CULTIVATE – Contextual AI for Understanding Cultural Themes and Assets in STEM

Khalid Mehtab Khan (kkhan@sfsu.edu), Department Of Computer Science, San Francisco State University

Faculty Supervisors: Dr. Anagha Kulkarni (ak@sfsu.edu), Department Of Computer Science, San Francisco State University

Summary of Research

STEM retention remains a significant challenge, particularly for students from underrepresented backgrounds. Many of these students possess valuable cultural strengths—such as resilience, community support, and aspirations—that shape their academic journeys. However, these assets are often overlooked in traditional educational support systems, leading to gaps in engagement and success. CULTIVATE addresses this issue by leveraging AI to learn from student experiences, recognize their cultural strengths, and provide meaningful insights that can help institutions and educators offer better support.

Our research aims to develop a model-driven understanding of Cultural Capital Themes (CCTs) in student narratives. Unlike rigid, predefined categories, CCTs are dynamic social constructs, evolving across different contexts and communities. Using Natural Language Processing (NLP) and AI, CULTIVATE identifies, structures, and refines these themes in a scalable, unbiased, and efficient manner.

We are working with state-of-the-art language models to analyze how students express their experiences, struggles, and achievements in STEM fields. Rather than depending on subjective human interpretations or classical machine learning methods, CULTIVATE's AI models learn directly from student essays, identifying patterns that reveal how cultural experiences shape STEM persistence and success. By capturing the depth and complexity of student narratives, the model provides a richer, more nuanced perspective on how Cultural Capital impacts STEM education.

This research has a twofold impact:

1. Helping institutions recognize and validate students' cultural strengths—ensuring that underrepresented voices are acknowledged and valued in STEM education.
2. Enabling educators to provide targeted interventions—allowing for the development of more inclusive mentorship, support structures, and resources tailored to students' unique backgrounds.

By learning from students' experiences and proactively addressing their challenges, CULTIVATE can foster a more inclusive, data-driven, and student-centered STEM environment.

CULTIVATE's Approach & CCT Identification

Our research focuses on 11 key Cultural Capital Themes (CCTs) that influence STEM retention among underrepresented students. These themes are social constructs—fluid, context-dependent, and interpreted differently across communities. While previous research, including Yosso (2005) and others, has explored aspects of Cultural Capital, our study compiles CCTs from multiple sources to systematically define them in an unbiased, scalable way.

The 11 Cultural Capital Themes (CCTs) we focus on are:

- Aspirational
- Navigational
- Perseverant
- Familial
- First-Generation
- Spiritual
- Attainment
- Community Consciousness
- Resistance
- Filial Piety
- Social

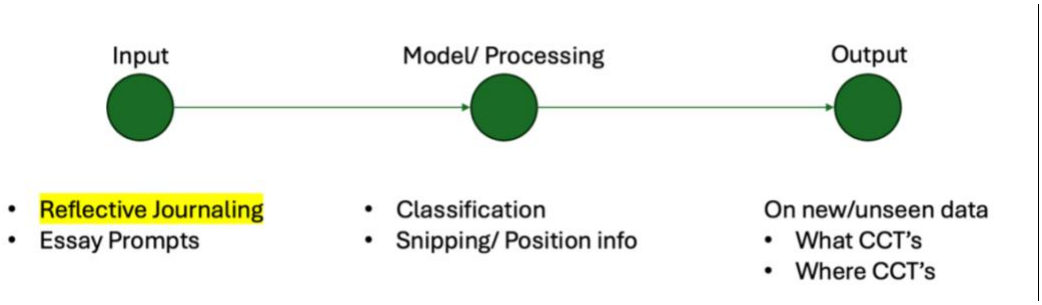

Figure 1: CULTIVATE's processing pipeline utilizing the Google/T5 transformer model for Cultural Capital Theme identification (Raffel et al., 2020). The pipeline processes reflective journaling and student essays as input, fine-tunes the model to learn contextual patterns in Cultural Capital expressions, and generates outputs for unseen and new STEM prompts, ensuring robust and unbiased theme recognition.

