MICHIGAN ROSS

Ross School of Business at the University of Michigan

Independent Study Project Report

Term: Spring/Summer 2024

Course: BE 399

Professor: Andreas Hagemann

Student: Alex Zhang

Title: Impact of LLMs on Academic Literature Synthesis: Influence and Oversight in Business Economics and Other Disciplines

Impact of LLMs on Academic Literature Synthesis: Influence and Oversight in Business Economics and Other Disciplines

Sp/Su2024 BE399 Independent Study

Faculty Advisor: Andreas Hagemann
Author: Alex Zhang

1. Introduction

The increasing integration of large-language-models (LLMs), such as OpenAI's GPT-series, Anthropic's Claude, or Meta's open-source Llama into academic writing is fundamentally reshaping the landscape of scholarly communication. This study examines the impact of LLMs on word frequencies and stylistic elements across various disciplines, identifying key semantic and linguistic areas of impact. Our findings reveal that LLMs are amplifying certain terminologies, introducing non-traditional language, and potentially standardizing the presentation of research methodologies and results. These changes have significant economic and ethical implications, providing a potential competitive advantage to institutions and researchers who effectively adopt these technologies, while also raising concerns about the homogenization of academic discourse and potential overemphasis on specific ideologies or research directions, such as prioritization of positive results over negative ones.

To address these challenges, we propose a high-level framework that outlines the necessary steps for managing the competitive and ethical implications of LLM usage in academia. This framework includes recommendations for ensuring that artificial intelligence technologies enhance academic productivity while preserving the diversity and originality of scholarly work. It also emphasizes the importance of developing ethical guidelines for the responsible use of AI in research, to prevent biases and ensure a balanced representation of perspectives in academic discourse. This study underscores the need for continued research and policy development to navigate the evolving role of LLMs in shaping the future of academic communication.

1.1 Background on LLMs (Large Language Models)

The advent of advanced AI language models has sparked significant interest and concern regarding their impact on academic writing practices. These large language models are developed by training on vast amounts of text data, allowing them to generate human-like text. The development of LLMs has been marked by several iterations, with the GPT (Generative Pre-trained Transformer) models by OpenAI being particularly notable, alongside Anthropic's Claude, Google's Gemini, and open-source counterparts like Meta's Llama series. As of July 2024, Meta's newest Llama 3.1 model represents an unprecedented advancement in open-source natural language processing, achieving text generation capabilities comparable to proprietary industry models. This development highlights the increasing accessibility and sophistication of LLMs, signaling the beginning of a transformative era in the use of AI for content creation and academic research.

1.2 Prevalence and Pointers of LLMs in Academic Literature

Preliminary investigations have highlighted intriguing findings regarding the potential involvement of AI language models in academic writing. Notable indicators include the overuse of certain words and phrases commonly associated with AI-generated text, such as "delve," "unleash," "tapestry," and "mosaic," from preliminary studies in the medical field[1] . Additionally, anecdotal evidence from academic communities has raised concerns about the possibility of AI assistance in scholarly communication[2][3], with notable LLM influences in peer-review processes[4].

Understanding the impact of GPT and similar models on academic writing is crucial for several reasons. It provides valuable insights into the evolving nature of scholarly communication in the digital age, highlighting the transformative role of technology in shaping research practices and presentation. Furthermore, it can reveal the extent of AI integration into academic research, from drafting to editing, and how these tools are being utilized by scholars along the entire publication process. This awareness also informs important discussions around ethical considerations and potential biases introduced by AI, underscoring the need for responsible usage and critical evaluation of these technologies within academic settings.

1.3 Purpose and Significance of the Study

While existing studies, such as the systematic review[5] by Khalifa and Albadawy (2024), have examined the positive impacts and risks of AI in assisting academic writing, there is a critical need to better understand and characterize the widespread utilization of AI tools in writing processes. This project aims to document potential influences of LLMs on word frequencies in academic literature across various disciplines. Initially focusing on the field of Business Economics, the study will expand to adjacent disciplines to conduct comparative analysis, allowing for a broader understanding of LLMs' semantic influences on scholarly communication.

This multidisciplinary approach is essential for researchers, educators, and policymakers seeking to navigate the evolving landscape of academic writing in the digital era. By examining how LLMs impact different fields, the study aims to codify more generalizable semantic and linguistic patterns, providing a foundation for a high-level ethical and competitive framework. This framework can help guide the responsible use of AI in academia, addressing issues such as originality, authorship, and the equitable use and appropriate oversight of these technologies.

Understanding the competitive dynamics and strategic implications of AI-assisted academic writing is particularly crucial in highly competitive fields like Business Economics. Staying updated with the latest research is vital for maintaining academic reputation, securing funding, and building collaboration networks. If GPT or similar AI models significantly influence word frequencies and stylistic elements in academic literature, researchers and institutions that effectively leverage these tools may gain a competitive advantage in producing and disseminating research. This could shape the landscape of academic research, making it imperative to explore and address the associated ethical and competitive challenges. By diving into these aspects, this project seeks to provide comprehensive insights into the evolving practices of academic research and communication in various fields.

1.4 Introduction to Methods and Technology Enablements

Advancements in technology have significantly enhanced the ability to conduct big data analysis, allowing researchers to efficiently manage large datasets and perform complex analyses. In this study, we utilize a range of natural language processing (NLP) techniques to assess the impact of LLMs on academic writing. Key methods include word frequency analysis and TF-IDF (Term Frequency-Inverse Document Frequency) metrics, which help identify normalized shifts in language use and the prominence of specific terms influenced by AI intervention, among other techniques.

1.5 Related Literature

The integration of LLMs into academic writing has sparked a range of studies examining their influence on various research disciplines. Two significant Stanford studies by Liang et al. (2024) analyzed the extent of LLM usage in scientific publishing, particularly focusing on computer science, engineering, and biomedical fields[9][10]. Their findings revealed that approximately 17.5% of computer science papers and 16.9% of peer review texts contained content generated by AI. The studies highlight the growing reliance on LLMs within the academic community, particularly in STEM fields, engineering, and biomedical research. They utilized robust statistical methods, including time series analysis and the examination of adjective frequencies, to detect shifts in language use before and after 2023. Additionally, the studies highlight the ethical considerations surrounding LLM usage, noting the varying policies of academic journals on AI-assisted writing. Liang's work clearly discovers the impact of LLMs, revealing the importance of robust bibliometric and statistical methods. Moreover, in an interview with supervising professor James Zou, he mentioned a particular focus of the research on "*Nature* family journals" in an interview, underscoring a need for extending such analyses across different disciplines to uncover potential cross-disciplinary biases and trends in linguistic generation[11].

Another relevant study by Movva et al. (2024) examined trends in LLM-related research across 17,000 papers on arXiv[12]. This research emphasized the growing interdisciplinary application of LLMs, with significant increases in topics related to societal impacts, human-computer interaction, and ethics. Movva et al. also explored the differences in LLM-related publications between industry and academic institutions, offering insights into the collaborative dynamics and policy influences shaping this research area.

Our study builds on the methodologies employed in these existing studies. We utilize similar statistical techniques, such as time series and word frequency analysis, to track changes in language use across a broader range of disciplines, including the humanities and social sciences. By extending the analysis beyond STEM fields, we aim to uncover subtle biases and trends in LLM-generated content that may be unique to different academic contexts.

While Liang et al. and Movva et al. provided foundational insights into the impact of LLMs within specific fields, our research seeks to offer a more comprehensive, cross-disciplinary perspective. By examining the influence of LLMs across various academic disciplines, we can identify overarching patterns and field-specific nuances in how AI-generated content is integrated into scholarly work. This interdisciplinary approach allows us to compare and contrast the effects of LLMs in diverse research environments, potentially revealing unique challenges and opportunities in each domain.

Both studies and broader discussion raise important ethical questions regarding the use of LLMs in academic writing. Zou et al. discussed the evolving policies of journals regarding AI assistance, highlighting the need for transparency and ethical guidelines. Similarly, Movva et al. explored the societal implications of LLMs, particularly in relation to policy and collaboration dynamics. Our study contributes to these discussions by examining how disciplines may address these ethical considerations and what policies are being developed to manage the integration of LLMs. We aim to provide actionable recommendations for maintaining academic integrity while leveraging the benefits of AI technologies.

By interfacing with existing studies, our research not only applies proven techniques but also expands the scope of analysis to include a wider range of academic disciplines. This allows us to offer a more holistic understanding of the impact of LLMs on scholarly communication and means of responsible intervention, management and oversight.

2. Methodology

2.1 Overview of Data Chosen

2.1.1 Dataset Creation and Software APIs

The primary data source for this study was the SemanticScholar, selected for its extensive dataset and user-friendly interface. Through their application programming interface (API), we were able to programmatically filter research papers by field, publication date, and citation count, ensuring that our analysis focuses on relevant and impactful studies. We chose SemanticScholar over alternatives like Google Scholar due to its comprehensive coverage and accessibility. A key consideration was balancing the number of papers retrieved with the API's usage limits, allowing us to curate a dataset of 1,000 research papers per discipline from both pre- and post-commercialization periods of GPT models.

For reference, a detailed software artifact with data processing methods are documented in our GitHub repository in the Appendix A.

2.1.2 Criteria for Research Paper Selection

The methodology for selecting and filtering papers involved several key steps:

- **Field Selection:** We initially focused on Business Economics, a critical area for timely and accurate research dissemination.
- **Journal Selection:** Papers were selected from top journals within each field, ensuring they had substantial citation counts to guarantee high-quality and widely recognized research.
- **Citation Count:** Priority was given to papers with significant citations, indicating their influence and relevance in ongoing scholarly conversations.

