

Tech Talk

Unlock the Possibilities of Generative AI for Health Communication Research

September 19, 2023

Robert Chew, MS

Mario Navarro, PhD

Stephanie McInnis, MA

Annice Kim, PhD, MPH





**New York City public schools
ban OpenAI's ChatGPT**

**ChatGPT passes exams from law and
business schools**

**Which is better — ChatGPT or a travel
agent? Here's our pick**

**Recruiters warn against overuse of AI tools such
as ChatGPT to write resumes, cover letters**

**ChatGPT may be coming for our jobs. Here are the 10 roles that AI is
most likely to replace.**



What is ChatGPT?

- “**Chat Generative Pre-trained Transformer**”
- A generative artificial intelligence (AI) chatbot that generates human-like text responses to questions, within seconds.
- ChatGPT is a type of **Large Language Model (LLM)** created by OpenAI.
- Though ChatGPT is the most famous, there are many LLMs, both publicly available and commercial, that are designed to generate text based on user instructions (prompts).



Webinar Overview

Potential benefits, opportunities, and use cases of generative AI tools for getting research insights faster

The strengths and limitations of ChatGPT compared to gold standard approaches for coding of unstructured text and generating plain language summaries

How federal health agencies are approaching the use of generative AI for communication research

Key areas of future research and consideration



Today's Presenters



Robert Chew

Senior Research Data Scientist
RTI International



Mario Navarro

Lead Health Scientist
U.S. Food and Drug Administration



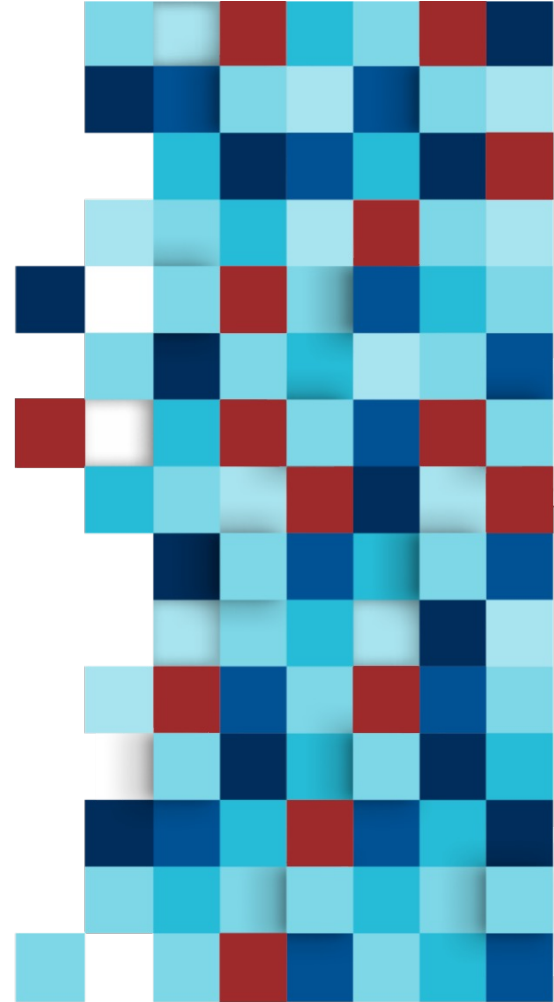
Stephanie McInnis

UX Researcher
RTI International

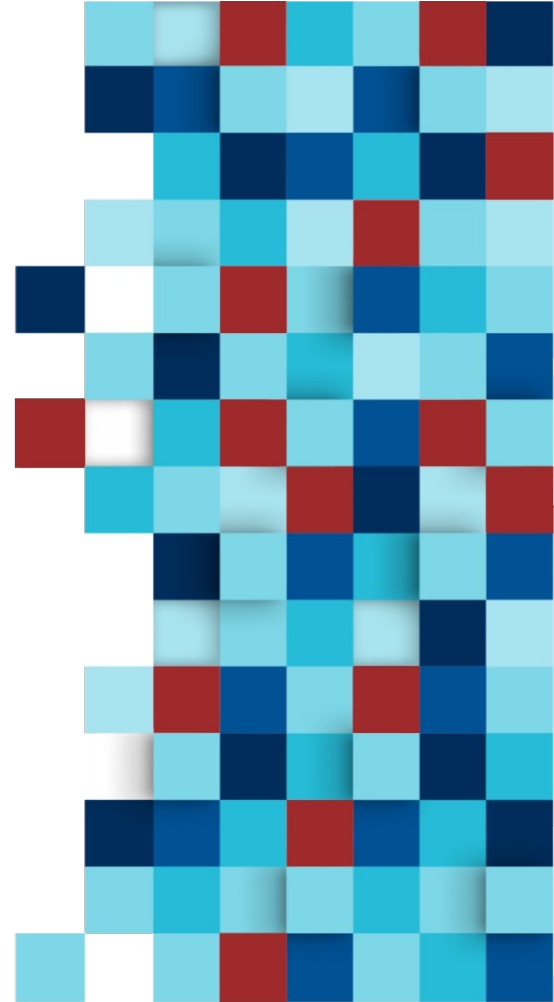
Data entered in generative AI interactions and models will be retained and could be incorporated into the model in the future. The data shared with these products can be linked to individual users. Non-public data should not be entered, or made available to, third-party websites that have not been approved.

