Utilizing OpenAI GPT for Healthcare Research

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

March 06, 2025
Applied Clinical Informatics Branch Meeting
Slides: https://wiki.nlm.nih.gov/confluence/display/LHC/LHC+Utilizing+OpenAI+GPT+for+Healthcare+Research

Table of Contents

Objectives

- An overview of LLMs in NLP - OpenAI ChatGPT, prompt engineering, OpenAI API (AI models).

- Utilizing LLMs (OpenAI GPT) in healthcare research:
  - The role of Natural Language Processing (NLP) in healthcare research.
  - NLP functionalities and limitations of OpenAI GPT.

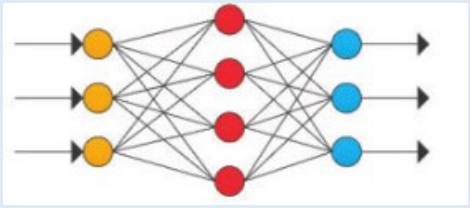
Scope

- Focuses on fundamental research in NLP and its applications in healthcare.
- Text-based NLP tasks and applications.

- Natural Language:
  The everyday language humans use naturally: spoken, written, or signed.
- Natural Language Processing (NLP):
  The use of computer to process and analyze text for various applications.

NN, DL, LLM, Generative AI

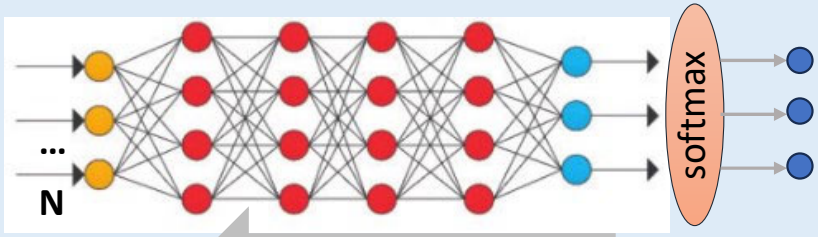
Neural Networks (NN)



Input feature | Hidden Neuron | Output Phase

: node
: link

Deep Learning (DL-NN)



Input Layer | Hidden Layer (more) | Output Layer

Large Language Model (LLM)

- Trained on a large amount datasets.
- **Billions of parameters**: weights, biases
- **Transformer** (2017)
  - A DL architecture
  - Encoder and decoder.
  - Attention mechanism.
  - Parallelization for fast training.
  - Understand and generate language.

- Generative AI: AI models that can create new content, including text, images, music, and other data.
- Most LLMs are generative models.
- ChatGPT is an LLM and also a generative AI model.

Popular Pre-trained LLMs

- OpenAI AI models (GPT)
  - OpenAI GPT-3.5: 175 billion parameters, 96 hidden layers
  - OpenAI GPT-4.o: 1.8 trillion parameters, 120+ hidden layers
  - OpenAI o1 & o3 models: for reasoning heavy tasks
- Meta, open-source AI Llama3 (Facebook, GPT)
  - Llama 2: 70 B parameters, 80 hidden layers
  - Llama 3.1: 8B,70B, 405 B parameters, 32-126 hidden layers
  - Liama 3.2: 1B, 3B, 11B, 90 B parameters, 16-80 layers
- Google, BERT (not generative, understanding, ~~G~~PT)
  - BERT-Base: 100 million parameters, 12 hidden layers
  - BERT-Large: 340 million parameters, 24 hidden layers

OpenAI ChatGPT

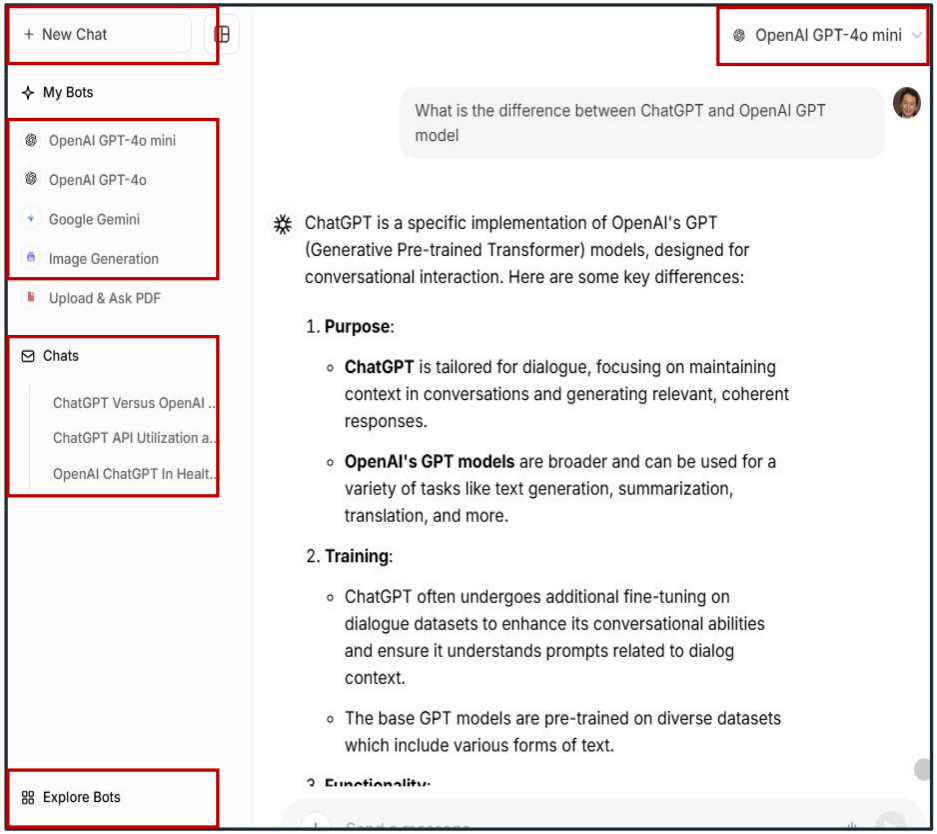
- ChatGPT – Chat Generative Pre-trained Transformer
  - Chat: interactive dialogue involving conversation and prompts.
  - Generative: creates new data or content.
  - Pre-trained: provides relevant responses without training the AI model.
  - Transformer: the underlying Neural Network architecture.

