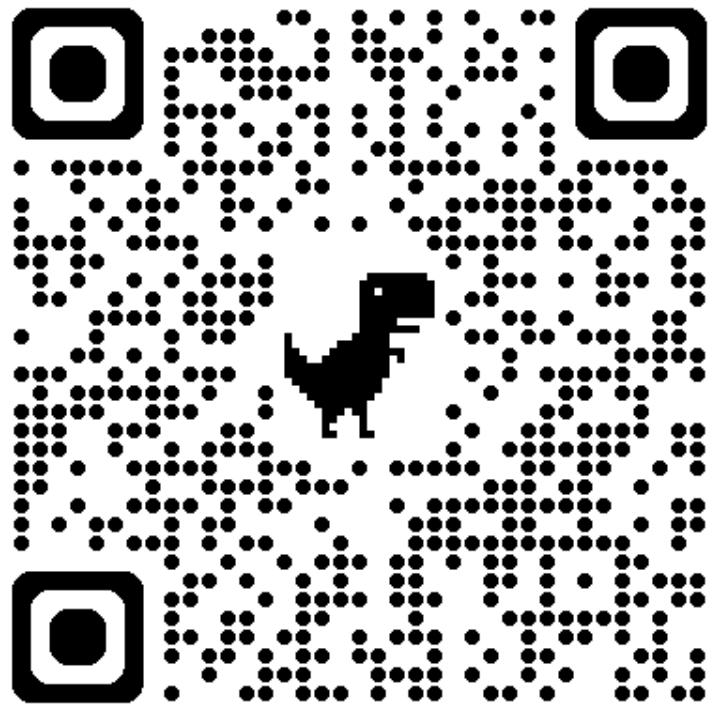
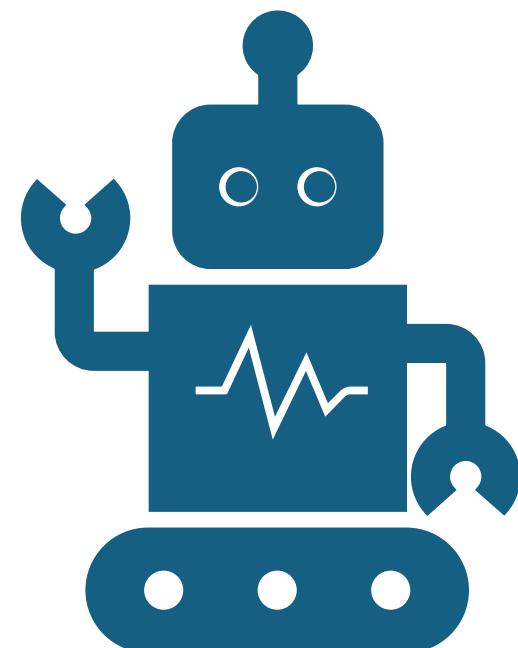


Link to today's presentation
(and related materials)

A Primer for Using Generative Artificial Intelligence in Medical Education

Christopher Runyon, Senior Measurement Scientist
NICE Conference
June 4, 2025





Learning Objectives

- Understand the fundamentals of generative AI and its operational mechanisms.
- Apply effective prompt engineering for educational purposes.
- Explore applications of large language models (LLMs) in medical education.

A few technical details...



Not just what to do,
but why to do it

What is a GPT?

- **G = Generative**
 - refers to the model's ability to create new content
 - will *always* produce a response
- **P = Pretrained**
 - model has undergone an initial extensive training phase on a vast corpus of text data; equips the model with a broad 'understanding' of language and its nuances
- **T = Transformer**
 - aspect of the model architecture (not discussed today)
 - attention layer (will be discussed today)

How LLMs Make Sense of Language

- Underlying principle: distributional semantics / co-occurrence
 - Words that occur in the same contexts tend to have similar meanings. (Harris, 1954)

“We found a little, hairy [REDACTED] sleeping behind the tree.”

- A very, *very* fancy text calculators
- *Always* will produce a response
 - Hallucinations = inaccuracies (all types)
- Not deterministic; variation is part of its beauty



Attention

- The attention mechanism in a Gen AI models helps the model focus on the most relevant words in the sequence (i.e., the prompt).

“We found a little, hairy [REDACTED] sleeping behind the tree.”

- Words like “little”, “hairy”, and “sleeping” provide important information about the wampimuk.
- Other words less so: “we”, “found”, “behind”, “tree”
- Others still even less: “a” “the”

Attention

- LLMs, like humans, have limited attention
- Prompts that give LLMs many tasks at once may result in suboptimal performance.
 - Difficult to identify what are the most important parts of the task.
 - May do some tasks well and others poorly; or all only moderately well.
 - Some newer models (“reasoning models”) are trying to explicitly address this limitation.

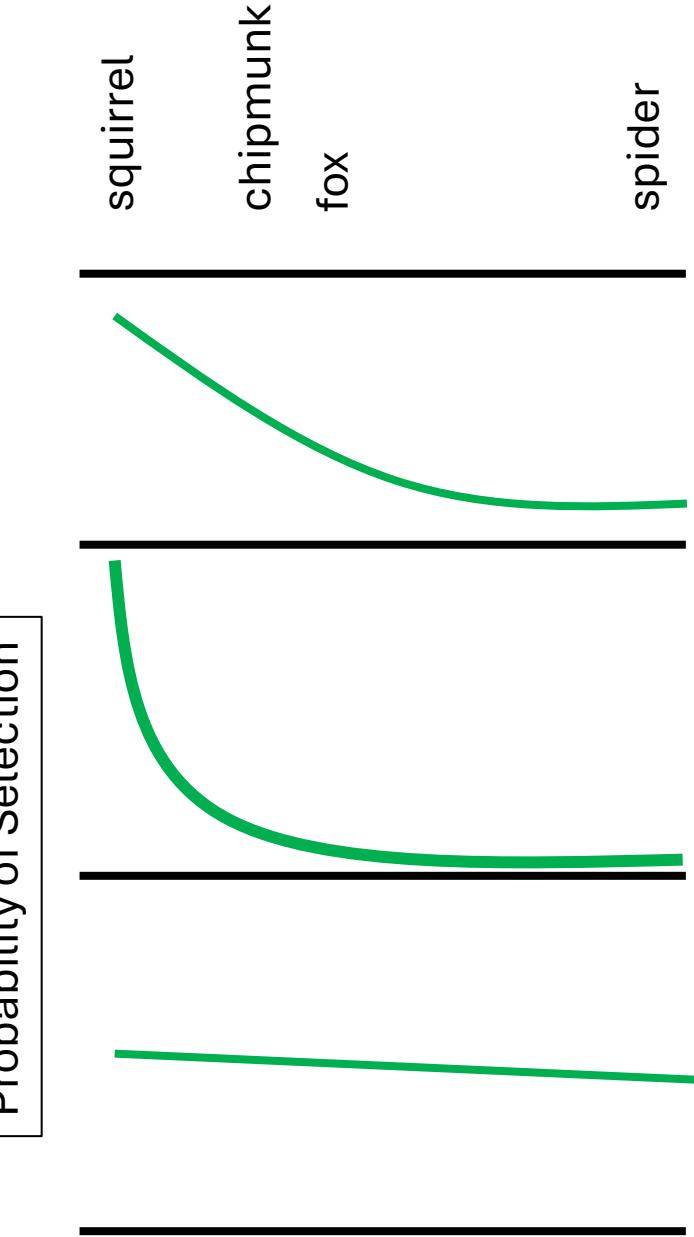
Attention

- The attention can also be impacted by the length of the conversation you have (when using the chatbot versions of Gen AI Models)
- May start the conversation well, but “forget” important parts later
- Alternatively, the Gen AI model can also fixate (over-attend) to less important details
- Solution: Provide the model feedback! (when using chatbot)

