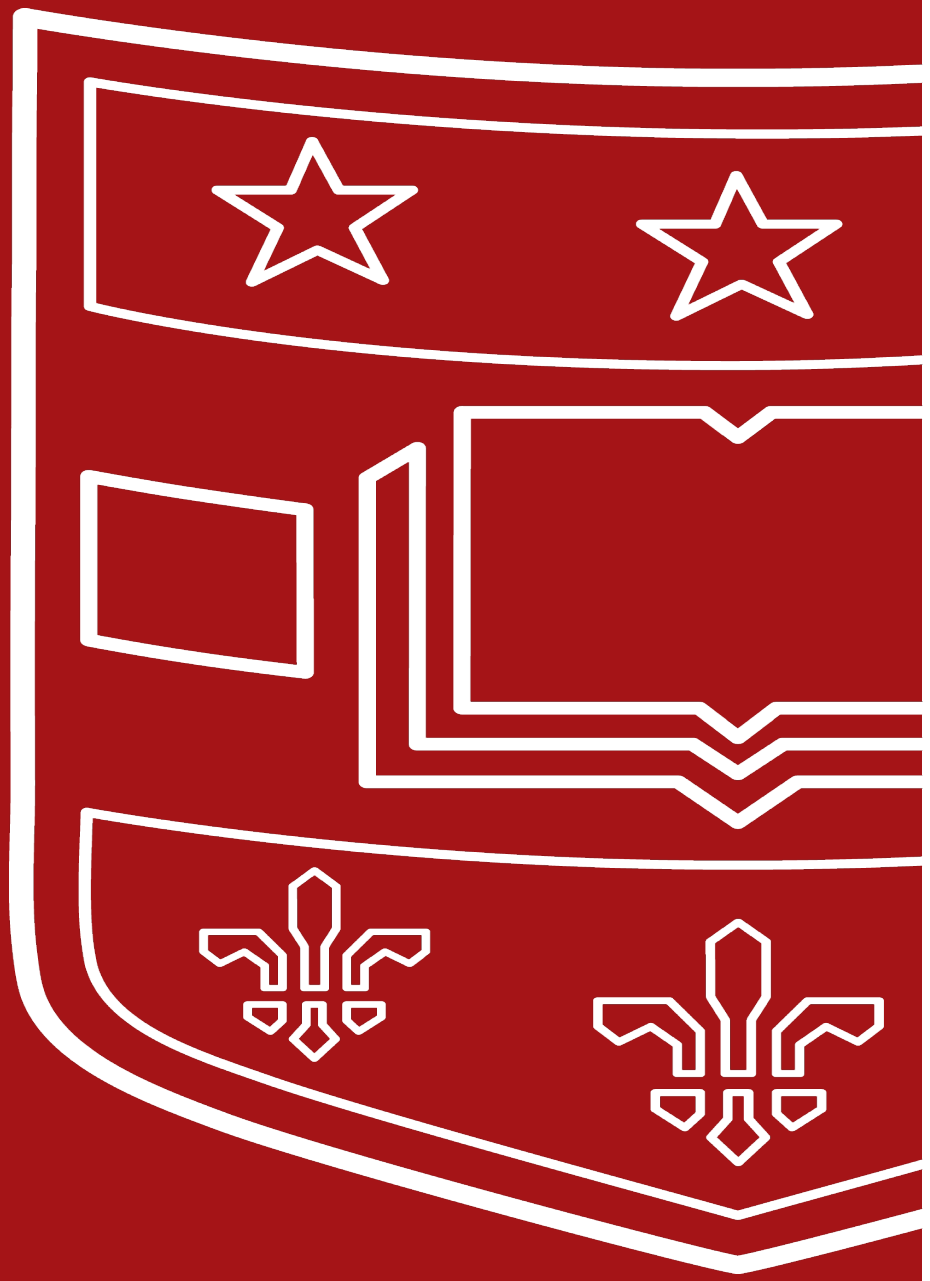
Reimagining Qualitative Research with Generative AI

Ruopeng An, Ph.D., MPP, FACE, FAAHB
Associate Professor
Faculty Lead in Public Health Sciences, Brown School
Faculty Fellow in AI Innovations for Education, Office of the Provost
Brown School & Division of Computational & Data Sciences



Washington University in St. Louis

Outline

- Beyond the dichotomy of quantitative vs. qualitative analysis
- Conventional quantitative analysis of qualitative data
- Text mining methods preceding modern (generative) AI
- Case studies using large language models to analyze qualitative data (sensitivity analysis, thematic analysis, ground theory, discourse analysis, data synthesis)

Data vs. Analysis

	Qualitative Data	Quantitative Data
Qualitative Analysis	Prevalent	Rare
Quantitative Analysis	Increasing	Prevalent

Conventional Quant Analysis on Qual Data

- Word count & frequency
- Word-by-respondent matrix
- Lexical richness (e.g., type-token ratio)
- Word cloud
- Free list and pile sorts
- Multidimensional scaling (MDS)
- Cluster analysis and taxonomy
- Similarity matrix and semantic network analysis

Word Count & Frequency

- **Definition:** A measure of how many times particular words appear in a text.

- **Example:** ”Obesity" appears 34 times in the Dietary Guidelines for Americans, 2020-2025.

Word-by-respondent Matrix

- **Definition:** A table showing the occurrence of words across different respondents' answers.

- **Example:** Respondent A mentions "quality" 10 times, B mentions it 5 times.

Type-token Ratio

- **Definition:** In a sentence or a paragraph, the ratio of the number of different words, called types, to the total number of words, called tokens. The ratio reflects the diversity of vocabulary (i.e., lexical richness) used in a text.

- **Example:** A child's story has a type-token ratio of 0.5.

Word Cloud

- **Definition:** A visual representation of word frequency in a text, with size indicating frequency.

- **Example:** "Pollution" is the largest word in an environmental report's word cloud.

Free List and Pile Sorts

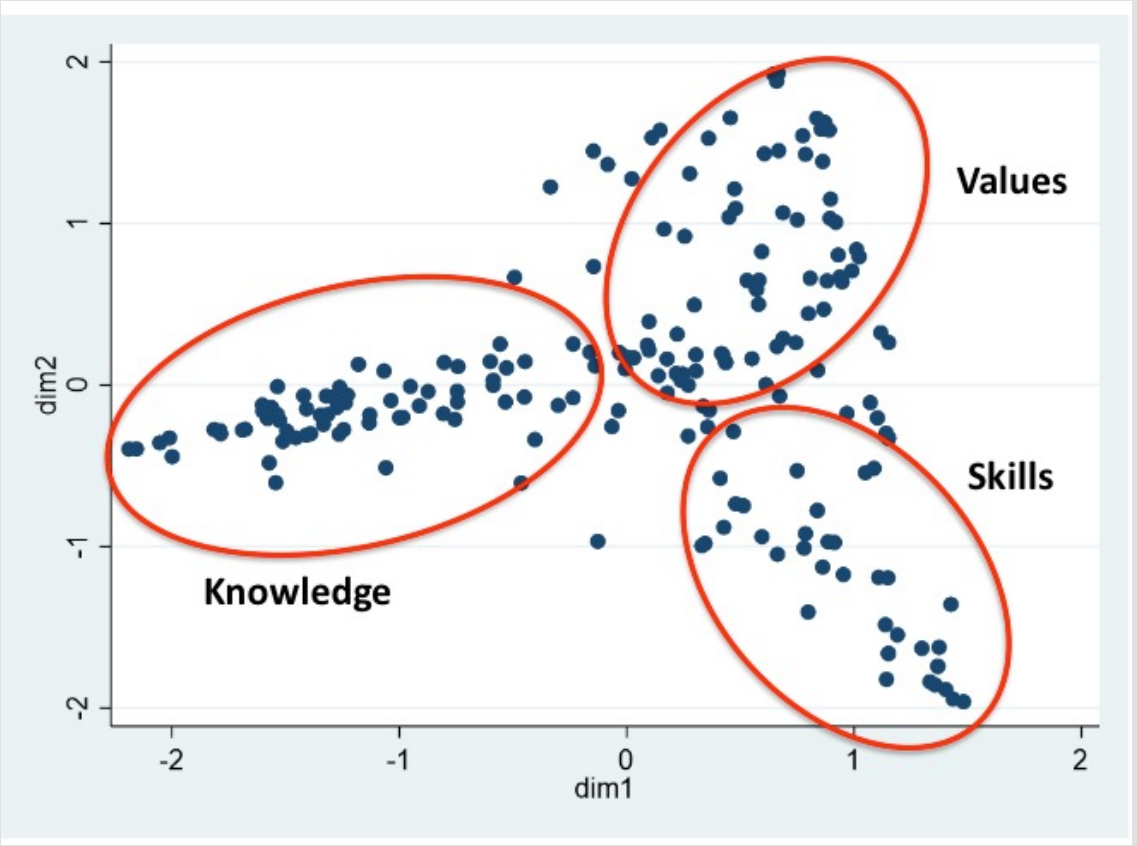
- **Definition:** Techniques for eliciting and categorizing items from respondents.