Chatbot & API (AI Models)

- Chatbot (APPs or Web browser): simulate conversation and provide responses.
- LLM API (AI models): allows developers to integrate the capabilities of LLMs.

Company	Chatbot	AI Model	APIs
OpenAI (Microsoft, 13M)	ChatGPT	GPT models: gpt-3.5-turbo, gpt-4o, gpt-4o-mini Reasoning models : o1, o1-mini, o3-mini	Yes
Microsoft	Copilot	gpt-4o, gpt-4o-mini	Yes
Gemini Trust Co (Google - Bard)	Gemini	2.0 Flash, 2.0 Flash-Lite	Yes
Anthropic AI (Amazon)	Claude	Claude: 3.5 Haiku, 3.7 Sonnet, Opus	Yes
Meta AI (Facebook)	Meta AI	Llama 3	Yes
NIH	CHiRP	**gpt-4o**	**No**

Chatbot - ChatGPT



- **Reasoning AI Models**: gpt-4o, gpt-4o-mini, o1, o1-mini, o3-mini, Gemini, Claude.
- **Search function** : web search engines.
- **Prompts** (retention):
  - session retention: retained within a single session.
  - long-term memory: use memory feature to save prompts for future reference.
  - engineering: saved prompts are used to enhance responses, not for training the model itself.
- **Other functions**: upload and parsing files, etc.

Prompt Engineering

- **Prompt**: the user's input text that guides the AI model to generate a relevant response.
- **Prompt Engineering**: the process of creating and refining input prompts to improve relevance and accuracy of AI model responses.

- A well-crafted prompt includes:
  - persona
  - task (mandatory)
  - context
  - examples
  - tone
  - desired output

Prompt Example

Prompt	Component
You are a computational linguist.	persona
Describe what NLP is	task
to a middle school student	context
in a casual tone.	tone
Make it like explaining that a search engine is a tool that helps you find information on the internet.	examples
Please keep it under 3 sentences.	output

Use Case: Identify SDOH from Provider's Notes

Social Determinants of Health (SDOH)

- Definition (WHO):
  - non-medical factors that influence health outcomes.
  - the conditions in which people are born, grow, work, live, and age.
  - such as employment, housing, transportation, parental status, relationship, and social support.

- Importance:
  - plays a critical role in health outcomes and significantly affect overall well-being.
  - addressing SDOH can enhance health and lead to better outcomes.

- Challenge: often underreported or buried within EHRs.

EHR: Structured and Unstructured Data

- Electrical Health Record (EHR):
  - a digital version of a patient's medical history.
- Structured data:
  - demographics (name, DOB, gender, address, contact information), vital signs (BP, HR, temperature, respiratory rate) , lab results, medication lists, problem lists, immunization records, allergy information, procedure codes, and more.
- Unstructured data:
  - clinical notes, provider's notes, progress notes, discharge summaries, consultation notes, consultation notes, caregiver notes, incident reports, care plans, and more.
  - patient symptoms, patient history, patient feedback, radiology reports, lab results, medication orders, and more.

Provider's Notes

- Provider's notes:
  - is an unstructured text document.
  - records a patient's medical encounter written by healthcare providers.
  - includes subjective information, objective information, assessments, plans, additional comments (e.g. SDOH).

EHR-01: Information Extraction

Patient Information

- **Patient Name:** John Doe
- **Age:** 45
- **Gender:** Male
- **Date of Birth:** [Insert DOB]
- **Contact Information:** [Insert Phone Number, Address]

Chief Complaint

Patient presents with high stress levels, difficulty accessing fresh food, and concerns about job insecurity.