- **Publication Date:** We included papers published between 2016-2022 and those published in 2023 onward, to analyze word frequencies before and after the commercialization of GPT models.

The dataset aimed to be statistically meaningful, comprising 1,000 papers from five broad academic disciplines, each contributing approximately 500,000 words. This scope ensures that the dataset is representative of high-quality research across diverse fields:

- **Business Economics:** Includes Economics, Business.
- **Health & Life Sciences:** Includes Biology, Chemistry, Medicine.
- **Humanities:** Includes Anthropology, Psychology, Sociology, History, Philosophy, Linguistics, Education, Art.
- **STEM Disciplines (Engineering Focused):** Includes Computer Science, Engineering, Materials Science, Physics, Mathematics.
- **Legal Space:** Includes Law, Political Science.

This approach provides a comprehensive and diverse dataset, enabling robust analysis of AI influence across different domains. We aimed for a dataset size of 1,000 papers to ensure a statistically significant sample, facilitating meaningful comparisons and insights into how LLMs might be influencing academic writing. The choice of selecting 1,000 papers per discipline was determined by balancing several factors:

- **Computational Constraints**: Processing a large number of research papers requires significant computational resources. By limiting the selection to 1,000 papers, we ensured that the data processing and analysis could be completed within a reasonable timeframe (e.g., order of hours/days) without overburdening our compute cluster. Relatedly, analysis on this dataset of this size allows us to perform complex analyses on a sufficiently large enough sample of words and papers, while preserving timeliness.
- **Statistical Robustness**: A dataset of 1,000 papers per discipline is large enough to provide a statistically significant sample, facilitating meaningful comparisons and insights. This size ensures that the dataset is representative of high-quality research across diverse fields and can capture a wide range of linguistic patterns influenced by LLMs.
- **Diminishing Returns**: Including more than 1,000 papers might yield diminishing returns in terms of additional insights while significantly increasing computational demands. Conversely, including fewer papers could compromise the robustness and generalizability of our findings. Therefore, 1,000 papers represent an optimal balance, providing a substantial amount of data for analysis while remaining manageable in terms of processing requirements.

Future Research and Improvements: The current study faced limitations due to local compute cluster constraints, affecting the ability to scrape and process data in parallel. Future research will focus on obtaining more granular insights and stratifying data by additional relevant features, enhancing the depth and precision of our analysis, whether it be further interdisciplinary insights, peer-review insights, and different modes of scholarly communication. Further exploration into other disciplines and broader datasets will also help to generalize the findings and refine the competitive and ethical framework proposed.

2.2 Word Frequency Analysis Methods

2.2.1 Text Parsing, Tokenization, and Frequency Indexing

The text processing pipeline for our analysis involved several critical steps to ensure data integrity and meaningful results:

- **Text Parsing:** We extracted the full text of each research paper, focusing specifically on the main content while excluding sections like references, footnotes, and appendices. This step was crucial to maintain the integrity of the dataset by focusing solely on the core scholarly content.
- **Tokenization:** The extracted text was split into individual tokens (words) using the NLTK library. Tokenization is essential for analyzing the frequency of each word in the corpus, enabling a detailed examination of language usage patterns.
- **Stopword Removal:** We removed common stopwords (e.g., "and," "the," "of") to concentrate on the meaningful content of the texts. This step reduces noise in the analysis, allowing us to focus on the words that carry the most semantic weight.
- **Lemmatization:** Words were lemmatized to their base forms (e.g., "running" to "run") using the WordNet lemmatizer. Lemmatization ensures that different forms of a word are treated as a single term, providing a more accurate representation of word usage and frequency.

These preprocessing steps are crucial for achieving dataset integrity and ensuring that our analysis captures the true linguistic patterns in the texts. By standardizing the text data, we reduce variability and enhance the reliability of our word frequency analysis[6].

2.2.2 Techniques to Measure Word Importance and Prevalence

To measure word importance and prevalence, we employed the TF-IDF (Term Frequency-Inverse Document Frequency) technique. TF-IDF is a statistical measure that assesses the importance of a word in a document relative to a collection of documents (the corpus)[7]. It helps identify terms that are not only frequent within a specific document but also distinctive when compared to the entire corpus.

- **TF-IDF Calculation[8]:** This technique combines two key components: Term Frequency (TF) measures how often a word appears in a document. It is calculated as the ratio of the number of occurrences of a word to the total number of words in the document. TF captures the relative importance of a word within a single document. Inverse Document Frequency (IDF) measures the importance of a word across the entire corpus by considering how common or rare the word is among all documents. IDF is calculated as the logarithm of the total number of documents divided by the number of documents containing the word. A higher IDF score indicates that the word is rarer across the corpus, thus potentially more informative.

The product of these two measures—TF and IDF—produces the TF-IDF score, which highlights words that are frequent in specific documents but not ubiquitous across all documents. This score helps in identifying potentially significant content by focusing on words that are not only frequent within a document but also provide distinguishing information when compared to the broader corpus.

Using TF-IDF allows us to mitigate the bias towards commonly used words and better interpret changes in word frequency. While TF-IDF itself does not "normalize" data in the conventional statistical sense, it does provide a way to prioritize words on relative frequencies. For example, an increase in the usage of a previously infrequent word can be more indicative of a meaningful trend than changes in more common words. This method enhances our ability to detect the subtle influences of LLMs on academic writing by identifying specific terms and phrases that have gained prominence in recent publications. This is crucial for understanding the nuanced ways in which LLMs might be shaping academic discourse.

2.3 Overview of Methods Used to Identify Selected-Word-Sets

2.3.1 Rationale for Choosing Specific Word Sets

We identified several word sets for targeted analysis, based on our hypothesis about LLM's influence on academic writing. The following table summarizes the word sets, example words, rationale behind choosing these words, and their significance:

Word Set	Example Words	Selection Rationale	Significance
Inference	conclude, suggest, indicate, imply, infer	Crucial in academic writing for presenting findings and interpretations	Indicative of how AI might influence scholarly conclusions
LLM-Primed	delve, unleash, tapestry, mosaic, multifaceted	Identified in preliminary studies as overused by AI models	Direct measure of LLM's impact on text generation
Methodology	method, analysis, data, experiment, model, algorithm	Fundamental in describing research approach	Reflects changes in how research methods are reported
Research Process	research, study, significant, result, finding	General terms related to academic research	Provides broader context for understanding AI's influence

The choice to focus on these specific word sets stems from an attempt to capture a wide range of linguistic and semantic influences that AI models might exert on academic writing. Each set was selected for its relevance in academic communication in addition to the potential to reveal subtle shifts in different language use that could be attributed to AI intervention. By analyzing these diverse categories, we aim to draw comprehensive insights into how LLMs might be reshaping various aspects of scholarly communication, from methodological rigor, presentation of inference, to the expression of research findings and interpretations.

Word sets are further enumerated in Appendix B.

2.4 Testing

2.4.1 Comparing Word Frequencies and Inference

In our study, we compared word frequencies from research papers published between 2016-2022 and those published in 2023 onward. This comparative analysis aimed to identify significant changes in word usage that may coincide with the commercialization and broader adoption of LLM models. Our methodology primarily involved a quasi-experimental approach, focusing on observed differences in word frequency patterns over time.

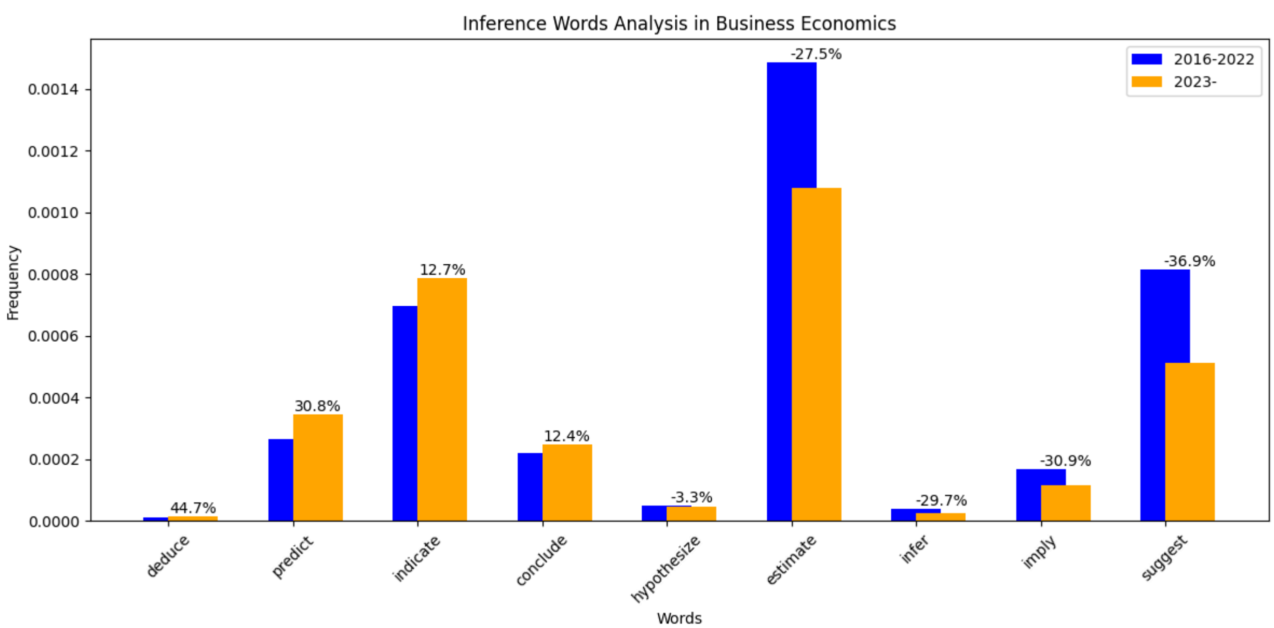
While our approach provided valuable preliminary insights, it is important to acknowledge its limitations. The methodology lacked the rigorous statistical controls typically required to definitively attribute changes in language patterns to LLM interventions. Future research should employ more robust statistical tests, such as time-series analysis, difference-in-differences, or causal inference techniques, to better isolate the effects of LLMs from other variables influencing academic writing trends. Nevertheless, our comparative analysis seeks to provide a preliminary foundation for developing a comprehensive framework to address the competitive and ethical implications of AI in scholarly communication. This framework will be crucial for guiding future research, policy-making, and the responsible integration of AI technologies in academia.