It has been shown that generative AI responses can be biased, limited, and misleading. There are reports of generative AI tools “hallucinating” work products, creating falsehoods, and being manipulated with simple tricks. Text and images produced by generative AI may also include previously copyrighted materials without acknowledgment and are thus open to charges of plagiarism and intellectual property theft. In addition, users need to be extremely cautious when prompting generative AI for software code or other machine-readable instructions, as there may be security vulnerabilities and other issues with the code. Code or scripts should not be used or executed in a sensitive network environment.

Current case studies discussed in this webinar are exploratory only and should not be used for any other purpose.



Using LLMs for Qualitative Coding





Background

Qualitative coding is a popular health communication research method for analyzing text.

However, coding can be **slow, expensive, and error prone.**



Collecting and Analyzing Qualitative Data

[Print](#)

Brent Wolff, Frank Mahoney, Anna Leena Lohiniva, and Melissa Corkum

Introduction

Qualitative research methods are a key component of field epidemiologic investigations because they can provide insight into the perceptions, values, opinions, and community norms where investigations are being conducted (1,2). Open-ended inquiry methods, the mainstay of qualitative interview techniques, are essential in formative research for exploring contextual factors and rationales for risk behaviors that do not fit neatly into predefined categories. For example, during the

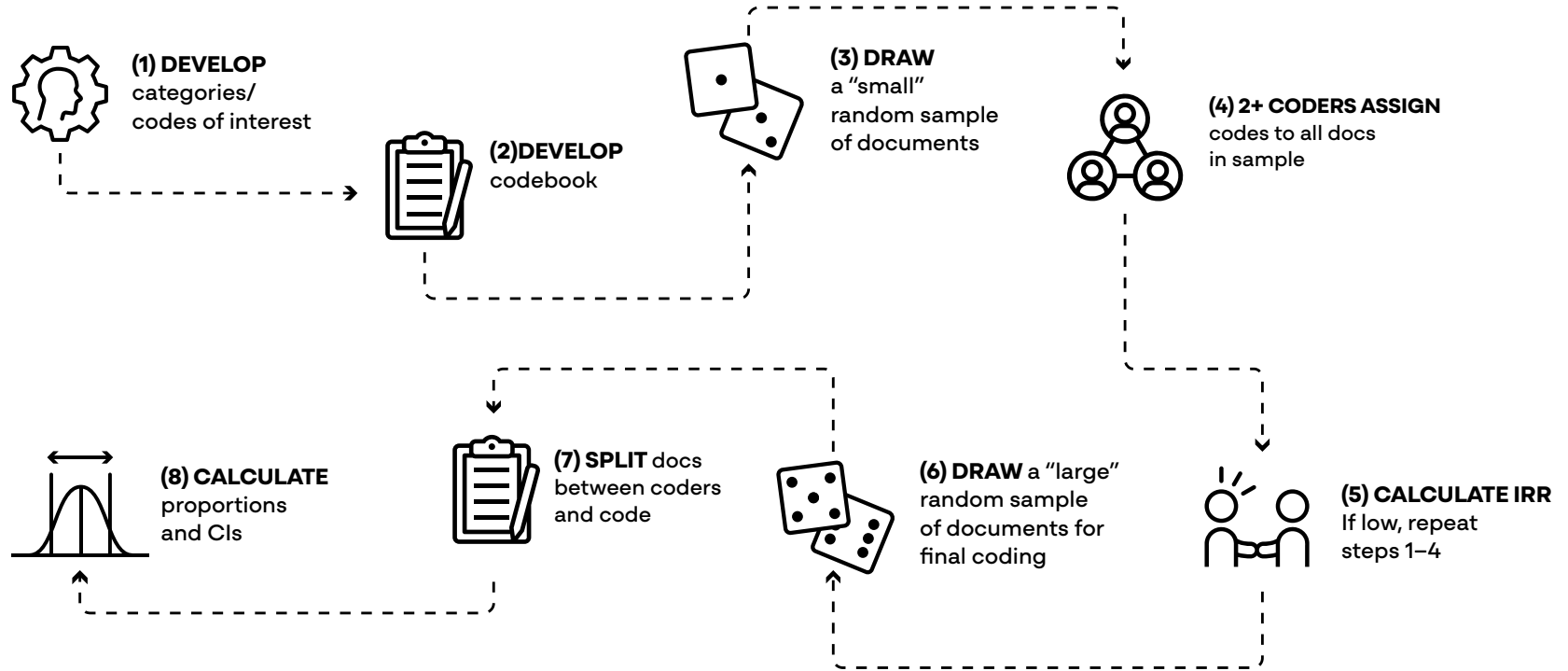
On This Page

[Choosing When to Apply Qualitative Methods](#)

[Commonly Used Qualitative Methods in Field Investigations](#)

<https://www.cdc.gov/eis/field-epi-manual/chapters/Qualitative-Data.html>

Deductive Coding





Research Question

How well does ChatGPT perform deductive coding compared to humans?

1. Inter-rater Reliability (IRR)
2. Coding Time



Publicly Available Datasets

Data Set	Doc Type	Mutually Exclusive	# Codes	# Docs	Notes
Trump Tweets	Tweets	No	13	2,083	Codebook written informally with short descriptions
Contrarian Claims	Blog Posts	Yes	28	2,904	Mutually exclusive, hierarchical code set. Codes nuanced and may have definitions with conceptual overlap
BBC News	News Articles	Yes	5	2,225	No formal codebook, only class names (e.g., business)
Ukraine Water Problems	Water Quality Reports	No	5	100	Brief codebook, but technically complex classes

Current case studies discussed in this webinar are exploratory only and should not be used for any other purpose.