Medical History

- **Chronic Conditions:** None reported.
- **Previous Surgeries:** None reported.
- **Medications:** Not currently taking any medications.

Social History

- **Living Situation:** Lives alone.
- **Employment:** Works in a low-paying job without health insurance.
- **Diet:** Reports difficulty accessing fresh food due to transportation issues.
- **Support System:** No family support nearby.
- **Lifestyle:** Sedentary lifestyle due to work constraints; struggles with maintaining a healthy diet.

Review of Systems

- **General:** Reports feeling stressed; no weight changes noted.
- **Cardiovascular:** No chest pain, palpitations, or edema.
- **Respiratory:** No shortness of breath or cough.
- **Gastrointestinal:** No significant changes in appetite; reports occasional gastrointestinal discomfort due to stress.
- **Musculoskeletal:** No joint pain or muscle aches reported.
- **Psychiatric:** High levels of stress related to job insecurity and living conditions; no history of mental health disorders reported.

Physical Examination

- **Vital Signs:**
  - Blood Pressure: [Insert BP]
  - Heart Rate: [Insert HR]
  - Respiratory Rate: [Insert RR]
  - Temperature: [Insert Temp]
- **General Appearance:** Appears anxious.
- **Cardiovascular:** Heart sounds normal; no murmurs detected.
- **Respiratory:** Clear upon auscultation; no wheezing or crackles.
- **Abdomen:** Soft, non-tender; normal bowel sounds.
- **Neurological:** Alert and oriented; no focal deficits.

Provider's Notes

John stays at a shelter and works part-times in a low-paying job without health insurance. He mentioned difficulty in accessing fresh food due to transportation issues. Reported high stress related to job insecurity and living conditions. No family support nearby.

Plan

1. **Referral to Social Services:** Connect with local food banks and community programs for assistance.
2. **Mental Health Support:** Encourage John to seek counseling services or support groups to manage stress and improve coping strategies.
3. **Follow-Up Appointment:** Schedule a follow-up in 4–6 weeks to assess progress and re-evaluate social support and stress levels.
4. **Health Education:** Provide resources on nutrition and stress management techniques.
5. **Consider Insurance Options:** Discuss potential avenues for obtaining health insurance coverage, including local programs.

Signature

[Provider's Name]
[Provider's Title]
[Date]

EHR-01: Provider's Notes to SDOH

Provider's Notes
John stays at a shelter and works part-times in a low-paying job without health insurance. He mentioned difficulty in accessing fresh food due to transportation issues. Reported high stress related to job insecurity and living conditions. No family support nearby.

1. **Housing Stability**: stays at a shelter.
2. **Employment and Income**: works part-time in a low-paying job.
3. **Health Insurance Access**: without health insurance.
4. **Food Security**: difficulty in accessing fresh food.
5. **Social Support**: no family support nearby.
6. **Stress and Mental Health**: high stress related to job insecurity and living conditions.

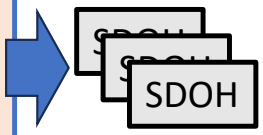
Identify SDOH from Provider's Notes

- A challenging NLP task (information extraction) because:
  - require an understanding the semantic meaning of the text.
  - consist of free (unstructured) text without a predefined standard format.
  - contain complexity in content.

- Two main research topics in AMIA (American Medical Informatics Association):
  - SDOH (health and healthcare equity working group)
  - Informatics in generative AI & LLMs (NLP working group)

NLP (w/o LLM) - Identify SDOH from Provider's Notes

1. Preprocessing: tokenizing, normalization, and other NLP techniques.
2. Keyword identification: experts develop a list of keywords and phrases related to SDOH.
3. NLP Models: gold standard, AI models, Iterative processes
   - Expert System, rule-based systems, pattern recognition.
   - Text Classification: classify SDOH in sentences.
   - Named Entity Recognition: identify terms related to SDOH.
4. Post-processing: refinement and validation of results.



LLM - Identify SDOH from Provider's Notes

EHR-01: Derived Medical Plan from SDOH

1. **Housing Stability**:
   Connect John with local housing assistance programs or shelters that may offer transitional housing support.
2. **Employment and Income**:
   Refer John to job training and placement services to improve his employment opportunities and potential income.
3. **Health Insurance Access**:
   Assist John in applying for public health insurance programs (e.g., Medicaid) or community health plans available for low-income individuals.
4. **Food Security**:
   Provide information on food pantries, community gardens, or nutrition programs that can help improve access to fresh food.
5. **Social Support**:
   Refer John to community support groups or organizations that can connect him with social services and peer support opportunities.
6. **Stress and Mental Health**:
   Recommend counseling or mental health services to address high stress levels and provide coping strategies. Offer resources for stress management programs if available.

EHR-01: Map SDOH to Z-Code

- Z-codes:
  - a set of classification codes used in the ICD system, specifically in the ICD-10-CM.
  - capture health-related conditions that are not classified as diseases or injuries.
  - individual: personalized treatment plans and preventative care strategies
  - identify factors that increase healthcare utilization and costs, (population analytics)

- ICD-10-CM: *International Classification of Diseases, 10th Revision, Clinical Modification*.

