Does Writing With Language Models (LLMs) Reduce Content Diversity?

Vishakh Padmakumar, He He

ICLR 2024

Motivation

Motivation

- Evidence that LLMs can influence the views of users during co-writing [1,2]

[1]Jakesch, Maurice, et al. "Co-writing with opinionated language models affects users' views." *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 2023.
[2] Bhat, Advait, et al. "Interacting with Next-Phrase Suggestions: How Suggestion Systems Aid and Influence the Cognitive Processes of Writing." *Proceedings of the 28th International Conference on Intelligent User Interfaces*. 2023.

Motivation

- Evidence that LLMs can influence the views of users during co-writing [1,2]
- As different users rely on the same model for suggestions, this creates an algorithmic monoculture[3]

[1]Jakesch, Maurice, et al. "Co-writing with opinionated language models affects users' views." *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 2023.

[2] Bhat, Advait, et al. "Interacting with Next-Phrase Suggestions: How Suggestion Systems Aid and Influence the Cognitive Processes of Writing." *Proceedings of the 28th International Conference on Intelligent User Interfaces*. 2023.

[3] Kleinberg, Jon, and Manish Raghavan. "Algorithmic monoculture and social welfare." *Proceedings of the National Academy of Sciences* 118.22 (2021): e2018340118.

Motivation

- Evidence that LLMs can influence the views of users during co-writing [1,2]
- As different users rely on the same model for suggestions, this creates an algorithmic monoculture[3]

Does collaborative with LLMs result in different users writing more similar text, reducing the overall diversity of content produced?

[1]Jakesch, Maurice, et al. "Co-writing with opinionated language models affects users' views." *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 2023.
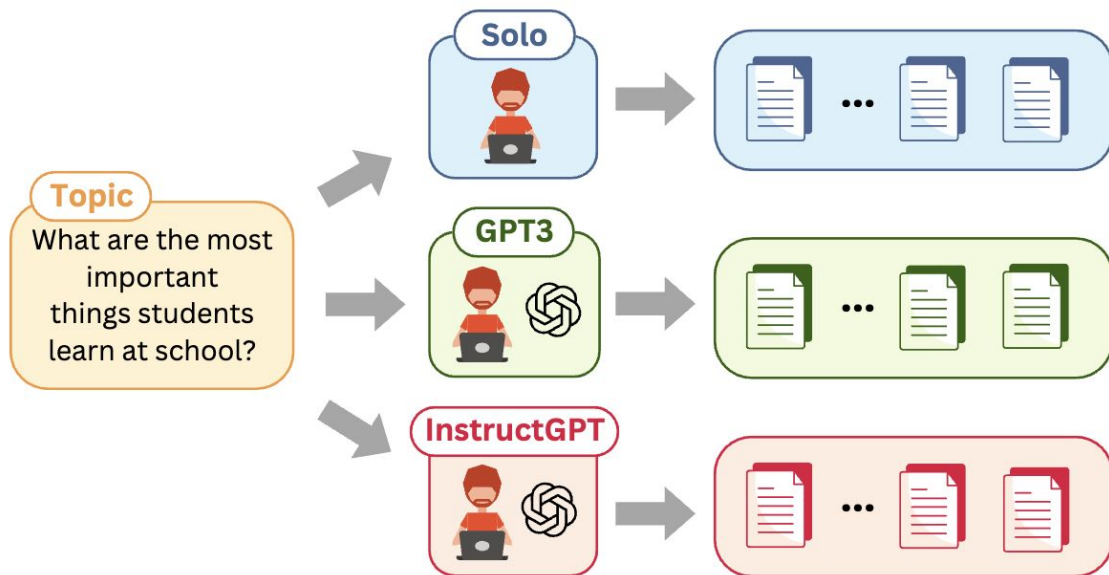[2] Bhat, Advait, et al. "Interacting with Next-Phrase Suggestions: How Suggestion Systems Aid and Influence the Cognitive Processes of Writing." *Proceedings of the 28th International Conference on Intelligent User Interfaces*. 2023.
[3] Kleinberg, Jon, and Manish Raghavan. "Algorithmic monoculture and social welfare." *Proceedings of the National Academy of Sciences* 118.22 (2021): e2018340118.

