

Graded Automation

Using Conformal Prediction to Safely
Deploy AI Tools in Medical Coding

Christopher Snyder, MD PhD



Washington University School of Medicine in St. Louis

Disclosure

In compliance with the requirements of the Continuing Medical Education (CME) accrediting body, I hereby declare that I have no financial relationships or affiliations with commercial interests to disclose.



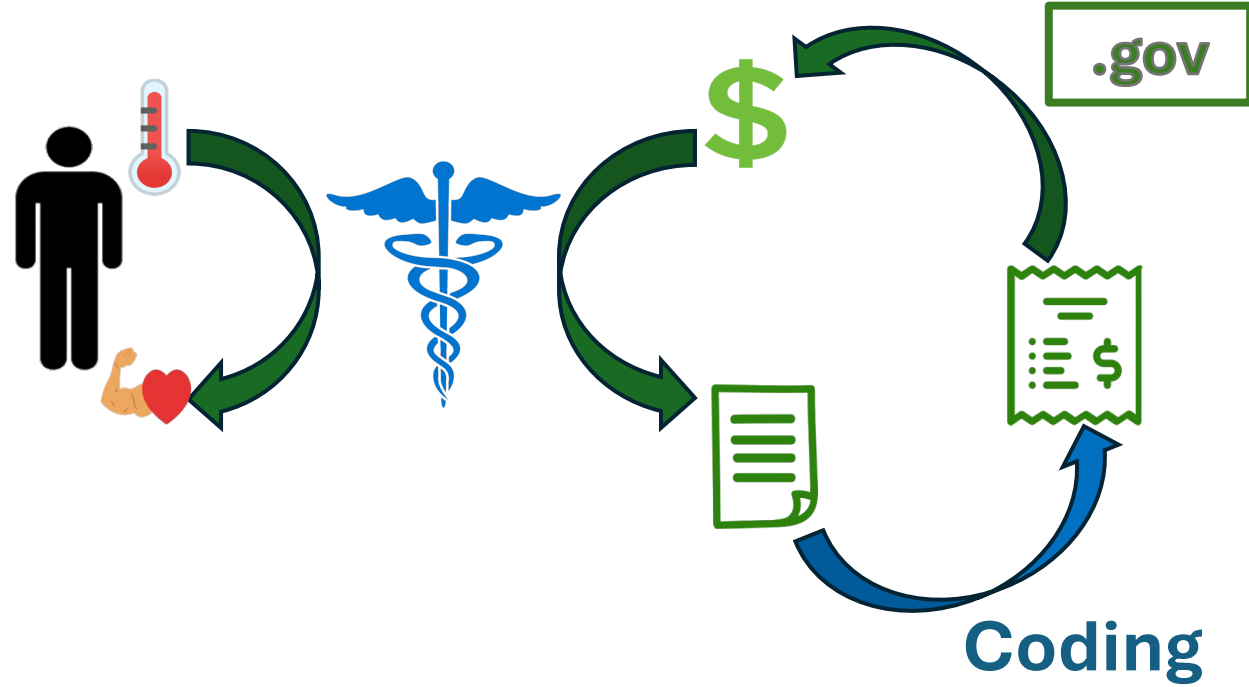
The Future of AI in Medicine

- Promising
- Unclear

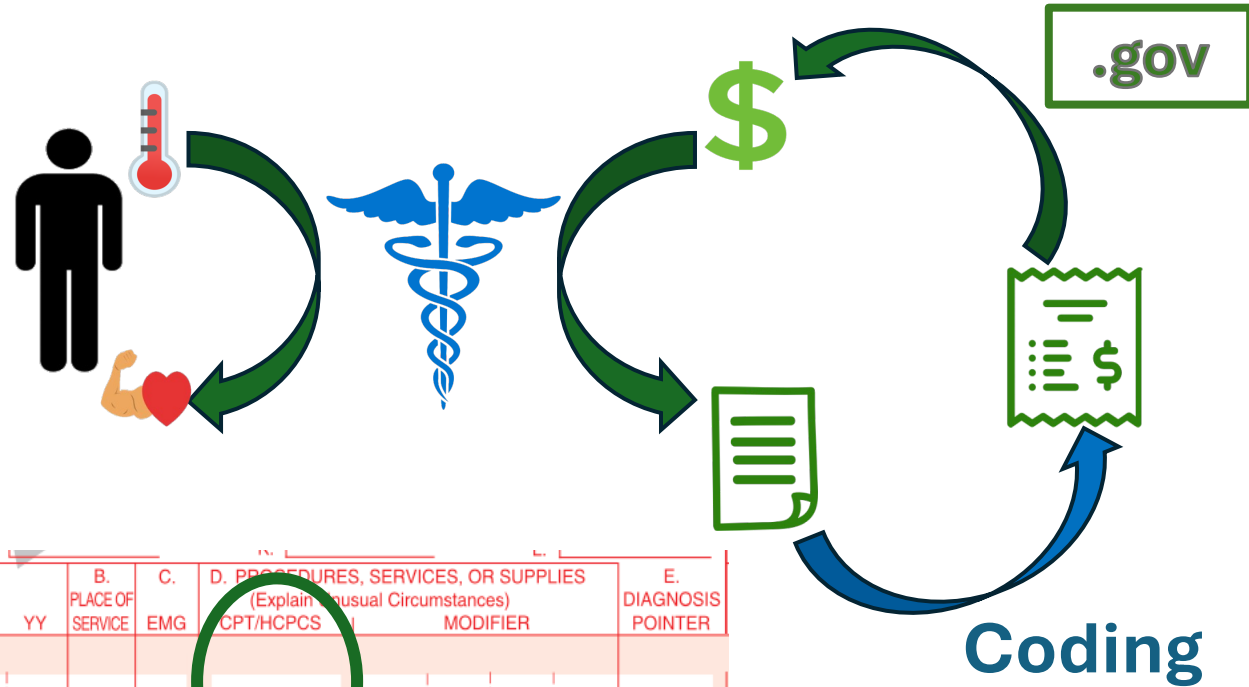


Medical Coding

- Essential
- Labor Intensive
- Worse with time



Medical Coding



Itemized Claim



