Welcome to

The Session Will Begin Shortly

START

LLMs for Qualitative and Mixed-Methods Social Network Analysis

Session 4: Research Designs with LLMs + Networks

Moses Boudourides

Faculty, Graduate Program on Data Science
Northwestern University

Moses.Boudourides@gmail.com

instats Seminar

Thursday, January 29, 2026
7:30 PM – 9:00 PM UTC

Northwestern | SCHOOL OF PROFESSIONAL STUDIES

Outline

Design patterns for integrating LLMs into Qualitative Analysis

Design patterns for integrating LLMs into Mixed-Methods

Validity, transparency, and ethical considerations

From Capabilities to Design

Previous sessions focused on:

- Interpretation-first networks
- What LLMs can and cannot do

Today: how to *design* research that uses LLMs responsibly.

Why Research Design Matters

Without explicit design:

- LLMs appear as black boxes
- Interpretation is displaced
- Validity is unclear

Design restores methodological control.

Division of Labor Principle

Central question:

> *Which tasks are delegated to LLMs, and which remain human?*

This division must be explicit and justified.

Breadth vs Depth

LLMs excel at:

- Breadth
- Consistency

Humans excel at:

- Depth
- Contextual interpretation

Design balances the two.

Design Patterns for LLM-Augmented SNA

- **Augmenting, not replacing!**
- These are not rigid templates, but flexible patterns for leveraging LLMs to augment the researcher's capabilities.
- They can be combined and adapted to fit specific research goals.

Pattern 1: Initial Data Exploration

- Use LLMs for a first pass on unstructured textual corpora, such as:
  - Small-N qualitative work
  - Interviews, ethnography, documents
  - Initial coding and concepts
- Identify key themes, entities, and potential relationships.
- Generate summaries to get a feel for the data.

Goal

- To develop an initial understanding that guides more focused qualitative inquiry, not to produce final results.
- This phase anchors interpretation.

Pattern 2: Systematic Coding at Scale

How to develop a robust coding scheme through in-depth qualitative analysis of a data sample:

- Develop a preliminary **codebook** from a purposive sample (definitions, inclusion/exclusion rules, and short examples).
- **Pilot** the codebook on additional cases; refine codes, resolve ambiguities, and record decision rules (audit trail).
- Prompt the LLM to **apply** the finalized scheme to the full corpus (with explicit formatting and justification requirements).
- **Validate** with human coders on a held-out subset (agreement, error analysis), then iterate: revise codebook/prompt and re-run.

Benefit

Combines the interpretive depth of qualitative coding with the scale and repeatability of computational analysis.

Pattern 3: Relationship Extraction and Network Construction

- Use LLMs to identify (and categorize) entities and the relationships between them from unstructured text.
- Extract information about the nature, direction, and strength of ties.
- Convert this extracted information into a structured network format (e.g., an edge list).

Crucial Step

The extracted network must be validated and refined by the human researcher.

Pattern 4: Anomaly Detection and Theoretical Insight

Finding the Unexpected

Use LLMs to identify surprising patterns, anomalies, or contradictions in the data that might not be apparent to a human reader.

- These anomalies can challenge existing assumptions and spark new theoretical insights.
- The researcher then investigates these anomalies using focused qualitative methods.

Pattern 5: Comparative Analysis Across Cases

- Use LLMs to extract the same variables and relational claims from texts across many cases (e.g., organizations, communities) using a shared template/codebook.
- Standardize outputs (entities, roles, tie types, evidence snippets) so cases become directly comparable rather than impressionistically summarized.
- Compare cases by identifying recurring motifs, contrasts, and outliers in relational patterns and mechanisms.
- Enables scalable, auditable comparative network analysis by coupling cross-case consistency with human interpretation of meaning and context.

Sequential Design: Qual → LLM → Qual

A powerful, integrated workflow:

1. **Qualitative Phase 1:** Conduct interviews or observations to develop a deep understanding of the context and key concepts.
2. **LLM Phase:** Use the insights from Phase 1 to guide a large-scale LLM analysis of a broader dataset.
3. **Qualitative Phase 2:** Return to qualitative methods to interpret, validate, and elaborate on the LLM findings.

An Example of Sequential Design in Practice: Mentorship Study

- **Phase 1:** In-depth interviews with mentors and mentees to understand the meaning of mentorship.
- **Phase 2:** Use an LLM to analyze thousands of company emails to identify mentorship ties and their characteristics based on themes from Phase 1.
- **Phase 3:** Conduct follow-up interviews to explore the organizational-level patterns revealed by the LLM.