Model Training and Cultural Capital Identification

```
prompt_text_attainment = (
    "Definition of Attainment Cultural Capital:\n"
    "Attainment cultural capital refers to the recognition, pursuit, and achievement of academic or professional s
    "It includes expressions of educational goals, career aspirations, perseverance in academics, and the value p
    "It can also include references to hard work, dedication, overcoming academic challenges, and striving for exc
    "Additionally, it reflects a mindset of progress, goal-setting, and structured efforts toward self-improvement

    "Essay:\n"
    f"{joined_essay}\n\n"

    "Task:\n"
    "Review the essay and identify every sentence that reflects attainment cultural capital as defined above. "
    "For each qualifying sentence, output it as a bullet point (starting with a dash '-' followed by a space). "
    "If no sentence qualifies, simply output the word 'None'.\n\n"

    "Response:"
)
```

Figure 2: AI model prompting process exemplified through Attainment Capital identification.

The process demonstrates how Large Language Models (LLMs) are guided to recognize and classify cultural capital expressions in student narratives.

- The Google Sentence-T5 model (Raffel et al., 2020) is fine-tuned on annotated STEM student essays to first understand and then detect cultural capital signal in text.
- The model learns meanings beyond keywords, ensuring robust semantic understanding of CCTs.

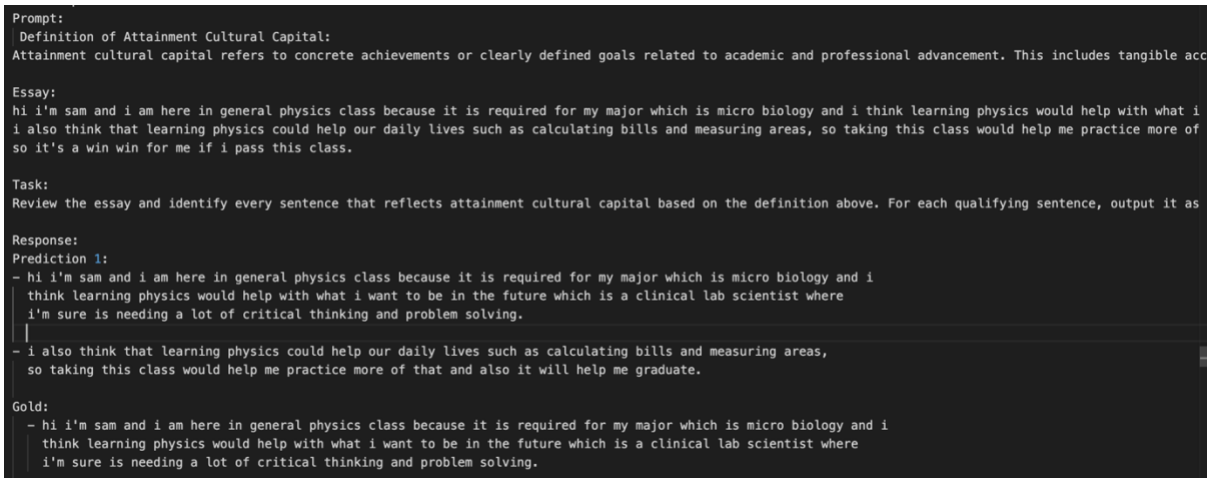

Figure 3: Model Output for Attainment Capital identification

- The AI model correctly identifies Cultural Capital sentences in student essays.
- In some cases, it outperforms human annotators, detecting sentences missed in the human-labeled dataset (gold standard).
- CULTIVATE's AI actively learns how Cultural Capital is expressed in student writing rather than matching predefined labels.