- **Example:** Participants list "apple," "banana," and "cherry," then sort into the "fruits" category.

Multidimensional Scaling (MDS)

- **Definition:** A statistical technique that visualizes the relative distances between items in a two-dimensional space.

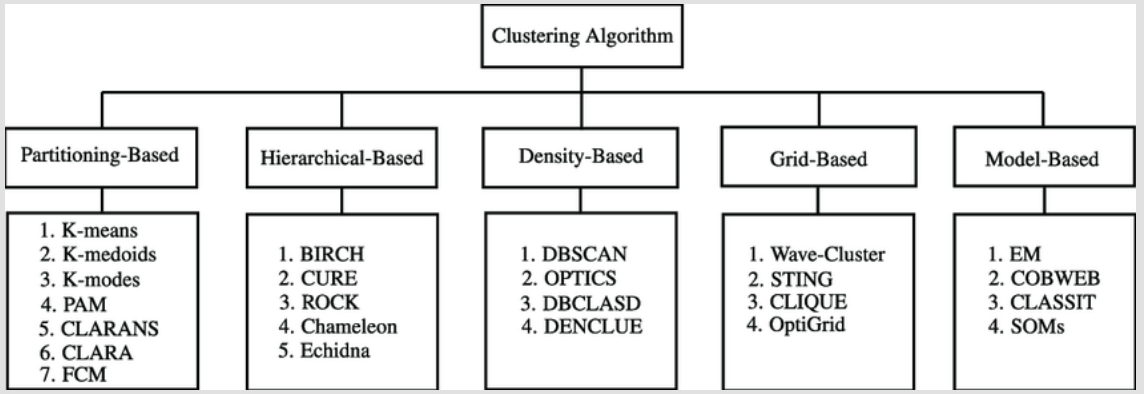
- Example: MDS plots various soft drinks based on taste test similarity ratings.



Source: Stark et al. at Brown School

Cluster Analysis and Taxonomy

- **Definition:** Grouping objects based on their characteristics to form a hierarchy.
- **Example:** Animal species are clustered into a taxonomy based on genetic traits.



Similarity Matrix and Semantic Network Analysis

- **Definition:** A matrix showing how similar objects are, used to analyze relationships.

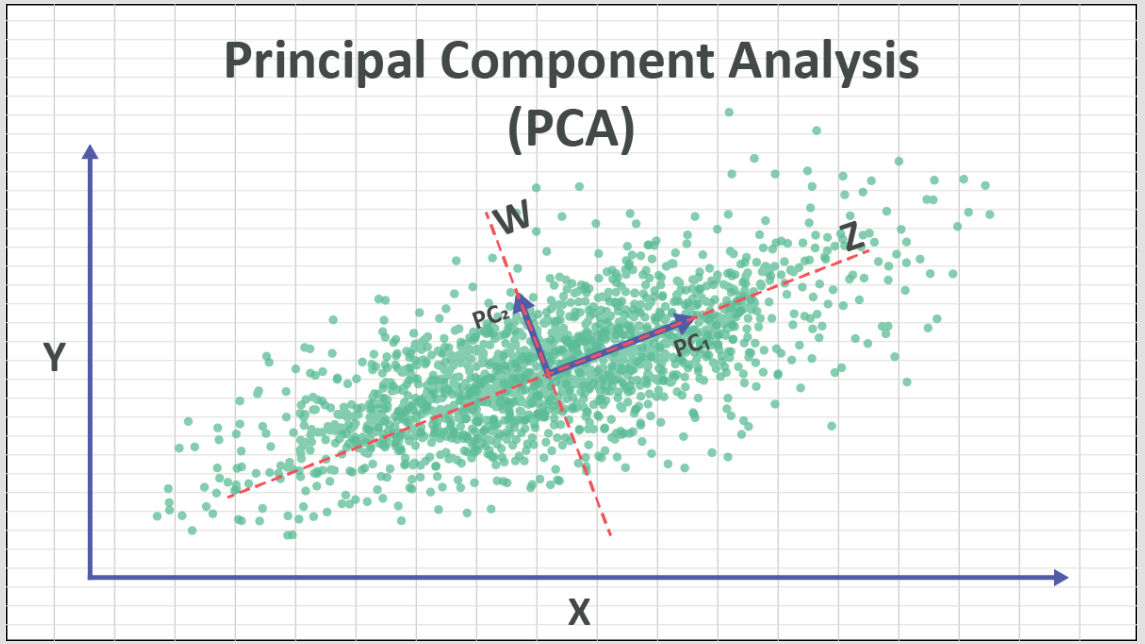
- **Example:** A matrix indicates "doctor" and "nurse" share semantic closeness in healthcare texts.

Text Mining Methods

- Dimensionality reduction
- Information retrieval
- Natural language processing (NLP, e.g., part of speech tagging)
- Named entity recognition (NER)
- Disambiguation
- Pattern identification
- Document clustering
- Coreference resolution
- Sentiment analysis
- Topic modeling
- Text preprocessing (e.g., tokenization, stemming, lemmatization)

Dimensionality Reduction

- **Definition:** Reducing the number of variables in data to the most significant ones.
- **Example:** Using PCA to simplify a dataset for easier visualization.



Information Retrieval

- **Definition:** The process of obtaining relevant information from a large repository.

- **Example:** Searching a database for articles on diabetes treatment options.

Part of Speech Tagging

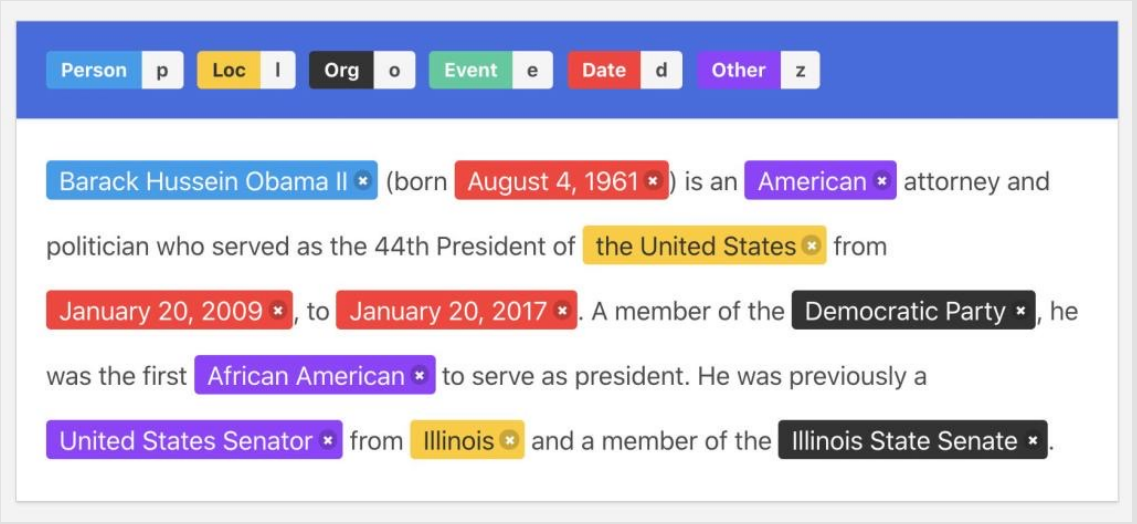
- **Definition:** The process of marking up a word in a text as corresponding to a particular part of speech.

- **Example:** In "The quick brown fox jumps over the lazy dog," 'jumps' is tagged as a verb (V) and ‘fox’ as the subject.

Named Entity Recognition (NER)

- **Definition:** Identifying and classifying key elements in text into predefined categories.

- **Example:** Extracting names of people and places from a news article.



Disambiguation

- **Definition:** Clarifying the meaning of words that have multiple interpretations.

- **Example:** Determining if "Apple" refers to the fruit or the tech company in a text.

