Exploring Bias and Creativity: A Comparative Study of Human and AI Generated Works with AI Priming*

**** Generated with ChatGPT 3.5***

Angela Chang, Dean Nakada, Kevin Stallone, Dave Zack

DATASCI 241: Experiments and Causal Inference (Summer 2023)

Professor Scott Guenther

August 15, 2023

Abstract

Recent advancements in large language models such as OpenAI's ChatGPT have caused rising concern due to our diminishing ability to distinguish GPT writing from that of a human. Previous studies have focused on said ability; however with the rise of discussion about ChatGPT in popular media, there have not been any with a focus on the effect the discussion has on people's ability to differentiate between authors. Our research aims to start filling this gap in the literature, and answer whether subjects will have a higher likelihood of being convinced something is written by a large language model such as ChatGPT when primed with the capacity of AI. Through our 2x2x2 questionnaire sent out to the general US population (n=667), we found that people were more likely to believe excerpts were written by AI than by humans when given information on the growing capacity of AI technology. These findings, in conjunction with the proliferation of news stories on the subject, indicate a growing capacity to believe content is AI-generated regardless of the source and highlight the necessity of new methods to distinguish between generated and human written content.

Introduction

With the spread and growth of AI technologies, worry and paranoia have begun to flourish; in one extreme case, a professor attempted to fail all of his students, believing that they had used ChatGPT to do their work (Ankel, 2023). Researchers are evaluating methods to identify AI-generated content, but questions remain on whether the rapid increase of news on or growing attention to developing AI capacity to generate written text affects how people judge content to be human or AI generated. In this experiment, we assessed how often people would judge writing as AI-generated, while randomly including a preface on the capacity of AI technology to prime subjects. This study seeks to answer: Does "priming" subjects with an article discussing the present capacity of AI affect their propensity to assign an article as being AI written?

Theory

As defined by the American Psychological Association (APA), priming is the "effect in which recent experience of a stimulus facilitates or inhibits later processing of the same or a similar stimulus" (*APA Dictionary of Psychology*, n.d.). Previous investigations have found that media coverage of various diseases affected perception of their severity and prevalence (Young et al., 2008). This example of media priming can be extended to present coverage of advancements in AI, most notably being OpenAI's "ChatGPT" based on the Generative Pre-trained Transformer (GPT). The natural language processing algorithm present in the GPT-4 architecture allows the AI to follow complex instructions and solve difficult problems with accuracy (OpenAI, 2023). Our team is primarily interested in discovering if priming subjects will affect the rate at which they determine an article was written by AI regardless of whether or not the article was actually written by AI. To elaborate, we believe that when a subject is primed by reading about the capabilities of generative AI, they will be more likely to reach the conclusion that any subsequently tasked passages are AI written.

Methodology

The primary research questions addressed in this study pertain to the impact of priming on two factors:

1. The likelihood of subjects perceiving a passage as written by AI
2. The accuracy of subjects' in determining if passage is written by AI or a human

Hypothesis

We hypothesize that exposure to an article about recent strides in AI capability can lead people to perceive that the passages they read are authored by AI. Additionally, we suspect that priming can influence the accuracy of their judgment on a passage's authorship; with information about generative AI, individuals may improve accuracy for AI-generated pieces but will have decreased accuracy for human-written pieces.

Null Hypothesis (H_0): $\mu_0 = \mu_a$

There is no significant difference between control and treatment groups when it comes to the rate of selecting AI as the passage author. As a result, there will be no significant difference in accuracy rates between control and treatment groups.

Alternative Hypothesis (H_a): $\mu_0 < \mu_a$ for AI rate

$\mu_0 < \mu_{a_1}$ for accuracy on AI-generated pieces, $\mu_0 > \mu_{a_2}$ for accuracy on human-written pieces

Exposure to an article about the recent advancements in the capabilities of AI, the treatment, will lead people to believe that passages they read are authored by AI. As a result, for the treatment group, accuracy on AI-generated pieces will increase, but accuracy on human-written pieces will decrease.

A secondary yet significant inquiry revolves around the potential variation in these effects based on passage styles. It is reasonable to assume that AI can generate more human-like texts for certain styles and genres of passages, which can make it more difficult for people to distinguish the writer. To attempt to assess this, we chose two distinct styles of passage, op-ed and fiction, with six different options in each, three human written and three AI generated. The goal was to ensure one passage type or topic was not the driver behind subjects' perception of the passage.

Experiment Design

To assess these questions, we conducted an online survey-based experiment.Our experiment was a posttest-only randomized experiment with a 2 x 2 x 2 factorial design. In the following ROXO grammar, X represents the treatment, X_1/X_2 indicates AI or human-written op-ed, and $X\__1$ /$X\__2$ indicates AI or human-written fiction piece. As a result, there are two outcomes represented, one for op-ed pieces and one for fiction pieces.

R	X	X_1	X_{11}	O	O
R		X_1	X_{11}	O	O
R	X	X_1	X_{12}	O	O
R		X_1	X_{12}	O	O
R	X	X_2	X_{21}	O	O
R		X_2	X_{21}	O	O
R	X	X_2	X_{22}	O	O
R		X_2	X_{22}	O	O

First, participants were randomly assigned to the control and treatment groups. Then they were randomly assigned an op-ed piece and then a fiction piece.

The survey comprised of three main sections:

1. Demographic questions
2. Priming article for the treatment group/placebo article for the control group
3. Passages (2)

The demographic questions collected basic subject information, including gender, age group, and highest level of completed education. We also inquired whether participants natively spoke English to exclude non-native English speakers from the analysis, as the study focused on assessing perception about English written passages. Additionally, the survey assessed the number of books written in English they read annually and any prior experience working in the technology sector, both of which could influence the study's outcomes. A simple screener question was also included to ensure the accuracy of responses, and participants who answered incorrectly were not included in the analysis.

In the priming phase, the treatment group read an article about the capabilities of generative AI and large language models. Subsequently, subjects answered a comprehension question about the article to confirm careful reading and compliance. The treatment passage, excerpted and adapted from both Reece Rogers's "How to Detect AI-Generated Text, According to Researchers" and Steven Johnson's "A.I. Is Mastering Language. Should We Trust What It Says?," is below:

> *AI-generated text, from tools like ChatGPT, is starting to impact daily life. Teachers are testing it out as part of classroom lessons. Marketers are champing at the bit to replace their interns. Memers are going buck wild. Me? It would be a lie to say I'm not a little anxious about the robots coming for my writing gig. (ChatGPT, luckily, can't hop on Zoom calls and conduct interviews just yet.)*

> ***With generative AI tools now publicly accessible, you'll likely encounter more synthetic content while surfing the web.*** *Chances are you have already interacted with a large language model if you've ever used an application — like Gmail — that includes an autocomplete feature, gently prompting you with the word "attend" after you type the sentence "Sadly I won't be able to…." But autocomplete is only the most rudimentary expression of what software like GPT is capable of. It turns out that with enough training data and sufficiently deep neural nets, large language models can display remarkable skill if you ask them not just to fill in the missing word, but also to continue on writing whole paragraphs in the style of the initial prompt. Algorithms with the ability to mimic the patterns of natural writing have been around for a few more years than you might realize.*

In order to keep the number of passages read throughout the survey consistent between control and treatment groups and to further assess the impact of priming subjects, the subjects in control received a placebo passage describing plant based meat in addition to the reading comprehension verification question or compliance check. The placebo passage, excerpted from "Emerging Technologies Set to Shape Next Generation of Plant-Based Meat," by Michael Dent, is below:

> *Plant-based meat has increasingly been making headlines in recent years, as product sales swell and industry investment continues to reach dizzying heights. "Plant-Based Meat 2021-2031", a new report by IDTechEx, explores the technologies that are shaping the plant-based meat industry, alongside the consumer and market factors that will decide whether plant-based meat can truly disrupt the $1 trillion global meat industry.*
>
> *Modern plant-based meat companies such as Beyond Meat and Impossible Foods have invested heavily into R&D and technology development to make their products as realistic as possible. Impossible Foods uses a genetically engineered strain of yeast to produce soy leghemoglobin, a key ingredient that makes its meat substitutes "bleed" and gives them a uniquely meaty flavor. To create its Beyond Burger product, Beyond Meat uses a food extrusion machine originally developed at the University of Missouri, which uses heat and pressure to force plant proteins into a fibrous, meat-like texture that resembles muscle fibers. Rather than using genetic engineering to produce its products, the Beyond Meat burger uses beet juice to replicate the bleeding from a real burger. Coconut oil and cocoa butter are used to provide marbling to further replicate the texture of real meat.*
>
> ***This technology development has helped to create a new generation of plant-based meat products that are slowly winning over meat-eating consumers****. Now, companies across the world are working to leverage a range of emerging technologies to help create the next generation of plant-based products.*

Participants were then presented with two passages in different styles: one op-ed and one fiction. This assessment aimed to investigate potential variations in the impact of priming depending on the style of passages. For each style, three versions of human-written passages were prepared alongside AI-generated versions produced by GPT-4, OpenAI's most advanced system. Multiple prompts were used during passage generation to ensure that human-written and AI-generated versions of each passage covered the same topic and had similar lengths. Random selection delivered one passage for each style to each participant.

After reading each passage, participants answered two questions: one about the quality of the passage's writing and either the persuasiveness of its arguments for the op-eds or the level of interest in reading the full story for the fiction pieces. The order of these two questions was randomized. Following the two passages and questions, participants were asked to identify whether each passage was written by an AI or a human and their confidence in their answers. These final questions were randomized to control for order effects and were intentionally placed at the end of the survey to minimize potential bias during passage reading. We used the resulting classifications to measure the potential outcomes for the treatment and control group: whether or not subjects believed the passage they read was AI-generated or human written. All passages and the full survey can be found in the appendix.

Sampling and Fielding

The sample consisted of 667 native English speakers residing in the United States, all aged 18 years and older. We partnered with PureSpectrum, a reputable private firm offering an open market research and insights platform known for its quality data and extensive audience pool across the country. Participant conditions were defined on the platform dashboard, with US census proportional quotas set for age group, gender, and educational attainment. Responses from 330 and 337 subjects were collected for the control and treatment groups, respectively. Participants selected to take the survey in exchange for compensation.

Of the respondents who completed the survey, 264 or 39.58% identified as male, while 392 or 58.77% identified as female.

Respondents were across the age spectrum from 18 to 70+:

- 65 or 9.75% were in the 18-22 range
- 107 or 16.04% were in the 23-29 range
- 128 or 19.19% were in the 30-39 range
- 110 or 16.49% were in the 40-49 range
- 121 or 18.14% were in the 50-59 range
- 87 or 13.04% were in the 60-69 range
- 47 or 7.05% were in the 70+ range
- 2 respondents did not provide their age

For highest completed educational attainment:

- 31 or 4.65% did not complete high school or receive a GED
- 203 or 30.43% had completed high school or achieved a GED
- 142 or 21.29% had completed some college/university without a degree
- 165 or 24.73% completed college or university
- 19 or 2.84% had some graduate school achieved
- 83 or 12.44% had a graduate degree or higher
- 24 or 3.56% did not provide their highest educational attainment or completed trade school or some other certification program

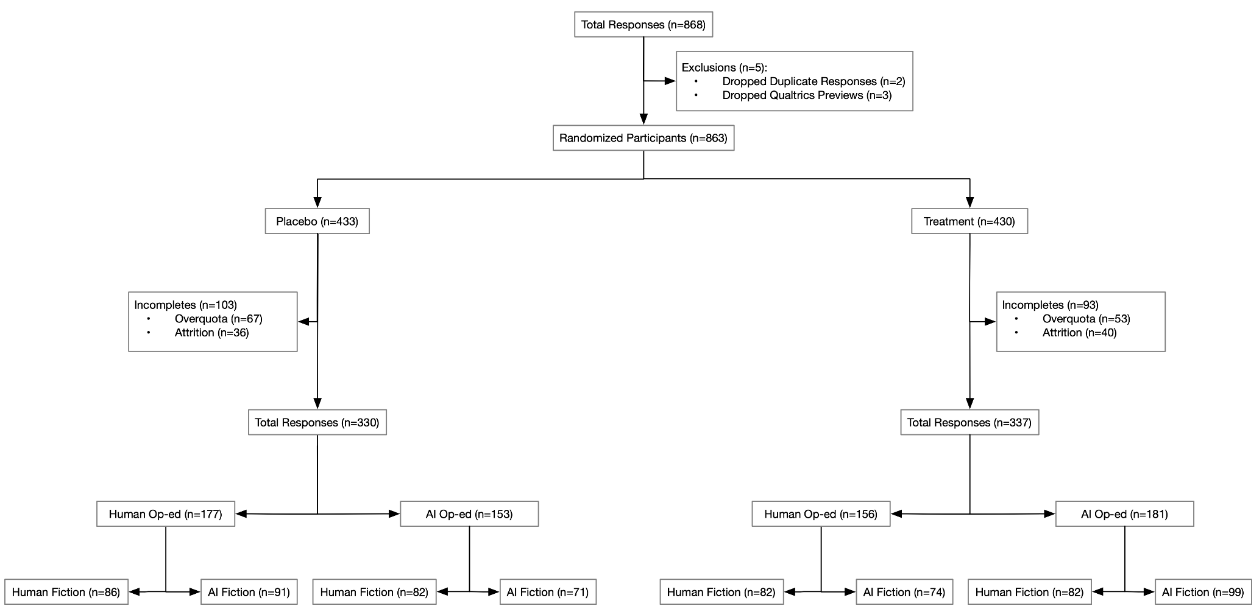
The majority of respondents reported they did not currently work in or had previously worked in tech (463 or 69.42%). Respondents also reported that they read anywhere from 0 to 5,000 books per year, though we regard the maximum with much skepticism. The majority of respondents, roughly 600, provided answers in the 0 - 50 books per year range.

The survey was created using Qualtrics, an experience management company's platform. The finalized survey version was linked to the PureSpectrum Sample Marketplace Platform for delivery. Surveying was conducted July 26, 2023 - July 29, 2023, with responses recorded on the Qualtrics platform in real-time for subsequent data analysis.

Within our completed subjects, we had a compliance rate, as measured through accurately answering the reading comprehension question given after the placebo or treatment passage, of 93.33% for the control group and 94.36% in the treatment group. This difference was statistically insignificant when using a two-sample t-test with a p-value of 0.5812. Given the high compliance rate, we reported and discussed the majority of average treatment effects (ATE) as-is, though we provided the local average treatment effect (LATE) or complier average causal effect (CACE) for reference in tables.

We additionally experienced attrition when conducting our experiment. The majority of incompletes shown in our consort document that follows this section was due to participants hitting an over quota wall, where they were not allowed to complete the survey. Roughly 39% of the incompletes were due to choice or attrition.

