



20
25

AMIA
**Informatics
Summit**

UNITING DATA-DRIVEN INFORMATICS

#IS25

DISCLOSURE OF CONFLICTS OF INTEREST

I have not had any relationships with ACCME-defined ineligible companies within the past 24 months.



AMIA 2025 Informatics Summit

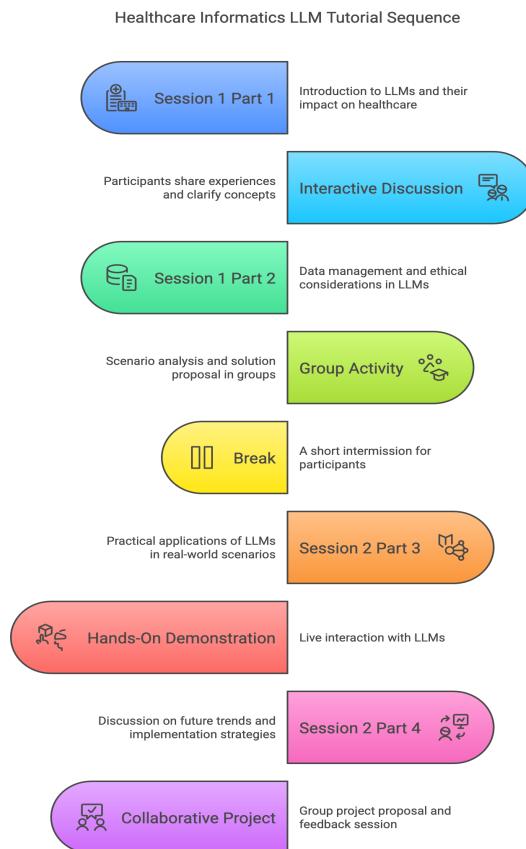
#IS25



Owl Health

**Harnessing Large Language Models in Healthcare
Informatics:
A Comprehensive Tutorial
Jeremy R. Harper, MS, FAMIA
Owl Health Works LLC, Indianapolis, Indiana, USA**

Welcome & Tutorial Overview



- Goals of the tutorial
- Housekeeping: breaks, interactive sessions, Q&A.

Mastering LLMs in Healthcare Informatics

Collaborative Project

Engaging in group projects and feedback sessions



Future Trends

Discussing emerging developments and integration strategies



Practical Applications

Identifying real-world use cases of LLMs



Data and Ethics

Exploring data management and ethical considerations



Foundations of LLMs

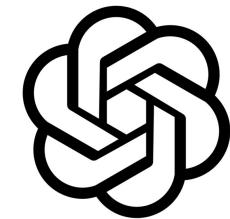
Understanding the history and impact of LLMs in healthcare



Free LLM's



groq



ChatGPT



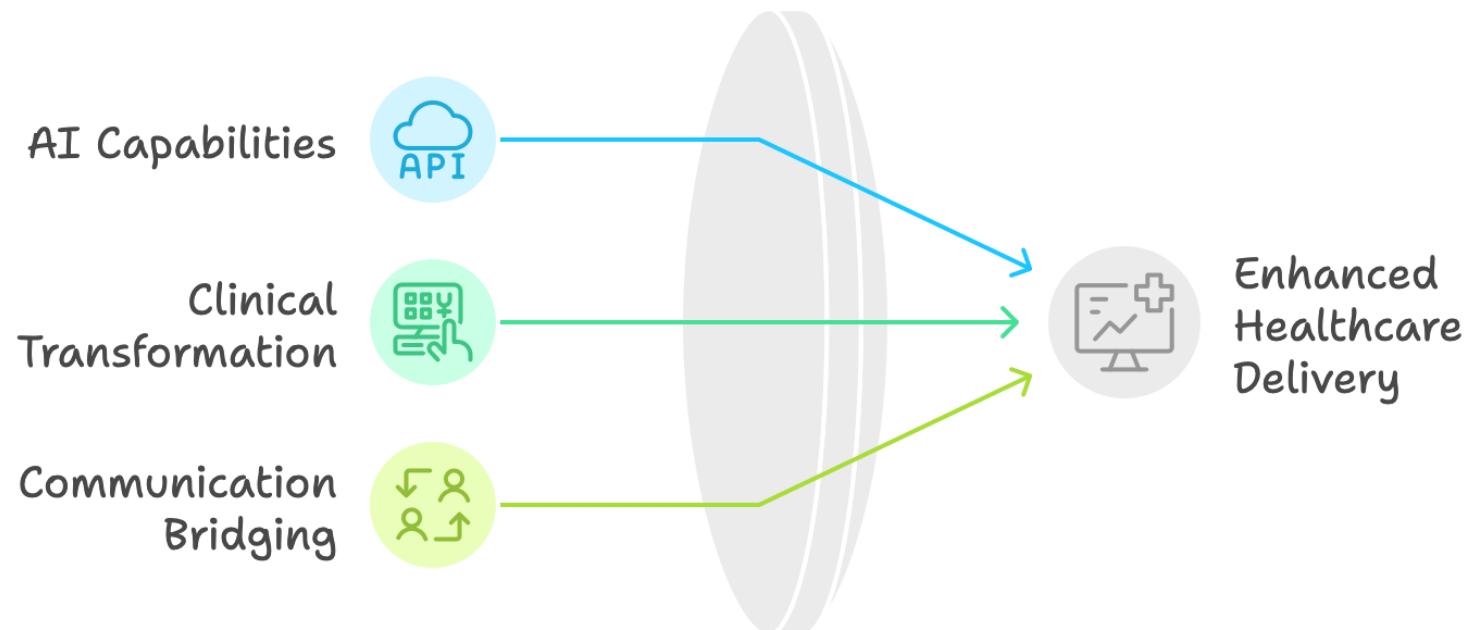
Gemini

https://github.com/Jeremy-Harper/AMIA_IS25

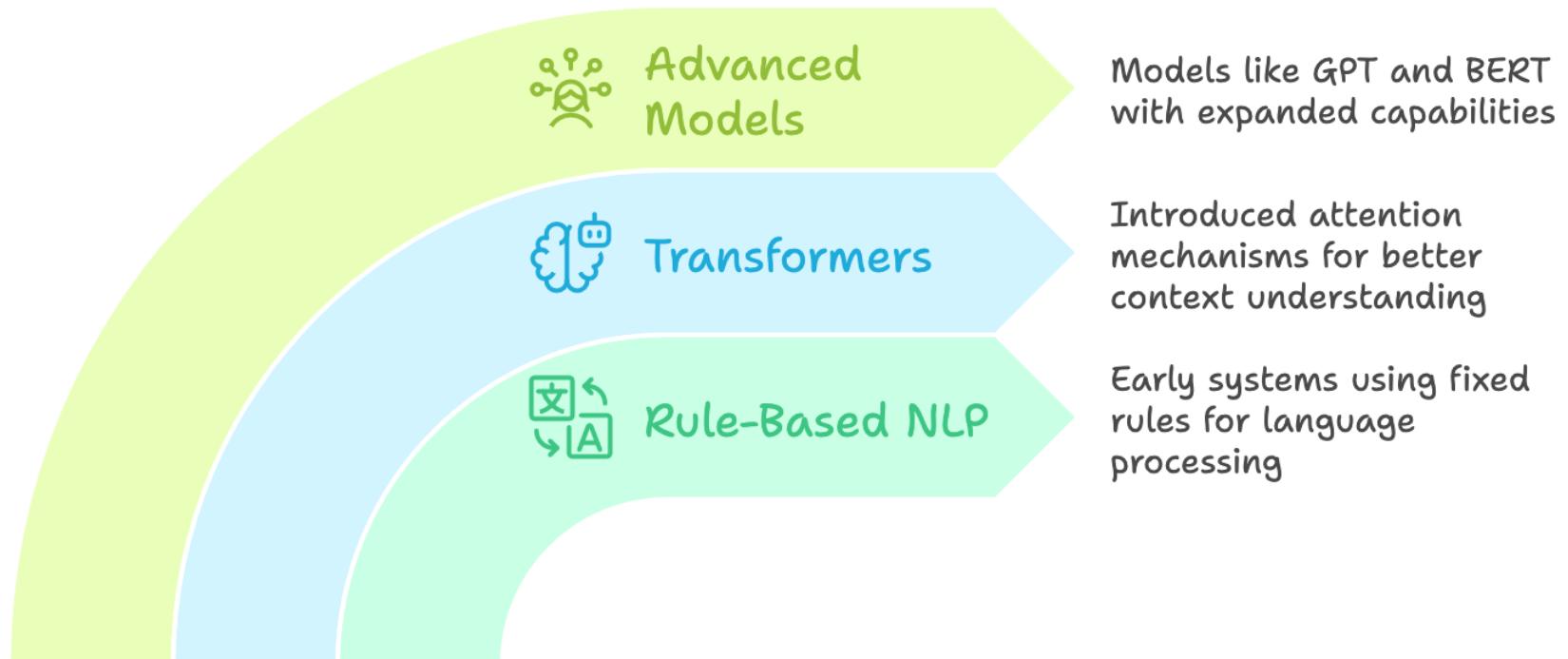


Why LLMs Matter in Healthcare

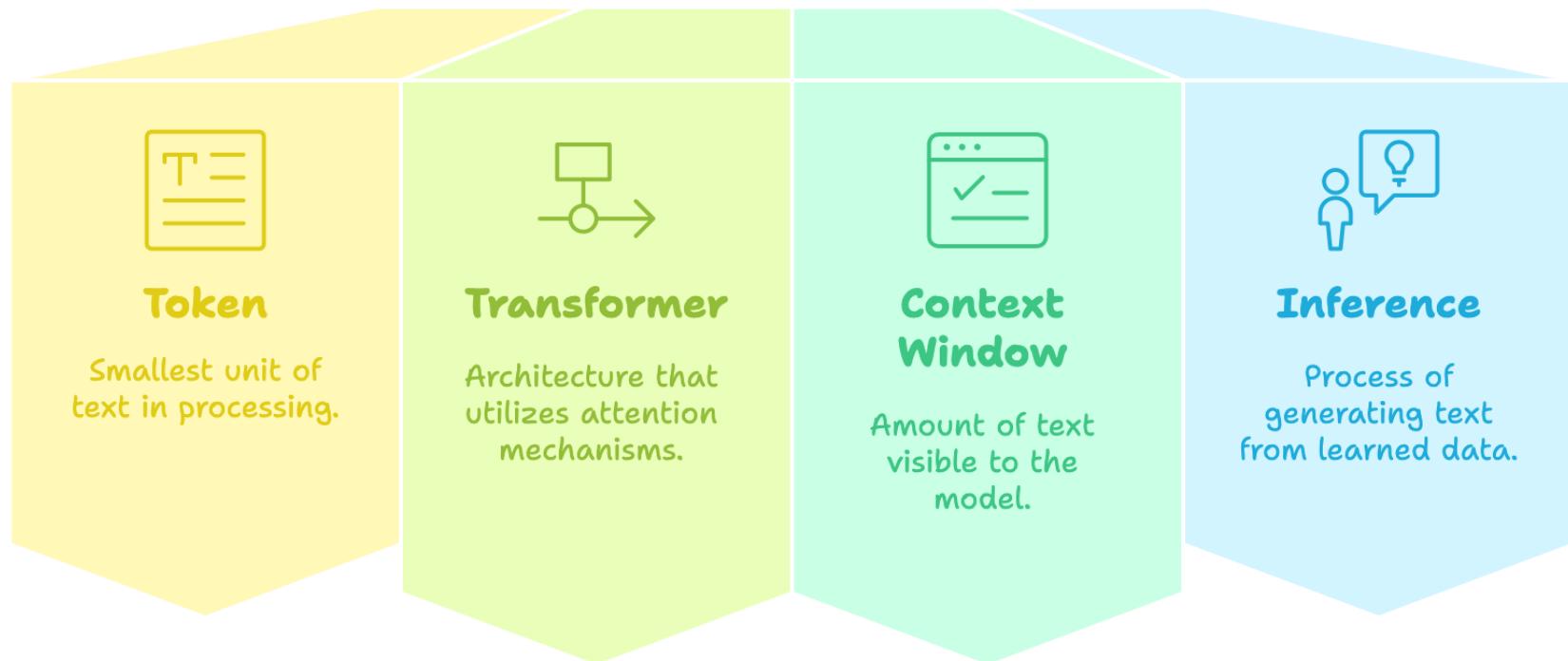
AI's Role in Healthcare Evolution



Evolution of Language Models

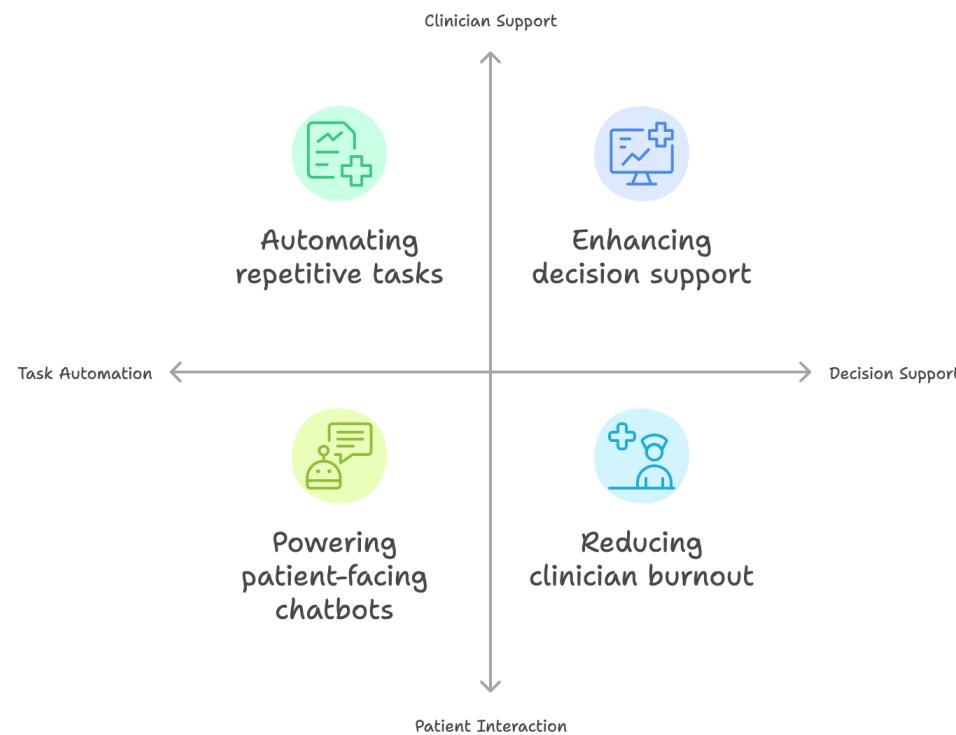


Core Concepts & Terminology

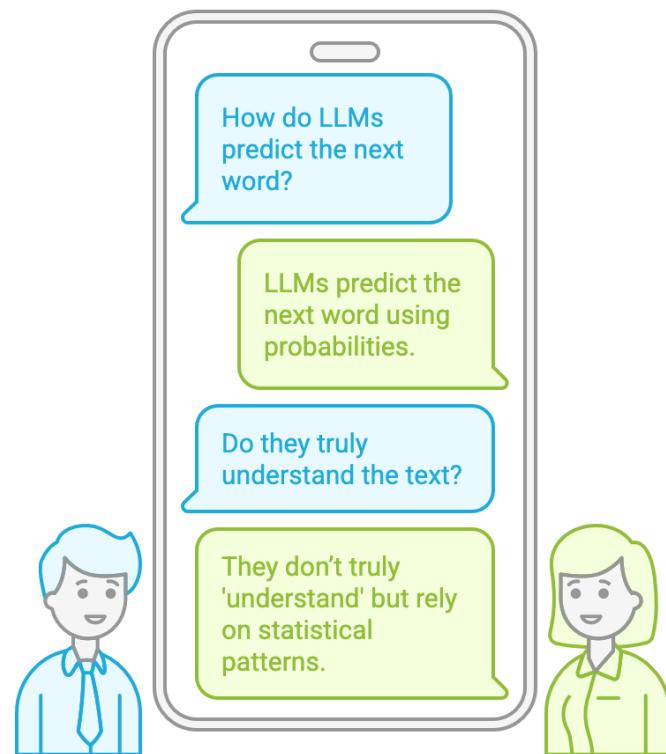


LLM Impact on Healthcare Informatics

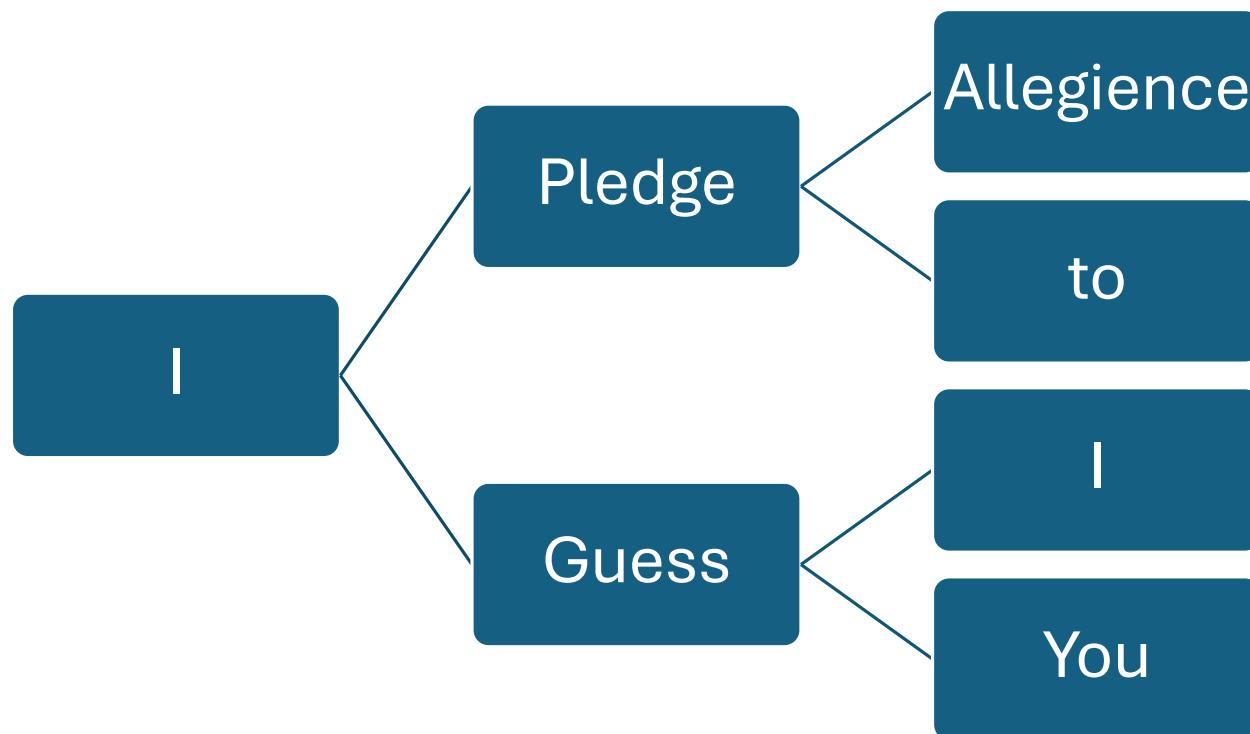
AI Applications in Healthcare



How LLMs 'Think' – The Next Word Prediction



Example



ChatGPT 4 ▾

Share

ChatGPT 4 ▾

Share

I pledge allegiance

It sounds like you're starting the Pledge of Allegiance! Here's the rest:

"I pledge allegiance to the Flag of the United States of America, and to the Republic for which it stands, one Nation under God, indivisible, with liberty and justice for all."