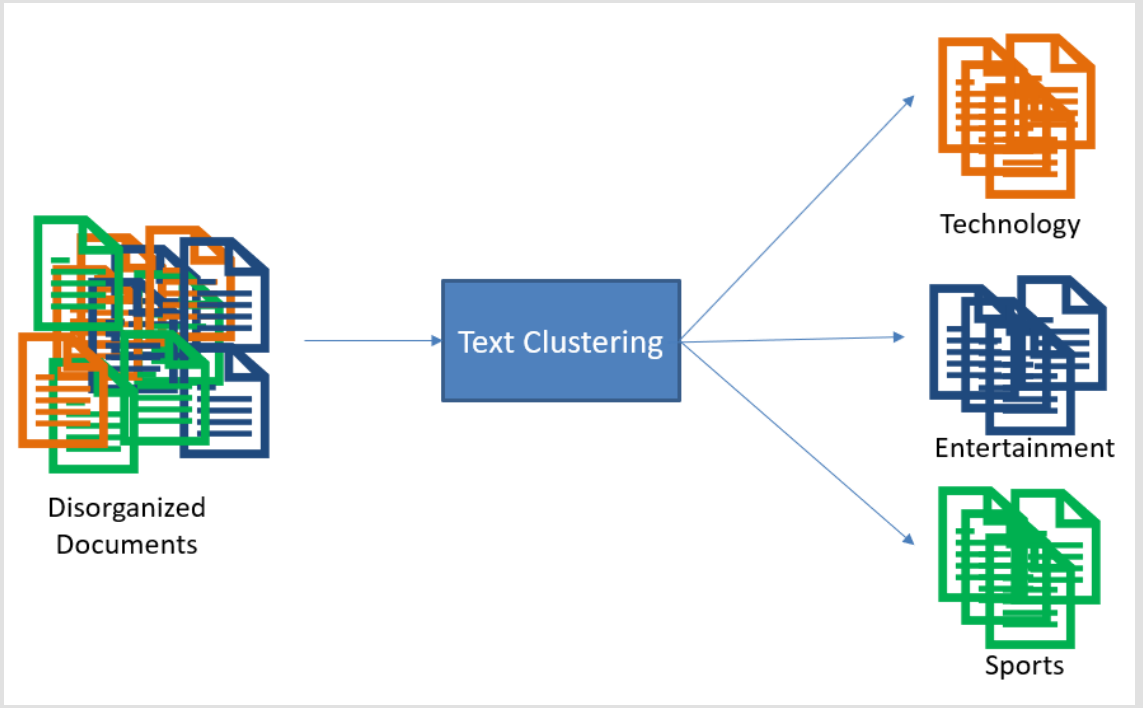
Pattern Identification

- **Definition:** Finding and extracting regularities or rules within data.

- **Example:** Detecting email or phone number formats in a document.

Document Clustering

- **Definition:** Grouping similar documents together based on content.
- **Example:** Organizing research papers by topic automatically.



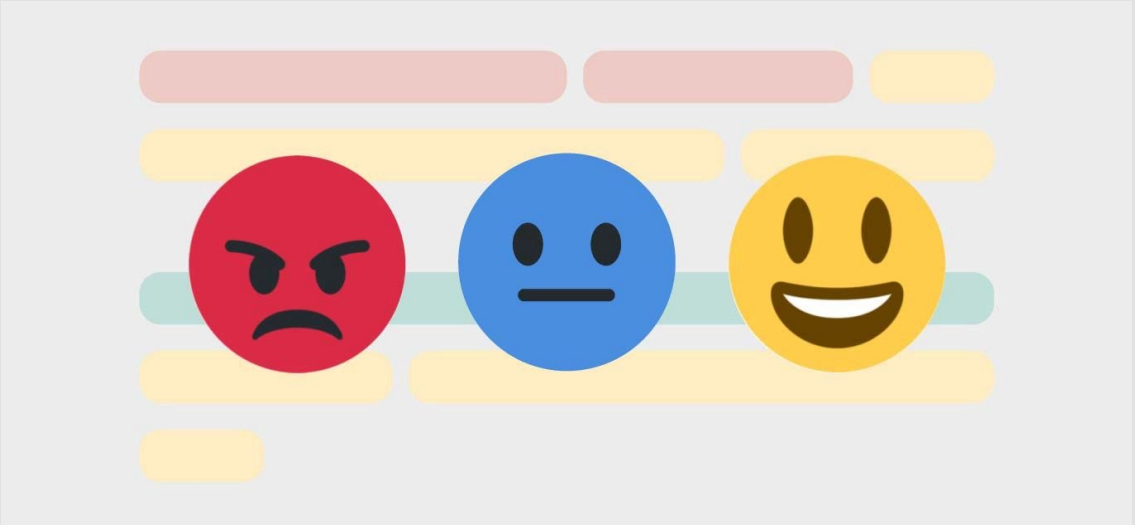
Coreference Resolution

- **Definition:** Determining when different words refer to the same entity in a text.

- **Example:** Recognizing that "Obama" and "he" in a paragraph are the same person.

Sentiment Analysis

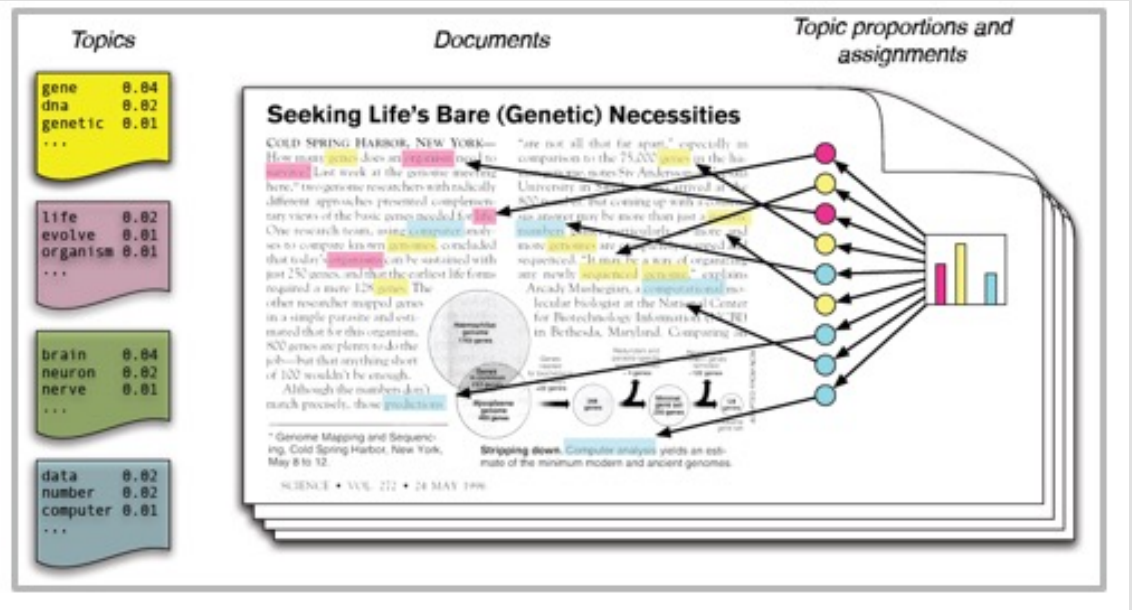
- **Definition:** Assessing the emotional tone behind a series of words.
- **Example:** Rating product reviews as positive, neutral, or negative.



Topic Modeling

- **Definition:** Discovering the abstract topics present in a collection of documents.

- **Example:** Finding common topics covered in thousands of news articles.



Text Preprocessing

- **Definition:** Preparing text for analysis by simplifying and standardizing it.

- **Examples:**
  - Lemmatization: Transform the word "running" to its base form "run."
  - Stemming: Reduce the word "fishing" to its stem "fish."

- **Note:** Lemmatization reduces words to their base or dictionary form while stemming simply removes affixes without regard to the word's actual meaning.

Generative AI for Qual Data Analysis

- **Scalability:** Analyze vast datasets swiftly, which is impractical for manual text analysis methods.
- **Contextual Understanding*:** Large language models grasp nuances and context, outperforming classic methods in tasks like sentiment analysis (e.g., using bag-of-words models) and topic modeling (e.g., using linear discriminant analysis or LDA).
- **Enhanced Accuracy*:** Leveraging deep learning, AI models reduce human error and bias in tasks like NER.
- **Integration of Multiple Methods:** AI can combine techniques (e.g., clustering, NLP) seamlessly for more robust analysis.
- **Adaptability:** AI models continuously learn, improving their analysis over time, unlike static traditional methods.

Key Features of Generative AI

- **Creates New Content:** Can generate text, images, and more, resembling human-made content.
- **Learns from Data:** Trains on large datasets to produce original outputs.
- **Adapts Across Domains:** Flexible in various fields like art, literature, and design.
- **Interactive and Responsive:** Can engage with users and iteratively refine its outputs.
- **Innovation Driver:** Fuels new possibilities in creativity and problem-solving.

Examples of Generative AI Models

- Text Generation Models
  - GPT (OpenAI): Offers conversational AI with a wide range of language understanding.
  - CLAUDE (Anthropic): Focuses on reliable and interpretable AI language models.
  - Hugging Face Transformers: A hub for various open-source models.

- Image Generation Models
  - Midjourney: An independent research lab's AI for creating images from textual descriptions.
  - Stable Diffusion: Generates high-quality images and supports creative expression.
  - DALL·E (OpenAI): Creates images from text captions.