Example Prompt

You are a qualitative coder who is annotating news stories. To code this text, do the following:

- First, read the codebook and the text.
- Next, decide which code is most applicable and explain your reasoning for the coding decision.
- Finally, print the most applicable code and your reason for the coding decision.

Use the following format:

Codebook:

```
---  
{codebook here} ← Coding instructions  
---
```

Text:

```
---  
{text here} ← Text document  
---
```

Code:

```
---  
business, entertainment, politics, sport, or tech  
---
```

```
Code: ← Coding decision and reason for decision
```

Human-Human vs. Human Model Agreement

Human-Human Agreement



Published
Data



Our Coded
Data

Human-Model Agreement



Published
Data



ChatGPT
Predictions

Agreement Metric

Gwet's AC1

Results: Reliability

ChatGPT often coded at levels of agreement comparable to humans

Table 7: Summary Benchmark Results Across Data Sets

Dataset	Code	Gwet's AC1		Tests of Randomness (p-value)
		Human-Human	Human-Model	
Trump Tweets	HSTG	0.96	0.18	0.19
Trump Tweets	ATSN	1.00	0.58	0.92
Trump Tweets	CRIT	0.73	0.76	0.00
Trump Tweets	MEDI	1.00	0.96	0.00
Trump Tweets	FAMY	0.97	0.96	0.00
Trump Tweets	PLCE	1.00	0.98	0.00
Trump Tweets	MAGA	0.99	0.98	0.00
Trump Tweets	CAPT	0.93	0.36	0.76
Trump Tweets	INDV	0.79	0.50	0.19
Trump Tweets	MARG	0.97	0.94	0.00
Trump Tweets	INTN	0.86	0.81	0.00
Trump Tweets	PRTY	0.81	0.76	0.00
Trump Tweets	IMMG	0.99	0.97	0.00
Ukraine Water	env_problems	0.23	0.64	0.62
Ukraine Water	pollution	0.59	0.55	0.62
Ukraine Water	treatment	0.84	0.88	0.00
Ukraine Water	climate	0.97	0.87	0.00
Ukraine Water	biomonitoring	0.51	0.86	0.00
BBC News	All	0.76	0.85	0.00
Contrarian Claims	All	0.65	0.59	0.00

Current case studies discussed in this webinar are exploratory only and should not be used for any other purpose.

Results: Reliability

Table 7: Summary Benchmark Results Across Data Sets

Dataset	Code	Gwet's AC1		Tests of Randomness (p-value)
		Human-Human	Human-Model	
Trump Tweets	HSTG	0.96	0.18	0.19
Trump Tweets	ATSN	1.00	0.58	0.92
Trump Tweets	CRIT	0.73	0.76	0.00
Trump Tweets	MEDI	1.00	0.96	0.00
Trump Tweets	FAMY	0.97	0.96	0.00
Trump Tweets	PLCE	1.00	0.98	0.00
Trump Tweets	MAGA	0.99	0.98	0.00
Trump Tweets	CAPT	0.93	0.36	0.76
Trump Tweets	INDV	0.79	0.50	0.19
Trump Tweets	MARG	0.97	0.94	0.00
Trump Tweets	INTN	0.86	0.81	0.00
Trump Tweets	PRTY	0.81	0.76	0.00
Trump Tweets	IMMG	0.99	0.97	0.00
Ukraine Water	env_problems	0.23	0.64	0.62
Ukraine Water	pollution	0.59	0.55	0.62
Ukraine Water	treatment	0.84	0.88	0.00
Ukraine Water	climate	0.97	0.87	0.00
Ukraine Water	biomonitoring	0.51	0.86	0.00
BBC News	All	0.76	0.85	0.00
Contrarian Claims	All	0.65	0.59	0.00

Our method was able to predict when ChatGPT fails at coding (p-values)

Current case studies discussed in this webinar are exploratory only and should not be used for any other purpose.

Results: Coding Time

ChatGPT substantially faster than humans, especially for long documents with many categories

Table 8: Coding Time per Document

Dataset	Coding Time (seconds / document)	
	Human Coder	LLM Coder
Trump Tweets	72	52
Ukraine Water	108	16
BBC News	72	4
Contrarian Claims	144	4

36x faster!

