

A Primer for Using ChatGPT in Medical Education



Christopher Runyon, Peter Baldin, Su Somay, Yiyun Zhou

Version 1.0 (April 4, 2024)

DISCLAIMER

The information presented in this document/presentation is solely for informational purposes and does not necessarily reflect the official stance or views of NBME. The opinions expressed herein are those of the authors/presenters and may not represent the views or policies of NBME. While every effort has been made to ensure the accuracy and reliability of the information provided, NBME makes no representations or warranties of any kind, express or implied, about the completeness, accuracy, reliability, suitability, or availability with respect to the content contained herein. Any reliance you place on such information is therefore strictly at your own risk.

Contents

Overview	3
Background	3
Courses / Tutorials	5
Reference Materials	5
Manuscripts and Blog Posts	5
Newsletters.....	6
Prompt Engineering	7
Overview	7
Courses / Tutorials	11
Reference Materials	12
Manuscripts and Blog Posts	14
Custom GPTs.....	14
Overview	14
Courses / Tutorials	15
Manuscripts	17
Glossary	22

Overview

This workshop is intended for students and educators in medicine who want a basic overview of LLMs and how they may be useful in medical education. Participants are not expected to have any experience using LLMs; however, we will also provide information that more experienced LLM users may find valuable. Using ChatGPT as an example, the workshop is structured to provide time for participants to practice basic tasks with this technology. The workshop materials are designed to help participants to further learn and experiment with ChatGPT after the workshop has concluded. We have several objectives in this workshop.

- Introduce the fundamentals of ChatGPT
- Illustrate principles for effective prompt engineering
- Explore and evaluate applications of ChatGPT in medical education

Background

GPT is a *large language model* (LLM), which is an *artificial neural network* based on a type of *machine learning* (ML) called *deep learning* (DL). It can perform a range of *natural language processing* (NLP) tasks. GPT models are a type of *artificial intelligence* (AI). Figure 1 shows the nested structure of these technologies.

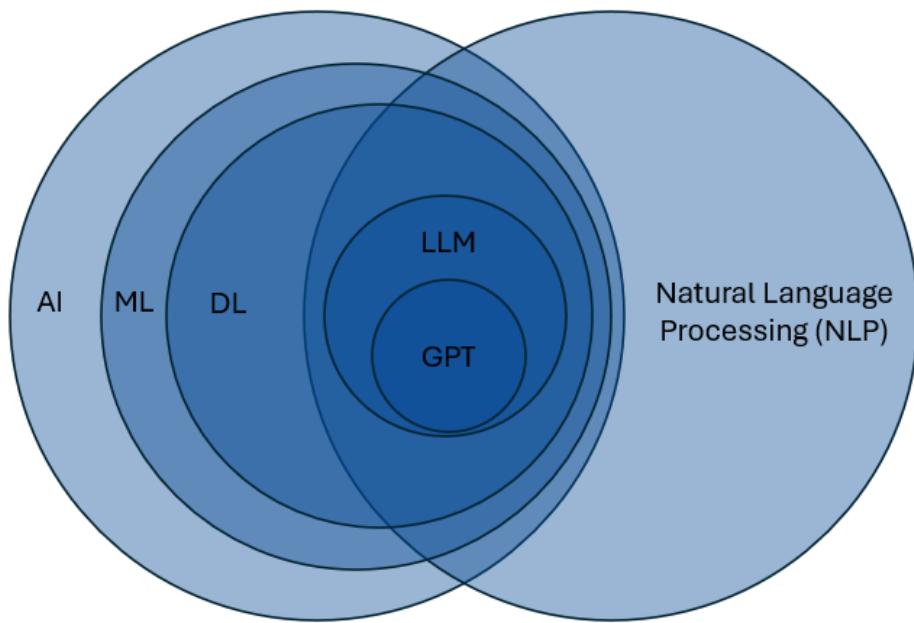


Figure 1. GPT models are language prediction models based on artificial neural networks built on the Transformer architecture. Here, GPT is shown in relation to the other related technologies. (AI = Artificial Intelligence; ML = Machine Learning; DL = Deep Learning; LLM = Large Language Models; GPT = Generative Pretrained Transformers.)

Some LLMs are trained—or *fine-tuned*—for a specific task. ChatGPT is a version of the GPT model that is optimized to interact with users in a conversational manner. Users interact by inputting *prompts*.

Since ChatGPT generates responses based on the prompts that are submitted, this has led to the notion of *prompt engineering*—the practice of producing the most effective prompts for eliciting specific kinds of responses from a chatbot.

Common Applications

Several applications of this technology in medicine have been identified. For example, Dave, Athaluri, and Singh (2023) suggest that ChatGPT may contribute to:

- Identification of research topics
- Diagnosis assistance
- Identify changes and new developments within specific fields of practice or study
- Assist patients in health management

Within medical education specifically, potential applications include:

- Personalized learning
- Skill development through simulation
 - Interactive case studies
 - Communication skills development
 - Ethical and professional development
- Medical knowledge database/resource
- Exam preparation and study assistance

Limitations

Despite these potential benefits, several challenges arise in the use of LLMs that have not yet been fully addressed. These include:

- Accuracy and reliability concerns (hallucinations)
- Bias and ethical concerns (potential for reproducing biases present in the training data)
- Data privacy and security
- Reliance on technology may degrade critical thinking and diagnostic skills development
- Regulatory and legal issues (e.g., liability for misinformation or malpractice)
- Exacerbate educational disparities arising from resource disparities
- Inadequate AI training (additional training required for effective use)
- Ethical use and patient consent
- Limited interpretability ("black boxes")

Courses / Tutorials

Coursera Course:

<https://www.coursera.org/learn/generative-ai-for-everyone>

Academic Medicine Podcast: *Do What You Do Better: Using AI Tools to Ease the Workload Burden on Faculty*

<https://academicmedicineblog.org/do-what-you-do-better-using-ai-tools-to-ease-the-workload-burden-on-faculty/>

AAMC Webinar Series:

<https://aamc.elevate.compartners.com/products/ai-in-academic-medicine-webinar-series>

Reference Materials

Ultimate ChatGPT Toolkit (by Prompts Daily):

<https://www.neatprompts.com/c/chatgpt-toolkit>

OpenAI provides some basic information about ChatGPT:

<https://help.openai.com/en/articles/6783457-what-is-chatgpt>

17 Tips to Take Your ChatGPT Prompts to the Next Level (by Wired)

<https://www.wired.com/story/17-tips-better-chatgpt-prompts/>

Manuscripts and Blog Posts

Newsletters

These daily newsletters span a wide range of AI topics, including leadership and investment in AI enterprises, emerging tools, regulations, and ethical considerations, significant publications alongside lighter research, and resources for training, reference, and quick guides.

