

Tools for Reflection: Creativity-Aware AI for High-Stakes Student Writing

Jinsook Lee
Cornell University

1. Motivation: Why I care about this challenge

For many high school students, college application essays are one of the few places where they can differentiate themselves beyond grades and test scores. In a high stakes decision, this writing becomes evidence of judgment, self awareness, and the ability to make meaning from experience. Creativity matters because it is how applicants turn ordinary moments into distinctive narratives. A surprising connection, a fresh framing, a specific voice, and a coherent perspective can make a student legible as a thinker rather than only as a resume. When essays become generic, students lose one of the only channels where individuality can be recognized, and this loss is especially consequential for students whose achievements are harder to signal through conventional metrics.

The rapid adoption of large language models complicates this moment. While LLMs provide fast and polished feedback, current systems often push writers toward typical phrasing and structure. A growing body of work shows that models tend to produce generic and homogenized outputs (Moon et al., 2025; Liang et al., 2024; Chakrabarty et al., 2024), nudging users toward standardized expression precisely when distinctive writing is most valuable (Huang, 2024; Beck & Godley, 2024). Other studies document consequential biases, including the flattening or penalizing of marginalized identities (Wang et al., 2025; Hofmann et al., 2024) and in group favoritism in human AI interactions (Hu et al., 2025). For many students, LLMs may be the most accessible and responsive feedback they receive. Those who rely heavily on AI generated content without additional human guidance, often because of resource constraints, may come to see the AI version as inherently superior even when it does not reflect their identity or lived experience. In trying to sound more polished, they risk blending into a homogenized style and losing an opportunity to think expansively about who they are becoming.

At its core, this is a problem of creativity and self understanding. Strong application essays require students to surface lived experiences, reflect critically, and shape them into coherent narratives. This raises questions that are central to this challenge. How can AI help students make their stories more creative and compelling? How can it create more moments of reflection, not only for a stronger application, but also for their growth as thoughtful and ethical human beings? What would it mean for the application process itself to become a life changing moment of self reflection regardless of admissions outcomes?

What makes this problem newly tractable is data and experimental leverage. I already work with large scale corpora of human authored and LLM authored admissions essays, including systematic generations across models and prompts that vary identity cues and background context. I believe this enables a rare triangulation. First, measuring where models collapse variation relative to human writing. Second, testing how creativity signals such as novelty, juxtaposition, narrative coherence, and voice shift under different prompting and assistance regimes. Third, examining whether these shifts interact with socioeconomic context, including whether AI assistance changes whose writing is perceived as authentic, compelling, or admissions worthy. Rather than treating creativity as an abstract ideal, this allows us to study it as a measurable signal with downstream consequences.

Throughout the fellowship, I aim to investigate what current AI technologies newly enable and to translate those capabilities into designs that expand rather than constrain human creativity. Specifically, I want to design AI writing support that operates near conceptual boundaries. Instead of offering only polished suggestions that funnel students toward a narrow template, I hope to build AI collaborators that help them uncover surprising and personally meaningful combinations of ideas while preserving voice, specificity, and social context.

2. Relevant expertise, qualifications, and perspective

My PhD training in Cornell Information Science is highly interdisciplinary. My research sits at the intersection of education, computational social science, and algorithmic fairness. I see this as an ideal vantage point for studying creativity in AI mediated writing. Broadly, I examine how AI systems reshape educational processes and equity with a particular emphasis on college admissions and other high stakes decision making contexts.

I have published in education focused venues such as AIED, LAK, and the British Journal of Educational Technology, in equity focused venues such as EAAMO, and in language model focused outlets such as COLM and the Journal of Big Data. My work has been supported by a Cornell Computational Social Science seed grant and the Siegel Public Interest PhD Impact Fellowship. I also serve as a reviewer for FAccT, COLM, the Journal of Learning Analytics, the International Journal of Artificial Intelligence in Education, and Frontiers in Psychiatry.

In *The Life Cycle of Large Language Models in Education: A Framework for Understanding Sources of Bias* (BJET), I propose a lifecycle perspective on how bias enters and persists in educational AI systems from data collection through deployment and feedback. In *Large Language Models, Social Demography, and Hegemony: Comparing Authorship in Human and Synthetic Text* (Journal of Big Data), my collaborators and I show how LLM generated text reflects and amplifies hegemonic demographic language patterns, shaping who models sound like.