3. Findings

3.1 Introduction:

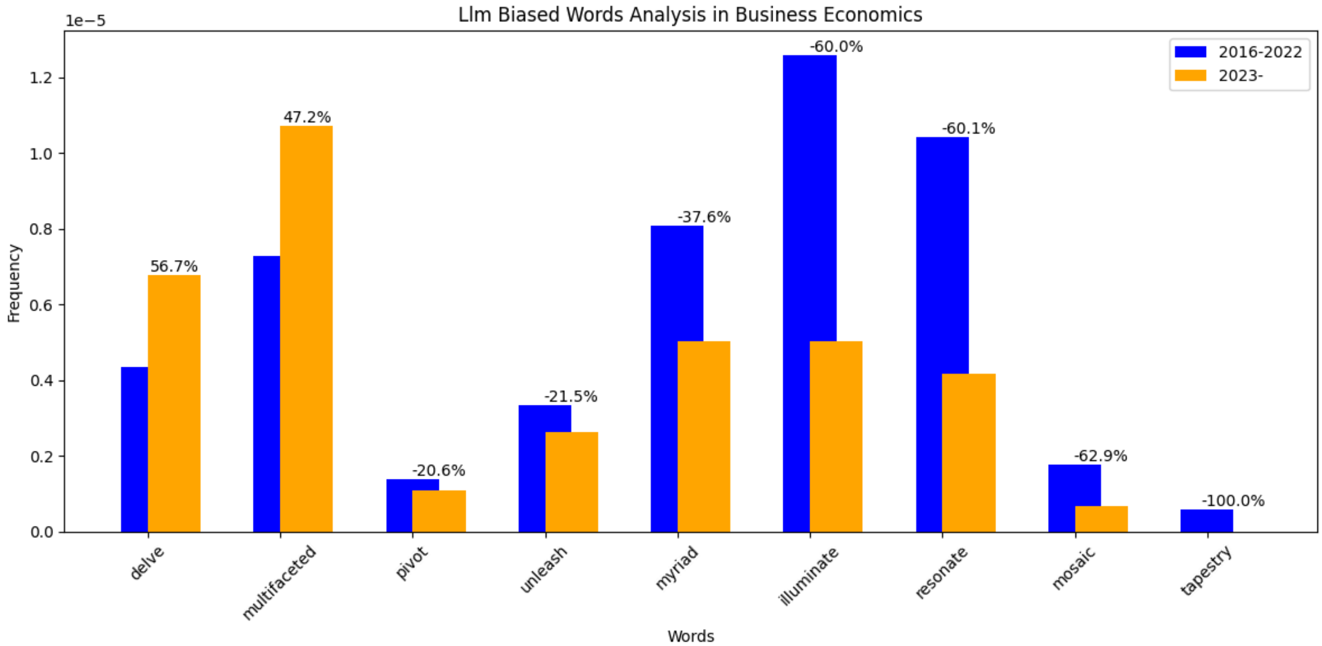
In our sample of 1,000 research papers across five research domains, significant trends in word frequencies were identified over time among the 500,000 word corpus for each discipline. The trends highlight changes in the use of specific words from the period 2016-2022 compared to 2023-. Key findings will be discussed in the following sections. To prime our analysis, a survey of the highest frequency words per analyzed discipline can be found in Appendix C.1.

3.2 Primer: Selected Trends in Business Economics



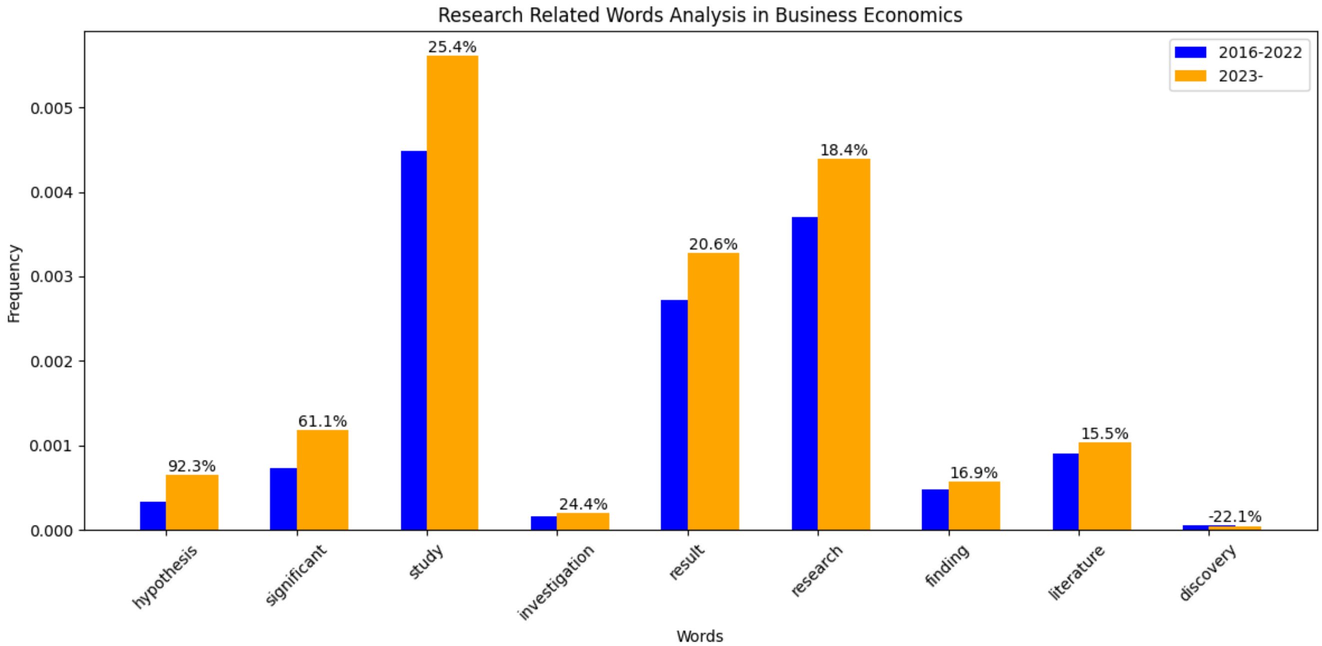
Inference frequency values from 2016-2022 to 2023-, sorted by % change decreasing

For *Inference* set, *estimate* and *suggest* display the greatest decreases in frequency among most commonly used inference words. *Predict*, *indicate*, and *conclude* display the greatest increases. *Deduce*, *infer*, and *imply* also exhibit significant changes, although they represent significantly smaller samples by virtue of their relatively lower average frequency value.



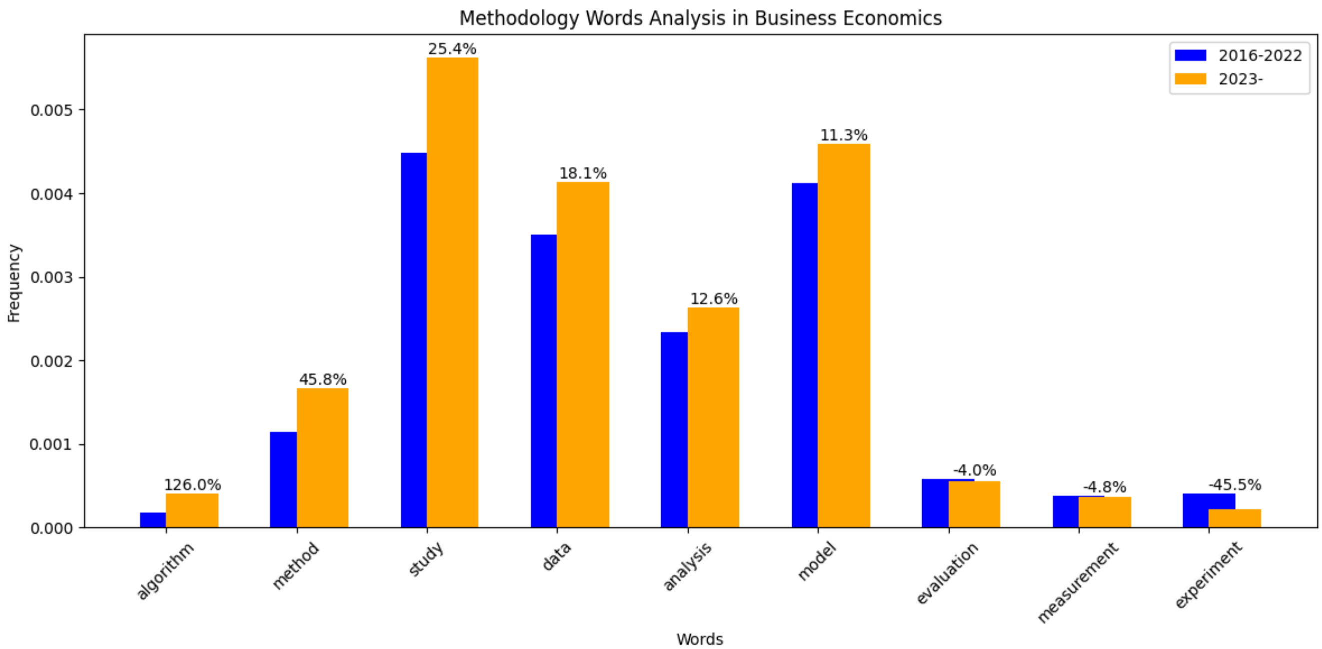
LLM-primed frequency values from 2016-2022 to 2023-, sorted by % change decreasing

On the selection of LLM-primed words, two words *delve* and *multifaceted* display the most significant increases of 1.5 fold, whereas other LLM-primed words like *myriad*, *illuminate*, and *resonate* see relatively significant decreases in frequency from the control to the target period. *Tapestry* and *mosaic* see greatest decreases in frequency values, ranging from greater than 50%, to 100% decreases.



