

hfma<sup>®</sup>

massachusetts-rhode island chapter

25<sup>th</sup> Annual Revenue Cycle Conference  
Tailgates, Touchdowns, & Revenue Cycle Championships!  
It's time.....to Reignite!

**REVOLUTIONIZING MEDICAL  
CODING WITH AI:  
OVERCOMING CHALLENGES  
AND ACHIEVING ACCURACY**

Jay Aslam, PhD

January 2024

# OUTLINE

## Motivation

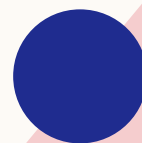
- Why focus on coding and charge capture?
- Why use AI?
- Governing principles

## Understanding the AI landscape

- AI, NLP, NLU, ML, DL, LLM/ChatGPT
- A gentle introduction to Machine Learning

## An ML-based medical coding automation solution

## Case-studies of real-world implementations



# WHY FOCUS ON CODING AND CHARGE CAPTURE?

## Coding is the most costly part of the RCM process

- Involving MDs does not make financial sense
- ~25% of RCM cost
- Coders are in high demand, highly compensated and in short supply
- Continuous training on new codes and regulations; certification and credentialing are basically required

## It is high impact and getting it wrong affects cost and revenue

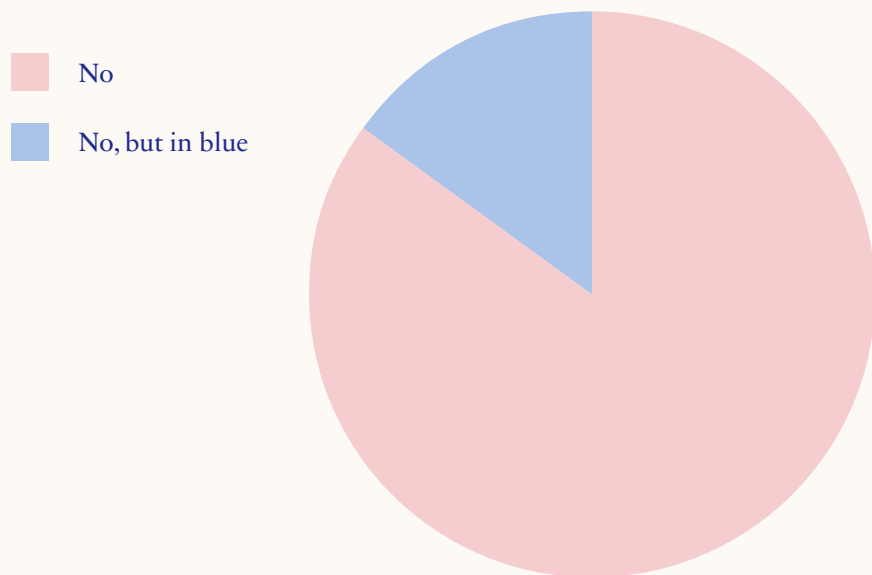
- Coding errors can lead to under-coding and lower revenue, and over-coding can lead to rejected claims, reprocessing or expensive penalties
- Coding is a part of the administrative burden experienced by providers, which results in less time with patients and lower yield

## It is large and complex with an enormous number of claims processed each year

- ~ 12 billion claims per year are processed by our healthcare system
- ~ 30% or 3.6 billion claims per year are related to physician-patient encounters
- Manually coding every claim is impractical; studies show 5% of billable services are missed altogether

# DATA FROM EVERY SINGLE MGH PHYSICIAN SURVEY

Do physicians enjoy, like, look forward to, or want to be involved in charge capture in any way, shape, or form?



I can't believe it! This is the 4th Y92.241 (hurt at the library, of course!) I've seen today.



# GOVERNING PRINCIPLES

## AT LEAST TWO LOGICAL APPROACHES TO TACKLING OUR PROBLEM:

### 1 Build a system and

- *train doctors* to document for that system (input-side) and/or
- *encode rules* to determine codes and make automation decisions from documentation (output-side)

#### Generally takes form as a rules-based expert system

- Cumbersome for coders/docs
- Brittle
- Difficult to maintain
- Difficult to determine accurate automation decisions