Large Language Models (LLMs) for Qual Data Analysis: Case Studies

- Use ChatGPT(GPT-4) to conduct sentiment analysis (few-shot learning)

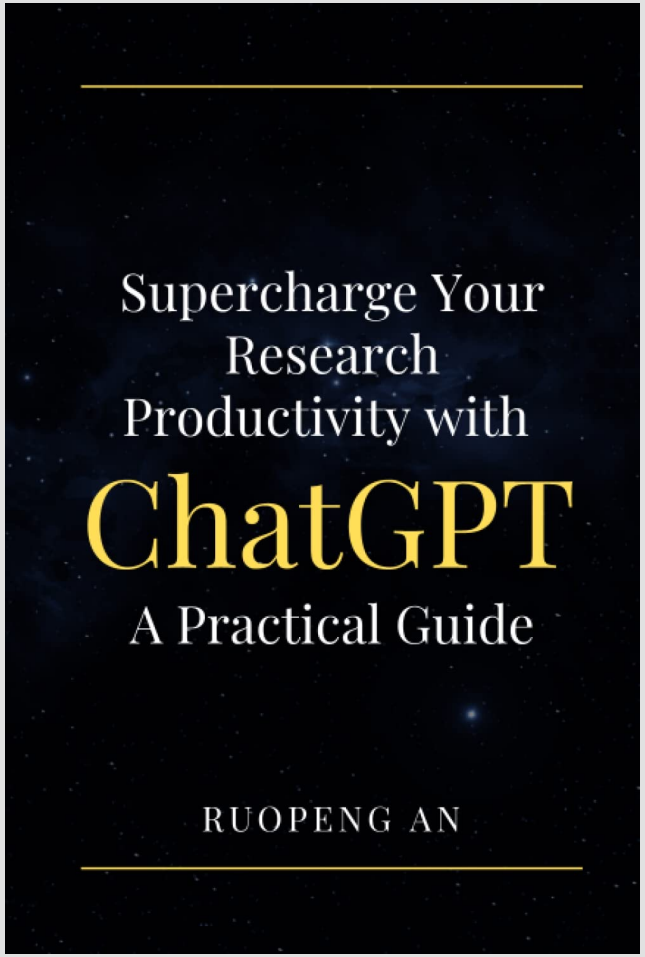
- Use ChatGPT (GPT-4) to conduct thematic analysis, ground theory, and discourse analysis (zero-shot learning and prompt engineering)

- Use OpenAI API (GPT-4) to identify and synthesize key themes and extract quotes from large volumes of interview and focus group transcripts

Demo from My Book

100 real-world examples covering 10 phases of research:

1. Identifying research topics and framing questions through an in-depth discussion with ChatGPT
2. Formulating and refining hypotheses based on the chosen research question
3. Undertaking literature reviews, covering all steps of a systematic review protocol
4. Selecting adequate research design and corresponding methodology
5. Developing valid, reliable, and efficient research tools
6. Handling every aspect of data collection, management, and ethics
7. Interpreting and analyzing both quantitative and qualitative data
8. Writing and refining research papers and reports
9. Addressing peer review comments
10. Disseminating study findings through mass and social media platforms



Amazon: https://a.co/d/6Y7C5G0

Qualitative Data Analysis: Barriers

Time consuming

Researchers must read lengthy transcripts to identify themes after many hours spent on recruiting and interviewing participants

Laborious

Researchers must manually annotating, summarizing, and synthesize lengthy transcripts to identify themes

Can Generative AI Help?

Identify and extract themes relevant to the research questions?

Provide detailed descriptions?

Extract the relevant quotes?

Our 4-step Approach

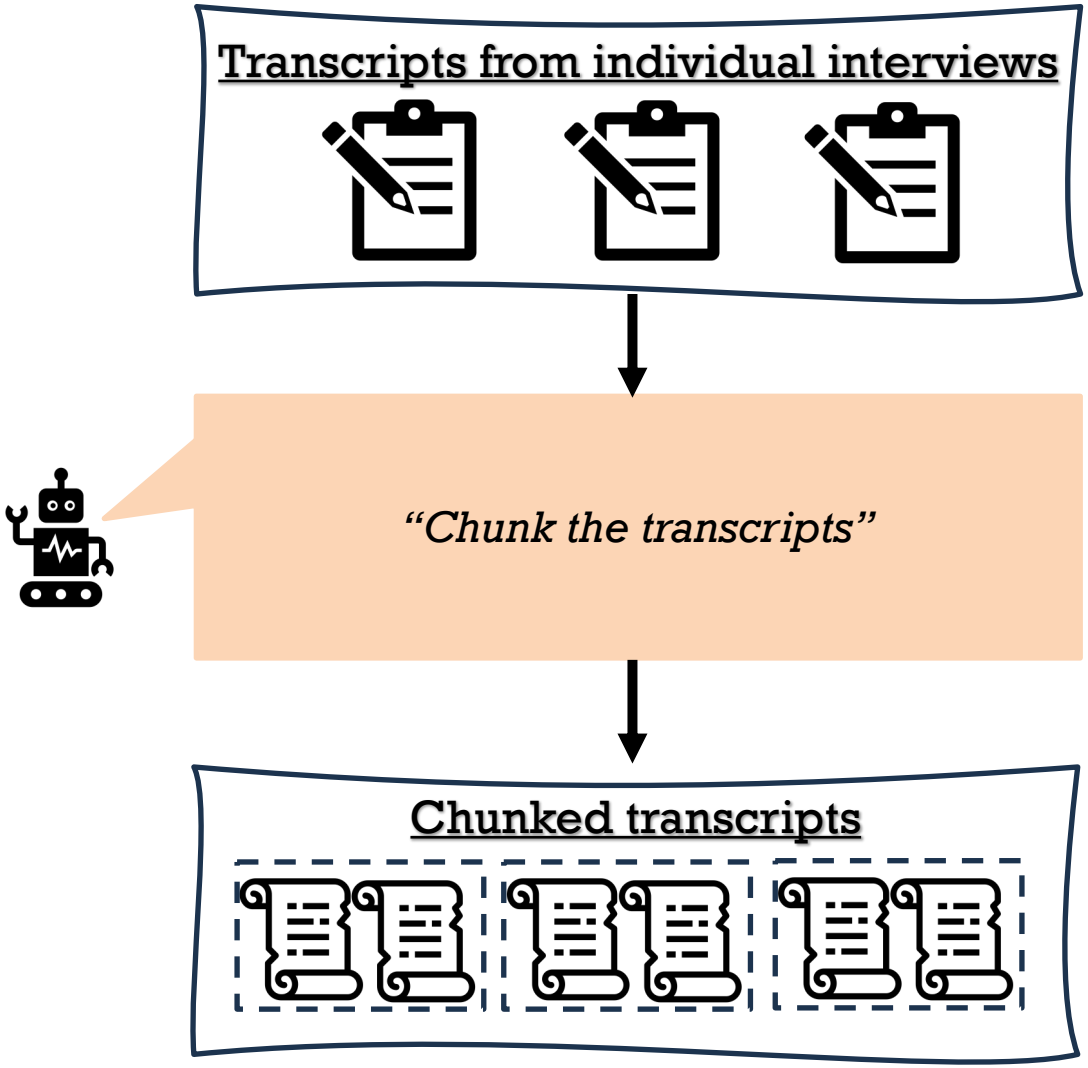
Step 1:
Chunking the transcripts

Step 2:
Extracting relevant themes

Step 3:
Synthesizing themes at the participant level

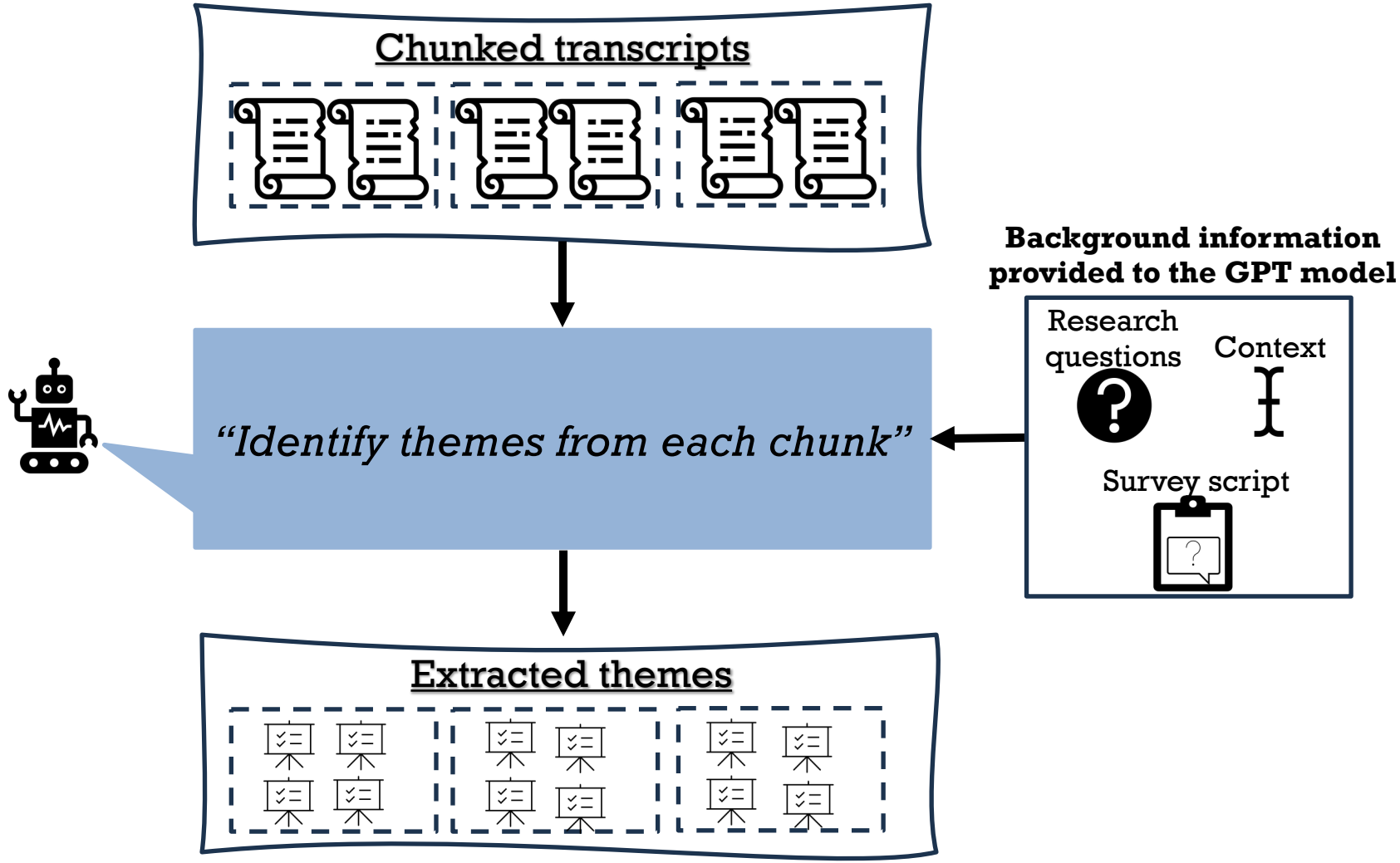
Step 4:
Integrating themes to the study level

