

# What Does Google Know? Plagiarism, ChatGPT, and Higher Education

Nicole Oster  
Punya Mishra

## Abstract

This quantitative study analyzed Google Trends data to examine recent interest in ChatGPT, plagiarism, and artificial intelligence (AI) within higher education. This study collected data on the search volumes from Google Trends for the terms "ChatGPT," "plagiarism checker," and "AI," focusing on cities with universities, to understand how interest in these topics has evolved over time and whether the factors of university enrollment, rank, and AI policies correlate with search trends. The analysis utilized descriptive statistics, linear regression models, and T-tests. The findings revealed a spike in interest for ChatGPT and related terms aligning with ChatGPT's public release. Cities with a higher percentage of undergraduate students and higher-ranked universities showed more interest in ChatGPT and AI. However, university policies (regarding the allowance or banning of AI) did not significantly correlate with search frequencies. The results suggest that as generative AI tools like ChatGPT are rapidly garnering interest and use in academic contexts, university policies might transition from allowing or banning the usage of these tools towards helping students understand how they can be used responsibly, ethically, and effectively.

*Key Words:* ChatGPT, artificial intelligence, plagiarism, Google Trends, higher education

## Introduction

The emergence of ChatGPT and generative AI has sparked significant attention and controversy in higher education (Author, 2023; Glaser, 2023; Cardona et al., 2023). Educators debate the ethical implications of these tools, ranging from issues of intellectual property and equitable access to bias (Cardona et al., 2023; Li et al., 2023; Baidoo-Anu & Ansah, 2023). Of widespread concern is academic integrity (Plata et al, 2023). Some educators fear that students will plagiarize and cheat, passing off the writing of generative AI as their own (Li et al., 2023). When examining ethical issues such as plagiarism, it may be difficult to ensure truthful responses, as students would not want to be accused of violating academic integrity. Although asking students directly garners questionable results, examining their Google searches may afford more accurate responses (Stephens-Davidowitz, 2014). Therefore, this analysis leverages Google Trends, which provides a large volume of anonymous data on the search patterns of users over time and across various contexts and locations. We look at how often people searched terms related to ChatGPT and plagiarism with a specific focus on cities with universities. Specifically, we seek correlations of factors such as university enrollment, university rank, and university policies to examine how interest in AI technologies like ChatGPT may vary based on the academic context.

## Literature Review

Since its release in November 2022, the generative AI tool ChatGPT has rapidly gained popularity and generated controversy. ChatGPT, a Large Language Model (LLM) generated by OpenAI, is a potentially transformative tool for the field of education with the capability to synthesize large

quantities of texts and offer customized learning experiences (Glaser, 2023). Open AI's LLM has been used to support a range of learning tools, including MagicSchool, Eduaide, Khanmigo, and EduChat, for lesson planning, student assessment, and emotional support (Extance, 2023).

This rapidly developing tool has major implications for educational practitioners. As the authors (2023) argue, educators must blend technology, pedagogy, and content knowledge, especially when using AI such as ChatGPT. Specifically, the authors note that ChatGPT's generative and social features necessitate updated teaching approaches and call for rethinking the role of AI in education to be more of a collaborative counterpart than merely a tool, acknowledging both the advantages and the challenges that it presents, such as misinformation and educational inequities (Author, 2023). In a discourse analysis conducted on Twitter conversations, teachers were found to have a generally positive outlook on ChatGPT as well as significant concerns, including "academic integrity, impact on learning outcomes and skill development, limitation of capabilities, policy and social concerns, and workforce challenges" (Li et al., 2023, p. 1). Knowing that ChatGPT is capable of rapidly generating large amounts of authoritative-sounding text, Li et al. (2023) point out the concerns that educational stakeholders have about academic integrity.

Given these concerns from the educators and researchers, it becomes important to know if there was indeed a greater interest in issues related to academic integrity after the release of ChatGPT (and other generative AI tools). This analysis leverages Google Trends data to examine the current interest in ChatGPT and related constructs plagiarism and AI among educational stakeholders. Analyzing Google Trends can provide powerful insights into public phenomena.

Google Trends provides normalized data on Google search terms. Google Trends allows for the comparison of the frequency of search terms over time and compared to other search terms. In addition, Google Trends allows users to filter search results based on various categories, such as education as well as by geographic location. In this study, Google Trends allowed us to observe the frequency of education-related searches in different cities for terms like ChatGPT, plagiarism checkers, and AI. This measure allows data to be precisely linked to geographic locations, ranging from national to city levels. In this analysis, it is considered as a measure of interest or concern in these topics within those geographic locations.