Research Process frequency values from 2016-2022 to 2023-, sorted by % change decreasing

For research inquiry relevant words, we see a substantial increase in the usage of words *hypothesis* and *significant*, exhibiting 1.5-2 fold increases. Generally, the majority of other Research Words increased in relative frequency as well, less *discovery*.



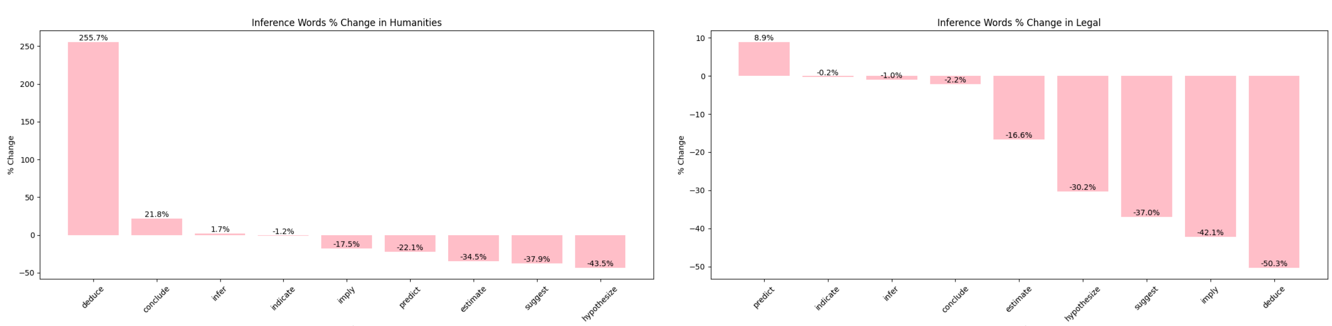
Methodology frequency values from 2016-2022 to 2023-, sorted by % change decreasing

Among *Methodology* words, *algorithm* and *method* exhibit the most significant positive change, whereas *experiment* exhibit the most significant negative change.

To investigate interesting trends across subject-matter differences and broaden our inference across disciplines, similar analyses were conducted across four adjacent fields: Health & Life Sciences, Legal, Humanities, and STEM. This comparative approach provides deeper insights into the prioritization and deprioritization of certain words across different domains.

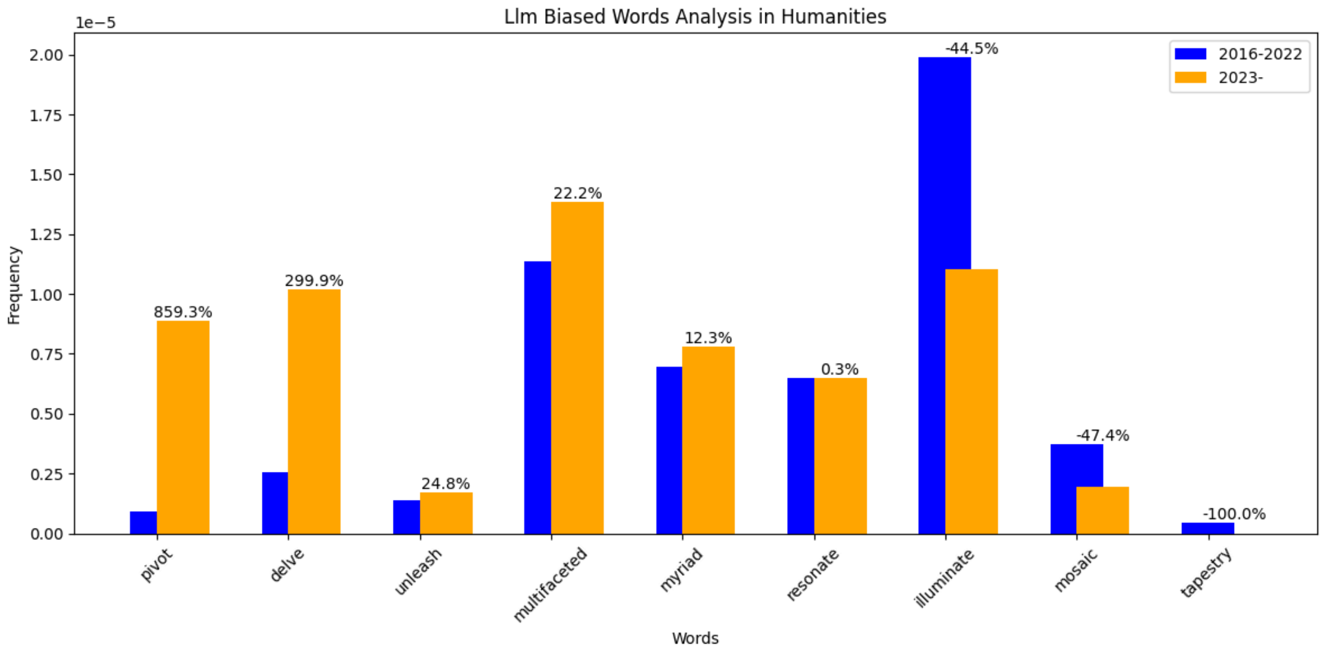
3.3 Cross-Domain

3.3.1 Selected Trends in Health & Life Sciences, Legal, Humanities, and STEM



Deduce represents the greatest positive and negative frequency % change from 2016-2022 to 2023-, in Humanities and Legal respectively

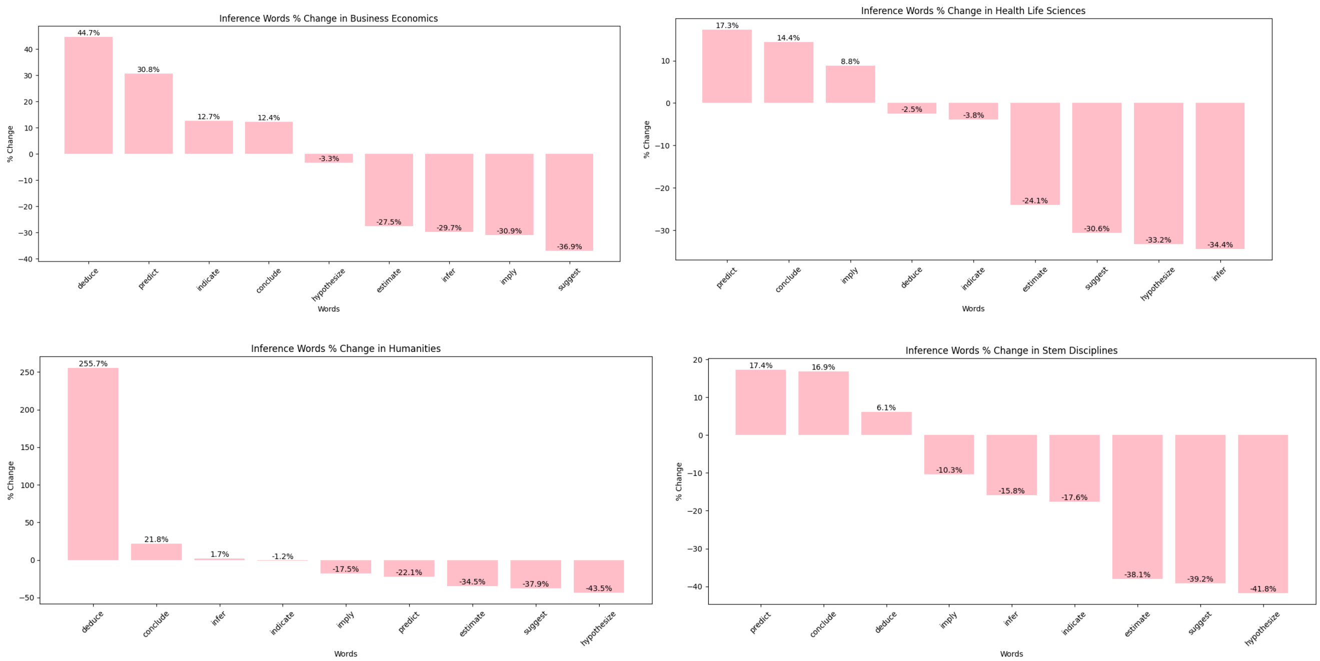
The word *deduce* showed a 255.7% increase in Humanities, the greatest increase among other Inference Words. In contrast, the Legal field experienced a -50.3% decrease in the same word, representing the greatest decrease. This contrast raises questions about domain-specific language trends and the influence of sample size or methodological approaches. Specific point-variation like this exists across different pairings of our five disciplines, potentially indicating different usage contexts or LLM training biases, or sample considerations.