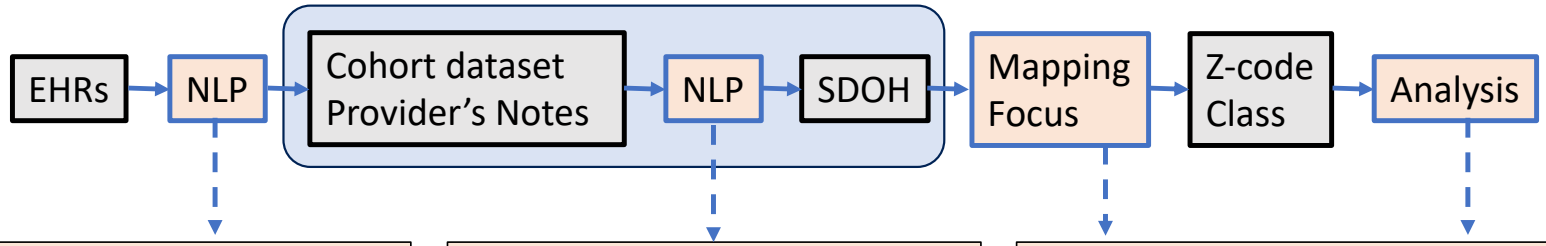
1. Housing Stability - Z59.0
2. Employment and Income - Z56.9
3. Health Insurance Access - Z59.7
4. Food Security - Z59.4
5. Social Support - Z63.9
6. Stress and Mental Health - Z91.29

EHR-01: Focus on One Group of SDOH (e.g. Homeless)

1.Housing Stability: stays at a shelter.
2.Employment and Income: works part-time in a low-paying job.
3.Health Insurance Access: without health insurance.
4.Food Security: difficulty accessing fresh food.
5.Social Support: no family support nearby.
6.Stress and Mental Health: high stress related to job insecurity and living conditions.

Yes

Use Case: NLP Pipeline



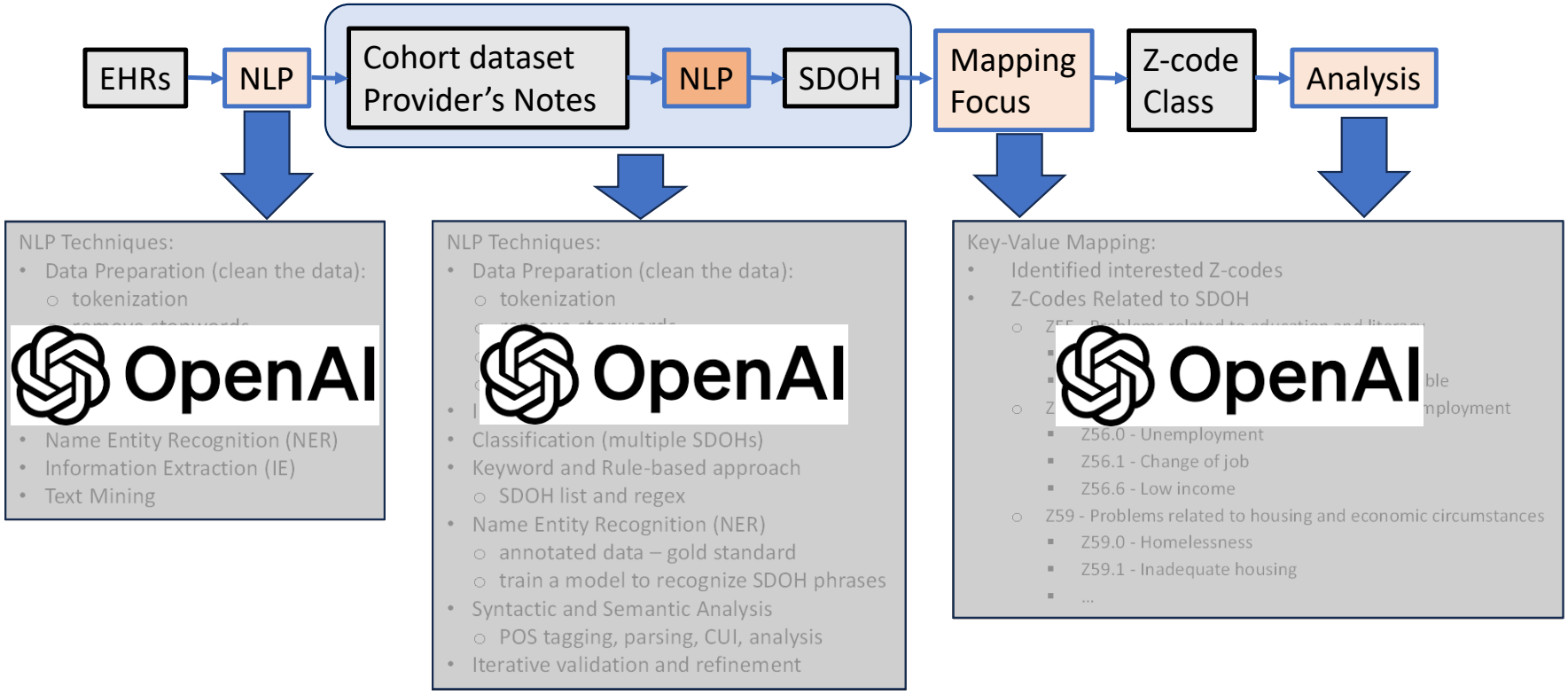
NLP Techniques:
- Data Preparation (clean the data):
  - tokenization
  - remove stopwords
  - normalization
  - etc.
- Keyword Extraction
- **Name Entity Recognition (NER)**
- **Information Extraction (IE)**
- Text Mining