Step 1: Chunking Transcripts

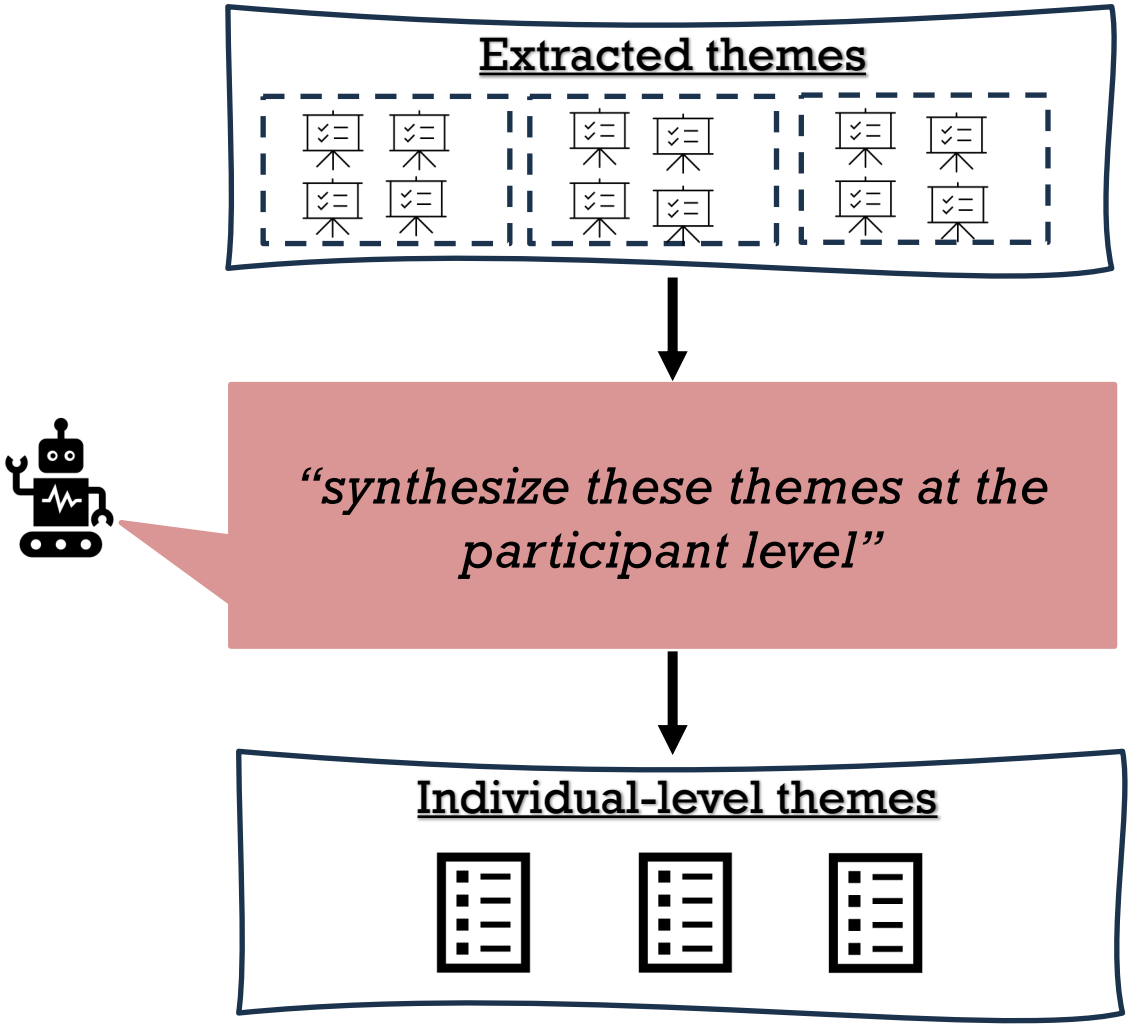


Why?
To deal with the limited context window of LLMs and the lost-in-the-middle problem

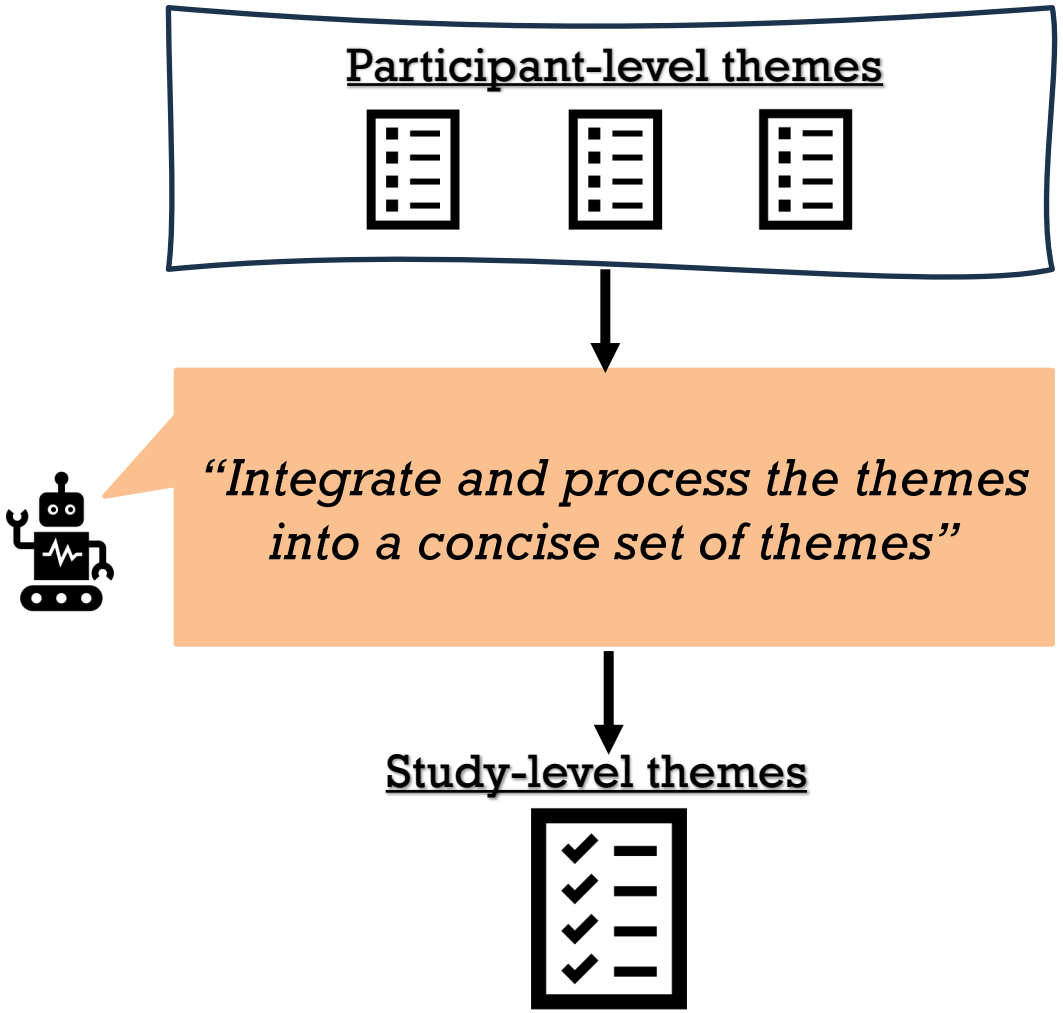
Step 2: Extracting Relevant Themes



Step 3: Synthesizing Themes at the Participant Level



Step 4: Synthesizing Themes to the Study Level



How Do We Evaluate Generated Themes?