For instance, Stephens-Davidowitz (2014) employed Google search data as a measure of the geographical distribution of racial animus in the United States. This method was beneficial, given that in addition to being able to gain authentic access to ethically sensitive information, Stephens-Davidowitz reports that their sample was "1.5 to 3 times larger than survey-based estimates" (2014, p. 26). While measuring racial bias may differ from measuring interest in technologies, both issues require a methodology that is sensitive to the ethically charged nature of both racism and plagiarism. The analysis described in this paper makes similar use of Google Trends data to gather a large volume of authentic information about people in the field of education's interests in generative AI and academic integrity.

### **Research Questions**

Specifically, we use Google Trends data to answer 4 questions. To gain a basic understanding of interest in ChatGPT and related terms, we ask, 1) How have the rates of education-related Google searches for *ChatGPT*, *plagiarism checkers*, and *AI* evolved from November 2018 to November 2023? Moreover, we were interested in whether there were any correlations between the term *ChatGPT* and the terms *plagiarism* and *AI*, particularly in cities with major universities. Therefore, we ask, 2) How much do Google searches for *ChatGPT* correlate with Google searches for *plagiarism checkers* in cities with ranked universities? As a follow up, we ask, 3) How much do Google searches for *ChatGPT* correlate

with Google searches for *AI* in cities with ranked universities? Finally, we ask, 4) How much do university characteristics (university enrollment, university rank, and university policies) correlate with Google searches for ChatGPT, plagiarism, and AI in these cities? Our method is described in greater detail below.

## Method

### Sampling

The study utilized normalized data from Google Trends, focusing on searches related to ChatGPT, plagiarism checkers, and AI within United States cities with universities. The sample included data from 95 universities (64 public and 31 private) across 33 states and Washington D.C. gathered from an existing list of top universities with policies about generative AI (Caulfield, 2023). Five universities were excluded from an initial existing list (Caulfield, 2023) of 100 universities because there was incomplete information for any of the other independent or dependent variables.

The time periods for data collection were from November 20, 2018 to November 20, 2023 for Research Question 1 and from November 1, 2022 to September 20, 2023 for Research Questions 2-4. These timeframes were selected to gain an understanding of how interest in generative AI and academic integrity have changed over time, specifically following the arrival of ChatGPT in November of 2022. To answer Research Question 1, we examined data across a longer period than Research Questions 2-4 to observe what rates and patterns of searches for these terms were both before and after the release of ChatGPT. We chose to observe this over five years to account for trends before COVID-19 in case this factor impacted searches. For Research Questions 2-4, the dates range from the month of the release of ChatGPT to the date of the initial data collection.

The sampling method was non-probabilistic convenience sampling that relied on publicly available data from Google Trends and additional existing data sources (Caulfield, 2023; The Best National Universities in America, n.d.; US Census Bureau, n.d.). This method was chosen due to its accessibility, relevance, breadth, and authenticity in measuring public interest and trends in education-related topics over time.

### Measures

Measures in this analysis include the (normalized) number of Google searches for “ChatGPT,” “plagiarism checker,” “AI,” and “shirt” and university factors.

#### ***Google Searches for “ChatGPT,” “Plagiarism Checker,” “AI,” and “Shirt”***

Google searches were used to measure interest in the terms “ChatGPT,” “plagiarism checker,” and “AI.” The search term “plagiarism checker” rather than “plagiarism” is used to measure interest in plagiarism and academic integrity as it corresponds to the use or interest in a specific tool for which users might actually search rather than plagiarism as an idea more broadly. Moreover, searches for the term “shirt” are measured to control for search volume. This measure is similar to the control variable “sweater” used by Stephens-Davidowitz (2014) but may be less likely to be influenced by variability in weather.

#### ***University-Related Factors***

**Percentage of Undergraduate Students in a City.** This serves as a measure of the student population's density in a city and is used to assess its correlation with searches for ChatGPT, plagiarism checkers, and AI. This measurement is calculated by comparing the enrollment of undergraduate students in each university in the analysis to the population of each city.

**University Rank.** This refers to the standing of universities in a city (ranging from 1-236), analyzed to see if there is any correlation with the frequency of specified Google searches.