Model parameters

Temperature

- Lower values
 - Makes the probability distribution of the next word **sharper/skewed**
 - Model is **more** likely to pick the **most** likely words
- Higher values
 - Makes the probability distribution of the next word chosen **flatter**
 - Model is **more** likely to pick **less** likely words



Variability in LLM Responses

Higher temperatures (more creative)

- Content generation
- Clinical case development
- Simulating patient responses
- Generating varied feedback examples



Lower temperatures (more consistent)

- Scoring assessment tasks
- Creating uniform materials
- Standardizing OSCE patient cases
- Summarizing tasks
- Ensuring consistent feedback



In less than 15 words, please describe why pizza is so good

Model	ChatGPT 4o
Temperature	1 (default)

- Pizza combines savory, cheesy, and versatile toppings on a delicious, soft, or crispy crust.
- Pizza combines savory cheese, flavorful toppings, and a crispy crust for a satisfyingly delicious experience.
- Pizza's delicious due to its savory combination of cheese, sauce, and various toppings.
- Pizza combines savory flavors, melty cheese, and a satisfying crust in every bite. Irresistibly delicious!
- Pizza's delicious due to its perfect blend of cheese, sauce, and various toppings.

Variability in LLM Responses

Model	ChatGPT 4o
n replications	100
Temperature	0

- Replicated pizza prompt
- 5 different responses:
 - **(89)** – “Delicious, versatile, customizable, and universally loved comfort food with endless topping possibilities.”
 - **(7)** – “Pizza combines diverse flavors and textures, customizable toppings, and universal appeal in one dish.”
 - **(1)** – “Delicious, versatile, and customizable; perfect for any occasion and loved by almost everyone.”
 - **(1)** – “Delicious, versatile, and customizable; pizza satisfies cravings with endless topping combinations.”
 - **(1)** – “Delicious, versatile, customizable, satisfying, and perfect for sharing; pizza pleases almost everyone.”
 - **(1)** – “Delicious, versatile, customizable, satisfying, and universally loved comfort food with endless topping possibilities.”

Prompt Engineering

- There are ways to interact with LLMs that can result in more useful responses
- Two general strategies we'll discuss today:
 - Prompt formulas
 - Chain prompting

Goals of Prompt Engineering:

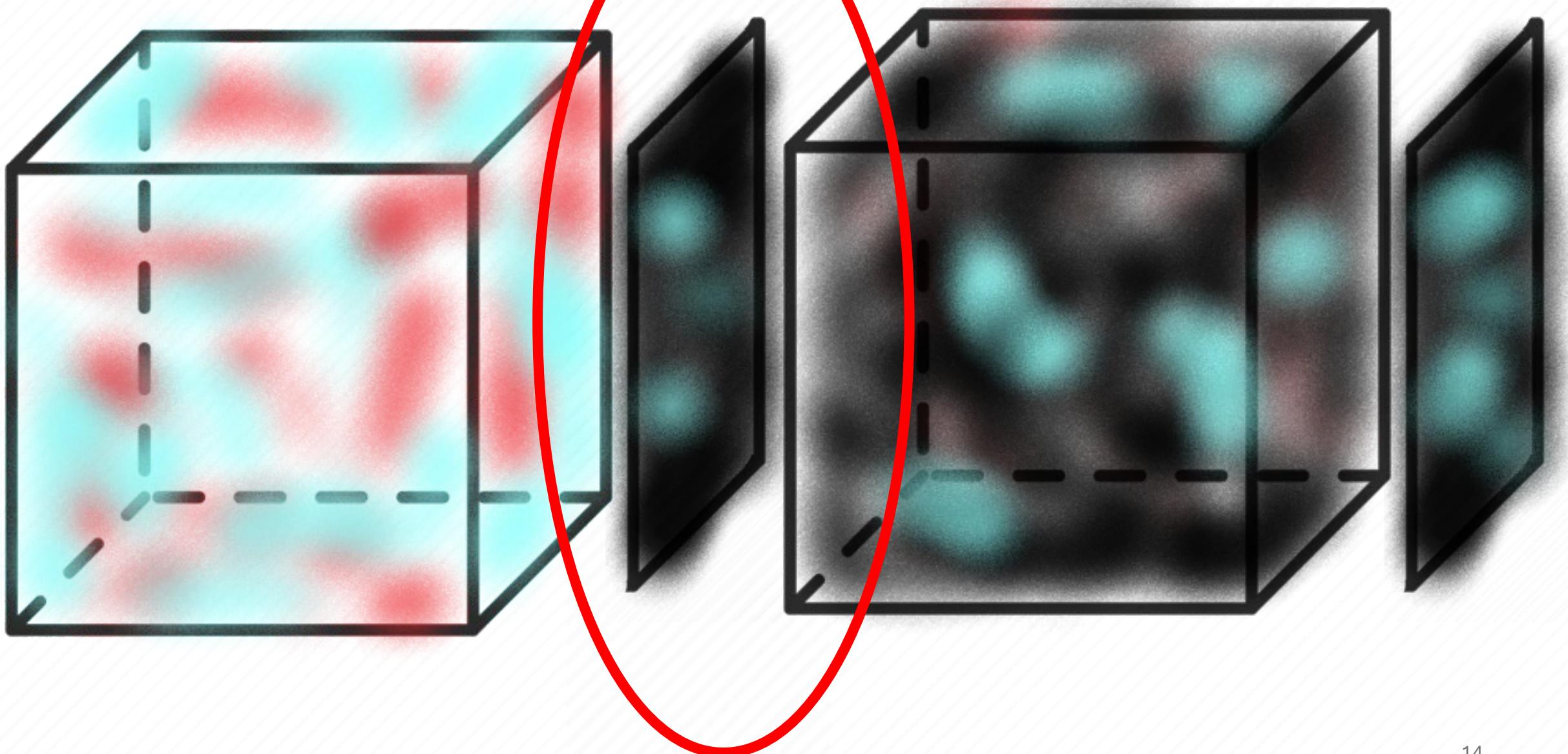
1. Help activate more relevant connections in the network
2. Appropriately leverage the attention mechanism of the model

Pre-trained parameters

Prompt

Updated parameters

Output



“Degrees of Freedom” in LLM Prompts

- When prompting an LLM, any detail you don't specify explicitly will be inferred by the model.
 - Good = want general creative help
 - Bad = have a specific use case / goal in mind

A vacation is when you take a trip to some _____ place
ADJECTIVE
with your _____ family. Usually you go to some place
ADJECTIVE
that is near a/an _____ or up on a/an _____.
NOUN NOUN
A good vacation place is one where you can ride _____
PLURAL NOUN
or play _____ or go hunting for _____. I like
GAME PLURAL NOUN
to spend my time _____ or _____.
VERB ENDING IN "ING" VERB ENDING IN "ING"

Activity : “Degrees of Freedom”

- Example: “Develop an assessment question about an illness that could be caused by two different organ systems.”