Evaluating qualitative studies can be tricky

We want themes to be close to researcher-identified themes and of high quality by public health research standards

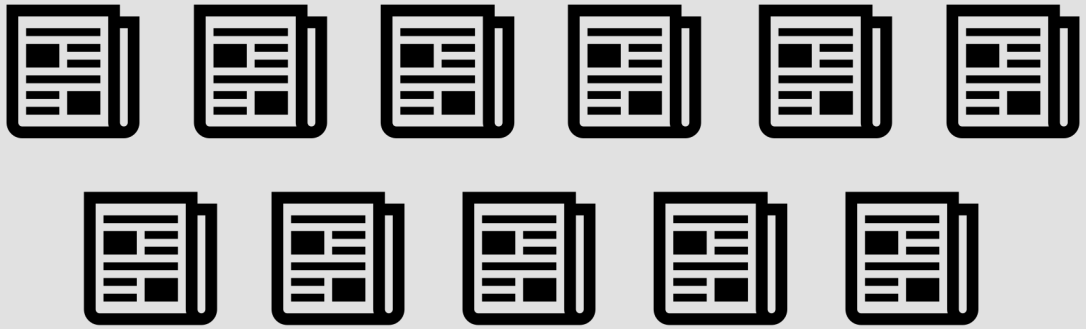
We DO NOT want themes to vague and possible hallucinated by the models

Evaluation Process Explained

5 human evaluators

Tasked with evaluating GPT generated themes on

11 published qualitative public health studies



Evaluated on 4 core-components

EVALUATED ON A 5-POINT LIKERT SCALE (0-4)

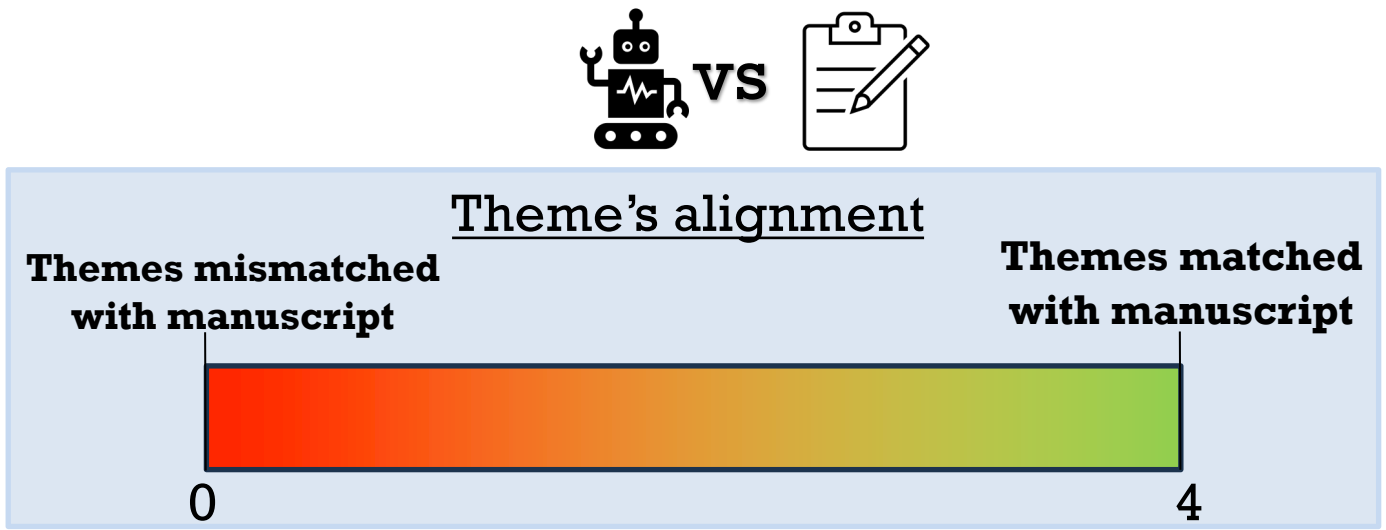
Alignment: Do the GPT-generated themes match the manuscript's published themes?

Succinctness: Are GPT-generated themes repetitive in nature?

Quality of explanations: Vague or descriptive?

Quality of quotes: Made-up or contextually relevant?

Evaluation Metric 1 – Alignment



Example from Ndupu *et al* on physical activity from university staff and students

Manuscript-listed themes	GPT-generated themes
Knowledge	Understanding and Awareness of Physical Activity
Environmental contexts and resources	Personal and Environmental Barriers to Physical Activity
Social Influences	Influence of Social Support and University Environment
Social/professional role and identity	
Beliefs about capabilities	Motivations for Physical Activity
Intention	Willingness and Desire for Change

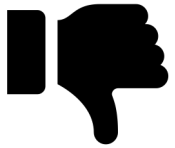
- <u>All</u> the manuscript-listed themes can be matched with the GPT-generated themes!
- 3 out of 4 evaluators gave this a rating of 4!

Ndupu, L. B., Staples, V., Lipka, S., Faghy, M., Bessadet, N., & Bussell, C. (2023). Application of theoretical domains framework to explore the enablers and barriers to physical activity among university staff and students: a qualitative study. *BMC Public Health*, *23*(1), 670.

Example from McFarren *et al* on feeding practices from Latinx mothers of children at risk of obesity

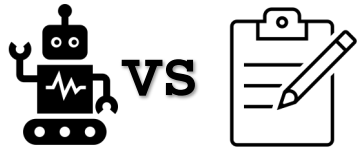
Manuscript-listed themes	GPT-generated themes
Feeding to appease or soothe and feeding when child request food at night	Child's Sleep and Soothe Patterns
	Influences on Feeding Practices
Food Choices including knowledge sources of nutritional information and use of juice and solid foods	Home Environment and Support Systems
Food and weight: Justification and perceptions of weight	Impact of Culture and Adaptation

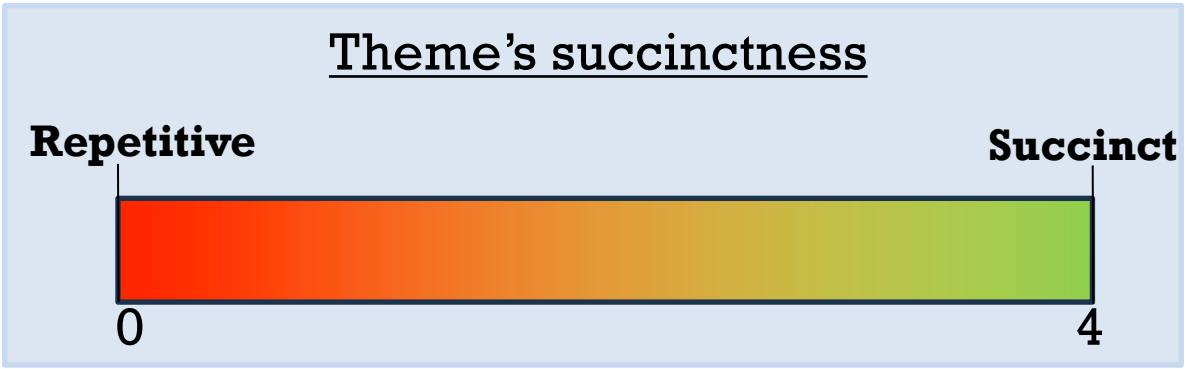
- The GPT-generated themes only matched one of the manuscripts identified themes!
- 3 out of 4 evaluators gave this a rating of 1!



McFarren, H., Vazquez, C., Jacobs, E. A., Dela Tejera, G., & Gray, M. (2020). Understanding feeding practices of Latinx mothers of infants and toddlers at risk for childhood obesity: A qualitative study. *Maternal & Child Nutrition*, *16*(3), e12983.

Evaluation Metric 2 – Succinctness





Example from Ndupu *et al* on physical activity from university staff and students

Manuscript-listed themes	GPT-generated themes
Knowledge	Understanding and Awareness of Physical Activity
Environmental contexts and resources	Personal and Environmental Barriers to Physical Activity
Social Influences	Influence of Social Support and University Environment
Social/professional role and identity	
Beliefs about capabilities	Motivations for Physical Activity
Intention	Willingness and Desire for Change

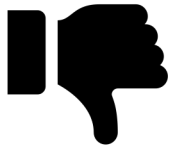
- GPT-generated themes here are uniquely succinct in their own right!
- 4 out of 4 evaluators gave this a rating of 4!

Ndupu, L. B., Staples, V., Lipka, S., Faghy, M., Bessadet, N., & Bussell, C. (2023). Application of theoretical domains framework to explore the enablers and barriers to physical activity among university staff and students: a qualitative study. *BMC Public Health*, *23*(1), 670.

