Developing a CSViewer for Education Application with Natural Language Interactions

Martin Q. Zhao[1][0009-0007-6869-7069], Hari S. Sanku[1], Troy H. Kleger[1], Rui Gong[1] and Qian Wang[2][0000-0002-3303-1183]
[1] Mercer University, Macon GA 31207, USA
[2] Texas A&M University College of Dentistry, Dallas, TX 75246, USA
zhao_mq@mercer.edu

Abstract. The CSViewer for Analysts application provides access to a comprehensive database collected from the Cayo Santiago rhesus monkey colony with 11000 subjects over the past 86 years. Assorted data selection, visualization and analytical features are added to its new version 1.2, and results from mining newly collected osteological measures revealed new skeletal and dental development models. To expose the intended knowledge model of the CS colony to public audiences, especially to science classes at colleges and schools, a CSViewer for Education edition is planned. Supporting queries in plain English is considered beneficial to help students to seek for answers. This paper presents initial experiments with the Claude language model. A dental checkup dataset is used to and queries in plain English are used to explore the dataset through Claude API and the results were integrated with CSViewer to use its charting features to display dental development trend of the CS monkey population. Further development based on natural language interactions enabling utilization of the generative AI features are to be continued.

Keywords: Knowledge Model, Cayo Santiago Rhesus Colony, AI in Education, Querying Database in Natural Language.

1 Introduction

The rhesus monkey (*Macaca mulatta*) colony at Cayo Santiago, Puerto Rico is one of the most useful primate resources in biomedical and anthropological research [1]. This colony is also the source of a rare skeletal collection with associated information about each animal's sex, age, and pedigree (up to eight generations). However, previously it was not easy for researchers to access data of interest from various data sources and match them together.

In 2019, a multi-year collaborative project was funded by an NSF grant and started to build a knowledge model of the CS rhesus population. A comprehensive database that integrates existing genealogical and demographic data with new osteological and pathological data collected from the skeletal collection through this project has been built.

A user-friendly software application, CSViewer for Analysts, is developed to allow researchers to access the entirety of the CS database using a system with menu-based navigation and sophisticated chart generation. Since its initial version V1.0 was

released and assessed by the research and education communities in April 2023, new subversions have been continuously updated and enriched to include more data managing and processing, analytic , and visualization functionalities.

Based on the integrated database, holistic perspectives of the CS rhesus colony have been revealed regarding matrilineal/patrilineal family lineage, headcount and reproduction dynamics, interactions among families and social groups. These analytical results have been included in CSViewer V1.1.1 as static charts, which can be used as a reference for researchers when selecting project-specific datasets using the newly added features. Essential analytical tasks can be performed on these datasets within CSViewer or the datasets can be saved as files and fed into other tools of researchers' choices for further analysis.

Several studies using the newly collected skeletal and dental datasets have provided new insights of the development trends of the CS rhesus monkeys [2-5]. Skeletal development trend lines can be best fitted with a stepwise regression model using two straight line segments. [6] Comparisons with other commonly used models showed lower root of mean squared errors (RMSE & RMSLE) with the stepwise model [7]. Guatelli-Steinburg et al. [3] presented relationships between teeth wear with various body measure(measures), including sex, body mass, and reproduction status. The same dental checkup dataset used in [3] could be further used to study long-term dental development trends for the CS rhesus monkeys going forward.

To expose the valuable CS monkey colony data and the derived knowledge model to a broader audience, especially to science educators and students in colleges and schools, it is considered beneficial to develop a CSViewer for Education edition. Sensitive information regarding monkeys' identity will be omitted in the for Education edition, which needs to be kept private to be compliant with HIPAA policies since monkeys have been treated as human patients.

This for Education edition can incorporate with recent generative AI features to allow for querying the database using a natural language (like English). Generative AI technology is poised to provide an innovative tool for promoting teaching and learning experience and excellence [8-12], and it is our intention to put this conceptual and technological advancement into practice. Initial experiments of this kind of effort will be discussed in this paper.

