Tracing Research Inequality in NLP: How Resource Disparities Shape Topic Trends and Methodological Diffusion via Citations

Anonymous ACL submission

Abstract

The growing resource gap between institutions raises critical questions about transparency, replicability, and inclusiveness in AI research. While some AI research topics remain accessible, research in areas such as large language models (LLMs) necessitate more resources such as computational power and data access: resources largely concentrated among industry companies and a few top universities. This study investigates research inequality in NLP by analyzing topic shifts, institutional resource gap, and citation intent patterns in papers from the ACL Anthology between 2010 and 2022. We identify 2016 as a critical turning point in NLP, marked by the rise of large language models (LLMs) and generative tasks, which have driven increased attention to topics such as Language Modeling, Generation, and Multimodality, while traditional areas like Machine Translation and Syntax/Parsing have declined. High-resource institutions are more likely to publish on these trending topics, as indicated by higher topic shift ratios. In contrast, low-resource teams are concentrated in declining topics. Citation intent analysis reveals that methodology-use citations, which indicate resource transfer, are decreasing over time, particularly in trending topics. This trend is especially pronounced in citations from low-resource to high-resource teams, suggesting that widening computational and infrastructural gaps limit the ability of low-resource institutions to adopt and build upon frontier research. These findings highlight a growing divide in NLP research participation and impact, underscoring the need for more inclusive and equitable research practices.

1 Introduction

Modern AI research demands increasing resources, especially access to large-scale infrastructure and datasets, creating a significant advantage to institutions with greater financial and computational capacity. For instance, in 2020, private enterprises reportedly spent over $80 billion on AI, while U.S. federal non-defense investment in AI-related research and development amounted to just $1.5 billion (Littman et al., 2022). This disparity has enabled well-resourced teams, especially those affiliated with major technology companies, to drive the development of increasingly sophisticated AI models. In contrast, many academic and public-sector institutions lack the resources necessary to reproduce, extend, or critically evaluate these advances (Patel, 2023), raising concerns about the inclusiveness and reproducibility of progress in the field.

This resource gap not only affects what institutions can build but also shapes what research questions they choose to ask (Movva et al., 2023). While industry actors often drive progress through proprietary models that require vast resources (Ahmed et al., 2023), academic and under-resourced teams often focus on problems that are more computationally tractable or theoretically grounded (Ignat et al., 2024; Togelius and Yannakakis, 2023).

Despite this significant resource disparity, the growing availability of open-source software frameworks, pretrained models, and benchmark datasets, has contributed to a broader participation in AI research (Gururaja et al., 2023), as evidenced by the influx of new authors in recent years (Movva et al., 2023). This raises a critical question: to what extent do high-resource teams, while pushing the frontier, also act as enablers (e.g., through the release of resources) for broader access? To better understand this dynamic, we focus on the NLP research community and address two research questions: 1) How do research topics differ between research teams from high-resource and low-resource institutions? And 2) To what extent has research from high-resource institutions lowered or heightened the barriers for low-resource teams in NLP?

To investigate the first question, we analyzed temporal shifts in topic distributions across institutions with varying resource levels, examining whether low-resource teams are increasingly constrained in the scope of topics they pursue.

Drawing on theories of citation that consider citations as framing devices and connectors of intellectual lineages (Latour, 1987), and previous studies that has used citation analysis to understand research dynamics (Huang et al., 2022; Jiang and Liu, 2023; Jones, 1994; Nishikawa, 2023), we investigate our second research question by analyzing citation patterns. Specifically, we examine how low-resource teams cite the work of high-resource teams, with a focus on methodological adoption, such as the use of models, datasets, or software developed by high-resource teams. This approach interprets methodology-use citations as a proxy for resource transfer from the cited institutions to the citing institutions, as the increasing prevalence of methodology-use citations has been linked to the growing availability of reusable technologies and evaluations (Jurgens et al., 2018; Jones, 1994).

2 Data and Methods

In this work, we synthesized data from various sources and trained two prediction models to generate variables for downstream analyses. To study research topic shift, we retrieved titles and abstracts of ACL Anthology papers from the ACL-OCL corpus (Rohatgi et al., 2023). Each paper's author affiliation metadata was retrieved from OpenAlex, an openly accessible database containing metadata on scientific research publications (Priem et al., 2022). To estimate each institution's resource level, we generated a proxy variable predicted by a machine learning model trained on research expenditure data and bibliometric features. Citation context data were obtained through Semantic Scholar's S2AG API (Wade, 2022). To analyze patterns of methodological diffusion, we fine-tuned a citation intent classifier to identify method-use citations, which are instances where the citing paper adopts tools, models, or methods from the cited work. The following subsections describe the above tasks respectively.