**University Policies.** This includes policies regarding the use of AI writing tools, whether they are allowed, banned, or left to individual instructors' discretion. The study investigates how these policies might influence the search patterns for ChatGPT, plagiarism checkers, and AI.

### **Construction of Measures**

Data were collected from various sources. The primary source of data was Google Trends, a tool that provides data on the popularity of specific search terms over time. The geographical focus was on cities with “top” (Caulfield, 2023) universities in the United States. Using Google Trends filter options to narrow down the search criteria, the data pertained to the category of Education, ensuring relevance to the research questions. The periods for data collection were from November 20, 2018 to November 20, 2023 (Research Question 1) and from November 1, 2022 (the month that ChatGPT arrived) to September 20, 2023 (Research Questions 2-4), capturing a broad range of recent trends. Specific search terms included "ChatGPT," "plagiarism checker," "AI," and "shirt" (as a control term to provide a comparative baseline).

Policies regarding the use of AI writing tools in universities were examined, sourced from an existing report (Caulfield, 2023). These policies were categorized by the report as allowed with citation (unless disallowed by instructors), decisions left to individual instructors, banned by default (with possible individual instructor allowance), or cases where there was no clear guidance/policy. University rank and enrollment data were sourced from a list of national universities (The Best National Universities in America, n.d.). Data on city populations and the percentage of undergraduate students in each city were obtained from the US Census Bureau (n.d.). This information was used to understand the demographic context of the search data, particularly the influence of student populations on search trends.

### **Results**

The study aimed to quantitatively assess how interest in ChatGPT, plagiarism checkers, and AI has changed over time and how various factors related to higher education institutions might correlate with this interest. The use of Google Trends data provided an authentic way to gauge public interest and sentiment at a large scale, while the application of linear regression and T-tests offered a means to explore potential correlations within the data. The data were analyzed through a descriptive analysis from Google Trends (RQ 1) and by conducting correlational analyses by running T-tests and linear regressions using the programming language and software R (RQs 2-4).

#### ***Research Question 1: How have the rates of Google searches for ChatGPT, plagiarism checkers, and AI evolved from November 2018 to November 2023 in the Google search category of education?***

For RQ 1, we conducted a descriptive analysis of Google Trends data for "ChatGPT," "plagiarism checker," "AI" (the dependent variables), and "shirt" (the control variable) over a time period, from November 20, 2018 to November 20, 2023, to see the evolution of interest in these topics. The analysis included examining trends, patterns, and fluctuations in search frequencies over time. This involved using Google Trends to analyze the evolution of interest in ChatGPT, plagiarism, and AI over time within the field of education. The analysis included examining trends, patterns, and fluctuations in search frequencies for these terms.

#### **Figure 1**

*ChatGPT, Plagiarism Checker, AI, and Shirt Google Trends*

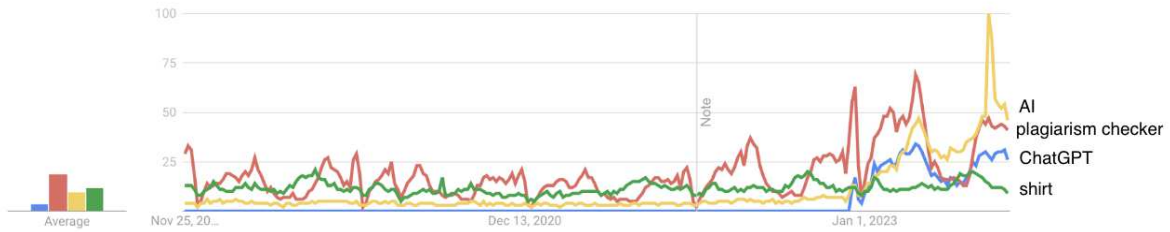


Figure 1 illustrates the evolution of searches for “shirt,” “ChatGPT,” “plagiarism checker,” and “AI” over time. To better understand the shifts in search volumes, we compare the search term “shirt” (green line) as a control variable to the other search terms. As Figure 1 shows, this line remains reasonably flat with minor ups and downs over the time period under study. The search term “ChatGPT” (blue line) shows a significant spike in searches, especially noticeable after November of 2022, which suggests a surge in interest for this term. This trend corresponds with the public release of ChatGPT, perhaps leading to increased public awareness or interest.

