

NIMLAS Workshop 2026

Practical Large Language Models (LLMs) for Social Science Research: A Hands-On Introduction for Working with Text, Survey, and Unstructured Data

Dr. Jinseok Kim
School of Information
University of Michigan

Topic 1. Welcome and Workshop Overview

Welcome!

- Today's workshop is designed for researchers with little or no programming experience.
- Throughout the workshop, you will learn by **using** LLMs rather than simply hearing about them.
- By the end of today, you will have:
 - Used open-source LLMs for common social science research tasks.
 - Learned when LLMs are appropriate—and when they are not.
 - Practiced writing effective prompts.
 - Learned how to evaluate LLM outputs critically.
 - Learned which – LLM vs Traditional ML – to choose for a research project.
- Our goal is not to become AI experts today. Our goal is to become informed and effective users of LLMs for research.

Workshop Roadmap

Table 1: Workshop Roadmap

Part 1. Experience LLMs First	Watch a live demonstration. Complete hands-on exercises using Hugging Face Chat. Explore common research applications.
Part 2. Understand How LLMs Work	What an LLM is (and is not). Capabilities and limitations. Prompt engineering. Evaluating LLM outputs.
Part 3. Apply LLMs to Research	Research applications for social science. Open-source models and local deployment. Choosing between LLMs and traditional machine learning.

Workshop Schedule (1)

Table 2: Workshop Schedule (1)

Time	Section	Purpose
0:00–0:10	1. Welcome and Workshop Overview	• Goals, agenda, software, expectations
0:10–0:55	2. Hands-on Exercise 1: What Can an LLM Do?	• Use Hugging Face Chat to analyze interview text, summarize documents, classify responses, extract information, answer questions, etc. • This gives participants a preview of the entire workshop. • Participants begin using Hugging Face Chat themselves with guided prompts.

Workshop Schedule (2)

Table 3: Workshop Schedule (2)

Time	Section	Purpose
0:55–1:20	3. What Is an LLM?	• Explain what participants just experienced. • Cover tokens, prediction, context windows, training at an intuitive level.
1:35–2:00	4. Capabilities and Limitations	• Strengths, hallucinations, bias, reasoning limitations, reproducibility. • Use examples from the demo.
2:00–2:25	5. Prompt Engineering	• How prompts influence outputs; prompt templates; iterative refinement.
2:25–2:50	6. Hands-on Exercise 2: Prompt Engineering	

Workshop Schedule (3)

Table 4: Workshop Schedule (3)

Time	Section	Purpose
3:05–3:25	7. Evaluating LLM Outputs	Human evaluation, verification, consistency checks, comparison across prompts/models.
3:25–3:45	8. Open-Source Models and Local Deployment	Why local models, privacy, hardware, installing open-source models, selecting appropriate models.
3:45–4:00	9. Traditional ML vs. LLMs, Resources, and Q&A	When to use each approach, future learning resources, final questions.

If the Workshop Is Interrupted

- Occasionally, technical issues such as an Internet outage or computer problem may occur.
- If that happens:
 - Please remain in the room.
 - I will reconnect using another computer within **5 minutes**, if possible.
 - If the issue is related to the Internet connection, I may need **up to 10 minutes** to reconnect using an alternative network.
- **While You Wait**, Please use this time to:
 - Review your notes.
 - Discuss today's activities with the people around you.
 - Try refining one of the prompts from today's exercises.
 - Write down any questions you would like to ask during the next discussion.
- Thank you for your patience. We will resume the workshop as soon as the technical issue has been resolved.

Topic 2: Hands-on Exercise 1: What Can an LLM Do?

1. Summarization
2. Information Extraction
3. Classification
4. Theme Generation
5. Data Transformation
6. Missing Data Imputation
7. Hallucination

Demo 1: Summarizing an Interview

- **Task:** Can an LLM quickly summarize this interview excerpt?
- **Scenario**
 - A social science researcher interviewed several participants about their experience working remotely after the pandemic.
 - The researcher wants a quick summary before beginning detailed qualitative coding.
- **Sample Interview Excerpt**
 - "Working from home has given me much more flexibility. I spend more time with my family and don't have to commute anymore, which saves me nearly two hours every day. However, I sometimes feel isolated because I don't interact with coworkers as often. It can also be difficult to separate work from personal life since my office is at home."

Demo 1: Sample Prompt

You are assisting a social science researcher.

Summarize the following interview response in 2–3 sentences.

Focus only on the participant's main experiences and avoid adding any new information.

Interview: “Working from home has given me much more flexibility. I spend more time with my family and don't have to commute anymore, which saves me nearly two hours every day. However, I sometimes feel isolated because I don't interact with coworkers as often. It can also be difficult to separate work from personal life since my office is at home.”

Demo 1: LLM Output

- The participant reported that working from home provides greater flexibility, allows more time with family, and eliminates a lengthy daily commute. At the same time, the participant experiences social isolation due to reduced interaction with coworkers and finds it challenging to maintain boundaries between work and personal life.
- **Items to Check**
 - Did the model capture the main ideas?
 - Was anything important omitted?
 - Did the model add information that was not present?

Demo 1: Interpretation & Takeaways

- **What the model did well**
 - Identified the central themes
 - Condensed a long response
 - Preserved the original meaning
 - Used neutral language
- **What researchers should still do**
 - Read the original transcript.
 - Verify that no important nuance was lost.
 - Check that no unsupported statements were introduced.
 - Remember that summaries are **starting points**, not research findings.
- **Key Takeaway**
 - LLMs can reduce the time needed to review qualitative data, but researchers (not the model) remain responsible for interpreting the evidence.

Demo 1: Researcher Notes

1. This example introduces several important ideas without requiring technical explanations
 - LLMs consider context.
 - They compress information instead of simply extracting sentences.
 - They may omit details depending on the prompt.
 - Researchers must verify summaries against the original text.
 - Good prompts include clear instructions (e.g., ‘avoid adding new information’).
2. LLMs are research assistants that accelerate analysis, not replacements for them.

