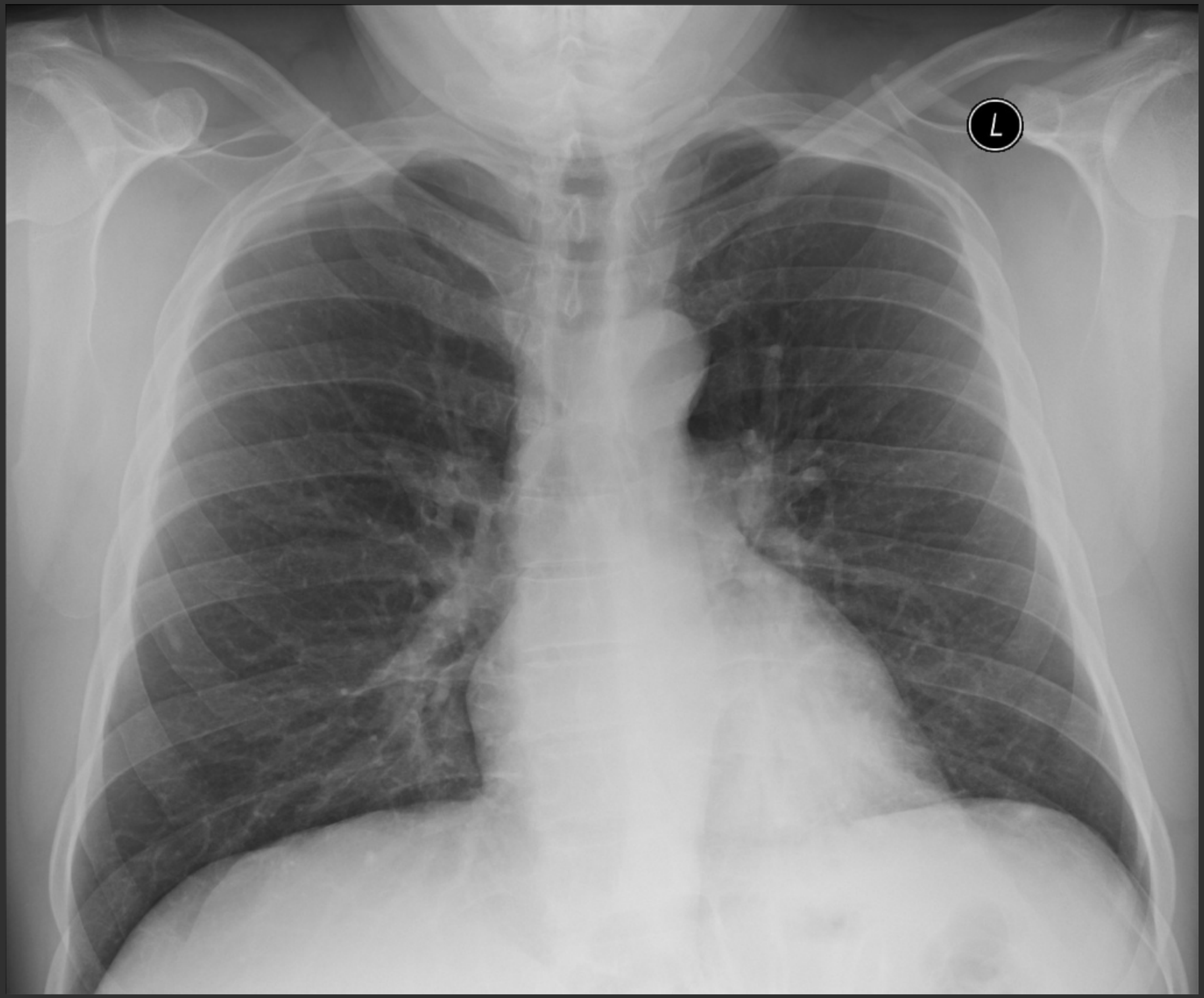


RLL pneumonia

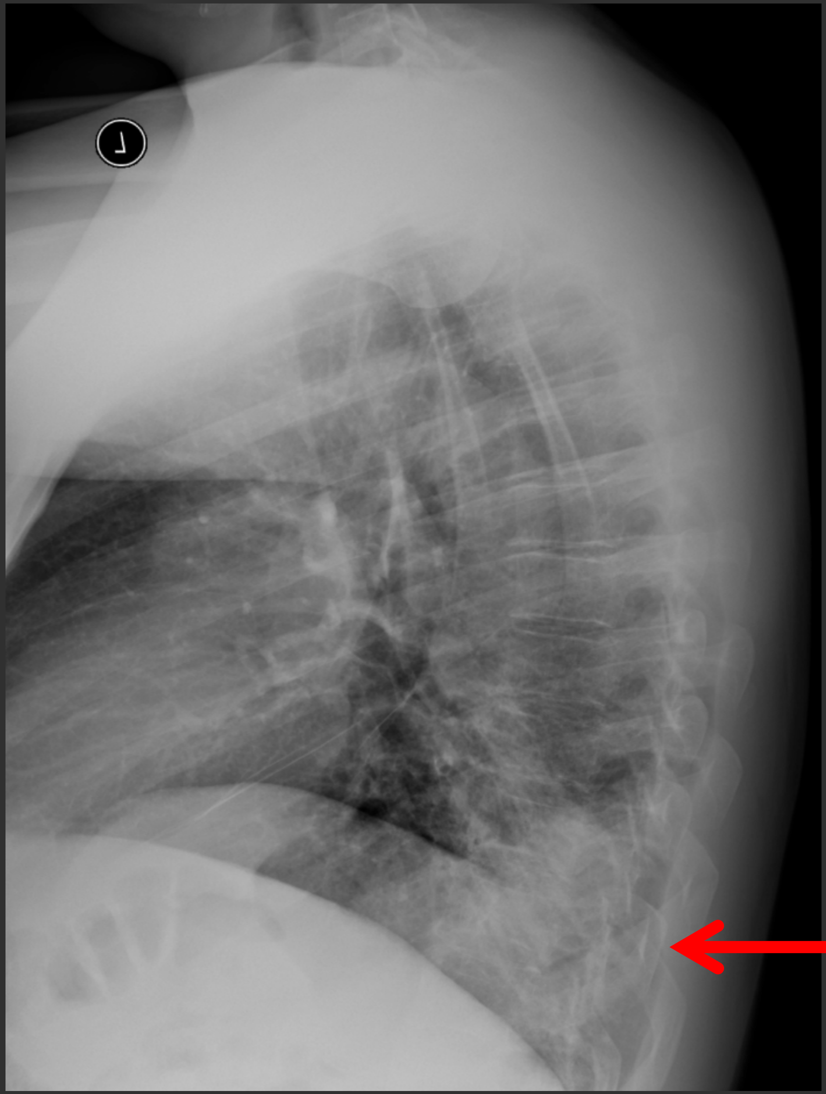