Consort Diagram



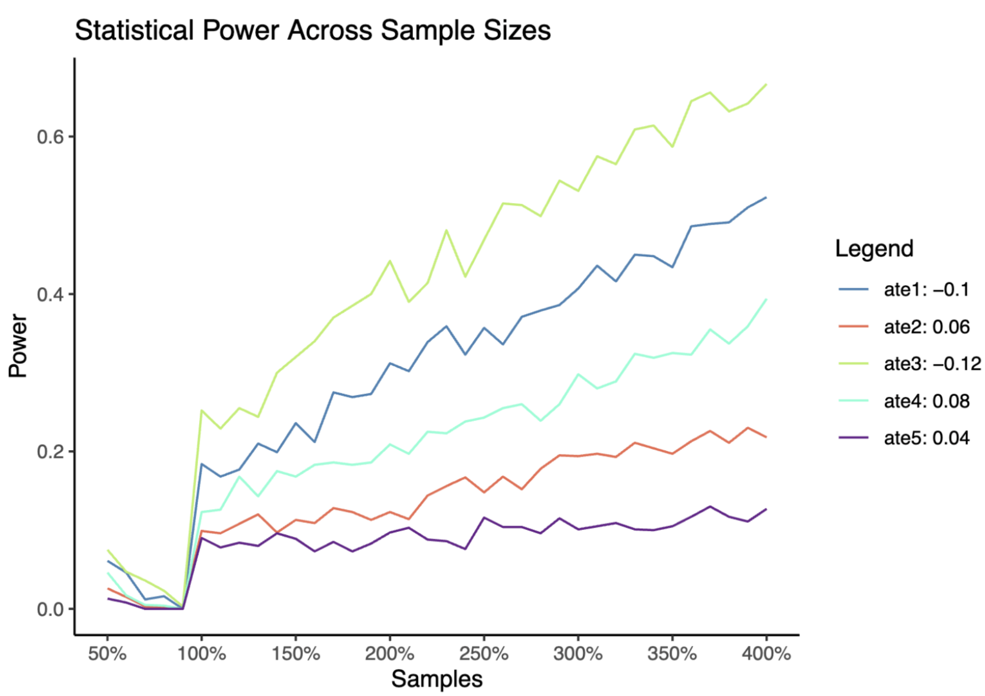
Power Analysis

We conducted a power analysis to estimate the anticipated sample needed to achieve adequate statistical power. Based on literature review of similar experiments and studies, we anticipated that the potential effect size for our treatment would fall in the range of 1-5%. In their 2021 study, Clark et al. tested the capacity of non-experts to differentiate between human-authored and machine-generated text using 50 texts written by humans and 50 texts generated by GPT 3, a widely employed natural language generation system (Clark et al., 2021). We based our expectations on the results from this experiment, mainly that accuracy would be around 50% or similar to chance.

Additionally, the authors explored the efficacy of three lightweight evaluator-training methods aimed at enhancing individuals' ability to identify machine-generated text. The impact of these interventions varied across methods and domains, with the most significant observed improvement in accuracy amounting to a mere four points. We used this four percent increase in accuracy as the basis for our possible true treatment effect of our experiment, 1-5% increase in the rate of judging a written piece to be AI-generated.

We generated a power analysis for five different ATE scenarios, randomly generating the binary data (1 for participants believing the passage to be AI generated or 0 for not) from a normal distribution of probabilities centered around the estimated ATE of 1%, 2%, 3%, 4%, and 5%.

We created 100 rows of random data and sampled randomly from 50% to 400% of the baseline sample size (n = 100) to generate p-values for 1000 randomizations from a t-test comparing treatment and control ATEs. Using the p-values, we then could calculate the statistical power for the experiment across potential sample sizes, 50 - 400, at increments of ten. Even at the highest ATE, 12% increase or decrease in rate of judging passages to be AI-written, we would achieve around 70% power at a sample size of 400. Given this, we would need at least a sample size of 500 in order to achieve statistical power with an ATE of 10% or higher. A lower ATE would likely require a sample of more than 1,000 for adequate power.



Based on these results, we anticipated needing as large a sample as our budget would accommodate. Estimating a cost per interview or survey completion of $0.70, we aimed to have a sample size of 650.

Results

Effects on AI-Rate

The first primary research question we sought to address was whether priming affects the likelihood of subjects perceiving a passage as generated by AI.

According to our analysis, the treatment group who were primed with an article about AI's capabilities exhibited an estimated 10.1% increase with a 95% CI [2.46% , 17.74%] in the proportion of subjects perceiving the op-ed. As seen in Table 1, 43.03% of the control group classified the op-ed piece as AI-written, while 53.12% of the treatment group did so. There was a 12.3% increase with a 95% CI [4.85%, 19.75%] for fiction passages, with 36.36% of the control group believing the piece was AI-written in comparison to 48.66% of the treatment group. The two estimated ATEs have standard errors of approximately 0.04. We reject the null hypothesis of no difference in perception between the control and treatment groups, assuming a significance level of 0.01.

Table 1	AI-Rate (Op-Ed)	AI-Rate (Fiction)
Control	43.03%	36.36%
Treatment	53.12%	48.66%
ATE/ITT	10.10%	12.30%
LATE/CACE	10.70%	13.04%

Subsequently, we investigated whether this treatment effect varied based on the authorship of the passage presented, human or AI. We created additional models featuring interaction effects between the treatment variable and passage author as well as a baseline model for comparison. An F-statistic and its corresponding p-value were calculated to contrast the sum of squared residuals between these two models. We found no significant interaction effects; that is, the AI rate increased, as seen in Table 2, but not significantly differently in regards to whether the passage was actually AI-generated or not.

Table 2	Op-Ed		Fiction	
Passage Author	**AI**	**Human**	**AI**	**Human**
Control AI Rate	39.87%	45.76%	35.80%	36.90%
Treatment AI Rate	55.80%	50.00%	45.66%	51.83%

The initial model (Column (1) of Table 3) pertains to op-ed passages without an interaction variable, while the subsequent model (Column (2) of Table 3) involves op-ed passages with an interaction variable. The F-test conducted on these two models yields an F-statistic of 2.29, which results in a p-value of 0.13. As the p-value is bigger than the conventional significance level 0.05, we fail to reject the null hypothesis that the average treatment effect (ATE) of priming is the same between reading a human-written or AI-generated op-ed.

Similarly, the F-test for the remaining two models related to fiction passages (Column (3) and Column (4) of Table 3) produces an F-statistic of 0.44, resulting in a p-value of 0.51.

Consequently, we are unable to reject the null hypothesis that the ATE of priming, when a subject reads an AI-generated fiction, matches the ATE of priming when a subject reads human-written fiction.

Table 3: *The proportion of subjects perceiving a passage as written by AI, categorized by passage type*

	AI Rate			
	Passage type: Op-ed		Passage type: Fiction	
	(1)	(2)	(3)	(4)
Treated (Priming)	0.101***	0.042	0.124***	0.149***
	(0.039)	(0.055)	(0.038)	(0.054)
Op-ed author: AI	0.0001	-0.059		
	(0.039)	(0.055)		
Treated * Op-ed author: AI		0.117		
		(0.077)		
Fiction author: AI			-0.037	-0.011
			(0.038)	(0.053)
Treated * Fiction author: AI				-0.051
				(0.076)
Constant	0.430***	0.458***	0.382***	0.369***
	(0.033)	(0.038)	(0.033)	(0.037)
Observations	667	667	667	667
R^2	0.010	0.014	0.017	0.017
Adjusted R^2	0.007	0.009	0.014	0.013
Residual Std. Error	0.498 (df = 664)	0.498 (df = 663)	0.491 (df = 664)	0.492 (df = 663)
F Statistic	3.416** (df = 2; 664)	3.045** (df = 3; 663)	5.686*** (df = 2; 664)	3.935*** (df = 3; 663)
Note:				*p<0.1; **p<0.05; ***p<0.01

The following regression (Table 4) presents the outcomes when considering demographic factors or other covariates. Column (1) reflects the overall effect of the treatment on op-ed passages, while Column (3) demonstrates the equivalent estimated effect for fiction passages. Column (2) and Column (4) represent the models incorporating demographic covariates of subjects, including gender, age group, educational background, and occupational history. None of these covariates, except for “Age: Prefer not to say,” which only includes 2 subjects, exhibited a significant association with the outcome variable.