2 Methodology

2.1 Recent Development in AI

Recent advancements in generative AI have revolutionized natural language processing (NLP), enabling human-like text generation and comprehension. Large-scale transformer-based models such as Claude from Anthropic leverage billions of parameters provide meaningful responses to user queries. The ability of these models to process complex queries and generate structured responses makes them valuable for educational applications. By integrating AI-driven natural language interactions, students can engage with datasets more intuitively, improving their learning experience.

2.2 Workflow Using the Claude Language Model

Claude is a state-of-the-art AI model designed to understand and generate human-like text. It is optimized for natural language comprehension and has been trained on a vast corpus, making it capable of interpreting plain English queries and converting them into structured database queries. This model will be used as the backend AI component in CSViewer for Education to facilitate natural language interactions. Queries are sent to the Claude API, which processes the input and returns relevant data points from the CS rhesus colony dataset. The responses are then formatted and visualized using CSViewer.

To enable seamless interaction between users and the dataset, we developed a Flask-based API in Python to act as a bridge between the Claude language model and the CSViewer Java application. The integration followed these steps (as shown in Fig. 1):

Flask API Development:
We implemented a lightweight REST API using Flask in Python [13]. The API accepts natural language queries as HTTP POST requests. It processes the input, sends it to Claude's API, and retrieves the AI-generated response. The processed response is returned as a structured JSON object, as seen in step 3 of the workflow chart in Fig. 1a.

Java-Based CSViewer API Call:
The CSViewer application, written in Java, sends HTTP requests to the Flask API. The response from the API is parsed and formatted within CSViewer, aligning with step 5 in Fig. 1a. The structured data is then visualized using the JFreeChart library [14] to generate trend charts, histograms, and line graphs, as illustrated in step 6.

2.3 Integration with the CSViewer App

This Flask-Java integration allows real-time AI-powered interactions, making CSViewer more accessible for students without SQL or programming knowledge. The AI-assisted learning experience enhances data exploration, enabling students to identify trends and patterns in scientific datasets effectively.

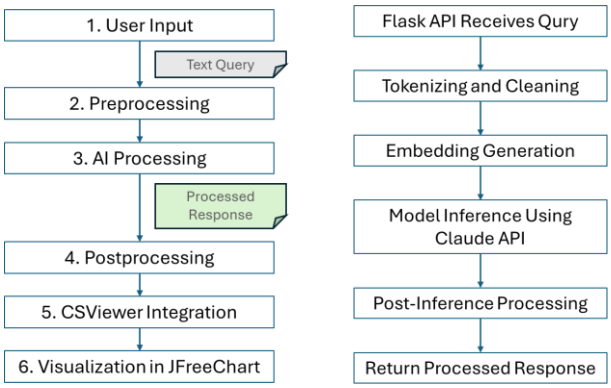
Data returned from Claude API calls can be processed and fed into CSViewer for Education to provide answers to the user. Details about responses in natural language will be implemented in later stages. In this paper, we will describe how the data can be used to generate charts using the visualization capability in CSViewer. The results will be presented in section 3.3.

3 Results & Discussions

3.1 Template for sending queries in English to Claude API

To facilitate seamless communication between users and the Claude API, we established a structured query processing workflow. This workflow enables students and

researchers to input queries in plain English, which are then translated into structured data outputs by the AI model.



a. Overall workflow b. Claude workflow

Fig. 1. Illustrations of Query processing and integration with CSViewer.

User query processing can be carried out in the following steps:

1. **User Query Submission:**
   - Users enter queries in natural language through the CSViewer interface.
   - Example: *"Show me the trend of dental checkups over the last decade."*
2. **Flask API Processing:**
   - The query is sent to the Flask-based API, which forwards it to the Claude API.
   - Claude interprets the query and retrieves the most relevant dataset insights.
3. **Response Formatting & Storage:**
   - The response from Claude is formatted into structured JSON data.
   - The JSON output is stored and processed for visualization in CSViewer.