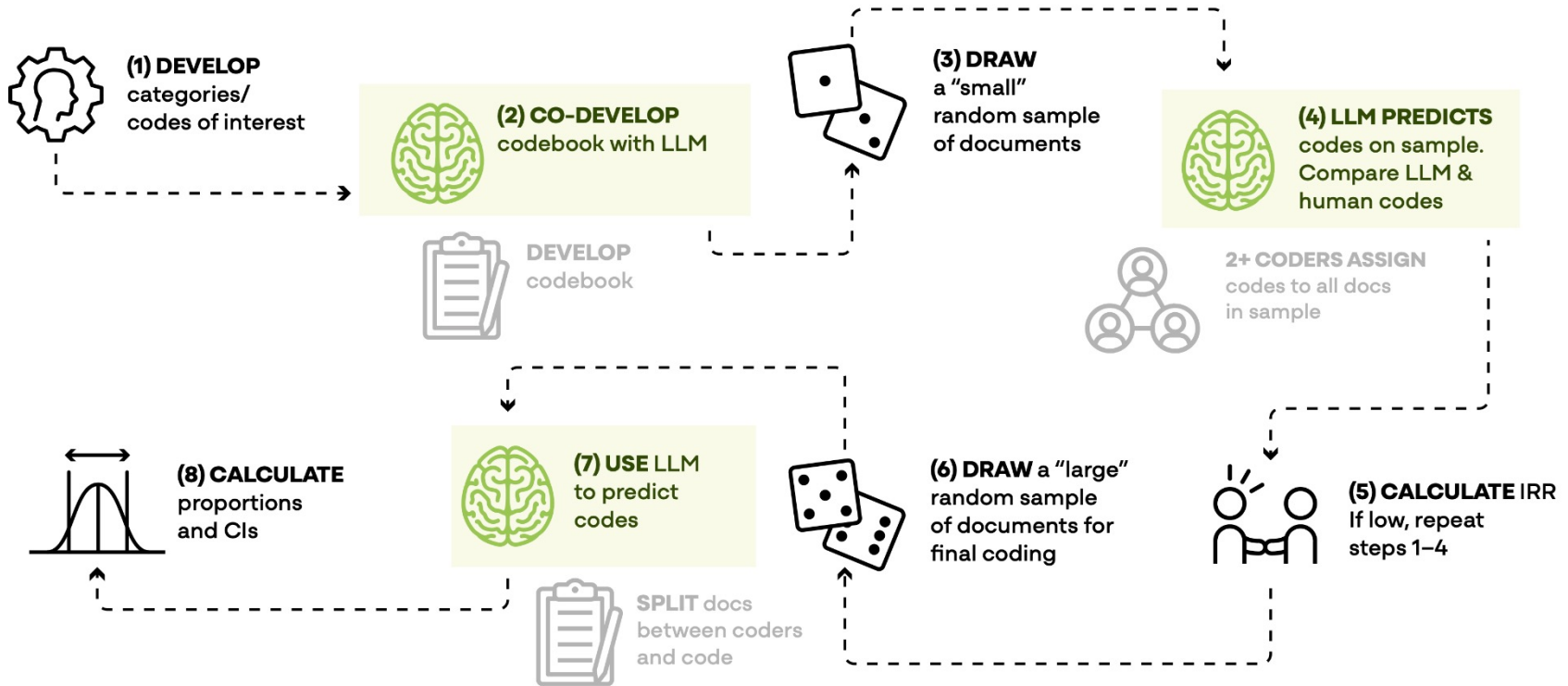
Current case studies discussed in this webinar are exploratory only and should not be used for any other purpose.



Discussion

- Based on coding time and reliability, ChatGPT appears promising for deductive coding.
- Use of ChatGPT for deductive coding will likely require different types of reporting and documentation for reproducibility and critique.
- **We do not consider ChatGPT as a replacement for qualitative researchers**, but rather, a tool to help accelerate the latter stages of deductive coding that tend to be more manually taxing and repetitive.

LLM-Assisted Content Analysis (LACA)





Limitations

- To match the original data sets, we forced ChatGPT to choose Yes / No or a single code (no “I don’t know” option).
- We only assessed ChatGPT and not a wider variety of LLMs.
- Implementing LACA would mean researchers read less documents, which may limit new theory development and discovering themes not proposed by the research team *a priori*.

Want to Learn More?

Check out the pre-print!

<https://arxiv.org/abs/2306.14924>

arXiv:2306.14924v1 [cs.CL] 23 Jun 2023

LLM-ASSISTED CONTENT ANALYSIS: USING LARGE LANGUAGE MODELS TO SUPPORT DEDUCTIVE CODING

Robert Chew, John Bollenbacher, Michael Wenger
Center for Data Science and AI
RTI International
{rchow, jmbollenbacher, mwenger}@rti.org

Jessica Speer, Annice Kim
Center for Communication and Media Impact
RTI International
{jspeer, akim}@rti.org

ABSTRACT

Deductive coding is a widely used qualitative research method for determining the prevalence of themes across documents. While useful, deductive coding is often burdensome and time consuming since it requires researchers to read, interpret, and reliably categorize a large body of unstructured text documents. Large language models (LLMs), like ChatGPT, are a class of quickly evolving AI tools that can perform a range of natural language processing and reasoning tasks. In this study, we explore the use of LLMs to reduce the time it takes for deductive coding while retaining the flexibility of a traditional content analysis. We outline the proposed approach, called *LLM-assisted content analysis* (LACA), along with an in-depth case study using GPT-3.5 for LACA on a publicly available deductive coding data set. Additionally, we conduct an empirical benchmark using LACA on 4 publicly available data sets to assess the broader question of how well GPT-3.5 performs across a range of deductive coding tasks. Overall, we find that GPT-3.5 can often perform deductive coding at levels of agreement comparable to human coders. Additionally, we demonstrate that LACA can help refine prompts for deductive coding, identify codes for which an LLM is randomly guessing, and help assess when to use LLMs vs. human coders for deductive coding. We conclude with several implications for future practice of deductive coding and related research methods.