User Study Format

User Study Format

- **Task:** Argumentative Essay writing (~300-500 words) on a [set of 10 open ended questions as collected by NYT](#)
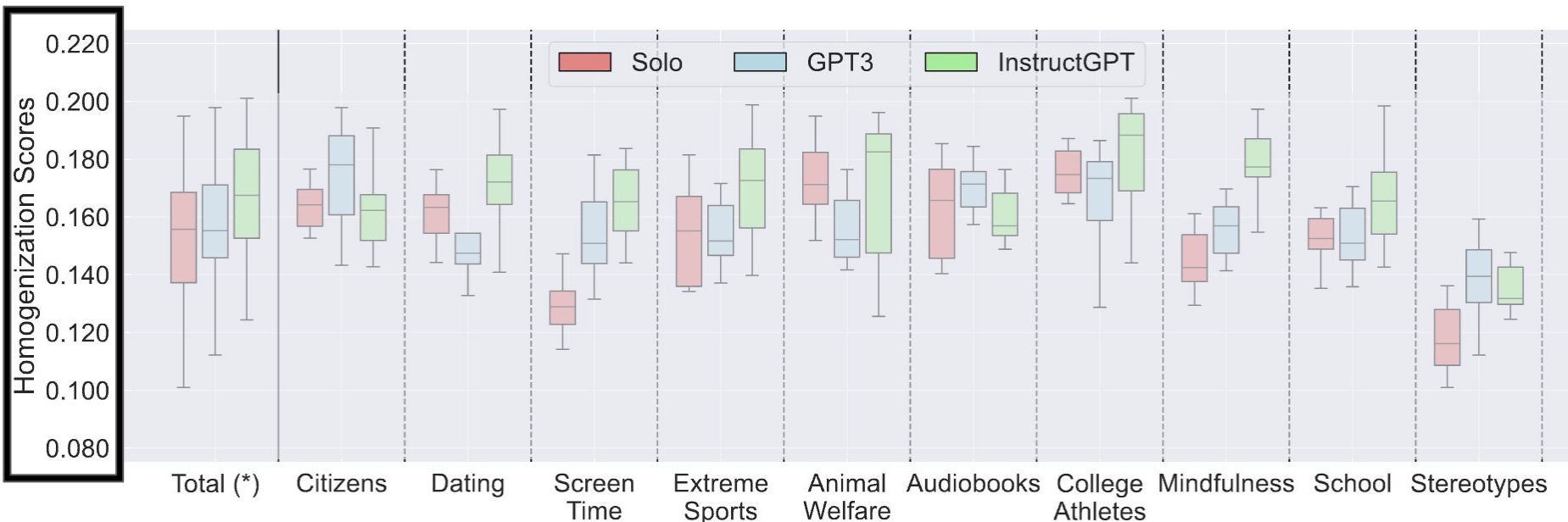
Example Essay