The search term “plagiarism checker” (red line) exhibits several peaks and valleys over time. The peaks occur in late April to early May and early December of each year. The valleys occur in late July and early August and late December and early January indicating interest levels that peak and then decline at the end of each academic semester. These fluctuations correspond with academic cycles, such as term start and end dates, when students and educators might more frequently conduct plagiarism checking. The peaks for this search term increased in November of 2022, which suggests that there was increased interest in “plagiarism checkers” around the time that “ChatGPT” emerged. The search term “AI” (yellow line) also shows variability over time with several peaks. The overall trend appears to be an upward trajectory in searches, reflecting growing interest in artificial intelligence. The peaks might correspond with notable events in the AI field, such as the arrival of ChatGPT in November of 2022.

Following this descriptive analysis of interest in ChatGPT, interest in plagiarism checkers, and interest in AI over time, this study also leverages statistical analysis to explore any correlations between these variables. Clearly, there are greater levels of searches for terms like ChatGPT, plagiarism checkers, and AI over time. However, it is unclear if this increased interest connects specifically with education. This leads us to Research Question 2, where we explore correlations between these variables to understand from where the searches originated.

***Research Question 2: How much do Google searches for ChatGPT correlate with Google searches for plagiarism checkers in cities with ranked universities?***

For RQ 2, we utilized a linear regression and T-tests to explore correlations between the interest in plagiarism checkers and interest in ChatGPT in cities with ranked universities. We subtracted the values for the control variable, the search term “shirt,” from the independent variable, the search term “ChatGPT,” to adjust for search volume. The dependent variable for RQ 2 was the search term “plagiarism checker” adjusted for the control variable, the search term “shirt.”

**Table 1**

*Searches for Plagiarism Checker and ChatGPT*

Searches for Plagiarism Checker
---------------------------------

Term	Estimate	Std. Error	t value	Pr(> t )
Searches for ChatGPT	0.43967	0.08553	5.14	1.51e-06 ***

Table 1 presents the findings related to the correlation between searches for ChatGPT and plagiarism checkers in city areas ( $b = 0.43967$ ). These findings suggest a significant ( $p = 1.51e-06$ ) positive correlation between the frequency of searches for ChatGPT plagiarism checkers in various city areas. This implies that as interest in ChatGPT increases in a particular city, there is a corresponding increase in the searchers for plagiarism-checking tools, possibly indicating heightened awareness or concern about plagiarism in the context of ChatGPT's growing popularity.

After exploring correlations between ChatGPT and plagiarism checkers, we examined whether interest in ChatGPT correlates with interest AI technologies more broadly in the university context.

***Research Question 3: How much do Google searches for ChatGPT correlate with Google searches for AI in cities with ranked universities?***

We used linear regression and T-tests to investigate correlations between interest in AI and interest in ChatGPT in cities with ranked universities using R. To adjust for search volume, we subtracted the values for the control variable, the search term “shirt,” from the independent variable, the search term “ChatGPT.” The dependent variable for RQ 3 was the search term “AI,” adjusted for the control variable, the search term “shirt.”

**Table 2**

*Searches for AI and ChatGPT*

Searches for AI				
Term	Estimate	Std. Error	t value	Pr(> t )
Searches for ChatGPT	0.70884	0.05045	14.05	< 2e-16 ***

The analysis in Table 2 indicates a strong positive ( $b = 0.70884$ ) and significant ( $p = < 2e-16$ ) correlation between searches for ChatGPT and AI in city areas. These findings imply a close relationship between the growing interest in ChatGPT and the broader field of AI in the selected city areas. The high correlation reflects the increasing relevance and impact of AI technologies, like ChatGPT, in these regions, possibly driven by educational interests.

Following analysis of the correlations between interest in plagiarism checkers, interest in ChatGPT, and interest in AI in the field of higher education, we examined the impact of specific university characteristics on interest in these terms.

***Research Question 4: How much do university characteristics (university enrollment, university rank, and university policies) correlate with Google searches for ChatGPT, plagiarism, and AI in cities with ranked universities?***

To answer RQ 4, we applied linear regressions and T-tests to check for correlations between variables using the program, R. For RQ 4, we conducted a separate analysis for each dependent variable. The dependent variables were the search terms “ChatGPT,” “plagiarism checker,” and “AI,” each adjusted for the control variable, the search term “shirt.” The independent variables were the percentage of undergraduate students in a city (calculated from the city population and undergraduate enrollment), university rankings, and university policies (whether AI tools are allowed, banned, or left to individual instructors' discretion).