AI Learning & STEM Impact

How the Model Internally Understands Cultural Capital Themes (CCTs)

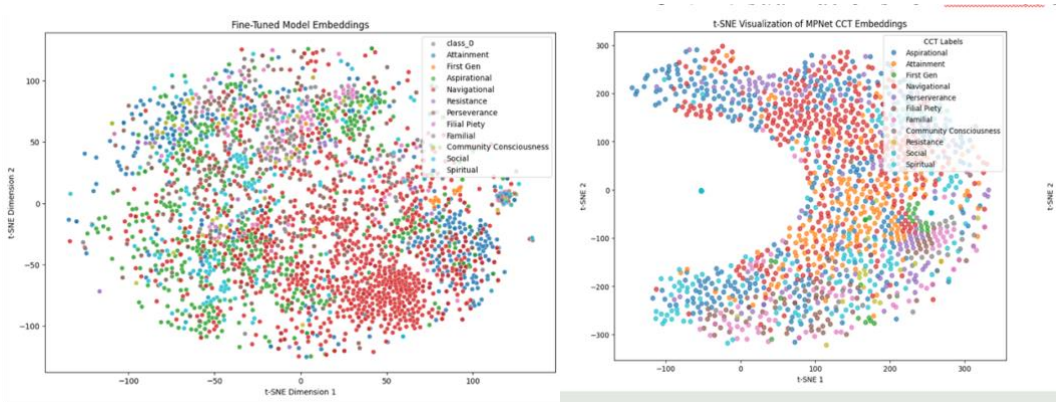

Figure 4: Visualizations of CCT Embeddings using tSNE method (van der Maaten & Hinton, 2008)

- The visualization in Figure 4 provides a peek into How the Model Internally Understands Cultural Capital Themes (CCTs).

- Overlapping clusters (e.g., Aspirational and Attainment Capital) reveal how themes share common linguistic patterns.

How Educators Can Use CULTIVATE to Support STEM Retention

Educators can leverage CULTIVATE's AI-learned representations to better understand student experiences:

- **Early Identification of At-Risk Students** - AI can flag students needing mentorship, peer support, or institutional guidance.
- **Strength-Based Mentorship Programs** - If a student exhibits strong Aspirational or Perseverant Capital, institutions can align research opportunities, internships, or scholarships accordingly.
- **Curriculum Development** - Reflective journaling assignments can help students identify and articulate their cultural strengths, fostering engagement and motivation.

Appendix: Bibliography & References

- Nayak, K., et al. (2020). Using text analytics on reflective journaling to identify cultural capitals for STEM students. 19th IEEE International Conference on Machine Learning & Applications(ICMLA), Miami, FL, USA,pp. 797-804. https://doi.org/10.1109/ICMLA51294.2020.00130
- Yosso, T. J. (2005). Whose culture has capital? A critical race theory discussion of community cultural wealth. Race Ethnicity and Education, 8(1), 69-91. https://doi.org/10.1080/1361332052000341006
- Tran, K., Barrera, A. M., Coble, K., Arreguin, M., Harris, M., Macha-Lopez, A., Perez, M., & Eroy-Reveles, A. (2022). Cultivating cultural capitals in introductory algebra-based physics through reflective journaling. Physical Review Physics Education Research, 18(2), 020139. https://doi.org/10.1103/PhysRevPhysEducRes.18.020139
- Acevedo, N., & Solorzano, D. G. (2021). An overview of community cultural wealth: Toward a protective factor against racism. Urban Education. https://doi.org/10.1177/00420859211016531
- Estrada, M., Eroy-Reveles, A., & Matsui, J. (2018). The influence of affirming kindness and community on broadening participation in STEM career pathways. Social Issues and Policy Review, 12(1), 258-297. https://doi.org/10.1111/sipr.12046
- Dunlap, L., Mandal, K., Darrell, T., Steinhardt, J., & Gonzalez, J. E. (2024). VibeCheck: Discover and quantify qualitative differences in large language models. arXiv Preprint. https://arxiv.org/abs/2410.12851
- Zhang, H., Wu, C., Xie, J., Kim, C., & Carroll, J. M. (2023). QualiGPT: GPT as an easy-to-use tool for qualitative coding. arXiv Preprint. https://arxiv.org/abs/2310.07061
- Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., ... & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research, 21*(140), 1-67. https://arxiv.org/abs/1910.10683
- t-SNE method: van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. Journal of Machine Learning Research, 9, 2579-2605. https://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf
- He, P., Liu, X., Gao, J., & Chen, W. (2021). He, P., Liu, X., Gao, J., & Chen, W. (2021). DeBERTa: Decoding-Enhanced BERT with Disentangled Attention. International Conference on Learning Representations (ICLR). https://arxiv.org/abs/2006.03654