Table 4: *The proportion of subjects who perceive a passage as written by AI*

	AI Rate			
	Passage type: Op-ed		Passage type: Fiction	
	(1)	(2)	(3)	(4)
Treated (Priming)	0.101***	0.104***	0.123***	0.123***
	(0.039)	(0.039)	(0.038)	(0.038)
Gender: Male		0.026		0.056
		(0.042)		(0.041)
Gender: Non-binary / Third gender		0.051		0.081
		(0.171)		(0.181)
Gender: Prefer not to say		-0.015		0.084
		(0.340)		(0.253)
Age: 23-29		-0.104		0.090
		(0.079)		(0.079)
Age: 30-39		-0.021		0.086
		(0.077)		(0.077)
Age: 40-49		-0.019		0.022
		(0.080)		(0.078)
Age: 50-59		-0.014		0.031
		(0.079)		(0.078)
Age: 60-69		-0.0003		0.065
		(0.085)		(0.083)
Age: 70+		-0.081		0.016
		(0.097)		(0.097)
Age: Prefer not to say		-0.482***		-0.391**
		(0.183)		(0.159)
High education		0.042		0.021
		(0.042)		(0.041)
Tech-related occupation		0.017		0.012
		(0.047)		(0.046)
Constant	0.430***	0.421***	0.364***	0.275***
	(0.027)	(0.071)	(0.027)	(0.070)
Observations	667	667	667	667
R^2	0.010	0.021	0.015	0.027
Adjusted R^2	0.009	0.002	0.014	0.008
Residual Std. Error	0.498 (df = 665)	0.500 (df = 653)	0.491 (df = 665)	0.493 (df = 653)
F Statistic	6.843*** (df = 1; 665)	1.088 (df = 13; 653)	10.450*** (df = 1; 665)	1.406 (df = 13; 653)
Note:				*p<0.1; **p<0.05; ***p<0.01

Effects on Accuracy

The second primary research question under our investigation concerns whether priming influences the accuracy of subjects' to identify human or AI authorship of a passage. We found no significant treatment effect, suggesting that the priming did not affect accuracy. Accuracy across treatment and control groups ranged from roughly 47 - 53%, or very close to random chance, as seen in Table 5:

Table 5	Accuracy Rate: Op-Ed	Accuracy Rate: Fiction
Control	47.58%	49.70%
Treatment	53.12%	46.88%

However, when examining potential heterogeneity of the treatment effect on accuracy using the same methodology as the analysis of the AI rate, we found potentially significant results. Accuracy increased for detecting an AI-written op-ed or AI-written fiction piece, though accuracy decreased in regards to the human-written pieces, as seen in Table 6:

Table 6	Op-Ed		Fiction	
Passage Author	AI	Human	AI	Human
Control Accuracy	39.87%	54.24%	35.80%	63.10%
Treatment Accuracy	55.80%	50.00%	45.66%	48.17%

Within Table 7, Column (1) and Column (2) comprise models pertaining to op-ed passages without and with interaction variables, respectively. The F-test applied to these models results in an F-statistic of 6.81, indicating a corresponding p-value of 0.009. With this p-value falling below the standard significance level of 0.05, we reject the null hypothesis that the ATE of priming on accuracy remains constant between instances where a subject reads an AI-generated op-ed and instances where they read a human-written op-ed.

The analysis conducted on fiction passages also revealed heterogeneity. Models for fiction without and with interaction variables are depicted in Column (3) and Column (4) of Table 7, respectively. The F-test applied to these models yields an F-statistic of 10.6, resulting in a p-value of 0.001. Since this p-value falls below the conventional significance level of 0.05, the null hypothesis can be rejected as well.

Table 7: *The proportion of subjects accurately perceiving the authorship of a passage.*

	Accuracy			
	Passage type: Op-ed		Passage type: Fiction	
	(1)	(2)	(3)	(4)
Treated (Priming)	0.058	-0.042	-0.025	-0.149^{***}
	(0.039)	(0.055)	(0.038)	(0.054)
Op-ed type: AI	-0.042	-0.144^{***}		
	(0.039)	(0.055)		
Treated * Op-ed type: AI		0.202^{***}		
		(0.077)		
Fiction type: AI			-0.148^{***}	-0.273^{***}
			(0.038)	(0.053)
Treated * Fiction type: AI				0.248^{***}
				(0.076)
Constant	0.495^{***}	0.542^{***}	0.569^{***}	0.631^{***}
	(0.033)	(0.038)	(0.033)	(0.037)
Observations	667	667	667	667
R^2	0.005	0.015	0.023	0.038
Adjusted R^2	0.002	0.010	0.020	0.034
Residual Std. Error	0.500 (df = 664)	0.498 (df = 663)	0.495 (df = 664)	0.492 (df = 663)
F Statistic	1.603 (df = 2; 664)	3.347^{**} (df = 3; 663)	7.686^{***} (df = 2; 664)	8.729^{***} (df = 3; 663)
Note:				*p<0.1; **p<0.05; ***p<0.01

For further validation of the estimate's specifics, we created two additional models: the ATE on accuracy for AI-generated op-eds and the ATE on accuracy for human-written op-eds (Table 8). The treatment increased the percentage of subjects who accurately discerned the authorship of an AI-generated op-ed passage by 15.9% with 95% CI [5.32%, 26.48%], (Column (2)). This alteration is notably significant, supported by a standard error of 0.05. Conversely, there is no significant ATE on accuracy detected for human-written op-eds (Column (1)).

Table 8 (left): *The proportion of subjects accurately perceiving the authorship of an op-ed passage, categorized by the author.*

Table 9 (right): *The proportion of subjects accurately perceiving the authorship of a fiction passage, categorized by the author.*

	Accuracy - op-ed only	
	Author: Human	Author: AI
	(1)	(2)
Treated (Priming)	-0.042	0.159***
	(0.055)	(0.054)
Constant	0.542***	0.399***
	(0.038)	(0.040)
Observations	333	334
R^2	0.002	0.025
Adjusted R^2	-0.001	0.022
Residual Std. Error	0.501 (df = 331)	0.495 (df = 332)
F Statistic	0.594 (df = 1; 331)	8.592*** (df = 1; 332)
Note:		*p<0.1; **p<0.05; ***p<0.01

	Accuracy - fiction only	
	Author: Human	Author: AI
	(1)	(2)
Treated (Priming)	-0.149***	0.099*
	(0.054)	(0.054)
Constant	0.631***	0.358***
	(0.037)	(0.038)
Observations	332	335
R^2	0.023	0.010
Adjusted R^2	0.020	0.007
Residual Std. Error	0.493 (df = 330)	0.491 (df = 333)
F Statistic	7.619*** (df = 1; 330)	3.380* (df = 1; 333)
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 9 portrays the models used to estimate the ATEs on accuracy for both AI-generated and human-written fiction. It is calculated that the treatment decreases the percentage of subjects who accurately identify the authorship of a human-written fiction passage by 14.9%, 95% CI [4.32%, 25.48%] (Column (1)). This observed change is statistically significant, with a standard error of 0.05. Moreover, the treatment is projected to raise the percentage of subjects correctly attributing the authorship of an AI-generated fiction passage by 9.9%, although the p-value associated with this estimate surpasses 0.05 (Column (2)).

This analysis indicates that priming appears to have varying effects on accuracy based on the authorship of the passage subjects read. In the case of op-ed passages, primed subjects exhibit a significant enhancement in accuracy when identifying AI-generated content. On the contrary, when it comes to fiction passages, primed subjects are projected to encounter notable difficulty in accurately identifying passages authored by humans.

Column (2) and Column (4) of following Table 10 present the models enriched with demographic covariates of subjects, encompassing factors such as gender, age group, educational history, and occupation background. Notably, although some segments exhibit a statistically significant association with the outcome variable, these findings warrant cautious interpretation due to its limited statistical power. In most cases, the p-value remains above 0.05 or the segment consists of a very limited number of subjects; for example, the age group “60-69” has a mere 87 subjects).