Demo 2: Extracting Information from an Interview

- **Task:** Can an LLM identify important information and organize it into a structured format?
- **Scenario**
 - A researcher is conducting interviews with first-generation college students about their university experiences.
 - Instead of reading hundreds of interviews to manually identify key information, the researcher wants the LLM to extract specific attributes from each interview.
- **Sample Interview Excerpt**
 - "I'm the first person in my family to attend college. During my first year, I struggled with time management and often felt overwhelmed by the workload. Joining a study group in my second semester helped me stay organized and improved my confidence. Now I'm considering applying to graduate school."

Demo 2: Sample Prompt

You are assisting a social science researcher. Read the interview excerpt below and extract the following information. Return your answer as a table with two columns: Attribute and Value.

Attributes:

- First-generation college student?
- Main challenge
- Helpful support
- Future educational goal

Interview: I'm the first person in my family to attend college. During my first year, I struggled with time management and often felt overwhelmed by the workload. Joining a study group in my second semester helped me stay organized and improved my confidence. Now I'm considering applying to graduate school.

If information is not explicitly stated, write "Not mentioned."

Demo 2: LLM Output

Table 5: Demo 2 LLM Output Review

Attribute	Value
First-generation college student	Yes
Main challenge	Time management and feeling overwhelmed by coursework
Helpful support	Joining a study group
Future educational goal	Considering applying to graduate school

- **Items to Check**

- Did the model correctly identify the requested information?
- Which answers came directly from the interview?
- Did the model make any assumptions?

Demo 2: Interpretation & Takeaways

- **What the model did well**
 - Located relevant information quickly
 - Converted narrative text into structured data
 - Ignored irrelevant details
 - Produced output that could easily be copied into a spreadsheet
- **What researchers should still do**
 - Verify every extracted field against the original interview.
 - Be careful when an answer is implied rather than explicitly stated.
 - Use consistent prompts when processing many interviews.
 - Consider reviewing a sample of outputs manually for quality control.
- **Key Takeaway:**
 - LLMs can transform unstructured qualitative data into structured variables, making large collections of interviews easier to organize, search, and analyze.
 - Human verification remains essential to ensure accuracy and consistency.

Demo 2: Researcher Notes

1. Structured outputs are often more useful than free-form text because they can be exported to spreadsheets, statistical software, or databases.
2. Prompts should specify exactly what fields to extract and what to do when information is missing (e.g., 'Not mentioned'). This reduces ambiguity and improves consistency.
3. The model is performing information extraction, not inference.
 - For example, it correctly reports that the participant is "considering applying to graduate school" rather than assuming they will attend graduate school.

Demo 3: Classifying Open-Ended Survey Responses

- **Task:** Can an LLM automatically assign categories to these responses using a predefined coding scheme?
- **Scenario**
 - A university administered a survey asking students: “What is the biggest challenge you face in your academic life?”
 - The researcher has collected **hundreds of responses** and wants to group them into meaningful categories before conducting further analysis.

Demo 3: Sample Survey Responses

Table 6: Sample Survey Responses

Response ID	Student Response
1	I work nearly 30 hours every week, so I don't have enough time to study.
2	I'm constantly worried about paying tuition and rent.
3	Many of my classes move so quickly that I struggle to keep up.
4	I often feel stressed and anxious during the semester.
5	Finding a quiet place to study is difficult because I live with three roommates.

Demo 3: Sample Prompt

You are assisting a social science researcher.

Classify each survey response into one of the following categories.

Categories: Time Management, Financial Concerns, Academic Difficulty, Mental Health, Study Environment, Other

Return your answer as a table with three columns: Response ID | Category | Brief Justification

Survey Responses:

1. I work nearly 30 hours every week, so I don't have enough time to study.
2. I'm constantly worried about paying tuition and rent.
3. Many of my classes move so quickly that I struggle to keep up.
4. I often feel stressed and anxious during the semester.
5. Finding a quiet place to study is difficult because I live with three roommates.

Demo 3: LLM Output

Table 7: Demo 3 LLM Output Review

Response ID	Category	Brief Justification
1	Time Management	Work schedule leaves limited study time.
2	Financial Concerns	Mentions tuition and living expenses.
3	Academic Difficulty	Difficulty keeping up with coursework.
4	Mental Health	Reports stress and anxiety.
5	Study Environment	Lack of a quiet study space.

- **Items to Check**

- Would you agree with every classification?
- Which responses were easy to classify?
- Which responses might reasonably fit more than one category?

Demo 3: Interpretation & Takeaways

- **What the model did well**

- Applied a predefined coding scheme consistently
- Produced standardized labels across multiple responses
- Worked much faster than manual coding
- Generated explanations that make the classifications easier to review

- **What researchers should still do**

- Develop the coding framework **before** asking the model to classify responses.
- Review a sample of classifications for accuracy.
- Some responses may fit multiple categories, even if the prompt requests only one.
- Refine category definitions or prompts if classifications are inconsistent.

- **Key Takeaway**

- LLMs can assist with coding qualitative responses, but they apply the coding scheme that researchers provide.
- Careful category design and human validation are essential for producing reliable research results.

Demo 3: Researcher Notes

1. Classification is one of the most common applications of LLMs in social science research.
 - e.g., coding interview excerpts, survey responses, social media posts, news articles, and policy documents
2. Unlike Demo 2 (the model extracted information), it is making a judgment by assigning a category.
3. Relating to inter-rater reliability.
 - If two human coders might disagree on Response 1, should we expect the LLM to always choose the correct category?
 - This indicates that disagreement may reflect ambiguity in the coding scheme rather than an error by the model (or human coder).
4. Small changes to the prompt (e.g., allowing multiple categories, defining categories more precisely, or providing example classifications) can significantly improve consistency and accuracy.

Demo 4: Identifying Themes Across Survey Responses

- Task: Can an LLM identify common themes across multiple responses without being given predefined categories?
- **Scenario**
 - A researcher asked students: **What has been the most positive aspect of your university experience?**
 - After collecting hundreds of responses, the researcher wants to identify common themes before developing a coding framework.

Demo 4: Sample Survey Responses

1. I've met friends who motivate me and make campus feel like home.
2. Learning from professors who are experts in their fields has been incredibly rewarding.
3. Being involved in student organizations helped me develop leadership skills.
4. The flexibility to explore different majors allowed me to discover what I really enjoy.
5. Working on research projects gave me confidence and practical experience.

Demo 4: Sample Prompt

You are assisting a social science researcher.