The Neuron (<https://www.theneurondaily.com/>)

TL;DR:AI (<https://tldr.tech/ai>)

Superhuman (<https://www.superhuman.ai/>)

Prompts Daily (<https://www.neatprompts.com/>)

Prompt Engineering

Overview

As noted above, ChatGPT generates responses based on *prompts* submitted by users. Prompts guide ChatGPT to focus on certain areas of what it has learned with goal of helping constrain the output to relevant information. Because of the important role prompts play in productively using ChatGPT, practices and strategies have been developed to produce the most effective prompts. These practices and strategies are referred to as *prompt engineering*.

Hidden Prompts

We will not spend much time on so-called hidden prompts but should the term arise, it usually refers to additional prompting that is submitted by an AI system to itself without being revealed to users. In the case of ChatGPT, it claims that OpenAI is transparent about its hidden prompts; but, in any case, these hidden prompts are generally intended to prevent the model from producing offensive output. Or as ChatGPT tells us:

...I do have guidelines I follow to ensure responses are helpful, respectful, and appropriate. These include prioritizing user safety, avoiding the creation or promotion of harmful content, and striving to provide accurate and relevant information.

Pre-prompts

More relevant for our use here is the idea of pre-prompts. Pre-prompts are instructions submitted to ChatGPT to guide its responses to subsequent prompts. Pre-prompts are designed to set the context or guide a model's behavior. Although there may be overlap in some cases, most pre-prompts can be classified into these five categories:

- **Context Setting.** Information that helps the model respond in a manner consistent with a specified scenario or domain. For example, "You're a tour guide," would prepare ChatGPT to answer questions related to tourism, landmarks, cultural information, etc.
- **Behavior Guidelines.** Instructions designed to influence how a model interacts with users. This could involve directives like maintaining a formal tone, avoiding certain topics, or prioritizing safety and appropriateness.
- **Model Configuration:** Parameters or settings adjusted before the interaction that affect the model's performance, such as specifying a particular language, domain expertise (e.g., medical), or creativity level. *Note:* Both *model configuration* and *behavioral guidelines* can influence the tone and focus of the conversation—the

key difference lies in their scope and intention. *Model configuration* is more about technical and operational adjustments for tailoring the model's capabilities and focus areas, whereas *behavior guidelines* are about ethical, social, and content-related boundaries and stylistic choices within those configured capabilities.

- **Session History.** A record of the previous exchanges between the user and the model within the same session. This history helps the model maintain continuity over the course of an interaction, remembering what has been discussed or requested previously. *ChatGPT does this automatically.*
- **Role Play:** Instructions that assign a role or character, such as pretending to be a historical figure, a fictional character, or having a certain profession. This shapes how the model constructs its responses according to the assumed identity.

Pre-prompts are a powerful tool for tailoring the model's outputs to specific needs or contexts. Consider an example.

User: *You are a professor at a medical school. [Context setting pre-prompt]*

ChatGPT: *As your professor at a medical school, I'm here to guide you through the complexities of medicine, from the basic sciences underpinning clinical practice to the nuanced patient care strategies that define exceptional physicians. Whether you need clarification on intricate physiological mechanisms, assistance in navigating the vast expanse of medical research, or advice on applying theoretical knowledge to clinical scenarios, consider me your dedicated resource. Let's embark on this educational journey together, ensuring you're equipped with the knowledge, skills, and ethical grounding to excel in your future medical career. How can I assist you in your studies today?*

User: *Explain this to your students: Walter's contusion hurt. [Main prompt]*

ChatGPT: *When we say "Walter's contusion hurt," we're touching upon a few fundamental aspects of medicine that are critical for medical students to understand. Let's break down this simple statement to explore its medical significance:*

What is a Contusion?

A contusion, commonly known as a bruise, is a type of hematoma of tissue in which capillaries and sometimes venules are damaged by trauma, allowing blood to seep, hemorrhage, or infiltrate into the surrounding interstitial tissues. Unlike cuts or lacerations, the skin is not broken in contusions, making them less

obvious in terms of injury severity but potentially just as painful or harmful, depending on their location and extent.

Pathophysiology of a Contusion:

The mechanism of injury in a contusion involves a blunt force striking the body, leading to compression and damage to the blood vessels beneath the skin's surface. This damage results in bleeding within the tissue, which causes the discoloration and swelling often observed with bruises. The body's response to this injury includes inflammation, which is part of the healing process but also contributes to the sensation of pain.

Pain in Contusions:

Pain in the context of a contusion arises from several factors. The initial trauma can activate nociceptors (pain receptors) in the skin and deeper tissues. The subsequent inflammation increases the sensitivity of these receptors, making the area tender to touch and movement. Additionally, the swelling associated with inflammation can put pressure on surrounding tissues and nerves, further exacerbating the pain.

Clinical Significance:

While most contusions are minor and heal without intervention, they can sometimes indicate more severe underlying issues. For example, a contusion in the abdomen or head can be a sign of internal bleeding or injury to internal organs. Therefore, understanding the presentation and potential complications of contusions is essential for clinical practice.

Management and Treatment:

The initial treatment for a contusion may include rest, ice, compression, and elevation (RICE) to reduce pain and swelling. However, medical evaluation may be necessary for severe cases to rule out more significant injuries or complications.

In explaining "Walter's contusion hurt," we're not only discussing the physiological and pathological aspects of a common injury but also emphasizing the importance of pain as an indicator of bodily harm and the necessity for a thorough assessment to ensure appropriate care. This underscores the broader principle in medicine that understanding the basic mechanisms of injury and the body's response is crucial for effective diagnosis and treatment.

Now, suppose we ask the same question with a very different pre-prompt:

User: You are a kindergarten teacher. [Context setting pre-prompt]

ChatGPT: As your kindergarten teacher, I'm here to create a world of wonder, curiosity, and learning for you. Our classroom is a magical place where we'll explore everything from the alphabet and numbers to the mysteries of nature and the joy of art. Imagine us going on adventures through books, discovering new words, counting stars, painting our dreams, and learning to be kind friends to each other. Every day is a chance to learn something new, make a friend, and grow a little bit more. Whether we're reading stories in our cozy corner or exploring the outdoors, I'm here to guide you with patience and a smile. Let's make today another beautiful day of learning and fun. What would you like to explore today?