Another quick look



Another quick look

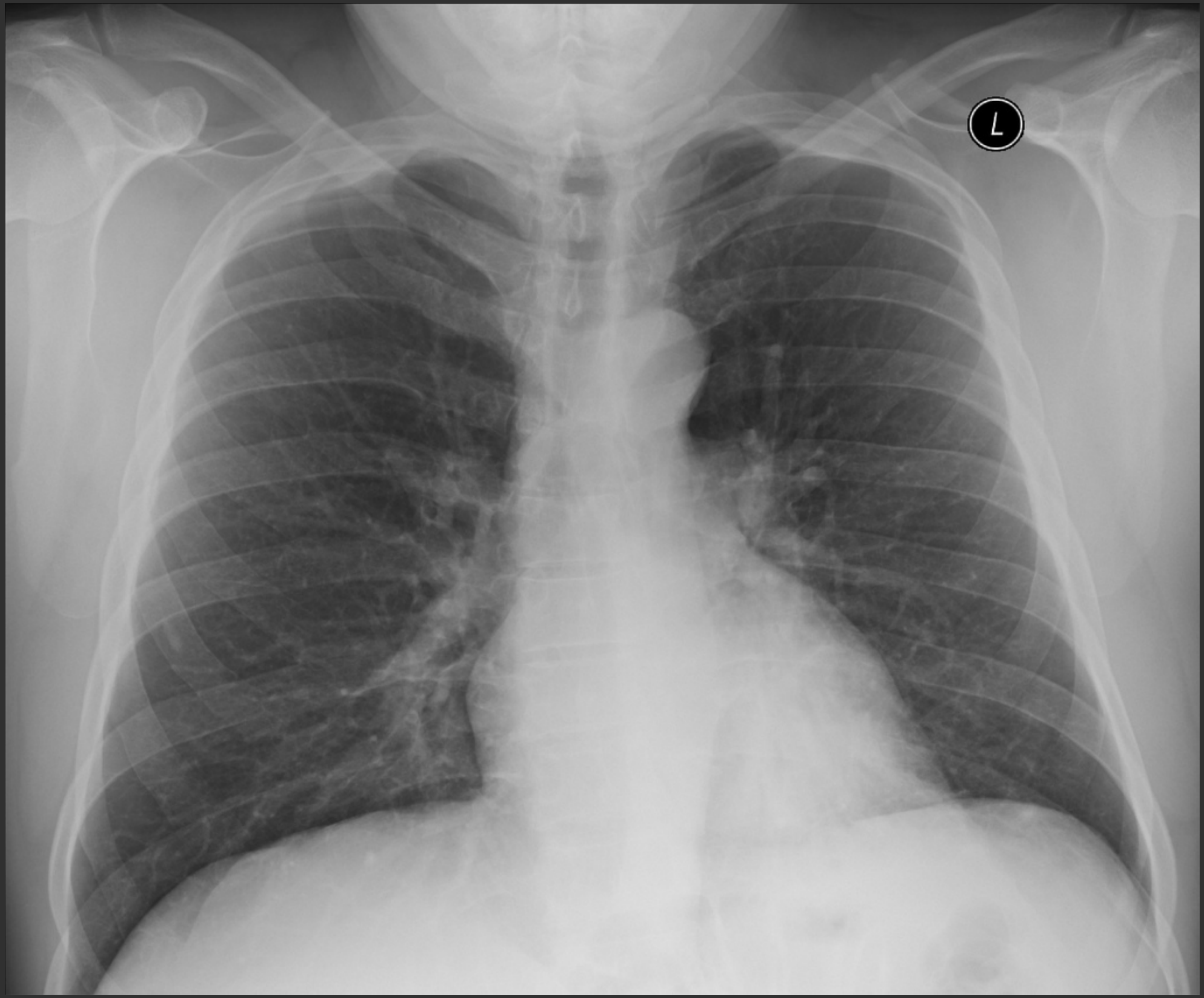


