

Human-Generated Content Achieves More Divergence than LLM-Generated Content: An Empirical Comparison of Human and ChatGPT Writing

Kibum Moon¹, Elizabeth Kronthal¹, Adam Green¹, Kostadin Kushlev¹
¹Georgetown University

Introduction

- Large Language Models (LLMs) like ChatGPT may enhance creativity but also raise concerns about a “homogenizing effect” – reducing idea diversity across groups of people who use the same AI model.
- We explored this phenomenon by comparing creativity reflected in human-written and AI-generated college admission essays at both individual and collective levels.

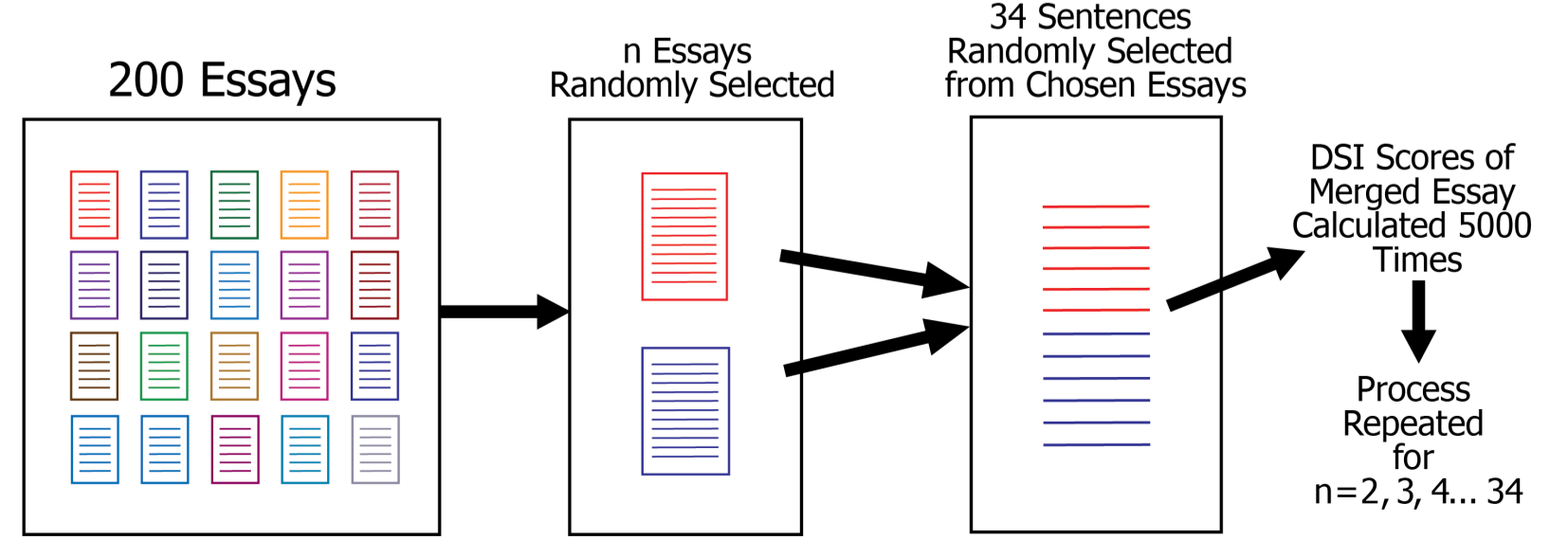
Methods

Data

- We analyzed 600 college admission essays: 200 generated by GPT-4 and 400 written by actual human applicants who applied to a private university from 2018 to 2022. The human-written essays were further divided into two groups: 200 from randomly selected general applicants and 200 from applicants with a diverse range of races and ethnicities.