Pivot surpasses *delve* in % frequency change from 2016-2022 to 2023-, in Humanities

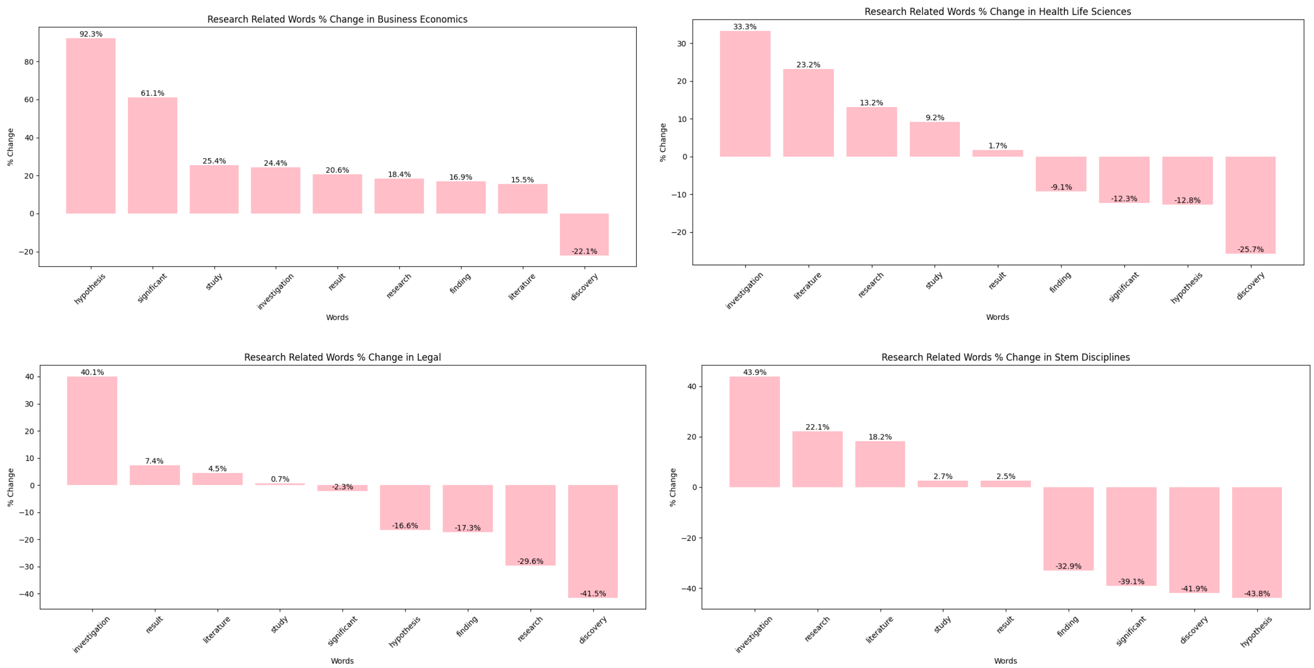
In Health & Life Sciences, Legal, Humanities, and STEM, LLM-primed words like *delve* showed consistent increases across disciplines, with a notable peak of over 1000% in STEM. Only in Health & Life Sciences, displayed a greater increase in the frequency of *pivot* at over 800% increase. In Legal, Humanities, STEM, and Business Economics, *pivot* saw less than 50% increases or decreases in frequency.

3.3.2 Synthesis



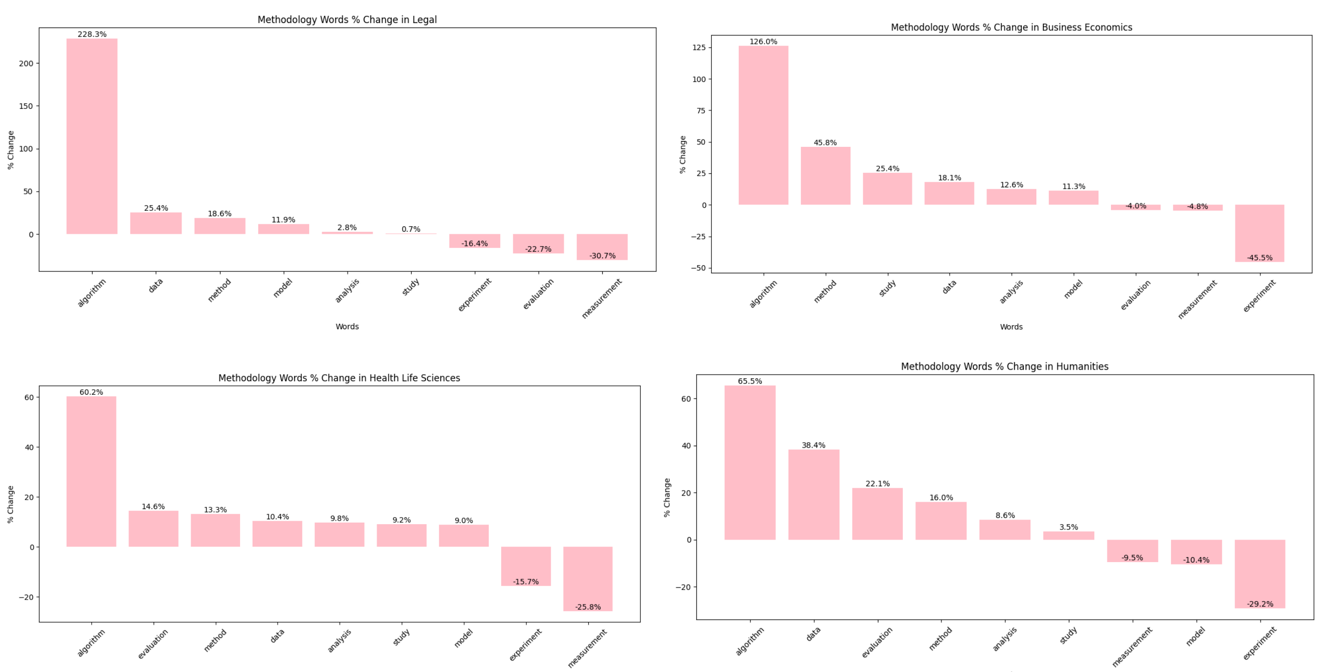
Inference frequency % changes across four selected disciplines

Comparing outcomes across disciplines, we observe the frequency of particular word batches being diminished or increased relatively consistently. Inference words *suggest*, *estimate*, and *hypothesize* consistently make the bottom five inference words in frequency change. For two disciplines Business Economics and Humanities, *deduce* exhibits the greatest frequency increases for their respective disciplines, from greater than 40% to greater than 250% respectively.



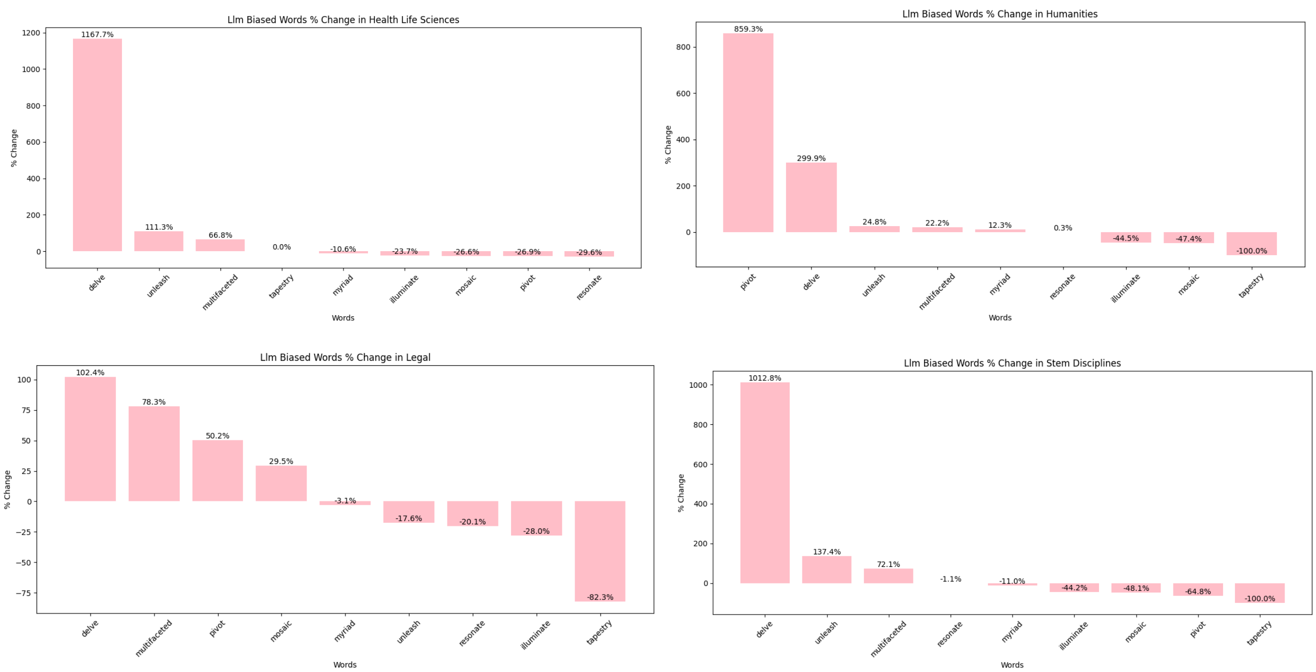
Research Process frequency % changes across four selected disciplines

Among the *Research Words* batch, *discovery* consistently decreased in four of five disciplines, less the Humanities discipline. *Investigation* often represents the words with greatest positive frequency change across disciplines. *Significant* and *Hypothesis* generally decreased in four of five disciplines. Interestingly, Business Economics was the only discipline in which both increased strongly as a pair.



Methodology frequency % changes across four selected disciplines

In *Methodology* batch, the word *algorithm* placed the most extreme increase in frequency among four disciplines, except STEM Disciplines.



