

Analyzing Generative AI Proficiency for ICD-10 Code Generation and Interpretation



Tulane

SCHOOL OF PUBLIC HEALTH
AND TROPICAL MEDICINE

Debra Winberg MSc MA, Dennis Xuan MSPH, Tiange Tang, MPH, Shelby Olin, MPH, Rong Rong, MPH,
Lizheng Shi, PhD

Tulane University Celia Scott Weathershead School of Public Health and Tropical Medicine, New Orleans, LA

INTRODUCTION

- ChatGPT is a **natural language processing tool** which creates human-like conversations, responds to questions, and creates written content.
- Opportunity present to utilize ChatGPT for **innovative research methodologies**
- ChatGPT is trained on published material, allowing it to find and parse relevant information, making it a useful tool for **identifying ICD codes**.
- As more academics and clinicians use the tool, gauging the accuracy of the information gathered by this automation becomes important

AIM

This study aimed to compare the efficacy of common, accessible Generative AI models in identifying ICD-10 codes.

METHODS

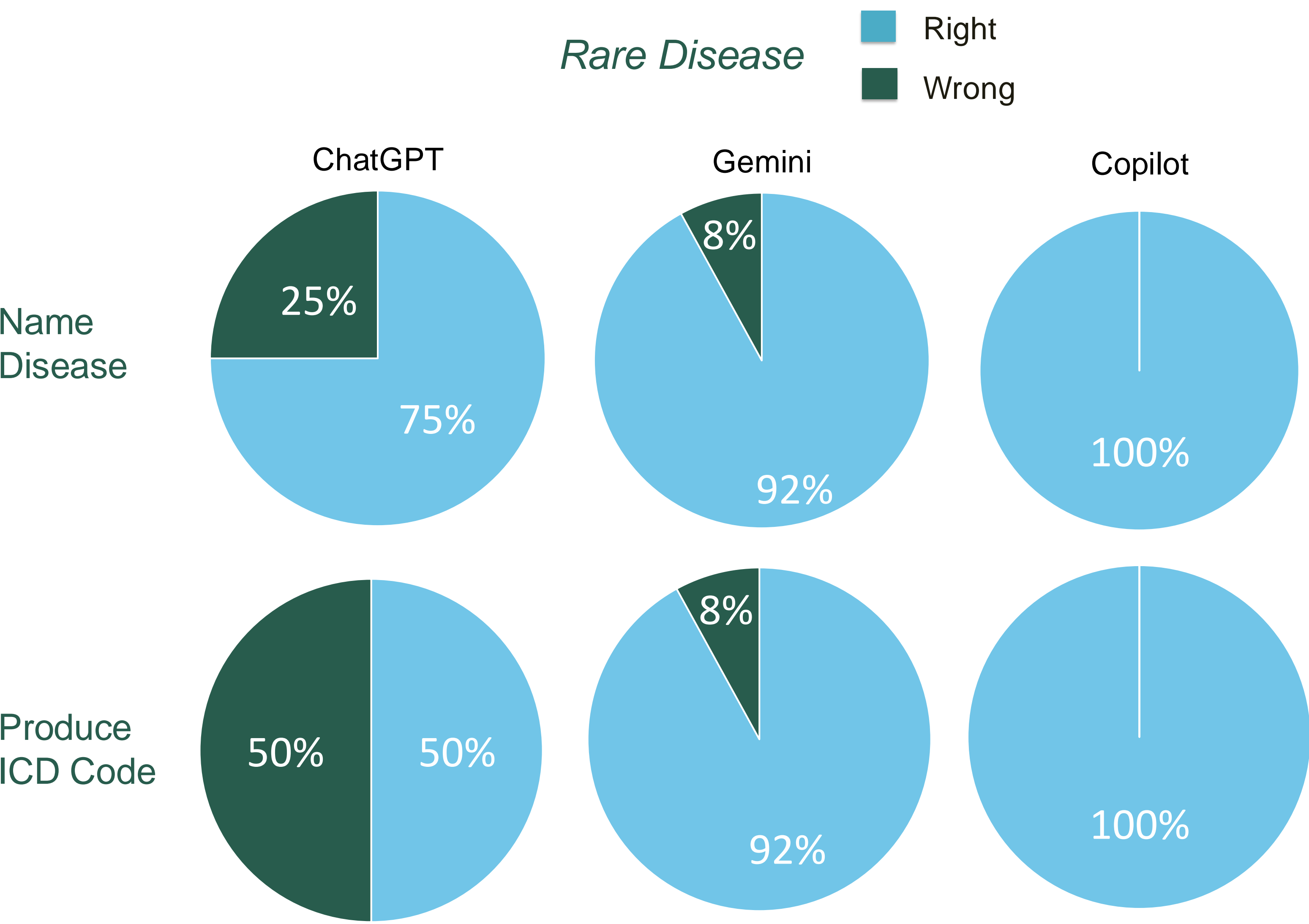
- Generative AI Models**
 - ChatGPT 4.0
 - Google Gemini
 - Microsoft Copilot
- ICD Codes and Diseases**