NLP Techniques:
- Data Preparation (clean the data):
  - tokenization
  - remove stopwords
  - normalization
  - etc.
- **Information Extraction (IE)**
- **Classification (multiple SDOHs)**
- Keyword and Rule-based approach
  - SDOH list and regex
- Name Entity Recognition (NER)
  - annotated data – gold standard
  - train a model to recognize SDOH phrases
- **Syntactic and Semantic Analysis**
  - POS tagging, parsing, CUI, analysis
- Iterative validation and refinement

Key-Value Mapping:
- Identified interested Z-codes
- Z-Codes Related to SDOH
  - Z55 - Problems related to education and literacy
    - Z55.0 - Illiteracy and low literacy
    - Z55.1 - Schooling unavailable and unattainable
  - Z56 - Problems related to employment and unemployment
    - Z56.0 - Unemployment
    - Z56.1 - Change of job
    - Z56.6 - Low income
  - Z59 - Problems related to housing and economic circumstances
    - Z59.0 - Homelessness
    - Z59.1 - Inadequate housing
    - …

Classification and Analysis: …

Use Case: OpenAI GPT Approach

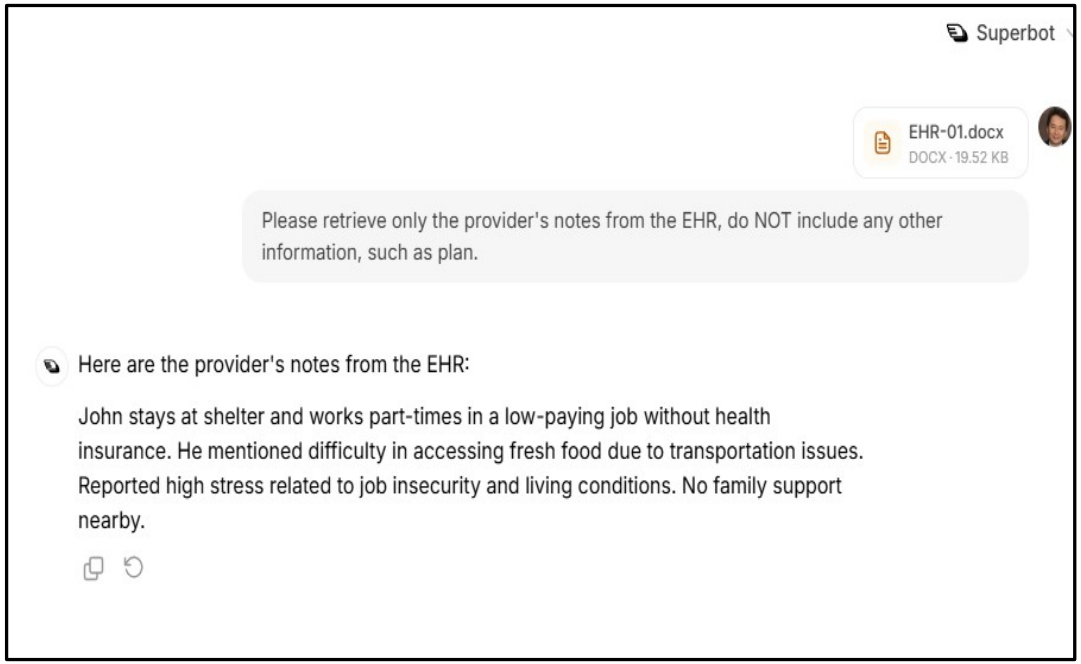


Extract Provider's Notes from EHR-01

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.**
- Please retrieve only the provider's notes from the EHR, do NOT include any other information, such as plan.

