Using AI for Qualitative Content Analysis

Josef Slerka, 4. 4. 2025

Qualitative Content Analysis

- **Vast Amount of Textual Data**
  News articles, social media posts, interviews, and surveys are proliferating at an unprecedented scale.
- **Central Goal**
  Systematically identify, code, and interpret patterns or themes in textual sources (Neuendorf, 2017).
- **Traditional Method**
  Manual coding guided by codebooks; can be time-consuming and relies on highly trained researchers.

Challenges of Manual Approaches

- **Time-Intensive**
  Large datasets require significant hours of human effort.
- **Subjectivity & Bias**
  Coder drift and personal interpretation can affect reliability (Krippendorff, 2013).
- **Resource Constraints**
  Small teams may struggle to process large corpora, limiting the scope of analysis.

Emergence of LLM Solutions

- **Large Language Models**
  GPT-4, Claude, and similar systems show promise for automating coding, classification, and thematic extraction (Chew et al., 2023).
- **Speed & Scalability**
  These AI tools handle extensive text corpora quickly—potentially **reducing** months of work to days.
- **Consistency Gains**
  Fixed prompts can reduce subjective bias by applying standardized coding rules (Bijker et al., 2024).

Toward Hybrid Approaches

- **Human-in-the-Loop**
  Researchers refine prompts and validate outputs, blending AI's speed with expert interpretation (Xiao et al., 2023).
- **Methodological Rigor**
  Prompt engineering, pilot testing, and iterative validation are key to maintain data quality (Wachinger et al., 2024).
- **Opportunities & Future**
  Ongoing development in LLMs is expanding their capabilities for summarization, memoing, and advanced content analysis.

Use of LLMs for Qualitative Coding

- Leveraging Large Language Models (LLMs) such as ChatGPT or GPT-4 for:
  - **Deductive coding** (predefined codes)
  - **Inductive coding** (emergent themes)
  - **Hybrid approaches** (mixed methods)

1. Deductive Coding with a Predefined Codebook

Scenario: You have established codes (e.g., “Confidence in Data,” “Policy Recommendation,” “Emotional Tone”). You want the LLM to classify new texts using these codes.

Sample Prompt:

> You are a **qualitative coding assistant**.
> **Codebook**: (1) Confidence in Data, (2) Policy Recommendation, (3) Emotional Tone.
> **Instruction**: For each excerpt, assign **one** code and provide a short rationale.

Workflow:

1. Provide the LLM with your code definitions.
2. Input the text excerpt.
3. The LLM outputs the **chosen code** + an **explanation**.
4. Compare to human coder results, refine prompt or codes if needed.

2. Inductive (Open) Coding of Interview Excerpts

Scenario: You have **no predefined** codes. The LLM identifies potential themes directly from the text.

Sample Prompt:

> "You are a *research assistant* analyzing interview data on remote work.
> Identify **3–5 distinct themes** in the following passage, each with a short label and 1-sentence rationale."

Example:

- *Improved Work-Life Balance* (the speaker discusses better personal/work separation)
- *Team Isolation* (the speaker feels disconnected from colleagues)
- *Increased Online Hours* (the speaker works more frequently online)

Benefit:

- Quickly surfaces **emerging categories** or ideas.
- Follow-up steps: **cluster** codes, refine them, or merge overlaps.

3. “Chain-of-Thought” Coding for Deeper Explanations

Scenario: You want the model to provide **step-by-step** reasoning for how it applies each code.

Sample Prompt:

> “You are a *qualitative researcher*.
> For each text, assign the best code from (Empathy, Justification, Critique) and produce a **chain-of-thought** explanation.
> Show the reasoning that led to your final code.”

Why It Helps:

- Improves **transparency**—researchers see exactly how the AI interprets text.
- Facilitates **error-checking**—if the reasoning doesn't match the chosen code, you can adjust definitions or the prompt.

4. Hybrid Deductive + Inductive Approach

Scenario: You have **some** predefined codes but want the LLM to propose *new codes* when needed.

Sample Prompt:

> “Use these existing codes (Financial Constraint, Personal Motivation, Environmental Impact).
> If none apply, **create a new code** and define it in 1–2 sentences.
> Provide a one-sentence rationale for each code you assign.”

Example:

- Excerpt about saving money and reducing carbon footprint:
  - **Financial Constraint + Environmental Impact**
  - No new code needed.

Outcome:

- Balances the **structure** of deductive coding with the **flexibility** of inductive discovery.

5. Validation & Agreement Checking

Scenario: After the LLM codes a dataset, compare its results with human-coded data.

Sample Prompt:

> "Below are 10 excerpts with LLM labels and human labels. For each disagreement, decide which label is more accurate and explain why."

Usage:

- Reveals where the LLM systematically differs from human coders.
- Guides **prompt refinements** (e.g., clarifying definitions or adding new instructions).
- Produces **agreement metrics** like Cohen's κ or accuracy.

Our Use Case: Map of media

Deductive Coding with a Predefined Codebook

https://www.nfnz.cz/mapa-medii/

What is original content?

Criterion ORIGINALS in code book

This criterion serves to differentiate between media outlets that provide original content and those that predominantly publish repurposed articles without significant added value. Sources of repurposed information may include agencies, other media outlets, social networks, or press releases. Translations from foreign media are also considered repurposed articles.