User: Explain this to your students: Walter's contusion hurt. [Main prompt]

ChatGPT: Alright, kids, let's talk about something that happened to our friend Walter. Walter got a boo-boo, called a contusion, but you might know it better as a bruise. Have you ever bumped into something really hard and later noticed a purple or blue spot on your skin? That's what happened to Walter.

When we get a bruise, it's because some tiny blood vessels under our skin get squished and break open a little bit. That's why the area turns a different color. And you know what? It can hurt because our body is telling us, "Hey, something happened here!"

But don't worry, our bodies are amazing at fixing these boo-boos. If you or a friend gets a bruise, it's important to be gentle with it. Sometimes, putting something cool on it can make it feel better. And after some time, the color changes from purple to green and then yellow until it's all gone. It's like a little magic trick our body does to heal itself!

So, if you see Walter and he seems a little sad about his bruise, you can tell him it's okay, it will get better soon, and that you're there to play gently with him. It's always nice to be kind and understanding when our friends have ouchies.

We'll talk about pre-prompts further below when we introduce custom GPTs.

Prompt Chaining

Prompt chaining is a technique where a series of prompts or questions are built upon each other to help ChatGPT leverage the session history to produce responses that are more relevant or thorough or that solve a problem step by step. Each prompt in the chain builds on the response to the previous one, allowing for a nuanced exploration of a subject or the completion of a multi-stage task.

It might be helpful to consider an example of prompt chaining focused on diagnosing a condition based on symptoms. This chain demonstrates how a series of questions can guide ChatGPT to narrow down the possible diagnosis through a logical, step-by-step process.

- **Initial Prompt:** "Explain the common causes of chest pain."
- **Follow-up Prompt #1:** "Among the causes you mentioned, which are typically associated with pain on the left side of the chest and worsen with deep breaths?"
- **Follow-up Prompt #2:** "For the causes related to left-sided chest pain that worsens with deep breaths, what additional symptoms might suggest it's pericarditis?"
- **Follow-up Prompt #3:** "Given the possibility of pericarditis, what diagnostic tests would you recommend to confirm the diagnosis?"
- **Follow-up Prompt #4:** "Assuming the diagnosis of pericarditis is confirmed, what are the standard treatment options available?"

These prompts illustrates how prompt chaining can move from general information about a symptom to a more focused inquiry into a specific condition, followed by questions about diagnosis and treatment. Each prompt in the chain builds upon the information provided in the previous responses, which helps ChatGPT provide increasingly specific information relevant to the medical scenario. This prompt chain including ChatGPT's responses is given in the Appendix.

Courses / Tutorials

Open AI provides its own prompt engineering guide:

<https://platform.openai.com/docs/guides/prompt-engineering>

Reference Materials

ChatGPT Prompt Frameworks

Unlock the full potential of ChatGPT

R-T-F

Prompt Example

Facebook Ad Marketer.

Design a compelling Facebook ad campaign to promote a new line of fitness apparel for a sports brand.

Create a storyboard outlining the sequence of ad creatives, including ad copy, visuals, and targeting strategy.

Act as a
[ROLE]

Create a
[TASK]

Show as
[FORMAT]

→ ROLE

→ TASK

→ FORMAT

T-A-G

Prompt Example

The task is to evaluate the performance of team members

Act as a Direct manager and assess the strengths and weaknesses of team members.

Goal is to improve team performance so that the average user satisfaction score moves from 6 to 7.5 in the next quarter.

Define
[TASK]

State the
[ACTION]

Clarify the
[GOAL]

→ TASK

→ ACTION

→ GOAL

B-A-B

Prompt Example

We're nowhere to be seen on SEO rankings

We want to be in top 10 SEO ranking in our niche in 90 days.

Develop a detailed plan for mentioning all the measures we should take also include list of top 20 keywords.

Explain Problem
[BEFORE]

State Outcome
[AFTER]

Ask for the
[BRIDGE]

→ BEFORE

→ AFTER

→ BRIDGE

C-A-R-E

Prompt Example

We are launching a new line of sustainable clothing

Can you assist us in creating a targeted advertising campaign that emphasizes our environmental commitment?

Our desired outcome is to drive product awareness and sales

A good example of a similar successful initiative is Patagonia's "Don't Buy This Jacket" campaign, which highlighted their commitment to sustainability while enhancing their brand image.

Give the
[CONTEXT]

Describe
[ACTION]

Clarify the
[RESULT]

Give the
[EXAMPLE]

→ CONTEXT

→ ACTION

→ RESULT

→ EXAMPLE

R-I-S-E

Prompt Example

Imagine you are a content strategist.

I've gathered detailed information about our target audience, including their interests & common questions related to our industry.

Provide a Step by Step content strategy plan identifying key topics based on our audience insights, creating an editorial calendar, and drafting engaging content that aligns with our brand message.

Aim is to increase our blog's monthly visitors by 40% and Enhance our brand's position as a thought leader in our industry.

Specify the
[ROLE]

Describe
[INPUT]

Ask for
[STEPS]

Describe the
[EXPECTATION]

→ ROLE

→ INPUT

→ STEPS

→ EXPECTATION

Created by:



Khizer Abbas, Growth Marketing specialist

ChatGPT Prompting Cheat Sheet

Use this Cheat Sheet to master prompting

Act as a [Role]	Create a [Task]	Show as [Format]			
<ul style="list-style-type: none">•Resume Writer•Marketer•Resume Reviewer•Teacher•LinkedIn Expert•HR Manager•Copywriter•Interviewer•Talent Acquisition Manager	<ul style="list-style-type: none">•Ghost Writer•Website Designer•Resume Reviewer•Accountant•Prompt Engineer•Financial Analyst•Therapist•Journalist•Investor	<ul style="list-style-type: none">•Resume from scratch•Resume Analysis•Attractive Cover Letter•Detailed LinkedIn Analysis•Short Job Description•Interview Questions•Headline•Article•Essay	<ul style="list-style-type: none">•Sales Copy•Analysis•SEO Keywords•Blog Post•Summary•Transcription•Recipe•Webpage•Book outline	<ul style="list-style-type: none">•A Table•A List•Summary•A Bullet Points•Spreadsheet•Plain Text File•Word Document PDF•Graphs	<ul style="list-style-type: none">•XML•Rich text•JSON•An Analogy