1 Introduction

Content analysis is widely used in qualitative research to analyze and interpret the characteristics of text, or other forms of communication, due to its systematic and unobtrusive nature [1]. Content analysis typically involves selecting a sample of text data, defining categories to classify the content, and then coding the content according to the categories with definitions. This is typically referred to as deductive coding in which researchers develop a coding scheme based on existing theories and research prior to the coding process. This is in contrast with inductive coding which involves not defining categories *a priori*, but rather identifying and naming categories that emerge from the text during the coding process. While more rigid, deductive coding is more well suited for generalizing results across studies [2].

Despite its strengths, deductive coding is a time-consuming process, particularly when coding substantial amounts of data [3] and for topics that may be nuanced or infrequently mentioned. Coding requires researchers to carefully read and code each piece of content, possibly multiple times, to ensure that they are accurately capturing all relevant information, properly interpreting the text, and applying the category definitions faithfully. This burden becomes magnified when developing and refining the codebook, training coders, and measuring inter-rater reliability to ensure code definitions are well-defined and can be coded consistently [4].

Recently, generative large language models (LLMs) [5, 6] have demonstrated remarkable progress toward achieving human-level performance on a number of natural language understanding and generation tasks [7]. For example, the

CENTER FOR TOBACCO PRODUCTS/OFFICE OF HEALTH COMMUNICATION & EDUCATION'S USE OF LARGE LANGUAGE MODELS (LLMS)

Mario Navarro, Shari Shojaie, Zach Cahn, Andie Malterud Center for Tobacco Products (CTP)

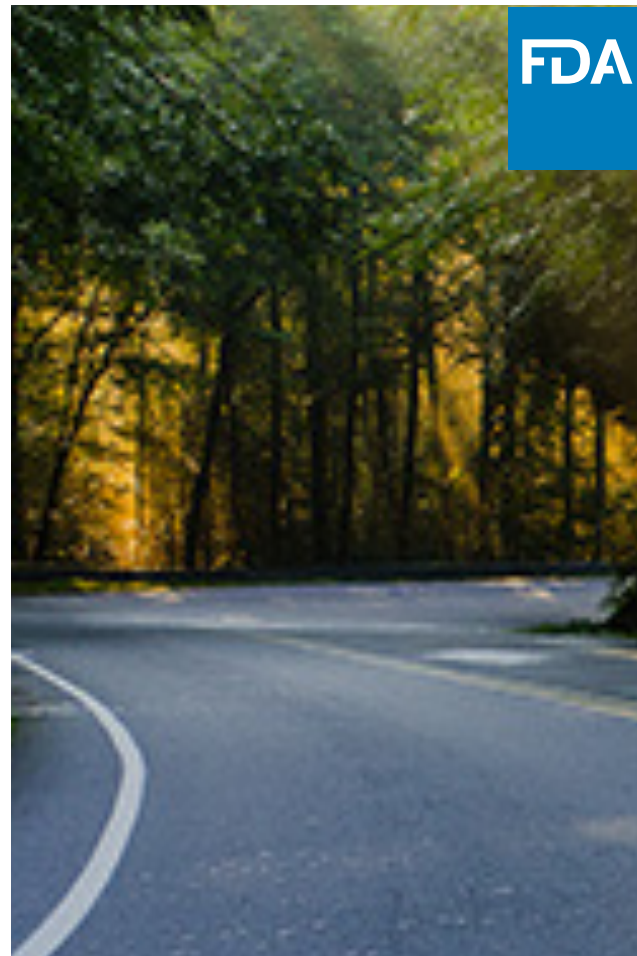
Disclaimer: This is not a formal dissemination of information by FDA and does not represent Agency position or policy.



FDA

OVERVIEW

- Presentation Takeaways
- Who are we?
- What are our team goals?
- Large Language Model (LLM) summary
- Marketing Intelligence Reports
- Things to consider/Next steps



- In the current presentation, I would like to answer the following questions:
 - What is the Market Intelligence (MI) Team within the Center for Tobacco Products (CTP)?
 - What are our goals?
 - How do large language models (LLMs) fulfill these goals?
 - What are our plans to use these models?
 - What methodologies do we use?



CTP REGULATORY AUTHORITIES, GOALS, AND OHCE



- CTP develops is a regulatory body that proposes new rules and regulations concerning the tobacco industry.
- In addition, CTP reviews tobacco applications (Office of Science), writing rules and guidances (Office of Regulations) enforces rules and regulations (Office of Compliance and Enforcement), and **informs the public about the harms (Office of Health Communication and Education).**

WHAT IS THE MARKETING INTELLIGENCE TEAM?

- The Market Intelligence (MI) team, an interdisciplinary team, is housed within the Division of Research and Evaluation and fulfills the urgent need for monitoring and surveillance of the tobacco product landscape to provide resonant messaging.
- Our team of 4 takes point on surveillance platforms and datasets. The team works with a contractor on the market intelligence reports.



- Understand youth, young adult, and current and former users' **perceptions, appeal, and use of current, new, and emerging tobacco products and their impact.**
- Understand youth, young adult, and current and former users' **perceptions of CTP Campaign messaging.**
- Understand youth, young adult, and current and former users' **understanding and sentiment of FDA policies and messaging.**
- Understand **tobacco industry marketing and advertising tactics and their impact** on youth, young adult and current and former users' perceptions, appeal and use of current, new, and emerging tobacco products.

HOW ARE WE USING GENERATIVE PRE-TRAINED TRANSFORMER (GPT) MODELS?



