# Utilizing LLMs for Healthcare Research

By Dr. Chris J. Lu NIH/NLM/LHNCBC/ACIB

April 16, 2025 Applied Clinical Informatics Branch Meeting

Slides: <u>https://wiki.nlm.nih.gov/confluence/display/LHC/LHC+Utilizing+OpenAI+GPT+for+Healthcare+Research</u>



National Library of Medicine Lister Hill National Center for Biomedical Communicati

# Table of Contents (cont.)

- Introduction
  - Objectives and Scope
  - o LLMs, Generative AI, ChatGPT
  - Chatbots and AI Models (API)
  - Prompt Engineering
- A Healthcare Research Use Case Identifying SDOH from Provider Notes
  - Information Extraction from Free Text
  - Traditional NLP Approach
  - LLM approaches OpenAl ChatGPT, OpenAl API, & Meta Al
- Discussion and Conclusion
- Q&A



National Library of Medicine Lister Hill National Center for Biomedical Communicatic

#### EHR-06: No SDOH in the Provider's Notes

Patient: Jane Doe

Date: 01/05/2025

Medical Record Number: 1234567

Provider: Dr. John Smith, MD

Reason for Visit: Follow-up appointment for hypertension management.

**Subjective:** Patient reports feeling generally well. Blood pressure remains well-controlled since last visit, with home readings averaging 120/80 mmHg. Reports adherence to prescribed medication regimen (lisinopril 20mg daily). States she is experiencing some difficulty affording her medication due to recent job loss. Expresses concerns about transportation to appointments, as her car is currently undergoing costly repairs. She mentions increased stress levels related to both her financial and transportation challenges.

#### Objective:

- Vitals: BP 122/78 mmHg, HR 72 bpm, RR 16 breaths/min, Temp 98.6°F (oral).
- **Physical Exam:** Within normal limits. No edema noted. Lung sounds clear bilaterally. Heart sounds regular rate and rhythm.
- Laboratory Data: Most recent lab work (dated 12/15/2024) showed normal lipid profile and blood glucose levels.

#### Assessment:

- Hypertension, well-controlled.
- Social Determinants of Health (SDOH) impacting health:
  - Financial Instability: Recent job loss resulting in difficulty affording medication.
     This is a significant barrier to adherence and overall health management.

- Transportation Barriers: Lack of reliable transportation is limiting access to medical care. This increases the risk of missed appointments and delayed treatment.
- Stress: Financial and transportation issues contribute to increased stress levels, potentially impacting blood pressure control.

#### Plan:

• Continue lisinopril 20mg daily. Schedule follow-up in 3 months.

#### SDOH Intervention:

- Discuss options for medication assistance programs (e.g., manufacturer coupons, patient assistance programs). Provide referral to social worker for assistance with accessing these programs.
- Explore options for transportation assistance (e.g., public transportation, ridesharing services, assistance programs). Refer to social worker to explore eligibility for transportation assistance programs.
- Recommend stress-reduction techniques such as regular exercise, mindfulness, or yoga. Provide appropriate educational material and potentially refer to a therapist for stress management support.
- Schedule follow-up lab work in 6 months.

**Provider's Note:** The patient's SDOH challenges are significant and directly impact her ability to manage her hypertension effectively. Addressing these issues is crucial for optimizing her health outcomes. Further investigation and collaboration with social work are essential to ensure that these barriers are mitigated.

Provider Signature:



#### **PN-06: Extract Wrong Information**

#### **Prompt:**

- Please retrieve the provider's notes from the HER.
- Please retrieve only the provider's notes from the EHR, **do NOT include any other information.**
- Extract only the provider's notes from the electronic health record (EHR). If no provider's notes are present, return "None".
- Please retrieve the provider's notes from the EHR, do NOT include any other information. Do NOT summarize. Return one word "None" if no provider's notes found in the provided EHR.