Is there a particular reason you're thinking about the pledge today?

< 2/2 > ⏪ ⏴ ⏵ ⏸ ⏹ ⏷

...to the flag of the United States of America, and to the Republic for which it stands, one Nation under God, indivisible, with liberty and justice for all.

That's the Pledge of Allegiance! Were you looking for information about its history or anything else related to it?

< 3/3 > ⏪ ⏴ ⏵ ⏸ ⏹ ⏷

ChatGPT 4 ▾

Share

ChatGPT 4 ▾

Share

I pledge allegiance

It sounds like you might be starting the Pledge of Allegiance! Do you want to continue or talk about something else related to it?

< 4/4 > ⏪ ⏴ ⏵ ⏸ ⏹ ⏷

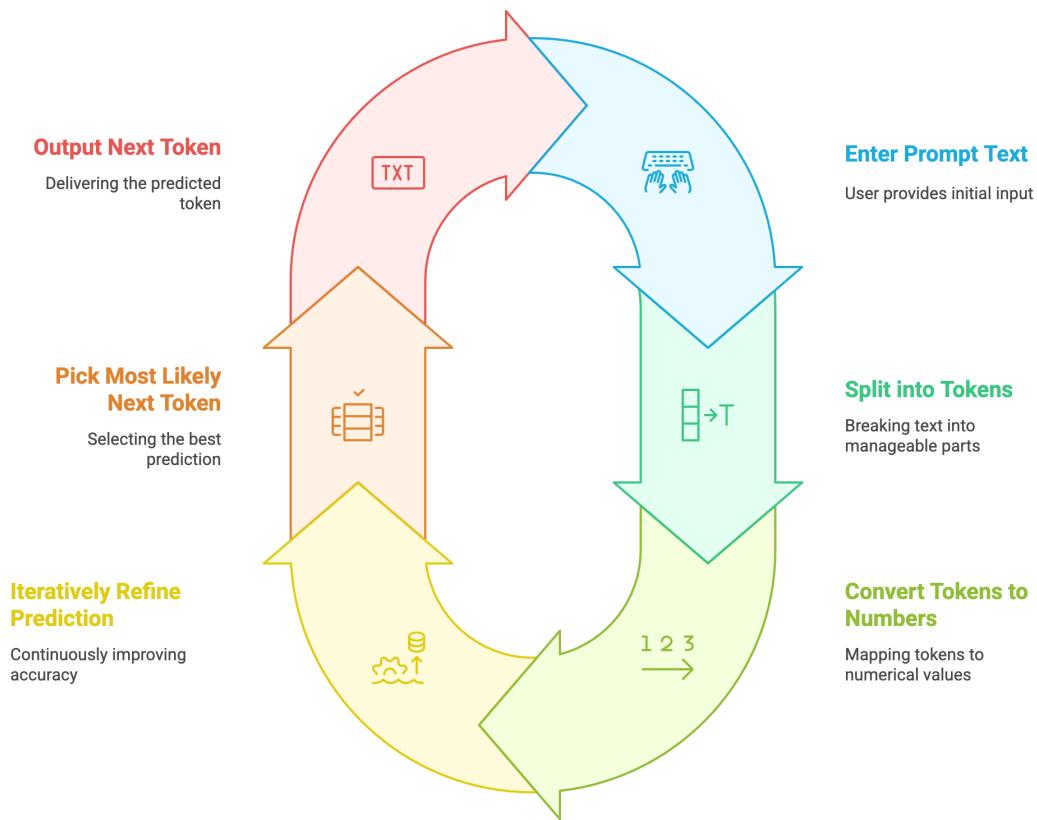
...to the flag of the United States of America, and to the Republic for which it stands, one nation, under God, indivisible, with liberty and justice for all.

That's the pledge! Are you reciting it for a specific occasion, or just brushing up on the words?

< 1/2 > ⏪ ⏴ ⏵ ⏸ ⏹ ⏷

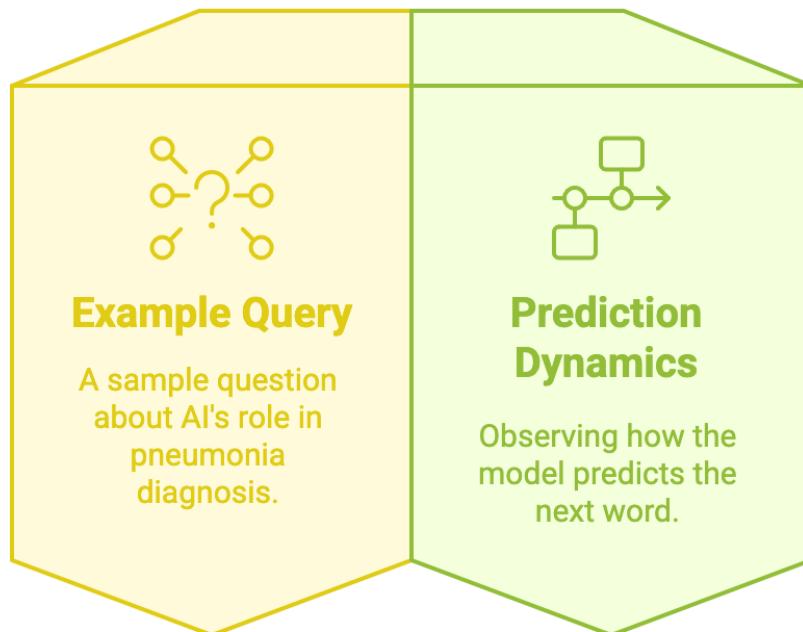
I pledge allegiance

Text Prediction Cycle



Hands-On Preview: Let's Query an LLM GPT-2

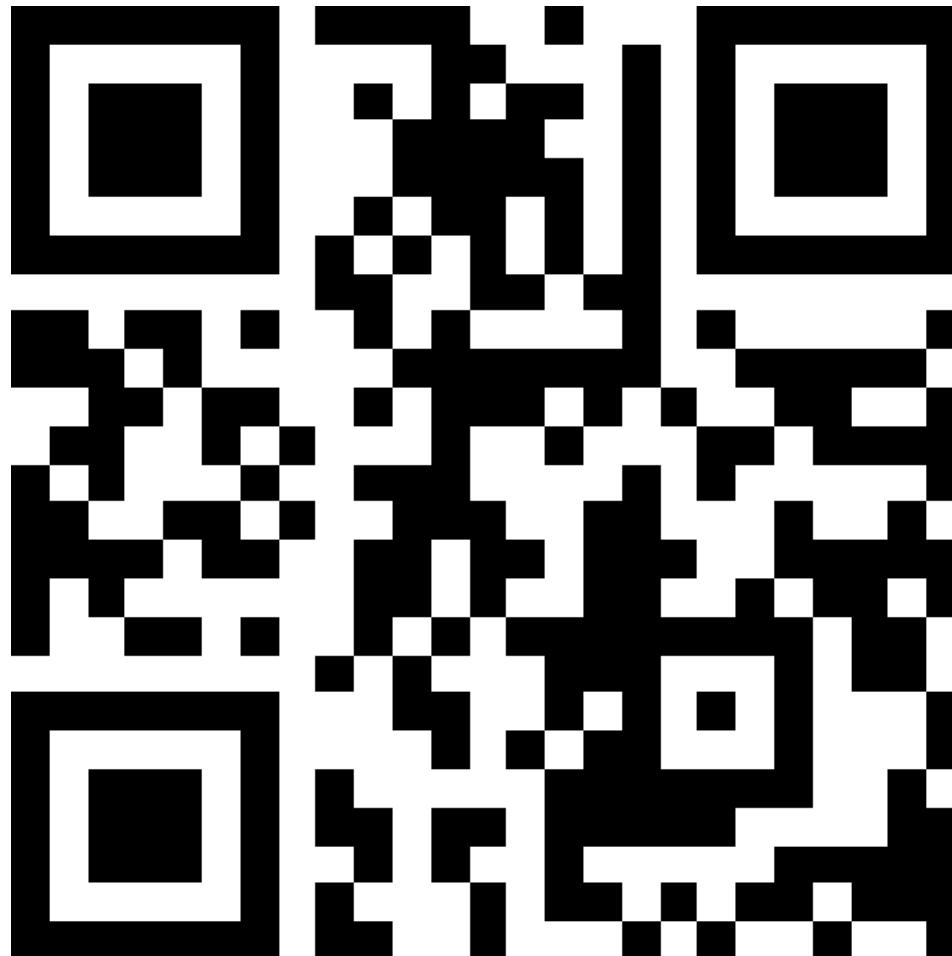
AI in Diagnosing Pneumonia



<https://tinyurl.com/y3ak7p48>

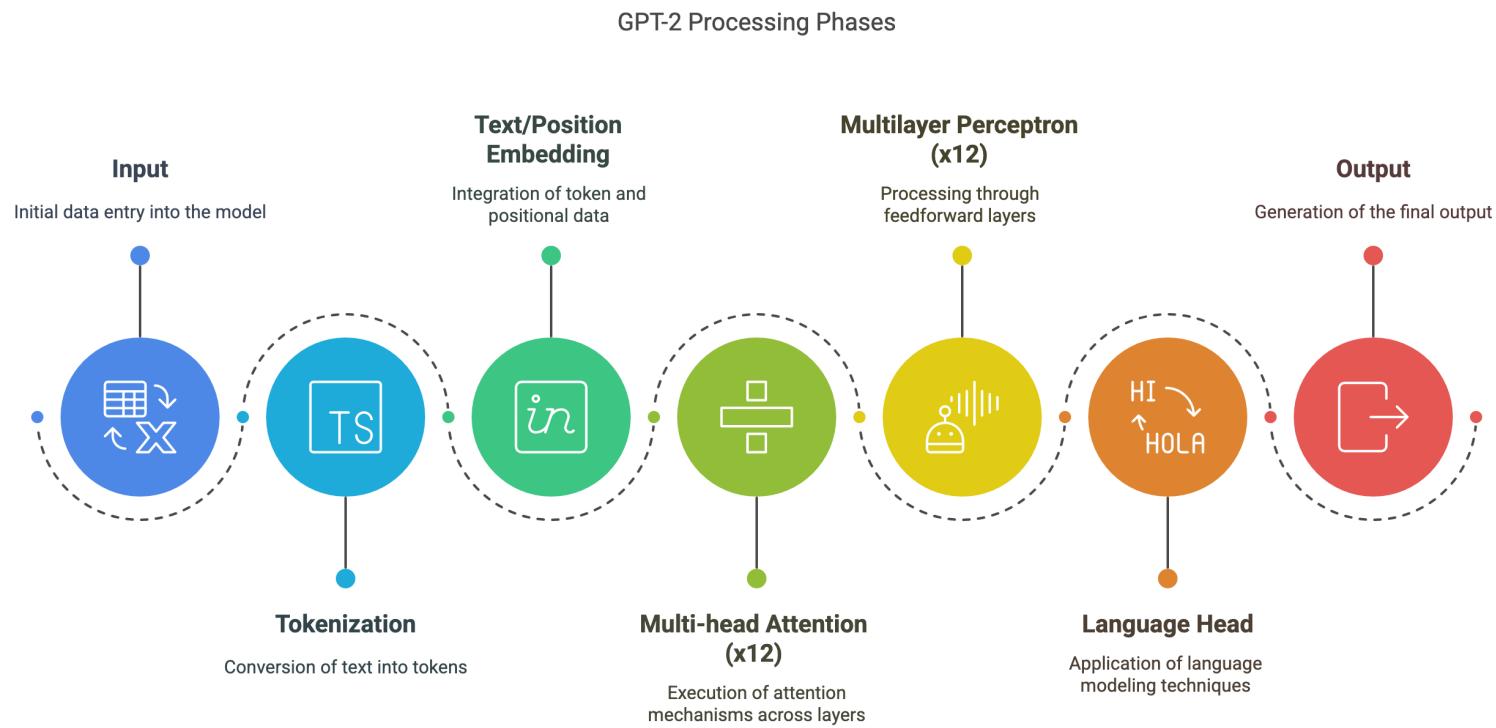


<https://tinyurl.com/y3ak7p48>



https://colab.research.google.com/drive/1BnH6uZUXpCalf23b1VyUL12mKmQ_d0z?usp=sharing

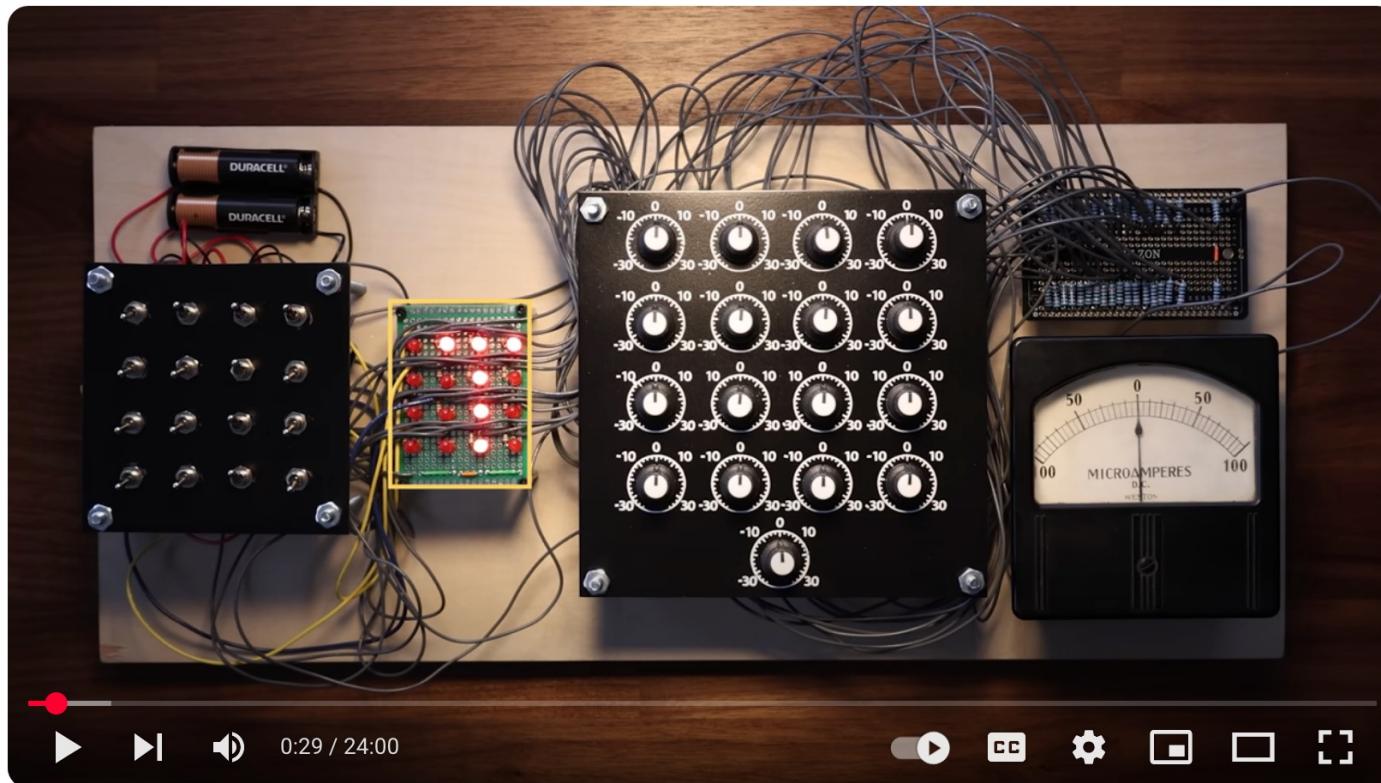
GPT-2



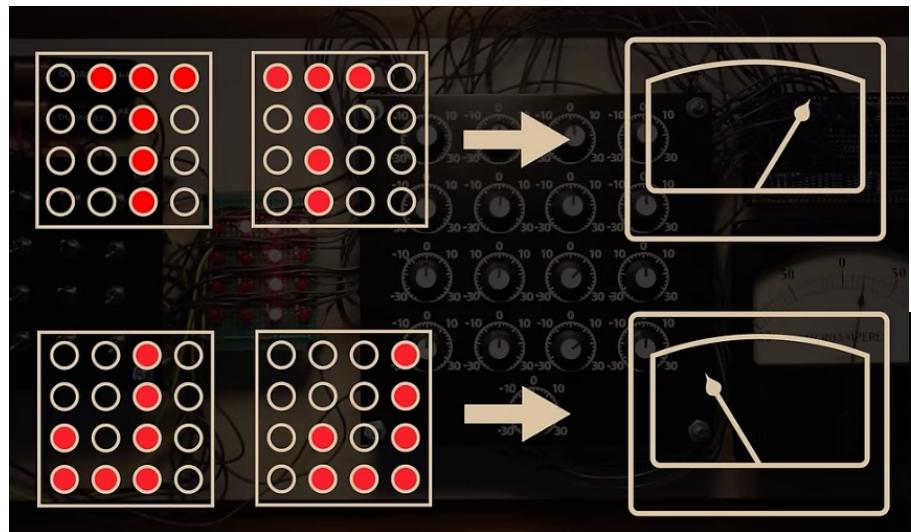
Internal Representation & Embeddings

- Words are converted into numeric vectors.
- Similar words cluster together in vector space.
- GPT-2 In Excel <https://arstechnica.com/information-technology/2024/03/once-too-scary-to-release-gpt-2-gets-squeezed-into-an-excel-spreadsheet/>
- 3Blue1Brown Youtube channel “Transformers”
<https://www.youtube.com/watch?v=wjZofJX0v4M&t=320s>