### 2 Build a system and

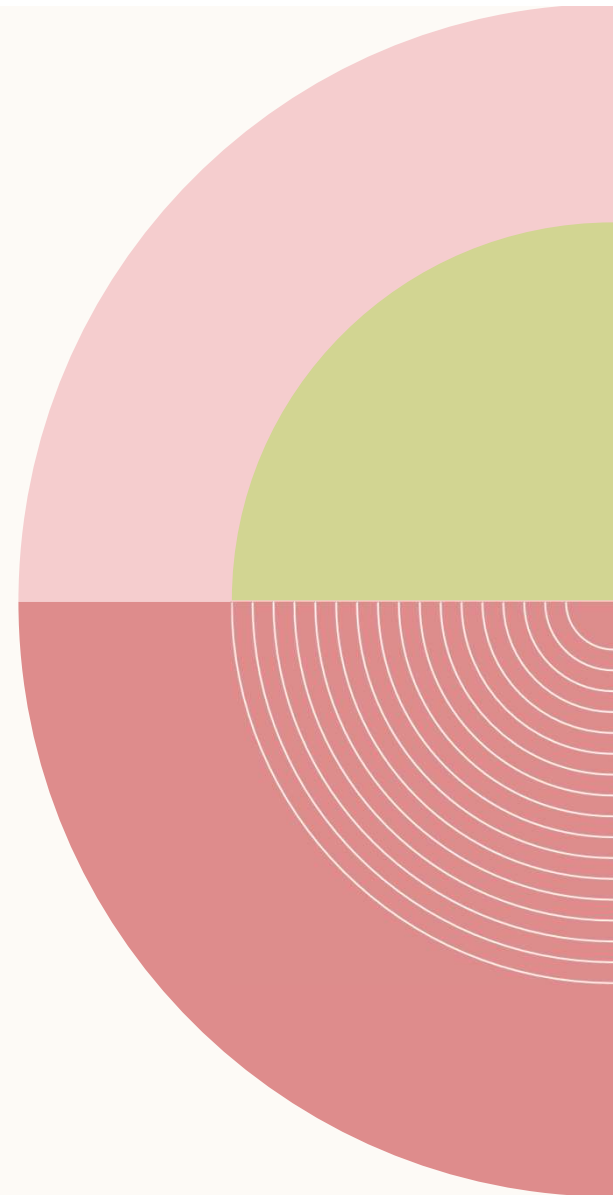
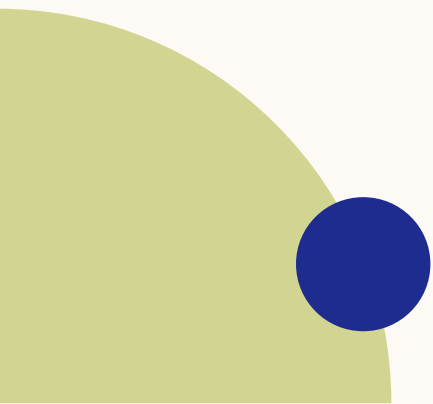
- *train that system* to understand doctors' documentation and
- *train that system* to automatically determine codes and make automation decisions

#### Generally takes form as an ML-based approach

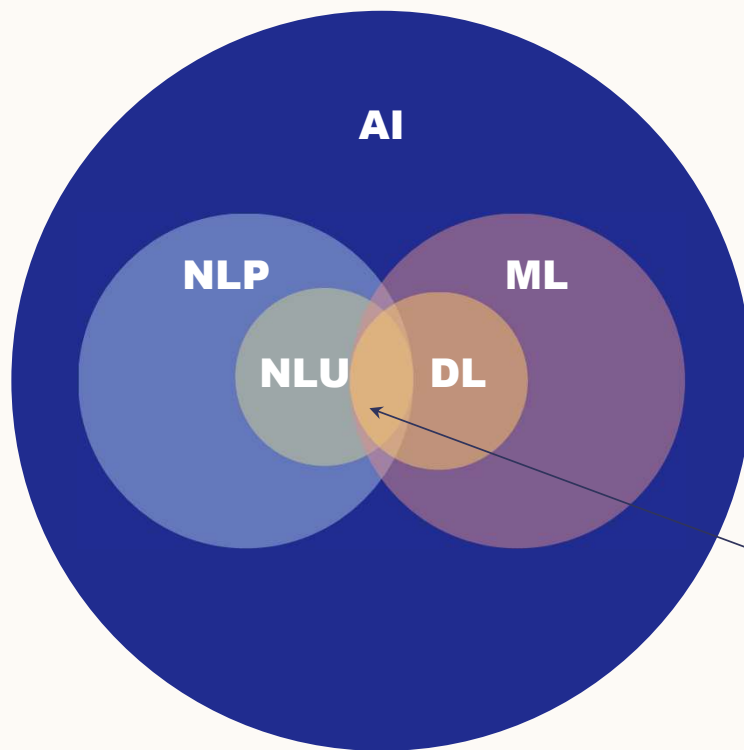
- Much less cumbersome
- Much more robust
- Much more accurate

# CMX TECHNOLOGY: GOVERNING PRINCIPLES

- We chose to go the modern ML-based route
- We had to devise a unique ML solution and invent many of its components
- But how does ML work and how can it be used to solve our task?
- And how is an ML-based solution different than an NLP or rules-based solution?