Original content typically includes:

- Original news findings/topics and interviews with relevant stakeholders;
- Follow-up on external findings supplemented with new information from relevant actors;
- Work with documents not yet publicly known that provide new information;
- Original journalistic content (commentary pieces, interviews).

Non-original content typically includes:

- Building on someone else's findings while only adding - background information without input from relevant actors;
- Monitoring other media, agencies, social networks, or - easily accessible public databases;
- Compilations from other texts that do not add new original information;
- Repurposed journalistic content.

The biggest problem is that it is time consuming, one website (120 articles) took the coder 12 hours.

So we tried an AI solution...

Prompt

> Please tell us what the main theme of the text is and rate the text according to the following criteria on a scale of 1 - 10, where 1 is a poor fulfilment of the criterion and 10 is an excellent fulfilment of the criterion and, if necessary, give examples of errors. These criteria are:
>
> 1. Originality and originality of content: was the article written and published by the medium or was it taken from another source? Write from which sources the information was taken. If the article quotes people, say whether they were contacted directly or quoted from other sources.
> 2. Value added by the author or medium: Has the article been substantially modified or expanded to include additional information, analysis, or commentary by the author? Does it include a statement made directly to the site or medium?

3. Source of information: Did the information come directly from the source (e.g., from an interview or own research) or was it taken from other media, news agencies, social media, or press releases? Articles citing one or no sources are usually not credible.
4. Overall rating: on a scale of 1-10, how would you rate this article overall?

Model Performance

- **Accuracy:** 86.73%
- **Precision:** 96.36%
- **Recall:** 82.81%

What do these numbers mean

- **Accuracy** tells us how often the AI was correct overall.
  → In 86.73% of cases, the AI correctly identified whether an article was original or not.
- **Precision** tells us how often the AI was right *when* it said something was original.
  → Out of 100 articles labeled as original by the AI, about 96 were truly original.
- **Recall** tells us how many real originals the AI actually found.
  → The AI successfully caught 83 out of every 100 truly original articles.

Costs and benefits

- We use Claude 3.5 Sonnet
- The price for 120 articles is $3.
- We're able to code hundreds of sites

Advantages

Advantages

Automation of Coding

LLMs can perform open and axial coding in minutes. For example, ChatGPT has been used to analyze survey responses and extract key themes.

Reducing Subjectivity

LLMs produce consistent outputs using fixed prompts, reducing bias and increasing reliability in qualitative analysis.

Speed and Scalability

Tasks that used to take weeks can now be done in hours. LLMs handle large volumes of text efficiently, enabling broader research designs.

Consistency and Reliability

LLMs apply prompts uniformly, avoiding human coder drift. Errors tend to be systematic rather than random, making them easier to detect.

Multi-lingual and Multi-domain Flexibility

LLMs can work across languages and domains, reducing the need for translation or specialized tools.

Generation of Explanations and Insight

LLMs can explain why a theme was identified, helping researchers interpret and write up findings.

Integration of Knowledge

LLMs may connect text to external knowledge, enriching analysis—though care must be taken to avoid adding information not in the original data.

Limitations & Challenges

Hallucinations

- **Invented Quotes & Sources**
  LLMs can produce plausible but entirely fabricated details.
- **Risk of Misinformation**
  May compromise data integrity if unchecked.
- **Mitigation**
  Always verify model outputs against original texts or external sources.

Domain-Specific Nuances

- **Missing Subtle Context**
  Jargon, cultural references, or historical contexts may be misinterpreted.
- **Limited Background Knowledge**
  The model’s training might not fully cover specialized fields.
- **Solution**
  Provide domain-specific prompts and consult experts to ensure accuracy.

Verification Needed

- **Human Review**
  Essential for confirming correctness, resolving ambiguities, and refining codes.
- **Complement, Not Replace**
  AI assists but does not eliminate the need for expert judgment.
- **Iterative Workflow**
  Test and iterate prompts on smaller data samples before large-scale coding.

Privacy and Data

- **Anonymizing Sensitive Data**
  Potential breach of confidentiality if raw transcripts are exposed to public APIs.
- **Ethical Compliance**
  Requires data protection strategies and possibly on-premises LLMs for sensitive content.
- **Responsible AI Use**
  Transparency about AI involvement and adherence to data-handling regulations.

BTW: AI was used for translation, language editing and some parts of the research in this presentation.

Thank you for your attention

References

1. Bazeley, P., & Jackson, K. (2013). *Qualitative Data Analysis with NVivo*. SAGE.
2. Bijker, R. et al. (2024). *ChatGPT for Automated Qualitative Research: Content Analysis*. JMIR.
3. Chew, R. F. et al. (2023). *LLM-Assisted Content Analysis: Using Large Language Models to Support Deductive Coding*. arXiv.
4. Dunivin, Z. O. (2024). *Scalable Qualitative Coding with LLMs: Chain-of-Thought Reasoning Matches Human Performance in Some Hermeneutic Tasks*. arXiv.
5. Krippendorff, K. (2013). *Content Analysis: An Introduction to Its Methodology*. SAGE.
6. Neuendorf, K. (2017). *The Content Analysis Guidebook* (2nd ed.). SAGE.

References

7. Wachinger, J. et al. (2024). *Prompts, Pearls, Imperfections: Comparing ChatGPT and a Human Researcher in Qualitative Data Analysis*. QHR.
8. Xiao, Z. et al. (2023). *Supporting Qualitative Analysis with Large Language Models: Combining Codebook with GPT-3 for Deductive Coding*. IUI Companion.