Parallel Design: LLMs + Humans

Concurrent Analysis

LLM analysis and human qualitative analysis proceed at the same time, with each stream of work informing the other.

- LLM findings can prompt new questions for the human analyst.
- Qualitative insights can reveal the limitations or biases of the LLM analysis.
- Creates a productive dialogue between computational and human interpretation.

Parallel Design Benefits and Analyses Divergence

Parallel Design Benefits:

- Triangulation
- Bias detection
- Reflexive comparison

Convergence and divergence both matter.

When Analyses Diverge:

- LLM bias
- Human blind spots
- Genuine ambiguity in data

Divergence is not failure.

An Example of Parallel Design in Practice: Online Community Study

LLM Analysis

- Topic modeling of forum posts.
- Sentiment analysis.
- Network extraction of user interactions.

Human Analysis

- Qualitative coding of a sample of posts.
- Interviews with community members.
- Digital ethnography.

Embedded Design: LLMs Within a Qualitative Framework

LLM as a Tool, Not a Phase

In this design, LLM analysis is not a separate stage but is integrated throughout a primarily qualitative research process.

- An LLM might assist with coding, memoing, or literature review as the researcher works.
- The LLM serves as a supportive tool to enhance the efficiency and reach of the qualitative researcher.

Fully Integrated Design

The Most Complex Design

Qualitative and computational elements are woven together at every stage of the research process in a continuous, iterative cycle.

- Requires careful coordination and a high level of methodological expertise.
- Can produce exceptionally rich and comprehensive findings.

Choosing a Design

The right design depends on:

- **Research Questions:** Are you focused on meaning, structure, or both?
- **Data Availability:** Do you have a large textual corpus suitable for LLM analysis?
- **Resources:** Time, budget, and personnel.
- **Researcher Expertise:** Align the design with your methodological strengths.

Validity, Transparency, and Audit Trails

Foundations of Rigor

Regardless of the design, rigor depends on:

- **Validity:** Ensuring your findings are credible and well-grounded.
- **Transparency:** Being explicit about your methods and decisions. Researchers should disclose:
  - Where LLMs were used
  - How outputs were filtered
  - What judgments were applied

  Transparency strengthens credibility.
- **Audit Trails:** Documenting your process so that others can understand and potentially reproduce it. Responsible design requires:
  - Prompt documentation
  - Versioning
  - Decision logs

  Audit trails enable scrutiny.

Validity in LLM-Augmented Research

We must consider multiple dimensions of validity:

- **Construct Validity:** Are we measuring what we think we are measuring?
- **Internal Validity:** Are our conclusions about the case at hand sound?
- **External Validity:** Can our findings be generalized?
- **Interpretive Validity:** Do our findings capture the meanings of the participants?
- **Relational Validity:** Do extracted ties reflect the *actual relation* intended (direction, type, strength, temporality), not just co-mention or rhetorical association?
- **Integrative Validity:** Are LLM outputs coherently integrated within our research without category drift or incompatible assumptions?
- **Computational Validity:** Is the LLM functioning as intended?

Member Validation and Feedback Loops

Checking In with Participants

Member validation—presenting findings to research participants for feedback—is essential.

- It strengthens interpretive validity.
- It provides accountability to the community being studied.
- It can reveal limitations or biases in the LLM analysis.

Common Design Pitfalls

Avoid:

- *Treating LLM output as ground truth:* LLMs can hallucinate, flatten nuance, and misread context—treat outputs as hypotheses to be checked.
- *Hiding prompts and corrections:* undocumented prompts, post-hoc edits, and silent cleaning undermine transparency and reproducibility.
- *Over-automating interpretation:* coding and memoing can be assisted, but meaning, boundary decisions, and theoretical claims remain human responsibilities.

Iterative Refinement and Reflexivity

- LLM-augmented research is not linear; it is an iterative cycle of analysis, interpretation, and refinement.
- This requires **reflexivity**—a constant critical reflection on how our tools and choices are shaping the findings.

Ethical Design Considerations

Ethics from the Start

Ethical considerations should not be an afterthought. They must be integrated into the research design from the very beginning.

- Data privacy and consent.
- Representing others' meanings
- Potential for harm or stigmatization.
- Equitable benefit to the community.
- Responsibility cannot be delegated.

Session 4 Summary and Looking Forward

- We explored five design patterns for integrating LLMs into Qualitative Analysis.
- We examined sequential, parallel, embedded, and fully integrated mixed-methods designs.
- We emphasized the crucial role of validity, transparency, and ethical considerations.

Next Session

We will move from design to practice, looking at the computational tools and ethical responsibilities involved in this work.

Questions and Discussion

Thank you!

Questions?

STOP