# PN-06: Hallucination

#### **Prompt:**

- Please identify SDOH from the provider's notes and **map identified SDOH to Z-code**.
- Analyze the provider's note for SDOH that may affect the patient's well-being, mapped identified SDOH to Z-code. Do Not include any explanation and details.

PN-06 (provider's notes 06): The patient's SDOH challenges are significant and directly impact her ability to manage her hypertension effectively. Addressing these issues is crucial for optimizing her health outcomes. Further investigation and collaboration with social work are essential to ensure that these barriers are mitigated.





National Library of Medicine Lister Hill National Center for Biomedical Communications

### EHR-06: Identify SDOH from EHR

#### **Prompt:**

- Please retrieve the provider's notes from the EHR:
- Please retrieve only the provider's notes from the EHR, do NOT include any other information.
- Analyze this EHR to identify SDOH that may affect the patient's well-being, mapped identified SDOH to Z-code without any explanation and details. Do NOT generate SDOH or Z-code if it is no SDOH or Z-code found in the provider's notes. Format your response as a JSON object with "z-code" as key and "SDOH" as value. Return "No SDOH/Z-code found" in such case.





# Python & LLM APIs

- Batch job execution:
  - $\circ~$  process a set of files in a single operation.
- File output:
  - $\circ~$  support for writing results to files.
- Structured output format for pipeline integration:
  - o convert results into structured formats, e.g. JSON objects
  - $\circ~$  enable seamless integration with downstream components.





# Python & LLM APIs: Experiment Platform

- Operating System: Linux (Ubuntu 22.04.5 LTS).
- Python version: 3.10.12.
- Text Editor: VIM (Vi IMproved 8.2) IDE independent.
- OpenAI API Access: API key required (account on hold as of 01.27.2025).
- Meta AI API: Llama 3.2 lightweight 3B model.
  - Host: lhc-lx-gpudev10.nlm.nih.gov.
  - run on-premise, locally, or on-device (no usage fee).
  - Capabilities: supports the OpenAl API protocol (Meta Al API post).
- No cloud platform or interactive development environment (IDE) were used.
- Cloud computing services such as Amazon AWS and Microsoft Azure offer access to AI models, including the OpenAI APIs and Meta AI Llama 3.2.

#### <u>Meta Al Llama</u>

- Llama: Large Language Model by Meta Al
  - a series of open-source AI models (LLMs) developed by Meta AI, designed for finetuning and flexible deployment across platforms.
  - multi-modal: text, images, (audio, video).
  - Meta AI (Meta Platforms, Inc.) is the parent company of Facebook.
  - Releases: 2, 3, 4





# Llama 3.2 Models Summary

	Light-weight LLM	Multi-modal VLM
Release Date	09.25.2024	09.25.2024
Model Size	1B and <mark>3B</mark>	11B and 90B
Context window	128K tokens	128K tokens
Vocabulary Size	128K	128K
Capabilities	Text-only	Text and images
Туре	Chat (instruct) model	Base & Chat model
G-P-T	Yes-Yes-Yes	Yes-Yes-Yes
Multilingual	8 languages	8 languages



# Key NLP Capabilities in Llama 3.2 (3B)

- Reading Comprehension (Language Understanding)

   Allows the model to understand text, identify key concepts and relationships, and generate relevant response.
- Commonsense Reasoning
  - Allow model to interpret implied meaning to support dialogue systems, decision-making, problem-solving with logical inferences.
- World Knowledge (knowledge cutoff: Dec. 2023)
  - Provides accurate, context-aware output based on up-to-date information.

# LLM API - Identify SDOH Project





## extractPN.data (prompt)

Please **extract** the provider's notes from the following EHR text. EHR text: ``**{ehr}**``

If no provider's notes are found, return "None". Return only the notes, without any additional information.