# THE AI LANDSCAPE



**AI:** Artificial Intelligence

**NLP:** Natural Language Processing

**NLU:** Natural Language Understanding

**ML:** Machine Learning

**DL:** Deep Learning

**LLM:** Large Language Models (e.g. ChatGPT)

# NLP VS. NLU VS. ML

NLP

## NATURAL LANGUAGE PROCESSING

“Pleural effusion” is a statistically interesting phrase

NLU

## NATURAL LANGUAGE UNDERSTANDING

“Pleural effusion” is a lung condition

ML

## MACHINE LEARNING

If pleural effusion with malignancy, code J91.0  
If pleural effusion in other conditions, code J91.8



# NLP VS. NLU VS. ML

NLP & NLU are largely about *extracting* and *assigning meaning* to words and phrases in the note.

- “pleural effusion” is a ”lung condition”

But you need rules to act on that information...

...what to do when you see the lung condition pleural effusion?

- In a traditional NLP/NLU-based system, you *manually craft* the rules
- In an ML-based system, you *automatically learn* the rules from data

**But how does ML do this?**



## ML INTRO

# MODEL GENERATION

An ML algorithm takes *labeled training data* as input and determines a *function of the features* that produces a *score* which is *correlated with the label*. This function is a *predictor*.

### CODING EXAMPLE

J91.0 score = 6.7 x [pleural effusion?] + 2.3 x [malignancy?] ...

J91.8 score = 7.5 x [pleural effusion?] - 1.3 x [malignancy?] ...

**High scores:** code is likely applicable

**Low scores:** code is likely not applicable

These predictors essentially encode *rules*, and these rules have been *automatically learned from data*.

But how do we make the output of these predictors *actionable*?



## ML INTRO

# CALIBRATION

Calibration takes these predictor scores and maps them to calibrated confidences, e.g.,

- Training cases with J91.0 score in range [10, 12] are 90% often coded with J91.0
- Training cases with J91.0 score in range [2, 4] are only 60% often coded with J91.0
- Training cases with J91.0 score in range [-8, -6] are merely 10% coded with J91.0.



## ML INTRO

# AUTOMATION

Calibrated confidences can be used for *automation*.

- J91.0 scores in range [10, 12] are 90% coded with J91.0
- J91.0 scores in range [2, 4] are only 60% coded with J91.0
- J91.0 scores in range [-8, -6] are merely 10% likely to be coded with J91.0.

If I want to automatically code cases with J91.0, but with at most 10% errors, then I can safely do so as long as my J91.0 score is in the range [10, 12] (or presumably higher).

**We build *machine learning* predictors and use calibrated confidences to make *automation decisions*.**

Thresholding these calibrated confidences allow us to intelligently make *automation vs. quality tradeoffs* to satisfy customer needs.



## AN ML-BASED SOLUTION

We solve the prediction problem for autonomous code prediction using machine learning.

- We use NLP/NLU/LLM to *extract information* from the notes
- We make that information *actionable* by *rules automatically learned via ML* from data
- Our ML is *glass-box* in the sense that every prediction can be traced back to the information in the note that caused that prediction to be made
- Every prediction has an associated *interpretable calibrated confidence* that enables user-defined automation at quality targets
  - *automation vs. quality trade-off*