Perceptron's



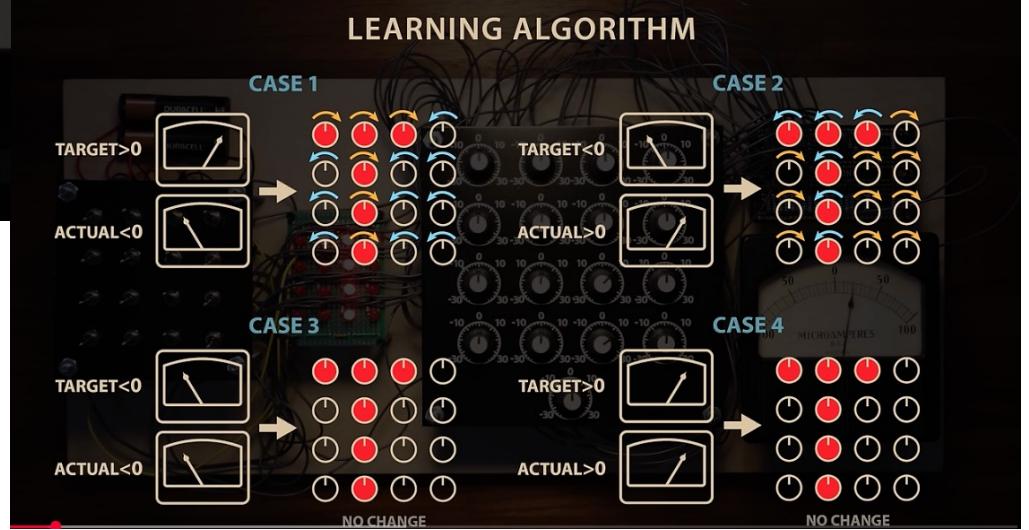
Perceptron's



ChatGPT is made from 100 million of these
[The Perceptron]

Welch Labs Youtube Channel

<https://www.youtube.com/watch?v=l-9ALe3U-Fg>



Internal Representation & Embeddings

| The | quick | brown | fox | jumps | over | the | lazy | dog |
|-----|-------|-------|-----|-------|------|-----|------|-----|
| 3.7 | 1.8 | 6.1 | 6.6 | 3.9 | 7.3 | 1.2 | 2.5 | 8.1 |
| 9.5 | 3.0 | 1.7 | 3.1 | 2.7 | 7.7 | 7.1 | 4.1 | 9.0 |
| 7.3 | 5.2 | 0.7 | 5.2 | 8.3 | 0.7 | 7.6 | 7.6 | 3.2 |
| 6.0 | 4.3 | 9.5 | 5.5 | 3.6 | 3.6 | 5.6 | 2.3 | 1.1 |
| 1.6 | 2.9 | 9.7 | 1.8 | 2.8 | 1.2 | 7.7 | 0.8 | 2.3 |
| 1.6 | 6.1 | 8.1 | 9.7 | 5.4 | 8.6 | 4.9 | 2.9 | 4.3 |
| 0.6 | 1.4 | 3.0 | 7.8 | 1.4 | 6.2 | 5.2 | 1.6 | 8.2 |
| 8.7 | 2.9 | 1.0 | 9.4 | 8.0 | 3.3 | 4.3 | 9.3 | 8.6 |
| 6.0 | 3.7 | 6.8 | 8.9 | 0.7 | 0.6 | 0.3 | 8.1 | 0.1 |
| 7.1 | 4.6 | 4.4 | 6.0 | 9.9 | 3.1 | 1.1 | 6.3 | 5.1 |
| 0.2 | 7.9 | 1.2 | 9.2 | 7.7 | 3.3 | 0.3 | 8.7 | 4.2 |
| 9.7 | 2.0 | 5.0 | 0.9 | 2.0 | 7.3 | 6.4 | 8.0 | 2.2 |
| 8.3 | 5.1 | 0.3 | 2.0 | 0.1 | 6.4 | 3.1 | 1.9 | 1.2 |
| 2.1 | 5.9 | 9.1 | 0.5 | 8.2 | 8.9 | 5.1 | 8.9 | 3.4 |
| 1.8 | 0.5 | 2.6 | 3.3 | 7.1 | 4.7 | 9.1 | 5.4 | 9.4 |

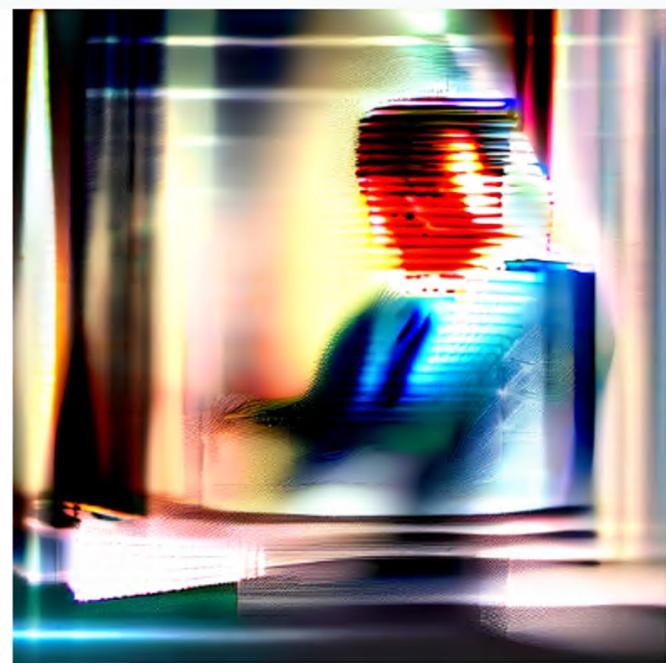


Diffusion – Different than LLMS (How?)



| | |
|----------------|--------------------------------|
| Sampling steps | <input type="text" value="5"/> |
| Batch count | <input type="text" value="1"/> |
| Batch size | <input type="text" value="1"/> |
| | <input type="text" value="7"/> |

Extra



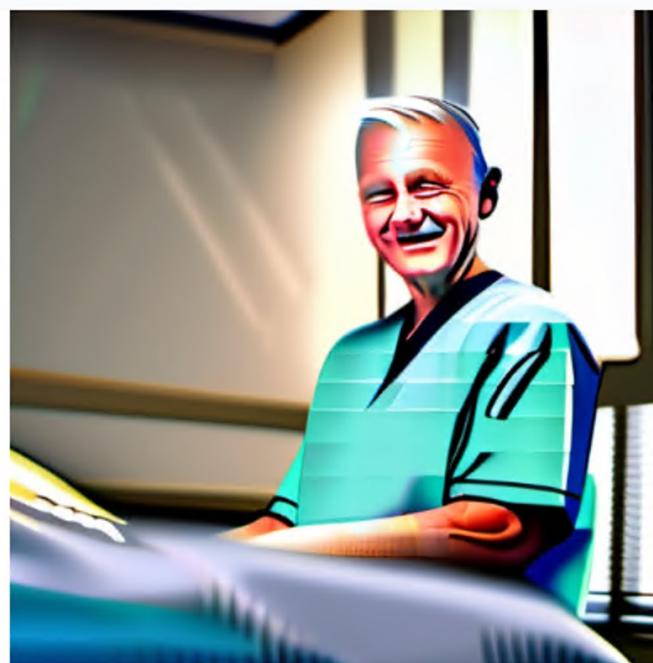
Sampling steps

Batch count

Batch size

7

Extra



Sampling steps

Batch count

↴

Batch size

↴

↴

Extra

↴



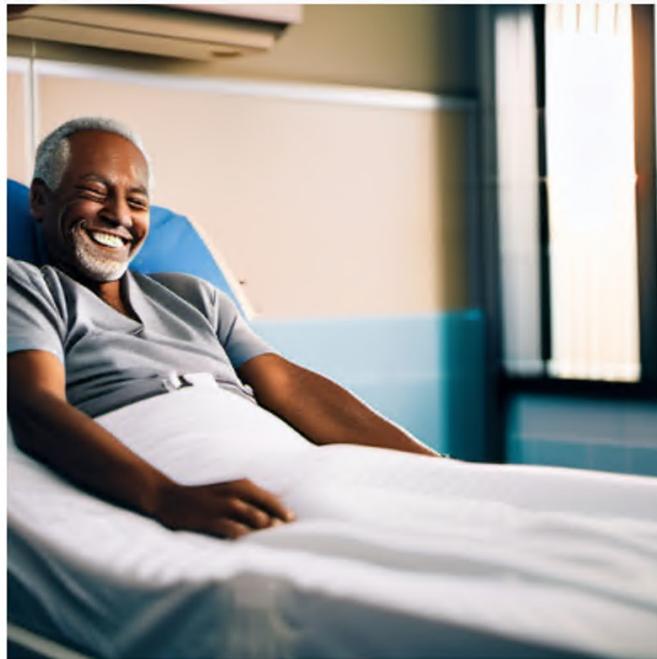
Sampling steps

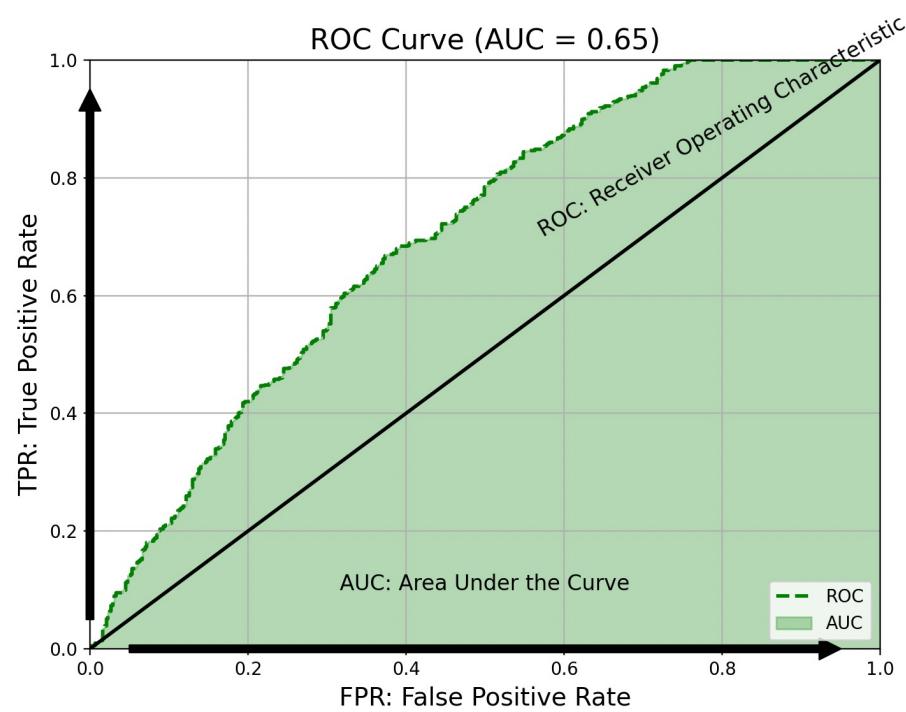
Batch count 

Batch size 

  Extra



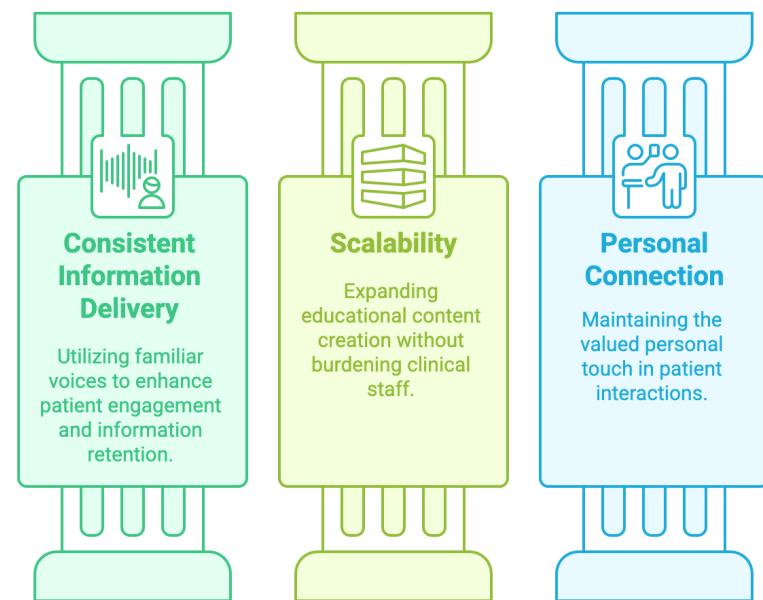




Voice Cloning Capabilities in Healthcare



Enhancing Patient Care Through Scalable Voice Cloning Technology



Pause for Questions

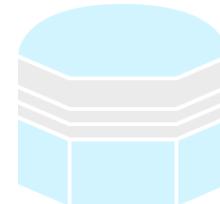
- Any clarifications on LLM so far?