Read the following survey responses.

Identify the major themes that appear across the responses.

For each theme,

- provide a short descriptive name,
- briefly explain the theme,
- list the response numbers that support it.

Survey Responses:

1. I've met friends who motivate me and make campus feel like home.
2. Learning from professors who are experts in their fields has been incredibly rewarding.
3. Being involved in student organizations helped me develop leadership skills.
4. The flexibility to explore different majors allowed me to discover what I really enjoy.
5. Working on research projects gave me confidence and practical experience.

Demo 4: LLM Output

Table 8: Demo 4 LLM Output Review

Theme	Description	Supporting Responses
Social Connections	Building friendships and a sense of belonging	1
Academic Mentorship	Positive learning experiences with faculty	2
Personal Growth	Developing leadership, confidence, and practical skills	3, 5
Academic Exploration	Discovering interests through flexible coursework	4

- **Items to Check**

- Would you have created the same themes?
- Should Responses 3 and 5 belong to the same theme?
- Are there alternative ways to organize these responses?

Demo 4: Interpretation & Takeaways

- **What the model did well**
 - Identified recurring ideas across multiple responses
 - Combined related responses into broader concepts
 - Produced a useful starting point for thematic analysis
 - Organized the data in an interpretable format
- **What researchers should still do**
 - Treat suggested themes as **provisional**, not final.
 - Compare multiple prompts or multiple models to assess consistency.
 - Review whether themes accurately represent participants' perspectives.
 - Refine, merge, split, or rename themes based on the research objectives and theoretical framework.
- **Key Takeaway**
 - **LLMs can generate candidate themes before researchers begin the coding process.**
 - **Theme development requires human judgment and domain expertise.**

Demo 4: Researcher Notes

1. LLM models can synthesize patterns across many observations, which is central to many qualitative research methods.
2. If we rerun the same prompt tomorrow or use a different open-source model, do you think we'll get the same themes?
 - The themes generated by an LLM are not “discovered truths.”
 - Different researchers or different prompts may produce different but equally reasonable thematic structures.
3. LLMs are collaborators that accelerate qualitative analysis rather than automated replacements for the researcher's interpretive role.

Demo 5: Converting Unstructured Text into a Structured Table

- **Task:** Can an LLM transform narrative text into structured data that can be analyzed in a spreadsheet?
- **Scenario**
 - A researcher is reviewing news articles about local community events.
 - Rather than manually recording key details from each article, the researcher wants the LLM to extract the information into a structured table for later analysis.
- **Sample News Excerpt:** On Saturday, April 12, approximately 250 residents gathered at Riverside Park for a community clean-up event organized by the Green Future Association. Volunteers collected litter, planted trees, and removed invasive plant species. According to the organizers, more than 400 pounds of trash were collected during the four-hour event.

Demo 5: Sample Prompt

You are assisting a social science researcher.

Extract the following information from the news article.

Return the results as a table with two columns: Field and Value.

Fields:

- Event Name
- Event Date
- Location
- Organizer
- Number of Participants
- Activities
- Amount of Trash Collected
- Event Duration

If a field is not explicitly mentioned, write "Not mentioned."

Article: On Saturday, April 12, approximately 250 residents gathered at Riverside Park for a community clean-up event organized by the Green Future Association. Volunteers collected litter, planted trees, and removed invasive plant species. According to the organizers, more than 400 pounds of trash were collected during the four-hour event.

Demo 5: LLM Output

Table 9: Demo 5 LLM Output Review

Field	Value
Event Name	Community clean-up event
Event Date	Saturday, April 12
Location	Riverside Park
Organizer	Green Future Association
Number of Participants	Approximately 250
Activities	Collected litter, planted trees, removed invasive plant species
Amount of Trash Collected	More than 400 pounds
Event Duration	Four hours

- **Items to Check**

- Did the model only extract information that appears in the article?
- Which fields required interpretation rather than direct copying?
- Which fields might need to be standardized before statistical analysis?

Demo 5: Interpretation & Takeaways

- **What the model did well**

- Converted narrative text into a structured dataset
- Located multiple types of information in a single pass
- Produced output that is ready to export to Excel or CSV
- Reduced repetitive manual data entry

- **What researchers should still do**

- Verify that every extracted value appears in the original text.
- Standardize values for consistency (e.g., approximately 250).
- Decide how to handle missing or ambiguous information before processing large documents.
- Remember that structured outputs should still undergo quality assurance before analysis.

- **Key Takeaway**

- LLMs can efficiently convert unstructured documents into structured datasets, enabling researchers to spend less time on manual data preparation and more time on analysis.
- Human review remains important to ensure accuracy and consistency across records.

Demo 5: Researcher Notes

1. This demonstration addresses one of the most time-consuming aspects of research: data preparation.
2. Distinction between information extraction (Demo 2) and data transformation (Demo 5)
 - **Demo 2:** The researcher extracts a few variables from a single interview to understand one participant.
 - **Demo 5:** The researcher designs a structured schema that can be applied across large documents, creating a dataset for quantitative analysis.
3. LLMs do not create a relational database!
 - They simply produce structured text.
 - Researchers remain responsible for designing the data schema, choosing variable names, standardizing formats (e.g., dates, numbers, categorical values, etc.), and validating the extracted data.

Demo 6: Imputing Missing Information

- **Task:** Can an LLM infer the missing value from the available information?
- **Scenario**
 - A researcher collected survey responses but some demographic information is missing because participants skipped optional questions.
 - The researcher wonders whether an LLM can infer missing values from the available information.

Sample Survey Data

Table 10: Sample Survey Data

ID	Major	Year	Campus Housing	Club Participation	GPA	First-Generation Student
101	Biology	Sophomore	Yes	Pre-Med Club	3.7	Yes
102	Computer Science	Senior	No	Robotics Club	3.8	No
103	Psychology	Junior	Yes	Mental Health Club	3.5	?

- Additional information for Student 103:
 - “Neither of my parents attended college, so I had to figure out many things on my own when applying to universities.”

Demo 6: Sample Prompt

You are assisting a social science researcher.

Below is a partially completed survey record.

Determine whether there is enough evidence to infer whether the participant is a first-generation college student.