24. A.	DATE(S) OF SERVICE						B. PLACE OF SERVICE	C. EMG	D. PROCEDURES, SERVICES, OR SUPPLIES (Explain Unusual Circumstances) CPT/HCPCS MODIFIER	E. DIAGNOSIS POINTER
	From	To								
	MM	DD	YY	MM	DD	YY				
1										
2										

Coding

Pathology Report



Clinical Diagnosis : "The patient is a 79-year-old woman with history of CML..."

Gross Description : "Specimens are received in two formalin containers..."

Final Diagnosis : "Bone marrow, right...CML, chronic phase-No excess blasts..."

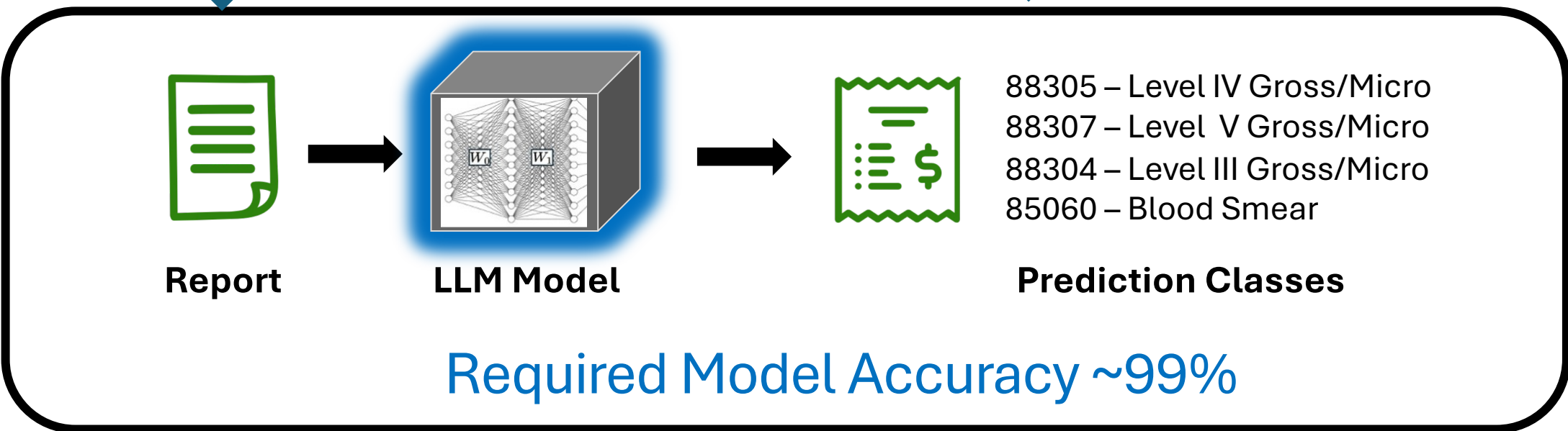
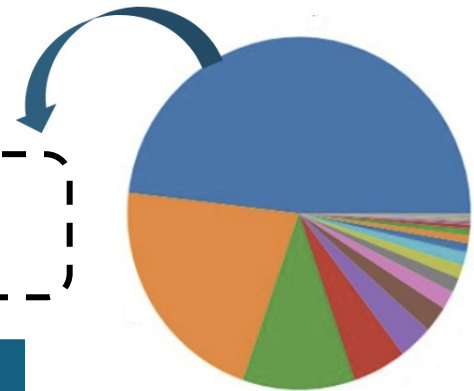
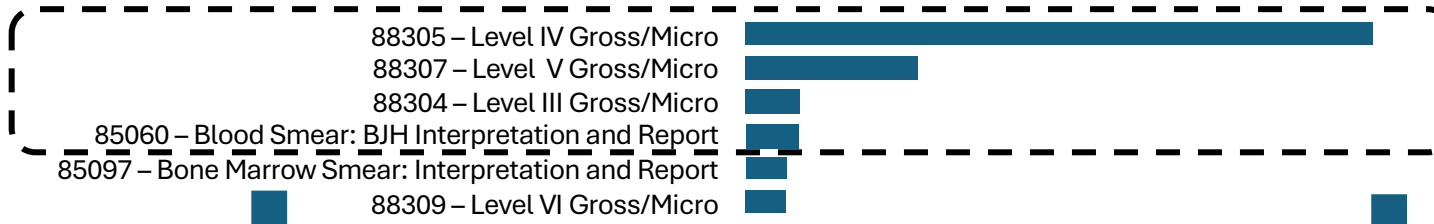
Can A Computer Do It?