LLM-primed frequency % changes across four selected disciplines

Select *LLM-primed words* exhibited the most extreme increases in word frequencies. *Delve*, consistently from greater than 50% increase in the Business Economics discipline, to over a 1000% increase in STEM disciplines. We see a similar extreme trend with *pivot*, in the Humanities discipline, showing greater than 800% increase.

4. Discussion

4.1 Summary of Findings and Interpretation

Our study reveals significant changes in the use of specific words across different research domains over time, reflecting evolving research focuses, methodologies, and possibly the influence of LLMs on academic writing. While some trends are consistent across disciplines, others are domain-specific, highlighting the complex interplay between language use and research context. These findings suggest that LLMs like GPT-series may be shaping academic writing in nuanced ways, including amplifying certain phrases, introducing new stylistic elements, and potentially influencing the reporting of research methodologies and results.

One of the more notable findings is the differential increase and decrease in the use of the word *deduce* across disciplines. The 255.7% increase in Business Economics contrasts sharply with a 50.3%

decrease in the Legal field. This variation could be driven by different methodological emphases—Business Economics may increasingly focus on quantitative, deductive reasoning, a trend potentially reinforced by the data-centric nature of LLM training corpora. In contrast, the Legal field might be shifting towards more narrative or case-based approaches, where deductive reasoning is less emphasized. This contrast highlights how LLMs, which learn from vast and diverse datasets, could be amplifying language styles prevalent in specific types of literature, thereby influencing academic discourse within those fields.

The increased frequency of terms like *delve*, *unleashed*, and *multifaceted* across multiple disciplines, especially in STEM, suggests an integration of more descriptive and exploratory language, possibly due to LLM influence. These models are trained on diverse datasets that include not only academic texts but also media and popular science writing, which often use richer, more illustrative language. As a result, the adoption of such language in academic contexts might reflect a broader trend towards making scientific literature more engaging in this way, a shift that LLMs may be accelerating by prioritizing these terms. Blending academic and more general language in this way may not be the best outcome for the academic space. In a space that traditionally values specificity and rigor, this illustrative style could risk reducing the precision that is crucial in scientific communication.

The strong positive or negative deltas in words like *hypothesis* and *significant* across different disciplines suggests a bias away or towards hypothesis-driven research and statistical validation depending on the field. This trend may reflect the influence of LLMs, which in certain fields, may highlight statistically significant results as key findings in certain fields, potentially reinforcing a focus on quantitative metrics in academic publishing. In tangential fields however, this trend may prove opposite, where LLMs may prioritize qualitative methods, and more quasi-experimental approaches due to subject matter differences. The decrease in the term *discovery* overall, however, suggests a potential shift away from exploratory research towards more structured and bounded approaches. This could be influenced by LLMs favoring more standardized and measurable expressions of research outcomes, thus potentially narrowing the scope of academic inquiry to more readily approachable studies.

Moreover, the rise in the use of terms like *algorithm* indicate a growing reliance on computational approaches, particularly in all fields except STEM. This shift may be partly attributed to the influence of LLMs, which are themselves products of complex algorithms and machine learning literature grounded in the computationally heavy engineering world. The integration of such models into academic workflows thus might enable the trickling of similar terminologies to all other disciplines. This trend could potentially lead to a homogenization of research methodologies, with a stronger focus on algorithmic and quantitative approaches at the expense of qualitative and theoretical methods, where they may see a worse fit. Additionally, the increased use of misfit technical jargon could make it challenging for interdisciplinary collaboration and communication.

4.2 Framework for Evaluating the Impact of LLMs on Academic Writing

To better codify the degree and form of impact LLMs are exerting on academic writing, we propose a framework encompassing several key dimensions that our findings capture:

Impact Dimension	Description	Methodology

Linguistic Amplification and Standardization	Evaluating whether LLMs are amplifying certain high-frequency words and phrases, leading to a standardization of academic language.	Analyzing the frequency and context of specific words across time and disciplines, comparing pre- and post-LLM adoption periods.
Introduction of Non Subject-Matter Language	Identifying the introduction of phrases and stylistic choices that are atypical for traditional academic writing but common in non-academic sources.	Assessing the prevalence of such terms and their impact on the tone and perceived rigor of academic publications.
Methodological/Ideological Result Reporting	Investigating how LLMs might be affecting the way methodologies and findings are described, potentially favoring certain terminologies or phrasings.	Conducting comparative analyses of methodological descriptions, noting shifts in language that align with LLM-influenced patterns.

The variability in the adoption of these language trends across different disciplines suggests that while LLMs may be influencing academic writing, their impact is mediated by the specific norms and priorities of each field. In turn, their impact varies. It may as well prove moot in the case of introducing incongruent style or wording in one domain, or explicitly negative, like in the case of introducing high impact words in inappropriate contexts, or incorrectly biasing ideas towards one research outcome over another. The potential downsides of these trends also include the risk of diminishing linguistic diversity and originality in academic writing. As LLMs standardize certain terminologies and styles, there is a concern that researchers might feel pressured to conform to these norms, potentially stifling creativity and the development of new theoretical frameworks. Additionally, the ethical implications of AI's role in academic writing must be considered, particularly concerning authorship and the authenticity of scholarly work. The increasing prevalence of AI-assisted writing tools raises questions about the ownership of ideas and the extent to which AI-generated content should be integrated into academic publications.

As AI continues to play a larger role in the production and dissemination of knowledge, it is crucial for the academic community to critically engage with these technologies, ensuring that their use enhances, rather than diminishes, the richness and diversity of scholarly communication. Further research is needed to explore the long-term impacts of LLMs on academic writing and to develop guidelines that balance the benefits of AI tools with the need for originality and ethical considerations in academic work. However, this high level framework of identifying and evaluating the particular patterns of potential LLM impact in academic literature synthesis can serve as a starting point for intervention, meta-review processes, and industry controls.

4.3 Competitive Dynamics in Academia

4.3.1 Technology Disruption

The rapid integration of LLMs into academic writing has introduced a profound shift in the landscape of scholarly communication. As adoption surged from late 2022 into 2023 and mid-2024, industry-wide disruption has not only proven timely in its arrival, but unprecedented in its magnitude. The speed and scale of this transformation necessitate appropriate policy oversight to maintain the integrity of the academic industry. This section presents a comprehensive framework for understanding and navigating the implications of LLMs, emphasizing firm-specific actions and industry-wide regulatory considerations.

4.3.2 Engendering an Efficient and Trustful Academic Industry: Component Frameworks

To harness the potential of LLMs while safeguarding the integrity of academic research, it is essential to evaluate and establish standards across several key dimensions. The goal is to leverage the unprecedented capabilities of hyper-scale LLM models to improve productivity in the research space while preserving the meritocratic, collaborative, and ethical practices that underpin the academic environment, or at least, purported ideals. The following table outlines the framework, highlighting actionable insights and strategies for implementation that zoom in on the key levers for LLM oversight under these goals.

1. Competition Framework for an Efficient Industry

Strategic Dimension	Description and Strategy
Leveraging LLMs for Enhanced Productivity and Innovation	**Actionable Insight:** Institutions and researchers should be able to leverage the capabilities of LLMs to gain a competitive edge. They should be able to integrate LLMs into their workflows, using these models for literature reviews, data analysis, and drafting research papers. LLMs can efficiently generate summaries, identify key themes, and suggest relevant citations, allowing researchers to focus on more complex tasks, idea synthesis, and innovation.
	Implementation Strategy: Establish dedicated AI support teams within academic institutions to assist researchers in adopting LLM technologies. Offer training programs and workshops to ensure proficiency in using these tools effectively. Encourage the practical and *safe* usage of LLMs in different research tasks.
Expanding Access to AI Technologies	**Actionable Insight:** Access to advanced AI tools like LLMs is currently unevenly distributed, favoring well-resourced institutions. Strategies are needed to democratize access and ensure equitable opportunities for all researchers.

	Implementation Strategy: Advocate for public and private funding to support the acquisition of AI technologies for under-resourced institutions. Develop or collaborate with open-source AI tools and platforms that are freely accessible, ensuring that researchers from all backgrounds can benefit from these advancements.
Ensuring Originality and Mitigating Risks of Homogenization*	**Actionable Insight:** While LLMs can enhance productivity, there is a risk of homogenization in academic writing styles, which can hinder research paper production and desired journal outcomes. Overreliance on these models or a lack of oversight in turn may lead to a lack of originality and a narrowing of research perspectives, hindering the outcomes of the research institutions. While encouraging the use and access to LLMs, these negative production risks must also be considered. * Directly coupled with ethical outcomes, outlined in *Ethical Framework*
	Implementation Strategy: Encourage the use of LLMs as supplementary tools rather than replacements for critical thinking and original writing. Institutions should implement guidelines promoting unique analytical frameworks and diverse perspectives. Regular peer reviews, originality checks, among other new review processes can help maintain the integrity and diversity of academic output.

The implications of integrating LLMs into academic writing extend beyond mere productivity gains. These technologies offer significant opportunities for innovation, such as enhancing literature reviews, accelerating data analysis, and streamlining the writing process. However, they also present challenges related to originality, collaboration, and perpetuating biases in research outcomes.