Return:

1. Your conclusion.
2. Your confidence (High, Medium, or Low).
3. The evidence supporting your conclusion.
4. If the evidence is insufficient, do not guess. Instead, answer “Cannot determine.”

Participant Information:

Major: Psychology

Year: Junior

Campus Housing: Yes

Club: Mental Health Club

GPA: 3.5

Interview Excerpt: “Neither of my parents attended college, so I had to figure out many things on my own when applying to universities.”

Demo 6: LLM Output

Conclusion

Yes

Confidence

High

Evidence

The participant states that neither parent attended college, which directly matches the common definition of a first-generation college student.

Items to Check

- Was the conclusion based on explicit evidence?
- Would you feel comfortable using this value in your dataset?
- Should this value be stored differently from an observed survey response?

Demo 6: Interpretation & Takeaways

- **What the model did well**
 - Used explicit evidence rather than guessing
 - Explained its reasoning
 - Reported a **confidence level (not a statistical term)**
 - Made the inference transparent
- **What researchers should still do**
 - Record that the value was **imputed**, not observed.
 - Keep the original value as missing.
 - Review imputed values before analysis.
 - Consider whether imputation is appropriate for the research question.
- **Key Takeaway**
 - LLMs can assist with imputing missing information when sufficient evidence exists.
 - Imputed values should never replace original observations without documentation and validation.

Demo 6: Researcher Notes

1. This example is evidence-based imputation, not prediction.
 - The participant explicitly states: “Neither of my parents attended college.”
 - The model is simply recognizing that this statement satisfies the definition of a first-generation college student.
 - Unlike statistical imputation (estimating missing values based on patterns across many observations), LLM-based imputation uses contextual information available for an individual case.
2. Now remove the interview excerpt and ask the same question again.
 - Many open-source models will respond with “Cannot determine.”
 - Some weaker models may instead guess: e.g., “Possibly yes.”
3. See the next demo on hallucination

Demo 7: When the LLM Gets It Wrong (Hallucination)

- Task: Can an LLM recognize when information is missing?
- **Scenario**
 - A researcher is reading an interview transcript and asks the LLM to identify the participant's occupation.
 - However, the participant never mentions their occupation.
 - Will the model admit that the information is unavailable, or will it invent an answer?
- **Sample Interview Excerpt:** “I’ve lived in this neighborhood for about fifteen years. I enjoy volunteering at the local food bank on weekends, and I spend most evenings helping my children with their homework. During the pandemic, I started gardening because it helped me relax.”

Demo 7: Sample Prompt 1

You are assisting a social science researcher.

Read the interview below and answer the following question.

What is the participant's occupation?

Interview: I've lived in this neighborhood for about fifteen years. I enjoy volunteering at the local food bank on weekends, and I spend most evenings helping my children with their homework. During the pandemic, I started gardening because it helped me relax.

Demo 7: Sample Prompt 2

You are assisting a social science researcher.

Read the interview below and answer the following question.

What is the participant's occupation?

If the occupation is not explicitly mentioned, answer: "Occupation not mentioned."

Interview: I've lived in this neighborhood for about fifteen years. I enjoy volunteering at the local food bank on weekends, and I spend most evenings helping my children with their homework. During the pandemic, I started gardening because it helped me relax.

Demo 7: LLM Output

- **Better Response**

- Occupation not mentioned.

- **Possible Hallucinated Response**

- The participant appears to be a teacher or community worker because they enjoy helping others and spending time with children.

- **Items to Check**

- Which response should a researcher trust?
- Where did the second answer come from?
- Does the interview actually support that conclusion?

Demo 7: Interpretation & Takeaways

- **What happened?**

- The second answer sounds reasonable.
- Unfortunately, none of it appears in the interview.
- The model guessed an occupation based on patterns learned during training.

- **What researchers should do**

- Ask models to distinguish between observations and inferences.
- Tell the model to answer "Not mentioned" when appropriate.
- Always compare important outputs with the original data.
- Never treat a plausible answer as evidence.

- **Key Takeaway**

- LLMs are designed to generate plausible language, not to guarantee factual accuracy. A confident answer is not necessarily a correct answer.

Demo 7: Researcher Notes

1. Hallucination is not “lying.”

- The model does not know it is making something up.
- It is predicting text that appears statistically plausible.

2. Prompt wording matters.

- Prompt A: What is the participant's occupation?
- Prompt B: If the occupation is not explicitly mentioned, answer “Occupation not mentioned.”

3. Verification is essential in using LLMs.

- If this interview were part of your dissertation, would you include 'teacher' as a finding? Absolutely not.
- Researchers, not LLMs, are responsible for the validity of research findings.

What Have We Learned?

- **LLMs can help researchers:**
 - Summarize qualitative data
 - Extract structured information
 - Classify open-ended responses
 - Identify themes across documents
 - Convert narrative text into analyzable datasets
 - Accelerate repetitive research tasks
- **But...**
 - They can also make confident mistakes.
 - Their outputs depend heavily on the prompts we provide.
 - Researchers remain responsible for validating results.
- **Next Question:** How do LLMs work, and why do prompts make such a big difference?

Topic 3: What Is an LLM?

We Just Used an LLM...

- **But What Actually Happened?**
- **Questions:**
 - Did the LLM search the Internet?
 - Did it understand the interview like a human?
 - Did it memorize millions of documents?
 - How did it generate those answers so quickly?
- By the end of this section, you'll understand:
 - what an LLM does
 - why it sometimes makes mistakes
 - why prompts matter.

Predicting the Next Word

- What word comes to mind?
- Suppose I write: **Peanut butter and** _____
- Now try this one: **Roses are red, Violets are** _____
- **Key points**
 - You weren't searching Google.
 - You weren't recalling one exact sentence.
 - You predicted what was **most likely** to come next.
 - An LLM works in a similar way: it predicts what text is most likely to come next.

LLMs Predict One Piece at a Time

- An LLM doesn't write an entire paragraph at once.
 - Instead, it repeatedly predicts the next small piece of text.
- Imagine it producing:
 - The → participant → reported → greater → flexibility → while → working → from
→ home
- Each prediction becomes part of the next prediction.
 - Thousands of these tiny predictions create a complete response.

What Is a Token?