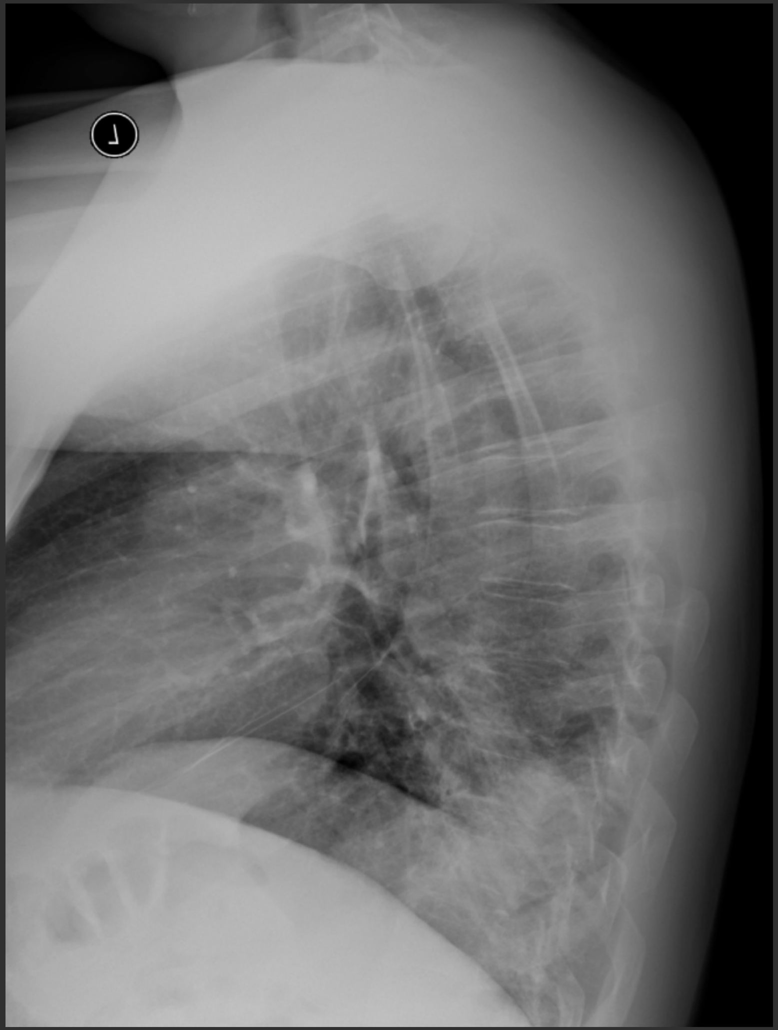
Need something else?