2. Ethical and Regulatory Framework for Maintaining Trust and Collaboration

Ethical Dimension	Description and Strategy
Addressing Ethical Concerns Around	**Actionable Insight:** The use of LLMs raises ethical questions about authorship, intellectual property, and the authenticity of scholarly work. A clear regulatory framework is needed to govern the use of AI in academic contexts. As industry collaboration is often strongly influenced by

Authorship and AI-Generated Content	authorship and citations for a given paper, being able to maintain the authenticity and attribution of academic work is essential to this standard.
	Regulatory Strategy: Establish industry-wide standards for AI-assisted writing, including transparency requirements for disclosing the use of AI tools. Create policies that define the roles and contributions of AI in the authorship process, ensuring that human authorship remains central and AI-generated content is properly credited and contextualized.
Monitoring and Evaluating the Impact of AI Integration	**Actionable Insight:** Ongoing monitoring and evaluation are crucial for understanding the long-term impact of AI integration in academic writing and research.
	Regulatory Strategy: Establish a framework for regular assessment of AI tools' impact on research productivity, innovation, and ethical standards. Metrics could include diversity in research outputs, originality of publications, and prevalence of AI-generated content. Regulatory bodies should publish regular reports on these assessments, providing transparency and accountability.
Mitigating Potential Biases in AI Models*	**Actionable Insight:** LLMs, trained on large datasets, may contain inherent biases affecting the representation of certain topics and perspectives, potentially marginalizing underrepresented voices.
	Regulatory Strategy: Implement guidelines for developing and using AI training datasets, emphasizing diverse sources and perspectives. Conduct regular audits of AI models to identify and mitigate biases. Establish independent oversight bodies to monitor and enforce these standards, ensuring AI technologies contribute positively to the diversity and inclusivity of academic discourse. * Directly coupled with production outcomes, outlined in *Competition Framework*

Promoting Ethical AI Use in Research	**Actionable Insight:** The rapid adoption of LLMs necessitates the development of ethical guidelines to prevent misuse and ensure responsible use.
	Regulatory Strategy: Develop comprehensive ethical guidelines covering various aspects of AI use in research, including data privacy, transparency, and academic integrity. Incorporate these guidelines into academic curricula to ensure future researchers understand the ethical considerations associated with AI technologies.

By implementing a comprehensive competition framework and regulatory strategy, the academic community can navigate these challenges effectively. This approach will help ensure that the benefits of AI are fully realized while mitigating potential risks, such as homogenization of academic discourse and the erosion of diverse perspectives. The ethical framework ensures that AI enhances the quality and integrity of academic work, maintaining trust and collaboration within the scholarly community. This dual approach fosters an environment where innovation and ethical considerations coexist harmoniously, advancing knowledge and preserving the rich diversity of academic discourse.

4.6 Limitations and Future Research

The integration of LLMs into academic writing presents unique challenges and opportunities. However, several limitations and considerations must be addressed to enhance the rigor and reliability of future research in this area. This section outlines specific pitfalls identified in our own data collection and analysis for this study, and suggests necessary steps to refine our methodologies and expand the scope of inquiry.

Limitations

1. Sample Selection and Representativeness:

- **Pitfall:** The selection of research papers may not adequately represent the full spectrum of academic writing, leading to potential biases in our findings. The dataset might be skewed towards certain journals, disciplines, or geographic regions, limiting the generalizability of the results.
- **Consideration:** Future research should ensure a more diverse and representative sample of academic literature, including papers from underrepresented disciplines and non-English publications. Additionally, stratified sampling methods could be employed to ensure that all relevant subfields are adequately represented. Moreover, paper selection criteria may also consider paper quality, distribution, and credibility of journals accepted, in addition to different kinds of academic literature beyond studies.

2. Temporal Variation and Short Study Period:

- **Pitfall:** The study period, focusing on changes from 2016-2022 to 2023-, may be too short to capture long-term trends and the full impact of LLMs on academic writing. Temporal variations in word usage might also be influenced by external factors such as major global events, rather than the adoption of AI tools.
- **Consideration:** Extending the study period to include earlier and later years could provide a more comprehensive understanding of trends over time. Time-series analysis techniques, such as ARIMA models or trend decomposition, can also help isolate genuine trends from short-term fluctuations.

3. Identification of LLM Influence:

- **Pitfall:** Distinguishing the influence of LLMs from other factors affecting academic writing (e.g., changes in research focus, editorial policies) is challenging. The observed changes in language use may not be solely attributable to LLMs but could also result from broader shifts in academic and research practices.
- **Consideration:** To better identify LLM influence, researchers could use controlled experiments where papers written with and without AI assistance are compared. Additionally, surveys or interviews with authors could provide insights into the extent and nature of AI use in their writing processes. Moreover, more rigorous statistical methods can be employed to identify a more ground truth impact of LLM influence, in the context of multiple other influencing factors.

Future Research Directions

1. Expanded Comparative Analysis:

- **Direction:** Future studies should compare the impact of LLMs across a wider range of disciplines, academic literature forms, and venues. This includes examining differences between humanities, social sciences, natural sciences, and technical fields, as well as comparing English-language publications with those in other languages. Moreover, it can investigate interesting differences in LLM usage between journals of various disciplines, geographies, qualities, and credibilities.
- **Methodology:** Use cross-disciplinary meta-analyses to assess the uniformity or divergence in LLM influence across these different spaces. This may involve a higher scale data collection pipeline and storage, in addition to robust machine learning modeling.

2. Ethical and Policy Implications:

- **Direction:** Explore more in-depth the ethical and policy implications of LLM integration in academia, focusing on issues like authorship, intellectual property, and the potential homogenization of academic discourse.
- **Methodology:** Conduct case studies and legal analyses to understand the implications of AI-generated content on copyright law and academic integrity. Investigate the impact on items like genuine citation-making, and plagiarism detection within journals. Develop policy recommendations for academic institutions and publishers regarding the ethical use of AI tools.

3. Exploring the Impact of LLMs on Research Collaboration:

- **Direction:** Investigate how the integration of LLMs influences research collaboration and network dynamics among scholars.
- **Methodology:** Analyze co-authorship patterns, citation networks, and collaborative projects before and after LLM adoption. Conduct surveys and interviews with researchers to understand changes in collaborative practices and the role of AI in facilitating or hindering these interactions.

In summary, while this study provides valuable insights into the influence of LLMs on academic writing, addressing these limitations and pursuing these future research directions will help build a more comprehensive and nuanced understanding of the continued impact of LLMs as they grow in adoption. By expanding the scope of analysis, refining methodologies, and exploring the ethical dimensions, researchers can better navigate the evolving landscape of academic communication in the age of AI.

5. Conclusion

This study has explored the profound influence of large language models, such as GPT-4, on academic writing across various disciplines. Our findings reveal significant shifts in language use, highlighting how these advanced AI tools are reshaping the landscape of academic communication. The amplification of specific terminologies, the introduction of non-traditional language, and the standardization in the presentation of research methodologies suggest a growing integration of LLMs in scholarly work.

The integration of LLMs presents both opportunities and challenges with notable economic implications. Institutions and researchers who effectively leverage these tools can gain a significant competitive advantage, enhancing their productivity, accelerating the research process, and improving the quality and accessibility of academic publications. This competitive edge can lead to increased funding, higher academic reputations, and stronger collaboration networks. However, the widespread adoption of LLMs also raises concerns about economic disparities in access to advanced AI technologies, potentially exacerbating inequalities between well-resourced and under-resourced institutions.

Moreover, the homogenization of academic language and the potential biases introduced by LLMs' training data could influence the direction of academic research and the dissemination of knowledge. These issues necessitate a critical evaluation of the ethical and strategic implications of LLM use in academia. Establishing clear guidelines and policies is crucial to ensure that AI tools are used responsibly, preserving the originality and integrity of scholarly work while promoting inclusivity and diversity in academic discourse.

In conclusion, while LLMs offer powerful tools for enhancing academic writing, their widespread and inevitable adoption must be managed carefully. Future research and policy development will be essential to harness the benefits of these technologies while mitigating potential economic and ethical challenges. By doing so, the academic community can ensure that LLMs contribute positively to the advancement of knowledge, fostering innovation and equity in the global academic landscape.

6. Reference

1. Gray, A. (2024). More "LLM keywords in quantities that can be seen from space": 'delve' (& in comments, 'mosaic', 'unleash', 'tapestry', 'multifaceted'). Retrieved from https://twitter.com/generalising/status/1774061899681869952
2. Cabanac, G. (2024). So #ChatGPT wrote the first sentence of this @ElsevierConnect article. Any other parts of the article too? How come none of the coauthors, Editor-in-Chief, reviewers, typesetters noticed? How can this happen with regular peer-review?. Retrieved from https://x.com/gcabanac/status/1767574447337124290?s=20
3. AI X BIO. (2024). AI Ethics Rise of Large Language Models in Academic Research: Implications for Reviewers and Writers. Retrieved from https://x.com/AIXBIO/status/1790576142425944120
4. Electric Eye. (2024). Use of #LLMs is starting to poison the peer-review process... (my fear: we could gradually see LLM-written papers being reviewed via LLM and –as this research shows– they are more likely to be higher ranked, more bullshit will end up being published). Retrieved from https://x.com/murakamiwood/status/1788522622633853080
5. Khalifa, M., & Albadawy, M. (2024). Using artificial intelligence in academic writing and research: An essential productivity tool. Computer Methods and Programs in Biomedicine, 100145. https://doi.org/10.1016/j.cmpbup.2024.100145
6. Manning, C. D., & Schütze, H. (1999). Foundations of Statistical Natural Language Processing. MIT Press.
7. Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to Information Retrieval. Cambridge University Press.
8. Ramos, J. (2003). Using TF-IDF to Determine Word Relevance in Document Queries. In Proceedings of the First Instructional Conference on Machine Learning.
9. Liang, W., Zhang, Y., Wu, Z., Lepp, H., Ji, W., Zhao, X., Cao, H., Liu, S., He, S., Huang, Z., Yang, D., Potts, C., Manning, C. D., & Zou, J. Y. (2024). Mapping the increasing use of LLMs in scientific papers. arXiv. https://doi.org/10.48550/arXiv.2404.01268
10. Liang, W., Izzo, Z., Zhang, Y., Lepp, H., Cao, H., Zhao, X., Chen, L., Ye, H., Liu, S., Huang, Z., McFarland, D. A., & Zou, J. Y. (2024). Monitoring AI-modified content at scale: A case study on the impact of ChatGPT on AI conference peer reviews. arXiv. https://doi.org/10.48550/arXiv.2403.07183
11. Kannan, P. (2024, May 13). How much research is being written by large language models? Stanford HAI. Retrieved from https://hai.stanford.edu/news/how-much-research-being-written-large-language-models
12. Movva, R., Balachandar, S., Peng, K., Agostini, G., Garg, N., & Pierson, E. (2024). "Topics, Authors, and Institutions in Large Language Model Research: Trends from 17K arXiv Papers." arXiv.