- **Our Problem:** Currently, we are deriving insights from large amounts of social media data From Twitter (X) and Reddit. We would like to pull themes from this data in a rigorous and consistent fashion. Doing this manually is not feasible.
- **LLM Solution:** Currently, we would like to use large language models, specifically GPT, to classify and qualitatively code social media data using a methodology, similar to social science research.
- Using methodologies found in social science research, we are our developing a new method to take traditional qualitative analysis and have them be performed at a very large scale.

HOW ARE WE USING GPT?

- Our work is primarily conducted through university partners and contractors.
- Rather than going through the graphical interface and website interface that you see (which uses GPT 3.5), our partners are going through the application programming interface, with a prompt function, using Python script in Jupyter notebook.
- They feed the individual Reddit/Twitter posts through the API as individual requests. Each request cost just several cents. The current project will be around \$30 for use of ChatGPT.

```
import openai
openai.api_key = 'YOUR_API_KEY'
messages = [ {"role": "system", "content":
              "You are a intelligent assistant."}
]
while True:
    message = input("User : ")
    if message:
        messages.append(
            {"role": "user", "content": message},
        )
        chat = openai.ChatCompletion.create(
            model="gpt-3.5-turbo",
            messages=messages
        )
        reply = chat.choices[0].message.content
        print(f"ChatGPT: {reply}")
        messages.append({"role": "assistant",
            "content": reply})
```

- We work with a contractor to monitor topics of interest based on CTP/OHCE areas of interest and goals **using social media data**. These are performed on a regular basis.
- We are working on a reports about ENDS products and discussions.
 - Previous reports: Menthol, Nicotine & Synthetic Nicotine, Blunts, Hispanic/Latino, Cigarettes.
- The sample of Twitter (X) and Reddit posts and comments is pulled using a query made up of e-cigarette keywords.
- We are just beginning to integrate LLMs into our reports.

- **Current research question:**
 - Can we focus on stories of e-cigarette users? In other words, are there ways to identify personal stories regarding e-cigarette experiences?
- We will analyze the following 6 datasets pulled from topic keywords:
 - (1) Mental Health on Twitter (2,162)
 - (2) Mental Health on Reddit (12,559)
 - (3) Quitting Vaping on Twitter (21,918)
 - (4) Quitting Vaping on Reddit (12,484)
 - (5) Product Switching on Twitter (2,552)
 - (6) Product Switching on Reddit (6,306)

This 5-step approach is modified from Xiao and colleagues (2023).

1. Create the codebook for deductive coding of personal narratives in mental health, quitting vaping, and product switching including definitions of a personal narrative and an example post.
2. Develop prompts for ChatGPT using an example-centered approach.
3. Human coders will manually label samples of posts as containing personal narratives for each of the topics (mental health, quitting vaping, and/or product switching) and each platform (X and Reddit). Then, assess agreement between human coders and with ChatGPT using Cohen's kappa.
4. Assess performance of ChatGPT and quality of codebook.
5. *Analyze and report on results of the ChatGPT classifications of personal narratives with volume analyses, natural language processing methods, and manual qualitative review.*

PROMPT INSTANCE EXAMPLE (X, WORK IN PROGRESS)



We will input the following series of prompts that will be tailored to each topic and platform:

Setup

Give a categorization response as if you were a social science researcher who has expertise in tobacco and vaping use in the United States

Definition of Code

You will be identifying whether tweets contain personal narratives of quitting vaping (given code and platform). “My newest goal is to quit vaping! It’s so damn hard tho But if I was able to quit hard drugs and alcohol this year, I know I can quit this damn nicotine addiction” (example text from codebook) is an example of a Personal Narrative of Quitting Vaping from Twitter (code) because this post discusses an author’s personal experiences with quitting vaping or e-cigarette use, including but not limited to mentions of quitting strategies, tips, or tricks or reports of time since last use of vape products and does not include mentions which only refer to quitting cigarette smoking or stopping use of other tobacco products.

Classification

Reading the following Tweet, does this text classify as a Personal Narrative of Quitting Vaping on Twitter? Please answer yes or no.

“quit vaping cold turkey 3 days ago cuz i was tired of walking around with fruity vapes 0 withdrawal symptoms, im invincible need a new vice now, im thinking dip.”