Key Healthcare LLM Use Cases

Healthcare LLM Applications Overview

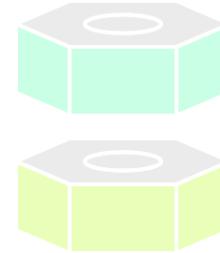
Summarizing Patient History

Condensing medical records for quick reference



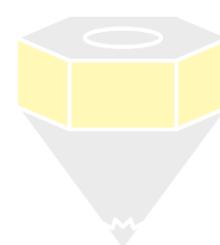
Translating Clinical Jargon

Simplifying medical terms for patient understanding



Generating Discharge Instructions

Creating clear post-hospitalization guidelines

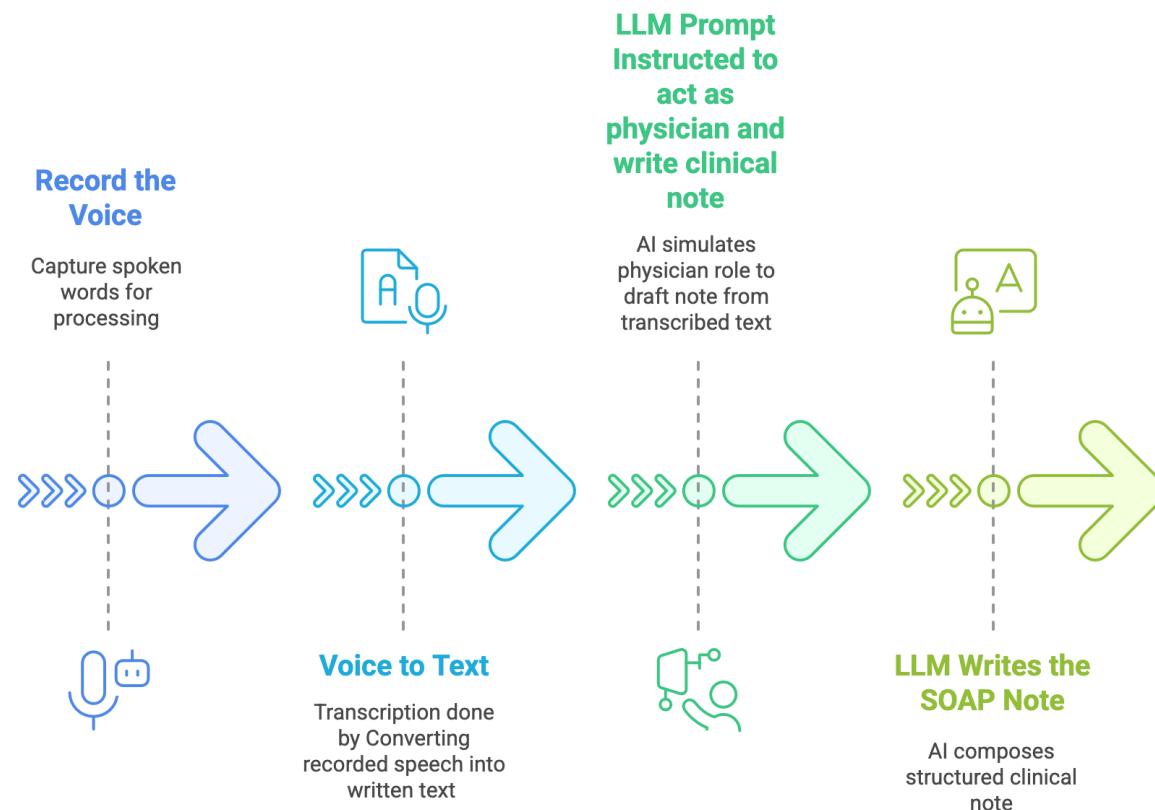


Extracting Data for Analytics

Pulling data for healthcare insights

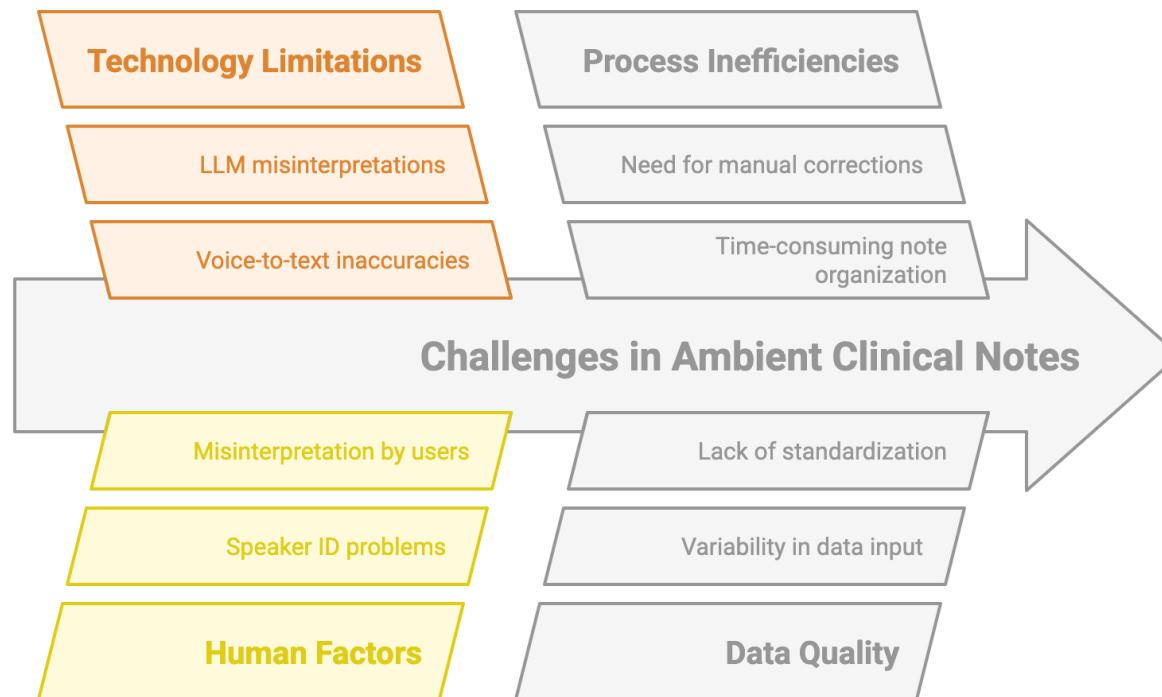


Ambient Listening Note Creation Process



Real-World Example: Ambient Clinical Notes

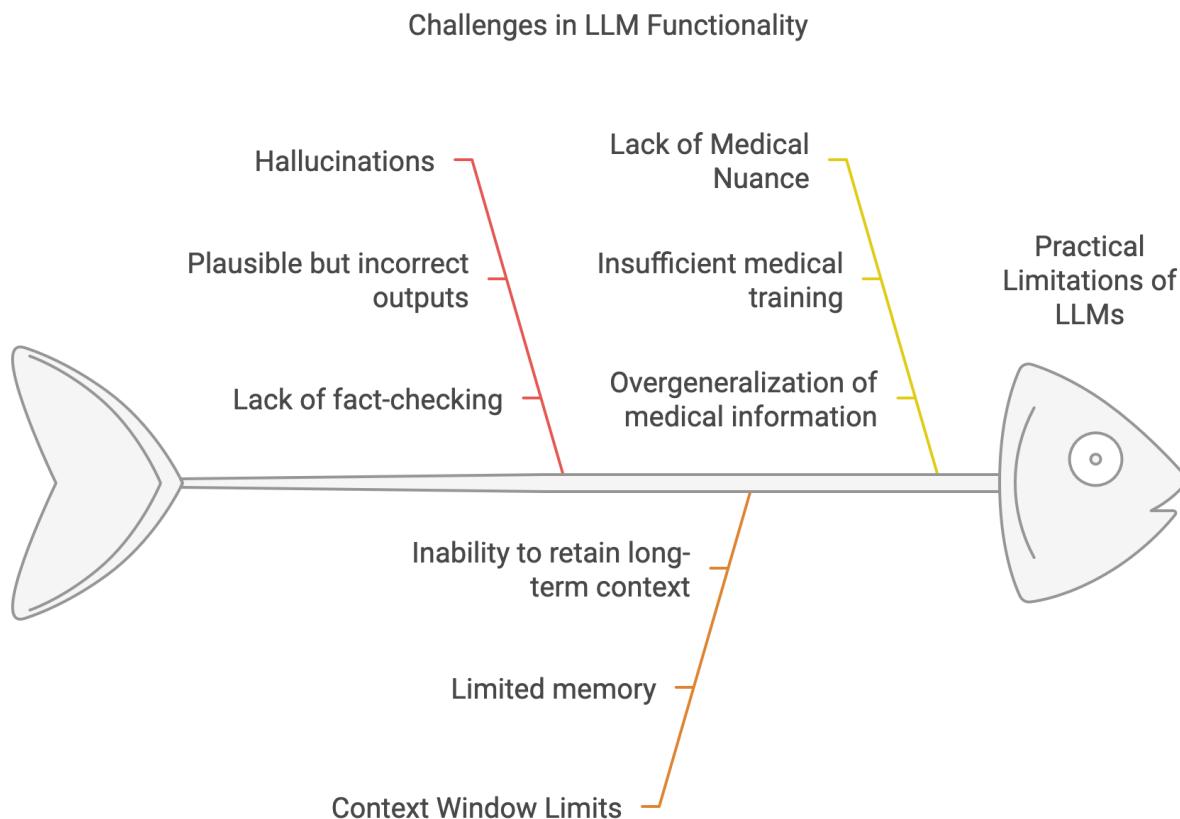
Analyzing Challenges in Ambient Clinical Notes



Discussion: What Surprises You About LLMs?

- Open floor for reactions and concerns.
- Discuss reliability, ethical issues, and bias.

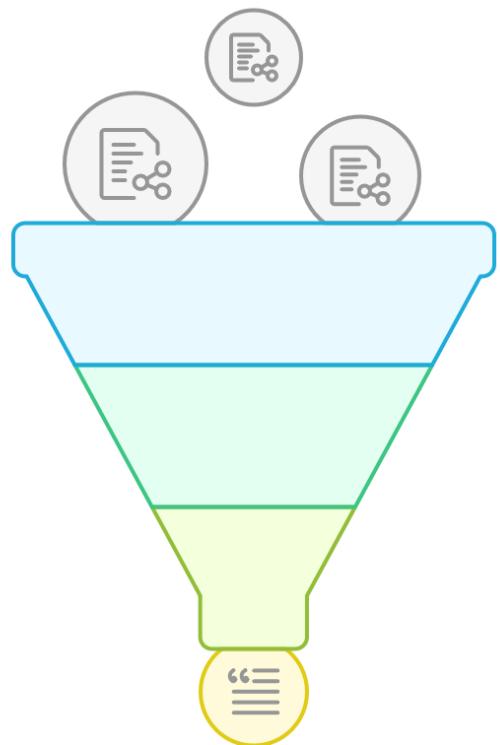
Practical Limitations & Current Challenges



Expert Audience Considerations

- Domain expertise is critical for oversight.
- LLMs can streamline research and EHR review.
- Potential to accelerate research and reduce manual work.

Case Study: Literature Review Automation



Abstract Scanning

LLMs scan and analyze abstracts



Summarization

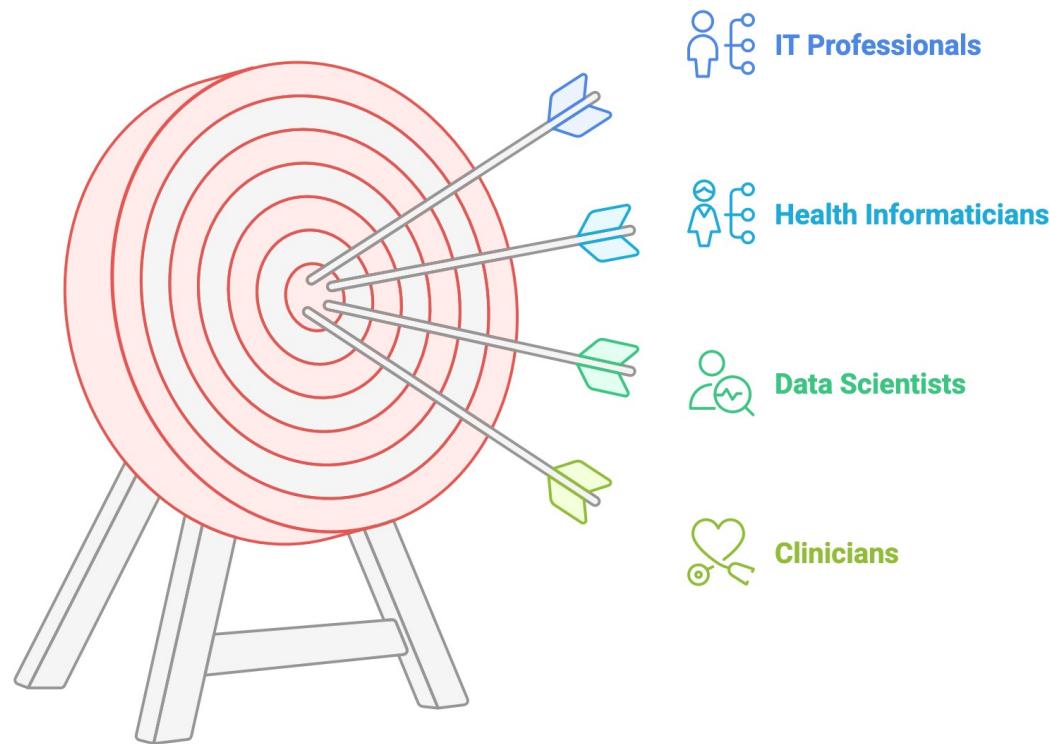
Generate concise summaries for reviews



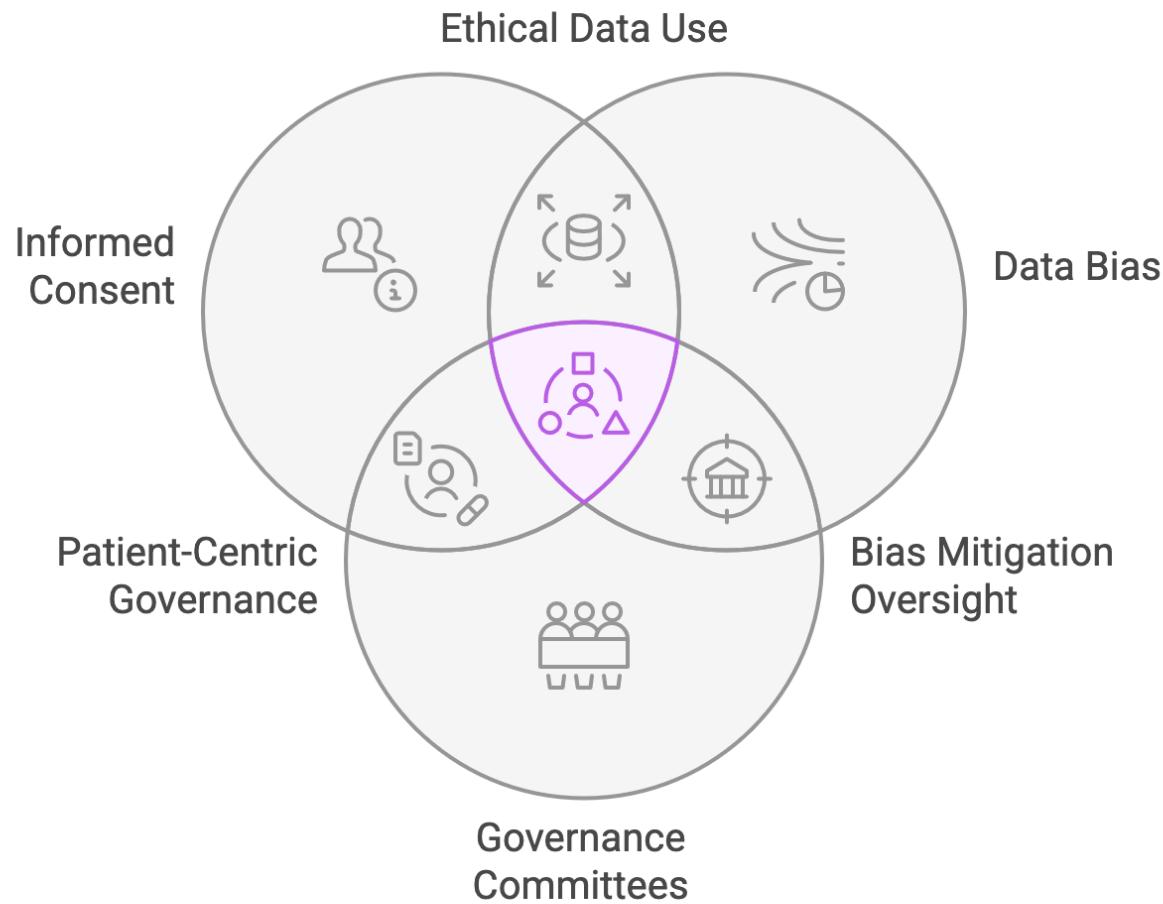
Nuance Omission Risk

Potential loss of critical details

Interdisciplinary Team Roles



Ethical Lens on LLM Introduction



Quick Recap: Segment 1

- LLMs predict words using statistical patterns.
- High potential in clinical workflows.
- Be mindful of limitations and biases.

I do not believe LLM's are going to result in Artificial General Intelligence

Pause for Questions

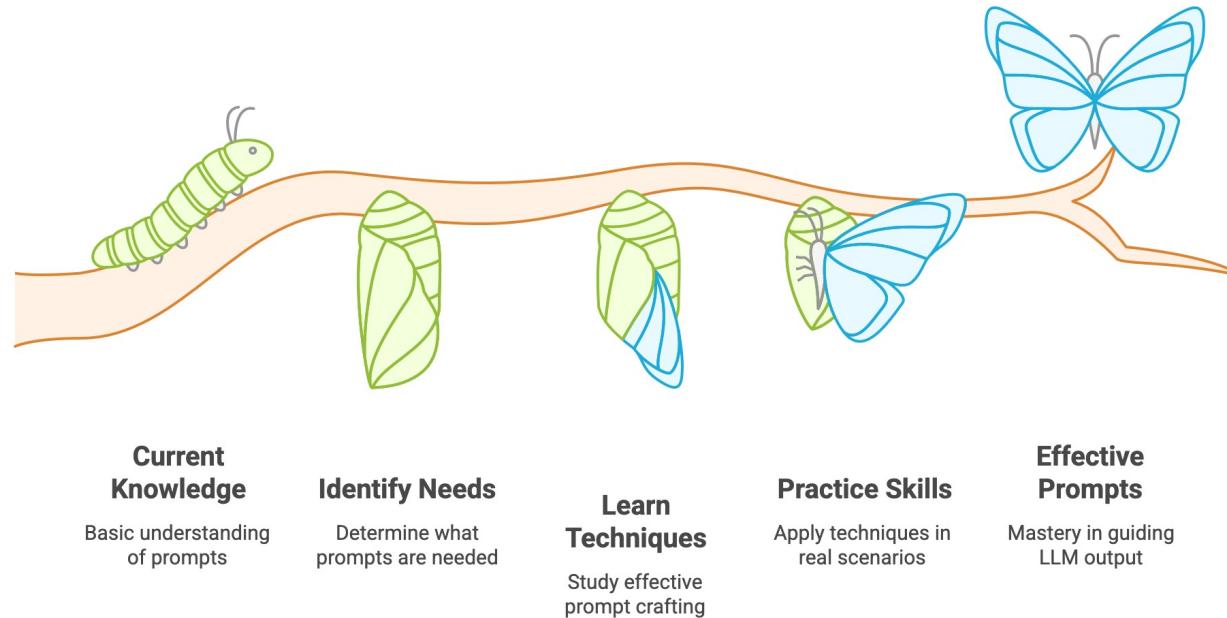
- Any clarifications on LLM so far?



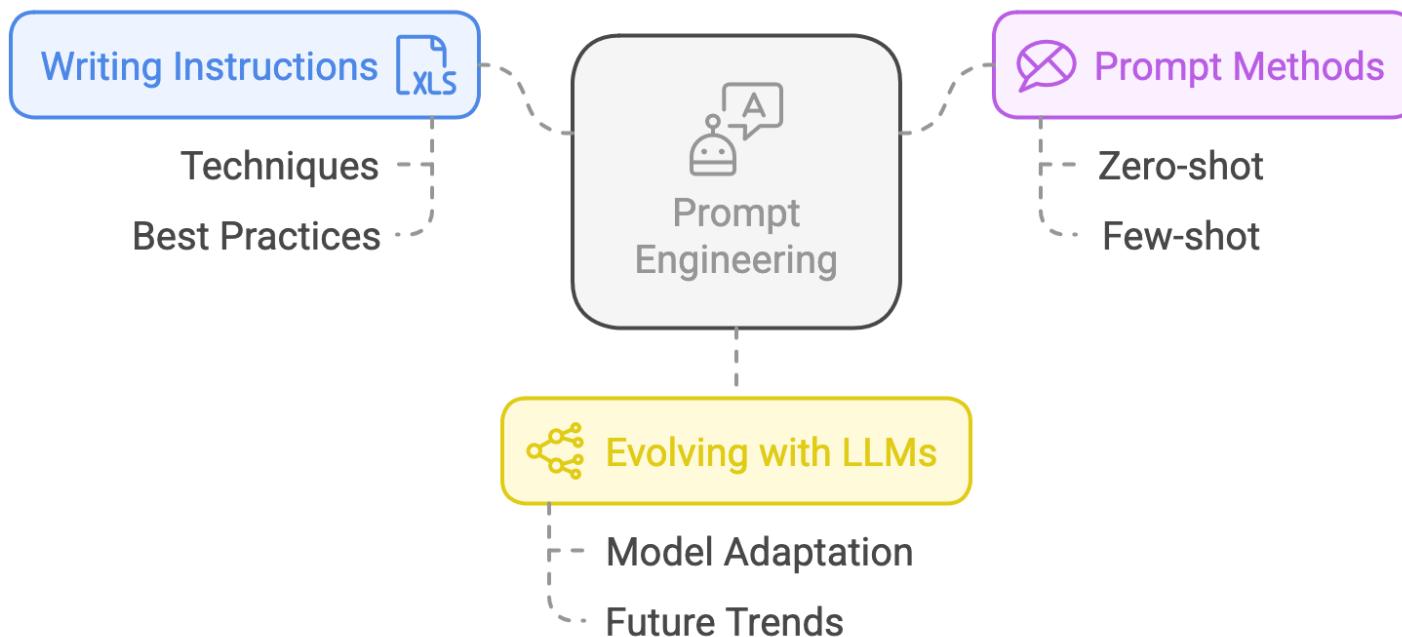
Owl Health

**Harnessing Large Language Models in Healthcare
Informatics:
A Comprehensive Tutorial
Jeremy R. Harper, MS, FAMIA
Owl Health Works LLC, Indianapolis, Indiana, USA**

Transition to Next Topic: Prompt Engineering



What is Prompt Engineering?



Prompt Anatomy



The Anatomy of an o1 Prompt

I want a list of the best medium-length hikes within two hours of San Francisco.

Each hike should provide a cool and unique adventure, and be lesser known.

For each hike, return the name of the hike as I'd find it on AllTrails, then provide the starting address of the hike, the ending address of the hike, distance, drive time, hike duration, and what makes it a cool and unique adventure.

Return the top 3.

Be careful to make sure that the name of trail is correct, that it actually exists, and that the time is correct.

--

For context: my girlfriend and i hike a ton! we've done pretty much all of the local SF hikes, whether that's presidio or golden gate park. we definitely want to get out of town -- we did mount tam pretty recently, the whole thing from the beginning of the stairs to stinson -- it was really long and we are definitely in the mood for something different this weekend! ocean views would still be nice. we love delicious food. one thing i loved about the mt tam hike is that it ends with a celebration (Arriving in town to breakfast!) The old missile silos and stuff near Discovery point is cool but I've just done that hike probably 20x at this point. We won't be seeing eachother for a few weeks (she has to stay in LA for work) so the uniqueness here really counts.