- LLMs Don't Read Entire Sentences: They read text in small pieces called tokens.
 - For example: **Working/from/home/has/given/me/more/flexibility/.**
 - Each piece becomes part of what the model processes.
- Sometimes a token is:
 - a whole word
 - part of a long word
 - punctuation
 - even a space
- **Why Should We Care? Because:**
 - longer prompts use more tokens,
 - larger documents require more tokens,
 - models have limits on how many tokens they can consider.

Context Window: How Much Can an LLM Remember?

- Imagine talking with someone who can remember only the last few pages of your conversation.
- Everything inside that memory is called the **Context Window**
 - It includes: your instructions, previous questions, uploaded documents, the model's earlier answers, etc.
- If the conversation becomes too long, older information may be forgotten.
- **Example**
 - Suppose you upload a 300-page report.
 - Some smaller models may not be able to consider all 300 pages at once.
- **Takeaway:** An LLM only works with the information inside its current context window.

Where Does the Knowledge Come From?

- Did the Model Memorize Everything? Not exactly.
- Before anyone uses it, the model has already read an enormous amount of text.
 - For example:
 - books
 - articles
 - websites
 - publicly available documents
 - computer code
- During training,
 - it learns language patterns,
 - not exact answers to every question.

Why Did the Demo Work?

- Think about Demo 3.
- We asked the model to classify survey responses.
 - Did it “know” that “I'm worried about paying tuition.”
 - belongs to Financial Concerns?
 - → No.
- During training, it...
 - learned many examples of similar language.
 - recognized a familiar pattern.
- **Takeaway:** LLMs are pattern recognizers

Why Can LLMs Make Mistakes?

- Suppose someone says: “I volunteer every weekend. I help children with homework.”
- Can you conclude that the person is a teacher?
 - But an LLM might. Why?
 - Because it predicts what sounds likely, not what has been proven.
 - This is exactly what we saw in Demo 7 (Hallucination).
- **Takeaway:** LLMs generate plausible language, not guaranteed facts.

Summary

- An LLM is a computer program that
 - has learned language patterns from enormous amounts of text,
 - predicts one token at a time,
 - uses information inside a limited context window,
 - recognizes patterns across text,
 - generates new responses instead of retrieving stored answers.
- An LLM is an extremely powerful pattern-completion engine rather than a search engine or a human expert.
- **Next practical question:** When should we trust those answers and when should we be cautious?

Topic 4: Capabilities and Limitations of LLMs

Should We Trust Everything an LLM Says?

- **In our demos, the LLM...**
 - summarized interviews.
 - classified survey responses.
 - identified themes.
 - extracted information.
 - created tables.
 - Then...it confidently invented an occupation that wasn't mentioned.
- **Question:** How can the same system be both amazingly useful and surprisingly wrong?

What LLMs Do Exceptionally Well

- **They can**
 - summarize long documents
 - classify text
 - extract information
 - organize messy data
 - identify themes
 - explain statistical results
 - draft reports
- **Why?**
 - Because these tasks depend on recognizing language patterns.
 - That's exactly what LLMs were trained to do.
- **Think of an LLM as a research assistant** who reads incredibly fast and writes incredibly fast.

LLMs Save Time

- Which task would you prefer?

Table 11: Comparing Workload Example

Option A	Option B
Read 500 interview responses one by one before starting analysis.	Ask the LLM to summarize, extract themes, and organize responses within minutes.

- The LLM doesn't replace the researcher. It reduces repetitive work.
- **Analogy:** A calculator didn't replace mathematicians.
 - It eliminated tedious arithmetic.
 - LLMs do something similar for language.
- **Takeaway:** LLMs help researchers spend less time preparing data and more time analyzing it.

LLMs Are Not Search Engines

- Suppose we ask: “Who won the Nobel Prize in Economics in 1978?”
 - If the model knows, it may answer correctly.
 - But how? Not because it searched Wikipedia or any info resource.
 - Instead, it generated the answer from patterns learned during training.
- **Important Difference**
 - A search engine finds documents.
 - An LLM generates text.
- **Takeaway:** An LLM generates answers. But it does not verify them.

Hallucinations

- **Remember this example?**

- Interview: "I volunteer every weekend."
- Question: What is this person's occupation?
- One possible answer: Teacher.
- Sounds reasonable. But, the interview never said that.

- **Why did this happen?**

- The model predicted what sounded plausible, not what was supported by evidence.
- **Researchers** using LLMs should ask “What evidence supports this answer?”

- **Takeaway:** Plausible is not the same as true.

Bias

- **Imagine these two interview excerpts.**
 - Participant A: **I recently became a father.**
 - Participant B: **I recently became a parent.**
 - Would the LLM respond the same way? Not necessarily.
 - Why? Because it learned from human-written text, and human language contains biases.
- **Bias can appear in**
 - wording, stereotypes, assumptions, unequal representation, etc.
- **Researchers** should check outputs carefully, especially when studying people.
- **Takeaway:** LLMs can inherit biases from the data used during training.

Reasoning Limitations

- LLMs often **appear** to reason. Sometimes they fail on simple logic.
- **Example:**
 - A farmer has 17 sheep. All but 9 die. How many sheep remain?
 - Correct answer: 9
 - Some models answer: 8
- **Why?** Because predicting language is not the same as logical reasoning.
- **Researchers should** be cautious when using LLMs for complex reasoning tasks.
- **Takeaway:** Fluent language does not always indicate correct reasoning.

Reproducibility

- **Question:** Suppose we ask LLMs the same question twice. Will we always receive the same answer?
- **Why?**
 - Many LLMs include randomness when generating responses.
 - Small prompt changes can also change the answer.
- **Researchers Should** save prompts, model name, model version, settings
 - Just as we save datasets and code.
- **Takeaway:** Good research requires reproducible prompts as well as reproducible analyses.

Human in the Loop

- **Think about every demo today.**
 - Who decided which interview to analyze?
 - Who wrote the prompt?
 - Who checked the answer?
 - Who interpreted the results?
- The LLM never replaced the researcher. It accelerated the researcher's work.
- **Analogy:** Autopilot assists a pilot. It doesn't eliminate the pilot.
- **Takeaway:** Researchers remain responsible for every conclusion reported in a study.