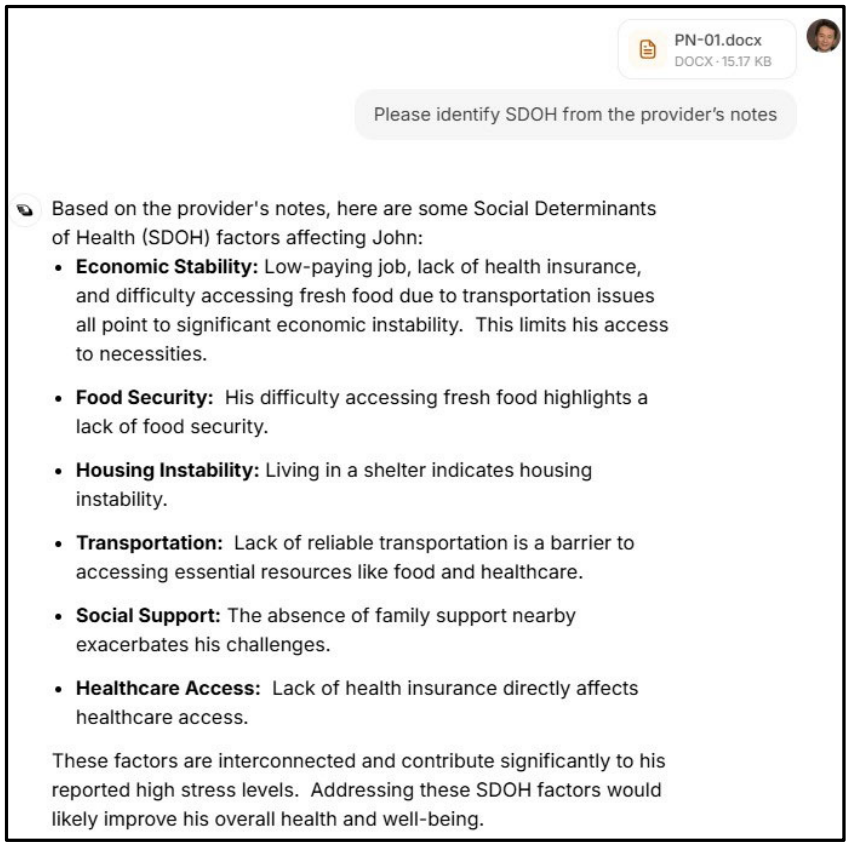
Prompt Engineering



PN-01: Identify SDOH, Overly Verbose

Prompt:

- Please identify SDOH from the provider's notes.
- Analyze the provider's note for SDOH that may affect the patient's well-being.

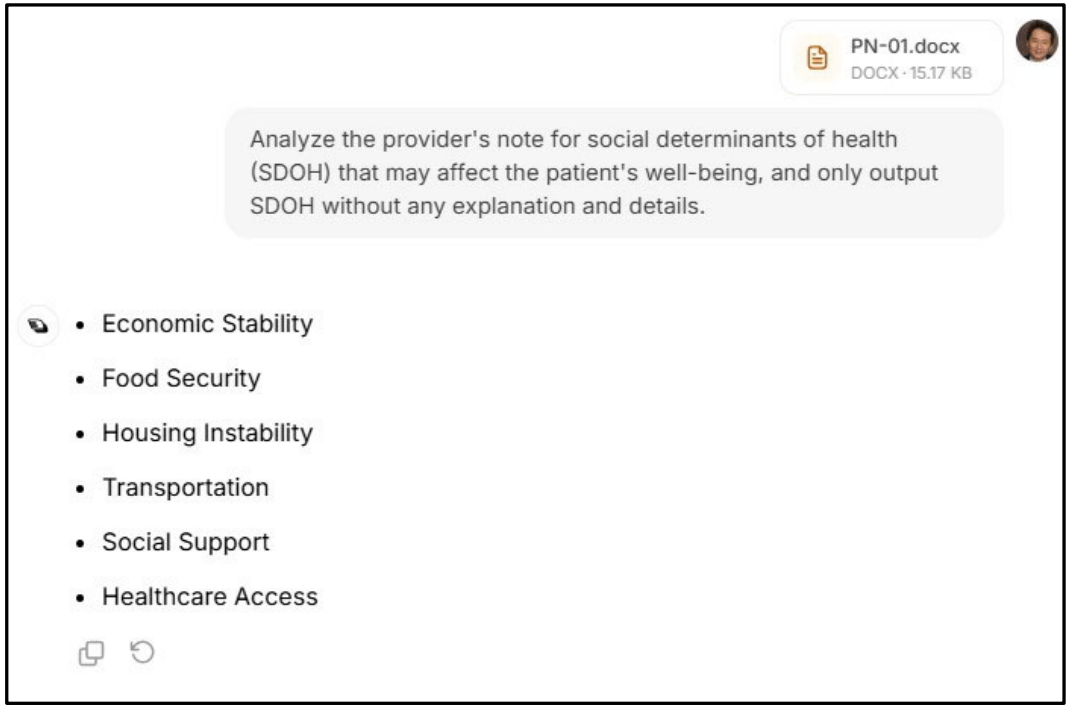


PN-01: Identify SDOH (Cont.)

Prompt:

- Please identify SDOH from the provider's notes.
- Analyze the provider's note for SDOH that may affect the patient's well-being.
- Analyze the provider's note for SDOH that may affect the patient's well-being, **and only output SDOH without any explanation and details**.