Table 10: *The proportion of subjects accurately perceiving the authorship of a passage by demographics.*

	Accuracy			
	Passage type: Op-ed		Passage type: Fiction	
	(1)	(2)	(3)	(4)
Treated (Priming)	0.055	0.059	-0.028	-0.032
	(0.039)	(0.039)	(0.039)	(0.039)
Gender: Male		0.030		0.015
		(0.042)		(0.042)
Gender: Non-binary / Third gender		-0.275^{*}		0.039
		(0.147)		(0.183)
Gender: Prefer not to say		0.392^{***}		0.266
		(0.100)		(0.325)
Age: 23-29		0.064		0.110
		(0.079)		(0.079)
Age: 30-39		0.068		0.048
		(0.077)		(0.076)
Age: 40-49		0.077		0.008
		(0.080)		(0.079)
Age: 50-59		-0.031		0.055
		(0.080)		(0.079)
Age: 60-69		-0.063		0.158^{*}
		(0.084)		(0.084)
Age: 70+		-0.026		0.081
		(0.099)		(0.098)
Age: Prefer not to say		0.389^{**}		-0.079
		(0.178)		(0.322)
High education		0.015		0.001
		(0.041)		(0.042)
Tech-related occupation		-0.067		0.051
		(0.047)		(0.048)
Constant	0.476^{***}	0.453^{***}	0.497^{***}	0.411^{***}
	(0.028)	(0.071)	(0.028)	(0.070)
Observations	667	667	667	667
R^2	0.003	0.024	0.001	0.013
Adjusted R^2	0.002	0.005	-0.001	-0.006
Residual Std. Error	0.500 (df = 665)	0.499 (df = 653)	0.500 (df = 665)	0.502 (df = 653)
F Statistic	2.047 (df = 1; 665)	1.243 (df = 13; 653)	0.527 (df = 1; 665)	0.673 (df = 13; 653)
Note:				*p<0.1; **p<0.05; ***p<0.01

Our survey contained additional questions on answer confidence, rating writing quality or effectiveness, and other features. The majority of these effects were not significant, but for more detail, tables on some of the results can be found in the appendix.

Discussion

We found that priming had a higher effect on responding that something was AI-generated than anticipated, roughly 10.1% for op-eds presented and 12.3% for fiction pieces presented. In terms of practical significance, for 100 people, 10 of those people would move towards declaring writing AI-generated in response to an op-ed or potentially other "factual" piece, while 12 people would declare writing AI-generated for a fiction or potentially other "creative" piece.

Priming did not seem to have an effect on accuracy overall in correctly identifying a human or AI writer. However, in line with the effects of priming on the AI rate, that increase in the AI rate pushes people to answer both more and less accurately, depending on the type of piece they read. Treated subjects became more accurate when reading an AI-generated op-ed by 15.95%, while they were less accurate when reading a human-written fiction piece by 14.9%. Secondarily, we found people are not particularly accurate at detecting AI to begin with; the average accuracy across control and treatment groups was in the range of 47-52%.

These results suggest that there is likely an effect from the proliferation of AI-focused news on the general public. People may now be more likely to believe content is AI-generated if they know large language models (LLMs) and text generators exist. Subsequently, people may treat text more critically with additional skepticism, especially as current LLMs do not necessarily include any fact-checking when generating text. On the other hand, we may see more extreme examples such as the professor who wanted to fail all his students due to erroneously believing their work was AI-generated. Priming, or knowledge on the capacity of AI, can have real-world implications in how people interact with and react to written work.

Limitations

While our results found that priming was statistically significant in changing the rate at which people would believe writing was AI-generated, our experiment has certain limitations.

For this experiment, we chose to present excerpts between 100-300 words in order to maximize subject compliance and participation. At this time, AI tools struggle with longer form writing. It is likely that the longer the piece we provided to people, the more easily it would be to detect that writing is AI-generated as opposed to human-written. We would not expect our results to hold for varying lengths; likely the longer the passage, the less we would expect people to believe a human-written piece would be AI-generated, regardless of priming.

Additionally, we used only one large language model, ChatGPT by OpenAI, to generate text for this experiment. Currently, this is one of the most well-known and possibly best performing text generators, but the technology is still limited. ChatGPT was trained through scraping publicly available web text and not trained in specific content. This means that the pieces generated through ChatGPT are not necessarily comparable to pieces written by an human expert, especially if pieces were to be longer, more complex, and reliant on specific facts or contextual knowledge outside of ChatGPT's training date range. It is possible that if subjects were topic experts for the opinion piece they were assigned, they could perceive the lack of expert knowledge or otherwise better detect AI-generation.

Using ChatGPT had additional limitations, particularly that fiction pieces were likely less comparable to their human counterparts than the op-ed AI versions. We avoided giving ChatGPT very specific and detailed prompts to avoid producing only the "best" that the tool could generate, but instead focused on what was easily produced or feasible for someone casually working with the tool. At this time, the technology does not seem as good at reproducing human written "creative" work, especially if the piece is less conventional, complex in tone or language, or unexpected in content, than more "factual" or professionally written in nature. Even when attempting to prompt ChatGPT more specifically, the tool generated very similar results utilizing similar language or literary style conventions and story structures.

Beyond the technology to generate the text, there are always limitations regarding surveys. While giving people text online via a survey was the easiest way to accomplish this experiment with the budget we had, we have limited guarantee that subjects seriously read the texts we provided. We attempted to address this issue by including screener questions to exclude random answering and a basic comprehension question on our assignment passage. Roughly 7% of subjects answered incorrectly on our comprehension question or compliance check in both treatment and control. Additionally, we implemented minimum time spent for each passage section, with respondents spending an average time of 4.79 minutes to complete the survey and an average of roughly one minute to read each passage and answer the following questions.

Our sample size also limited our ability to make strong conclusions on subgroups. In some of the narrowest buckets, such as those older than 70, or for the specific passage given, we had less than 100 subjects for that subgroup. We would need a larger sample size to have conclusive results for these subgroups and whether there were true heterogeneous effects. At the same time, the number of possible covariates we could measure means we had to be sensitive to the risks of multifactor experiments, where significance could be found simply due to random chance.

Conclusion

Our study sought to address two main topics:

- Whether priming affects the percent of subjects perceiving a passage as generated by AI
- Whether priming influences the accuracy of subjects' in identifying human or AI authorship

For the first topic, we found statistically significant evidence that priming subjects will cause them to perceive more passages as generated by AI across two different types of excerpts. This finding is consistent with previous literature on priming and our effect size was even greater than we initially estimated it would be.

For the second topic, we found no evidence that priming subjects will cause them to accurately identify whether or not a passage was AI generated. Interestingly enough, AI and specifically ChatGPT has advanced far enough to have people's accuracy hover around 50% for both op-ed and fiction passages. We expect as AI continues to advance this number to remain around the same.

References

Ankel, S. (2023, May 17). *A Texas professor failed more than half of his class after ChatGPT falsely claimed it wrote their papers*. Business Insider. https://www.businessinsider.com/professor-fails-students-after-chatgpt-falsely-said-it-wrote-papers-2023-5

APA Dictionary of Psychology. (n.d.). Dictionary.apa.org. https://dictionary.apa.org/priming

Clark, E. A., August, T., Serrano, S., Haduong, N., Suchin Gururangan, & Smith, N. A. (2021). *All That's "Human" Is Not Gold: Evaluating Human Evaluation of Generated Text*. https://doi.org/10.48550/arxiv.2107.00061

Dent, M. (2021, July 20). *Emerging Technologies Set to Shape Next Generation of Plant-Based Meat*. IDTechEx. https://www.idtechex.com/en/research-article/emerging-technologies-set-to-shape-next-generation-of-plant-based-meat/24262

Johnson, S. (2022, April 15). A.I. Is Mastering Language. Should We Trust What It Says? (Published 2022). *New York Times*. https://www.nytimes.com/2022/04/15/magazine/ai-language.html

OpenAI. (2023). *Transforming work and creativity with AI*. Openai.com. https://openai.com/product

Rogers, R. (2023, February 8). *How to Detect AI-Generated Text, According to Researchers*. Wired. https://www.wired.com/story/how-to-spot-generative-ai-text-chatgpt/

Young, M. E., Norman, G. R., & Humphreys, K. R. (2008). Medicine in the Popular Press: The Influence of the Media on Perceptions of Disease. *PLoS ONE*, *3*(10), e3552. https://doi.org/10.1371/journal.pone.0003552

Appendix

The data, processing, and analysis code for this experiment can be found here: https://github.com/am-chang/241experiment.