Example from Gray *et al* on sleep habits from toddlers at risk of obesity

Manuscript-listed themes	GPT-generated themes
Structured Sleep Habits	Sleep routines
	Nighttime walking
Feeding during the middle of the night	Feeding Transitions
	Feeding quantities
Co-sleeping	Impact of Culture and Adaptation

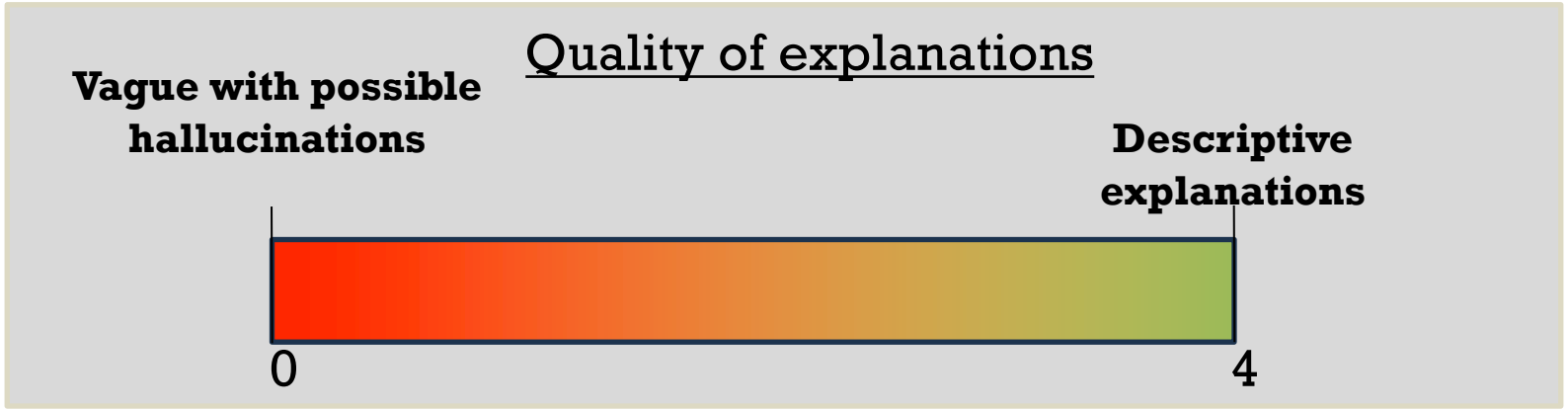
- GPT-generated themes of "Sleep routines" and "nighttime walking" can belong to the same theme of "structured sleep habits"!
- Same applies to the theme of feeding!
- 2 out of 4 evaluators gave this a rating of 2!



Gray, M. J., Vazquez, C. E., & Agnihotri, O. (2022). "Struggle at night–He doesn't let me sleep sometimes": a qualitative analysis of sleeping habits and routines of Hispanic toddlers at risk for obesity. *BMC pediatrics*, *22*(1), 413.

Evaluation Metric 3 – Quality of Explanations

Quality assessment



Example from Ndupu *et al* on physical activity from university staff and students

Theme:
Influence of Social Support and University Environment

Explanation:
The social environment, including friends and colleagues, in conjunction with the availability and quality of university-provided physical activity facilities profoundly impacts participant physical activity. This underlines the need for university-centered social and environmental interventions.

- Explanation is similar to the manuscript's explanation of the theme.
- Explanation is relevant to the GPT-generated theme and provides proper insight into the selected theme.
- All Evaluators gave it a rating of 4!

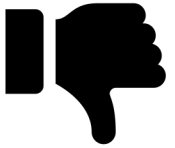
Ndupu, L. B., Staples, V., Lipka, S., Faghy, M., Bessadet, N., & Bussell, C. (2023). Application of theoretical domains framework to explore the enablers and barriers to physical activity among university staff and students: a qualitative study. *BMC Public Health*, *23*(1), 670.

Example from Gray *et al* on sleep habits from toddlers at risk of obesity

Theme:
Feeding Transitions: Mothers struggle to create consistent sleep routines for their toddlers

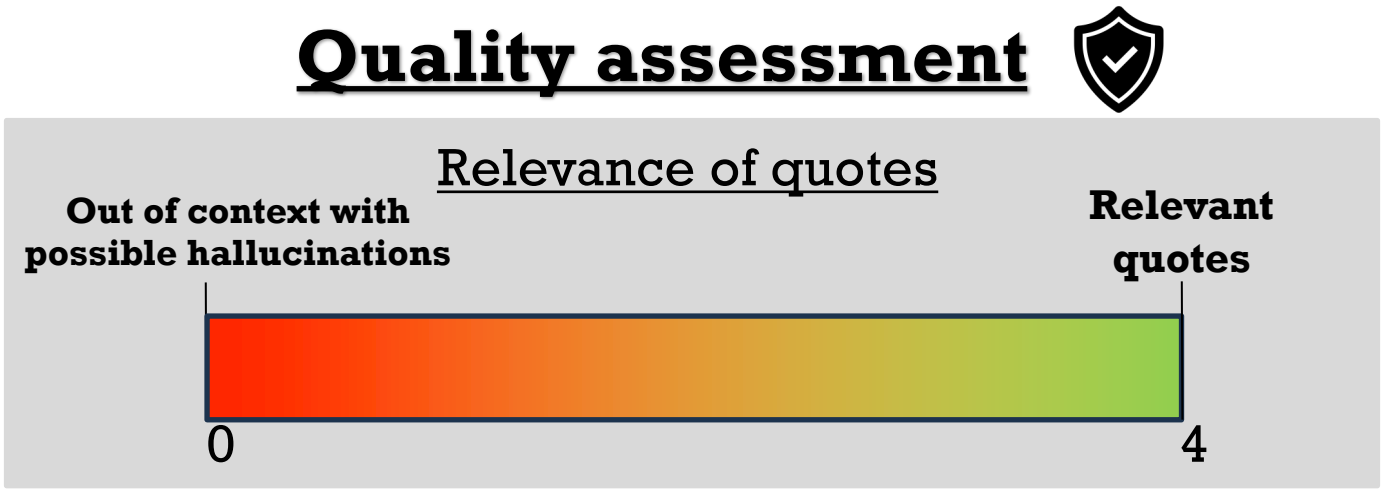
Explanation:
Such transition is not always smooth and involves understanding whether a child is feeding due to hunger or comfort.

- Explanation can be seen as irrelevant.
- Explanation is brief and does not provide any insight.
- Most evaluators gave it either a rating of 2.



Gray, M. J., Vazquez, C. E., & Agnihotri, O. (2022). "Struggle at night–He doesn't let me sleep sometimes": a qualitative analysis of sleeping habits and routines of Hispanic toddlers at risk for obesity. *BMC pediatrics*, *22*(1), 413.

Evaluation Metric 4 – Quality of Quotes



Example from Ndupu *et al* on physical activity from university staff and students

Theme:
Influence of Social Support and University Environment

Quote:
...you feel, "Okay, I got a workout partner, let's go ahead and go with it," rather than you don't really feel as motivated if you're going to go by yourself...

- Quote is relevant and informational to the theme identified by the GPT model!!!
- 3 Evaluators gave it a rating of 4!

Note: Quote shortened to be more presentable

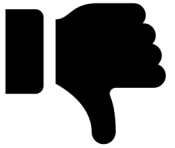
Ndupu, L. B., Staples, V., Lipka, S., Faghy, M., Bessadet, N., & Bussell, C. (2023). Application of theoretical domains framework to explore the enablers and barriers to physical activity among university staff and students: a qualitative study. *BMC Public Health*, *23*(1), 670.

Example from Gray *et al* on sleep habits from toddlers at risk of obesity

Theme:
Feeding Transitions: Mothers struggle to create consistent sleep routines for their toddlers

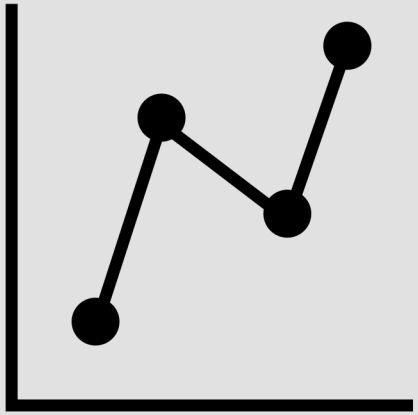
Quote:
"How did you go about building your routine? What time do you try to put him/her to sleep at night?"

- Quote does not seem relevant and informational to the GPT-generated theme
- Evaluators gave it either a rating of 1 or 2



Gray, M. J., Vazquez, C. E., & Agnihotri, O. (2022). "Struggle at night–He doesn't let me sleep sometimes": a qualitative analysis of sleeping habits and routines of Hispanic toddlers at risk for obesity. *BMC pediatrics*, *22*(1), 413.

Let's explore our evaluations quantitatively!