This structured approach enables natural language-driven data exploration, allowing students to retrieve complex dataset insights without prior SQL or programming knowledge.

3.2 Generating a Heatmap in CSViewer

The integration between Claude AI and CSViewer was implemented using a Flask API in Python and Java-based visualization tools. Data received by the CSViewer app can be stored as a list of tuples with Age (of animal when examined), Status (of dental development stage), and Count (of animals in that category). In the dataset used for this pilot study, there are 684 animals with ages ranging between 0 and 28. Eight levels are identified as representative dental development stages, as shown in Table 1.

Table 1. Categories for CS Rhesus Monkey Dental Development Status.

Level	Abbreviation	Description
0	0-N	No teeth
1	1-S	Some deciduous teeth
2	2-L	Full deciduous teeth – except certain teeth
3	3-D	Full deciduous teeth
4	4-G	Full deciduous teeth – with certain permanents
5	5-M	Full permanent teeth – missing certain permanents
6	6-P	Full permanent teeth
7	7-B	Full permanent teeth – with certain loss/broken

The CSViewer app uses JFreeChart API [14] to create various charts. To generate a heatmap as shown in Fig. 2, the headcounts by age by status code are stored with a DefaultXYZDataset. Specifically, Age and Status code are assigned to x and y series, while headcount is assigned to the z series. A SpectrumPaintScale class was created to customize the PaintScale interface in JFreeChart API.

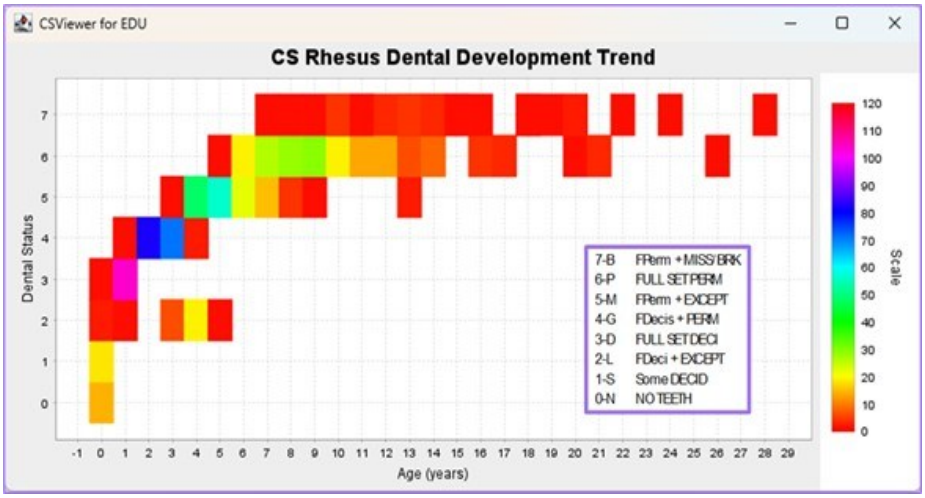


Fig. 2. Dental development trends from the CS rhesus dental checkup data.

3.3 Discussions

The integration of Claude AI with CSViewer demonstrates the power of natural language-driven data interaction. By leveraging AI-generated insights, students can:

- Efficiently explore datasets without SQL or programming skills.
- Quickly generate visual trends and analytics based on historical data.
- Enhance learning experiences by transforming data queries into structured charts.

This AI-powered data exploration enhances accessibility and usability for students, researchers, and educators. Future work may involve optimizing query refinement and expanding charting capabilities for even richer data visualization.

4 Conclusions & Future Work

Capabilities in CSViewer for Analysts have been introduced and the intention of its "for Education" edition described. With this CSViewer for Education edition, students in science classes as well as public can use a natural language like English to explore the valuable CS database and learn knowledge about primates. Initial experiments have shown the potential of using the Claude language model to transform queries in English into SQL for retrieving information from the CS rhesus database. Charts that help visualize the results can be generated with existing modules in CSViewer. Future development can make more sophisticated interactions using natural language as commonly available in current generative AI systems.