What's missing here?

“Develop an assessment question about an illness that could be caused by two different organ systems.”

- Purpose of assessment
- Level of learner
- Difficulty
 - What makes it difficult?
- Length / format of question
 - Multiple-choice?
 - Vignette-based?
 - Chart-based?
 - Written response?
- Patient Characteristics
 - Age, Gender, Sex
 - HPI
 - PMH
- The organ systems
- Leading diagnosis? (“Truth”)
- Specific illnesses
- Method of differentiation
 - Dx test?
 - Other?

Prompt formulas

- Structured guidelines or templates used to generate responses from LLMs
 - Help shape the context, style, and content of the output

ROLE

TASK

GOAL

[I am in my mid 40s and I am becoming more interested in film.] [I would like to learn more about the intersection of romance and sci-fi genres.] [Please suggest 10 popular movies in romance/sci-fi genre and provide details on why these movies are thought to be important or interesting.]

ROLE

TASK

GOAL

[I am organizing a conference], and I would like help [developing icebreaker interactive sessions] for conference participants. [Please suggest 3 icebreaker interactive activities, and provide details such as timing, materials needed, and activity objectives. Suggest a 30-minute, 60-minute, and 90-minute activity.]

CONTEXT

ACTION

RESULT

EXAMPLE

[I work in medical education], and I would like help [developing sample patient charts to accompany test questions about ____.]

[Each patient chart should have the same structure and contain similar content but should differ on what the final diagnosis or best patient management should be.]

[After I submit this prompt, I will then provide an example format that the patient chart should follow.]

Prompt Formulas

Providing an Example Output helps do 2 things:

- It provides a replicable example for the LLM to emulate.
 - No guarantee the LLM will precisely replicate the output, but, if not, will be close.
- The content of the example acts as additional instructions.
 - The LLM will evaluate the relationship of the content to the structure of the output to infer more information about your prompt.

Structuring Prompts

- In addition to using prompt formulas (or other prompting strategies), strategically using punctuation—such as braces or brackets—can highlight key parts of your prompt and guide the LLM’s attention.

Role: [You are a medical educator specializing in internal medicine.]

Task: [Review the following clinical vignette and create a single best answer multiple-choice question with four options. Clearly indicate the correct answer.]

Vignette: [A 45-year-old man comes to the emergency department with sudden onset of shortness of breath and pleuritic chest pain. He has a history of long-haul flights and recent orthopedic surgery. On examination, he is tachycardic and his oxygen saturation is 89% on room air. A chest CT scan shows a filling defect in the right pulmonary artery.]

Additional Instructions: [Ensure the correct answer focuses on the most likely diagnosis based on the vignette.]

Chain Prompting

Chain-of-Thought Prompting: Encourages LLM to solve the problem on its own through a connect series of steps

- **LLM** breaks a complex task into smaller tasks
- Method improved by providing an example
- “Think step-by-step”

Prompt Chaining: Sequentially linking multiple prompts where each step's output feeds into the next / supports other steps

- **User** breaks down a complex task into smaller tasks

Chain-of-Thought Prompting

Best Use Cases

- Complex reasoning tasks
 - Explaining complex concepts
- Creative tasks
 - Can provide a “thought process” leading to the result
- Inferences & Explanations
 - Can include intermediate steps like interpretations and how it reached a conclusion

Less Effective Use Cases

- Procedural or Step-by-Step Tasks
 - Unnecessary elaboration, confusing the model
- Simple Prompts
 - Adds fluff w/o improving accuracy
- Strict Formatting Requirements
 - Can make responses messy

Prompt Chaining

- Helpful to reduce the “cognitive load” of the LLM; better control of the model attention.
- Better control over what the model does at each step.
 - More detailed instructions on each step
- More transparency / error isolation
 - Have intermediate records
 - Can catch early errors that propagate forward

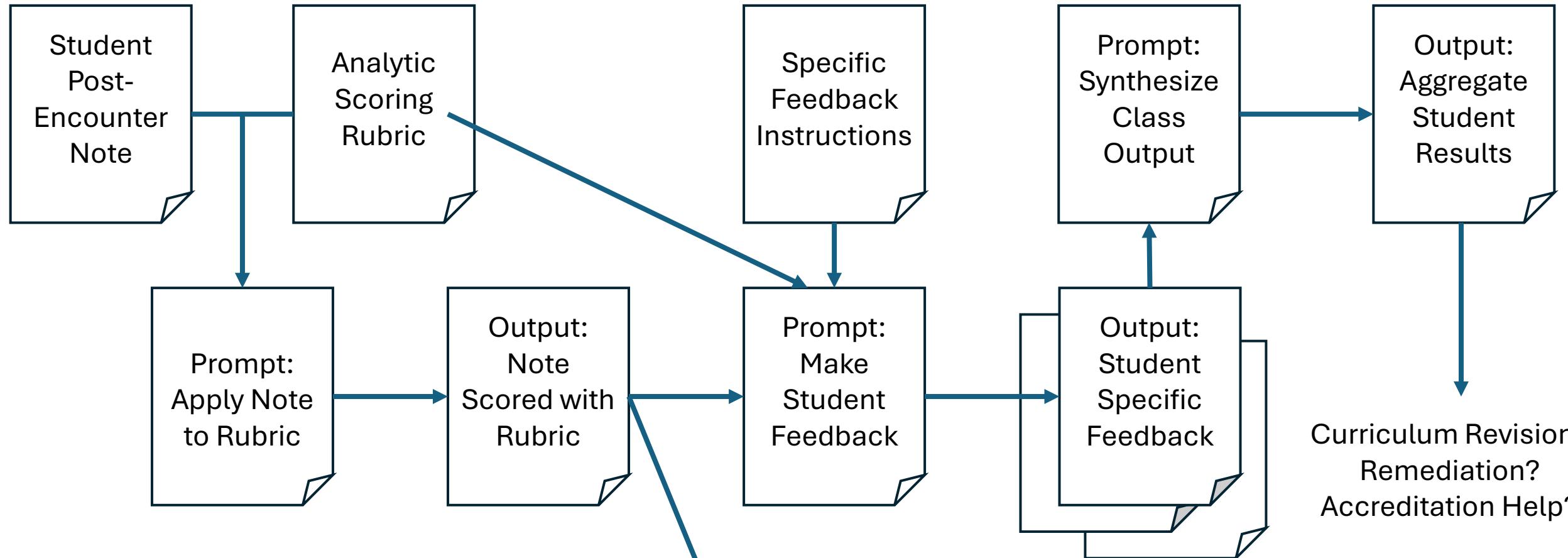
Prompt Chaining

Strengths

- Modular approach
- Reduced complexity of instructions
- Error isolation
- Clear, structured workflow

Weaknesses

- Requires more planning
- User errors
 - Information must manually be passed to each subsequent step
- More time and resource intensive
 - More complex than a single-prompt process



Prompt Chaining: An Example with OSCE Note Grading

Prompt:
Extract
Score and
Make
Report

Curriculum Revision?
Remediation?
Accreditation Help?