Goal

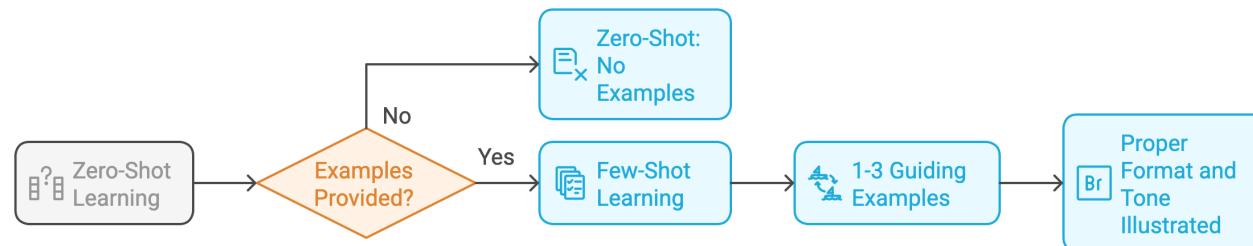
Return Format

Warnings

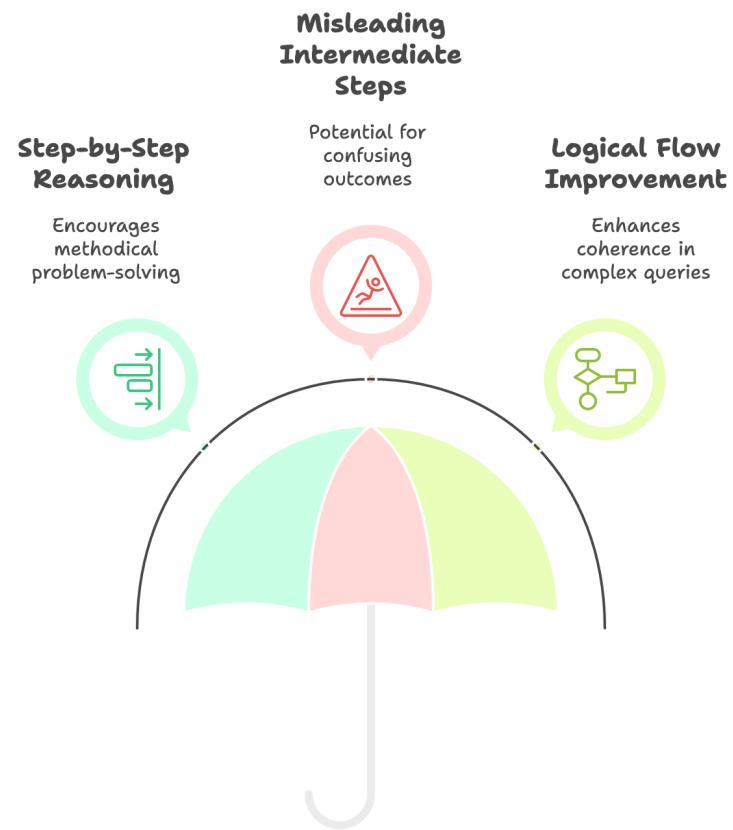
Context Dump

Zero-Shot vs. Few-Shot

- Zero-shot: No examples provided.
- Few-shot: Provide 1–3 guiding examples.
- Examples help illustrate proper format and tone.



Chain-of-Thought Prompting



Zero Shot

- "Given the following patient symptoms: persistent cough for 3 weeks, fever in the evenings, unexplained weight loss of 10 pounds in the last month, and occasional night sweats, what potential diagnoses should be considered? Please rank them by likelihood and explain your reasoning."

One Shot

I'd like you to suggest potential diagnoses based on patient symptoms. Here's an example of how I'd like you to respond:

Example: Patient symptoms: Fatigue, excessive thirst, frequent urination, blurry vision
Potential diagnoses:

1. Diabetes Mellitus (Most likely) - The combination of polydipsia, polyuria, fatigue and visual changes strongly suggests elevated blood glucose levels.
2. Hyperthyroidism - Can cause fatigue and polydipsia, but would typically present with weight loss and heat intolerance.
3. Chronic kidney disease - Could explain fatigue and polyuria, but would typically present with edema and hypertension.

Now, please analyze this new case: Patient symptoms: severe headache, stiff neck, fever, sensitivity to light, confusion"

Few Shot

I need help identifying potential diagnoses from patient symptoms. Please follow these examples:

Example 1: Symptoms: Chest pain radiating to left arm, shortness of breath, nausea, sweating Diagnoses:

1. Myocardial infarction - Classic presentation with chest pain radiating to left arm with associated autonomic symptoms.
2. Angina pectoris - Similar to MI but typically triggered by exertion and relieved by rest.
3. Aortic dissection - Can present with severe chest pain, though typically described as tearing and radiating to the back.

Example 2: Symptoms: Fever, productive cough with yellow sputum, chest pain when breathing deeply, fatigue Diagnoses:

1. Pneumonia - Fever with productive cough and pleuritic chest pain is highly suggestive.
2. Acute bronchitis - Possible but less likely to cause pleuritic chest pain.
3. Tuberculosis - Should be considered especially with risk factors or chronic symptoms.

Example 3: Symptoms: Right lower quadrant abdominal pain, nausea, low-grade fever, loss of appetite Diagnoses:

1. Appendicitis - Classic presentation with RLQ pain and migration from periumbilical region.
2. Ileitis (Crohn's disease) - Can mimic appendicitis but typically has longer history.
3. Ovarian cyst (in females) - Can cause RLQ pain but typically without systemic symptoms.

Now, analyze this new case: Symptoms: Sudden onset vertigo, nausea, horizontal nystagmus, hearing loss in right ear"

Chain Of Thought

Chain of Thought Diagnostic Approach

Initial Prompt:

"A 45-year-old female presents with progressive fatigue over 3 months, joint pain primarily in hands and wrists, morning stiffness lasting more than an hour, and a butterfly-shaped rash across her cheeks and nose. Let's think through potential diagnoses step by step."

Step 1: Categorize the Symptoms

"First, let's categorize these symptoms:

Constitutional: Progressive fatigue (3 months)

Musculoskeletal: Joint pain in hands/wrists, morning stiffness > 1 hour

Dermatological: Butterfly-shaped rash on face

Timeline: Chronic, progressive over months

These symptoms suggest a systemic inflammatory process rather than an infectious or acute condition."

Step 2: Consider Organ Systems Involved

"The presentation involves multiple organ systems:

Skin (rash)

Joints (pain, stiffness)

Constitutional (fatigue)

Multi-system involvement points toward autoimmune conditions rather than organ-specific diseases."

Step 3: Analyze Pattern Recognition

"The specific pattern of:

Symmetric small joint involvement

Prolonged morning stiffness

Malar/butterfly rash

Systemic symptoms

This constellation strongly suggests connective tissue disease, particularly in the lupus-rheumatoid spectrum."

Step 4: Differential Diagnosis Ranking

"Based on the analysis, my differential diagnosis would be:

Systemic Lupus Erythematosus (SLE) - The malar rash is highly specific, and when combined with joint symptoms and fatigue, strongly suggests SLE.

Rheumatoid Arthritis - The joint symptoms fit, but the facial rash is not typical of RA.

Mixed Connective Tissue Disease - Could explain multiple system involvement.

Dermatomyositis - Can cause facial rash but typically includes muscle weakness.

Drug-induced lupus - Would need medication history to evaluate."

Step 5: Recommended Workup

"The appropriate next steps would include:

ANA titer and pattern

Anti-dsDNA antibodies

Anti-Sm antibodies

Complement levels (C3, C4)

Complete blood count

Comprehensive metabolic panel

Urinalysis to check for proteinuria/hematuria

ESR and CRP to assess inflammation

Based on these results, a rheumatology referral would be indicated for confirmation and management."

Good Prompting Principles

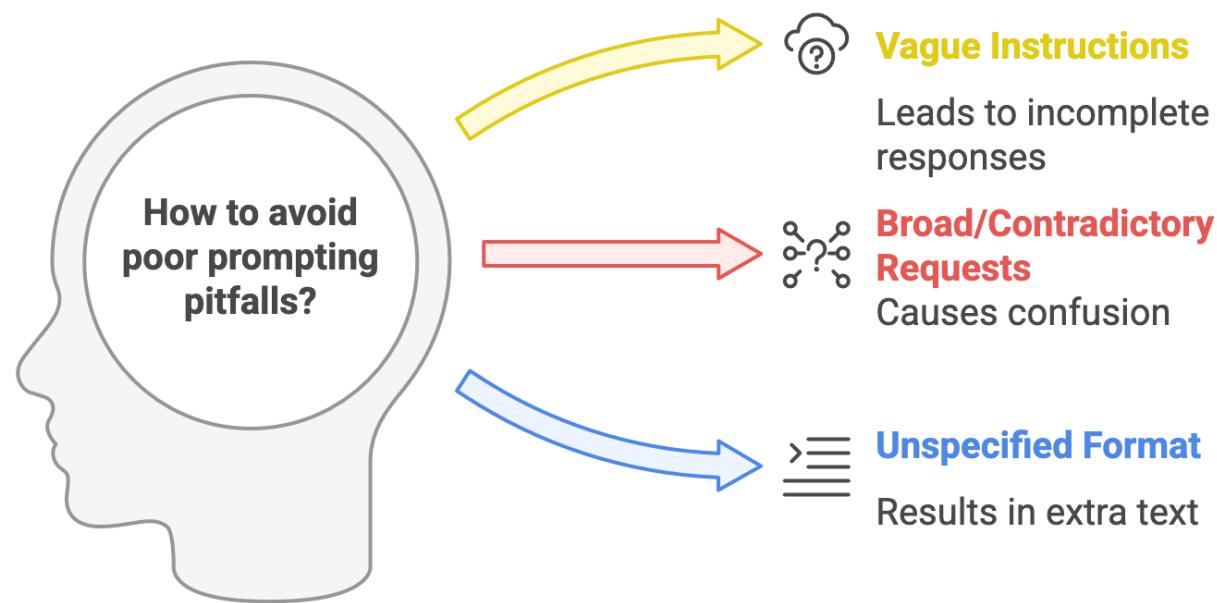


Shout Out: Lab Results Summary Prompt

Create a brief prompt that would help a doctor quickly assess a patient's blood work when anemia is suspected. Your prompt should include

- Role:
- Context:
- Instruction:

Pitfalls of Poor Prompts



Hands-On Exercise #1 Instructions

- In groups, draft a prompt (Or a process) for how you would normalize two clinical datasets.
- Share outcomes in 10 minutes.

How do I build this prompt?

- Demo

Interactive Moment: Live Prompt Demo

- Show a naive prompt vs. a refined prompt.
- Compare differences in output clarity and accuracy.
- Discuss why adding constraints helps.

(Hands-On) Prompt Engineering in Practice

- Experiment with how extra details change output style.
- Note any shifts in tone and inclusion of extraneous data.

Discussion: Observations from Hands-On

- Did increased specificity improve accuracy?
- Were there hallucinations or irrelevant details?
- Discuss variations in tone or style.

This Explains LLM Agents in 2 minutes

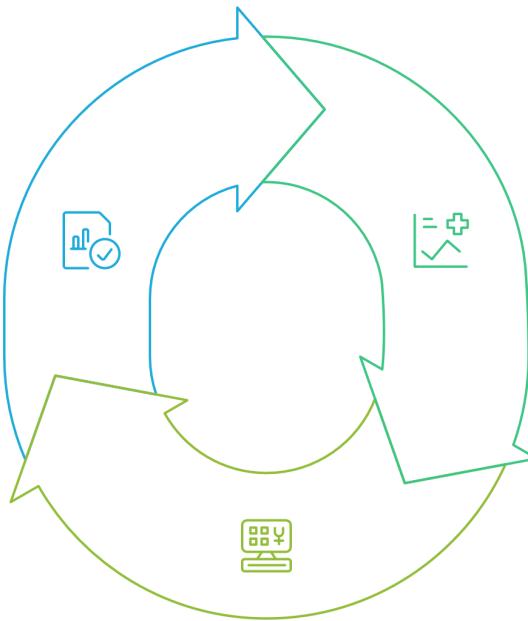
Agent 3 Diagnostic Evaluator

Agent 3 prompt: Assesses the refined diagnosis for accuracy, comprehensiveness, and clinical relevance, providing a quality score and specific improvement areas

Agent 1 Analyze Symptoms

Agent 1 LLM Prompt says: Process a patient's description of symptoms and generates a list of possible conditions

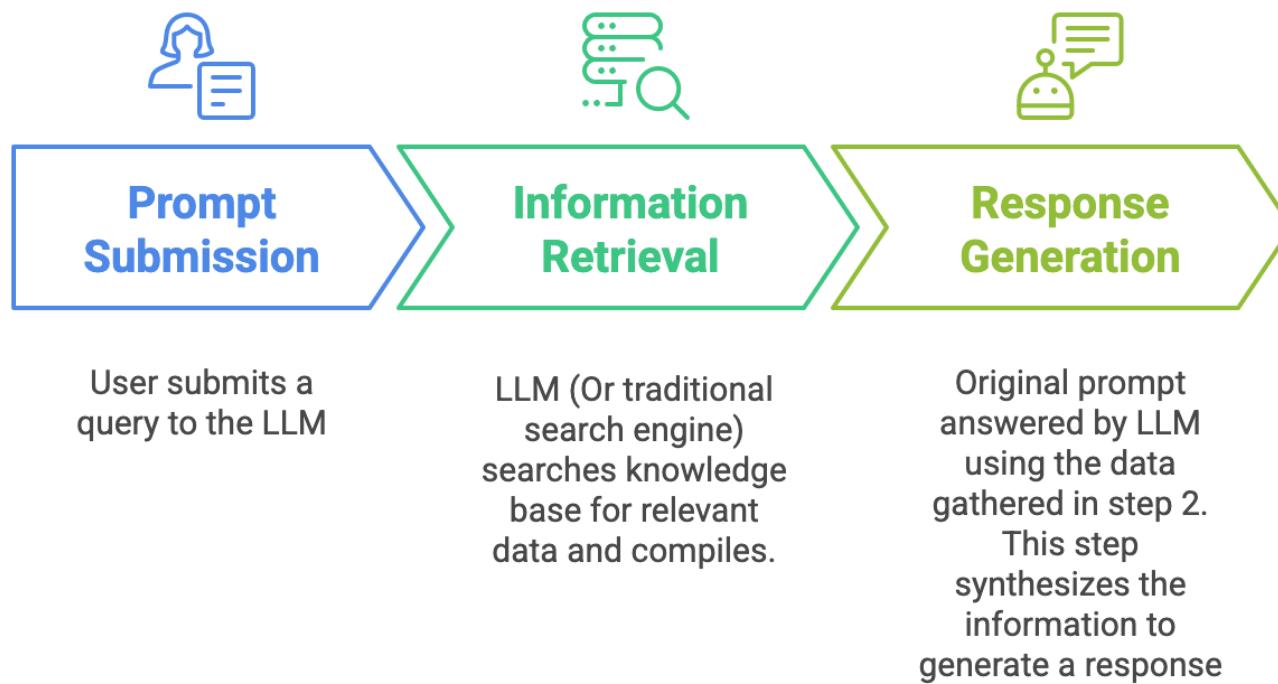
If Passed in from Agent 3 it follows the evaluators instructions when regenerating



Agent 2 Medical Specialist

Agent 2 uses an LLM prompt that says: Take the output of 1 and refine the initial analysis by incorporating medical knowledge to prioritize conditions based on likelihood and severity

RAG LLM in 2 Minutes





Owl Health

**Harnessing Large Language Models in Healthcare
Informatics:
A Comprehensive Tutorial
Jeremy R. Harper, MS, FAMIA
Owl Health Works LLC, Indianapolis, Indiana, USA**

Prompt Opportunities for Policy or Research

- Summarize new research papers succinctly.
- Generate policy briefs with clear bullet points.
- Risk: Using outdated knowledge if model isn't current.

Strategies for Clinical Prompts

- Clearly define scope: e.g., 'Focus only on medication interactions.'
- Include relevant patient history if needed.
- Validate outputs against real data.

Common Prompt Engineering Mistakes

- Overloading the prompt with unnecessary details.
- Ignoring the model's context size limitations.
- Expecting multi-step logic without guidance.

Prompt Versioning & Iteration

- Start simple and refine iteratively.
- Maintain records of best-performing prompts.
- Re-test after model updates or domain changes.

- I need you to act as a large language model prompt engineer and create a verbose prompt to do the following. <DESCRIPTION OF THE THING YOU WANT IT TO DO>. Again I do not want you to do those instructions I want you to build the large language model prompt like a prompt engineer that will achieve the task in a single shot prompt.

Ethical Prompting

De-identification Techniques

Reduces risk by removing identifiable information from training data.

Aggregate and Anonymize

Minimizes data exposure by combining and anonymizing data sets.

Re-identification Risks

Highlights the importance of being aware of potential re-identification.



Example Prompt

- Generic: "Based on the patient's ethnicity and socioeconomic background, determine the most likely diseases they might have."
- Specific: "If a patient is a white male, living in a low-income area, determine the most likely diseases he might have. Focus on conditions commonly associated with poor nutrition, lack of access to healthcare, and high rates of smoking or substance abuse in low-income white populations."

Example: Prompt for Hypothesis Generation

- Provide a clear research question.
- Ask the model for known risk factors or data gaps.
- Evaluate the provided references or disclaimers.

Summary: Prompt Engineering Key Takeaways

- Concrete instructions plus context yield better results.
- Iteration and versioning are essential.
- Ethical and secure usage is critical in healthcare.