LLL pneumonia

Why we sometimes need two views





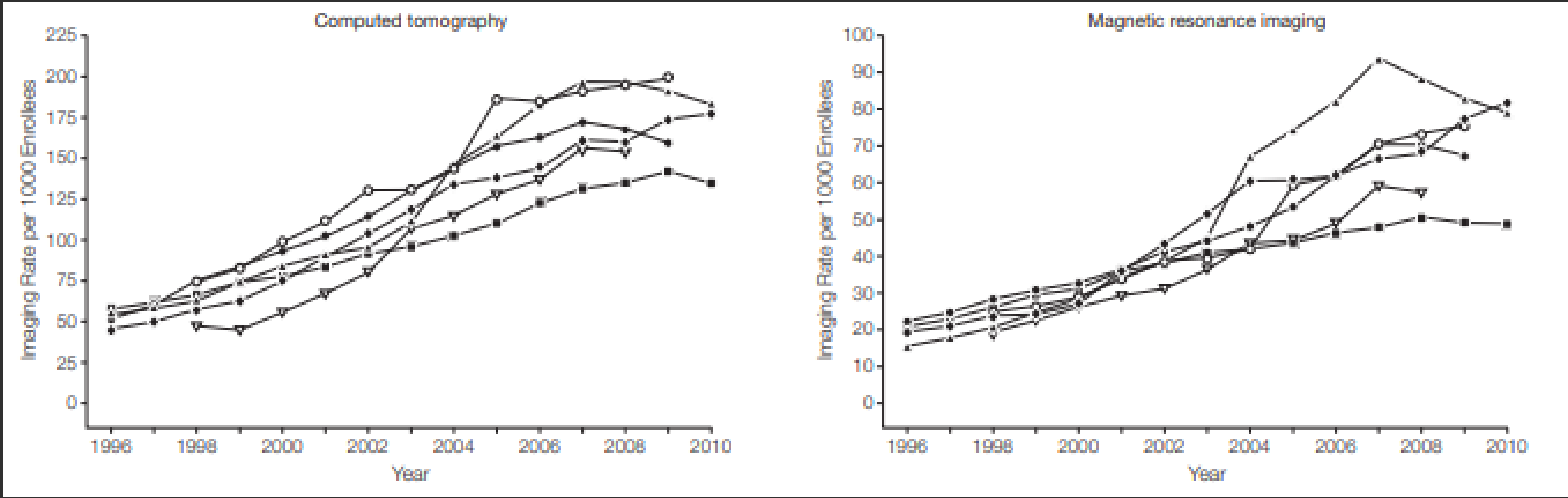
How fast?

- We have had radiologists who could make that LLL pneumonia diagnosis in about 200 milliseconds
- Is that time enough for visual search?
- What kind of visual search?

What those developing ML/AI systems should understand about radiologists

- We are extremely good detectors
- We are generally very fast at detecting abnormalities on images
- Almost all of us – both in academic and private practice – have been experiencing a steady increase in examination volume
- Increasing reports of “burnout” among radiologists and other physicians

Growth in CT and MRI



Smith-Bindman R, Miglioretti DL, Johnson E, et al: Use of Diagnostic Imaging Studies and Associated Radiation Exposure for Patients Enrolled in Large Integrated Health Care Systems. JAMA 2012; 307(22): 2400-2409.

It's not about money

- Radiologists are among the highest-paid specialist physicians
- However, in a survey of our Department faculty, 30% said that they would be willing to work fewer hours even if it meant a cut in pay

Good detectors? When are we not so good

- Mammography: Current cancer miss rate is about 13% (false negative)
- But, only about 30% of breast biopsies are positive for cancer
- Can we reduce the false negative rate without further increasing the false positive rate?
- Some early work with AI suggests that this may be the case

Why do we miss things?

- In mammography, eye-tracking studies have shown that, in missed cancers, the radiologist did fix his or her gaze on the lesion
- Other reasons include:
  - The satisfaction of search problem
  - Fatigue and decrease of vigilance
  - Lack of familiarity with new imaging techniques (cited in early higher breast cancer miss rates with digital breast tomography)

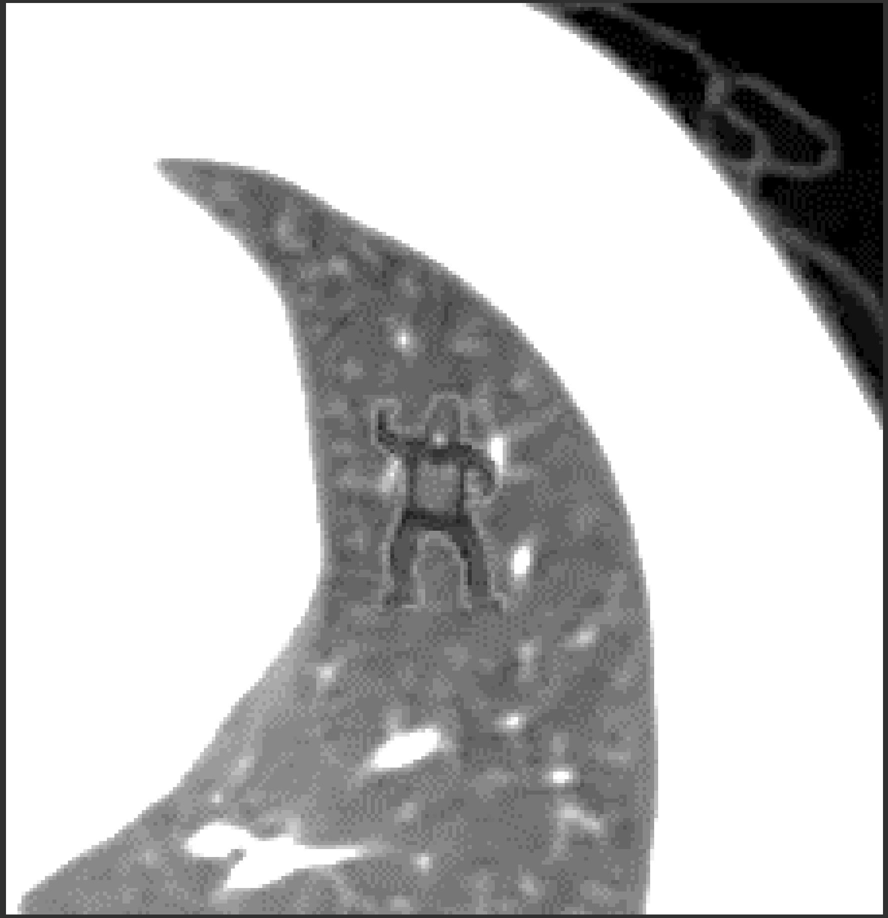
Why do we miss things?

