

# Welcome to **instats**

The Session Will Begin Shortly

# START

# LLMs for Qualitative and Mixed-Methods Social Network Analysis

## Session 3: LLM Capabilities and Mixed-Methods Designs

Moses Boudourides

*Faculty, Graduate Program on Data Science  
Northwestern University*

[Moses.Boudourides@gmail.com](mailto:Moses.Boudourides@gmail.com)

instats Seminar

Thursday, January 29, 2026  
6:00 PM – 7:30 PM UTC

Northwestern | SCHOOL OF PROFESSIONAL STUDIES

# Outline

Capabilities and limitations of LLMs for qualitative relational analysis

Validation, transparency, and reflexivity

Ethical challenges in LLM-augmented research

# From Qualitative Interpretation to Computational Assistance

Previous sessions established:

- ▶ Networks are interpretive constructions
- ▶ Meaning precedes measurement

Today: *What can LLMs contribute to this process?*

# What LLMs Fundamentally Do

LLMs are:

- ▶ Statistical language models
- ▶ Trained to predict plausible text

**LLMs operate on *language*, not on social relations.**

# LLMs as Relational Text Analyzers

## A New Lens on Text

- ▶ LLMs are not just text generators
- ▶ beyond simple word counts, they can detect and analyze the *relationships* between *entities* in text
- ▶ they can *categorize*, i.e., distinguish entities and relationships types (e.g., kinship vs. professional) based on linguistic cues
- ▶ they can identify and predict complex patterns and structures in text
- ▶ they can scale qualitative analysis to a speed and efficiency that's difficult for humans to match
- ▶ **they cannot model networks directly!**

# Entity Extraction, Relationship Identification, and Categorization

- ▶ **Entity Extraction:** Identifying key actors, organizations, names, places, events or concepts in the data
- ▶ **Relationship Identification:** Identifying relationships and connections between these entities (e.g., trust, power dynamics, rivalry etc.)
- ▶ **Entity or Relationship Categorization:** Classifying the types or characteristics of extracted entities or identified relationships
- ▶ **Interpretation remains a human-led task:** LLMs identify *potential* relationships, entities, predicates, and categories for researcher review

## Example

From a news article, an LLM could extract that "Company A acquired Company B", "Companies A and B belong to certain sectors", and that "Person X is the CEO of Company A."

# Entity Extraction vs. Relational Interpretation

## Extraction

Identifying that a connection exists.

LLMs are good at extraction, but interpretation still requires human judgment.

## Interpretation

Understanding the meaning, nature, and quality of that connection.

# Coding, Memoing, and Theory-Building

- ▶ **Coding** (systematic labeling): LLMs can assist with coding large volumes of data—primarily text, but not only—based on a researcher-developed scheme.
- ▶ **Memoing** (reflective notes): LLMs can help generate memos by summarizing key themes and patterns.
- ▶ **Theory-Building:** By identifying unexpected patterns, LLMs can spark new theoretical insights.

# Coding Is Not Classification

In qualitative research:

- ▶ **Codes evolve:** coding is iterative—labels shift as new cases, exceptions, and emerging themes reshape what the researcher is trying to capture.
- ▶ **Categories are contested:** categories are rarely neutral; they reflect theoretical commitments, local meanings, and boundary disputes, and different analysts can defend different cuts of the same material.

LLMs can **propose** candidate codes and surface alternative framings, but they cannot **resolve** interpretive disagreements or replace the researcher's judgment about what a category should mean.

# Memoing and Theory Building

Memoing involves:

- ▶ **Reflective interpretation:** articulating what a passage suggests, noting ambiguities, and recording the analyst's evolving hunches, questions, and decisions.
- ▶ **Linking data to theory:** connecting observations to concepts and mechanisms, proposing explanations, and tracking how evidence supports, refines, or challenges a theoretical frame.

LLMs can **draft** plausible analytic memos and suggest connections, but researchers must **revise and own the meaning**—checking claims against the data, sharpening concepts, and making accountable interpretive choices.

# Computational Grounded Theory

LLMs enable:

- ▶ **Scaled pattern detection:** rapid scanning of large corpora to surface recurring themes, contrasts, and outliers that may warrant closer human interpretation.
- ▶ **Systematic memo support:** consistent drafting of summaries, comparisons, and cross-case prompts that help sustain an audit trail and keep analytic thinking moving.

But grounded theory remains a **human practice**: researchers still decide what counts as a concept, how categories relate, which comparisons matter, and what theoretical claims are justified by the data.

## Abductive Reasoning

- ▶ **Abduction** is the process of inferential reasoning from an observation to the most likely (best) explanation through data-theory iteration
- ▶ LLMs can support this by identifying surprising patterns or *anomalies* in data that demand explanation (e.g., actors with similar structures but different identities)
- ▶ These “surprises” may serve as catalysts for new theoretical insights

# Sentiment and Affect Analysis

- ▶ LLMs can analyze the emotional tone of text, providing insights into the affective dimensions of relationships.
- ▶ This can be used to understand trust, conflict, and solidarity within a network.