Funding Sources

This project is supported by NSF grant to M.Q.Z and Q.W. (NSF 1926402, 1926601). The CPRC Skeletal Collection has been supported by National Institutes of Health NIH contracts NIH 5P40OD012217.

References

1. Francis, G., Wang, Q. History of Health at Cayo Santiago – An investigation of Environmental and Genetic Influences on the skeletal remains of the introduced Rhesus Macaque (Macaca mulatta) colony. In: American Journal of Primatology. 87: 87:e23722. DOI: 10.1002/ajp.23722 (2025).
2. Zhao, M. Q., Maldonado, G., Kensler, T. B., Kohn, L. A.P., Guatelli-Steinberg, D., Wang Q.: Conceptual Design and Prototyping for a Primate Health History Model. In: Advances in Computer Vision and Computational Biology, Springer: New York, p.511-522 (2021).
3. Guatelli-Steinberg, D., Watson, J., Samuel, L., Showalter, E., Lerner, F., Dixon, E., Kensler, T. B., Francis, G., Maldonado, E., Kohn, L. A.P., Zhao, M. Q., Wang, Q. Revisiting linear enamel hypoplasia in Cayo Santiago rhesus macaques (Macaca mulatta): How a stress marker relates to environment and maternal lineage in a rare rhesus colony with known life history. American Journal of Primatology. 87:e23692. DOI: 10.1002/ajp.23692 (2025).
4. Zhao, M. Q., Novak, C. T., Gong, R., Kaur, M., Francis, G., Kensler, T. B., Guatelli-Steinberg, D., Maldonado, E., Kohn, L. A.P., Wang Q. CSViewer for Analysts: I. Building an Integrative Database and Knowledge Model for the Cayo Santiago Rhesus Macaque Colony and its Derived Skeletal Collections. In: American Journal of Primatology. (Accepted 2025).
5. Zhao, M. Q., Gong, R., Kaur, M., Kundu, S., Francis, G., Kensler, T. B., Guatelli-Steinberg, D., Maldonado, E., Kohn, L. A.P., Wang Q. CSViewer for Analysts: II. Analytic Tools and Visualization towards Founder Lineages, Social Groups and Reproduction Dynamics of the

Cayo Santiago Rhesus Macaque Colony. In: American Journal of Primatology. (Accepted 2025).
6. Zhao, M. Q., Novak, C., Widener, E. R., Patel, R. A., Gong, R., Francis, G, Wang, Q.: Enhancing Data Analytics and Visualization Support in CSViewer for Analysts - Version 1.1: Access to an Integrative Database and Knowledge Model of Cayo Santiago Rhesus Macaques. In: Journal of Data Science and Intelligent Systems, Vol. 00(00) 1–9 DOI:10.47852/bonviewJDSIS42023383 (2024).
7. Zhao M. Q., Gong, R., Wang, Q. Skeletal Growth Modeling for the Cayo Santiago Rhesus Macaques with Stepwise Regression Supported in CSViewer for Analysts. In: Proceeding of 13th Computing Conference, London (2025).
8. Khan, S. Brave New Words: How AI Will Revolutionize Education (and Why That's a Good Thing). New York City: Viking Press (2024).
9. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. *Attention is all you need.* Advances in neural information processing systems, 30. (2017).
10. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). *Language models are few-shot learners.* NeurIPS 2020.
11. Luckin, R. Machine Learning and Human Intelligence: The Future of Education for the 21st Century. UCL Press. (2018).
12. Woolf, B. P. Building Intelligent Interactive Tutors: Student-centered Strategies for Revolutionizing E-learning. Morgan Kaufmann (2010).
13. Grinberg, M. (2018). *Flask Web Development: Developing Web Applications with Python.* O'Reilly Media.
14. Sandner, U. *JFreeChart: A Developer Guide.* Object Refinery Limited (2008).