HOW TO BUILD A CHAIN PROMPT WITH EXAMPLE

1. Insert first prompt: Give me a summary of this document [insert or copy paste document text]
2. Modify the output: Use the summary above and write a 500 word piece that explains the topic to beginners
3. Modify the tone: Change the tone of the answer above and make it sound more professional
4. Modify the format: Convert the answer above into text for a presentation with 1 slide for each key point

GENERAL PROMPTS

Rewrite this text and make it easy for a beginner to understand: [insert text].
I want to [insert task or goal]. Generate 5 for ideas for [insert task or goal].
Explain [insert topic] in simple and easy terms that any beginner can understand.
Summarize the text below and give me a list of bullet points with key insights and the most important facts.
Proofread my writing above. Fix grammar and spelling mistakes. And make suggestions to improve the clarity of my writing.
Can you provide a detailed explanation of [Topic]? Please include key concepts, historical context, and current relevance.

PROMPTS FOR MARKETERS	PROMPTS FOR DEVELOPERS
<p>List [insert number] ideas for blog posts about [insert topic] Create a 30 day social media calendar about [insert topic] Generate landing page copy for [insert product description] Write 5 pieces of Facebook ad copy for [product description] Generate 5 persuasive subject lines for an email about [insert email description]</p>	<p>Help me find mistakes in my code: [insert your code] Explain what this snippet of code does: [insert code snippet] What is the correct syntax for a [statement or function] in [programming language]? How do I fix the following [programming language] code which [explain the functioning]? [insert code snippet]</p>
PROMPTS FOR SALES	PROMPTS FOR DESIGNERS
<p>Generate 10 ways to generate leads for [product description] Create a personalized sales email for potential customers. Include [topic, brand name, promo offers, etc.] Write a sales landing page description for [product description] Generate 5 personas I should include in my outreach for [X] Generate a script to use when cold-calling [insert persona]</p>	<p>What are some interactions to consider when designing a [insert app or website description] Create a user persona for [describe product] Generate 10 questions for a user interview regarding [topic] Create a user journey for [insert app and persona description] Generate UI/UX design requirements for [describe feature]</p>
PROMPTS FOR RESEARCH	PROMPTS FOR CUSTOMER SERVICE
<p>Identify the top 20 companies in [insert industry] by revenue What are the top trends in [insert industry] for 2023? Find me the best-reviewed software for [insert task] Summarize the annual financial statement of [insert company] Summarize this research paper and give me a list of the key insights: [insert research paper text]</p>	<p>Create a template for an email response to customers inquiring about [product]. What are the most frequently asked questions about [topic]? Create a help page that explains how to use [your product]. Summarize the following knowledge base article to give step-by-step instructions: [insert article]</p>

10 Best Prompting Tools

[PromptDr/ve](#)

[Geniea](#)

[CreativAI](#)

[Public Prompts](#)

[Prompt Perfect](#)

[Promptist](#)

[Maker Box](#)

[Trickle](#)

[PromptBase](#)

[PromptInterface.ai](#)

Manuscripts and Blog Posts

Principled Instructions Are All You Need for Questioning LLaMA-1/2, GPT-3.5/4 (authors? Years? How do I want to show this? Summary? ChatGPT summary?)

<https://arxiv.org/pdf/2312.16171v1.pdf>

Custom GPTs

Overview

Custom GPTs are specialized versions of the GPT models, tailored for specific tasks, industries, or data sensitivities. They're built upon the foundation of large, general-purpose GPT models developed by organizations like OpenAI, but are fine-tuned or customized to better suit particular use cases.

OpenAI's platform provides a flexible and powerful way to create custom GPT models. Users can build highly specialized applications using OpenAI's tools and APIs that facilitate this customization process by enabling fine-tuning of models using specific datasets or instruction-based approaches. Here's how custom GPTs work in the context of OpenAI:

1. Access to OpenAI API: The journey typically begins with access to the OpenAI API, which provides programmatic access to various GPT models. OpenAI's API supports a range of models with different capabilities and sizes, allowing users to select the one that best fits their requirements in terms of complexity, speed, and cost.
2. Choice of customization approach:
 - a. Fine-tuning: OpenAI allows users to create custom versions of their models through fine-tuning, where the user provides a dataset specific to their domain or task. This dataset is used to adjust the model's parameters specifically for the user's requirements, enhancing its performance on relevant tasks.
 - b. Instruct GPT: Another approach is to use variants of the models that are optimized to follow instructions in natural language. This allows users to guide the model's outputs through prompts without needing a custom dataset or fine-tuning process.
3. Dataset preparation: For fine-tuning, the user needs to prepare a dataset that is representative of the task or domain. This involves collecting, cleaning, and possibly annotating text data. The quality and relevance of this dataset are crucial for effective customization.
4. Fine-tuning process: If the user opts for fine-tuning, they will use the dataset to train a custom model. OpenAI provides guidelines and tools to assist with this process, which

involves submitting the dataset through the API and running the fine-tuning process. This adjusts the model's weights to better suit the specific data and tasks.

5. Testing and evaluation: After fine-tuning, the custom model is tested to ensure it meets the desired criteria for performance, such as accuracy, coherence, and relevance. OpenAI's tools enable users to evaluate model outputs and make adjustments as necessary.
6. Integration and deployment: Once satisfactory, the custom model can be integrated into applications or services through the OpenAI API. This involves calling the API with specific prompts or instructions and handling the model's responses in the application's context.
7. Maintenance and updates: Over time, it may be necessary to update the custom model to address changing requirements or improve performance. This could involve re-fine-tuning the model with new data or adjusting the way prompts are constructed.

Courses / Tutorials

For those interested in learning how to create and implement custom GPT models, particularly with OpenAI, there are several high-quality resources available from both official OpenAI documentation/tutorials and third-party courses and guides.