The appendix contains:

1. Supplementary Model Results
2. Passage Excerpts
3. Full Survey

Supplementary Model Results

We asked participants a number of questions related to writing quality/effectiveness. Many of these results were not statistically significant, but we have included some model results for reference, though they should be treated with skepticism as it is possible ratings were measured post-treatment and therefore possibly affected by the priming. More models were created during the analysis process, though the results were not included in this appendix.

AI Rate Models Including Interaction Effect with Quality, Persuasiveness/Interest

Table A1: *The proportion of subjects perceiving a passage as written by AI including passage quality*

	AI Rate			
	Passage type: Op-ed		Passage type: Fiction	
	(1)	(2)	(3)	(4)
Treated (Priming)	0.101^{***}	0.098	0.123^{***}	0.109
	(0.039)	(0.071)	(0.038)	(0.074)
Op-ed high quality	-0.032	-0.035		
	(0.042)	(0.060)		
Treated * Op-ed high quality		0.005		
		(0.085)		
Fiction high quality			-0.095^{**}	-0.105^{*}
			(0.043)	(0.061)
Treated * Fiction high quality				0.020
				(0.086)
Constant	0.453^{***}	0.455^{***}	0.432^{***}	0.440^{***}
	(0.040)	(0.050)	(0.041)	(0.052)
Observations	667	667	667	667
R^2	0.011	0.011	0.023	0.023
Adjusted R^2	0.008	0.007	0.020	0.018
Residual Std. Error	0.498 (df = 664)	0.498 (df = 663)	0.490 (df = 664)	0.490 (df = 663)
F Statistic	3.703^{**} (df = 2; 664)	2.466^{*} (df = 3; 663)	7.759^{***} (df = 2; 664)	5.183^{***} (df = 3; 663)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table A2: *The proportion of subjects perceiving a passage as written by AI including passage persuasiveness / interest*

	AI Rate			
	Passage type: Op-ed		Passage type: Fiction	
	(1)	(2)	(3)	(4)
Treated (Priming)	0.104***	0.133***	0.122***	0.120**
	(0.039)	(0.047)	(0.038)	(0.048)
Op-ed persuasive	-0.063	-0.015		
	(0.041)	(0.059)		
Treated * Op-ed persuasive		-0.092		
		(0.082)		
Fiction interesting			-0.093**	-0.096*
			(0.039)	(0.054)
Treated * Fiction interesting				0.005
				(0.079)
Constant	0.449***	0.435***	0.397***	0.398***
	(0.030)	(0.033)	(0.030)	(0.034)
Observations	667	667	667	667
R^2	0.014	0.015	0.024	0.024
Adjusted R^2	0.011	0.011	0.021	0.019
Residual Std. Error	0.497 (df = 664)	0.497 (df = 663)	0.490 (df = 664)	0.490 (df = 663)
F Statistic	4.593** (df = 2; 664)	3.478** (df = 3; 663)	8.022*** (df = 2; 664)	5.341*** (df = 3; 663)
Note:				*p<0.1; **p<0.05; ***p<0.01

Accuracy Models Including Interaction Effect with Quality, Persuasiveness/Interest

Table A3: *The proportion of subjects accurately perceiving the authorship of a passage including the author*

Estimated ATE on accuracy by passage quality

	Accuracy			
	Passage type: Op-ed		Passage type: Fiction	
	(1)	(2)	(3)	(4)
Treated (Priming)	0.055	0.149^{**}	-0.028	-0.097
	(0.039)	(0.071)	(0.039)	(0.074)
Op-ed high quality	-0.007	0.059		
	(0.042)	(0.060)		
Treated * Op-ed high quality		-0.132		
		(0.085)		
Fiction high quality			0.006	-0.042
			(0.043)	(0.062)
Treated * Fiction high quality				0.096
				(0.087)
Constant	0.480^{***}	0.434^{***}	0.492^{***}	0.527^{***}
	(0.040)	(0.050)	(0.042)	(0.052)
Observations	667	667	667	667
R^2	0.003	0.007	0.001	0.003
Adjusted R^2	0.0001	0.002	-0.002	-0.002
Residual Std. Error	0.500 (df = 664)	0.500 (df = 663)	0.501 (df = 664)	0.501 (df = 663)
F Statistic	1.034 (df = 2; 664)	1.494 (df = 3; 663)	0.273 (df = 2; 664)	0.587 (df = 3; 663)
Note:				$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table A4: *The proportion of subjects accurately perceiving the authorship of a passage including passage persuasiveness / interest*

Estimated ATE on AI rate by passage effectiveness

	Accuracy			
	Passage type: Op-ed		Passage type: Fiction	
	(1)	(2)	(3)	(4)
Treated (Priming)	0.058	0.112^{**}	-0.029	-0.025
	(0.039)	(0.047)	(0.039)	(0.048)
Op-ed persuasive	-0.067	0.020		
	(0.041)	(0.060)		
Treated * Op-ed persuasive		-0.166^{**}		
		(0.083)		
Fiction interesting			-0.033	-0.028
			(0.040)	(0.057)
Treated * Fiction interesting				-0.009
				(0.081)
Constant	0.496^{***}	0.470^{***}	0.509^{***}	0.507^{***}
	(0.030)	(0.033)	(0.031)	(0.035)
Observations	667	667	667	667
R^2	0.007	0.013	0.002	0.002
Adjusted R^2	0.004	0.009	-0.001	-0.003
Residual Std. Error	0.499 (df = 664)	0.498 (df = 663)	0.500 (df = 664)	0.501 (df = 663)
F Statistic	2.323^{*} (df = 2; 664)	2.910^{**} (df = 3; 663)	0.590 (df = 2; 664)	0.397 (df = 3; 663)
Note:				$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Passage Excerpts

All AI-generated passages were generated using ChatGPT version 3.5 or 4.0.

Op-ed Excerpts

Charter Schools - Human Written; Excerpted from: "Charter Schools Won't Save Education" by Gary Orfield, January 2, 1998

The charter school movement has swept the country, offering what many say is a simple, low-cost answer to the educational crisis. If bureaucracy and rigidity are to blame for failing schools, then why not contract groups of educators and businesses to run their own schools, using public money?

Well, it's not that simple. Charter schools are not the panacea their supporters make them out to be. Indeed, these schools are not well regulated and often fail to serve students or their communities fairly or well. Furthermore, the flexibility and innovation ideally offered by charter schools can be achieved with fewer risks within public school systems.

Support for charter schools comes from high places and cuts across party lines. This support is misplaced. A charter, after all, is not an educational program. It is a school that uses public money to advance a privately defined vision of education. In one school that vision may be a positive plan put into place by dedicated teachers. In another, a biased or sectarian group may have a disturbing agenda.

Of course, problems exist in public schools as well. But at least there are more systems in place to detect corruption or unapproved curriculums. Renegade charter schools would be much harder to rein in.

Although charter schools are required to obey the same Federal regulations that apply to public schools, recent research shows that some do not provide for disabled children, while others ignore the rights of students who need instruction in English as a second language.

Healthcare - Human Written; Excerpted from "American Health Care Is Broken. Major Hospitals Need to Be Part of the Solution" by Donald M. Berwick and Michelle A. Williams, 2023

Let's first look at the data: The U.S. now spends more than $4 trillion a year on health care. That's nearly 20% of gross domestic product. Yet U.S. life expectancy lags literally dozens of other nations—including Portugal, Slovenia, and Turkey—by as much as seven years. If trends continue, we will drop to 64th in the world in life expectancy by 2040, though we will continue to spend significantly more per capita than nearly any other nation.

Diagnosing this failure is not difficult. Nearly all the money we spend on health care goes to pay for medical interventions. But clinical care is responsible for at most 20% of health outcomes. The overwhelming majority of factors that determine an individual's health are embedded in the world around them: How many bus transfers they need to reach a store that sells fresh vegetables. Whether the windows in their workplace let in light and fresh, clean air. How often they face the stress and pain of discrimination because of the color of their skin.

