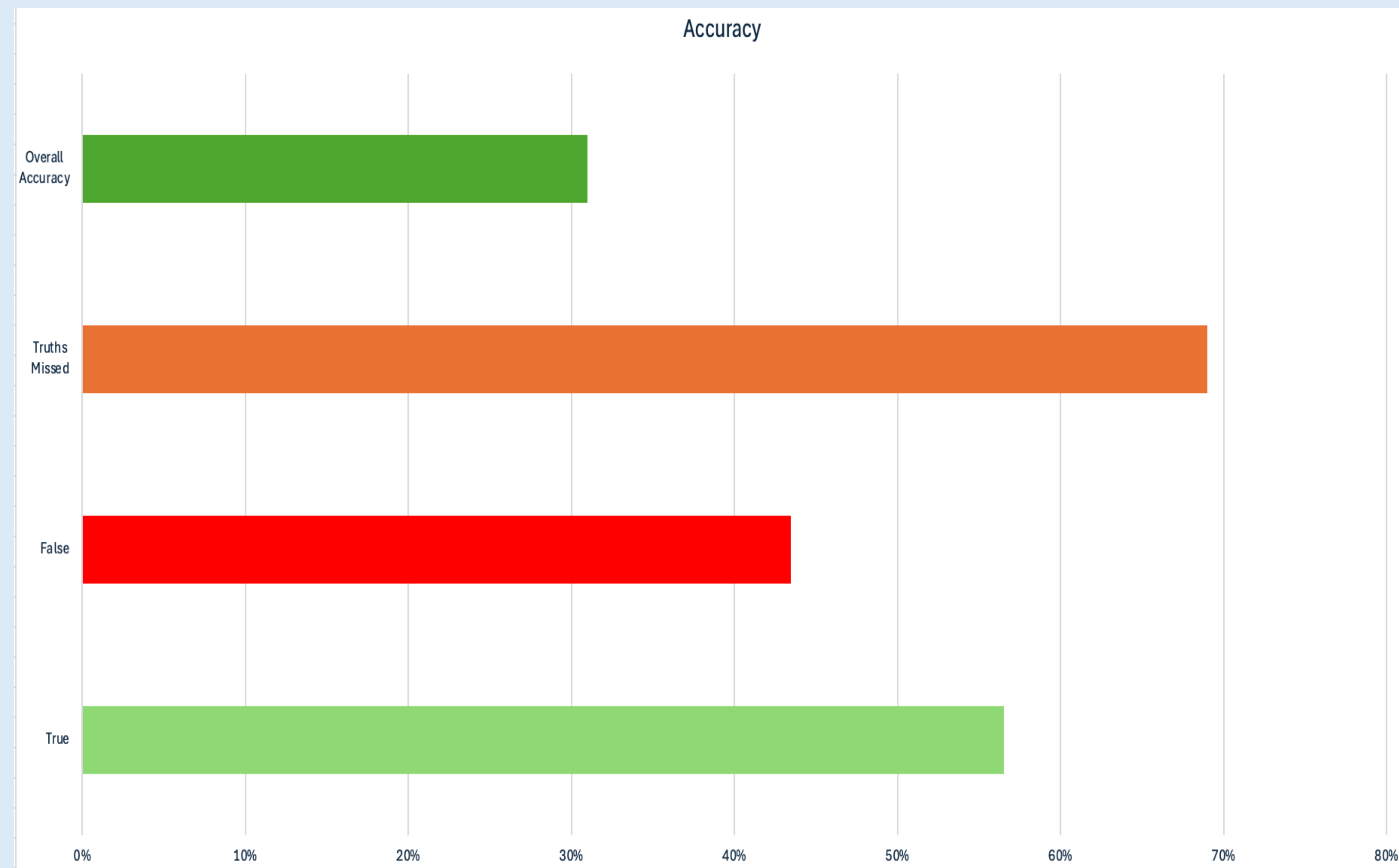


# Open – Ended Survey Response Analysis Using LLMs

Camden Crum and Brian Wilcoxon

## Abstract:

This project pursued the goal of using a large language model (LLM) to effectively sort a representative sample of survey responses. We aimed to analyze the following questions: Under what conditions could using a LLM to analyze qualitative data be effective? How can we best leverage these models in qualitative research? How can we improve upon existing work at FSU using LLMs to classify open-ended survey data? This project is relevant in many different areas of work because AI is useful for making long and tedious processes more efficient. The goal of this is to lessen the number of hours humans spend analyzing data or performing tasks that AI can effectively accomplish within a matter of seconds. By using a Snowflake database and recently introduced features, we were able to create a pipeline that could handle over 2000 survey responses and sort them into categories that fit each response more consistently than if done by a disjoint group of individuals. In doing this, we made it easier to analyze large amounts of data in a shorter amount of time. The next steps in terms of using these results would be to present our work to other organizations on campus and fully integrate our pipeline into survey analysis at FSU. The results and findings of this research suggest that AI can be used to expedite qualitative coding work, giving organizations greater consistency with data analysis while saving time.