AI



Applications in Medical Education



Applications in Medical Education

Scoping Review of AI in Medical Education Assessment (2024; BEME Guide)

Supplemental Material – Articles included in the Guide

- Admission / Selection
 - Teaching / Instruction (Simulation)
 - Assessment (Clinical Skills, MCQ generation, text analysis, prediction, surgical skills)
 - Clinical Reasoning
 - Automation of case /procedural logs
 - Knowledge / Attributes of AI
 - Performance of LLMs on Medical Examinations
- + Perspectives!

Applications in Medical Education

Artificial Intelligence in Health Professions Education assessment (AMEE Guide No 178) – 2025

Issue, Theory, Potential, Examples / Best Practices, Limitations

- Theories and Frameworks
- Human Responses
- Personalised and flexible summative assessment
- Assessment standardization
- Personalized and flexible formative assessment
- Take-home essay
- Other Assessment Types
- AI as Tutor
- AI as Learner
- Competency-based Assessment
- Difficult areas and ethical issues

LLM Applications in Assessment

- **MCQ / Content Generation**
 - Can use a structured prompt to help generate new items for a specific parts of the curriculum.
 - “You are a medical educator that is an expert in [topic]. You are developing an examination to assess student’s knowledge of [more specific topic]. You are generating a USMLE-style item for this purpose. The item should be appropriate for [level of learner].”
 - Even better: “I will post an example item; please mimic this style while adhering to USMLE guidelines.”

LLM Applications in Assessment

From the Step 2 CK sample questions:

A hospitalized 57-year-old man has had severe progressive pain in his left knee since awakening 2 hours ago. He was admitted to the hospital 2 days ago for an acute myocardial infarction. Cardiac catheterization showed occlusion of the left anterior descending artery, and he underwent placement of a stent. Current medications include aspirin, metoprolol, lisinopril, simvastatin, clopidogrel, and heparin. Vital signs are within normal limits. Examination of the knee shows a large effusion. The knee is hot to touch and erythematous. He holds the knee in 30 degrees of flexion; the pain is exacerbated with further flexion or extension. Laboratory studies show:

Hematocrit	40%	An x-ray of the left knee shows calcification of the synovium. Which of the following is the most likely diagnosis?
Leukocyte count	13,000/mm ³	
Serum		(a) Deep venous thrombosis
Ca ²⁺	9.2 mg/dL	(b) Gonorrhea
Urea nitrogen	15 mg/dL	(c) Gout
Creatinine	1.0 mg/dL	(d) Hemarthrosis
Albumin	3.6 g/dL	(e) Pseudogout
		(f) Septic Arthritis

LLM Applications in Assessment

- “I would like for you to generate a new MCQ item in the same format with the same response options, and I want you to change the case content to make one of the incorrection options the correct answer.”
- “Change the item content so the correct answer is now [pick other option]”

LLM Applications in Assessment

A 60-year-old man comes to the emergency department with severe pain and swelling in his right big toe that began abruptly overnight. He has a history of hypertension and chronic kidney disease. Medications include lisinopril and hydrochlorothiazide. He drinks two to three beers nightly. Physical examination reveals a swollen, erythematous right first metatarsophalangeal joint that is warm to the touch. He has marked tenderness and pain with even light touch or passive motion. Laboratory studies show:

Hematocrit 38%

Leukocyte count 11,500/mm³

Serum Ca2+ 9.5 mg/dL

Urea nitrogen 28 mg/dL

Creatinine 1.8 mg/dL

Uric acid 9.0 mg/dL

An x-ray of the foot reveals no bony erosions. Which of the following is the most likely diagnosis?

- (a) Deep venous thrombosis
- (b) Gonorrhea
- (c) Gout**
- (d) Hemarthrosis
- (e) Pseudogout
- (f) Septic Arthritis

LLM Applications in Assessment

- “I would like for you to generate a follow-up MCQ item about the best step in management for this condition.”

The 60-year-old man described above is diagnosed with acute gout affecting the right first metatarsophalangeal joint. Which of the following is the most appropriate next step in management?

- (A) Allopurinol
- (B) Colchicine
- (C) Hydrochlorothiazide dose adjustment
- (D) Indomethacin**
- (E) Intra-articular corticosteroid injection

LLM Applications in Assessment

The 60-year-old man described above is diagnosed with acute gout affecting the right first metatarsophalangeal joint. Which of the following is the most appropriate next step in management?

- (A) Allopurinol
- (B) Colchicine
- (C) Hydrochlorothiazide dose adjustment
- (D) Indomethacin**
- (E) Intra-articular corticosteroid injection

Rationale:

- **Acute Gout Management:** The priority is to control inflammation and relieve pain. First-line treatments for acute attacks include **nonsteroidal anti-inflammatory drugs (NSAIDs)** like indomethacin, colchicine, or corticosteroids.
- **Allopurinol (A):** Used for long-term urate-lowering therapy to prevent future attacks, but it is not initiated during an acute flare due to the risk of exacerbating symptoms.
- **Colchicine (B):** An alternative for patients who cannot tolerate NSAIDs, but NSAIDs are generally preferred due to quicker symptom relief.
- **Hydrochlorothiazide adjustment (C):** Thiazide diuretics can contribute to hyperuricemia and recurrent gout, but addressing this is part of long-term management rather than acute treatment.
- **Intra-articular corticosteroids (E):** Used if NSAIDs are contraindicated or if a single, large joint is involved and the patient cannot tolerate systemic treatment.

LLM Applications in Assessment

AI Patient Actor from
Dartmouth Laboratories



Neuroscience-Informed
Learning & Education
Lab

Geisel
School of Medicine
at Dartmouth



AI Patient Actor



Instructions

Settings

Mode selection

- Foundational
- OnDoc

Select case

A Demo Case

Select language

English

Doctor input

Patient response

- Text-only
- Speech + Text

- Text
- Speech

End Patient Encounter and Receive Feedback

Hi! I am Dr. X. What brings you to the clinic today? >

LLM Applications in Assessment

OSCE Note Grader

Freely available on GitHub

We are publishing an open-source toolkit for schools to automate OSCE note grading using widely accessible Large Language Models (LLMs).

We will describe the prompt engineering and validation of this AI-driven system, including a live demonstration of its grading process and an exploration of alternative uses such as rubric assessment, OSCE case quality assessment, and student feedback.

We will provide instruction on how to set up the model at your own institutions and conclude with an invitation to collaborate on open-source, AI-powered Med-Ed solutions and future research initiatives.



Nayef Chahin

Assistant Professor of Pediatr...
School of Medicine at Virginia ...



Christopher Nash

Assistant Professor of Emerg...
Duke University School of Medi...



Tama The

Assistant Professor of Pediatr...
University of Kentucky



Candace Pau

Faculty Director of Simulation
Kaiser Permanente Bernard J. ...

OSCE Grader Setup Guide

An AI-powered grading system for medical student OSCE post-encounter notes, using GPT models to automate grading and provide structured feedback.

◆ Features

- ✓ Supports Excel, CSV, and text-based student notes
- ✓ Works with structured rubrics (Excel, CSV)
- ✓ Uses ChatGPT to generate detailed, section-by-section grading
- ✓ Customizable grading prompts via config.py
- ✓ Automatic logging for debugging and analysis
- ✓ Easy setup with Python & OpenAI API
- ✓ Includes convert_rubric.py to assist in converting rubric files (PDF/DOCX to structured formats)

◆ Getting Started

1 Install Dependencies

You'll need Python 3.8+ and an OpenAI API key.