These are the social drivers of health—and for far too long, our health care systems have largely ignored them.

Air Quality - Human Written; Excerpted from "Opinion: Air quality disasters aren't going away. Here's how we should protect ourselves" by Abraar Karan, Devabhaktuni Srikrishna, and Ranu Dhillon, 2023

Outdoor air quality is a more complicated problem the U.S. has chased for decades. The need to reduce, monitor and regulate air pollution was a fundamental driver for the Nixon administration to create the Environmental Protection Agency. Today, the EPA has a list of "criteria air pollutants" — including carbon monoxide, lead, nitrogen oxides, ground-level ozone, particulate matter and sulfur oxides — that must be maintained within permissible levels. The EPA is also authorized under the 1970 Clean Air Act to regulate the emissions of hazardous pollutants. But actual implementation falls primarily to states and federal territories. As we saw with states' pandemic responses, this inherently produces fragmented strategies with weak oversight and varying quality of safeguards.

In California, both the California Air Resources Board and the state public health department are tasked with managing outdoor air quality. During catastrophic events such as massive wildfires, the Department of Forestry and Fire Protection is also involved. That leaves multiple agencies with sometimes conflicting goals and approaches to monitor and regulate outdoor air, and our homes, schools and businesses are often ill-prepared for air quality emergencies despite their frequency.

The fragmented approach also forces individuals to rely on stop-gap measures during crises. Last week's major public health messaging on the East Coast was for folks to stay indoors. Those with that option can purchase portable air filters for as little as $250 with less than $10 in monthly energy bills, or use do-it-yourself options popularized during the pandemic that are constructed from box fans, filters and other hardware store materials. (During New York City's recent haze, HEPA purifiers went out of stock, making DIY cleaners particularly useful.) Indoor air quality meters, including those that measure common pollutants and gasses such as fine particulate matter and carbon dioxide, will be crucial as severe weather events become more frequent.

Charter Schools - AI Written; ChatGPT Prompt: Write an op-ed against charter schools in the us with a focus on the lack of regulation

In the vast mosaic of America's education system, charter schools stand out as a controversial fragment. Initially introduced in the 1990s as an alternative to traditional public schools, charters promised innovative educational methods, competitive results, and the allure of choice. However, as charter schools increasingly became a substantial part of the American educational landscape, concerns about their regulation—or lack thereof—became increasingly stark.

Charter schools, unlike their public counterparts, often function independently of the local school district. They receive government funding but enjoy considerable operational freedom, which proponents argue allows for greater creativity and innovation. But with autonomy comes responsibility, and it's here where charter schools fall critically short.

The principal issue lies within the regulation deficit. With charter schools able to operate with far fewer constraints compared to traditional public schools, they can often sidestep critical accountability measures. The vast diversity in charter school operations, in terms of their curriculum, staffing, admissions, and disciplinary policies, reveals a system that can effectively elude comprehensive and consistent oversight.

One might argue that student performance could serve as a check against malpractice. However, academic results from charter schools are decidedly mixed. A 2013 study from the Center for Research on Education Outcomes (CREDO) at Stanford University found that while some charters outperform traditional public schools, many do not, with charters producing “significantly worse” results in some cases. The stark inconsistency highlights the uneven quality of education, largely stemming from a lack of uniform standards and oversight.

Healthcare - AI Written; ChatGPT Prompt: Write an op-ed within 200 words about the US healthcare system, shedding light on its heavy spend on medical interventions and its failure to provide enough service for social drivers of health.

The U.S. healthcare system, notorious for its extravagant expenditure, paradoxically fails to deliver an equivalent impact on population health. We dedicate nearly 18% of our GDP on healthcare, surpassing every other nation, yet we struggle with poor health outcomes and stark inequities.

The heart of the issue lies in the misalignment of spending. Our system is heavily biased towards expensive medical interventions, often reactive, instead of proactively addressing social determinants of health. While we invest in cutting-edge surgeries, we overlook essential factors like safe housing, nutritious food, and social support that have a profound impact on health.

Healthcare costs are escalating, but Americans are not healthier. The reason is simple: health isn't merely the absence of disease; it's a holistic blend of physical, social, and economic well-being. We can't medicate our way out of a problem that's deeply rooted in societal structures.

It's high time we rebalance our healthcare investments to acknowledge social drivers of health. A system that values prevention as much as treatment, that addresses root causes as much as symptoms, can lead to both healthier populations and fiscal prudence. The transition won't be easy, but it's a vital step towards a more equitable, efficient healthcare system.

Air Quality - AI Written; ChatGPT Prompt: Write an op-ed focused on why air quality control is complicated in the United States

When considering environmental challenges, it is easy to underestimate the complexity and significance of air quality control. Breathable, clean air is as vital as clean water and fertile soil for life on Earth, yet the intricacies of maintaining it—particularly in the United States—are often overlooked.

The issue of air quality control in the U.S. is multifaceted, involving a delicate balance of environmental science, public policy, economics, and social equity. Unraveling this complex issue calls for an understanding of its unique factors and obstacles.

At the heart of this challenge is the geographical diversity and enormity of the United States. The topography, climate, and regional industrial activities all influence local air quality, making it a considerable task to develop a 'one size fits all' strategy. What works in the dry, hot deserts of Arizona may not work in the humid, cool forests of Washington. This necessitates a regionally adaptive approach that takes into account the distinctive attributes of each area.

The U.S. also grapples with the fact that pollutants don't respect state or national borders. Emissions produced in one state can affect air quality in another, leading to interstate and even international disputes over responsibility and remediation. This spatial dilemma requires robust cooperation between states and nations, something that's often challenging to achieve given the different economic and political interests involved.

Fiction Excerpts

Fiction Piece 1 - Human Written; Excerpted from "Hunger" by Lan Samanatha Chang

Ruth grew tall and almost painfully slender; in the sixth grade, she began to practice on Tian's full-sized violin. At fourteen, she won a city-wide competition for young musicians. As the winner, she would perform a concerto, accompanied by the orchestra that had sponsored the competition. Tian had chosen the Sibelius concerto, and for many, many months he and Ruth practiced those wondrous loops of sound, the joyous slithering movements of Sibelius.

He treated her as cruelly as he did himself - with complete disregard for her age and temperament… She would play and sob for hours, but emerge from the room without a word. She believed, she had known from infancy, that she held him in her hands. Now he had replaced his tenderness with this stern passion and she followed him there, believing the source of his sternness lay in love.

Fiction Piece 2 - Human Written; Excerpted from "The Visitor" by Lydia Davis

Sometime in the early summer, a stranger will come and take up residence in our house. Although we have not met him, we know he will be bald, incontinent, speechless, and nearly completely unable to help himself. We don't know exactly how long he will stay, relying entirely on us for food, clothing, and shelter.

Our situation reminds me that a leathery-skinned old Indian gentleman once spent several months with my sister in London. At first he slept in a tent in her back yard. Then he moved into the house. Here he made it his project to rearrange the many books in the house, which were in no particular order. He decided upon categories—mystery, history, fiction—and surrounded himself with clouds of smoke from his cigarettes as he worked. He explained his system in correct but halting English to anyone who came into the room. Several years later he died suddenly and painfully in a London hospital. For religious reasons, he had refused all treatment.

This Indian visitor of my sister's also reminds me of another old man—the very old father of a friend of mine. He had once been a professor of economics. He was old and deaf even when my friend was a child. Later he could not contain his urine, laughed wildly and soundlessly during his daughter's wedding, and when asked to say a few words rose trembling and spoke about Communism. This man is now in a nursing home. My friend says he is smaller every year.

Like my friend's father, our visitor will have to be bathed by us, and will not use the toilet. We have appointed a small, sunny room for him next to ours, where we will be able to hear him if he needs help during the night. Some day, he may repay us for all the trouble we will go to, but we don't really expect it. Although we have not yet met him, he is one of the few people in the world for whom we would willingly sacrifice almost anything.