Official OpenAI Resources

1. OpenAI documentation: The best starting point is the OpenAI API documentation. It provides comprehensive guides, API references, and best practices for using GPT models. The documentation includes sections on fine-tuning, prompt engineering, and safety best practices.
2. OpenAI quickstart guides: OpenAI often publishes quickstart guides and tutorials that help new users get up and running with their API. These guides cover a range of topics from basic usage to more advanced topics like fine-tuning.
3. OpenAI community and workshops: OpenAI occasionally hosts workshops and webinars on how to effectively use their models, including custom GPTs. Additionally, the OpenAI Community Forum is a valuable resource where users share tips, tutorials, and support each other in solving specific problems.

Third-party Courses and Tutorials

1. Hugging Face's transformer library: Although not a course, the Hugging Face Transformers library documentation includes tutorials and guides for working with transformer models, including GPT. They often provide easy-to-follow notebooks that show how to fine-tune models on custom datasets.
2. Coursera and Udemy: Platforms like Coursera and Udemy offer courses on AI and machine learning that sometimes cover topics related to GPT and custom language models. Courses such as "Natural Language Processing with Transformers" can be particularly useful.

3. YouTube tutorials: YouTube hosts a wealth of tutorials on GPT and custom models. Channels like Sentdex, Hugging Face, and others often post high-quality videos that guide you through the process of working with GPT models, including coding tutorials and conceptual explanations.
4. Medium and Towards Data Science: Many data scientists and developers write about their experiences with custom GPT models on blogging platforms like Medium and Towards Data Science. These articles can provide practical insights and code snippets for specific use cases or challenges.
5. GitHub repositories: Exploring GitHub for projects related to GPT and fine-tuning can be incredibly educational. Many developers share their custom GPT projects, complete with code, documentation, and sometimes even blog posts explaining their process.
6. When choosing a resource, consider your learning style and the specific aspects of custom GPT models you are interested in, such as fine-tuning, prompt engineering, or application integration. Keep in mind that the field of AI and the capabilities of GPT models are rapidly evolving, so it is beneficial to look for the most current resources or those that are regularly updated.

Manuscripts

Medical Education AI Articles:

Stone, Reel, Farrell (2021) Artificial intelligence meets clinical skills teaching

<https://doi.org/10.1111/medu.14636>

Gin, ten Cate, O'Sullivan, Hauer, Boscardin (2021) Exploring how feedback reflects entrustment decisions using artificial intelligence <https://doi.org/10.1111/medu.14696>

Johnsson et al (2023) Validity evidence supporting clinical skills assessment by artificial intelligence compared with trained clinician raters <https://doi.org/10.1111/medu.15190>

Kripalani, Grimmer, Wang (2023) Med versus machine: Using ChatGPT in team-based learning <https://doi.org/10.1111/medu.15226>

Hornos et al (2024) Online clinical reasoning simulator for medical students grounded on dual-process theory <https://doi.org/10.1111/medu.15322>

Fawns, Schuwirth (2023) Rethinking the value proposition of assessment at a time of rapid development in generative artificial intelligence <https://doi.org/10.1111/medu.15259>

van der Niet, Bleakley (2020) Where medical education meets artificial intelligence: 'Does technology care?' <https://doi.org/10.1111/medu.14131>

Webster (2020) Artificial intelligence and the adoption of new technology in medical education <https://doi.org/10.1111/medu.14409>

Pearce, Chiavaroli (2023) Rethinking assessment in response to generative artificial intelligence <https://doi.org/10.1111/medu.15092>

Ng, Tay, Wilding, Ng, Han (2024) Transforming curriculum mapping: A human-AI hybrid approach <https://doi.org/10.1111/medu.15331>

Medical Teacher:

Masters (2019) Artificial intelligence in medical education

<https://doi.org/10.1080/0142159X.2019.1595557>

Hess, Cupido, Ross, Kvern (2023) Becoming adaptive experts in an era of rapid advances in generative artificial intelligence <https://doi.org/10.1080/0142159X.2023.2289844>

Mihalache, Huang, Popovic, Muni (2023) ChatGPT-4: An assessment of upgraded artificial intelligence chatbot in the USMLE <https://doi.org/10.1080/0142159X.2023.2249588>

Zuckerman et al (2023) ChatGPT for assessment writing

<https://doi.org/10.1080/0142159X.2023.2249239>

Wang, Paidisetty, Cano (2023) The next paradigm shift? ChatGPT, artificial intelligence, and medical education <https://doi.org/10.1080/0142159X.2023.2198663>

Costa-Dookhan, Maslej, Donner, Islam, Sockalingam, Thakur (2023) Twelve tips for NLP in medical education program evaluation <https://doi.org/10.1080/0142159X.2024.2316223>

Divito, Katchikian, Gruenwald, Burgoon (2023) The tools of the future are the challenges of today: The use of ChatGPT in problem-based learning medical education

<https://doi.org/10.1080/0142159X.2023.2290997>

Gordon at al (2024) A scoping review of artificial intelligence in medical education: BEME Guide No. 84 <https://doi.org/10.1080/0142159X.2024.2314198>

Masters (2023) Ethical use of Artificial Intelligence in Health Professions Education: AMEE Guide No. 158 <https://doi.org/10.1080/0142159X.2023.2186203>

Kiyak, Kononowicz (2024) Case-based MCQ generator: A custom ChatGPT based on published prompts in the literature for automatic item generation

<https://doi.org/10.1080/0142159X.2024.2314723>

Masters (2023) *Medical Teacher's* first ChatGPT's referencing hallucinations: Lessors for editors, reviewers, and teachers <https://doi.org/10.1080/0142159X.2023.2208731>

Masters et al (2024) Twelve tips on creating and using custom GPTs to enhance health professions education <https://doi.org/10.1080/0142159X.2024.2305365>

Indran, Paranthaman, Gupta, Mustafa (2023) Twelve tips to leverage AI for efficient and effective medical question generation: A guide for educators using ChatGPT

<https://doi.org/10.1080/0142159X.2023.2294703>

Academic Medicine:

Cheung et al (2022) Optimizing Clinical Reasoning Assessments with Analytic and Holistic Ratings: Examining the Validity, Reliability, and Cost of a Simplified Patient Note Scoring Procedure <https://doi.org/10.1097/ACM.0000000000004908>

Booth et al (2023) Competency-Based Assessments: Leveraging Artificial Intelligence to Predict Subcompetency Content <https://doi.org/10.1097/ACM.0000000000005115>

Van de Riddler et al (2023) Finding the Place of ChatGPT in Medical Education <https://doi.org/10.1097/ACM.0000000000005254>