Building on this work, my recent projects focus directly on college admissions essays. In *Poor Alignment and Steerability of Large Language Models* (COLM workshop), I show that LLMs struggle to adapt writing style to diverse applicant identities even under rich prompting, and that attempts at steering can worsen homogenization. In ongoing work titled *Algorithmic Voices in College Admissions? How Large Language Models Reshape Admission Chances across Socioeconomic Groups*, I examine how LLM written essays may alter perceived merit across socioeconomic groups and reveal how models implicitly encode advantage.

Across these projects, my orientation is interdisciplinary and critically engaged with mainstream model design. I combine NLP and generative modeling with insights from education research, fairness and equity, and social demography in order to ask not only what models output, but also for whom and with what consequences. I see the purpose of education as cultivating independent and holistic thinkers, and I approach LLMs through that lens. I treat them not only as tools for efficiency, but also as socio technical systems that can narrow or expand the conceptual and creative boundaries within which young people learn to tell their stories. This perspective aligns closely with the ideal collaborator profile for this challenge. I bring hands on experience with generative models, comfort publishing across AI and education venues, and a strong interest in using AI to explore conceptual boundaries and foster originality while consistently questioning whose innovation these systems ultimately serve.

3. Proposed research directions within the challenge

3.1 Mapping where AI collapses creativity

I will build an empirical map of where current models fail to explore conceptual boundaries in admissions essays. Using corpora of human authored and LLM authored essays, I will represent texts in a multi view embedding space that captures topic, rhetorical structure, lexical diversity, complexity, and narrative coherence. I will quantify homogenization by measuring concentration around typical modes of writing and compare this to where human essays show coherent novelty. I will identify regions where students naturally connect distant concepts or identities while still producing grounded and meaningful narratives, and I will compare these regions to where models most often generate text. The outcome will be a creativity boundary map that highlights where variation collapses relative to where human creativity thrives and will identify conceptual regions where boundary pushing support is most needed.

3.2 Learning a boundary sensitive creativity objective

Building on the creativity boundary map, I will develop training objectives and selection strategies that increase coherent novelty while discouraging templated writing. One approach will use a discriminator that detects generic admissions style templates and overused rhetorical patterns. This discriminator can

be used for reranking or as a training signal so that the model avoids highly typical phrasing while remaining coherent and grounded. A second approach will use metric learning to represent coherent novelty directly, learning an embedding where distance reflects meaningful juxtaposition rather than topic drift. This representation can support contrastive selection, reward modeling, or reranking that favors outputs that are novel but still narrative consistent. A third approach will focus on voice preservation through contrastive fine tuning, anchoring outputs to a student's draft by matching stylometric and rhetorical features such as rhythm, sentence length distribution, idioms, and code switching markers while varying structure or emphasis. Across approaches, the goal is not to maximize novelty in the abstract, but to increase originality without sacrificing truthfulness, specificity, and voice.

3.3 Orchestrated boundary pushing writing support

I will prototype AI writing support that emphasizes reflective moments and boundary exploration rather than full automation. I will explore an orchestration approach in which specialized components collaborate. A grounding component will extract a student's concrete experiences, constraints, and priorities. A boundary explorer will propose a small number of distant but plausible connections among experiences, identities, and contexts. A critique component will flag cliché, exaggeration risk, and identity flattening. A voice preservation component will rewrite while maintaining measurable features of the student's style. A reflection component will pose targeted questions that help students interpret what their experiences mean and why they matter. This design creates structured opportunities for students to discover surprising directions in their own stories while maintaining ownership and authenticity.

These strands will set up the project for at least one co-authored paper with Microsoft researchers in venues such as CHI, COLM, or AIED, focused on AI tools that better support creative writing for formal essays.

4. Collaboration with Microsoft researchers and impact

I am excited about the opportunity to collaborate with the Microsoft Research Accelerator team and the researchers associated with this challenge. With Shannon Monroe and Matt Corwine, I would explore model design and prototyping strategies that incorporate conceptual boundary awareness into generative systems for text based creativity support. With Richard Banks and Sean Rintel, I see strong synergies in understanding how people use AI tools in sensitive identity loaded creative practices such as admissions writing and in designing interactions that foreground agency, reflection, and ownership. With Neeltje Berger, I would connect perspectives from HCI, education, and creativity research to develop evaluation approaches that move beyond simple preference judgments and instead capture changes in how students see their possibilities and identities.