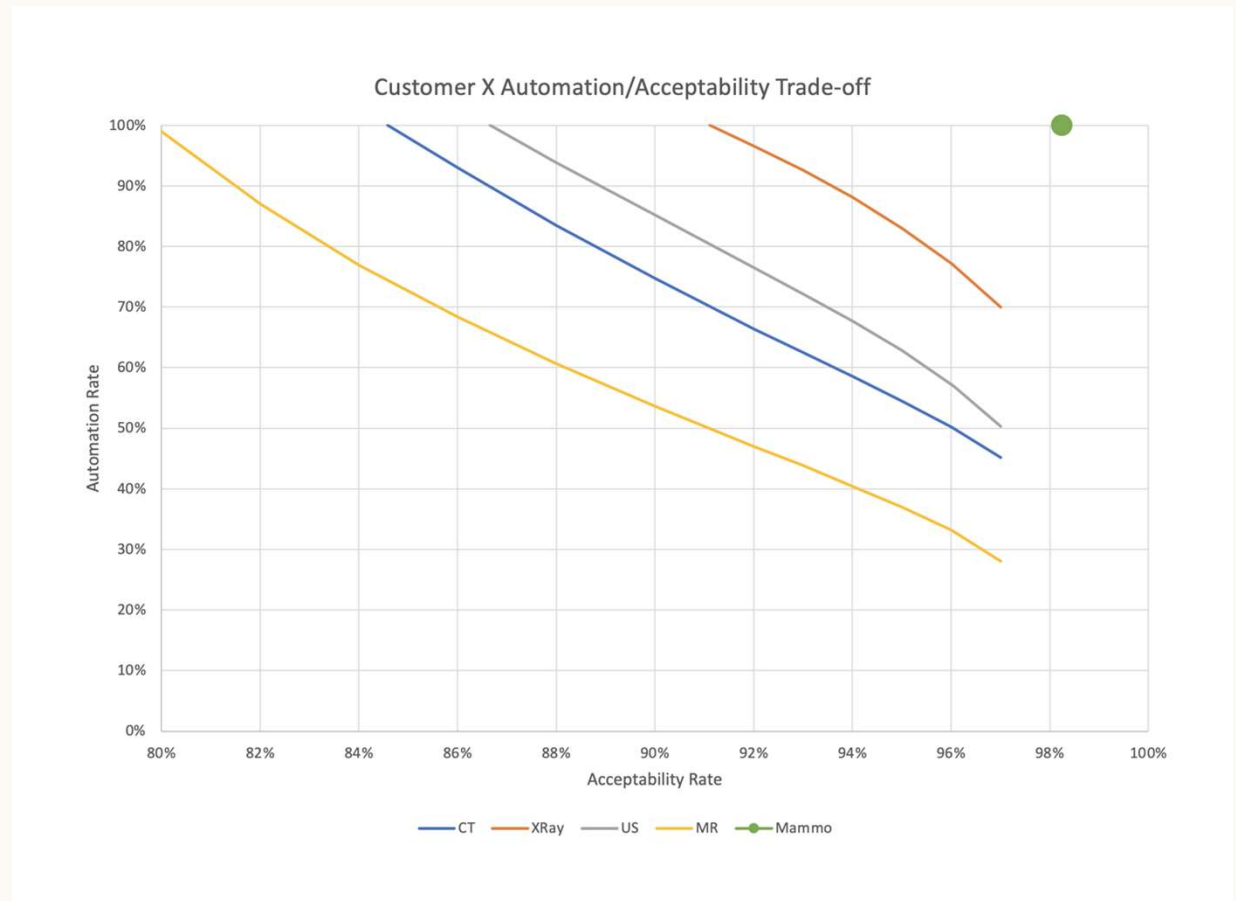
## ML INTRO

# CALIBRATION

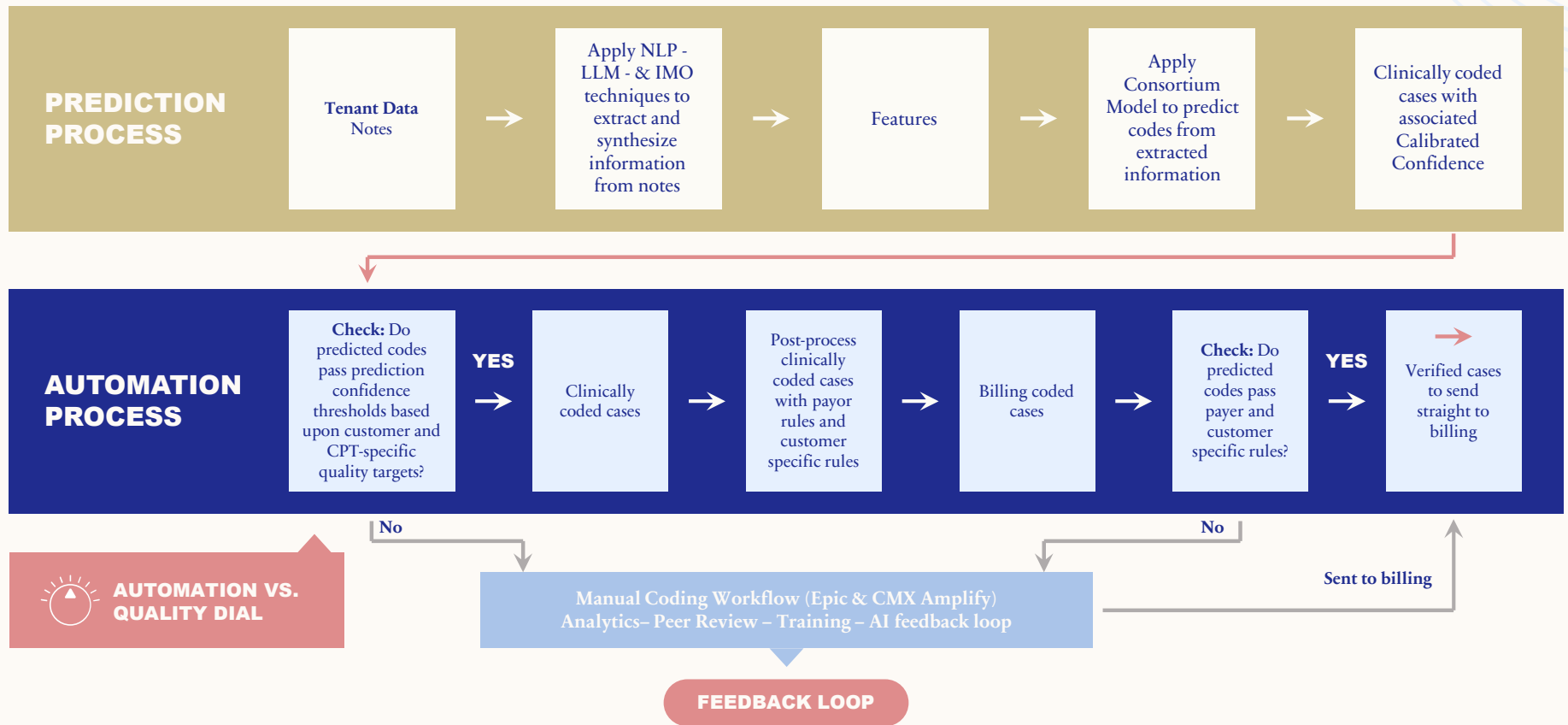
Calibration takes these predictor scores and maps them to calibrated confidences, e.g.,

- J91.0 scores in range [10, 12] are 90% coded with J91.0
- J91.0 scores in range [2, 4] are only 60% coded with J91.0
- J91.0 scores in range [-8, -6] are merely 10% likely to be coded with J91.0.