- Distractions (phone calls, curbside consults, called away to see a patient or perform a procedure)
- Inattentional blindness
  - All of you have probably seen the demonstration of the gorilla walking through people passing basketballs – very few notice until told
  - This idea was further tested on radiologists

Find the lung nodule

Did you see it?

Did you see the dancing gorilla?



Drew T, Vo ML, Wolfe JM: The invisible gorilla strikes again: Sustained inattentional blindness in Expert observers. Psychol Sci 2013; 24(9): 1848-1853.

Missing the gorilla

- Do not feel badly if you did not see it – 83% of radiologists did not (and nearly 100% of non-experts did not)
- Eye tracking showed that the majority of radiologists did fix their gaze on the gorilla
- Inattentional blindness is the visual miss of something when engaged in a different task
- Do you really want to text while driving?

The stopped clock

- Have you ever glanced at a clock with a second hand and the clock seems to be stopped – until you wait longer than a second and you see the hand tick again?
- This is known as chronostasis and occurs because perception is suspended during eye saccades (or else our visual cortex would have to process blurry images). When the saccade stops, the signal that tells the visual system to resume processing results in a short delay.

Perception bandwidth

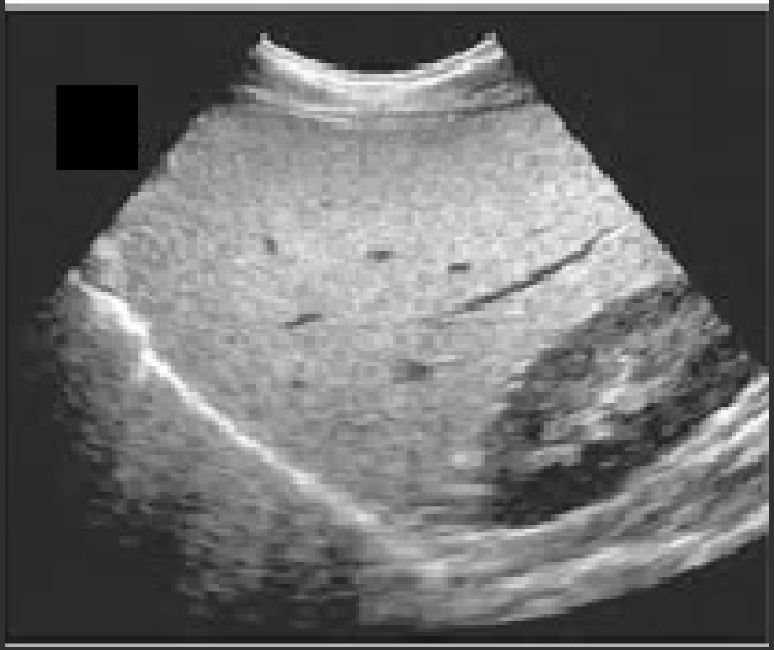
- Effectively, our perception systems have a bandwidth
- It is why we really cannot multitask; what seems to be multitasking is switching between tasks not actually performing two or more simultaneously
- AI systems have a bandwidth too, but processing is very much faster than in our nervous systems (for some things)

What else is hard for us to see?