Fiction Piece 3 - Human Written; Excerpted from *Stranger* by Sherwood Smith and Rachel Manija Brown

Something flickered at the edge of sight. He jammed his heel into the dirt to stop himself, scanning warily. Stone. Dust. A hardy sprig of tarweed fluttering in the breeze. Maybe that had been it.

A black claw slashed at his eyes, its serrated edges glinting with oily poison. He threw himself backward. A segmented leg emerged from a shadowy fissure; then a large, black-furred tarantula squeezed out and landed with a thump, sending up a puff of dust. Its mandibles, as long as the blades of Ross's knives, clicked together at knee height as the spider lunged at him.

Ross snatched up a loose piece of granite. No point wasting one of his precious daggers. The throw hit the tarantula in its furry abdomen. It curled up, chittering angrily.

He edged past, then picked up speed until the gully curved ahead, out of sight. When he reached the rocky outcropping, gravel and dry weeds crunched under his feet.

Fiction Piece 4 - AI Written; ChatGPT Prompt: Write a story with less than 300 words about a person who grew up in a family whose parents or other members are dominant.

Amelia grew up in the shadow of her parents' brilliance. Her father was a renowned surgeon, her mother, a celebrated physicist, and her older siblings, champions in their respective fields. All of them stood tall, their voices firm and resounding. In their world of assertive conversations and challenging ideas, Amelia was a quiet echo.

Amelia was more of a dreamer, a solitary spirit often lost in her own world of art and imagination. In a family that thrived on logic and empirical evidence, her intuitive, creative nature often felt out of place. She yearned for acceptance, for her voice to be heard amidst the powerful symphony that was her family.

With time, Amelia found her canvas of expression in music. She channeled her thoughts and feelings into melodies, her gentle voice a contrast to the harsh timbre of intellectual debate. As her fingers danced across the piano keys, she painted a world that transcended the dominance of her family's intense logic.

Fiction Piece 5 - AI Written; ChatGPT-Prompt: Write a 300 word short story about caring for an incontinent family member.

Sarah sat in the dimly lit room, her eyes tired and heavy. The faint sound of raindrops tapping on the windowpane echoed the rhythm of her exhausted heart. It had been three long years since her beloved grandmother, Margaret, had been diagnosed with dementia. The disease had robbed Margaret of her memories, her lucidity, and her ability to care for herself.

As Sarah glanced at Margaret, now frail and vulnerable, she couldn't help but feel a deep sense of responsibility and love. The task of caring for an incontinent family member was not easy, but Sarah was determined to provide the best care she could.

Every day, Sarah would wake up at the crack of dawn to assist Margaret with her morning routine. With gentle hands and a tender smile, she would help her grandmother dress, bathe, and perform all the necessary tasks she could no longer manage on her own. The rituals had become routine, and Sarah's love for her grandmother gave her the strength to carry on.

The challenges of dealing with incontinence were both physical and emotional. Sarah would clean up accidents, change bed sheets, and help Margaret maintain her dignity throughout it all. There were moments when frustration threatened to overwhelm her, but Sarah reminded herself that Margaret was not to blame. It was the disease that had taken away her control.

Fiction Piece 6 - AI Written; ChatGPT Prompt: Write a page from a young adult novel where a character is attacked by a giant spider.

Suddenly, there was a rustling in the trees above us. I looked up, straining my eyes to penetrate the darkness. It was probably just the wind, or a squirrel. Or so I hoped.

"I think we should head back," I suggested. My voice trembled, betraying my attempt at nonchalance.

Zach laughed. "Afraid, Riley?" He teased, his voice echoing in the eerie silence.

Just then, there was a louder rustling, and then, something large and monstrous plummeted from the treetops. Its arrival was heralded by a thud that vibrated through the forest floor.

My heart seemed to stop as I saw the monstrous silhouette. Eight spindly, hairy legs, a body larger than any living thing had the right to be, and two beady eyes reflecting the scant moonlight. A spider, but not just any spider, a giant spider. It was the size of a small car, black and glossy, with fangs that glistened threateningly.

Screaming, I stumbled backward, nearly tripping over a gnarled root. The spider advanced, its bulbous body heaving, the chittering noise it made sending chills down my spine.

"Riley, run!" Zach shouted, but my feet felt as if they were glued to the ground. Terror was a potent adhesive.

The spider lunged, its massive body launching forward with an alarming speed. Its fangs sunk into the tree trunk where I had been standing just seconds ago. Adrenaline kicked in, and I darted aside, barely missing the creature's attack.

Full Survey

An asterisk indicates the question or answer was shown in random order.

Section 1: Demographic Questions

First, please answer the following demographic questions.

1. Which age group do you belong to?*
    a. Under 18
    b. 18-22
    c. 23-29
    d. 30-39
    e. 40-49
    f. 50-59
    g. 60-69
    h. 70+
    i. Prefer not to say
2. What is your gender?*
    a. Male
    b. Female
    c. Non-binary/third gender
    d. Prefer not to say
3. What is the highest level of education you have completed?*
    a. Didn't complete high school / receive GED
    b. High school diploma / GED
    c. Trade school/certification program
    d. Some college/university
    e. College/university degree
    f. Some graduate school
    g. Graduate school degree or higher
    h. Prefer not to say
4. Are you a native English speaker?*
    a. Yes
    b. No
5. How many books (any genre) in English do you read per year?*
    a. (Open-ended response, numeric value > 0 only)

(This question was not randomized but shown last.)

6. Do you currently work in or have you previously worked in the technology sector?
    a. Yes
    b. No

Section 2: Screener

7. What is 12 x 2 (enter as a number)?

   a. *(Open-ended response, no text/integer validation: if responder put in a value anything other than 24, the survey was terminated.)*

Section 3: Assignment

In this survey you will be tasked with reading 3 passages and then answering some questions about each.

Please read the following excerpt and then answer the subsequent question about it.

(*Respondents were randomly assigned to treatment or control. Treatment was shown the treatment passage only, while control was shown the control passage only.)*

3A. Treatment Branch

(Treatment Passage)

8. What was the previous text about?
   a. Generative AI
   b. Plant-based meat
   c. Drones
   d. Solar panels

3B. Control Branch

(Placebo Passage)

9. What was the previous text about?
   a. Cars
   b. Drones
   c. Plant-based meat
   d. Solar panels

Section 4 Evaluation Text: Op-Ed

Please read the following excerpt and then answer the subsequent questions about it.

(*Respondents were randomly shown one of six op-ed pieces.)*

10. Please rate how well/poorly written the above passage is.*
    a. (Likert scale of 1 - 10, where 1 represents worst quality and 10 best quality)
11. How persuasive are the arguments presented?*
    a. Not persuasive at all
    b. Slightly persuasive
    c. Moderately persuasive
    d. Very persuasive
    e. Extremely persuasive

Section 5 Evaluation Text: Fiction

Please read the following excerpt and then answer the subsequent questions about it.

(*Respondents were randomly shown one of six fiction pieces.)*

12. Please rate how well/poorly written the above passage is.*
    a. *(Likert scale of 1 - 10, where 1 represents worst quality and 10 best quality)*
13. How interested would you be in reading the full story associated with this excerpt?*
    a. Not interested at all
    b. Slightly interested
    c. Moderately interested
    d. Very interested
    e. Extremely interested

Section 6 AI Check

14. Do you think the passage about (*op-ed passage shown topic)* was:*
    a. AI-generated*
    b. Human-written*
15. How confident are you in your answer?
    a. Not confident at all
    b. Slightly confident
    c. Moderately confident
    d. Very confident
    e. Extremely confident
16. Do you think the passage about (*fiction passage shown topic)* was:*
    a. AI-generated*
    b. Human-written*
17. How confident are you in your answer?
    a. Not confident at all
    b. Slightly confident
    c. Moderately confident
    d. Very confident
    e. Extremely confident