**Table 3***Searches for ChatGPT*

Searches for ChatGPT				
Term	Estimate	Std. Error	t value	Pr(> t )
Percentage of undergraduate students in city	0.420541	0.094715	4.44	2.57e-05 ***
University rank	-0.00204	0.000474	-4.303	4.31e-05 ***
Individual instructors decide	0.039651	0.117011	0.339	0.736
Banned by default, individual instructors may allow	-0.039155	0.125205	-0.313	0.755
No clear guidance/policy	-0.068043	0.121651	-0.559	0.577

First, we analyzed correlations between interest in ChatGPT in a city and university characteristics. Table 3 illustrates whether there are correlations between Google Searches for ChatGPT and university rank, enrollment, and policies. The analysis of the percentage of undergraduate students in a city and ChatGPT searches indicates a moderate to strong positive ( $b = 0.42054$ ) and significant ( $p = 2.57e-05$ ) correlation. This finding suggests that cities with higher proportions of undergraduate students tend to exhibit increased searches for ChatGPT. A slight positive ( $b = -0.00204$ ; note that university rank is a descending variable) and significant ( $p = 4.31e-05$ ) correlation was observed regarding university rank. This implies that higher-ranked universities might have slightly higher search frequencies for ChatGPT. Compared to policies allowing the use of AI writing tools, there was no significant difference in searches for ChatGPT given alternative policies.

Likewise, we conducted an analysis on the correlation between interest in plagiarism checkers in cities with universities and the percentage of undergraduate students in the city, university rank, and university policies.

**Table 4***Searches for Plagiarism Checker*

Searches for Plagiarism Checker				
Term	Estimate	Std. Error	t value	Pr(> t )
Percentage of undergraduate students in city	0.091242	0.1060514	0.86	0.391904
University rank	0.0004892	0.0005308	0.922	0.359137
Individual instructors decide	0.002449	0.1401909	0.017	0.986101
Banned by default, individual instructors may allow	0.01877	0.1310159	0.143	0.886405
No clear guidance/policy	-0.1081845	0.1362108	-0.794	0.429167

The analysis of factors influencing searches for plagiarism checkers in university cities yielded the findings illustrated in Table 4. These findings suggest that none of the factors examined, including the percentage of undergraduate students in a city, university rank, or university policy, show a statistically significant correlation with the frequency of searches for plagiarism checkers in city areas.

We performed a similar analysis on the correlation between searches for AI in cities with universities and the percentage of undergraduate students in the city, university rank, and university policies.

**Table 5**

*Searches for AI*

Searches for AI				
Term	Estimate	Std. Error	t value	Pr(> t )
Percentage of undergraduate students in city	0.1905282	0.0881789	2.161	0.0334 *
University rank	-0.0018166	0.0004413	-4.116	8.58e-05 ***
Individual instructors decide	0.0606917	0.116565	0.521	0.6039
Banned by default, individual instructors may allow	0.0501935	0.1089363	0.461	0.6461
No clear guidance/policy	0.001683	0.1132557	0.015	0.9882

There is a moderate positive ( $b = 0.1905282$ ) and significant ( $p = 0.0334$ ) correlation between the percentage of undergraduate students and searches for AI. A slight positive ( $b = -0.0018166$ ; note that university rank is a descending variable) and statistically significant ( $p = 8.58e-05$ ) correlation is observed with university rank (as rank is a descending variable). This suggests that higher-ranked universities might have slightly higher search frequencies for AI.

The findings suggest that the percentage of undergraduate students in a city positively influences searches for AI and that university rank has a slight but significant correlation with AI searches. Decisions by individual instructors, specific university policies, or the absence of clear guidance do not show significant correlations with the frequency of AI searches in university cities.

### Limitations

Google Trends data does not provide access to raw data, so the analysis relied on its already normalized data, which may not fully capture the nuances of search behavior. Furthermore, the control variable "shirt" may not perfectly account for general search trends, potentially affecting the accuracy of the correlation analysis. Additionally, the approximation in the percentage of undergraduate students in city data does not account for students living outside university cities, possibly skewing the representation of student density. Finally, the universities and their corresponding cities were limited to the universities in Caulfield's report on university policies.