- Our visual processing has difficulty characterizing textures beyond second-order statistics
- We do reasonably well differentiating first order statistics (e.g., mean and variance)
- Less well at second order statistics (relationships between pixels)

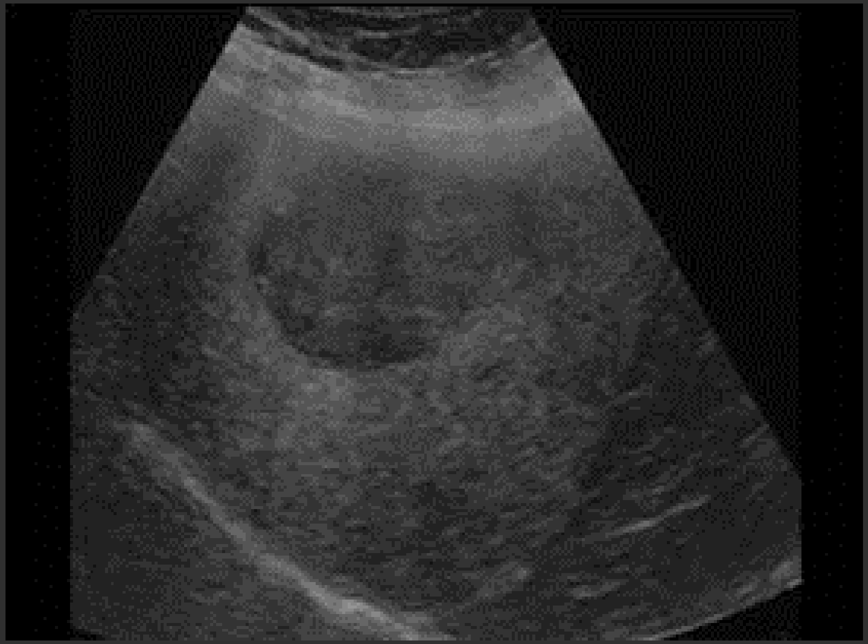
Why is texture important?

- Change in texture of tissues (e.g., liver, renal cortex, thyroid, etc.) often occurs in different disease states



Texture

- This is an example of a difference in first order statistics. The image with the bright liver has a higher mean pixel value
- How about these?



Can you see any difference?

- These look very similar – their first order statistics are probably very close
- What do we do to figure this out?
  - We turn to additional imaging methods (MRI, CT)
  - Or the use of agents that are specific for cells in a lesion
  - Or exploit the often different vascularity of benign compared to malignant lesions

AI and texture analysis

- It is easy for a computer algorithm to compute higher-order statistics for an image
- AI can then be trained to classify images based on differences in these higher-order statistics
- This is something our visual systems cannot do

Picking things out

- Distinguishing tissues in images leads to another task with which we need help
- We are often asked to compare the size of an abnormality before and after (or during) treatment
- We make measurements, but this can be very difficult
- The task of selecting tissues with different characteristics and separating them is segmentation

Segmentation

- Research in this goes back for many decades
- After all, if you want to do analysis of a lesion or have an AI system characterize it, you have to find it in a background of other tissues
- This is another area in which AI could benefit us
- Typically, though, the “gold standard” is a human observer; I always ask those doing segmentation work, “How do you know your segmentation algorithm is not *better* than a human?”

How about things that challenge AI?

- An anecdote about my amazing a resident
- Abstraction
- Diverse experience
- The two (conflicting?) approaches to AI
- Jokes – fun with AI

Amazing a resident

- I was reviewing an abdominal ultrasound with a resident
- He properly said that the patient had cirrhosis
- But he wondered why a 25-year old who did not drink, did not have hepatitis, and was not on any medication toxic to the liver - would have cirrhosis
- I asked the resident, “Check to see if he had congenital heart disease.”

Amazing a resident

- He looked through the chart and said, “Yes he did!”
- I said, “I’d guess it was hypoplastic left heart syndrome”
- The resident was amazed – he asked me how I knew. I followed this with another question
- Did he have a procedure called a Fontan?

Amazing a resident

- At this point, I think the resident thought I must have reviewed this case before, but I said no and explained my “detective” work
- Hypoplastic left heart syndrome can be ameliorated to the point where the child can have a near-normal level of activity
- The surgery is called a Fontan procedure
- However, the procedure results in increased pressure in the hepatic veins

Amazing a resident

- That persistent high hepatic vein pressure results in development of cirrhosis
- An AI system could easily make the diagnosis of cirrhosis, but what about the rest?
- It might then produce a gamut – a list of all causes of cirrhosis, but that is not useful to the referring physician. She or he needs a specific diagnosis to establish treatment.