START:

50k Reports



CPT
Codes



Result 1: No, AI Cannot Easily Code Reports

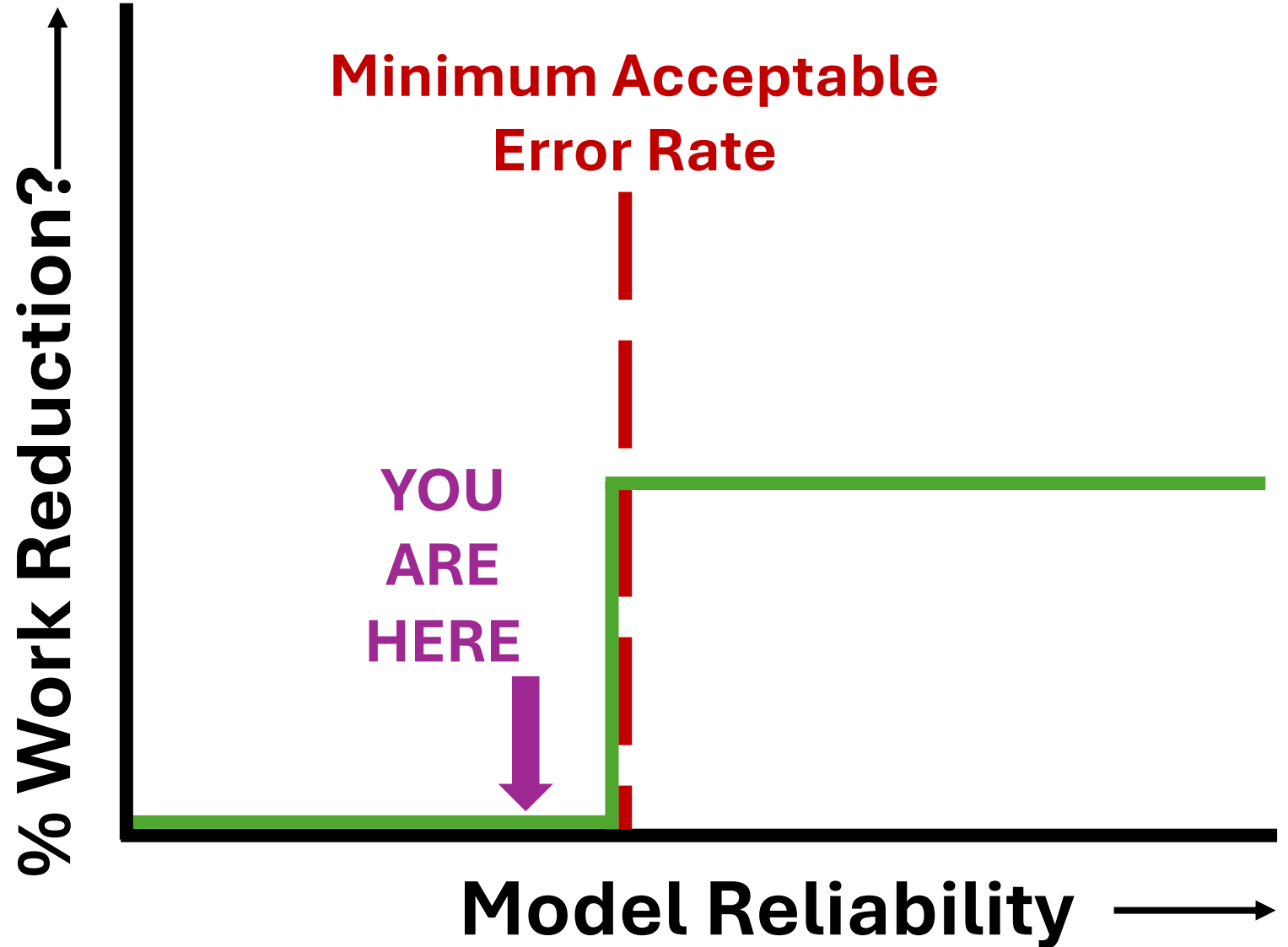


- *Prior work reaches similar conclusions [1]*
- *What does this mean for the future of AI in Pathology (Report Coding)?*