2.1 Modeling Research Topic Shift

The ACL-OCL corpus includes the full text of 73k papers from the ACL Anthology up to September 2022. We selected the papers published since 2010, since AI research intensified in the past 10-15 years (See Figure 1).

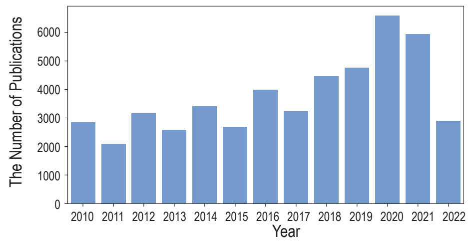

Figure 1: The number of papers included in ACL-OCL dataset each year.

We used paper titles and abstracts as input for topic modeling. We first embedded each paper's title and abstract using the SPECTER 2 language model (Cohan et al., 2020). The resulting embeddings were then reduced in dimensionality using UMAP. HDBSCAN was then applied to further clustered the dimension-reduced embeddings into topic clusters.

Using topic coherence (Röder et al., 2015) as the evaluation metric, we compared BERTopic (Grootendorst, 2022) under various parameter settings. The highest coherence score was achieved by a BERTopic model configured with 275 neighbors, 125 UMAP components, and a minimum cluster size of 275 for HDBSCAN.

We further validated the model by manually reviewing sample articles and representative keywords from each topic cluster, comparing them with the ACL submission topics. Using this comparison, we were able to assign each topic cluster a label based on keywords from the ACL topic list (see Table 3 in the appendix). Finally, each paper was assigned a topic label based on the highest topic probability generated by the topic model.

We then calculated the topic shift ratio to measure the annual change in a research topic's popularity, i.e. whether it gained or lost attention, using the following equation:

$$Topic\ Shift\ Ratio = \frac{P(Paper\ is\ assigned\ to\ topic\ X \mid Paper\ Published\ in\ year\ Y)}{P(Paper\ is\ assigned\ to\ topic\ X \mid Paper\ published\ before\ year\ Y)} \tag{1}$$

If the topic shift ratio for Topic X is greater than 1 in Year Y, it means that Topic X became more prevalent in Year Y, compared to its prevalence in the years before.

A slightly modified equation was designed to compare the popularity of a topic before and after a cutoff year, e.g. 2016:

$$Topic\ Shift\ Ratio = \frac{P(Paper\ is\ assigned\ to\ topic\ X \mid Paper\ Published\ after\ year\ Y)}{P(Paper\ is\ assigned\ to\ topic\ X \mid Paper\ published\ before\ year\ Y)} \tag{2}$$

2.2 Estimating Institutional AI Resources

We trained a regression model to estimate an institution's AI resource. The training data comes from the 2023 Higher Education Research and Development (HERD) Survey. The HERD Survey is an annual census of 501 U.S. colleges and universities that expended at least $150,000. The survey includes data for various research areas. We used expenditure in the area of information and computer science as a proxy measure for an institution's AI research resource.

Bibliometric features have long served as tools in the science of science (Fortunato et al., 2018) and the scientometric community (Leydesdorff and Milojević, 2012). Employing these features allows for investigations of the characteristics and dynamics inherent in scientific activities and entities. We extracted author affiliations for each ACL paper from OpenAlex, an openly accessible database containing metadata on scientific research publications. For each institution, we aggregated 15 bibliometric features, including (1) basic counts, i.e. the number of publications, citations, co-institutions, and researchers for each institution; (2) researcher seniority for each institution including the mean, median, min, and max researcher h-index; (3) outbound citation targets, such as the number of unique authors, institutions, and publishing venues (such as journals, conference proceedings); (4) outbound citations aggregated at different research entity level, such as author-, institution-, and publishing venue-level citations. See Table 5 in the appendix for a full list of features and their definitions.