# AUTOMATION AT QUALITY



# A COMPLETE ML-BASED AUTOMATION SOLUTION



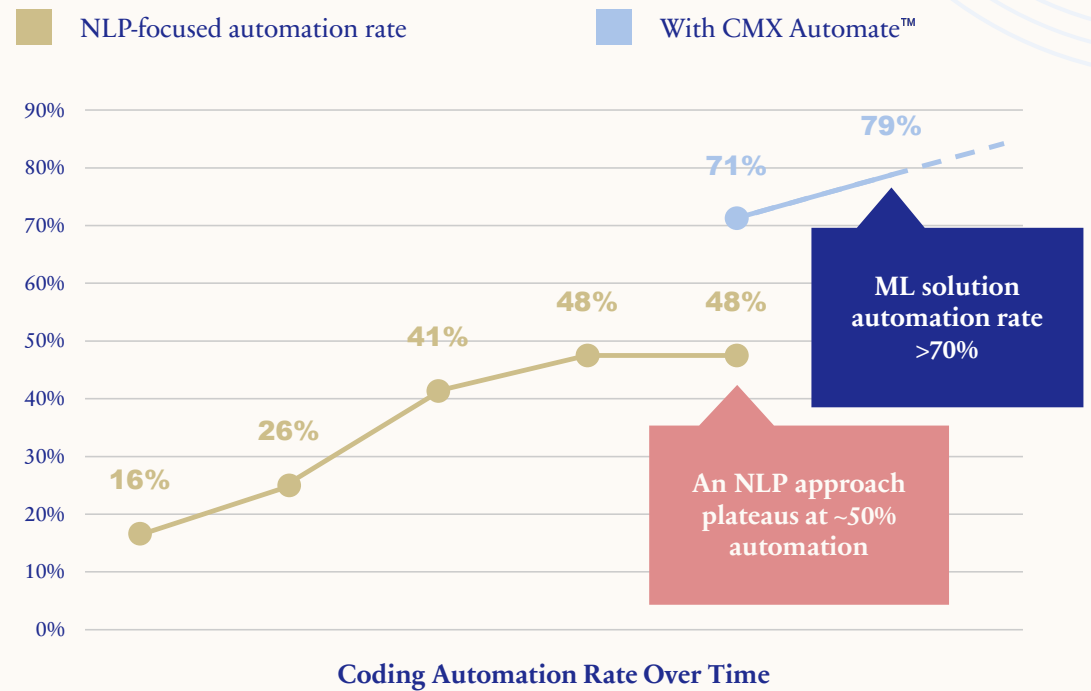


Revolutionizing Medical Coding with AI

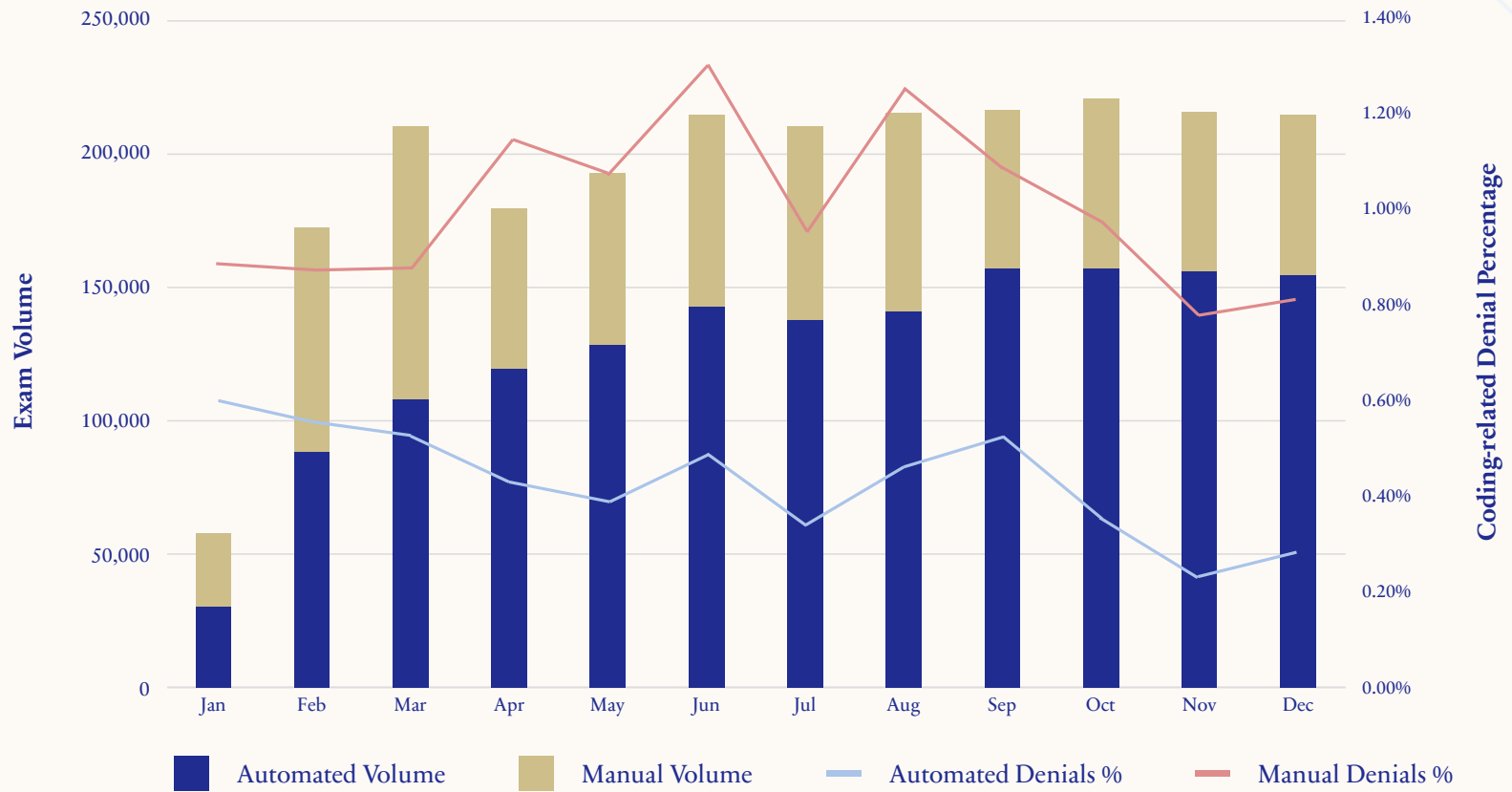
# RECENT RESULTS OF AUTOMATION

The CMX Autonomous-Coding platform outperforms the NLP-centric solutions by combining the power of AI: NLP, NLU, DL, ML.

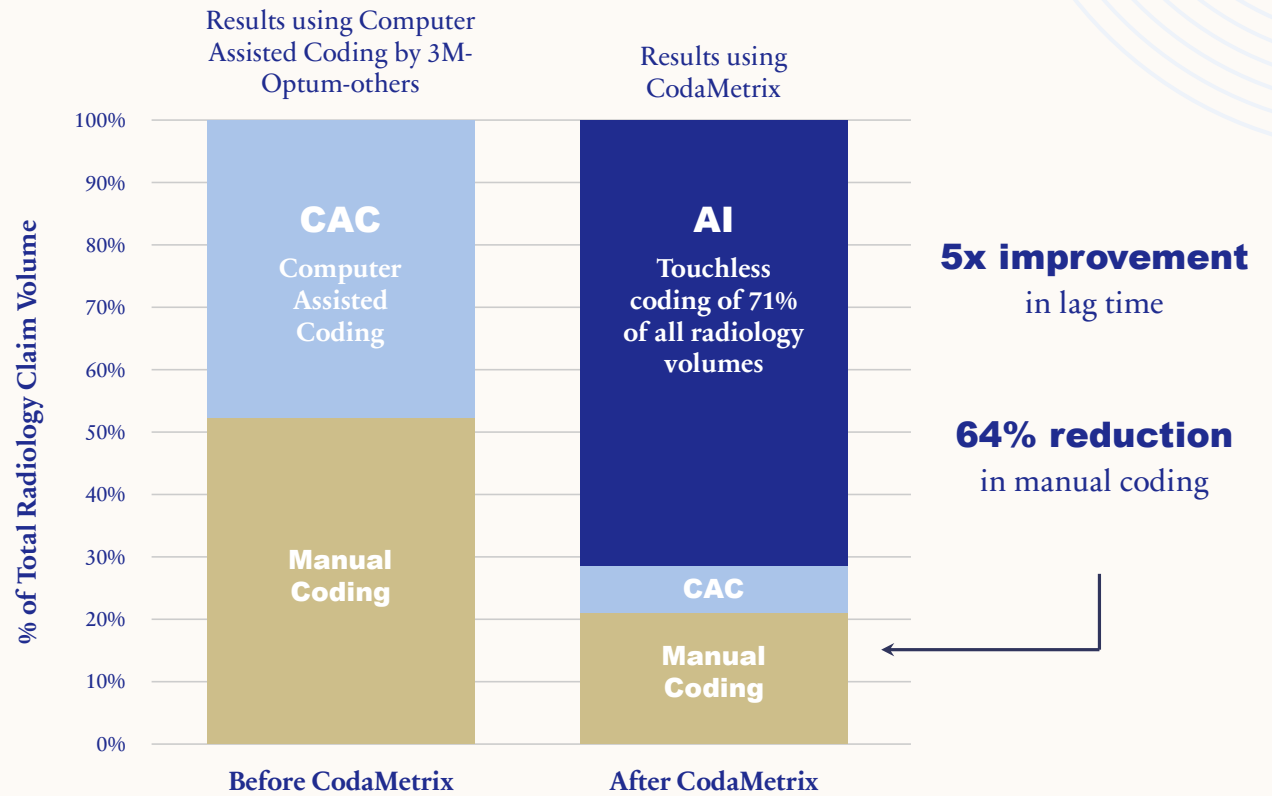
## CU MEDICINE IMPLEMENTATION OF CODAMETRIX



# MASS GENERAL BRIGHAM'S CY2021 TOTAL ACCESSIONS (~2.1M)



# CU MEDICINE CMX RESULTS



The background features a dark blue horizontal band at the top. Below it, a large white circle overlaps with a large light pink circle. The pink circle contains several thin, white, concentric curved lines. The word "QUESTIONS" is centered in the white area.

