

CAN AI REPLACE HUMAN CODERS FOR SOCIAL SCIENCE CATEGORIZATION?

Basic Idea of the Paper

Utilize open-ended survey responses to:

1. Establish a descriptive framework of how effectively Large Language Models (LLMs) categorize these responses compared to traditional methods.
2. Contrast the accuracy and efficiency of LLMs with human coders in the categorization process.
3. Analyze how the effectiveness of LLMs varies depending on factors such as the complexity of survey questions, language nuances, length of response, and the model's training data.
4. Extract best practices from both LLM and human categorization techniques to propose an optimal hybrid approach.

Conclude with broader implications for the future use of LLMs in qualitative data analysis and the potential for enhancing their categorization capabilities.

Sequence of Analysis

1. **Technique Evaluation:** Employ various categorization techniques with the AI on the open-ended survey data, subsequently analyzing and distinguishing the most effective methods.
2. **Response Analysis:** Delve into the categorized data to pinpoint which responses the AI has accurately categorized and which it struggled with, providing a comprehensive overview of its performance.
3. **Survey Design Recommendations:** Based on the results of the response analysis, formulate actionable recommendations for the future design of open-ended survey questions to optimize categorization efficacy.
4. **Package Development and Publication:** Construct and release a Python package encapsulating all the efficacious techniques identified, streamlining the process for future applications and offering an accessible tool for researchers and developers.

Introduction:

- **Research question:**
 - How effectively do Large Language Models (LLMs) categorize open-ended survey responses compared to traditional methods?
- **Importance and gaps in the current literature.**
- **Current contribution:**
 - Highlight the unique approach of using AI for categorization.
 - Present the aim of extracting best practices.
 - Emphasize the eventual publication of a Python package to implement the findings.
 - * In addition, emphasize a custom fine-tuned model for this specific task.

- **Literature review:**
- **Limitations of current categorization methods:**
 - Include costs of human coders.
 - Include traditional “machine learning” or NLP methods.
- **Exploration of AI in survey response categorization.**
- **Expected contributions and advancements of using LLMs.**

Method:

- Describe the UCNets dataset.
- Overview of the AI models and techniques used.
- Description of the evaluation and validation process.
 - Describing how we determined what “the best” coding is.

Results:

- Analysis of the efficacy of various AI categorization techniques.
- Analysis of which responses are easiest and hardest to categorize.
 - And which categories produce the best outcome.
- Recommendations derived from these results on how to design open-ended survey questions.
- Introduction to the functionalities of the Python package.
 - And the customized fine-tuned model.

Discussion:

- Recap of findings.
- Implications for survey design and data analysis.
- Future of AI in survey data categorization.
- Limitations and potential biases.
 - Costs money but lower than human coders.

Appendix:

- In-depth details about the AI model architectures used.
- Advanced statistics, formulas, and findings.
- Detailed methods for the Python package.
- Case studies.