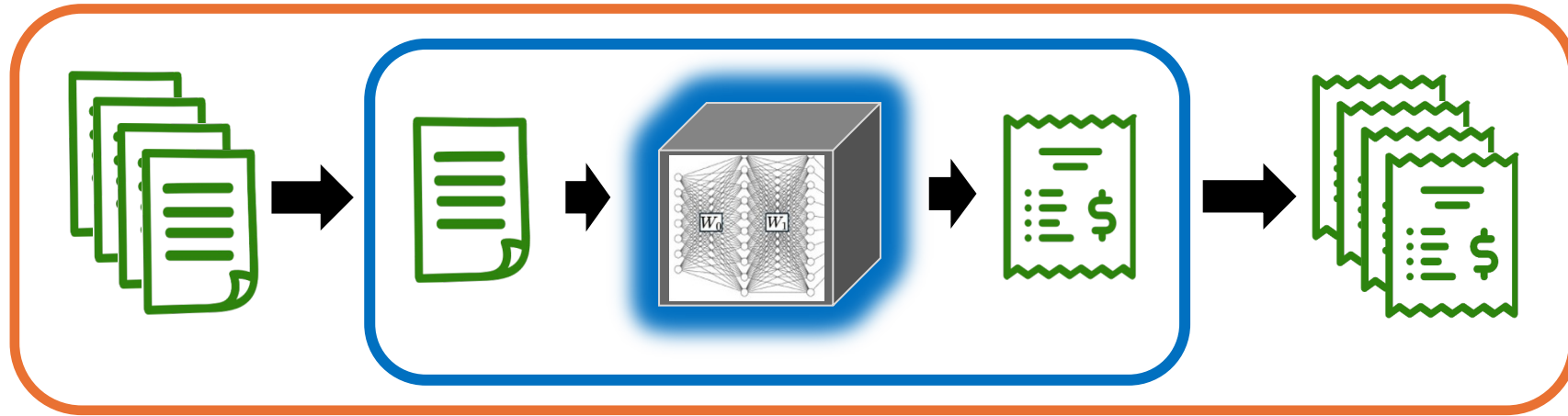
Is this True?

95 < 99%
Accuracy → 0% Work
Reduction

(Verifying CPT Codes is
No Easier than Coding
from Scratch)

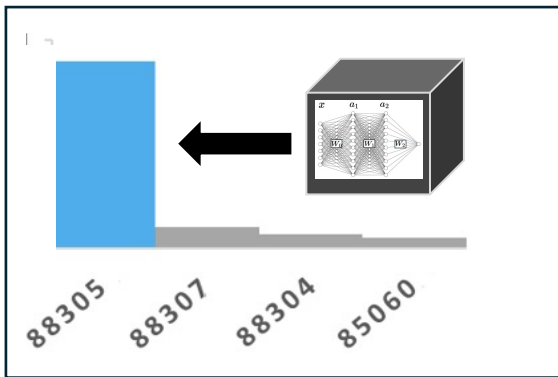


What if We Think Outside the Box?