Feng, Shen (2023) ChatGPT and the Future of Medical Education
<https://doi.org/10.1097/ACM.0000000000005242>

Bair, Norden (2023) Large Language Models and Their Implications on Medical Education
<https://doi.org/10.1097/ACM.0000000000005265>

Turner, Hashimoto, Vasisht, Schaye (2023) Demystifying AI: Current State and Future Role in Medical Education Assessment <https://doi.org/10.1097/ACM.0000000000005598>

Ray (2024) ChatGPT: A Game Changer in Academic Medicine's Artificial Intelligence Revolution
<https://doi.org/10.1097/ACM.0000000000005486>

Hoang (2024) How Automation and ChatGPT Will Affect Competency Definitions and Assessment Tools <https://doi.org/10.1097/ACM.0000000000005483>

Boscardin, Gin, Golde, Hauer (2024) ChatGPT and Generative Artificial Intelligence for Medical Education: Potential Impact and Opportunity <https://doi.org/10.1097/ACM.0000000000005439>

Patino et al (2024) The Promise and Perils of Artificial Intelligence in Health Professions Education Practice and Scholarship <https://doi.org/10.1097/ACM.0000000000005636>

Chen, Chen, Liu (2024) Impact of Democratizing Artificial Intelligence: Using ChatGPT in Medical Education and Training <https://doi.org/10.1097/ACM.0000000000005672>

Hashimoto, Johnson (2023) The Use of Artificial Intelligence Tools to Prepare Medical School Applications <https://doi.org/10.1097/ACM.0000000000005309>

Assorted:

Chavez et al (2023) Chat Generative Pre-trained Transformer: why we should embrace this technology <https://doi.org/10.1016/j.ajog.2023.03.010>

Zhang, Pei, Zhen, Li, Liang (2023) Chat Generative Pre-Trained Transformer (ChatGPT) usage in healthcare <https://doi.org/10.1016/j.gande.2023.07.002>

Sauder, Tritsch, Rajput, Schwartz, Shoja (2024) Exploring Generative Artificial Intelligence-Assisted Medical Education: Assessing Case-Based Learning for Medical Students
<https://doi.org/10.7759%2Fcureus.51961>

Moritz, Romeike, Stosch, Tolks (2023) Generative AI (gAI) in medical education: Chat-GPT and co <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10407583/pdf/JME-40-54.pdf>

Li, Li (2023) Generative artificial intelligence in medical education: way to solve problems
<https://doi.org/10.1093/postmj/qgad116>

Heng, Teo, Tan (2023) The impact of Chat Generative Pre-trained Transformer (ChatGPT) on medical education <https://doi.org/10.1093/postmj/qgad058>

Shoja, Van de Ridder, Rajput (2023) The Emerging Role of Generative Artificial Intelligence in Medical Education, Research, and Practice [10.7759/cureus.40883](https://doi.org/10.7759/cureus.40883)

Karabacak, Ozkara, Margetis, Wintermark, Bisdas (2023) The Advent of Generative Language Models in Medical Education <https://mededu.jmir.org/2023/1/e48163>

Eysenbach (2023) The Role of ChatGPT, Generative Language Models, and Artificial Intelligence in Medical Education: A Conversation with ChatGPT and a Call for Papers <https://mededu.jmir.org/2023/1/e46885/>

Nagi et al (year) Applications of Artificial Intelligence (AI) in Medical Education: A Scoping Review <https://pubmed.ncbi.nlm.nih.gov/37387115/>

Cooper, Rodman (2023) AI and Medical Education – A 21st-Century Pandora’s Box <https://www.nejm.org/doi/full/10.1056/NEJMp2304993>

Zhang et al (2023) AI in Medical Education: Global situation, effects and challenges <https://link.springer.com/article/10.1007/s10639-023-12009-8>

Tsang (2023) Practical Applications of ChatGPT in Undergraduate Medical Education <https://doi.org/10.1177/23821205231178449>

Preiksaitis, Rose (2023) Opportunities, Challenges, and Future Directions of Generative Artificial Intelligence in Medical Education: Scoping Review <https://doi.org/10.2196/48785>

AAMC resource bundles: <https://communities.aamc.org/resource-bundles/artificial-intelligence>

AI NEJM: <https://ai.nejm.org/>

<https://www.medbiq.org/initiatives/international-advisory-committee-artificial-intelligence>

AMIE: A research AI system for diagnostic medical reasoning and conversations :

https://blog.research.google/2024/01/amie-research-ai-system-for-diagnostic_12.html

<https://www.forbes.com/sites/williamhaseltine/2023/12/15/medical-text-written-by-artificial-intelligence-outperforms-doctors/?sh=5207f37679ca>

Need to make sure all these are included:

Manuscripts

Masters, K., Benjamin, J., Agrawal, A., MacNeill, H., Pillow, M. T., & Mehta, N. (2024). Twelve tips on creating and using custom GPTs to enhance health professions education. *Medical Teacher*, 1-5.

Benítez, T. M., Xu, Y., Boudreau, J. D., Kow, A. W. C., Bello, F., Van Phuoc, L., Wang, X., Sun, X., Leung, G. K., Lan, Y., Wang, Y., Cheng, D., Tham, Y. C., Wong, T. Y., & Chung, K. C. (2024). Harnessing the potential of large language models in medical education: promise and pitfalls. *Journal of the American Medical Informatics Association : JAMIA*, 31(3), 776–783.

Bakkum, M. J., Hartjes, M. G., Piët, J. D., Donker, E. M., Likic, R., Sanz, E., de Ponti, F., Verdonk, P., Richir, M. C., van Agtmael, M. A., & Tichelaar, J. (2024). Using artificial intelligence to create diverse and inclusive medical case vignettes for education. *British journal of clinical pharmacology*, 90(3), 640–648.

Lower, K., Seth, I., Lim, B., & Seth, N. (2023). ChatGPT-4: Transforming Medical Education and Addressing Clinical Exposure Challenges in the Post-pandemic Era. *Indian journal of orthopaedics*, 57(9), 1527–1544.

Holderried, F., Stegemann-Philipps, C., Herschbach, L., Moldt, J. A., Nevins, A., Griewatz, J., Holderried, M., Herrmann-Werner, A., Festl-Wietek, T., & Mahling, M. (2024). A Generative Pretrained Transformer (GPT)-Powered Chatbot as a Simulated Patient to Practice History Taking: Prospective, Mixed Methods Study. *JMIR medical education*, 10, e53961.