Summary for Topic 4

- **When Should We Use LLMs?**

Table 12: Comparing LLM Use Scenarios

Excellent For	Be Careful With
✓ summarizing ✓ extracting information ✓ coding text ✓ identifying themes ✓ organizing data ✓ drafting reports	⚠ factual claims ⚠ missing evidence ⚠ bias ⚠ complex reasoning ⚠ reproducibility

- **Key takeaway:** Use LLMs to accelerate your thinking, not to replace your thinking.
- **Next practical question:** Can we make LLMs produce better answers?

Topic 5: Prompt Engineering

Why Does One Prompt Work Better Than Another?

- Same interview. Same model. Different results.

Table 13: Comparing Prompts

Prompt A	Prompt B
Summarize this interview	Summarize in three sentences. Separate positive and negative experiences. Do not add information.

- Why?
- **This Session's Goal:** Learn how to write prompts that produce better answers, more consistent answers, more useful research outputs.

Think of Prompting as Giving Instructions

- Imagine asking two people to bake a cake.
- **Person A** hears
 - Bake a cake.
- **Person B** hears
 - Bake a chocolate birthday cake for eight people.
 - Use strawberries for decoration.
 - Finish before 5 PM.
- Who is more likely to produce what you wanted?
- The same idea applies to LLMs: Prompt engineering is simply giving clear instructions.

Good Prompts Answer Four Questions

- **Before writing a prompt**, ask yourself:

Table 14: Good Prompt Questions

What task?	What data?	What output?	Any rules?
Summarize? Classify? Extract? Compare?	Interview? Survey? News article? Research paper?	Paragraph? Table? JSON? Bullet list?	Maximum length? No assumptions? Use evidence only?

- **Takeaway:** Every good prompt clearly defines the task, the data, the output, and the rules.

Give the Model a Role

Table 15: Comparison of Action Item

Instead of	Try
Summarize this interview.	You are assisting a social science researcher. or You are an experienced qualitative researcher.

- **Why?:** The role gives the model context about vocabulary, writing style, priorities.
- Think back to our demos: Almost every prompt began with something like “You are assisting a social science researcher.”
- **Takeaway:** Giving the model a role often improves consistency and focus.

Tell the Model Exactly What You Want

Table 16: Comparing Prompts

Poor prompt	Better prompt
Analyze this interview.	Identify three major themes. Quote supporting evidence. Return a table.

- The more specific the instructions, the less guessing the model has to do.
- **Remember Demo 5:** Instead of asking Organize this article, we specified exactly which fields to extract.
- **Takeaway:** Specific prompts reduce ambiguity.

Specify the Output Format

- Imagine collecting 1,000 interview summaries. Would you rather receive
 - 1,000 different writing styles?
 - 1,000 tables with the same columns?
- **Researchers love** structured outputs because they are easy to review and analyze.
- **Examples:** Markdown table, Bullet list, Numbered list, CSV, JSON
- **Takeaway:** Tell the model what the answer should look like, not just what it should say.

Tell the Model What NOT To Do

- Sometimes the most important instruction is what to avoid.
- **Examples**
 - Do not infer missing information.
 - If the answer is unknown, write “Not mentioned.”
 - Use only information in the interview.
 - Do not use outside knowledge.
- Remember Demo 6
 - Prompt 4: **If the evidence is insufficient, do not guess. Intead, answer “Cannot determine.”**
 - That single instruction greatly reduced hallucination.
- **Takeaway:** Constraints are often just as important as instructions.

Prompt Engineering Is Iterative

- Rarely does the first prompt become the final prompt.
- Researchers usually improve prompts step by step.
- **Example**

Table 17: Prompt Development Example

Prompt 1	Summarize.
Prompt 2	Summarize in three sentences.
Prompt 3	Summarize separately positive and negative experiences.
Prompt 4	Summarize separately positive and negative experiences. Quote supporting evidence.

- **Takeaway:** Prompt engineering is an iterative process of refinement.

A Simple Prompt Template

For most research tasks, this template works well

Table 18: Prompt Template

Role	You are assisting a social science researcher.
Task	Identify the major themes.
Data	Interview transcript...
Output	Return a table.
Rules	Use only evidence from the interview. If information is missing, write "Not mentioned."

Summary for Topic 5

- **Five Habits of Good Prompt Writers**

1. Clearly define the task.
2. Give the model a role.
3. Specify the desired output.
4. Add important constraints.
5. Refine the prompt through iteration.

- **The Most Important Lesson:** Prompt engineering is not programming. It is learning to communicate clearly.

- **Takeaway:** If a prompt is vague, the model has to guess. If a prompt is clear, the model has much less guessing to do (less hallucinations)

Topic 6: Activity - Improving a Prompt

Let's Improve a Prompt Together

- Suppose we have the following interview.
 - “Working from home has given me much more flexibility. I spend more time with my family and don't have to commute anymore, which saves me nearly two hours every day. However, I sometimes feel isolated because I don't interact with coworkers as often. It can also be difficult to separate work from personal life since my office is at home.”
- **The first prompt is simply: Analyze this interview.**
- **Question: How could we improve this prompt?**
- **Activity**
 - Use the Huggingface Chat to prompt the simple one and get the result
 - Rewrite the prompt for **2 or 3 minutes**. Try the chat again with the rewritten prompt and compare the result
 - Share your comparison with participants

A Hint (The Five Questions)

- Don't worry about writing a “perfect” prompt. Try to answer these five questions.
1. **Role:** Who should the model be?
 2. **Task:** What exactly should it do?
 3. **Data:** What information should it analyze?
 4. **Output:** How should it present the answer?
 5. **Rules:** What should it avoid?

One Possible Prompt

You are assisting a social science researcher.

Read the interview below.

Identify the three main themes discussed by the participant.

For each theme,

- provide a short description,
- include one supporting quotation,
- indicate whether the theme reflects a positive or negative experience.

Return the results as a table.

Use only information explicitly stated in the interview.

If something cannot be determined, write “Not mentioned.”

Reflection

- Most groups probably wrote different prompts. That's expected.
 - Prompt engineering is not about finding one perfect prompt.
 - It is about writing prompts that are clear, reproducible, appropriate for the research task.
- **Question:** If we changed only one sentence in the prompt, would the answer change?
 - This is why researchers should save their prompts just like they save their datasets, analysis code, and research notes.
- **Key Takeaway:** A prompt is part of your research methodology. If someone cannot reproduce your prompt, they may not be able to reproduce your results.