These features reflect an institution's research output and impact in the NLP community. The number of publications and the number of researchers are used to represent the capacity of each institution: a larger university usually has more faculty members and students, and thus generates more publications. A researcher has an h-index of h if they have published h papers, each of which has been cited at least h times. Aggregated researcher h-index data for an institution indicates its research capacity and prestige. Citations from other research teams, institutions, and publishing venues represent the reputation of the institution in the research community. Conversely, citations made by the institution's researchers reflect their engagement with and awareness of the field, serving as a proxy for their research capability too.

If a paper involved multiple institutions, it was counted toward each affiliated institution.

Using the bibliometric features and expenditure data as training data, we trained and cross-validated linear regression and random forest regression models with different hyper-parameters. The model with the best performance is a random forest regression model with the maximum depth set to 20, minimal samples split set to 2. The model achieves 0.407 R-squared on the testing dataset and 0.907 R-squared on the training dataset.

Using the best prediction model and institution-level bibliometric data for all institutions that ACL-OCL authors are affiliated with, we predicted a pseudo-expenditure value for each affiliation as a proxy for the amount of their AI resource. The feature importance of the model is shown in Table 4 in the appendix. We found that the features "Number of citations" and "Number of publications" are among the most important features. It makes sense since research spending should be positively correlated with the size and capacity of institutions.

2.3 Identifying Methodology-use Citations

For each citation to ACL-OCL papers, we retrieved the citation context, or the citation sentence, from the S2AG database. We applied the method proposed by (Shui et al., 2024) because it doesn't require external data such as author and affiliation information to achieve performance comparable to the state-of-the-art. We fine-tuned a SciBERT model using SciCite and ACL-ARC data as a multi-task learning task.

The ACL-ARC dataset (Jurgens et al., 2018) provides annotations for citation intents with six classes, including Extends, Future, Motivation, Compares, Uses, and Background, for 1,969 citation sentences from 10 ACL Anthology articles.

The SciCite data set includes 11,020 citation sentences from computer science and medicine articles sampled from the Semantic Scholar corpus (Cohan et al., 2019). The SciCite data schema was simplified based on ACL-ARC, after removing citation intent categories that are rare or not useful for meta-analysis of scientific publications. SciCite includes three categories: *background information*, *use of methods*, and *comparing result*. Here we refer to them briefly as *background*, *methodology*, and *result* citations.

Using SciCite as the main training set and ACL-ARC as the auxiliary training set, the resulting model has achieved a macro 0.86 F1 score on the SciCite dataset, with balanced precision and recall values on all categories. This result is comparable to (Paolini et al., 2024). See Table 1 for category-level performance.

Table 1: Performance for the citation intent classifications task.

Class	Precision	Recall	F1 Score
Background	0.89	0.91	0.90
Methodology	0.81	0.79	0.80
Result	0.88	0.87	0.87
Macro			0.86

Using the fine-tuned citation intent classification model, we predicted citation intent for each citation context retrieved from S2AG.

3 Result

3.1 Research Topic Shifts in NLP

Figure 2 illustrates annual changes in NLP research topics from 2012 to 2022, revealing several clear trends. Most notably, 2016 emerges as a key turning point: LLM-driven topics, such as *Generation* and *Language Model*, have gained significant popularity since 2016, while formerly core NLP areas like *Machine Translation* and *Syntax/Parsing* have been declining.

This shift reflects a reorientation in the NLP field following architectural breakthroughs: the Transformer architecture was introduced in 2017 (Vaswani et al., 2017), followed by GPT-1 in 2018 (Radford et al., 2018) and BERT in 2019 (Devlin et al., 2019). These innovations contributed to the rise of general-purpose models and generative tasks (Ma et al.; Touvron et al., 2023; Lewis et al., 2021; Ramesh et al., 2021), drawing attention away from narrower, task-specific approaches, demoralizing researchers in those areas (Togelius and Yannakakis, 2023).

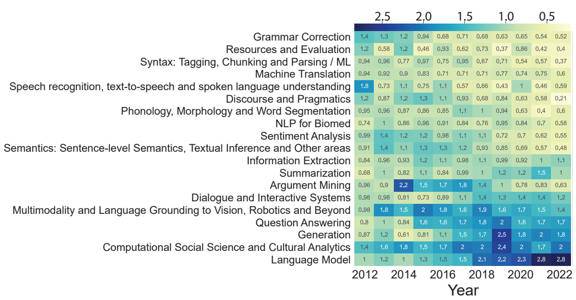

Figure 2: Topic shift ratios for each topic by year from 2012 to 2022.