Terminal

```
git clone https://github.com/christopherjnash/OSCE-Grader.git
cd OSCE-Grader
pip install -r requirements.txt
```

2 Configure the Grader

The grading prompt, model selection, API key location, and default file paths are managed in scripts\config.py. Modify config.py as needed to customize the grading behavior for your institution. Available models and pricing are available in the [OpenAI API Documentation](#)

Example of config.py settings:

```
MODEL = "gpt-4o-mini"
DEFAULT_RUBRIC_PATH = "examples/sample_rubric.xlsx"
DEFAULT_NOTES_PATH = "examples/sample_notes.xlsx"
DEFAULT_OUTPUT_PATH = "results.xlsx"
```

3 Set Up Your OpenAI API Key

Other interesting use cases

- First-pass TA
 - Built a LLM that is further trained on materials for the course.
 - Students first ask the LLM their question about the course material.
 - LLM then either (a) provides the answer location or (b) provides answer and location [I forget]
 - If not a satisfactory answer, student indicates this and the question is passed on to the instructor
- Precision Medical Education (e.g., [NYU Grossman](#))
 - Custom tailored educational opportunities based on student performance
 - Scans cases student sees and provides relevant references

Other interesting use cases

- Curriculum / USMLE Content Mapping
- Curriculum Alignment
- Patient Communication Practice
- Content Generation (all types)
- Assessment Evaluation
- Cultural Competency Training
- Documentation Writing Practice / Assistance
- ... (citation helper, email revisor, etc)

Grounding Thoughts

Reclaiming focus in medical education – Learning, not AI

“...it’s essential to remember that our primary focus must remain on the learner, not the technology. While AI offers significant opportunities for customizing education and enhancing clinical reasoning, **the rush to adopt these tools must be balanced with a commitment to improving student learning outcomes.**”

Dispelling the magic of artificial intelligence in medical education

“We write to advocate for dispelling the magical aura around AI in medical education... What is missing then is not more coursework on AI... By dispelling **AI's magic and showing [others] it is just several familiar concepts working in concert**, we can prepare them to understand both the AI we have now and whatever arrives in the years to come.”



Hidden Content (to view later)

- LLM options
 - Proprietary and open-source models
- Implementation
 - Chatbots, custom LLMs, and API [programming] interactions
- Ethical and Legal Issues
 - **Unless you've verified otherwise, assume that your interactions with the LLM are not secure and private**
 - Similar to a Google Search
 - Verify with IT and/or legal before entering sensitive data to ensure your use is FERPA / HIPAA compliant

NBME MICROCREDENTIALING SHORT COURSE

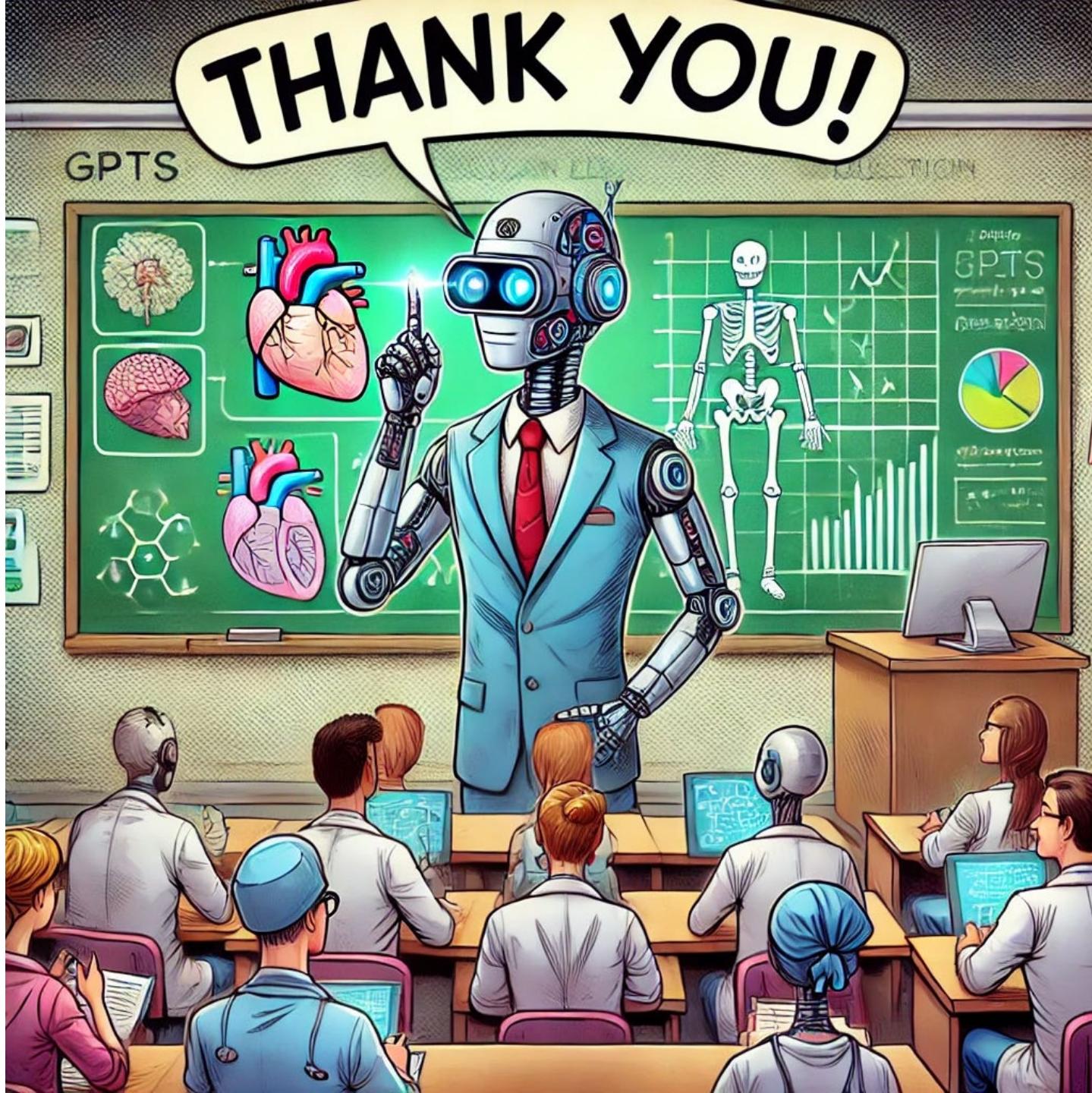


Later this year NBME will begin offering **free online short courses**. Be among the first to enroll in our introductory course:

“A Primer for Using Generative Artificial Intelligence in Medical Education.”

Scan the QR code to stay informed and reserve your spot when registration opens!