# **QUESTIONS**

**THANK YOU!**  
**CODAMETRIX**

# WHAT ABOUT LLMS (E.G. CHATGPT)?

Language Models are designed to understand human language.

- A practical example we see every day is *auto-completion*: “Thank you very \_\_\_”

Large Language Models (LLMs) are trained on *enormous* amounts of data, and they understand human language, computer programming, art & images, and a host of other things very well.

- LLMs, out-of-the-box, are not particularly good at *medical coding* (“hallucinations”).

However, LLMs are extremely good at *information extraction*, and with further specific *pre-training* and *fine-tuning*, they will likely perform quite well in (parts of) medical coding.

## LLM EXAMPLE

# FINDINGS SECTION OF A COLONOSCOPY

The perianal and digital rectal examinations were normal. A moderate amount of stool was found in the entire colon, interfering with visualization. Lavage of the area was performed using 1 liter of sterile water, resulting in clearance with good visualization. Many small and large-mouthed diverticula were found in the sigmoid colon.

A 1 mm polyp was found in the cecum. The polyp was sessile. The polyp was removed with a jumbo cold forceps. Resection and retrieval were complete. Verification of patient identification for the specimen was done.

A 8 mm polyp was found in the ascending colon. The polyp was sessile. The polyp was removed with a saline injection-lift technique using a hot snare. Resection and retrieval were complete. Verification of patient identification for the specimen was done. Estimated blood loss was minimal.

A 4 mm polyp was found in the descending colon. The polyp was sessile. The polyp was removed with a cold snare. Resection and retrieval were complete. Verification of patient identification for the specimen was done. Internal hemorrhoids were found during retroflexion.

## LLM EXAMPLE

# FINDINGS SECTION OF A COLONOSCOPY

The perianal and digital rectal examinations were normal. A moderate amount of stool was found in the entire colon, interfering with visualization. Lavage of the area was performed using 1 liter of sterile water, resulting in clearance with good visualization. Many small and large-mouthed diverticula were found in the sigmoid colon.

A 1 mm polyp was found in the cecum. The polyp was sessile. The polyp was removed with a jumbo cold forceps. Resection and retrieval were complete. Verification of patient identification for the specimen was done.

A 8 mm polyp was found in the ascending colon. The polyp was sessile. The polyp was removed with a saline injection-lift technique using a hot snare. Resection and retrieval were complete. Verification of patient identification for the specimen was done. Estimated blood loss was minimal.

A 4 mm polyp was found in the descending colon. The polyp was sessile. The polyp was removed with a cold snare. Resection and retrieval were complete. Verification of patient identification for the specimen was done. Internal hemorrhoids were found during retroflexion.





## LLM EXAMPLE

# PROMPT ENGINEERING FOR CHATGPT

In the following medical report, find all mentions of polyp.

For each mention, return size of the polyp, body location of the polyp, the technique used to remove the polyp, and the original text span from which the answer is extracted, and the character ranges for the text span.

The answer should be in a json format. The keys in the json output should be "size", "body\_location", "removal\_technique", "text\_span" and "character\_span".

## LLM EXAMPLE

# CHATGPT RESULT

```
{
  "mentions": [
    {
      "size": "1 mm",
      "body_location": "cecum",
      "removal_technique": "jumbo cold forceps",
      "text_span": "A 1 mm polyp was found in the cecum. The polyp was sessile. The polyp was removed with a jumbo cold forceps.",
      "character_span": {
        "start": 817,
        "end": 920
      }
    },
    {
      "size": "8 mm",
      "body_location": "ascending colon",
      "removal_technique": "saline injection-lift and hot snare",
      "text_span": "A 8 mm polyp was found in the ascending colon. The polyp was sessile. The polyp was removed with a saline injection-lift technique using a hot snare. Resection and retrieval were complete. Verification of patient identification for the specimen was done. Estimated blood loss was minimal.",
      "character_span": {
        "start": 946,
        "end": 1138
      }
    },
    {
      "size": "4 mm",
      "body_location": "descending colon",
      "removal_technique": "cold snare",
      "text_span": "A 4 mm polyp was found in the descending colon. The polyp was sessile. The polyp was removed with a cold snare. Resection and retrieval were complete. Verification of patient identification for the specimen was done.",
      "character_span": {
        "start": 1164,
        "end": 1300
      }
    }
  ]
}
```