NLP capabilities	Requirement	
Reading Comprehension	Some – extract	
Commonsense Reasoning	Little – if, return None	
World Knowledge	No	



# identifyPN.data (prompt)

You are a clinical expert. Please analyze the following Provider's notes text to <b>identify</b> Social Determinants of Health (SDOHs) and the associated reason that may affect patient's well-being. Provider's notes text: "'{pn}"		
Find the Z-code for each identified SDOH from ICD-10-CM (International Classification of Diseases, 10th Revision, Clinical Modification) coding system.		
Please format the response <b>as a list of JSON object with indent 2</b> <b>with SDOH, Z-code, reason as the keys.</b> Return an empty list if no SDOH is identified. Please do NOT include extra information or explanation except for the JSON object.		

NLP capabilities	Requirement	
Reading Comprehension	more – analyze, identify, find, etc.	
Commonsense Reasoning	more – SDOH, Z-code, JSON, if, return None	
World Knowledge	Yes – SDOH, Z-code, list, JSON, indent	



### ehr processor.py: Batch Jobs





#### pn processor.py: Structured Output





National Library of Medicine Lister Hill National Center for Biomedical Communications

#### **Result Comparison on Llama 3.2 and GPT-40**

API: Llama 3.2 (3B)	Chatbot: Chat GPT-4o	
<pre>{     "SDOH": "Unstable Housing",     "Z-code": "L00-L09",     "reason": "Living in a shelter with limited access to basic necessities" }, {     "SDOH": "Food Insecurity",     "Z-code": "C02-C08",     "reason": "Difficulty accessing fresh food due to transportation issues" }, </pre>	<pre>{     "SDOH": "Housing instability (shelter residence)",     "Z-code": "Z59.0",     "reason": "John stays at a shelter."     },     {         "SDOH": "Employment and occupational concerns",         "Z-code": "Z56.0",         "reason": "Works part-time in a low-paying job without health insurance."     }, </pre>	
<pre>{     "SDOH": "Employment and Income",     "Z-code": "P00-P99",     "reason": "Low-paying job without health insurance, leading to financial stress" },</pre>	{ "SDOH": "Food insecurity", " <b>Z-code": "Z59.4",</b> "reason": "Difficulty accessing fresh food due to transportation issues." },	
{ "SDOH": "Social Isolation", "Z-code": "S06-S09", "reason": "No family support nearby, contributing to high stress levels" }	<pre>{     "SDOH": "Lack of adequate social support",     "Z-code": "Z63.31", &lt;= Absence of family member due to military deployment     "reason": "No family support nearby." }, {</pre>	
	"SDOH": "Stress related to life circumstances", "Z-code": "Z65.8", "reason": "High stress related to job insecurity and living conditions."	



# Comparison on Llama 3.2 and GPT-40

Feature	Meta AI, Llama 3.2	OpenAl, GPT-4o
Release Date	September, 2024	March, 2024
Parameters	1B, <mark>3B</mark> , 11B, 90B	Estimated 200B+
Context Window	128K tokens	128K tokens
Vocabulary Size	128K	200K (199,997)
Multimodal	Text + vision	Text + audio + image+ video
Multilingual	8 languages	77 languages
G-P-T	Yes-Yes-Yes	Yes-Yes-Yes
Deployment Options	Edge device and cloud-based	Cloud-based
Accessibility	Open source	Proprietary



National Library of Medicine

## **Prompt Engineering**

- Objective: to enhance the response from the AI model.
- Processes:
  - o improve the communication with model to get better response.
  - adjust the wording, structure, and context in a prompt.
  - utilize one-shot or few-shot prompting are often employed.
  - apply chain-of-thought prompting to solve complex reasoning problems.
  - $\circ~$  is an iterative process aimed at better response.
  - o does NOT constitute fine-tuning or re-train the AI model.
- Example:
  - prompts retained in a session in ChatGPT is a form of prompt engineering, they do not involve training the model.