Generated a list of common diseases (diabetes, hypertension, hyperlipidemia, obesity, etc) and rare diseases(Canavan disease, yellow nail syndrome, Maffucci syndrome, etc). Found the associated ICD codes through literature review or ICD code books
- Generative AI ICD Codes**

For each disease separately asked (after clearing memory)

 - What is the ICD-10 code for XXX
 - What is the disease associated with XXX ICD-10 code?

RESULTS



Microsoft Copilot was able to accurately provide the ICD code or the disease name/family for all 10 diseases. Gemini was right the majority of the time, but ChatGPT performed the poorest. ChatGPT performed worse when asked to produce ICD-10 codes.

Common Disease		
Example GPT Response	Example Gemini Response	Example Copilot Response
The ICD-10 code for obesity is E66. This category includes several subtypes based on the specific characteristics of obesity.	The ICD-10 code for fatigue depends on the context and whether there's a known cause.	The specific ICD-10 code for chronic obstructive pulmonary disease (COPD) with acute exacerbation is J44.11

All three programs were able to get the correct ICD code and disease name, however it usually **only able to give broad categories for all three programs** and not the specific codes. In the cases where it did give high levels of specificity, it **did not include every code**

Notable Observations or Responses		
ChatGPT	Gemini	Copilot
At times it gave the incorrect ICD code, but gave information for the correct disease	Responded at times with: "I am a language model and don't have the capacity to help with that"	Responded at times with "Please consult with a healthcare professional for a comprehensive understanding or diagnosis."
Often would only provide ICD codes to higher levels of specificity (e.g. would give E11.2 not E11.21)	Provides pictures and disease information	Explained that "These codes are valid for the submission of HIPAA-covered transactions from October 01, 2023 through September 30, 2024"
Did not accurately give responses for ICD codes in the 2023-2024 version	Became more accurate overtime	Slowest at providing response
	For common diseases, sometimes admitted was not giving extensive code list, other times did not	Provided consistent citations

DISCUSSION

- Generative AI is more useful in giving the disease associated with an ICD code than at producing an ICD code
- For generative AI to be useful, it needs to be trained on the latest ICD code books and tested on consistency
- Additionally, generative AI needs to be able to provide all ICD codes associated with a disease (e.g. not only E11.1 for type 2 diabetes) or make it clear that it is not an extensive list
- It was helpful that copilot provided citations

CONCLUSIONS

Generative AI models are a good starting point for searching and identifying ICD codes. However, researchers need to recognize which version of the ICD book a model was trained on and the specificity of the target ICD code.

LIMITATIONS

- There are multiple ways to code rare diseases
- Used ChatGPT more which could have impacted its history and ability to predict future results
- Used same GPT account, but different copilot and Gemini accounts

Contact

Contact Debra Winberg at dwinberg@tulane.edu for more information