Together, we could prototype and test tools that move away from a simple finish my essay mode and toward reflective and boundary expanding AI collaborators. I hope this work will not only advance computational creativity, but also yield design principles for AI writing systems in other high stakes creative domains.

5. Preliminary progress and fit with the fellowship

Given the size and richness of the data I work with, including essays and linked applicant information across many years, I am already running large scale analyses on two dedicated servers to track changes in lexical diversity and complexity in admissions essays over time. Early results show clear signs that essays are becoming more homogenized in terms of vocabulary and readability. This fellowship would allow me to focus my time on this project, deepen these analyses, and extend my current work on homogenization and algorithmic voices into explicitly creativity focused and boundary focused AI systems, while drawing on expertise in computational creativity and model design that goes beyond what is available in my current lab. It would also support the co development of new interaction paradigms with direct relevance for education technology companies, universities, and policymakers who are trying to make sense of AI assisted admissions.

In short, my research goal is to understand and redesign AI systems so that they amplify, rather than overwrite, human creativity and diversity of expression. This goal fits closely with the fellowship's focus on tools that foster innovation. I am excited about the possibility of working with Microsoft researchers to rethink what it means for AI to support bold and original thinking in writing, starting from one of the most consequential genres in young people's lives, the college admissions essay.

6. Reference

Beck, S. W., & Godley, A. J. (2023). "What Makes You, You": The Discursive Construction of the Self in US College Application Essays. *American Journal of Education*, 129(4), 539–564.

Bower, J., Acar, S., & Kursuncu, U. (2023). Measuring creativity in academic writing: an analysis of essays in advanced placement language and composition. *Journal of Advanced Academics*, 34(3-4), 183-214.

Chakrabarty, T., Laban, P., & Wu, C. S. (2024). *Can AI writing be salvaged? Mitigating Idiosyncrasies and Improving Human-AI Alignment in the Writing Process through Edits*. arXiv preprint arXiv:2409.14509.

Drosos, I., Williams, J., Sarkar, A., Wilson, N., Rintel, S., & Panda, P. (2025, June). Dynamic Prompt Middleware: Contextual Prompt Refinement Controls for Comprehension Tasks. In *Proceedings of the 4th Annual Symposium on Human-Computer Interaction for Work* (pp. 1-23).

Hofmann, V., Kalluri, P. R., Jurafsky, D., & King, S. (2024). *AI generates covertly racist decisions about people based on their dialect*. *Nature*, 633(8028), 147–154.

Hu, T., Kyrychenko, Y., Rathje, S., Collier, N., van der Linden, S., & Roozenbeek, J. (2025). *Generative language models exhibit social identity biases*. *Nature Computational Science*, 5(1), 65–75.

Huang, T. J. (2024). *Translating authentic selves into authentic applications: Private college consulting and selective college admissions*. *Sociology of Education*, 97(2), 174–192.

Lee, J., Alvero, A. J., Joachims, T., & Kizilcec, R. F. (2025). Poor Alignment and Steerability of Large Language Models: Evidence from College Admission Essays. *arXiv preprint arXiv:2503.20062*.

Liang, W., Izzo, Z., Zhang, Y., Lepp, H., Cao, H., Zhao, X., Chen, L., Ye, H., Liu, S., Huang, Z., McFarland, D. A., & Zou, J. Y. (2024). *Monitoring AI-modified content at scale: A case study on the impact of ChatGPT on AI conference peer reviews*. arXiv preprint arXiv:2403.07183.

Moon, K., Green, A. E., & Kushlev, K. (2025). Homogenizing effect of large language models (LLMs) on creative diversity: An empirical comparison of human and ChatGPT writing. *Computers in Human Behavior: Artificial Humans*, 100207.

Wang, A., Morgenstern, J., & Dickerson, J. P. (2025). *Large language models that replace human participants can harmfully misportray and flatten identity groups*. *Nature Machine Intelligence*, 1–12.