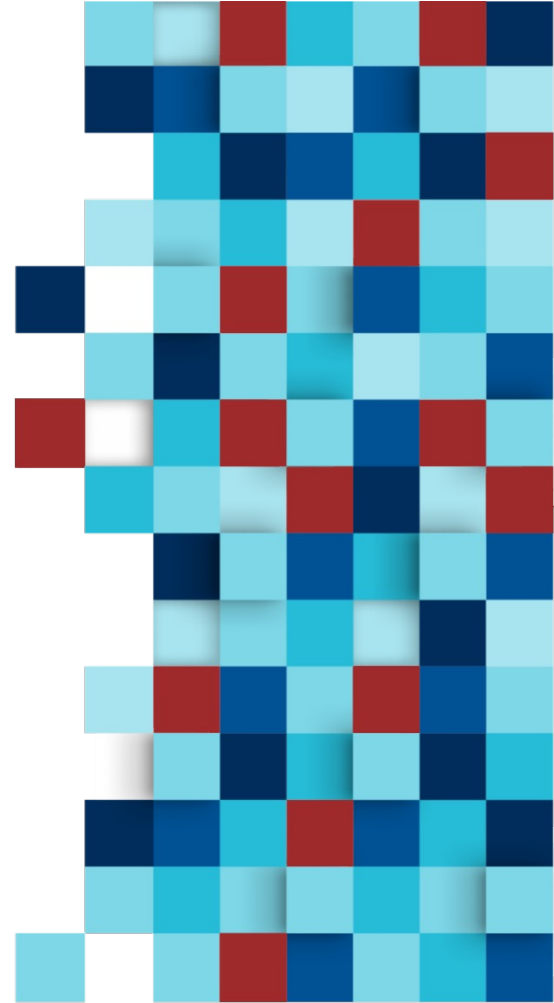
- Ethical considerations
 - We are limited to using public data and information. Anything that is processed through ChatGPT is saved on OpenAI's servers and may be accessible to others. Focus group or survey research should not be analyzed by these large language models.
 - FDA cannot share any internal or confidential information through ChatGPT. Currently, staff cannot directly work with ChatGPT. HHS is working on integration with security and privacy protocols. As mentioned earlier, this work is done through outside collaborations.
- Other considerations
 - There is still little known how these models work and how they determine their conclusions.
 - While there has been some recorded literature on prompt development and how to best elicit responses, it is still a nascent field of work.
 - At this time, LLMs are not a replacement for traditional qualitative coding or supervised learning models, which may provide higher accuracy but use more development and execution time.

- As mentioned previously, this is just one component of our social media analysis.
- This presentation covered the methodology that will be used to code social media posts that utilized personal narrative. This will provide more insight on how individuals deal with their own mental health and vaping behavior journey.

Xiao, Z., Yuan, X., Liao, Q. V., Abdelghani, R., & Oudeyer, P.-Y. (2023). Supporting Qualitative Analysis with Large Language Models: Combining Codebook with GPT-3 for Deductive Coding. IUI'23 Companion: Companion Proceedings of the 28th International Conference on Intelligent Users Interfaces. ACM Digital Library.

<https://dl.acm.org/doi/10.1145/3581754.3584136>

Using LLMs for Developing Plain Language





Case Study Overview

- Leverage ChatGPT to adapt health content into plain language
- Experimented with 25 different materials
 - Webpages (12)
 - Technical reports (5)
 - Manuscripts (5)
 - Consent forms (3)
- Developed a ChatGPT prompt
 - Some content was tested with multiple prompts (e.g., different intended audience)
- Evaluated by multiple measures
 - Readability scores, CCI, qualitative analysis



Case Study Overview

Research Questions

- Can ChatGPT help develop plain language content?
- What are the opportunities and limitations for using ChatGPT to develop plain language content?
- What areas should future research explore?

Specifically Interested In

- If and how well can ChatGPT improve content formatting (headings, bulleted lists, information chunking etc.)?
- If and how well can ChatGPT identify accurate alternatives for technical or scientific terminology?

Literacy in the U.S.

- Almost 130 million adults in the U.S. read at or below a 6th-grade level.
- Individuals with low literacy are four times more likely to report health status as “fair” or “poor.”



What is Plain Language?

- A strategic response to reduce the complexity of information and improve health literacy
- A way of writing that is understandable on the first read (PLAIN, 2010)





Plain Language at RTI

INSIGHTS

Beyond Plain Language: Creating Effective Communication Materials for People with Intellectual and Developmental Disabilities

Molly M. Lynch, Linda Squiers, Sidney Holt, Jeff Novey

Creating Communication Materials for People with IDD and ELL

Adults with intellectual and developmental disabilities and extreme low literacy (IDD/ELL) are a diverse audience with unique communication needs. Even when information is created using current plain language and clear communication standards, people with IDD/ELL may still have challenges understanding it. Before you start creating materials for this audience, it's important to better understand who you're writing for and how their communication needs may differ from other audiences.

Generative AI





ChatGPT Prompt

For this task, you will **rewrite a provided text according to specific guidelines to improve its clarity for the target audience**. The three key elements are: the **original text to rewrite, the target audience's characteristics, and the communication principles** to adhere to. Read through these, and then rewrite the original text with the target audience in mind, following the provided communication principles. Be concise and do not add any new information not covered in the original text. Use Markdown formatting.

"target audience": "The **primary audience** for this document is the general public in the United States. They may be presumed to have high school education, basic computer literacy, and approximately an 8th grade reading level.",

"original text":

clear communications principles":“



Analysis Measures

Evaluator Assessed

- CDC's Clear Communication Index (CCI)
- Accuracy
- Language & tone
- Qualitative notes

Mathematical Calculation

- McLaughlin's SMOG
- Flesh-Kincaid Grade Level
- Flesh Reading Ease



General Observations

- Formatting is denoted by symbols
 - (# = headings, ## = subheading, and ** = bold)
- Some adaptations were too brief and incomplete as a result
- Some inconsistencies appeared across the documents, despite using the same prompt
- There were instances where inclusive language could have been improved but did not occur
- ChatGPT did not demonstrate proficiency in finding alternatives for highly technical terms and phrases
- Terms were often defined when the definition was in the source content
- Reading level decreased in almost all of the materials

Webpage Outcomes

- Website content was sometimes chunked appropriately
- Some section titles worked while others were confusing
- When content was removed from the original version, some pieces of information were lost
- When developing content for audiences with lower literacy, some word choices were confusing and sentences were incomplete
- Multiple instances of carrying over stigmatizing language from source content (e.g., obese)