## Caution

Sentiment analysis is culturally specific and can be easily misinterpreted.

**Validation is crucial.**

# Bias in LLM Training Data

## “Garbage In, Garbage Out”

- ▶ LLMs are trained on vast amounts of text from the internet, which contains societal biases and cultural asymmetries related to race, gender, and other social categories
- ▶ These biases are reflected in the LLM's output: they are *amplified at scale!*

# Hallucination and Factual Accuracy

## Plausible but False...

LLMs can “hallucinate,” generating information that is plausible-sounding but factually incorrect. This is a major risk in research, *especially dangerous on qualitative inference*.

## The Need for Validation

All LLM output must be rigorously validated against the source text and, if and where possible, against human judgment.

# Semantic Homogenization

## The Risk of Flattening Meaning

- ▶ LLMs may smooth over linguistic variations and collapse diverse, nuanced meanings into a few uniform categories, a process called **semantic homogenization**.
- ▶ This can obscure local idioms or minority interpretations and important cultural and contextual differences by erasing and misrepresenting minority meanings.
- ▶ Researchers must guard against this to preserve qualitative richness. *Semantic homogenization contradicts qualitative aims!*

# Reflexivity and Transparency

- ▶ **Reflexivity:** Researchers must be critically aware of how the LLM is shaping their analysis and must ask:
  - ▶ Why this prompt?
  - ▶ Why this coding scheme?
  - ▶ LLM use itself should be understood as another plausible methodological choice.
- ▶ **Transparency:** Researchers must be transparent about how they used the LLM, including:
  - ▶ Prompt documentation
  - ▶ Output validation
  - ▶ Version tracking
  - ▶ Interpretive checkpoints

The human is the analyst.

The LLM is the research  
assistant.

### LLMs Are Not Analysts

LLMs do not:

- ▶ Possess theoretical sensitivity
- ▶ Understand social context

**LLMs cannot replace qualitative judgment.**

# Division of Labor in LLM-Augmented Research

## LLM Tasks

- ▶ Breadth, consistency, and speed of analysis
- ▶ Speed and scale
- ▶ Pattern recognition

## Human Tasks

- ▶ Depth of analysis
- ▶ Interpretation
- ▶ Meaning, theory, and ethical judgment

# Validation and Quality Control

- ▶ **Human-in-the-loop validation:** A human researcher reviews a sample of the LLM's output to assess its accuracy.
- ▶ **Inter-rater reliability:** Comparing the LLM's coding to that of a human coder.
- ▶ **Triangulation:** Comparing the LLM's findings to those from other data sources or methods.
- ▶ **Cross-LLM validation:** Comparing the findings from a certain LLM to those obtained from the use of other LLMs.

# Audit Trails and Reproducibility

## Document Everything

To ensure reproducibility and accountability, researchers must create a detailed audit trail of their research process.

- ▶ The exact prompts used
- ▶ The version of the LLM used
- ▶ The validation procedures
- ▶ The code used to process the LLM's output

# Ethical Challenges in LLM-Augmented Research

- ▶ **Privacy:** How is sensitive data protected when sent to a third-party API?
- ▶ **Consent:** Have participants consented to their data being analyzed by an LLM?
- ▶ **Bias:** How can we prevent the LLM from amplifying existing societal biases?
- ▶ **Responsible Use:** How might the findings be used, and how can we prevent harm?

# Data Privacy and Consent

## A Major Concern

Sending data to third-party LLM providers raises significant privacy concerns. Researchers must ensure that they have the appropriate data security measures in place and that they have obtained informed consent.

# Responsible Use and Accountability

- ▶ Researchers are accountable for the outputs of the LLMs they use.
- ▶ They must consider the potential societal impacts of their research.
- ▶ They must be prepared to justify their methodological choices.

# LLMs in Mixed-Methods Designs

LLMs can be integrated into all of the mixed-methods designs we have discussed:

- ▶ **Sequential:** Use an LLM to analyze a large dataset after initial qualitative work.
- ▶ **Parallel:** Compare the findings of an LLM analysis to a concurrent human analysis.
- ▶ **Embedded:** Use an LLM as a tool to assist with coding and memoing within a qualitative study.

## Session 3 Summary and Looking Forward

- ▶ We have examined the capabilities and limitations of LLMs for qualitative relational analysis.
- ▶ We have discussed the importance of validation, transparency, and reflexivity.
- ▶ We have highlighted the ethical challenges that arise when using LLMs in research.

### Next Session

We will explore specific research designs that integrate LLMs and networks.

## Questions and Discussion

Thank you!

## Questions?

# STOP