Example Essay:
Topic: How Worried Should We Be About Screen Time During the Pandemic?
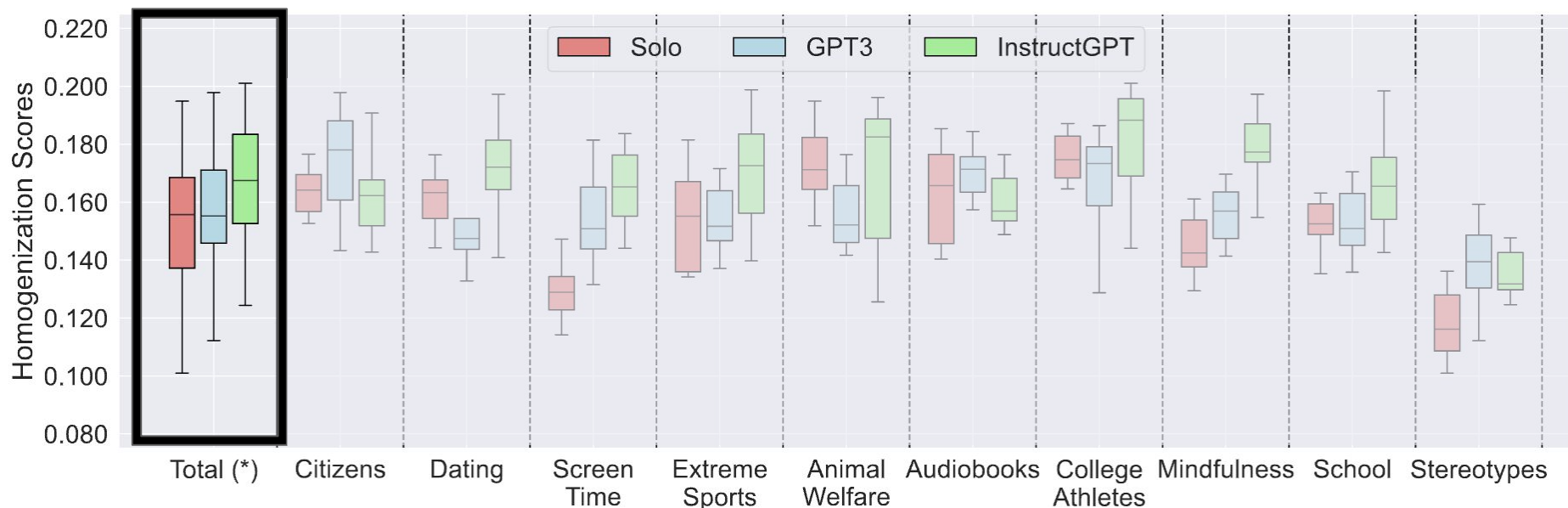
While I believe the concerns regarding children's screen time are valid, I believe it is somewhat biased to not take this problem, which is a genuine issue right now, as an everyone problem, [skipped] I know that I, along with many other teenagers, would like nothing more than to go back to school, play sports outside, meet new people, and such. [skipped] They are also places for new ideas, watching college lectures, and political discourse [skipped]

We compare the 'homogenization' i.e. average similarity to other the essays from the same setup
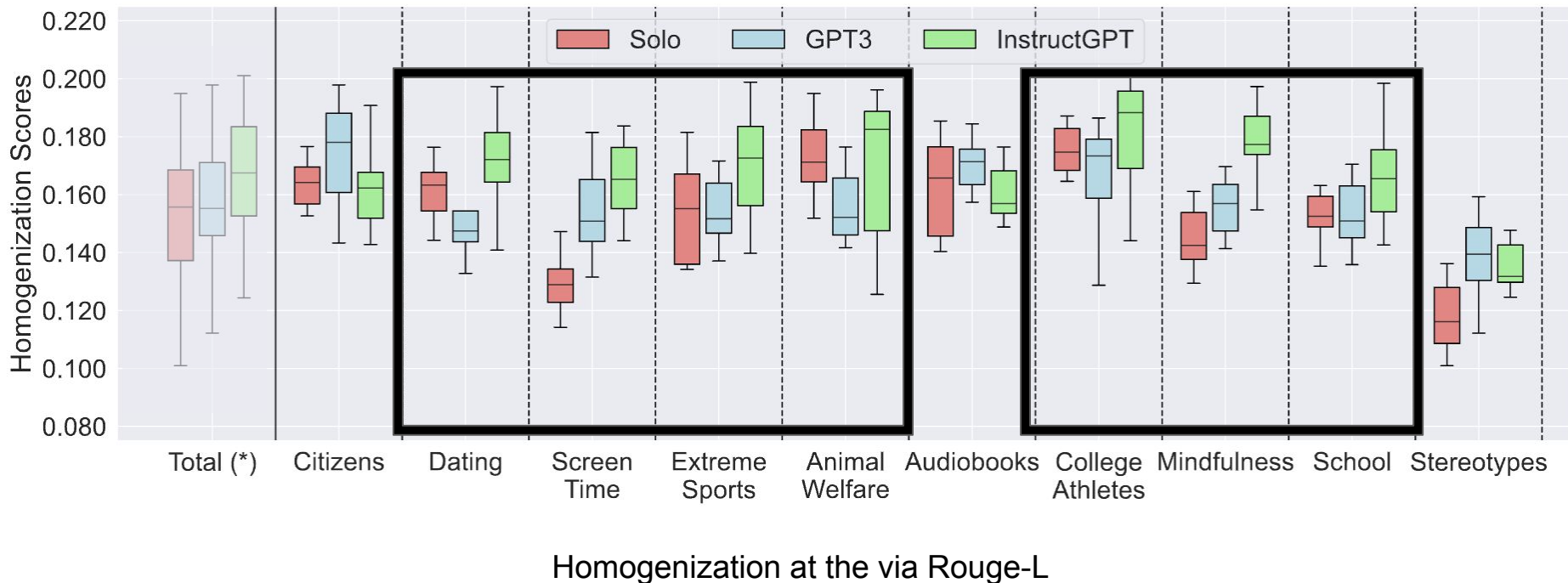