Abstraction

- Radiologists learn to apply principles learned in one imaging situation to others
- A radiologist learns that a benign ovarian tumor called a dermoid may have a fat-fluid level on ultrasound and MRI
- Seeing a fluid-fluid level in a knee MRI and determining that it is a fat-fluid one, she may work out that it is due to an injury of the knee

A shout out to Brad Erickson for the fluid-fluid level idea

Diverse experience

- Creating a machine learning system that is the equivalent of a general radiologist is a very difficult task
- Radiologists are typically able to interpret imaging studies from many imaging devices and across different specialized areas
- You would have to build multiple AI applications and integrate them to do something similar

The two schools of AI thought

- “Symbolists” – mostly AI with which we are familiar: Develop systems based on logical rules and models of the world
- “Connectionists” – build systems based on neural networks modeled on, or inspired by, biology
- A combination of these approaches has been developed by Deepmind, MIT, and IBM and is called the “Neuro-Symbolic Concept Learner” and learns in a manner not dissimilar from the way a child learns.

Man J, Gan C, Kohli P, Tenenbaum JR, Wu J: The Neuro-Symbolic Concept Learner: Interpreting Scenes, Words, and sentences From Natural Supervision. ICLR 2019 conference paper.

The Neuro-Symbolic Concept Learner

- Learns basic object-based concepts
- Learns relational concepts
- Combines these to answer questions based on visual cues
- Could this be a way for AI systems to learn abstraction? The authors seem to be doing this.

Radiology’s pain points

- Improving diagnostic accuracy – that’s what grabs attention and has led to the claims about radiologists being out of work
- But there are other areas that could greatly improve our efficiency and quality
- And improve service to patients as a part of this

Speech wreck ignition

- Yes, that should be “speech recognition”
- As good as it is, it is still not close to a human transcriptionist when it comes to accuracy
- We keep lists of our favorite errors:
  - “The examination was performed horribly”
  - “The patient has men of Russia”
  - “The most likely diagnosis is wannabees”

Speech recognition

- Adding to our workload is that we are now editors
- The speech recognition systems make word substitution errors, not spelling errors, so they are difficult to find
- Some errors are dangerous – we have had our system leave out the word “no”

Finding information in the EMR

- I may spend as much, or more, time looking for information in the EMR as I do interpreting an exam
- The history given to me is “abnormal liver function tests” – you would think the EMR would automatically pull up those lab values but it does not
- If I don’t get enough history, I go searching through the notes to find out more

Critical results reporting

- Certain conditions require a telephone call to the physician responsible for the patient
- Trying to reach the appropriate person (not uncommonly at 4:30 PM on a Friday afternoon) can take an hour.
- This is still a manual procedure and done (at our facilities) by radiologists

Scheduling examinations

- We have “special needs” patients
- It may take longer for them to have the pre-procedure or examination work done
- Yet we tend not to learn from this, so these patients get scheduled in the same length time slot as someone who does not need extra time
- An AI system could learn from readily available information in our systems to adjust scheduling appropriately

Thanks to Raym Geis for this example

The challenge of the new

- Genomics and radiomics are vastly expanding the dimensionality of images
- For any given image or even for a given pixel, there is the potential to have non-imaging data associated with it
- Having that dimensionality available can be of great diagnostic and therapeutic benefit
- But it needs to be correlated with diagnoses and outcomes – help from our pathologists, oncologists, and geneticists

The challenge of the new

- The information dimensionality becomes so complex that for us to perceive it holistically is very difficult if not impossible
- We will need assistance with this

The example from architecture, engineering, and construction

- How do you think a construction project like the new Hudson Yards in New York gets built?
- There are thousands of contractors, each with a schedule
- There are thousands of manufactures who need to deliver their products on time
- There are inevitable changes in requirements; those need to be made known to all the participants

How is all this accomplished?

- Building information modeling (BIM)
- All the contractors, suppliers, finance managers, architects, engineers, legal experts – have access to a uniform, continuously updated model of the project
- Any change in design or schedule is made available in real time to all concerned

BIM is not new

- Roots in the 1970s
- Practical systems in the 2000s
- Mandated by the Federal Government for almost all construction projects
- So, why can’t we apply this to healthcare?
- We think of the EMR as doing this, but it clearly does not
- There is not enough intelligence in the EMR

Emotional work

- Think of what nurses do in addition to the usual duties
- The same for flight attendants
- And home care workers
- They provide a level of emotional support

The toll-taker experiment

- **Toll takers on one day would touch the hand of the driver when giving change; on the alternate day, they would not.**
- **Drivers were surveyed and their opinions about the toll takers was assessed. Those who had gotten change on the day the toll takers touched their hands rated the toll takers much more favorably than the same toll takers on the non-touch day**

Automation and emotional work

- There is an opinion that emotional work will not be replaced by automation or AI
- Those whose work has an emotional component will not likely lose out to a robot

AI and jokes

- Humor is very complex
- Creating jokes is not a trivial task for computers and machines, so far, cannot do this
- There are, however, riotously funny results of AI gone wrong

Some unintentional AI humor

This rather bizarre picture of a cat was synthesized by an AI algorithm that searched examples on the Internet. The weird text is because most cat photos online have text on them, so the AI interpreted the text as being part of the cat.