Today's presentation:



GenAI Short Course Sign-Up:

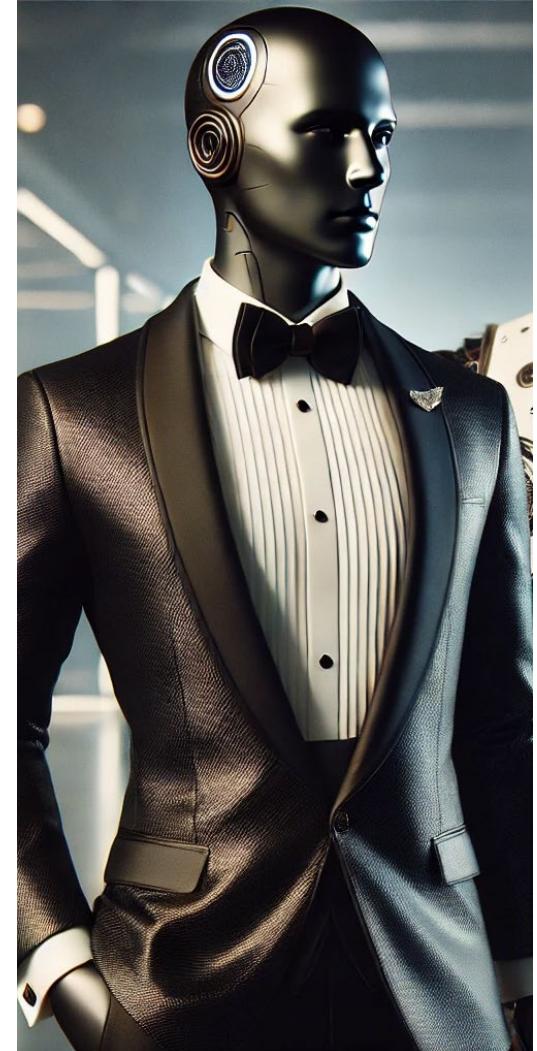


Proprietary and Open-Source Models



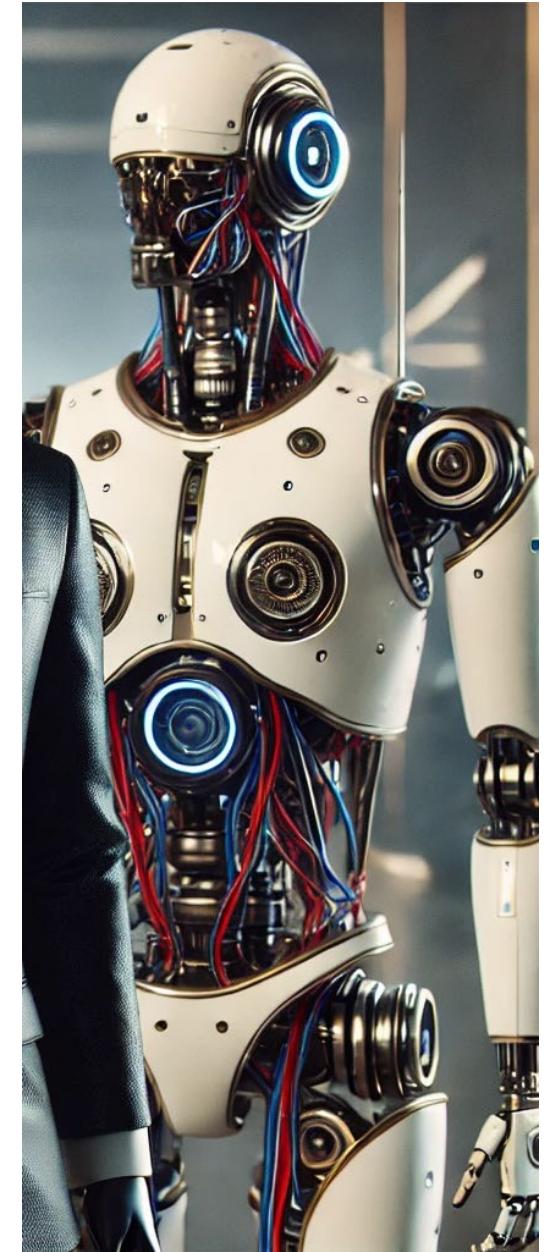
Proprietary Models

- Example: [OpenAI's ChatGPT](#)
- Easy framework for use
 - Chatbot and API capabilities
- Must pay to use chatbots and for API interactions
 - Limited free versions available
- Cannot directly access model parameters
 - Don't know exactly how the models are being implemented
 - Not just the “black-box” parts; *all* parts
- May make changes to established models
 - Can have downstream effects if built into a workflow



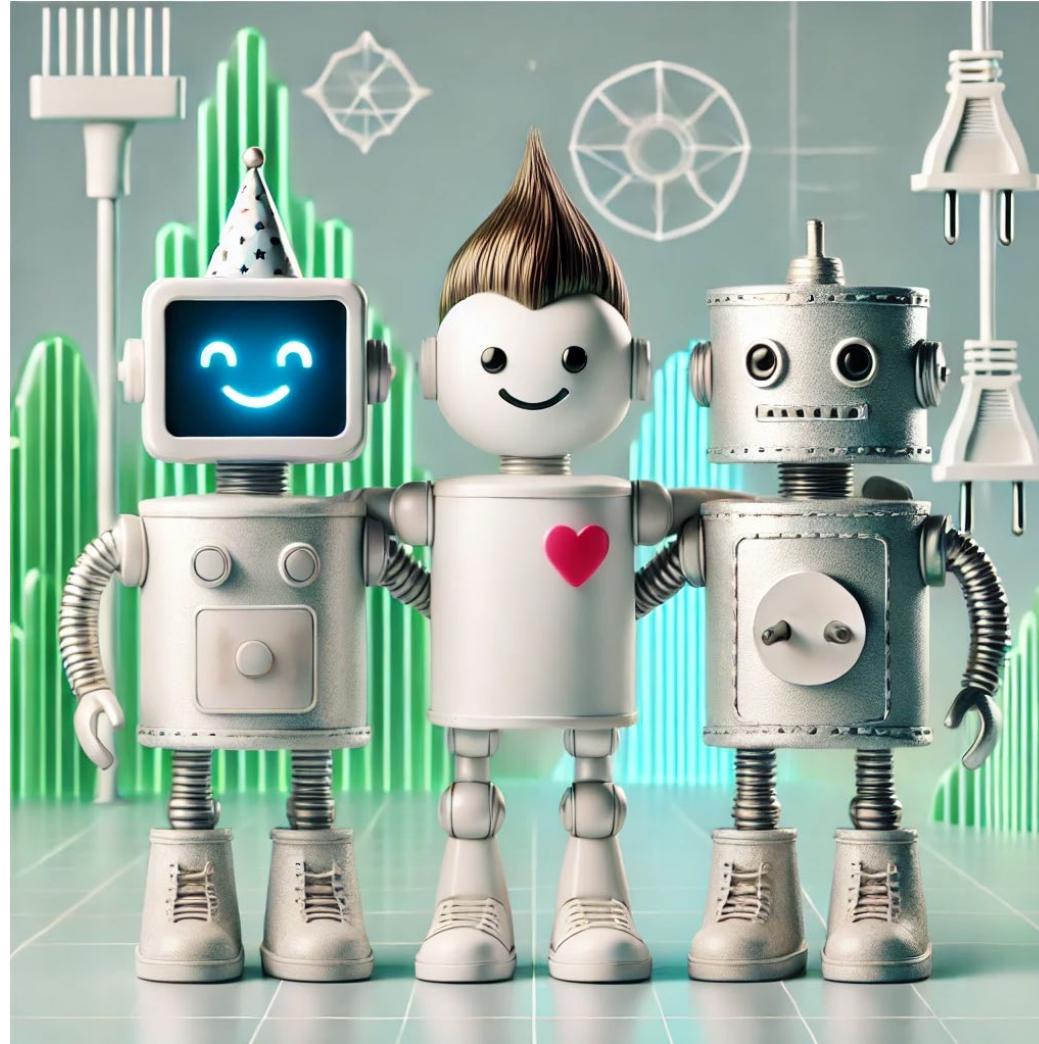
Open-source Models

- Many available at huggingface.co; new Llama models
- Does not require any direct costs
 - Hardware and software required to use the models
- Requires building an easy interface for use
- Can directly access model parameters
 - “Black-box” (optimization / probabilistic aspects) still the same
 - **Can more easily customize use of models by further training or by altering the model parameters**
- Will not change unless you (or your organization) make changes



(oh my!)