Topic 7: Evaluating LLM Outputs

The LLM Gave an Answer...But Is It Good Enough?

- Think back to today's demos.
 - The model
 - summarized interviews
 - classified responses
 - generated themes
 - extracted information
- But... how do we know the answers are correct?
- **This Session's Goal:** Learn a simple framework for evaluating LLM outputs before using them in research.
- **Takeaway:** Generating an answer is easy. Evaluating it is the researcher's responsibility.

Four Questions Every Researcher Should Ask

- Whenever an LLM gives an answer, ask four questions.
 1. Is it **correct**?
 2. Is it **supported by evidence**?
 3. Is it **complete**?
 4. Is it **consistent**?
- These four questions apply to every research project.
- **Takeaway:** Don't ask, "Is this answer impressive?" Ask, "Is this answer trustworthy?"

Check the Evidence

- Remember our interview summary.
 - **Original interview:** Working from home has given me more flexibility...
 - **Summary:** The participant values flexibility but sometimes feels isolated...
 - **Question:** Where did every sentence (or phrases) come from?
- If you cannot point to supporting evidence in the original data, the answer should not appear in your findings.
- **Takeaway:** Every important conclusion should be traceable to the original data.

Compare Multiple Prompts

- Suppose we ask
 - **Prompt A:** Identify the themes.
 - **Prompt B:** Identify three themes and quote supporting evidence.
- Will the outputs be identical?
- Which output would you trust more?
- Better prompts often make evaluation easier.
- **Takeaway:** If changing the prompt changes the answer dramatically, investigate why.

Compare Multiple Models

- Imagine asking two different models the same question.
 - **Model A** identifies four themes.
 - **Model B** identifies five.
- Should we panic? No.
- Instead, compare agreements, disagreements, and supporting evidence
 - Agreement does not guarantee correctness.
 - Disagreement does not automatically mean someone is wrong.
- **Takeaway:** Comparing models helps identify uncertain cases that deserve closer human review.

Consistency Matters

- Suppose we classify 500 survey responses. Tomorrow, we classify them again. Will we obtain the same results?
- Researchers should check consistency, especially when processing large datasets.
- **Save**
 - prompts
 - model name
 - model version
 - settings
- **Takeaway:** Consistency is part of research quality.

Human Evaluation

- Imagine this workflow: LLM produces summaries → Researcher reviews summaries → Researcher edits summaries → Final dataset
- Who made the final decision?
- Think of the LLM as a first reviewer, not the final reviewer.
- **Analogy:** Spell-check finds many mistakes. You still proofread your paper.
- **Takeaway:** Human judgment remains the final quality control step.

A Practical Evaluation Checklist

- Before using an LLM output, ask yourself:
 - ☐ Is every important claim supported?
 - ☐ Did the model avoid unsupported assumptions?
 - ☐ Is anything missing?
 - ☐ Is the output reproducible?
 - ☐ Would I feel comfortable publishing this result?
- If the answer to the last question is “No,” review the output again.
- **Takeaway:** A simple checklist can prevent many research mistakes.

Topic 8: Open-Source Models and Local Deployment

Why Run an LLM on Your Own Computer?

- Suppose you are analyzing:
 - Confidential interview transcripts
 - Medical narratives
 - Student records
 - Internal company documents
- Would you upload these data to an online AI service?
- **This Session's Goal**
 - Understand: Why researchers use local LLMs
 - When local models are preferable to cloud-based models
 - How to get started with running LLMs on your own computer
- **Takeaway:** Local LLMs allow researchers to analyze sensitive data while keeping the data on their own computers.

Cloud vs. Local LLMs

- Two Ways to Use an LLM

Table 19: Two Ways to Use an LLM

Cloud LLM	Local LLM
Your data are sent to a remote server	Your data remain on your computer
Requires an Internet connection	Can work offline
No installation required	Requires one-time installation
Limited control over the environment	Full control over the environment

- Which would you choose?
 - Drafting a blog post?
 - Summarizing confidential interview transcripts?
- **Takeaway:** Local models provide greater privacy and control, while cloud models offer convenience.

Why Researchers Use Local Models

- **Advantages of Local LLMs**

- **Privacy:** Sensitive data never leave your computer.
- **Offline capability:** No Internet connection is required.
- **Cost:** Most open-source models can be used without ongoing subscription fees.
- **Control:** You choose the model, software, and update schedule.

- **Common Research Applications**

- Interview transcripts
- Focus group discussions
- Clinical narratives
- Internal organizational reports
- Student assignments

- **Takeaway:** Privacy and data governance are among the strongest reasons researchers choose local LLMs.

Are Local Models as Good as Cloud Models?

- There is no perfect choice

Table 20: Cloud versus Local Models

Cloud Models	Local Models
Often achieve higher performance	Often use smaller models
Usually faster on powerful servers	Speed depends on your hardware
Require Internet access	Can run completely offline
Easy to start using	Require installation and model download

- **Choosing a model involves balancing:**
 - Performance
 - Privacy
 - Cost
 - Convenience
- **Takeaway:** The best model depends on your research requirements, not on benchmark scores alone.

How Do I Choose an Open-Source Model?

- Before selecting a model, ask yourself:
- **Research Questions**
 - What research task am I performing?
 - How accurate does the model need to be?
 - Are my documents confidential?
 - How long are my documents?
- **Technical Questions**
 - Will the model fit on my computer?
 - Is it actively maintained?
 - Does its license allow my intended use?
- **Takeaway:** Select the model that fits your research task, hardware, and privacy requirements.

Running a Local LLM Is Easier Than You Think

- **Typical Workflow**

- Install a local LLM application.
- Download an open-source model.
- Open the application.
- Load your documents.
- Begin interacting with the model.

- **During This Workshop**

- You will receive a guide manual to run a local LLM so you can implement on your own computer after the workshop.

- **Takeaway:** Running a local LLM no longer requires advanced programming skills.

Will My Laptop Be Powerful Enough?