Homogenization at the via Rouge-L

Writing with InstructGPT results in the highest average homogenization or most similar essays

Homogenization at the via Rouge-L

InstructGPT has the highest median homogenization in 7 out of 10 topics

Homogenization at the via Rouge-L

RQ: Does Writing With LLMs Reduce Overall Diversity?

Formalize Diversity Using Unique Information

We calculate the diversity of a set of essays as the fraction of unique information in them

Formalize Diversity Using Unique Information

We calculate the diversity of a set of essays as the fraction of **unique information** in them

How Can We Compare the Content of Essays?

How Can We Compare the Content of Essays?

Example Essay:
Topic: How Worried Should We Be About Screen Time During the Pandemic?

While I believe the concerns regarding children's screen time are valid, I believe it is somewhat biased to not take this problem, which is a genuine issue right now, as an everyone problem, [skipped] I know that I, along with many other teenagers, would like nothing more than to go back to school, play sports outside, meet new people, and such. [skipped] They are also places for new ideas, watching college lectures, and political discourse [skipped]

Comparing the Content of Essays via Summarization

Raw Text Level

While I believe the concerns regarding children's screen time are valid, I believe it is somewhat biased to not take this problem, which is a genuine issue right now, as an everyone problem, [skipped] I know that I, along with many other teenagers, would like nothing more than to go back to school, play sports outside, meet new people, and such. [skipped] They are also places for new ideas, watching college lectures, and political discourse [skipped]

Comparing the Content of Essays via Summarization

Raw Text Level

While I believe the concerns regarding children's screen time are valid, I believe it is somewhat biased to not take this problem, which is a genuine issue right now, as an everyone problem, [skipped] I know that I, along with many other teenagers, would like nothing more than to go back to school, play sports outside, meet new people, and such. [skipped] They are also places for new ideas, watching college lectures, and political discourse [skipped]

- The problem of screen time should be considered an everyone problem, not just a student one
- Social media can be used for educational purposes
- Limiting screen time may not be effective in the long run
- Parents should trust their teenager more and not worry too much about their screen time

Key Point Level

Formalize Diversity Using Unique Information

We calculate the diversity of a set of essays as the fraction of **unique information** in them

Content Diversity:

Information Unit: Key Points

Diversity Measure: Fraction of Unique Clusters of Key Points

Example of Clustering of Key Points

Essay 1:
1. They help to develop self-awareness, stress reduction and emotional regulation
2. It's important to make sure practices are inclusive and voluntary so that students don't feel forced into them
3. Mindfulness and meditation can be personalized for each individual

….

Essay 2:
1. They should be implemented in a culturally neutral and straightforward manner and information on their benefits should be provided to students.
2. Focus should be on the scientific principles behind mindfulness and meditation as well as self-care, emotional regulation, and stress.
3. Mindfulness and meditation should not be forced or used to guide or persuade students towards particular beliefs.

Example of Clustering of Key Points

Essay 1:

1. **They help to develop self-awareness, stress reduction and emotional regulation**
2. It's important to make sure practices are inclusive and voluntary so that students don't feel forced into them
3. Mindfulness and meditation can be personalized for each individual

….

Essay 2:

1. They should be implemented in a culturally neutral and straightforward manner and information on their benefits should be provided to students.
2. **Focus should be on the scientific principles behind mindfulness and meditation as well as self-care, emotional regulation, and stress.**
3. Mindfulness and meditation should not be forced or used to guide or persuade students towards particular beliefs.
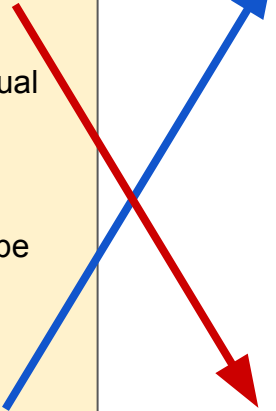
Cluster 1

Example of Clustering of Key Points

Essay 1:
1. **They help to develop self-awareness, stress reduction and emotional regulation**
2. **It's important to make sure practices are inclusive and voluntary so that students don't feel forced into them**
3. Mindfulness and meditation can be personalized for each individual

….