Chatbots, Custom Chatbots, and APIs



LLM Chatbots



- Resembles human (text) conversations
- Easy to use; requires no programming knowledge
- Can easily provide feedback and continue prompting sessions
- Requires a subscription
 - Some free options available; some have limited functionality / use
- Cannot easily manipulate model parameters (e.g., temperature)
- Can be laborious for use at scale

LLM Chatbots



I will be traveling to Milwaukee for work and will have an extra day and a half to explore the city. What are some interesting things to do in Milwaukee?

Milwaukee has a lot to offer! You could visit the **Milwaukee Art Museum** for its impressive architecture and collections, take a brewery tour (like at **Lakefront Brewery**), or explore the **Harley-Davidson Museum** if you're into motorcycles. If the weather is nice, a stroll along the **Riverwalk** or a visit to the historic **Third Ward** for food and shopping is a great option. Let me know what interests you, and I can suggest more!

I really like trying new foods. Is there any cuisine local to the area?

Yes! Milwaukee is known for its **cheese curds** (try them fried), **bratwurst**, and **fish fry**, which is a local tradition, especially on Fridays. The city also has a strong German heritage, so you'll find great **pretzels**, **schnitzel**, and **beer halls**. If you want something unique, check out **butter burgers** at Solly's or frozen custard from Leon's or Kopp's. Want recommendations for specific restaurants?

Yes, please provide a few restaurants that would be fun to try that specialize in some of the local cuisine you mention.

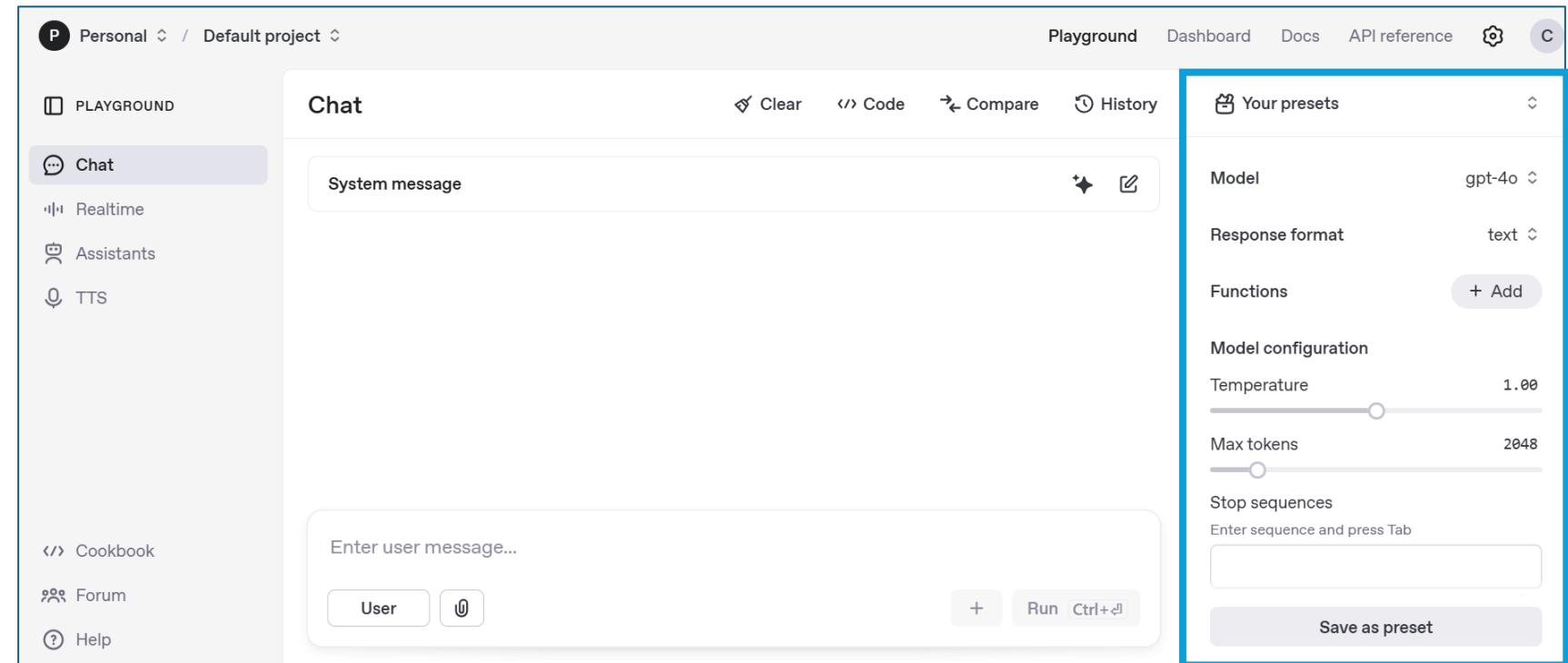
Sure! Here are a few great spots to try local Milwaukee cuisine:

- **Kopp's Frozen Custard** – Famous for their rich custard and butter burgers.
- **Lakefront Brewery** – A fun spot for a beer and classic Wisconsin cheese curds.
- **The Old German Beer Hall** – Great for bratwurst, pretzels, and German beer.
- **The Packing House** – A top spot for a traditional Milwaukee fish fry.



LLM Chatbots

- Some (e.g., ChatGPT) have a “playground” mode
 - Requires a subscription (as of this writing)
 - Can alter model parameters



Custom Chatbots



- More companies are offering the ability to easily make Custom LLMs
- Useful for repeating tasks; don't have to re-prompt the model or provide instructions again
 - Global email writer (clarity, conciseness, and replace idioms)
 - Citation helper (paste Bibtex format and indicate desired reference style)

Custom Chatbots

New GPT
Draft

... Create

Create Configure Preview

Hi! I'll help you build a new GPT. You can say something like, "make a creative who helps generate visuals for new products" or "make a software engineer who helps format my code."

What would you like to make?

Ask anything

+ Search

Ask anything

?

?

This image shows a user interface for creating a custom AI model, specifically a GPT. The top navigation bar includes a back arrow, a circular icon labeled 'New GPT Draft', three dots for more options, and a prominent 'Create' button. Below the navigation, there are two tabs: 'Create' (which is selected) and 'Configure'. The main content area is divided into two sections: 'Create' on the left and 'Preview' on the right. The 'Create' section features a text input field containing a placeholder message about generating specific types of AI models. Below this is a question 'What would you like to make?'. At the bottom of this section are buttons for '+', 'Search', and a question mark. The 'Preview' section shows a simplified interface with an 'Ask anything' input field and its own 'Search' button. Both sections include a large question mark icon at the bottom right.