Kiyak, Y. S., & Kononowicz, A. A. (2024). Case-based MCQ generator: A custom ChatGPT based on published prompts in the literature for automatic item generation. *Medical teacher*, 1–3. Advance online publication.

Preiksaitis, C., & Rose, C. (2023). Opportunities, Challenges, and Future Directions of Generative Artificial Intelligence in Medical Education: Scoping Review. *JMIR medical education*, 9, e48785.

Milad, D., Antaki, F., Milad, J., Farah, A., Khairy, T., Mikhail, D., Giguère, C. É., Touma, S., Bernstein, A., Szigiato, A. A., Nayman, T., Mullie, G. A., & Duval, R. (2024). Assessing the medical reasoning skills of GPT-4 in complex ophthalmology cases. *The British journal of ophthalmology*, bjo-2023-325053. Advance online publication.

Glossary

large language model (LLM): a type of machine learning model called a neural network that is designed to understand, generate, and work with human language. These models are trained on vast amounts of text data from the corpus they are trained on and can perform a wide range of language-related tasks.

fine-tuning: the process of further training a pre-trained model on a smaller dataset related to a specific activity or domain in order to improve the model's performance on tasks related to that activity or domain.

prompt: an input statement or question that guides an AI system's response generation. Prompts help the system determine the direction and boundaries of its output, influencing its creativity, style, and relevance.

prompt engineering: the practice of producing effective prompts. Prompt engineering may involve designing, testing, and refining prompts in order to improve effectiveness.

prompt chain: A series of prompts built upon each other that allows ChatGPT to produce responses informed by the session history.

Stuff to fit in somewhere / yet to be organized:

<https://learn.microsoft.com/en-us/azure/ai-services/openai/concepts/advanced-prompt-engineering?pivot=programming-language-chat-completions>

Medprompt:

<https://github.com/microsoft/promptbase?tab=readme-ov-file>

Supporting paper:

<https://arxiv.org/pdf/2311.16452.pdf>

MMedLM “Multilingual Language Model for Medicine”

<https://github.com/magic-ai4med/mmedlm>

<https://github.com/microsoft/promptbase?tab=readme-ov-file>

More research needed to contextualize this: <https://github.com/microsoft/promptbase>

Appendix

An example of a prompt chain.

User: Explain the common causes of chest pain. [**Initial Prompt**]

ChatGPT: Chest pain is a symptom that can be attributed to a range of conditions, some of which are life-threatening, while others are relatively benign. Understanding the common causes of chest pain is crucial in the medical field, not only for diagnosis but also for determining the urgency and type of treatment required. Here are several common causes of chest pain, categorized by their primary systems:

Cardiac Causes

1. Myocardial Infarction (Heart Attack): This occurs when blood flow to a part of the heart is blocked for a long enough time that part of the heart muscle is damaged or dies.
2. Angina Pectoris: Caused by reduced blood flow to the heart muscle, leading to chest pain or discomfort, often triggered by physical activity or stress.
3. Pericarditis: Inflammation of the pericardium, the sac-like covering of the heart, which can cause sharp, stabbing chest pain that may feel worse with deep breathing or lying down.
4. Aortic Dissection: A serious condition involving a tear in the aorta's inner layer, causing severe chest or back pain.

Pulmonary Causes

1. Pulmonary Embolism: A blood clot that has traveled to the lungs, blocking blood flow and causing chest pain, often with shortness of breath.
2. Pneumothorax: Also known as a collapsed lung, this occurs when air leaks into the space between the lung and chest wall, causing chest pain and difficulty breathing.
3. Pneumonia: An infection that inflames the air sacs in one or both lungs, which can fill with fluid or pus, causing chest pain, especially when you cough or breathe deeply.

Gastrointestinal Causes

1. Gastroesophageal Reflux Disease (GERD): A condition where stomach acid frequently flows back into the tube connecting your mouth and stomach (esophagus), causing a burning chest pain.
2. Peptic Ulcers: Open sores that develop on the inside lining of your stomach and the upper portion of your small intestine, causing a burning stomach pain that may extend up to your chest.

3. Gallstones: Hardened deposits in the gallbladder that can block the flow of bile, causing pain that may radiate to the chest.

Musculoskeletal Causes

1. Costochondritis: Inflammation of the cartilage that connects a rib to the breastbone (sternum), causing chest pain that may worsen with movement or deep breathing.
2. Muscle Strain: Overuse or injury to the chest muscles can cause chest pain.

Psychological Causes

1. Panic Attacks: These can mimic heart attack symptoms, including chest pain, heart palpitations, and shortness of breath.

It's essential for healthcare professionals to conduct a thorough history and physical examination, and sometimes diagnostic testing, to determine the cause of chest pain due to the wide range of possible etiologies. The management and treatment will vary significantly depending on the underlying cause.

User: Among the causes you mentioned, which are typically associated with pain on the left side of the chest and worsen with deep breaths? [Follow-up Prompt #1]

ChatGPT: Pain on the left side of the chest that worsens with deep breaths can be particularly concerning, as it might indicate conditions that require prompt medical attention. Among the causes I mentioned, the following are more commonly associated with such symptoms:

1. Pericarditis: Inflammation of the pericardium, the two thin layers of a sac-like tissue that surrounds the heart, can cause sharp, stabbing pain on the left side or middle of the chest. The pain often worsens with deep breathing, lying flat, and may improve when sitting up and leaning forward.
2. Pneumothorax (Collapsed Lung): Although pneumothorax can cause pain that is felt on either side of the chest depending on where the collapse occurs, it can cause sharp, sudden chest pain on the left side if the left lung is affected. The pain can worsen with deep breaths due to the increased pressure on the affected lung.
3. Pulmonary Embolism: A blood clot that travels to the lungs can cause chest pain that is often sharp and may become worse with deep breaths, coughing, or even eating. The pain is usually sudden in onset and can be accompanied by shortness of breath, rapid heart rate, and sometimes coughing up blood.

It's important to note that while these conditions commonly present with pain on the left side of the chest that worsens with deep breaths, chest pain can be a symptom of various other

conditions as well, including myocardial infarction (heart attack) which typically requires immediate medical attention. Therefore, any new, severe, or persisting chest pain should be evaluated by a healthcare professional to rule out serious conditions.