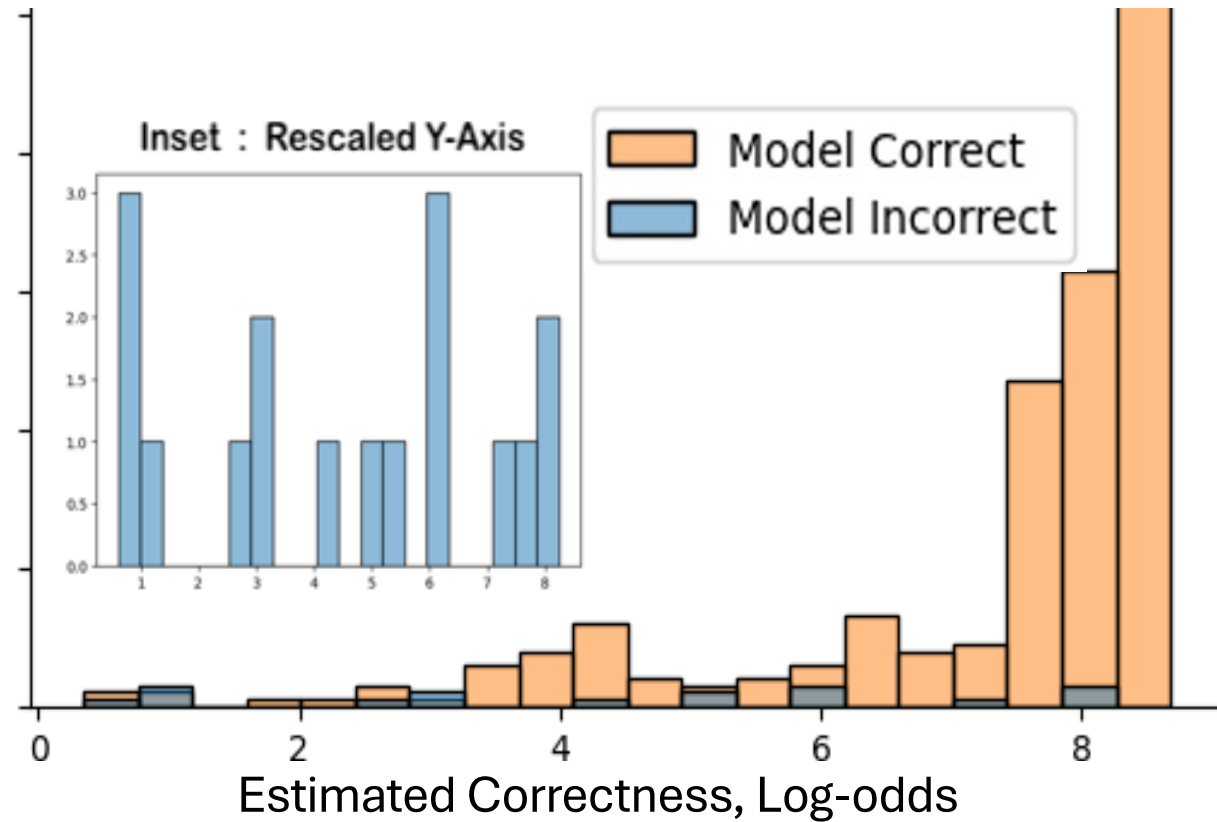


- *Are AI **Processes** Limited by AI **Models**?*

Result 2: The LLM Knows When it Doesn't Know



Magnitude Predictive of Correctness

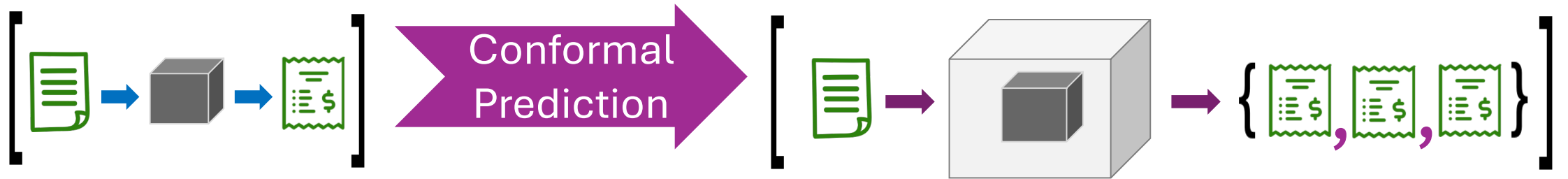


*→ Delegation Thresholds
can be Calibrated to
Improve Performance*

Conformal Prediction

An Established Technology

Any Model



Conformal Model

Result : Single Output

Performance : Variable

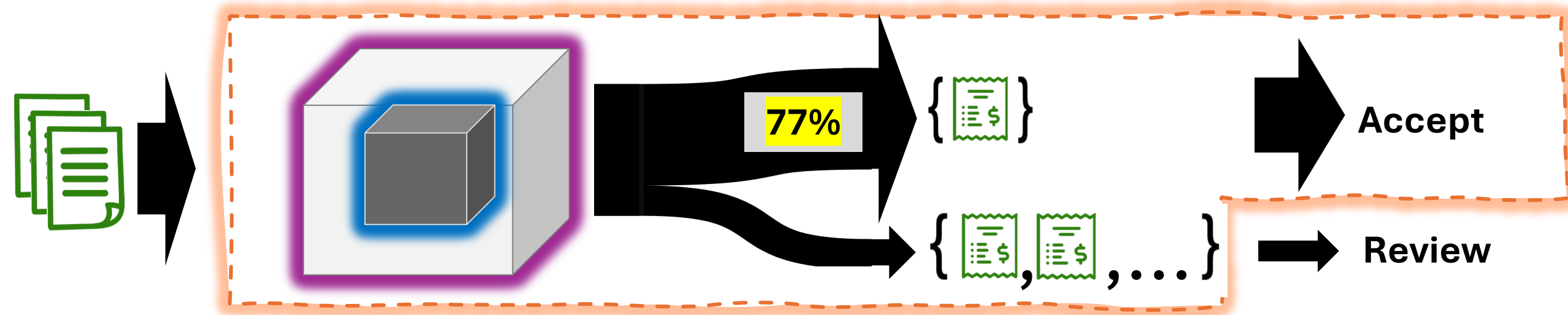
Deployment : “All Or Nothing”

Result : Confidence Set

Performance : User-Specified

Deployment : Can be Triaged

Application of Conformal Prediction to CPT Coding:



<i>Model</i>	
95 %	Accuracy
0 %	Abstain

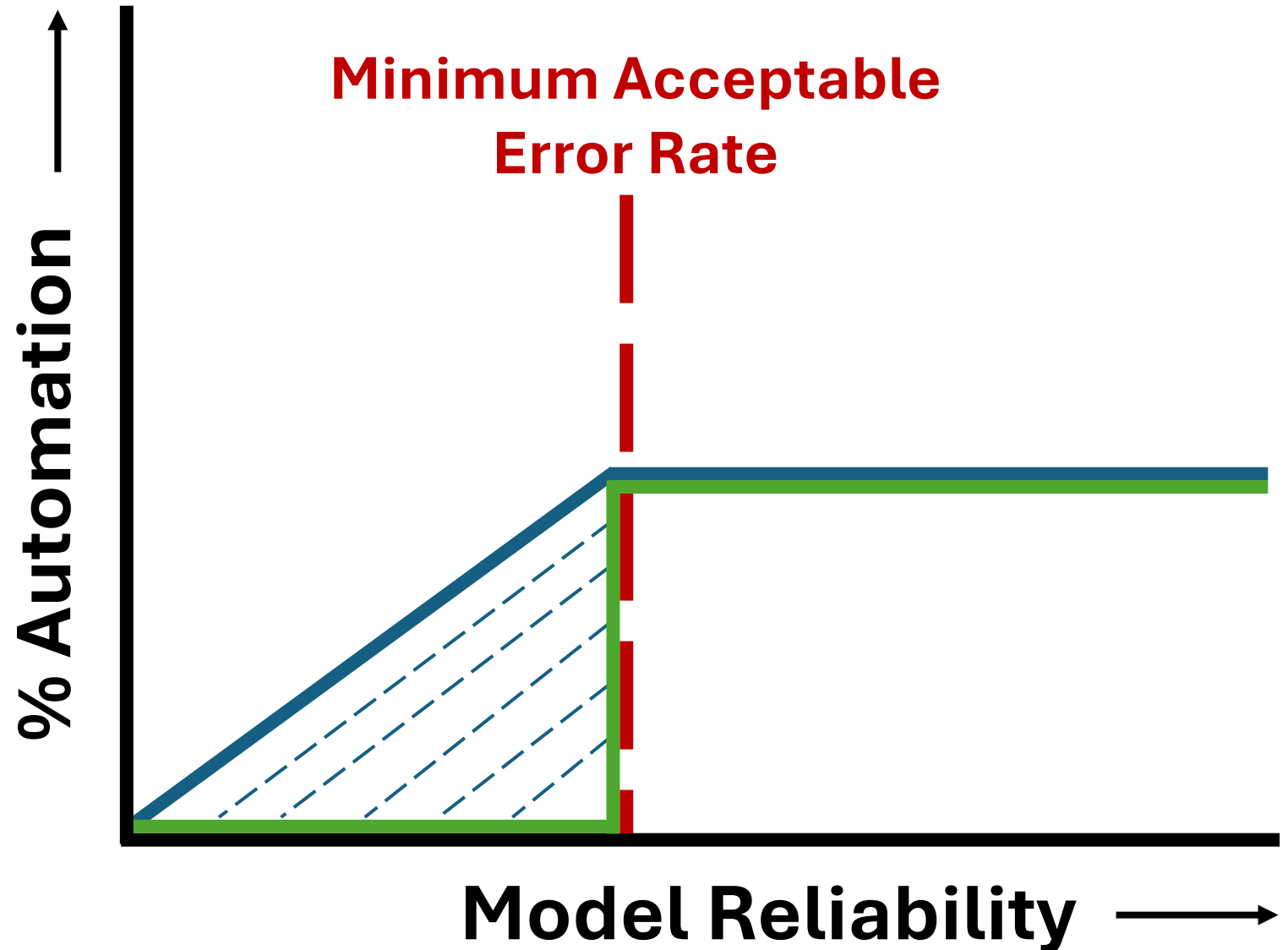
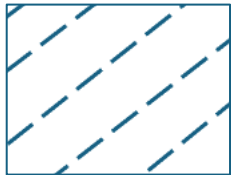