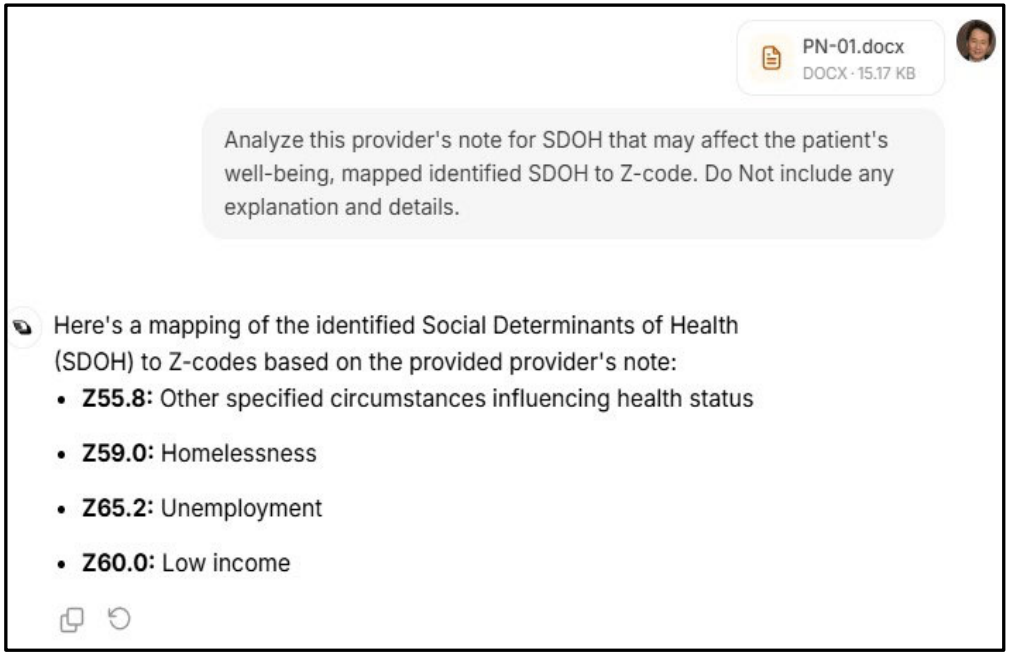
Desired Output



PN-01: Map SDOH to Z-code

Prompt:

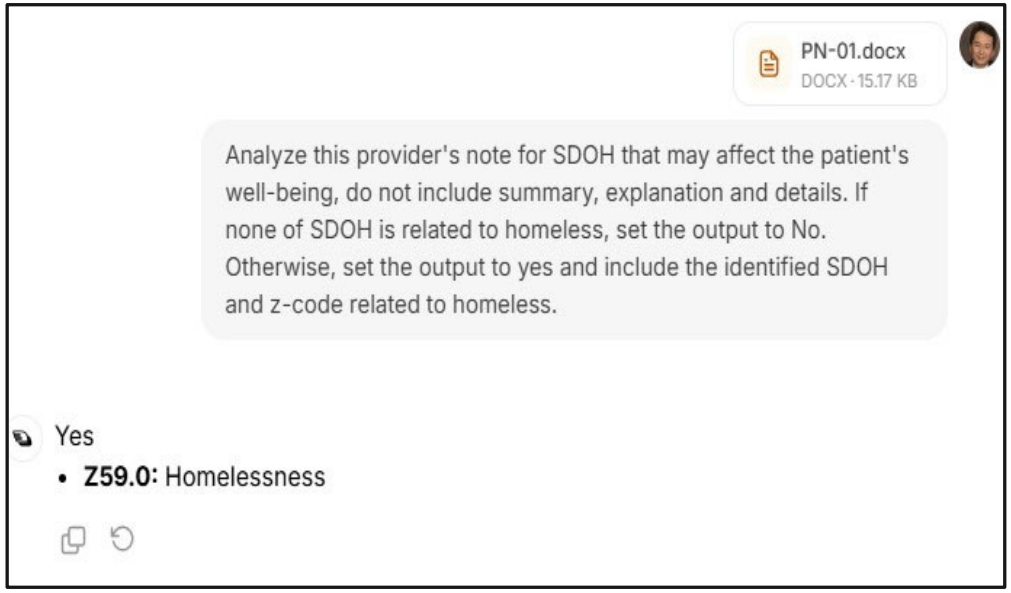
- Analyze this provider's note for SDOH that may affect the patient's well-being, **mapped identified SDOH to Z-code.** Do Not include any explanation and details.



PN-01: Focus on SDOH of “homeless”

Prompt Engineering:

- Analyze this provider's note for SDOH related to homelessness that may affect the patient's well-being.
- Analyze this provider's note SDOH and identify if the patient is homeless, output with one word, yes or no.
- Analyze this provider's note for SDOH that may affect the patient's well-being, do not include summary, explanation and details. **If none of SDOH is related to homeless, set the output to No. Otherwise, set the output to yes and include the identified SDOH and z-code are related to homeless.**



EHR-09: No Provider's Notes

Patient: Robert Johnson
Date: 01/05/2025
Medical Record Number: 9876543
Provider: Dr. Anya Sharma, MD
Reason for Visit: Annual physical examination and follow-up for type 2 diabetes.