- **The answer depends on the model**

Table 20: Smaller versus Larger Model Comparison

Smaller models	Larger model
Require less memory Run faster Produce lower-quality responses in some tasks	Produce stronger results Require more memory and storage May run more slowly on consumer hardware

- **Recommendation:** Start with a model that runs comfortably on your current computer. Upgrade only if your research requires additional performance.
- **Takeaway:** The largest model is not always the best choice for your workflow.

A Typical Research Workflow Using a Local LLM

- A local LLM can support many stages of a research project.

Table 21: LLM’s Role Per Research Stage

Research Stage	How a Local LLM Can Help
Prepare data	Clean and organize text
Explore data	Summarize documents
Analyze data	Extract variables and identify themes
Organize results	Produce structured tables
Write reports	Draft summaries and findings

- Throughout the entire workflow, your sensitive documents remain on your own computer.

Learn Principles, Not Just Models

- Open-source LLMs are evolving rapidly.
 - Today's most popular model may be replaced next year.
 - Fortunately, the skills you learned today will continue to apply.
- **Skills that transfer across models**
 - Writing effective prompts
 - Evaluating outputs critically
 - Choosing appropriate models
 - Protecting sensitive data
 - Verifying results
- **Takeaway:** Learn enduring research skills rather than memorizing the features of a particular model.

Summary for Topic 8

- The best model is not the newest or the largest. It is the one that best supports your research objectives while respecting your privacy, computational resources, and methodological requirements.

Table 22: If-Do Chart for LLM Tool Selection

If you...	Start with...
Want to experiment with LLMs	Hugging Face Chat
Want to analyze confidential text	A local LLM (e.g., LM Studio or Ollama)
Want the easiest installation	LM Studio
Want to integrate LLMs into Python workflows later	Ollama

Topic 9: Traditional ML vs LLMs

Start with the Research Question, Not the AI Tool

- A common question is: "Should I use an LLM or traditional machine learning?"
- The better question is: "What am I trying to accomplish?"
- Different research goals require different methods

Table 23: Choosing AI Methods For Research Goals

Research Goal	Good Starting Point
Understand or generate text	LLM
Predict future outcomes	Traditional ML
Classify large structured datasets	Traditional ML
Summarize interviews	LLM
Extract variables from text	LLM

Traditional ML Excels at Prediction

- **Traditional machine learning** is often **preferred** when your primary goal is to predict, classify, or estimate outcomes from structured data.
- **Examples** in Social Science
 - Predict whether a student will graduate on time.
 - Predict voter turnout using demographic variables.
 - Estimate the risk of employee turnover.
 - Forecast housing prices or unemployment rates.
 - Detect fraudulent survey responses.
 - Predict disease outcomes from structured health records.
- **Why?** Traditional ML models are designed to learn relationships between variables and optimize predictive accuracy.
- **Takeaway:** If your research question begins with “Can we predict...?”, traditional machine learning is often the better starting point.

LLMs Excel at Understanding Language

- LLMs are especially useful when the data are primarily **text** and the goal is to interpret or organize information.
- **Examples in Social Science**
 - Summarize interview transcripts.
 - Code open-ended survey responses.
 - Identify themes in focus groups.
 - Extract variables from policy documents.
 - Draft literature summaries.
 - Explain statistical results in plain language.
- **Why?** LLMs were trained to recognize patterns in natural language rather than numerical relationships.
- **Key Takeaway:** If your research question begins with "**What does this text say?**", an LLM is often the better starting point.

Comparing LLMs and Traditional Machine Learning

Table 24: Comparing LLMs and Traditional ML

Question	Traditional ML	LLM
Works best with structured data	Yes	Sometimes
Understands natural language	Limited	Excellent
Generates new text	No	Yes
Predicts future outcomes	Excellent	Limited
Easy to explain model predictions	Often easier	Sometimes difficult
Processes millions of records efficiently	Yes	Usually slower
Requires labeled training data	Usually	Often not
Handles qualitative research naturally	Limited	Excellent

A Simple Decision Framework

- Before choosing a method, consider the following questions. **Table 25: Questions for Selecting AI Methods**

Question	If YES...
Is my data primarily text?	Consider an LLM
Is my data mostly structured variables?	Consider Traditional ML
Do I need to predict future outcomes?	Traditional ML
Do I need to summarize or interpret language?	LLM
Is reproducibility critical?	Traditional ML may be preferable
Is privacy a major concern?	Consider a local LLM

- Ask Yourself**
 - What type of data do I have? What question am I trying to answer? How important are accuracy, interpretability, and scalability?
- Takeaway:** Choosing a method is about balancing your research objectives, data, and practical constraints.

Sometimes the Best Solution Is Both

- **Hybrid Workflows:** Many research projects may combine LLMs and traditional machine learning.
- **Example:** A researcher has **100,000 open-ended survey responses**.
 - **Step 1:** Use an LLM to code 1,000 responses.
 - **Step 2:** Review and refine the codes.
 - **Step 3:** Train a traditional classifier using the reviewed labels.
 - **Step 4:** Classify the remaining 99,000 responses efficiently.
- **Benefits**
 - LLMs reduce manual coding effort.
 - Traditional ML provides speed and scalability.
 - Researchers maintain quality through human review.
- **Takeaway:** LLMs and traditional machine learning are often complementary rather than competing approaches.

Researcher Notes

- Distinction between **prediction** and **understanding**.
 - **Do you want to predict something that hasn't happened yet?** → Think traditional ML.
 - **Do you want to understand or organize language that already exists?** → Think LLM.
- Social science research often includes **both** structured and unstructured data. A mixed-methods project might involve:
 - interview transcripts,
 - demographic variables,
 - survey scales,
 - administrative records.
- Researchers can use LLMs to transform unstructured text into structured variables and then use traditional machine learning or conventional statistical methods for prediction, hypothesis testing, or causal analysis.
- **Final takeaway:** LLMs are **one component of a broader analytical toolkit**, rather than as replacements for existing methods

Resources

- NIMLAS 2026 Workshop Manual: Creating a Hugging Face Account and Testing HuggingChat
- NIMLAS Workshop 2026 Demo Prompts
- NIMLAS Workshop 2026 Manual: Installing Ollama and Running Your First Gemma Model Using the Ollama App
- NIMLAS Workshop 2026 Manual: Evaluating Open-Source Large Language Models for Research

Thank you!