User: For the causes related to left-sided chest pain that worsens with deep breaths, what additional symptoms might suggest it's pericarditis? [Follow-up Prompt #2]

ChatGPT: Pericarditis, the inflammation of the pericardium (the fibrous sac surrounding the heart), presents with several characteristic symptoms in addition to left-sided chest pain that worsens with deep breaths. Recognizing these additional symptoms can help suggest a diagnosis of pericarditis. They include:

1. Characteristic Chest Pain:
 - a. The pain is typically sharp and stabbing.
 - b. It often worsens with lying down or taking deep breaths and may improve when sitting up and leaning forward.
2. Fever: Mild fever is common in pericarditis, reflecting the inflammatory process.
3. Pericardial Friction Rub: A distinctive, scratchy, grating sound, best heard with a stethoscope on the chest, resulting from the pericardial layers rubbing against each other.
4. Heart Palpitations: Some individuals may experience irregular heartbeats or a feeling of the heart "fluttering," which can occur due to the irritation of the heart's surrounding tissue.
5. Shortness of Breath (Dyspnea): Especially when reclining; this symptom arises from the difficulty in breathing due to the pain or accumulation of fluid in the pericardial space (pericardial effusion), which can compress the heart (cardiac tamponade).
6. Swelling: Swelling in the legs or abdomen might occur if pericarditis leads to pericardial effusion, affecting heart function and fluid circulation.
7. Fatigue: General malaise and tiredness can also accompany pericarditis, reflecting the body's response to inflammation.

It's important to note that while these symptoms can suggest pericarditis, diagnostic confirmation usually requires an echocardiogram, ECG, blood tests for inflammatory markers, and sometimes additional imaging or procedures to assess for fluid around the heart and its effects. Treatment of pericarditis often involves managing the inflammation and pain, as well as addressing any underlying causes.

User: Given the possibility of pericarditis, what diagnostic tests would you recommend to confirm the diagnosis? [Follow-up Prompt #3]

ChatGPT: If pericarditis is suspected based on the clinical presentation, several diagnostic tests can be utilized to confirm the diagnosis, assess the severity, and identify possible underlying causes. The recommended diagnostic tests for pericarditis include:

1. **Electrocardiogram (ECG or EKG):** An ECG is often the first test done. It can show specific changes in the heart's electrical pattern that suggest inflammation of the pericardium. Changes in the ECG pattern can vary depending on the stage of pericarditis.
2. **Echocardiogram:** This ultrasound examination of the heart can visualize the heart's structure and function, checking for the presence of pericardial effusion (fluid in the pericardial space) and assessing the heart's pumping action to see if the inflammation has affected the heart's ability to function properly.
3. **Chest X-ray:** While a chest X-ray cannot diagnose pericarditis directly, it can show if the heart is enlarged, which might suggest a pericardial effusion, or rule out other causes of chest pain, such as pneumonia or a pneumothorax.
4. **Blood Tests:** Various blood tests can be useful. Tests for markers of inflammation (like C-reactive protein (CRP) and erythrocyte sedimentation rate (ESR)), white blood cell count, and cardiac enzymes (which might be elevated in cases where the heart muscle has been affected) can support the diagnosis. Blood cultures and tests for specific antibodies might also be done to investigate infectious or autoimmune causes.
5. **Cardiac MRI (Magnetic Resonance Imaging):** Cardiac MRI can provide detailed images of the heart's structure and function, offering valuable information on the extent of inflammation and helping to rule out other heart conditions.
6. **CT Scan (Computed Tomography):** A CT scan can be useful for diagnosing pericarditis, particularly if complications are suspected, such as a pericardial effusion or constrictive pericarditis. It can also help rule out other causes of chest pain, such as aortic dissection.

In some cases, pericardiocentesis (draining fluid from the pericardial space with a needle under careful monitoring) might be performed, both as a therapeutic and diagnostic procedure to analyze the pericardial fluid for signs of infection, cancer cells, or other underlying conditions.

The choice of diagnostic tests will depend on the clinical situation, including the severity of symptoms, the presence of possible complications, and the suspected underlying cause of pericarditis.

User: Assuming the diagnosis of pericarditis is confirmed, what are the standard treatment options available? [Follow-up Prompt #4]

ChatGPT: Once the diagnosis of pericarditis is confirmed, the treatment plan typically focuses on relieving symptoms and addressing the underlying cause, if known. The standard treatment options for pericarditis include:

Medications

1. **Anti-inflammatory Drugs:** The primary treatment for acute pericarditis involves non-steroidal anti-inflammatory drugs (NSAIDs), such as ibuprofen, aspirin, or indomethacin, to reduce inflammation and relieve pain. The choice of NSAID and the duration of treatment depend on the patient's symptoms and response to therapy.
2. **Colchicine:** Often used in addition to NSAIDs, colchicine can reduce the inflammation of the pericardium and decrease the risk of recurrent pericarditis. It's recommended for most patients with acute pericarditis, barring contraindications.
3. **Corticosteroids:** In cases where NSAIDs and colchicine are ineffective or not recommended (e.g., due to contraindications or side effects), corticosteroids such as prednisone may be used. They are also indicated in pericarditis caused by autoimmune diseases.
4. **Antibiotics:** If the pericarditis is determined to be bacterial, antibiotics will be prescribed based on the specific type of bacteria identified.

Additional Treatments

1. **Pericardiocentesis:** This procedure involves the removal of excess fluid from the pericardial space (if present and causing significant symptoms or cardiac tamponade) using a needle and catheter. It can provide immediate relief of symptoms related to the fluid pressure on the heart.
2. **Pericardectomy:** Rarely, in cases of recurrent or constrictive pericarditis that do not respond to medical therapy, surgical removal of the entire pericardium (pericardectomy) may be considered to relieve symptoms and improve heart function.

Monitoring and Follow-up

Regular follow-up with a healthcare provider is crucial to monitor the response to treatment, adjust medications as necessary, and detect any complications early. The duration of treatment can vary, ranging from a few weeks to several months, depending on the severity of the condition, the response to treatment, and whether there are recurrent episodes.

Lifestyle modifications, such as rest and avoiding strenuous activities, may be recommended during the acute phase to reduce symptoms and prevent exacerbation of the condition.

The choice of treatment strategy is tailored to each patient's specific situation, taking into account the cause of pericarditis, the presence of complications, and any underlying health conditions.