### **Prompt Evaluation**

- Objective: obtain an optimized prompting that the response is:
   o in the desired format.
  - o high quality.
- Processes:
  - Establish an annotated gold standard data for evaluation (a smaller set).
  - Define measurement metric: accuracy, precision, recall, F1 score, Cohen's Kappa (K), Matthews Correlation Coefficient (MCC), ROC-AUC (Receiver Operating Characteristic - Area Under Curve).
  - An iterative validation processes to improve prompts.

# **Challenges & Strategies**

- Inconsistency (difficult to debug):
  - o same prompt might yield different results, especially in a complicated case.
- Overly verbose (redundant):
  - o responses may contain excessive detail or redundant reasoning.
- Hallucination (inaccurate):
  - o factual inaccuracies, fabricate information, **making up**, or misleading content.
- The model is highly sensitive to the input phrases and struggles with ambiguity, resulting in misinterpreting the user's intent. Use specific instructions and examples, identify patterns, reinforce context, limit and format output accordingly.

# Security & Privacy



#### **NIH Policy**

- Data sanitization: remove sensitive data.
- Data privacy: strong authentication and access controls
- De-Identification: remove identifying information.

#### **OpenAl**

- Strict data access controls: only authorized personnel.
- User's data is not used for training AI models without user consent.
- Data retention policies are aligned with privacy policy.

#### **Data Security:**

- compliance with Service Organization Controls (SOC) 2 Type 2. 0
- cloud Security Alliance Security Trust Assurance and Risk (STAR). 0
- **Privacy Policy:** ٠

•

- Compliance with applicable privacy laws. 0
- GDPR: The General Data Protection Regulation. 0
- CCPA: The California Consumer Privacy Act. 0
- https://openai.com/security-and-privacy/
- https://openai.com/policies/row-privacy-policy/



## **OpenAl Cost Report & Suggestions**

- Charged by tokens and models.
- Total expense in January 2025: \$47.84
  - GPT model: average of \$2.38 per day
  - O1 model: \$7.24 per day
- Suggestions:
  - Use the Chatbot or Llama 3.2 for experimenting with prompt.
  - Use OpenAl APIs: GPT-40 for NLP development.
  - Use OpenAI o1 or o3 models for complex problems (reasoning).
  - Do NOT run large sets of files until you are ready (tested on a smaller set).
  - $\circ~$  Limit the max. tokens for output.



National Library of Medicine Lister Hill National Center for Biomedical Communicatio

### <u>Conclusion</u>

- LLMs appear highly promising for NLP tasks in healthcare research.
  - **Pre-trained** AI models: reduce development efforts.
  - High quality training data: leads to superior response.
  - Low learning curve: prompt engineering.
  - User-friendly APIs: via access key, the OpenAI package, and proxy setup.
  - **Rich NLP Capabilities**:

summarization (w/wo focus), information extraction, sentiment analysis, text classification, question answering, negation detection, and more.

- Rapid model advancements: from GPT to o1, o3 and Llama (from 3 to 4) models.
- Cost considerations: use suggested strategies.



### Future Work

- To experiment with different LLMs:
  - Meta Al Llama 3.2 (3B) -> 3.3 (70B)
- To experiment with a mid-scale task:
  - Antonym Annotation: An evaluation on the linguistic performance of LLMs in NLP.
    - Gold Standard: A collections of 13K antonyms from the Lexicon.2025 release, annotated by linguists.
    - Classification on four tags: canonicity, domains, types and negations.
    - Antonyms requires knowledge of linguistic, semantic, and statistics.
    - A compelling research subject for AMIA, MedInfo, HealthInfo conferences.
- To explore a more complicate task:
  - $\circ~$  To apply in healthcare (NLP) projects.



# <u>Q & A</u>

- Thank you!
- URL: <u>https://wiki.nlm.nih.gov/confluence/display/LHC/LHC+Utilizing+OpenAI+GPT+for+Healthcare+Research</u>