Essay 2:
1. They should be implemented in a culturally neutral and straightforward manner and information on their benefits should be provided to students.
2. **Focus should be on the scientific principles behind mindfulness and meditation as well as self-care, emotional regulation, and stress.**
3. **Mindfulness and meditation should not be forced or used to guide or persuade students towards particular beliefs.**
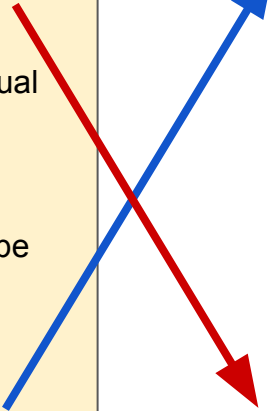
Cluster 1

Cluster 2

Example of Clustering of Key Points

Essay 1:
1. **They help to develop self-awareness, stress reduction and emotional regulation**
2. **It's important to make sure practices are inclusive and voluntary so that students don't feel forced into them**
3. Mindfulness and meditation can be personalized for each individual

Essay 2:
1. They should be implemented in a culturally neutral and straightforward manner and information on their benefits should be provided to students.
2. **Focus should be on the scientific principles behind mindfulness and meditation as well as self-care, emotional regulation, and stress.**
3. **Mindfulness and meditation should not be forced or used to guide or persuade students towards particular beliefs.**
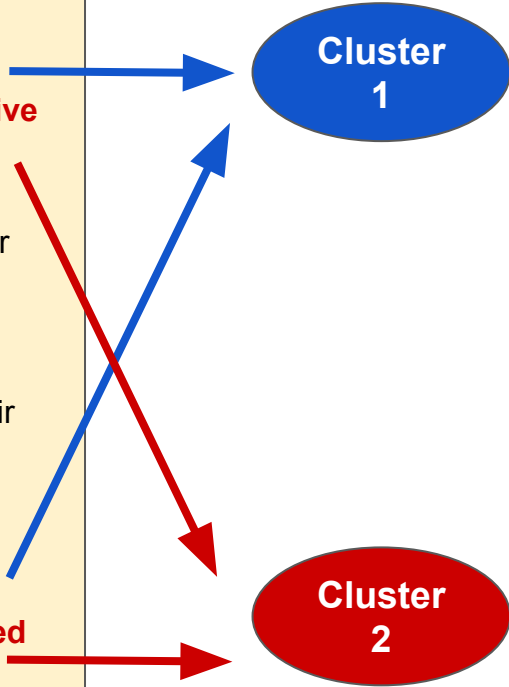
Cluster 1

Cluster 2

Diversity =

Fraction of Unique Key Points =

4 / 6 = 0.66

Results

Thresholds	Solo	GPT3	InstructGPT
0.5	0.982	0.971	**0.950**
0.6	0.941	0.927	**0.877**
0.7	0.792	0.779	**0.738**
0.8	0.543	0.514	**0.494**

(a) RougeL

Varying The Thresholds Impacts Cluster Size

Thresholds	Solo	GPT3	InstructGPT
0.5	0.982	0.971	**0.950**
0.6	0.941	0.927	**0.877**
0.7	0.792	0.779	**0.738**
0.8	0.543	0.514	**0.494**

(a) RougeL

Writing with InstructGPT reduces key point diversity across all thresholds

Thresholds	Solo	GPT3	InstructGPT
0.5	0.982	0.971	**0.950**
0.6	0.941	0.927	**0.877**
0.7	0.792	0.779	**0.738**
0.8	0.543	0.514	**0.494**

(a) RougeL

Takeaways

- Collaboration with InstructGPT makes users write more similar essays, reducing the overall diversity as well

- This effect is not observed with GPT3 highlighting that the bump in performance from tuning the model on human feedback comes at the cost of more homogeneous content