Subjective: Patient reports feeling well overall. States he is experiencing occasional fatigue but attributes it to increased work demands. Reports good compliance with his prescribed medication regimen (metformin 1000mg BID, glipizide 5mg daily). Denies chest pain, shortness of breath, or dizziness. He reports maintaining a healthy diet and exercising regularly, 3-4 times per week.

Objective:

- **Vitals:** BP 130/85 mmHg, HR 70 bpm, RR 18 breaths/min, Temp 98.6°F (oral), Weight 185 lbs, Height 5'10".
- **Laboratory Data:**
  - Fasting Blood Glucose: 140 mg/dL (slightly elevated)
  - HbA1c: 7.2% (slightly elevated)
  - Lipid Panel: Total Cholesterol 190 mg/dL, LDL 115 mg/dL, HDL 45 mg/dL, Triglycerides 150 mg/dL
- **Physical Exam:** General appearance is good. No acute distress. Heart sounds regular rate and rhythm. Lungs clear to auscultation. Abdomen soft, non-tender. No edema noted.

Assessment:

- Type 2 Diabetes Mellitus, fairly well controlled. Slight elevation in fasting blood glucose and HbA1c suggests need for minor adjustments to diabetes management.
- Mild hyperlipidemia.

Plan:

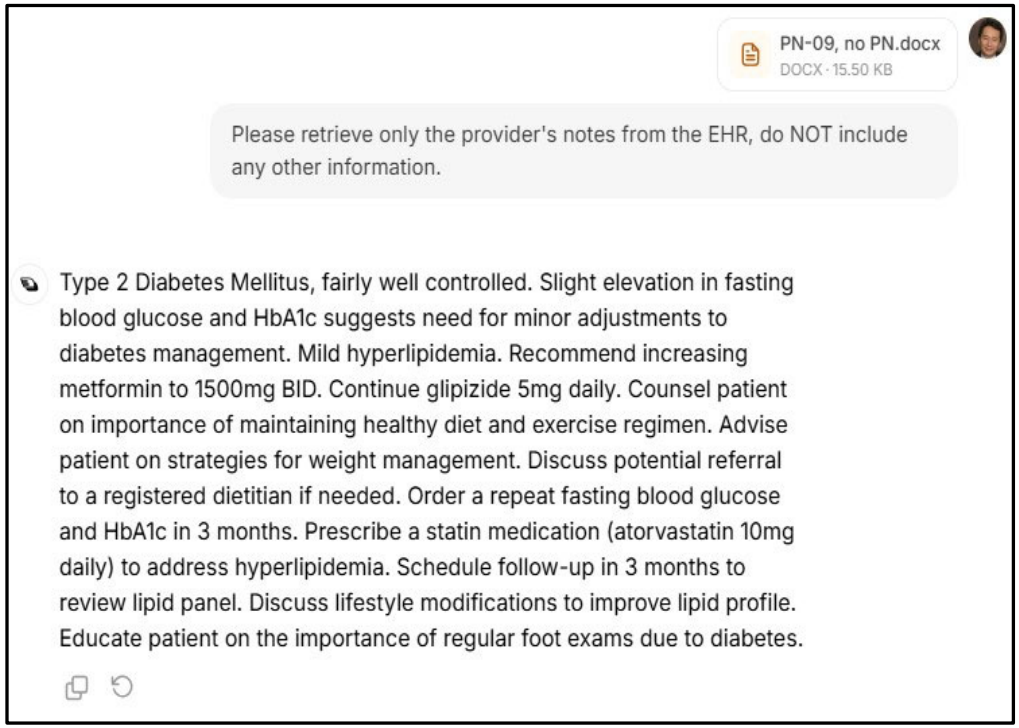
- Recommend increasing metformin to 1500mg BID.
- Continue glipizide 5mg daily.
- Counsel patient on importance of maintaining healthy diet and exercise regimen.
- Advise patient on strategies for weight management. Discuss potential referral to a registered dietitian if needed.
- Order a repeat fasting blood glucose and HbA1c in 3 months.
- Prescribe a statin medication (atorvastatin 10mg daily) to address hyperlipidemia. Schedule follow-up in 3 months to review lipid panel. Discuss lifestyle modifications to improve lipid profile.
- Educate patient on the importance of regular foot exams due to diabetes.

Provider Signature:_____

PN-09: Generative Summarization

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.**
- Please retrieve the provider's notes from the EHR, do NOT include any other information. **Return "None" if no provider's notes found in the provided EHR.**
- Please retrieve the provider's notes from the EHR, do NOT include any other information. **Return "None" if no provider's notes found in the provided EHR. Do NOT generate provider's notes from summarization.**



PN-09: Specific Instructions

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.
- 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.