### Discussion

The findings offer insights into current interest in ChatGPT, plagiarism checkers, and AI within higher education with implications for university policymakers and faculty members. There is a notable increase in interest in ChatGPT, particularly in cities with higher percentages of undergraduate students or higher-ranking universities. This trend suggests a correlation between the academic environment and the curiosity, concerns, or use of AI technologies such as ChatGPT. A similar increase in interest is observed for plagiarism checkers and AI in higher education. This reflects the growing awareness and potential implications of AI tools in academic integrity and educational practices. Notably, the analysis found no significant difference in interest based on university policies towards ChatGPT, AI, or plagiarism checkers. This suggests that institutional policies may not be a primary factor of interest or concern regarding these technologies among the academic community.



These results align with the theory that technological advancements, particularly in AI, are rapidly influencing educational domains. The increased interest in ChatGPT and AI reflects the growing importance of these tools in educational settings. The lack of a significant correlation between university policies and search interests challenges the assumption that institutional directives strongly influence academic engagement with new technologies. University policymakers and instructors might consider the implications of both the consistent usage of these tools regardless of university policies as well as the generally high rate of interest and usage. Specifically, instead of creating policies governing whether students are permitted to use these tools (that may not be followed), university policymakers and faculty members might transition towards guiding students in how to use these tools responsibly, ethically, and responsibly.

Finally, this study demonstrates the value of utilizing anonymized, large-scale data sets of search queries to investigate potentially sensitive topics that may be difficult to directly study. By leveraging Google Trends data, the analysis was able to explore people's authentic interests in ethically complex issues like plagiarism. The patterns in Google searches offer insights that direct surveys or self-reports could miss. This study underscores the benefits of employing large-scale data sets to indirectly examine public interest in topics that are challenging to observe first-hand due to their ethical sensitivity.

### **Conclusion**

This analysis suggests a correlation between the academic environment (undergraduate enrollment and university ranking) and interest in AI technologies, regardless of university policies. University policies might shift from allowing or banning generative AI tools toward facilitating appropriate student use.

## References

- Author (2023). TPACK in the age of ChatGPT and Generative AI. *Journal of Digital Learning in Teacher Education*, 39(4), 1–17. <https://doi.org/10.1080/21532974.2023.2247480>
- Baidoo-Anu, D., & Ansah, L. O. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI*, 7(1), 52-62.
- The Best National Universities in America*. (n.d.). US News & World Report. Retrieved November 29, 2023, from [https://www.usnews.com/best-colleges/rankings/national-universities?schoolName=Virginia+Tech&\\_sort=rank&\\_sortDirection=asc](https://www.usnews.com/best-colleges/rankings/national-universities?schoolName=Virginia+Tech&_sort=rank&_sortDirection=asc)
- Cardona, M. A., Rodríguez, R. J., & Ishmael, K. (2023). *Artificial Intelligence and the Future of Teaching and Learning*. Office of Educational Technology.
- Caulfield, J. (2023, April 24). *University Policies on AI Writing Tools | Overview & List*. Scribbr. <https://www.scribbr.com/ai-tools/chatgpt-university-policies/>
- Extance, A. (2023). ChatGPT has entered the classroom: How LLMs could transform education. *Nature*, 623(7987), 474–477. <https://doi.org/10.1038/d41586-023-03507-3>
- Glaser, N. (2023). Exploring the Potential of ChatGPT as an Educational Technology: An Emerging Technology Report. *Technology, Knowledge and Learning*. <https://doi.org/10.1007/s10758-023-09684-4>
- Google Trends. (n.d.). Google Trends. Retrieved November 29, 2023, from <https://trends.google.com/trends/explore?cat=74&date=today%205-y&geo=US&q=plagiarism%20checker,chatgpt,ai,shirt&hl=en>
- Li, L., Ma, Z., Fan, L., Lee, S., Yu, H., & Hemphill, L. (2023). *ChatGPT in education: A discourse analysis of worries and concerns on social media* (arXiv:2305.02201). arXiv. <http://arxiv.org/abs/2305.02201>
- Plata, S., De Guzman, M. A., & Quesada, A. (2023). Emerging Research and Policy Themes on Academic Integrity in the Age of Chat GPT and Generative AI. *Asian Journal of University Education*, 19(4), 743-758.
- Stephens-Davidowitz, S. (2014). The cost of racial animus on a black candidate: Evidence using Google search data. *Journal of Public Economics*, 118, 26–40. <https://doi.org/10.1016/j.jpubeco.2014.04.010>
- US Census Bureau (n.d.). *City and Town Population Totals: 2020-2022*. Census.Gov. Retrieved November 29, 2023, from <https://www.census.gov/data/tables/time-series/demo/popest/2020s-total-cities-and-towns.html>