<i>Process</i>	
99 %	Accuracy
23 %	Abstain

Significance of This Work

Conventional Wisdom

Our Contribution

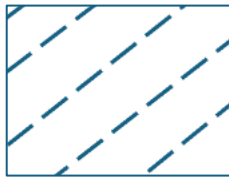


Significance of This Work

Conventional Wisdom

Our Contribution

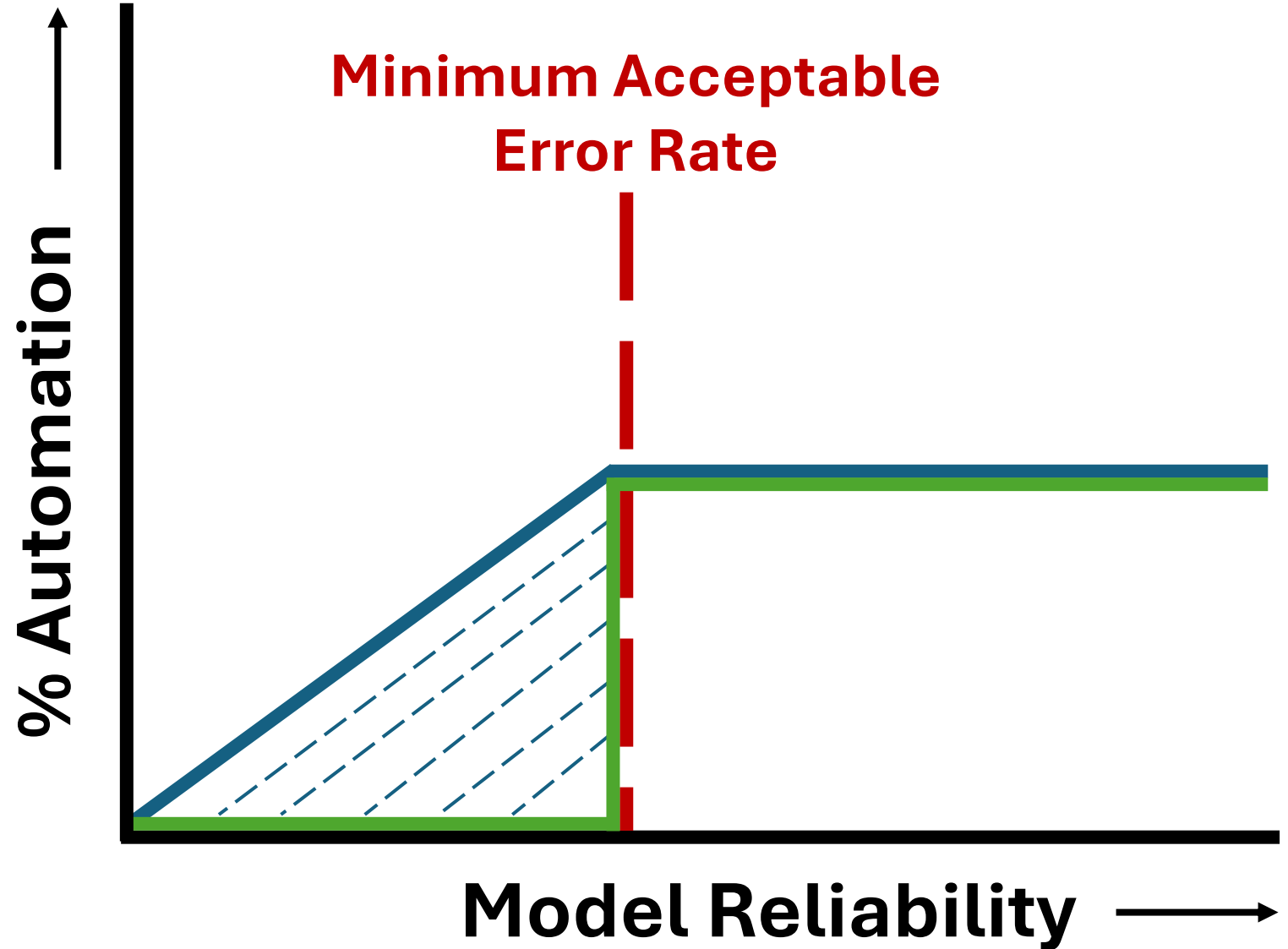
Graded
Automation



Nod to ACGME—

“Graded Responsibility”

*gradual increases in “trainee”
autonomy with improvement*



Acknowledgements and Thanks

Pathology Informatics

- Dr. Victor Brodsky

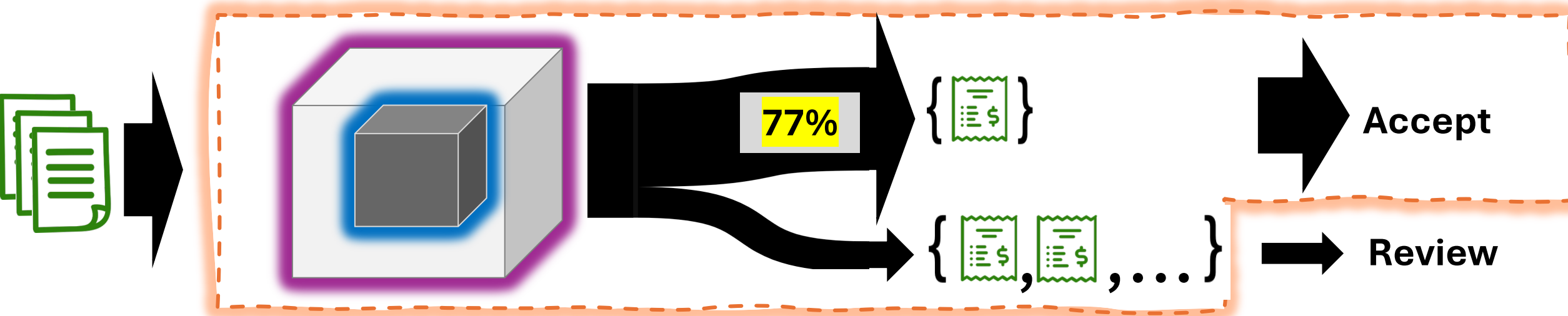
Medical Coding Team

- Amanda Martsolf
- Kelly Morgan



Follow-Up 1:

“So did you ever succeed at making your model better than 95%?”