Health care facility



Health care place




Webpage Example

****Food Poisoning: What You Need to Know****



****Food poisoning can make you feel really sick, but it usually gets better on its own. However, in some cases, it can lead to serious health problems. If you think you have food poisoning, it's important to take care of yourself and contact your doctor if needed.****

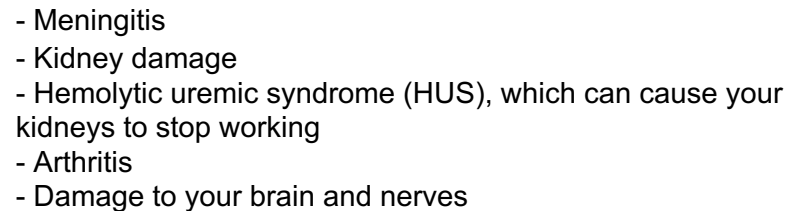


Food poisoning happens when you eat food that has harmful germs in it, like Salmonella or E. coli. The signs that you might have food poisoning include:

- Diarrhea
- Stomach pain or cramps
- Feeling nauseous
- Throwing up
- Fever

If you're throwing up or have diarrhea, make sure to drink lots of water or other fluids. This helps stop you from getting dehydrated, which means your body doesn't have enough water.

While most people get better from food poisoning, some germs can cause serious health problems that need to be treated in a hospital. These can include:

- 
- Meningitis
 - Kidney damage
 - Hemolytic uremic syndrome (HUS), which can cause your kidneys to stop working
 - Arthritis
 - Damage to your brain and nerves

Manuscript Outcomes

- Overall, the structure of background, key points, and findings were clear
- Some of the summaries were very short and lacked information about methods and analysis
- Some interpretations of technical jargon were confusing – sometimes there was no interpretation and technical terms appeared in the output
- Consider figure limitations

***“after recovery from
COVID-19”***



***“after the acute phase of
COVID-19”***



Manuscript Summary Example

Understanding the Impact of COVID-19 on Children and Adolescents

Key Points

- The impact of COVID-19 on children and adolescents after the acute phase of the infection is still being studied.
- The symptoms and health problems that persist or appear 4 weeks or more after the infection are referred to as Postacute Sequelae of SARS-CoV-2 infection (PASC).
- There is a need for more research to understand the incidence, symptoms, and duration of PASC in children and adolescents.
- The study presented here is part of a larger initiative to understand, treat, and prevent PASC.

About the Study

The study used health record data from several large children's health care organizations across the United States. The data included children and adolescents who tested positive for COVID-19 and those who tested negative. The aim was to identify the symptoms and health conditions related to PASC and the medications used to treat them in the 1 to 6 months following the infection.

Findings

The study found that children and adolescents who tested positive for COVID-19 experienced a range of symptoms in the months following their infection. These included changes in smell and taste, chest pain, fatigue, and fever among others. Some also developed health conditions such as myocarditis (inflammation of the heart muscle), acute respiratory distress syndrome, and myositis (inflammation of the muscles).

Medications used to treat these symptoms and conditions included cough and cold preparations, nasal decongestants, corticosteroids with antiseptics, opioids, and decongestants.

The study also found that the burden of PASC was higher in children younger than 5 years, those requiring intensive care during the acute illness, or those with a complex chronic condition. The risk of PASC was highest during the March to June 2020 time period.

Conclusion

The study suggests that the impact of PASC on children and adolescents may be different from adults. More research is needed to fully understand these differences. This study is part of ongoing efforts to better understand, treat, and prevent PASC in children and adolescents.



Strengths and Weaknesses

Strengths

- Condensing large amounts of content
- Chunking content
- Maintaining the meaning of original content
- Avoiding adding new content
- Active voice

Weaknesses

- Inclusive language
- Omitting too much content
- Finding alternatives to and defining medical and health care terminology
- Interpretation of the content
- Consistency




Considerations

- The materials we worked with were publicly available and therefore did not need substantial tone or style changes. It is unclear if ChatGPT can adapt the tone and style of content.
- ChatGPT inputs can be a maximum of 4,096 characters (e.g., manuscripts are typically longer).
- ChatGPT cannot interpret images and figures. This is a consideration when content has multiple images and figures that communicate meaning to readers.




Limitations

- Our reviewers are trained in plain language, but there were multiple reviewers.
 - Our prompts did not specify the length that the content needed to be.
 - We did not include another iteration of testing with an updated prompt after our evaluation.
 - We did not conduct testing with the intended audience to assess how outputs resonate with them.
- 



Where Can More Research be Done?

- Our experiment highlights that multiple iterations of testing is needed to better understand why and when inconsistencies occur.
 - Additional experimentation and adaptation of prompts is needed better serve plain language needs.
 - Research with end-users and target audiences to better understand how adapted content resonates with them.
 - Large language models are constantly evolving so ongoing research is needed to keep a pulse on changes.
- 



Use Cases

- Potentially useful tool to generate a **working first draft** of plain language content.
 - Manuscript summaries
 - Webpage content
- Generative AI, including ChatGPT, should only be used as a supplemental tool.
- Human review and adaptation is still very much required.

Thank You

Robert Chew | rchew@rti.org

Stephanie McInnis | smcinnis@rti.org

Mario Navarro | mario.navarro@fda.hhs.gov

Annice Kim | akim@rti.org