7. Appendix

A. Software Artifacts:

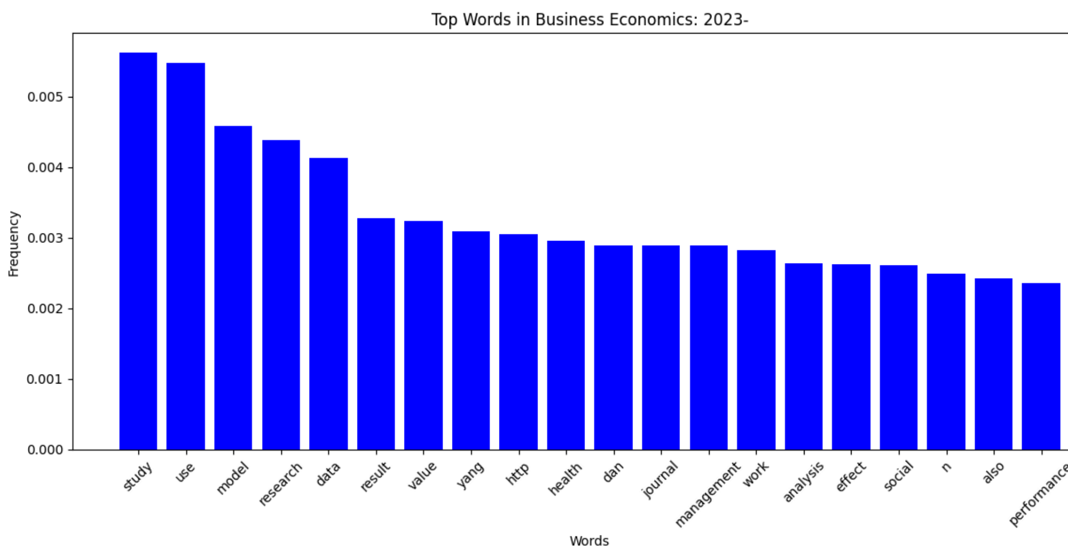
Github Repository: azhang315/scholar: Included: (github.com)
OpenSource Scholarly Data**:** Semantic Scholar Public API (2024) via https://www.semanticscholar.org

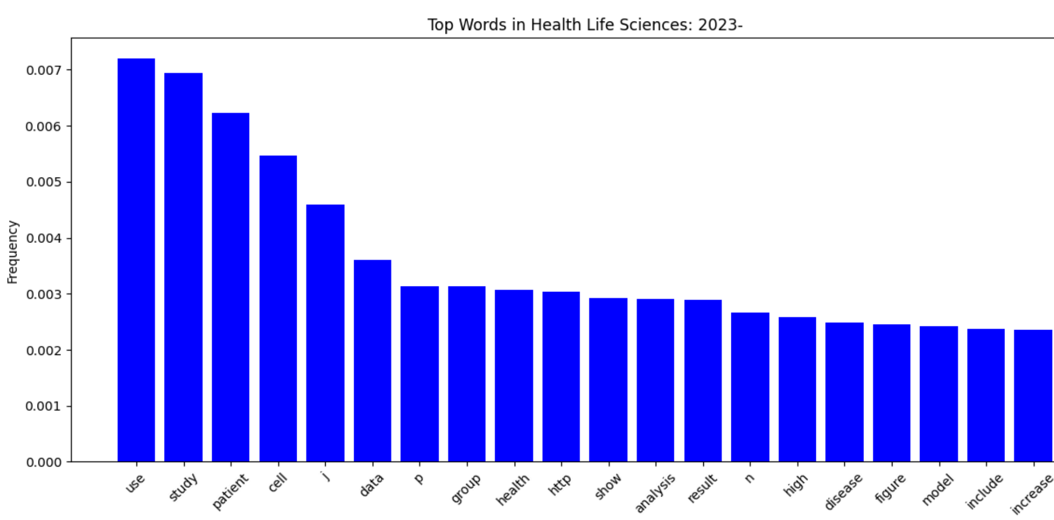
B. Word Sets

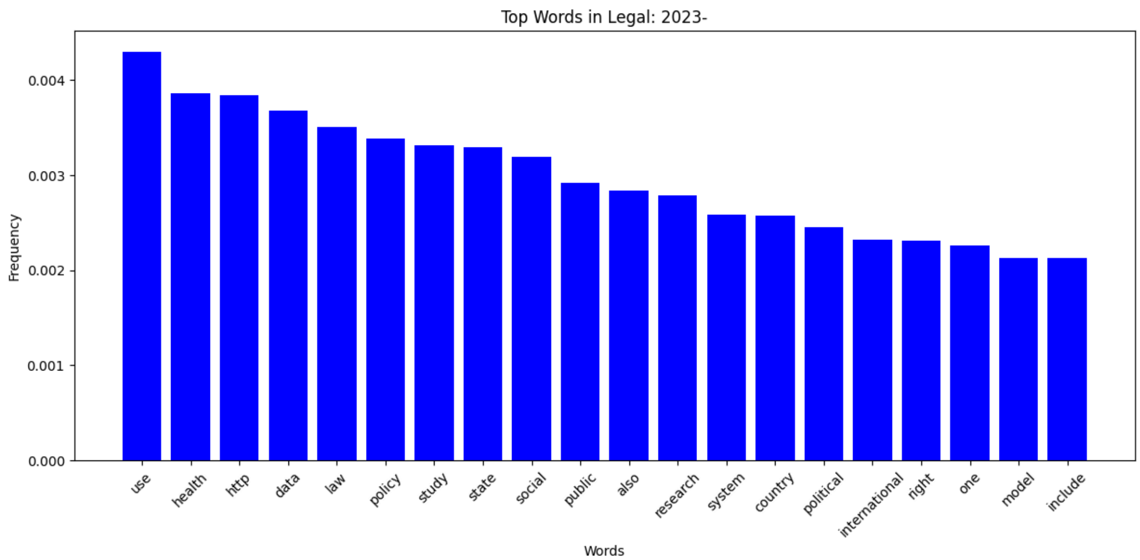
Inference: deduce predict indicate conclude hypothesize estimate infer imply suggest
LLM-Primed: delve unleashed multifaceted resonate myriad illuminate mosaic pivot tapestry
Methodology: method model data algorithm study evaluation analysis experiment measurement
Research-Process: investigation,research,literature,study,result,finding,significant,discovery,hypothesis

C. Additional Figures

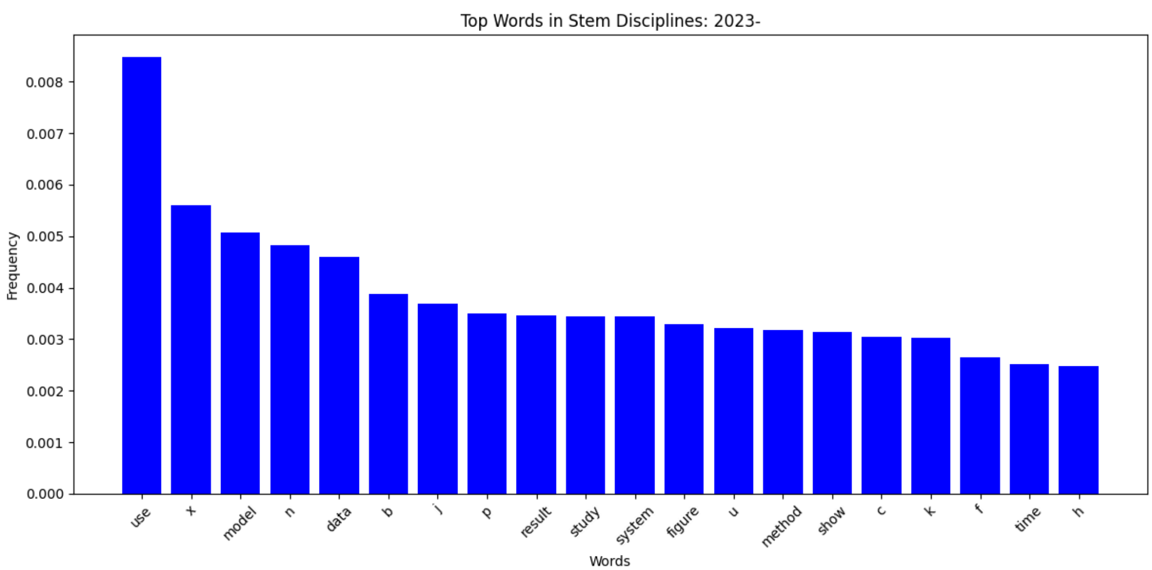
C.1. Observed Word Frequencies by Discipline descending







Top Words in Legal: 2023-
Frequency
0.000
0.001
0.002
0.003
0.004
use
health
http
data
law
policy
study
state
social
public
also
research
system
country
political
international
right
one
model
include
Words



Top Words in Stem Disciplines: 2023-
Frequency
0.000
0.001
0.002
0.003
0.004
0.005
0.006
0.007
0.008
use
x
model
n
data
b
j
p
result
study
system
figure
u
method
show
c
k
f
time
h
Words