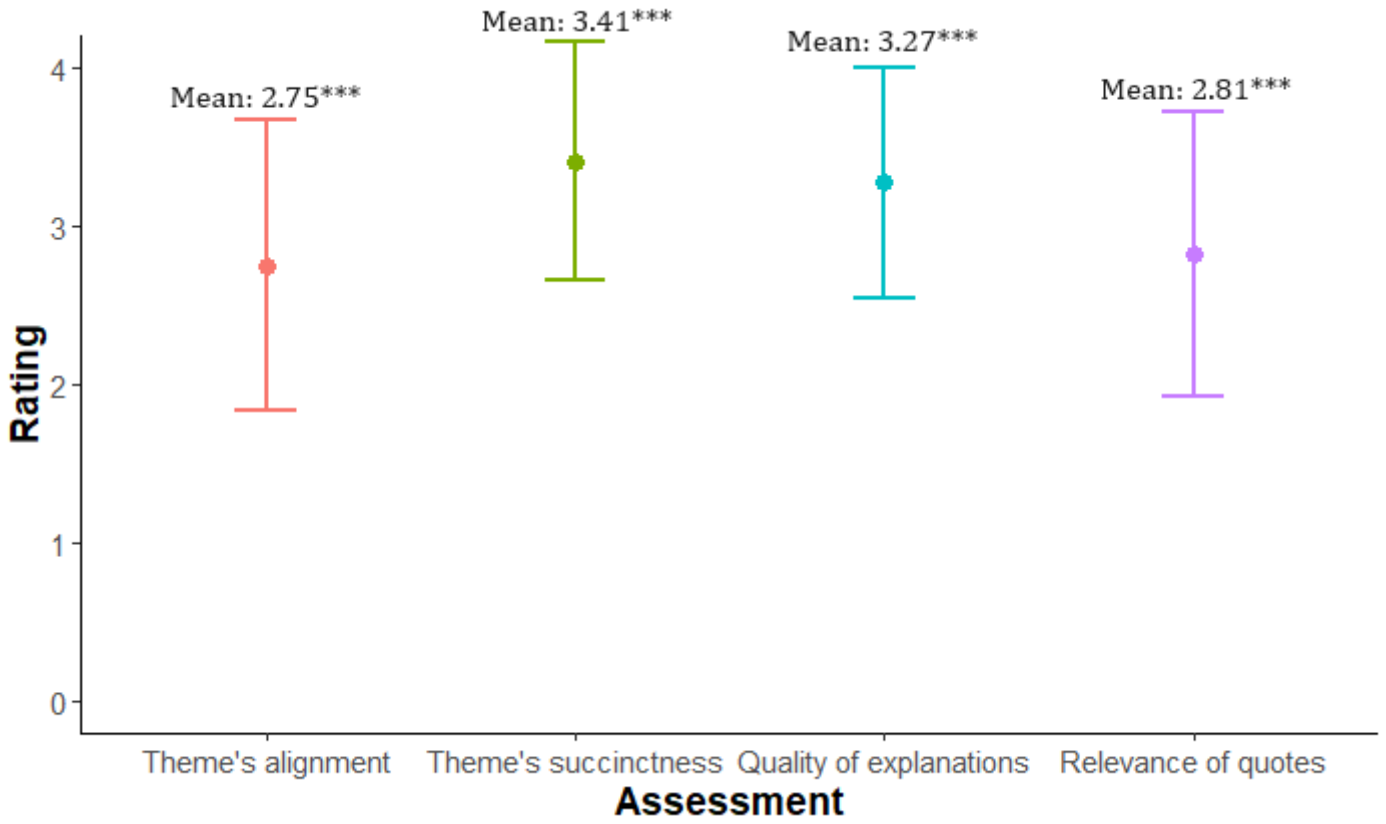


Interrater Reliability: ICC

Level	Inter-cluster correlation (ICC)	p-value
Overall	0.845	<0.001
Theme's alignment	0.648	0.01
Theme's succinctness	0.581	0.03
Quality of Quotes	0.893	<0.0001
Quality of explanation	0.650	0.01

Note: The ICC3k score assesses the reliability of the average scores of evaluators across multiple targets

Quan analysis reveals that GPT performs adequately



*Note: * p<0.05, ** p<0.01, *** p<0.001 based on t-test*

Conclusion

- LLMs can assist researchers in extracting themes from qualitative interviews
- Themes generated from LLMs can be used to complement field-specific domain knowledge (e.g., obesity intervention, patient care)
- LLMs and other modern AI tools become increasingly powerful, opening many exciting opportunities for qual data analysis

Interested in Learn More About AI?

[Artificial Intelligence Applications for Health Data Advanced Learning Certificate Program at WashU]

To learn more about the certificate program and participants' testimonials:

https://aicademe.publish.library.wustl.edu/advanced-learning-certificate/

Program Schedule

- Online-only Program via Zoom
- 15 weeks, 3 hours per week
- Weekly project-based assignment
- Early Sep – mid Dec, 2024
- Contact Prof. Ruopeng An: ruopeng@wustl.edu

Program Content (Weeks 1 and 2)

- An overview of AI
- Learn how to code in Python
- Use NumPy and Pandas to do data wrangling
- Use Matplotlib to do data visualization

Program Content (Weeks 3-7: Machine learning)

- End-to-end ML project using Scikit-Learn
- Classification tasks
- Regression tasks
- Model training and validation
- Support vector machines
- Decision trees
- Ensemble methods (e.g., random forest, XGBoost)
- Dimensionality reduction
- Unsupervised learning

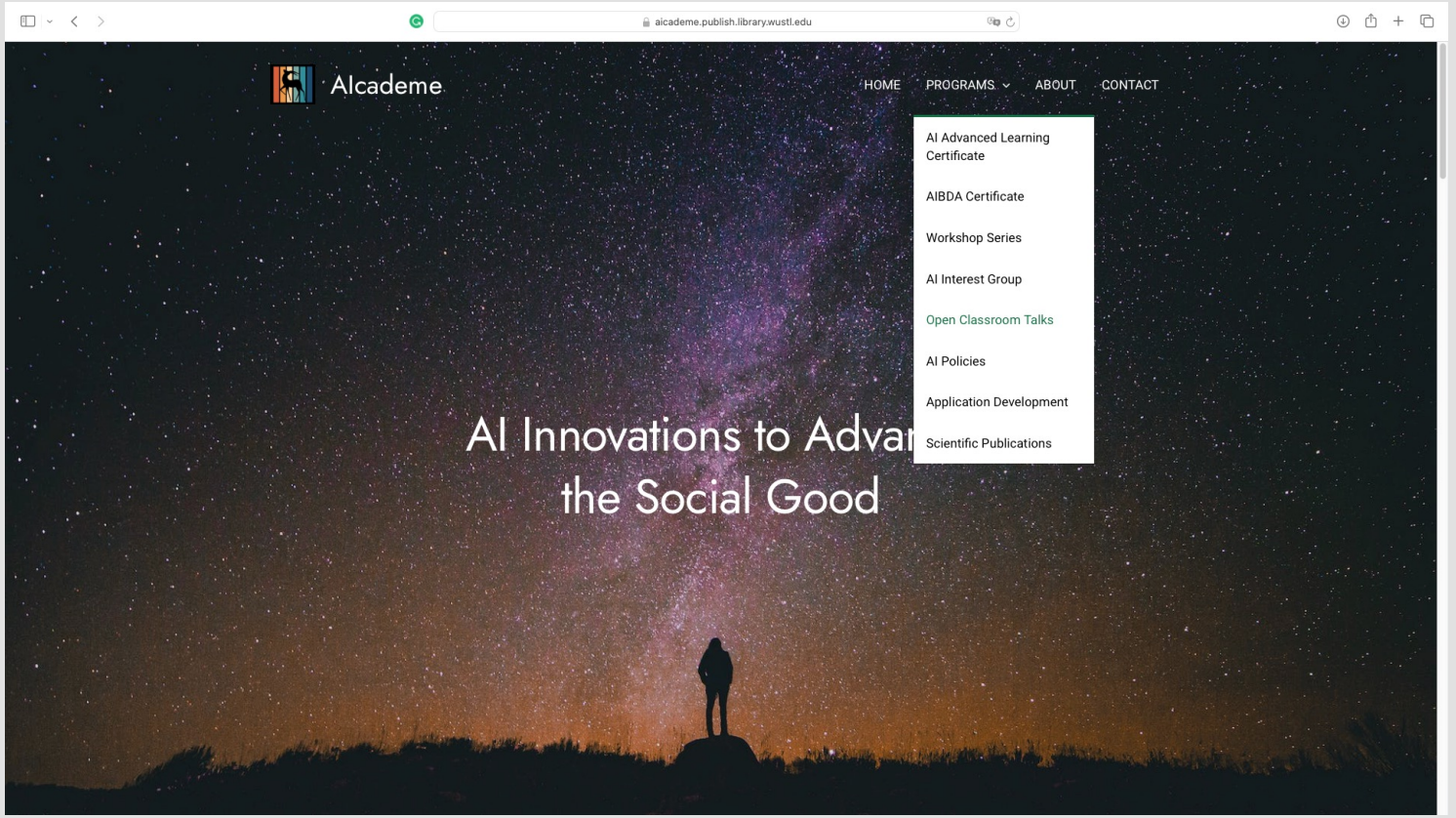
Program Content (Weeks 8-15: Deep learning)

- Neural network basics
- Computer vision: image classification, object detection, and image segmentation
- Time series forecasting
- Recommender system
- Natural language processing: sentiment analysis, text summarization, question-answering, chatbot, translation
- Generative deep learning for image and text generation
- Synthetic data generation

Course Format

- Seamless conjunction of lecture presentations and hands-on lab sessions
- Google Colab Jupyter notebook with free GPU-computing
- All code provided as templates for AI model prototyping
- Weekly assignment focusing on applying AI models to address real-world problems

Today's Recordings & Slides on My Website



https://aicademe.publish.library.wustl.edu