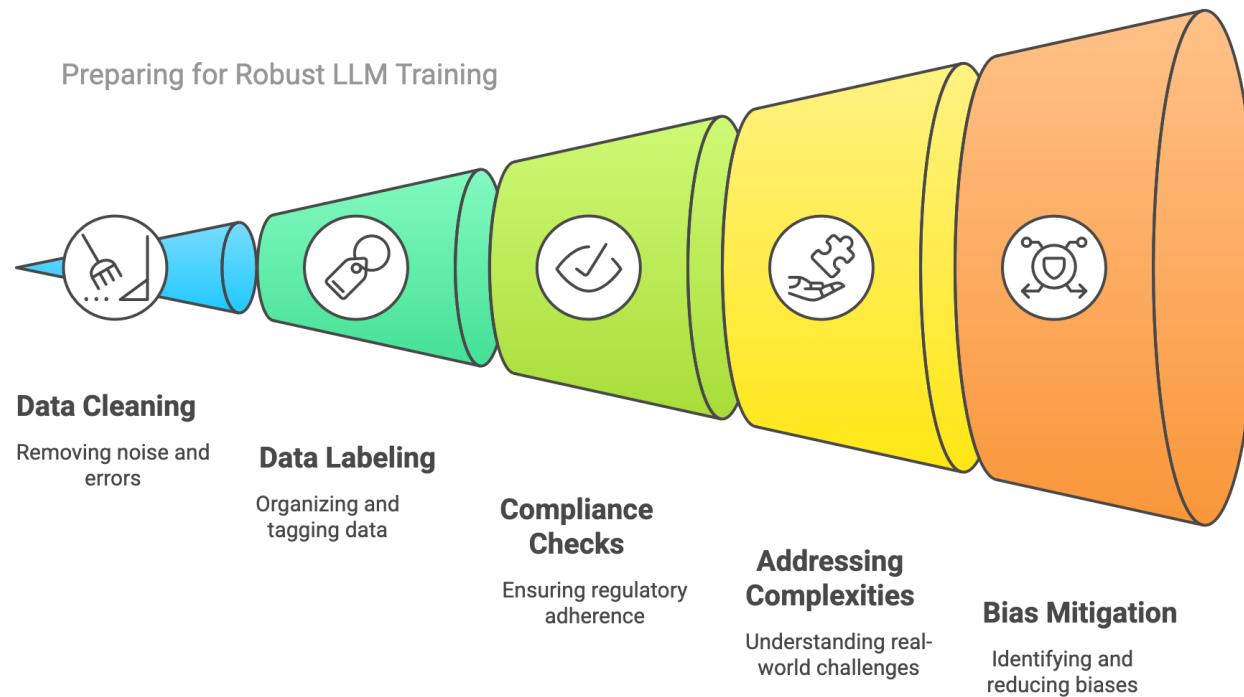
Owl Health

**Harnessing Large Language Models in Healthcare
Informatics:
A Comprehensive Tutorial
Jeremy R. Harper, MS, FAMIA
Owl Health Works LLC, Indianapolis, Indiana, USA**

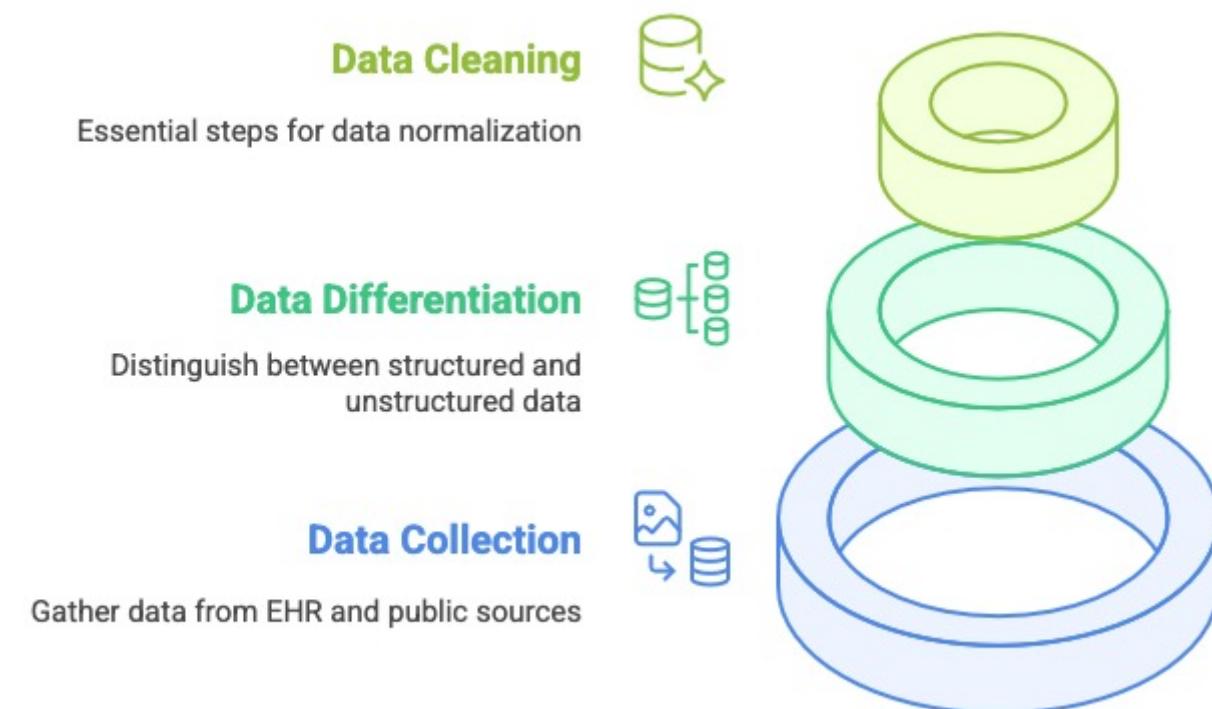
Segment 3: Data Preparation & Ethical Considerations - Overview

- Data sourcing and preprocessing.
- Regulatory compliance (HIPAA, IRB).
- Bias detection and mitigation strategies.

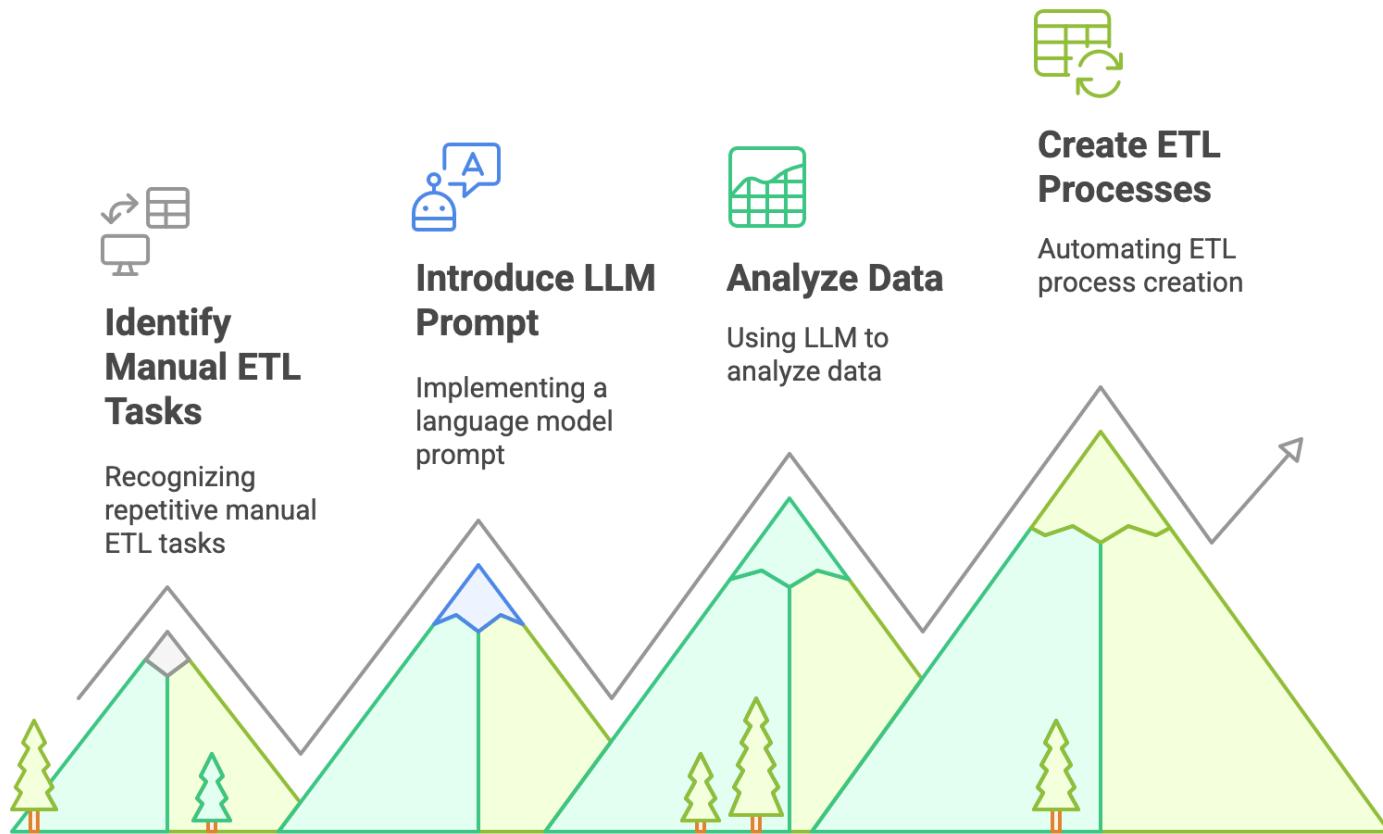
Next: Data Prep & Ethical Considerations



Data Management for LLMs



Automating ETL Processes with LLM Prompts



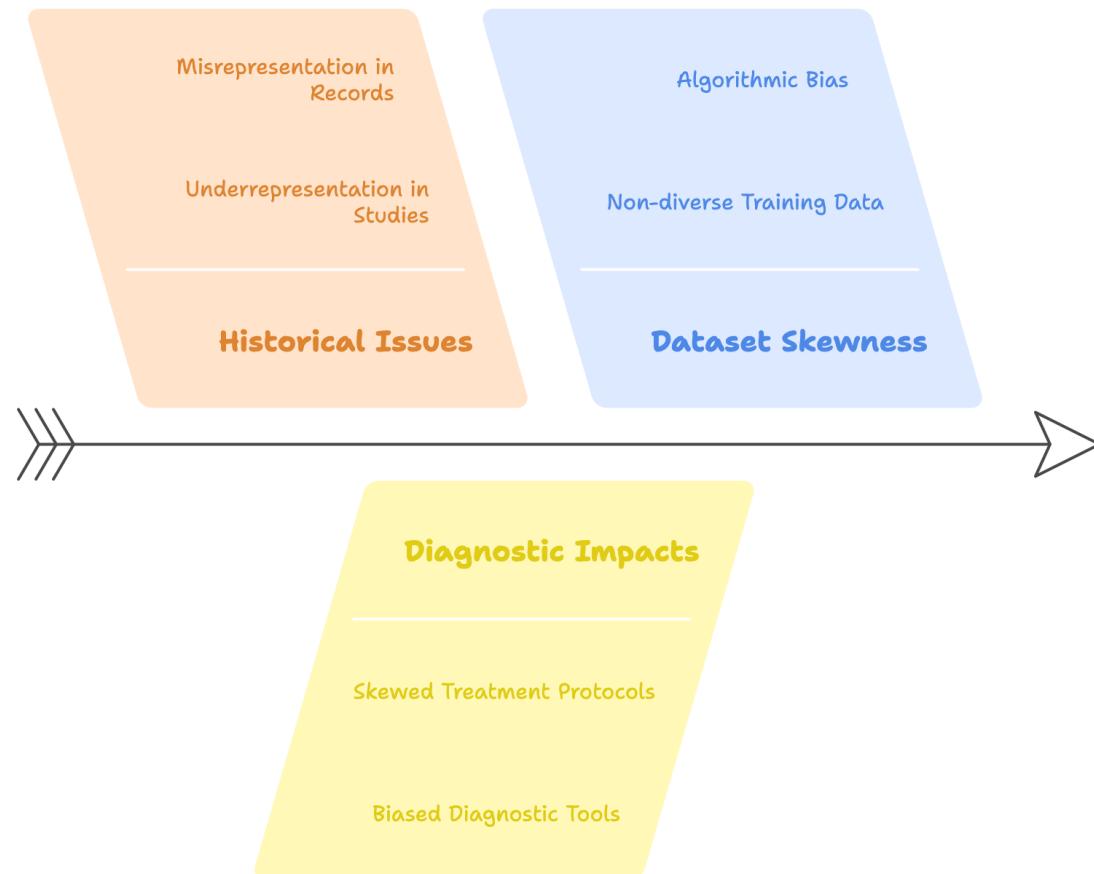
Handling PHI and Sensitive Data

- Use de-identification techniques for training data.
- Aggregate and anonymize data where possible.
- Be aware of potential re-identification risks.

Regulatory Frameworks

- HIPAA: Rules for Protected Health Information (PHI) in the US.
- GDPR: Data minimization and the right to be forgotten in the EU.
- IRB oversight for clinical research.

Bias in Healthcare Data



Methods for Detecting Bias

- Perform statistical checks (e.g., distribution analysis).
- Test on known sensitive attributes.
- Involve diverse experts in evaluation.

Image Generator bias



Mitigation Strategies

- Balance data or oversample underrepresented groups.
- Post-process outputs to remove biased language.
- Establish continual monitoring and feedback loops.

Hands On

- Ask an LLM to generate a synthetic dataset for you
-

Please generate a table of a typical ambulatory clinic patient load.
I want the table to have 100 rows. I want demographics.
Name/Age/Gender/Race/Height/Weight/City/State/Employer

Interactive Discussion: Ethical Dilemma Cases

- What scenario's have you seen or do you imagine?



Owl Health

**Harnessing Large Language Models in Healthcare
Informatics:
A Comprehensive Tutorial
Jeremy R. Harper, MS, FAMIA
Owl Health Works LLC, Indianapolis, Indiana, USA**

Role of Governance Committees

- Establish AI ethics boards in hospitals.
- Conduct regular reviews of algorithm performance.
- Mandate oversight and update check-ins.

Legal Implications of LLM Mistakes

Managing Risks and Responsibilities in LLM Clinical Use



Overcoming Data Quality Issues

Data Processing Techniques



Terminology Standardization

Standardizing terminologies like SNOMED and ICD is essential.



NLP Techniques

Utilizing advanced NLP techniques helps in cleaning free-text notes effectively.



Continuous Curation

Ongoing curation is crucial for maintaining robust performance.

Balancing Privacy and Utility

- Over-deidentification can reduce model accuracy.
- Consider federated learning or on-premise training.
- Stakeholder buy-in is critical.

Practical Tips for Ethical Data Prep

- Begin with a small pilot using a limited data subset.
- Document every data-preparation step for accountability.
- Establish institutional oversight early on.

Bridge to Next Session: Applications & Hands-On

- Next: Explore real-world applications of LLMs.
- See how data prep influences final outputs.

Break Announcement (Segment 3 End)

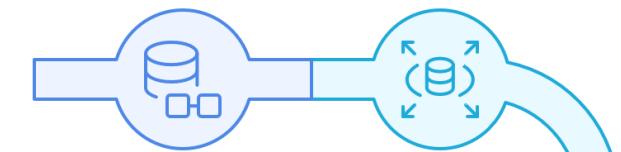
- 30-minute break.
- Feel free to ask any questions offline.
- Next segment: Practical usage in clinic, research, and administration.

Schedule (Restart: 10:30am)

Mastering LLMs in Healthcare Informatics

Session 1 Part 1

Foundations of Large Language Models in Healthcare Informatics

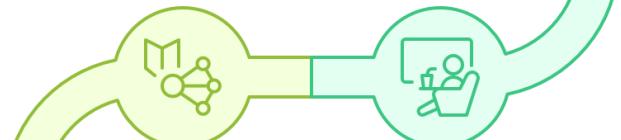


Session 1 Part 2

Data Preparation and Ethical Considerations

Session 2 Part 3

Practical Applications of LLMs

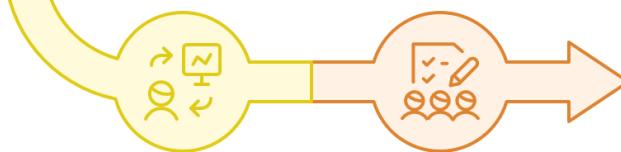


Break

Refreshment Break

Session 2 Part 4

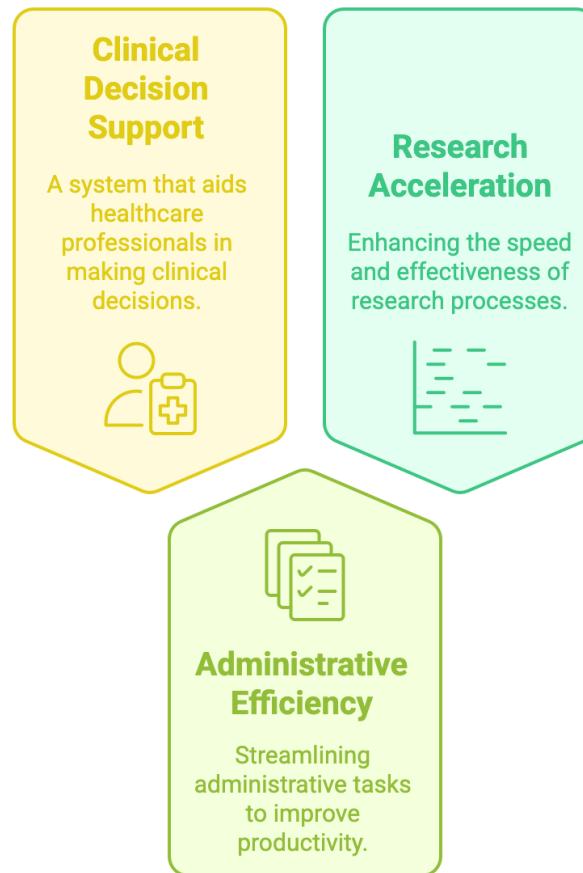
Future Trends and Implementation Strategies



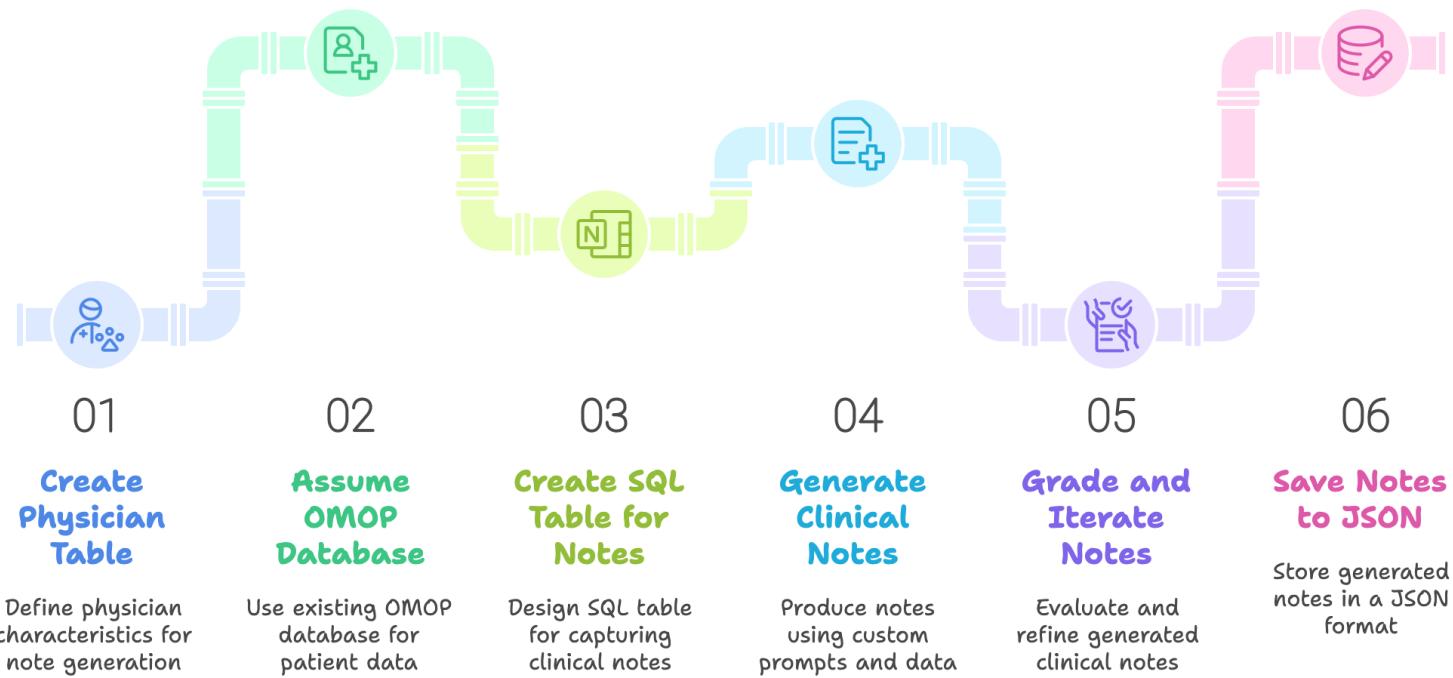
Session 2 Part 5

Collaborative Project and Wrap-Up

Real-World LLM Use Cases



Workflow for Generating Synthetic Clinical Notes

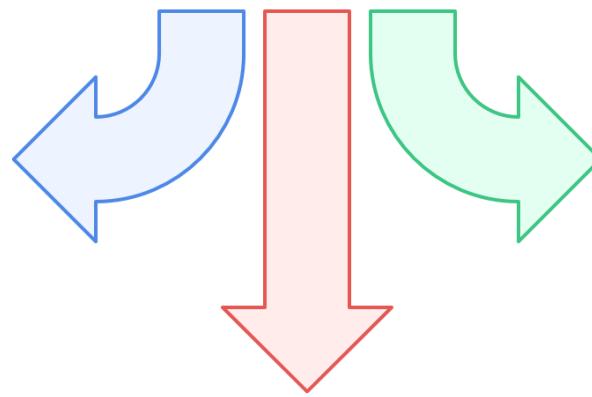


Clinical Decision Support (CDS)

How to effectively use Clinical Decision Support (CDS) systems?