## Introduction:

In qualitative research, analyzing open-ended survey responses is a time intensive task that takes a lot of manual labor. It often requires hours of coding and manual categorization of each response. Due to this, there is a growing need for a more efficient but still reliable method to sort these responses. This project explores the capability of using a large language model within a Snowflake SQL environment to classify and organize over 2,000 open-ended survey responses into certain categories. By using databases and AI analysis, we aimed to determine how LLMs can effectively support qualitative research and how they can be implemented to fit any set of survey responses the LLM is given. Our work demonstrates the potential that AI driven analysis has in assisting qualitative while maintaining accuracy. By automating a labor intensive task such as this, the model offers a framework that organizations and universities can adopt to streamline their operations and be able to allocate human effort toward different tasks that are not able to be automated by AI.

## Procedure:

Below is a small sample of the 50+ labels and descriptions that we gave the LLM to sort the survey responses. We gave the model these examples and descriptions and asked it to sort survey responses accordingly based on the descriptions. We started with a specific table of data that we already had manual labels for and compared it to the generated labels to test the accuracy. We then changed the code to make it so it had testing examples and would work with any set of survey responses we gave it.

## Results:

- Overall Accuracy: ~33%
- Truths Missed (In groundtruth but not in predicted): ~68%
- False (In predicted but not groundtruth): ~44%
- True: (In predicted and groundtruth): ~56%
- Accuracy if random: ~2%
- Sample Size: 276

## Acknowledgements:

- Brian Wilcoxon
- Hugh Long

## References

- Arthur, M. M. L. (2021). *21 Qualitative coding*. In *Social Data Analysis: Qualitative and Quantitative Approaches*. Rhode Island College. <https://pressbooks.ric.edu/socialdataanalysis/chapter/qualitative-coding-2/>
- Snowflake, Inc. (n.d.). *SQL command reference*. Snowflake Documentation. <https://docs.snowflake.com/en/sql-reference-commands>
- Snowflake, Inc. (n.d.). *Snowflake key concepts and architecture*. Snowflake Documentation. <https://docs.snowflake.com/en/user-guide/intro-key-concepts>

```
{'label': 'admin responsiveness-positive', 'description': 'administrative responsiveness to student concern or needs'},
{'label': 'admin responsiveness-negative', 'description': 'administrative responsiveness to student concern or needs'},
{'label': 'campus resources general-positive', 'description': 'Resources provided by campus (online/in person) incl. clubs, campus rec, intramural sports, buildings'},
{'label': 'campus resources general-negative', 'description': 'Resources provided by campus (online/in person) incl. clubs, campus rec, intramural sports, buildings'},
{'label': 'personal growth-positive', 'description': 'personal growth relating to time spent/experiences at FSU'},
{'label': 'personal growth-negative', 'description': 'personal growth relating to time spent/experiences at FSU'},
{'label': 'parking-negative', 'description': 'university parking'},
{'label': 'parking-positive', 'description': 'university parking'},
{'label': 'community-positive', 'description': 'of or relating to a sense of community, can be friends, resources, staff or other'},
{'label': 'community-negative', 'description': 'of or relating to a sense of community, can be friends, resources, staff or other'},
{'label': 'belonging-positive', 'description': 'of or relating to a sense of belonging at FSU. Usually mentioned by name or we fall to community'},
{'label': 'belonging-negative', 'description': 'of or relating to a sense of belonging at FSU. Usually mentioned by name or we fall to community'},
{'label': 'career opportunities-positive', 'description': 'of or relating to career opportunities as a result of FSU experience. These DO NOT include things like career fairs. Includes mentioned or perceived career opportunities available to students, actual or perceived'},
{'label': 'career opportunities-negative', 'description': 'of or relating to career opportunities as a result of FSU experience. These DO NOT include things like career fairs. Includes mentioned or perceived career opportunities available to students, actual or perceived'},
{'label': 'career resources-positive', 'description': 'of or relating to career resources afforded to students, career center, career fairs, resume building'},
{'label': 'career resources-negative', 'description': 'of or relating to career resources afforded to students, career center, career fairs, resume building'},
{'label': 'academic advising availability-positive', 'description': 'mention of knowledge/lack of ability to meet with an academic advisor'},
{'label': 'academic advising availability-negative', 'description': 'mention of knowledge/lack of ability to meet with an academic advisor'},
{'label': 'academic advisor performance-positive', 'description': 'academic advisor performance - ability to effectively advise student mentioned - can be any advisor, if an advisor is mentioned nebulously it may be faculty or staff.'},
{'label': 'academic advisor performance-negative', 'description': 'academic advisor performance - ability to effectively advise student mentioned - can be any advisor, if an advisor is mentioned nebulously it may be faculty or staff.'},
```