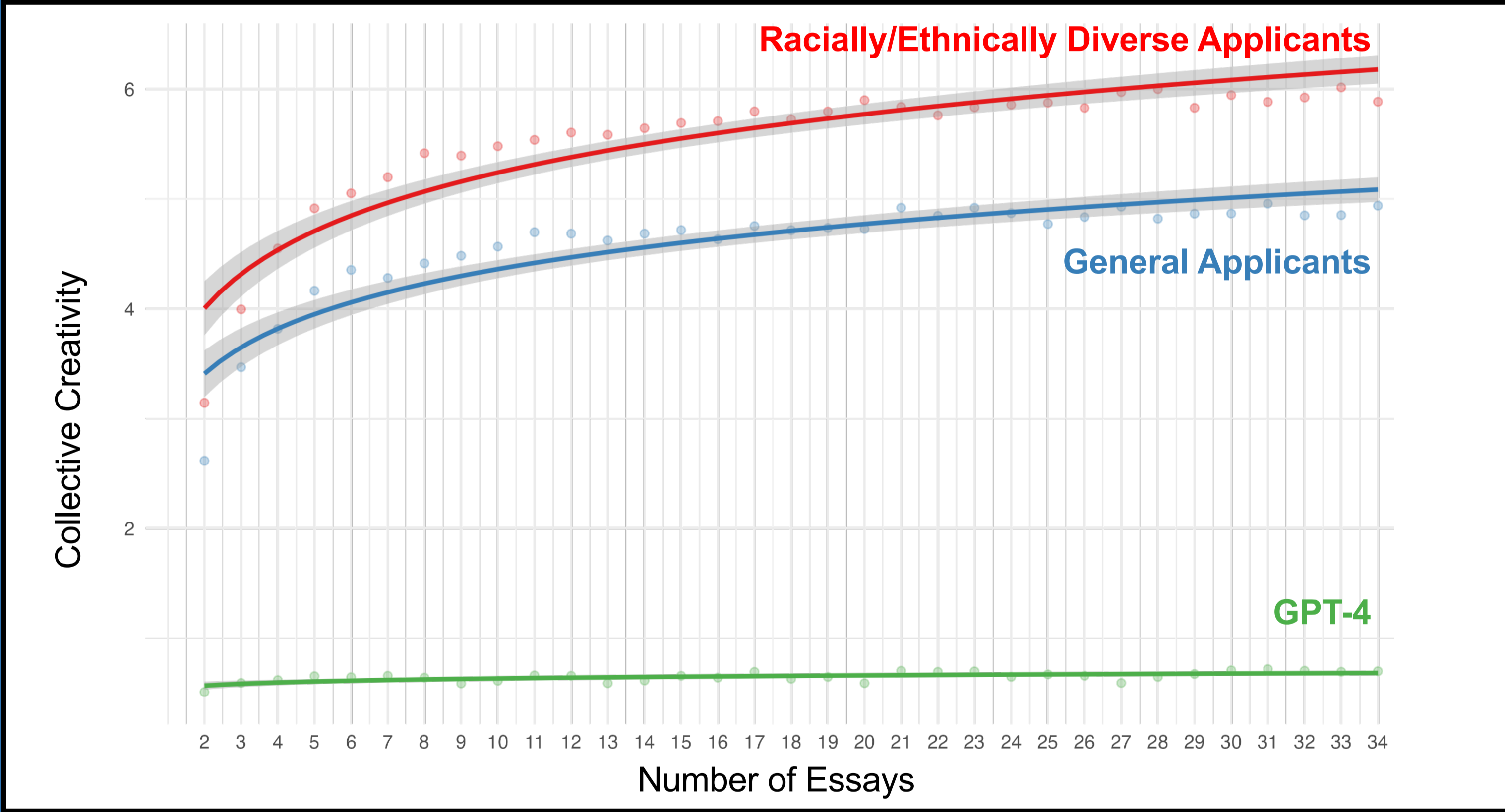
Measures

- Individual Creativity.** We computationally assess the creativity level of each essay by measuring idea diversity utilizing a semantic distance approach – *Divergent Semantic Integration* (DSI).
- Collective Creativity.** We defined collective creativity as the cumulative creative output of a group. We assessed collective creativity by measuring the incremental semantic diversity as we pooled together a progressively larger number of essays.



Does ChatGPT homogenize the diversity of ideas?

Each additional human essay adds a greater semantic diversity of ideas than each additional GPT-4 essay.



Note. Collective creativity represents the unique semantic diversity incremented by each additional essay compared to the average semantic diversity of individual essays. The semantic diversity of essay content was measured by Divergent Semantic Integration (DSI; Johnson et al., 2022). The fitted curves show increases in collective creativity as the number of essays increases. The interaction between authorship and the log-transformed number of essays was significant, suggesting that the cumulative creative outputs were the largest among racially/ethnically diverse human group ($p < .01$), followed by the general human group ($p < 0.5$). Shades represent 95% confidence intervals.



Scan for the Online Material



The Digital Health and Happiness Lab
 Georgetown University

Results

Individual Creativity

- Human-written essays were more semantically diverse than GPT-4-generated essays for regular (Cohen's $d = 0.44$, $p < .001$) and racially diverse applicants (Cohen's $d = 0.31$, $p = .004$).
- There was no difference in semantic diversity between essays from the regular applicants and the racially/ethnically diverse applicants (Cohen's $d = 0.13$, $p = .193$).

Aggregated Creativity

- Each additional human essay adds a greater diversity of ideas than each additional GPT-4 essay does (see Table 1).
- Within the human-authored essays, the increase in semantic diversity was more pronounced for those from the racially/ethnically diverse group than for those from the general applicants' group.

Table 1. Regression Analysis Predicting Collective Creativity.

	Model1	Model2
Authorship [Diverse Group]	0.946 ***	0.474 **
Authorship [GPT-4]	-3.938 ***	-2.458 *
Log(Number of Essays)	0.467 ***	0.592 ***
Authorship [Diverse Group] × Log(Number of Essays)		0.167 **
Authorship [GPT-4] × Log(Number of Essays)		-0.551 ***
Observations	99	99
R ² (Δ R ²)	0.982	0.993 (0.011)

Note. * $p < .05$ ** $p < .01$ *** $p < .001$. The reference group for the authorship effect was the regular human group. We applied a logarithmic transformation to the number of essays to capture the non-linear and declining effect of the number of essays on changes in aggregated creativity.

Discussion

- GPT-4 did not produce creative content that is comparable to that of individual humans.
- Moreover, the GPT-4 did not match the unique and diverse ideas generated by a collective of humans.
- These findings highlight the risk of a "homogenizing effect" on creativity through the repeated use of a specific LLM.
- Our findings indicate that an overreliance on AI models at the societal level could result in a diminished diversity of creative ideas.
- Conversely, promoting racial and ethnic diversity can enrich the diversity of ideas in creative outputs.