Use for Diagnosis

CDS can suggest differential diagnoses and treatment guidelines, enhancing clinical decision-making.



Summarize Data

CDS can summarize data from labs, imaging, and notes, providing a comprehensive view of patient information.

Awareness of the risk of over-reliance is crucial to avoid missing edge cases.

Example CDS Workflow



Administrative Efficiency

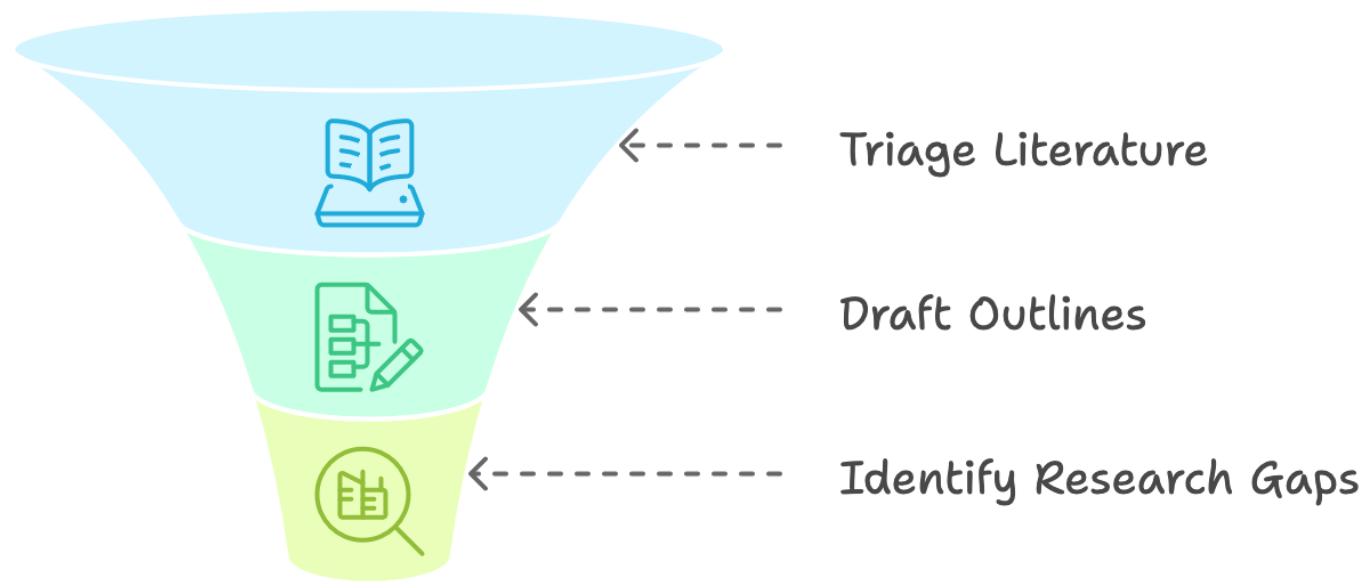
Automate
Summaries

Patient
Reminders

Translate
Notes

Research Acceleration Example

Literature Review



Focused Research
Strategy

Hands-On Demonstration Setup

- Show a user-friendly LLM interface (e.g., ChatGPT/Hugging Face).
- Step-by-step: Input → Output demonstration.
- Participants encouraged to try on their devices.

Live Demo: Summarizing a Clinical Note

- Paste a de-identified clinical note.
- Prompt: 'Summarize key diagnoses, treatments, follow-up steps'.
- Evaluate formatting and completeness.

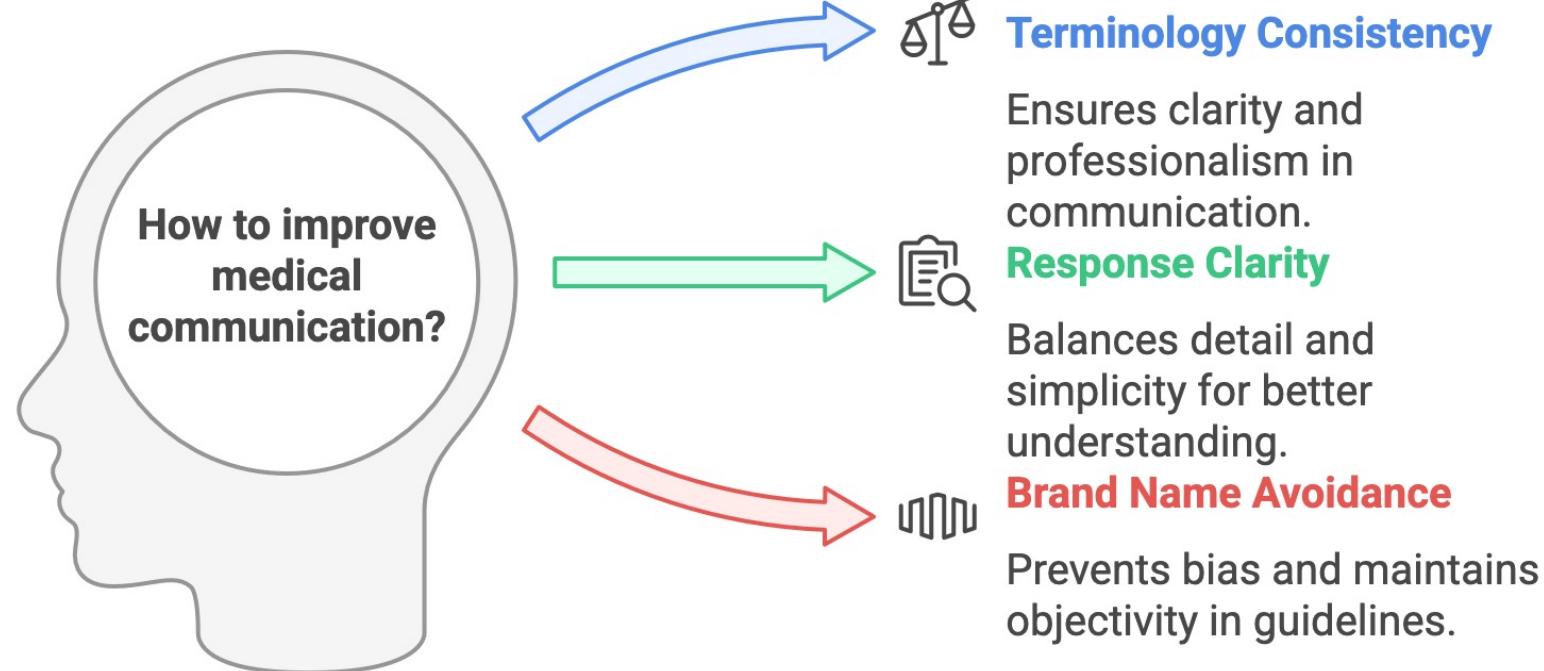
Demo Discussion: Accuracy vs. Speed

- Assess whether the output is coherent and correct.
- Identify missing or extraneous details.
- Discuss potential prompt enhancements.

(Hands-On #2) Interactive Demo: Your Turn

- In small groups, craft a prompt for a sample clinical scenario.
- Evaluate the LLM output and discuss improvements.

Common Issues Encountered Live



Future Trends in LLM Clinical Use

- Voice-based interactions (ambient listening).
- Real-time analytics for telemedicine.
- Integration with wearable device data.

Implementation Challenges

- EHR integration complexity.
- Building clinician trust in the system.
- Keeping the model updated with new guidelines.

Efficiency Gains vs. Liability

LLM Output Considerations



Time Savings

Highlights the efficiency gained by reducing documentation efforts.



Risk Assessment

Discusses potential dangers of incorrect outputs from LLMs.



Human Oversight

Emphasizes the importance of human review on final outputs.

Real-World Experience: Automated Documentation

- Trial to generate discharge notes in a mid-sized hospital reduced discharge note time by 40%.
- 5% of notes contained errors requiring manual review.

Discussion: Implementing in Your Institution

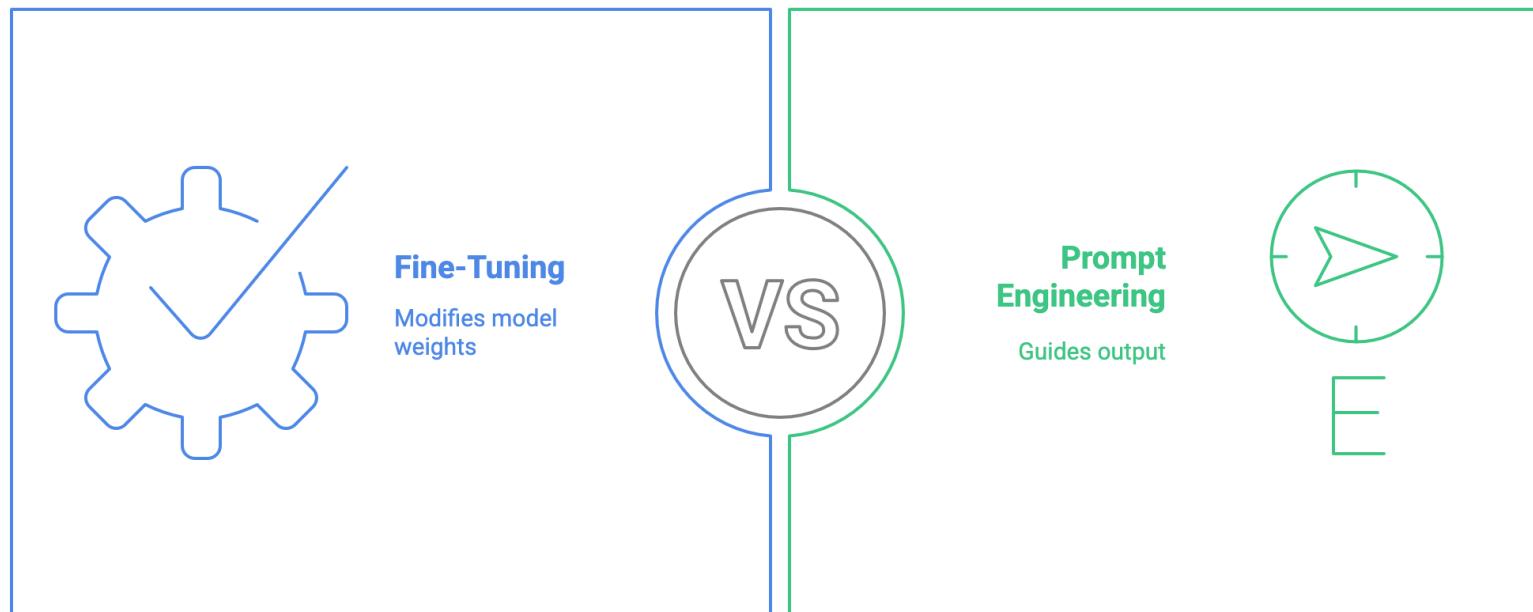
- Which clinical areas offer the highest payoff?
- Identify key stakeholders: IT, clinical leads, compliance.
- Pilot with a small group before full rollout.

Segment 5: Advanced LLM Architecture & Implementation Strategies

- Fine-tuning & domain adaptation.
- Model deployment: Cloud vs. on-premise.
- Handling performance at scale.

Fine-Tuning vs. Prompt Engineering

Choose the best approach for healthcare applications



Potential for Continual Learning

How to manage and update medical models?

Update Models

Keeps models current with new data, improving accuracy and relevance.

Explore Online Learning (RAG)

Allows models to adapt in real-time but requires careful implementation to avoid errors.

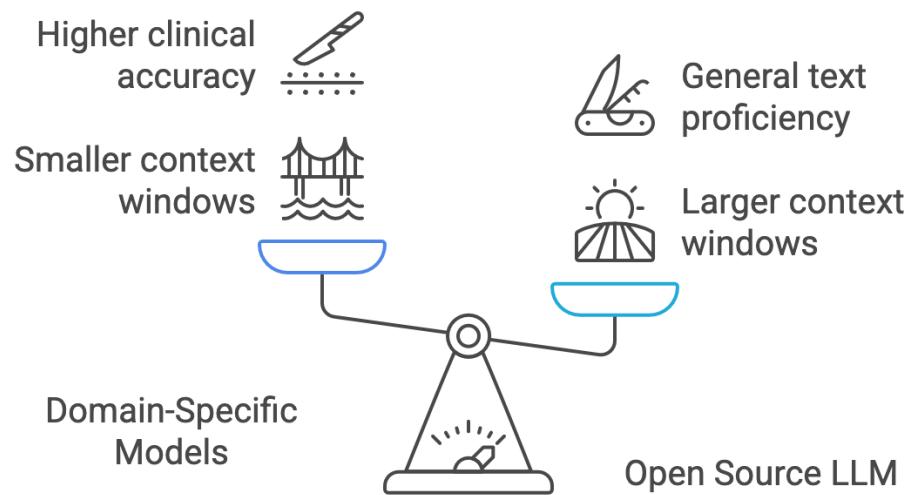
Ensure Knowledge Retention (Finetuning)

Preserves valuable insights from previous data, preventing loss of knowledge.



Domain-Specific Models (BioBERT, ClinicalBERT)

- Pretrained on specialized medical corpora.
- More accurate for clinical text than general models.



Comparing Model Specialization and Context Size

Data Requirements for Fine-Tuning

Data Preparation for Fine-Tuning

Labeled Data Acquisition

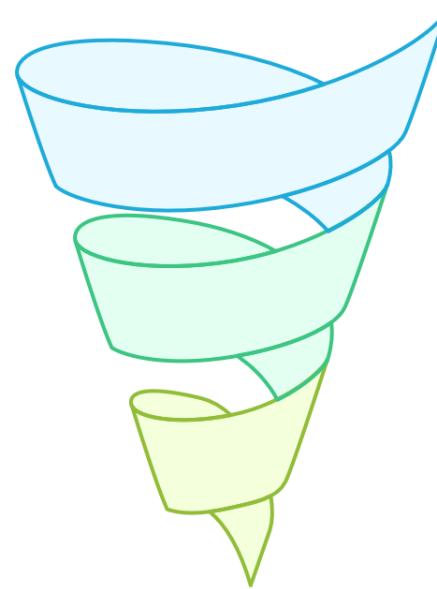
Collecting data with appropriate labels

Volume Expansion

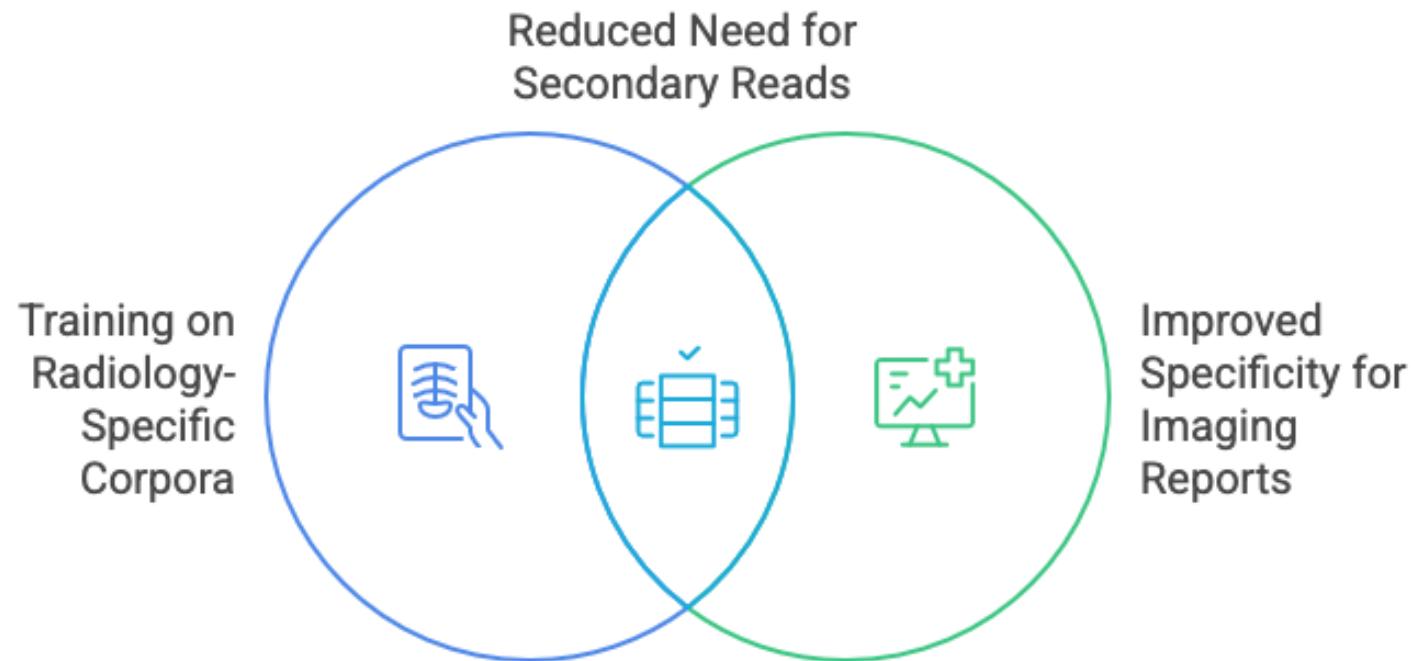
Increasing data quantity for generative tasks

