

Using Biterm Topic Modeling to Explore Gender Differences in Secondary Students' Wonderings about AI Concepts

Doreen Mushi, North Carolina State University, drmushi@ncsu.edu

Jie Chao, Concord Consortium, jchao@concord.org

Shiyan Jiang, North Carolina State University, sjiang24@ncsu.edu

Rebecca Ellis, Concord Consortium, rellis@concord.org

Carolyn P. Rosé, Carnegie Mellon University, cprose@cs.cmu.edu

Kenia Wiedemann, Concord Consortium, kwiedemann@concord.org

Abstract: The study reveals gender-based differences in how secondary students express curiosity in an AI curriculum. The study used biterm topic modeling to explore questions posed by male and female learners (n=239) in Grades 8-12. The findings revealed that keywords related to model accuracy and prediction such as *Words* and *Weights* were prevalent in both groups. Key variations were observed in technical terms, such as *Machines*. The discussions highlight the need for socially relevant AI curriculums.

Introduction

The rapid growth of Artificial Intelligence (AI) as a field has imposed a critical need for delivering educational programs that would foster the development of knowledge and skills in the area. Empirical research has indicated that designing and developing socially relevant and student-centered AI curriculums comes from understanding how students learn through the contextualization of their needs (Alvarez et al., 2023; Eguchi et al., 2021). This study is framed by the formative assessment approach, which also has foundations in design-research theory (Sandoval, 2014). It is fundamental to assess student learning and development of interests throughout an instructional unit and adapt subsequent instruction to align with the student's attainment and interests. Students' responses reveal information about their thought processes, interests, identities, and conceptions about the learning topics. The study's main purpose is to identify themes in questions that Grades 8-12 students ask as they learn AI concepts and how these themes differ across female and male gender groups. The following research questions are addressed in this study: RQ1: What keywords and their associations were present in students' responses to a prompt about text classification models? RQ2: How do these keywords and their associations differ across gender groups?

Methods

This study is based on an AI curriculum module designed for secondary students to learn AI concepts in their English language Arts (ELA). The participants included female (n=133) and male (n=106) students in Grades 8-12 who took part in a 7-hour module designed to explore AI concepts and practices in the context of text classification models. The data collection involved extracting students' responses from Activity 2's open-ended question labeled "What questions or wonderings do you have about the model?". The study deployed Biterm Topic Modeling (BTM) (Yan et al., 2013) to predict emerging themes in the students' questions from a corpus of open-ended responses. For this case, we went for clustering occurrences of nouns "NOUN", "PROPN" and "ADJ" with hyperparameter K set at 4.

Results

The results of the analysis provide valuable insights into the context and scope of responses that both male and female students posed after completing a series of activities on text classification models. In the frequency distribution charts (see Figure 1), it can be observed that keywords such as *Words*, *Model*, *Negative*, and *Positive* had the highest occurrence numbers in the analyzed responses from both student groups. The questions posed by male students were more technical and included field-specific terms such as *Machines* and *Accurate*. The results of female students indicate that the majority of their questions were interested in *understanding* how the model worked and how it assigned weights to specific words. The four-topic BTM model validated the observations from the keywords and showed an association of common words in both male and female questions (see Figure 2). For instance, in the green theme, male responses had associated the model with classification, text, and accuracy, while in female responses, the model was associated with words. Another unique observation is that female responses used more verbs (such as *pick*, *improve*, and *determine*) compared to male responses that used human action words.

Figure 1
Frequent Words From Male and Female Students' Responses

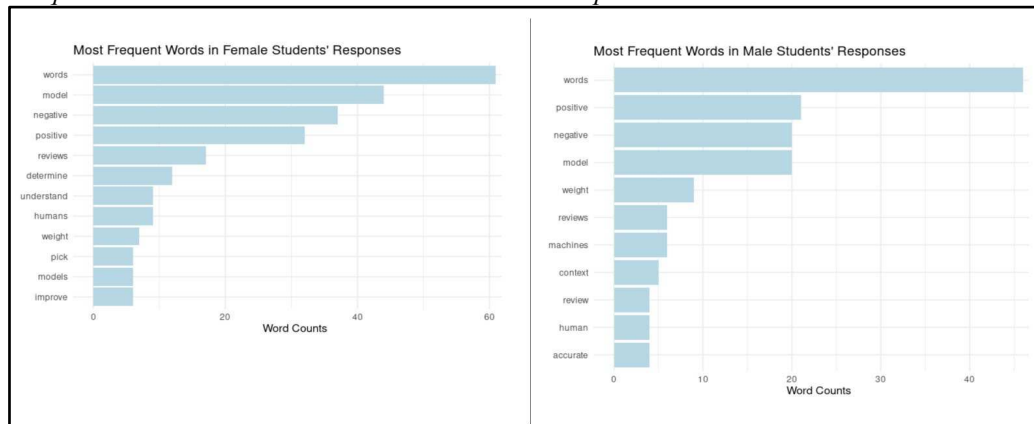
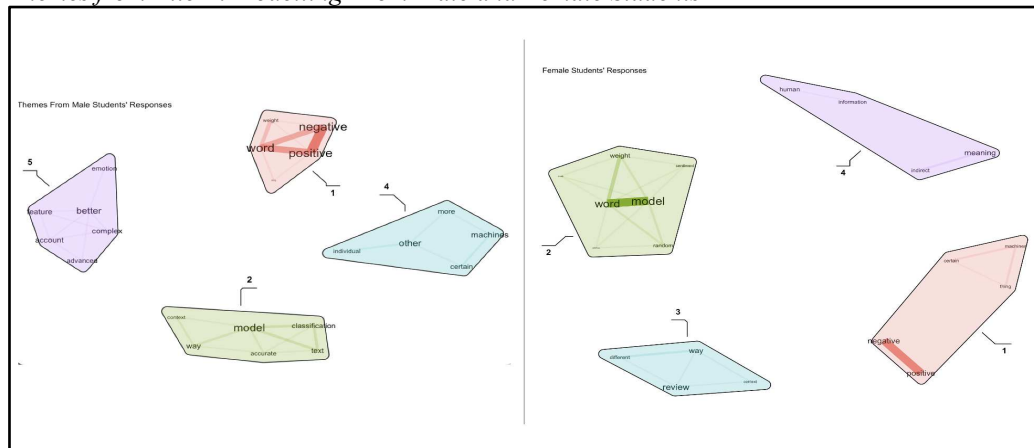


Figure 2
Themes from Biterm Modelling From Male and Female Students



Conclusion

AI Education for K-12 presents ample opportunities for young learners to nurture their curiosity and develop their skills in AI concepts. The identified themes can inform the efficiency of the AI module's cohesiveness and guide the future design of advanced and socially relevant AI curriculums.

References

- Alvarez, L., Gransbury, I., Cateté, V., Barnes, T., Ledécz, Á., & Grover, S. (2022, June). A Socially Relevant Focused AI Curriculum Designed for Female High School Students. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 36, No. 11, pp. 12698-12705).
- Eguchi, A., Okada, H., & Muto, Y. (2021). Contextualizing AI education for K-12 students to enhance their learning of AI literacy through culturally responsive approaches. *KI-Künstliche Intelligenz*, 35(2), 153-161.
- Sandoval, W. (2014). Conjecture mapping: An approach to systematic educational design research. *Journal of the learning sciences*, 23(1), 18-36.
- Yan, X., Guo, J., Lan, Y., & Cheng, X. (2013, May). A biterm topic model for short texts. In *Proceedings of the 22nd international conference on World Wide Web* (pp. 1445-1456).

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. DRL-1949110. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.