An AI comes up with dessert recipes

@@@@@ Noo No CA Cracker DECOR

cookies, deserts, holiday

1/2 cup brown sugar
1/2 cup sugar (or beaten
1 1/2 cup wheated sugar
1 tablespoon baking soda
1 cup brown sugar
3/4 cup brown sugar
2 egg whites
1/4 cup sugar
1 cup packed brown sugar
1 cup brown sugar
3 each egg whites
2 1/2 cup sugar
1 tablespoon powdered sugar

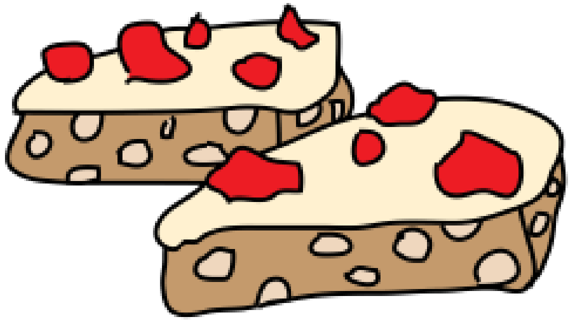
Cream the egg mixture and vanilla. Beat mixture to mix well. Add egg whites and vanilla and flAvour.
ADDININTA MONFA Mincecio: Mapkamares thoroughan Cookies.

Yield: 12 bars

Yield: 1 servings

----- 1 teasp'ove 9x95

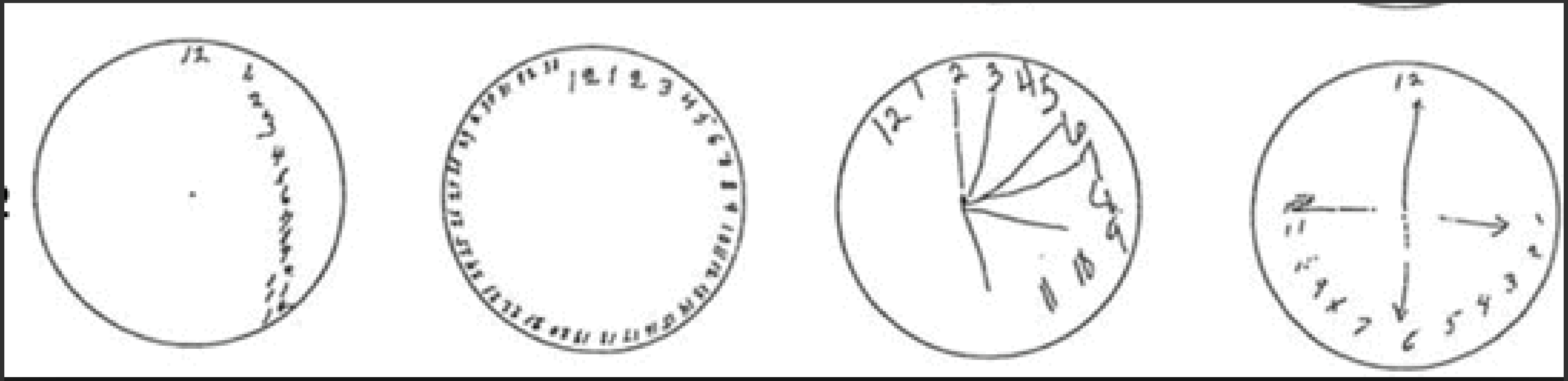
1/2 tablespoon liquid sugar cry or figs
3 cup flour
1/2 teaspoon baking soda
1/2 teaspoon cinnamon
1/2 teaspoon nutmeg
1/2 teespoon cinnamon; cocoa
1 cup pegar
3 cup granulated sugar
1 fresh meat carefus, topping
1 biscotti
5 acd: (375) (24 browned)



: my pressing

An AI and paint color names

The AI timekeepers



Conclusion

- Radiologists are very good at what they do
- There are limitations based on perception and workload
- AI can help with the “pain points” of radiology
- AI can make us better, not replace us – if we learn to use AI and understand its strengths and limitations
- We need to understand how information from new sourcss and new therapeutics can improve care of patients.

Image credits

Liver:

https://www.ultrasoundpaedia.com/segments/

https://www.semanticscholar.org/paper/Clinical-relevance-of-reporting-fatty-liver-on-in-Mahale-Prabhu/a5d67e59ba832a5825e0340a310a2cf4de2a5385

http://www.ultrasoundcases.info/Case-List.aspx?cat=129

HCC: Bagley JE, Paul DE, Halferty S, DiGiacinto: The use of contrast-enhanced ultrasonography for the characterization of focal liver lesions. J Diagn Med Sonography 2017; 33(6): 500-511.