Data Cleaning

Ensuring data quality through cleaning



Example: Fine-Tuning for Radiology Summaries



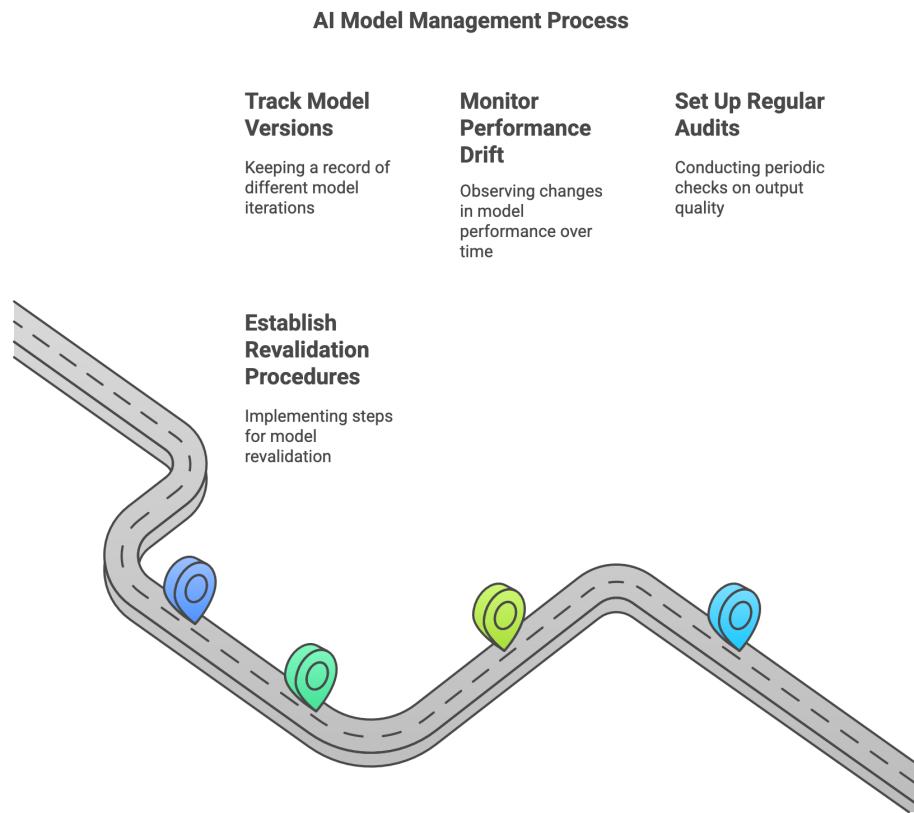
Cloud Deployment vs. On-Premise

- Cloud: scalable with lower hardware overhead.
- On-premise: offers tighter PHI control, though costlier.
- Hybrid solutions (e.g., VPC) can combine benefits.

Runtime Performance & Scalability

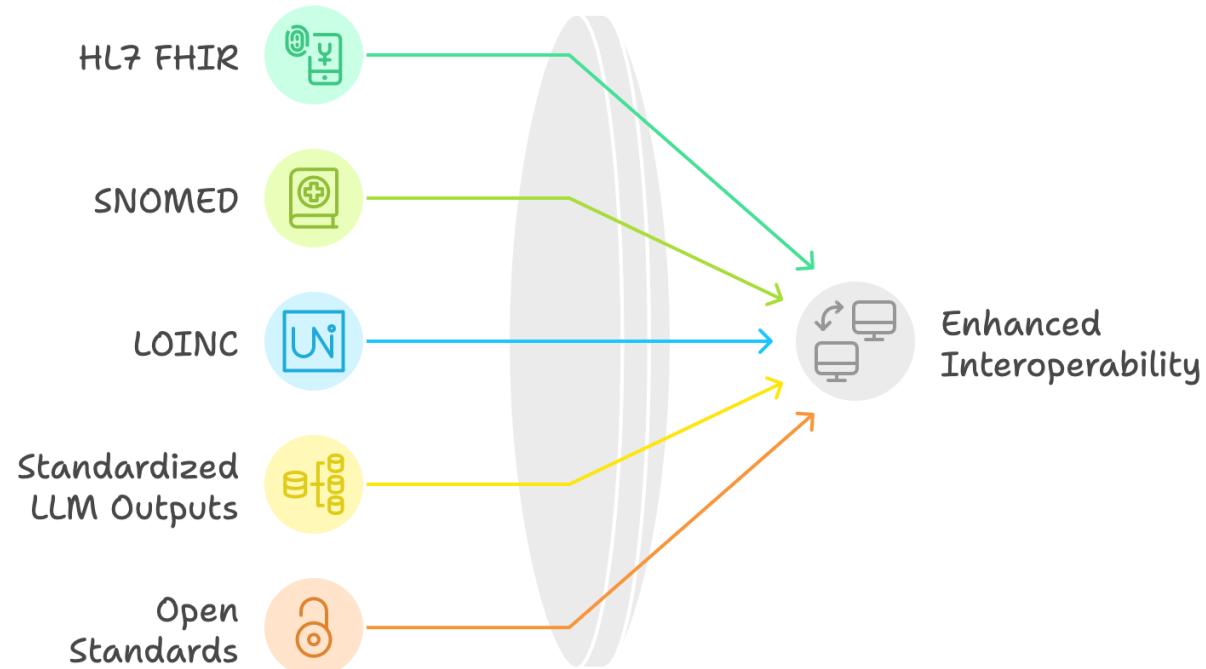
- Utilize GPU/TPU acceleration for faster inference.
- Batch processing can optimize large tasks.
- Watch for memory and concurrency bottlenecks.

Monitoring & Maintenance



Integrating LLM into Existing EHRs

Achieving Seamless Data Integration



Real-World Example: Ambient Notes at Scale

- A large hospital system processed 100k patient visits in 6 months.
- Achieved a 25% reduction in documentation time.
- Occasional errors require manual review.

Risk Management Strategies

Controlled Pilot Deployment

Implementing new strategies in a limited, controlled setting to assess effectiveness.

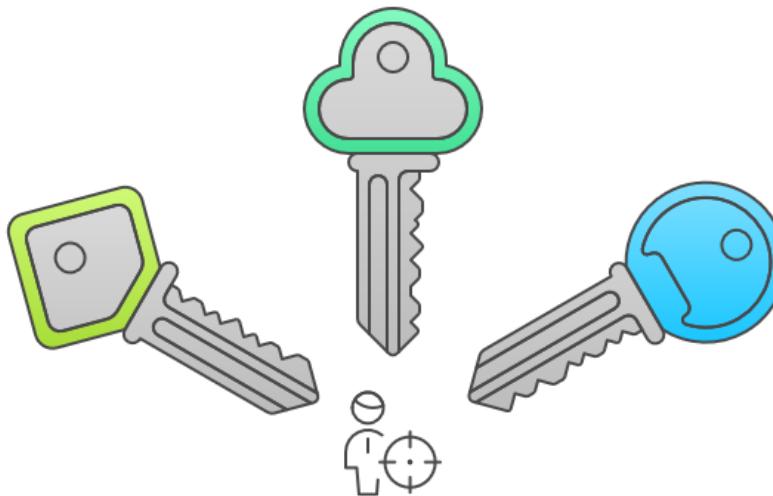
Risk Management Strategies

A/B Testing

Comparing different approaches using control groups to identify the most effective solution.

Fail-Safe Mechanisms

Ensuring safety and reliability through manual oversight and backup systems.



Scaling Across Multiple Departments

- Different specialties may require tailored models.
- Customize prompts or fine-tuned models for each department.
- Central oversight ensures consistency.

Team Requirements for Success

- Data scientists/ML engineers for model training.
- Domain experts for evaluation and validation.
- Project managers for coordination between departments.

Security Considerations Revisited

- Implement strict access control measures.
- Log all user interactions with the LLM.
- Consider prompt-level encryption for sensitive information.

Case Study: Large-Scale EHR Integration

- Phased implementation: pilot → partial rollout → full deployment.
- Key lessons: strong clinical champion and iterative prompt refinement.
- High ROI achieved over 12 months, but with significant upfront costs.

Discussion Prompt: Which Architecture Suits Your Setting?

- Cloud versus on-premise for your data sensitivity.
- General LLM vs. domain-specific fine-tuned models.
- Share reflections on your current environment.

Quick Knowledge Check

- Name one advantage of using a domain-specific LLM.
- Name one challenge with scaling the model.
- Reflect on a similar scenario in your institution.

Hands-On: Collaborative Implementation Planning

- Choose a healthcare domain (e.g., oncology, cardiology).
- Outline a pilot project using an LLM.
- Identify data sources, prompt strategies, and other constraints.

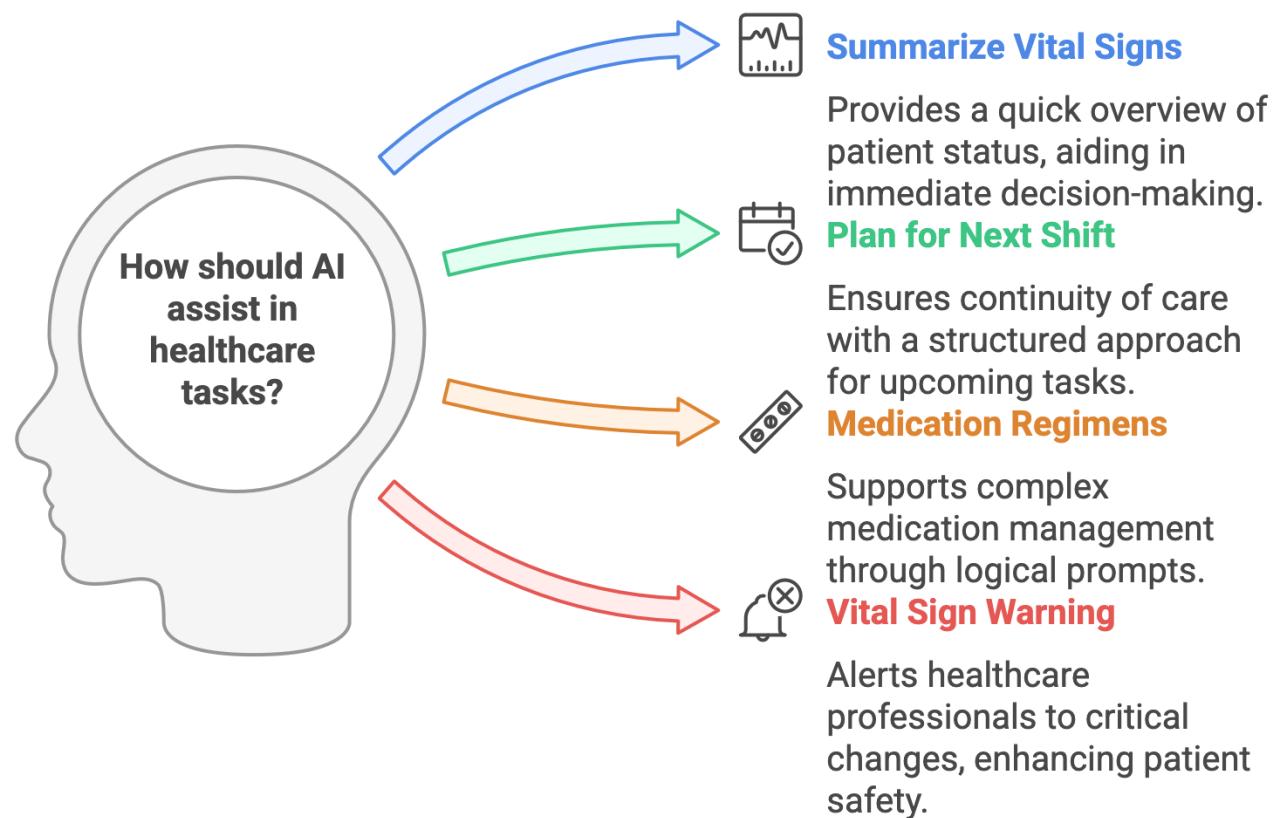
Instructions for Group Work

- Step 1: Define the scope & problem statement (5 minutes).
- Step 2: Decide on data requirements and initial prompt ideas (5 minutes).
- Step 3: Outline risk management and success metrics (5 minutes).

Example Scope: Nurse Shift Summaries

- Problem: Nurses spend hours on shift documentation.
- LLM approach: Summarize vital signs, events, and medication changes.
- Data from daily nurse logs and patient EHR extracts.

Prompt Examples to Consider



Potential Ethical/Privacy Flags

- Real-time detection of personal details that shouldn't be shared.
- Avoid over-summarization that loses important context.
- Watch for implicit biases in how patient conditions are described.

Group 1 Sample Slide (Hypothetical)

- Domain: Oncology research data summarization.
- Data Source: 5 years of tumor registry data.
- Key Prompt: 'Focus on tumor size changes and treatments.'

Group 2 Sample Slide (Hypothetical)

- Domain: Rural telemedicine triage.
- Approach: LLM-based triage chatbot.
- Key Challenge: Low bandwidth and translation issues.

Feedback & Iteration Steps

- Start with a small pilot (10–20% of users).
- Gather metrics: time saved, error rates, satisfaction.
- Iterate your prompt and model design based on feedback.

Segment 7 Overview: Future Trends & Risk Management

- Emerging LLM updates (e.g., GPT-4.5, 5).
- Advanced regulatory and security concerns.
- Strategies for sustainable AI adoption.

Emerging LLM Capabilities

- Larger context windows (100k+ tokens (2 million if you're gemini)).
- Improved reasoning via chain-of-thought.
- Integration with multi-modal data (text + images).

Multi-Modal Healthcare Applications

- LLM + image analysis (radiology).
- LLM + sensor data for real-time monitoring.
- Creates more holistic patient summaries.

Regulatory Shifts & AI Legislation

- Potential FDA-style approvals for AI tools.
- Evolving privacy laws (e.g., GDPR).
- Possibility of licensing requirements for advanced LLMs.

Risk of Model Staleness

- Rapidly changing medical knowledge.
- Older models may become outdated.
- Scheduled re-training or knowledge injection as solutions.

AI Explainability & Trust

- Clinicians want to see the reasoning behind outputs.
- Transformers are often 'black boxes'.
- Methods to provide partial chain-of-thought explanations.

Handling Edge Cases & Rare Diseases

- LLMs may be less knowledgeable on low-prevalence conditions.
- Encourage referencing official guidelines.
- Ensure human experts verify rare-case outputs.

Cybersecurity Threats with Evolving LLM

- Risk of model infiltration to extract sensitive data.
- Prompt injection attacks to bypass restrictions.
- Ongoing patching and vulnerability scanning are required.

Culture Change & Adoption Barriers

- Staff may fear job displacement.
- Resistance to new technology in clinical workflows.
- Ongoing education and sharing success stories are key.

Sustainable AI Strategies

- Start small and demonstrate clear ROI.
- Plan for scalability with multi-year budgets.
- Include resources for continuous training and re-validation.

Example of Long-Term LLM Roadmap

- Phase 1: Documentation pilot.
- Phase 2: Decision support and partial automation.
- Phase 3: Fully integrated AI across departments.

Risk Management Checklist

- Include patient consent and disclaimers in tools.
- Establish clear escalation paths for errors.
- Ensure appropriate insurance or liability coverage.

Discussion Prompt: Long-Term AI Vision

- Will LLMs become standard members of healthcare teams?
- How can we maintain the human touch in care?
- Future synergies: integration with wearables, genomics, etc.

Building Expertise in Your Team

- Cross-train clinicians on basic AI concepts.
- Hire data scientists with healthcare domain expertise.
- Establish partnerships with academic institutions.

Monitoring & Post-Deployment Studies

- Collect usage metrics: frequency and override rates.
- Measure patient safety outcomes and cost savings.
- Share results with the broader community.

Interoperability & Standards

- Use HL7 FHIR, SNOMED, and LOINC for data exchange.
- Standardize LLM outputs for easy parsing.
- Minimize vendor lock-in with open standards.

Final Thoughts on Future LLM Directions

- Continual refinement of language understanding.
- Integration of vision-language and speech-language models.
- Regulatory frameworks will shape future usage.

Group Exercise: Comprehensive LLM Plan

- Combine data prep, prompt design, and domain adaptation.
- Address ethical, regulatory, and future expansion issues.
- Build on earlier group activity insights.

Detailed Instructions

- Step 1: Summarize your domain problem and scope (5 minutes).
- Step 2: Outline how the LLM fits—from data ingestion to UI (10 minutes).
- Step 3: Identify risk management measures and next steps (5 minutes).

Example Final Plan Outline

- Rationale and objectives.
- Technical architecture (fine-tuning? API integration?).
- Implementation timeline.
- Evaluation metrics for success.

Emphasize Realism

- Consider budget constraints.
- Factor in staff training and acceptance.
- Availability of computing resources (HPC or cloud).

15-Minute Group Working Time

- Prepare a concise 2–3 minute final presentation.
- Focus on key points of your plan.

Live Support & Questions

- Instructor(s) available for real-time clarifications.
- Option to test prompts on a shared interface.

Example Final Slide from Group 3 (Hypothetical)

- Domain: Pediatric telehealth triage.
- LLM tasks: Summarize parent statements and flag red signals.
- Pilot in 2 clinics with error rate measurement.

General Q&A

- Open floor for broader questions on LLMs, policy, and advanced topics.

Post-Tutorial Resources

- Key references: Brown et al., Devlin et al., Harper's LLMs for Healthcare (Publishing with Taylor and Francis July 2025).
- Online courses and code repositories (Hugging Face, PyTorch).
- Further reading on ethics and AI governance.
- Jeremy posting on youtube @jeremylearns, Linkedin LLM Newsletter

Suggestions for Further Study

- Join AMIA working groups on AI.
- Participate in local hackathons or code nights.
- Subscribe to AI/ML journals with a healthcare focus.
- Dedicated youtube channels

Stay Connected

- LinkedIn: Jeremy Harper, Owl Health Works LLC.
- Engage in knowledge-sharing beyond this tutorial.
- Look out for future conferences and publications.

Final Words on LLM Potential

- We are on the cusp of a major transformation in healthcare.
- LLMs empower clinicians and patients alike.
- Remain vigilant to maintain safety and accuracy.