<i>Model</i>	
95 %	Accuracy
0 %	Abstain

Conformal Prediction Application

<i>Process</i>	
99 %	Accuracy
23 %	Abstain

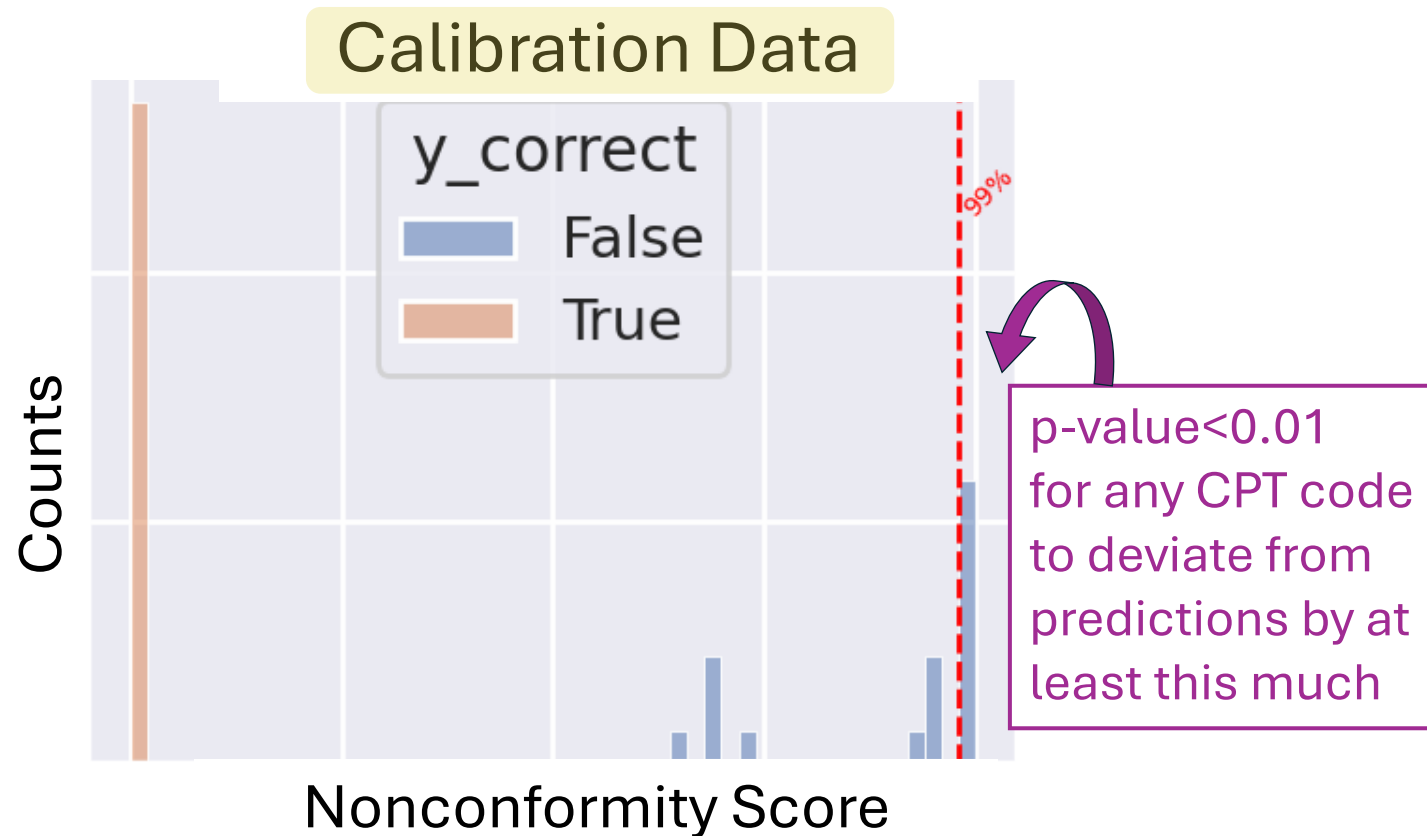
Follow-Up 2:

“What is the nature of the multi-element confidence set predictions?”

Case Review	A { III, V }	B { III, IV }
Situation	“gallbladder” & “Liver”	“... ‘Mitrofanoff stoma polyp’...”
Background	Merits Both Code III and IV	Merits either III or IV; depends on Polyp Site: Nasal(III) or GI (IV)
Interpretation	Appropriate Uncertainty: Report Contains Multiple Codes	Appropriate Uncertainty: Insufficient Evidence to Distinguish Further

Clinically Meaningful Uncertainty! (Irreducible)

A Reference Interval for Hallucination: An Interesting Perspective for the Pathologist



Automation is Not the End Goal

Graded
Automation