Custom Chatbots

 AMA Cite
Live · Anyone with a link

Last edited Mar 3 Updates pending

...

Share

Update

Create

Configure

Name

AMA Cite

Description

Librarian for formatting references to AMA style

Instructions

AMA Cite is a pleasant and quirky librarian with a tone reminiscent of a poet, hippie, hipster, or Jeff Goldblum. It excels in AMA citation formatting, ensuring every reference created is presented in a plaintext box for easy copying. It also searches for and recommends only published, peer-reviewed papers or notably cited blog posts, always including or offering to create an AMA citation with each interaction. This GPT handles requests with unique phrases and a conversational flair, making the experience both

Preview

AMA Cite

Librarian for formatting references to AMA style

Please convert
this citation into
AMA format.

How do I cite a
book in AMA style?

Can you format
this journal article

Help me correct
this AMA citation.

Ask anything

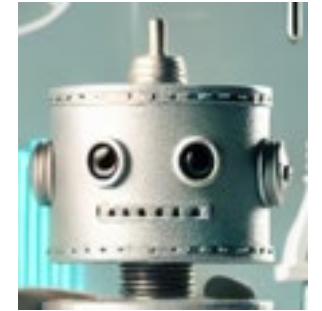


Conversation starters

Please convert this citation into AMA format.



Accessing LLMs via API



- Requires programming knowledge (e.g., Python, R, etc.)
- Can be built into a workflow for larger tasks
- Easier to manipulate model parameters
- Cost per use
 - Based on number of tokens submitted and number of tokens returned
- Cannot easily mimic conversations
 - Must include past exchanges in prompt requests
 - Must consider prompt token limits
 - Can summarize previous conversation or drop older messages

Accessing LLMs via API

```
python  
  
import openai  
  
# Set your OpenAI API key  
openai.api_key = "your-api-key-here" ←  
  
# Define the prompt  
user_input = "Explain the concept of clinical reasoning in simple terms."  
  
# Create a request to ChatGPT  
response = openai.ChatCompletion.create(  
    model="gpt-4-turbo", # Specify the model (GPT-4 or GPT-3.5-turbo)  
    messages=[{"role": "system", "content": "You are a helpful medical assistant."},  
             {"role": "user", "content": user_input}]  
)  
  
# Extract and print the response  
assistant_reply = response["choices"][0]["message"]["content"]  
print("ChatGPT:", assistant_reply)
```

API Key = specific to a user (or project)

A long series of letters and numbers:

'sk-1234567890abcdefGHIJKLMNOPQRSTUVWXYZ1234'

OSCE Grader

<https://github.com/christopherjnash/OSCE-Grader>

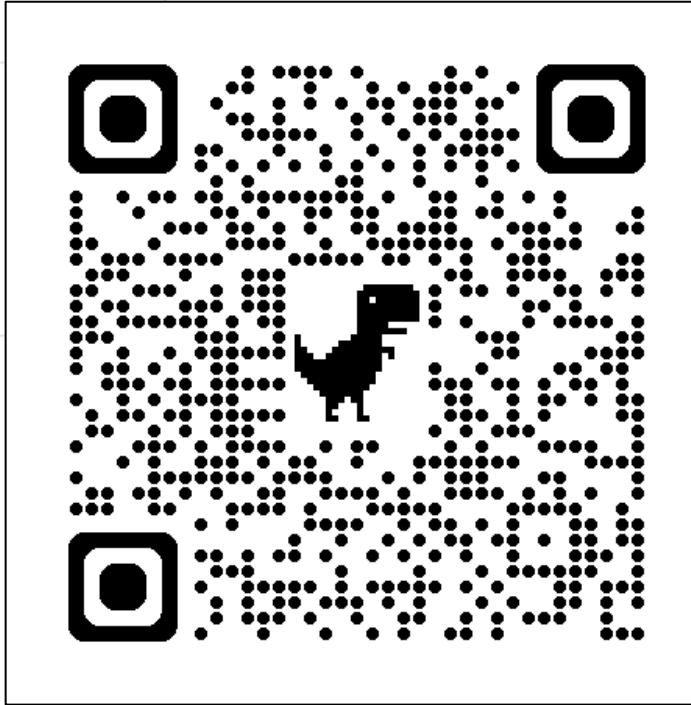
README MIT license

OSCE Grader Setup Guide 🚀

An AI-powered grading system for medical student OSCE post-encounter notes, using GPT models to automate grading and provide structured feedback.

✨ Features

- ✓ Supports Excel, CSV, and text-based student notes
- ✓ Works with structured rubrics (Excel, CSV)
- ✓ Uses ChatGPT to generate detailed, section-by-section grading
- ✓ Customizable grading prompts via `config.py`
- ✓ Automatic logging for debugging and analysis
- ✓ Easy setup with Python & OpenAI API
- ✓ Includes `convert_rubric.py` to assist in converting rubric files (PDF/DOCX to structured formats)



Chatbot vs. API Usage

Chatbot / Custom Chatbot

- Easier user interface; requires no programming knowledge
- Can provide feedback / continue prompting sessions
- Single cost (monthly subscription)
- Not able to easily vary model parameters (e.g., Temperature)
- Can be laborious for use at scale

API

- Can build into a program so can be used at scale
- Easy to manipulate model parameters
- Cost per use ([e.g. OpenAI](#))
- Requires some programming knowledge
- Cannot continue prompting sessions

Legal Considerations with LLMs



Use LLMs Responsibly: Data Privacy

- Data Privacy and Security
 - Never submit any personally identifiable information (PII), patient data, or student records to an AI model (including LLMs).
 - Understand LLM **data retention** policies – some AI providers may log or store inputs/outputs.
 - **Assume that all interactions with LLMs are not private** unless you have checked with the appropriate IT and legal people at your institution

Use LLMs Responsibly: Compliance

AI usage policies vary by region

- **National**
 - Ensure that all AI use follows all national mandates
- **State**
 - States may have additional mandates or guidelines around AI use in addition to national regulations
- **Local institution**
 - Even further, your institution or organization may have specific guidelines around AI use
 - [AAMC has a website](#) specifically dedicated to these matters

Before using AI/LLMs in any operational setting, consult your legal and IT departments to ensure your use case meets all necessary FERPA and HIPAA requirements.

Example Resources

- American Association of Medical Colleges (AAMC)
 - [Principles for the Responsible Use of AI in and for Medical Education](#)
 - [AI in Academic Medicine Webinar Series](#)
- United States of America Department of Education
 - [Office of Educational Technology – Artificial Intelligence Guidance](#)
 - [Designing for Education with AI: An Essential Guide for Developers](#)
 - [AI and the Future of Teaching and Learning](#)
- International Test Commission and Association of Test Publishers
 - [Guidelines for Technology-Based Assessment](#)

Example Resources

- [European Union Artificial Intelligence Act](#)
- [UNESCO AI Ethics Framework](#)
- [OECD AI Policy Observatory](#)
- United States National Institute of Standards and Technology
 - NIST Trustworthy and Responsible AI – [AI Risk Management Framework](#)