In addition to these LLM-centered topics, Figure 2 also shows the emergence of interdisciplinary and application-oriented topics, such as *Computational Social Science and Cultural Analysis*, *Multimodality*, and *Dialogue Systems*. This trend supports recent findings on diversification of NLP research directions (Gururaja et al., 2023).

To further quantify these shifts, we calculated topic shift ratios before and after the cutoff year 2016. Figure 3 visualizes annual shifts for the five most trending and five most declining topics, based on their shift ratios between pre- and post-2016 periods.

Figure 2 also shows plateauing or moderate growth in some topics, such as *Summarization*, *Information Extraction*, and *sentiment analysis*, suggesting topic saturation. This pattern aligns with LLMs' strong performance on these tasks. The progress of these maturing areas may require new angles, such as multilinguality, interpretability, or social applications.

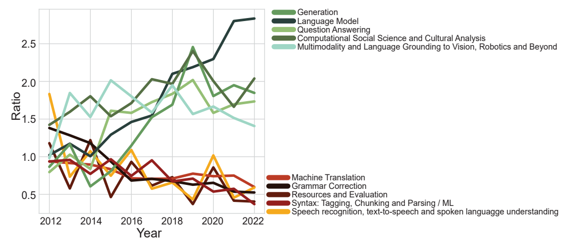

Figure 3: Annual trends in topic ratios for the five most trending and five most declining topics.

3.1.1 Topic Differences Between Low-resource and High-resource Teams

Next, we compared the distribution of research topics between high-resource and low-resource teams. Using our regression model to estimate each institution's AI resource level, we classified the top 10% of institutions as high-resource, and the remaining 90% as low-resource. We define a research team as the group of authors on a single paper. The team's resource level is determined by the highest-ranked institution among the authors' affiliations. We then assigned each paper a topic shift ratio, based on its publication year and its assigned research topic (see Figure2).

We found that papers from high-resource teams were associated with significantly higher topic shift ratios than those from low-resource teams, indicating that high-resource teams are more likely to publish on trending topics. This difference is statistically significant, as shown by a Mann–Whitney U test (U = 230,316,303.0, p < 0.001).

We then compared the topic distributions between low-resource and high-resource research teams by counting the number of papers published in each topic and applying chi-squared tests to assess statistical differences. Figure 4 shows the residuals from these tests for papers published in 2016 ($\chi^2 = 70.913$, $p < 0.0001$), 2018 ($\chi^2 = 55.188$, $p < 0.0001$), and 2020 ($\chi^2 = 58.455$, $p < 0.0001$). The topics are ordered by their overall trend, with those declining in popularity near the top and trending topics near the bottom (based on topic shift ratios calculated in 2016 and held consistent across all panels).

Positive residuals (in blue color) indicate over-representation by high-resource teams, and negative residuals (in red color) indicate over-representation by low-resource teams. The pattern across all three years suggests a persistent topic divide: low-resource teams are more concentrated in declining topics, while high-resource teams increasingly dominate emerging and computationally intensive areas.

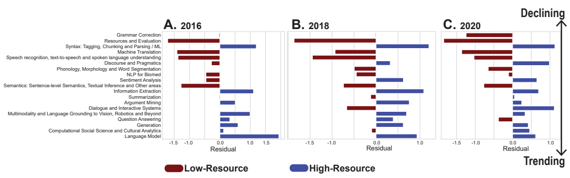

Figure 4: Topic differences between low-resource and high-resource institutions. Residual of chi-squared test comparing topic distribution between papers published by low-resource teams and high-resource teams for papers published after 2016 (A), 2018 (B), and 2020 (C).

3.2 Comparing Citation-Mediated Resource Transfer in Trending Topics and Declining Topics

Considering methodology-use citation as an indicator of "resource transfer" from one institution to another, we analyzed the intent of citations to the ACL Anthology papers. Overall, about one half of citations are background information, about one third on methdology use, and the remaining on result discussion. Figure 5A presents the proportions of these citation intent types, normalized by the number of citations per year, from 2010 to 2022, illustrating a trend that background citations are increasing (Mann Kendall test: $\tau = 0.923, p < 0.0001$), while methodology-use citations (Mann Kendall test: $\tau = -0.718, p < 0.001$) and result citations (Mann Kendall test: $\tau = -0.923, p < 0.001$) have been declining.

As NLP literature expands, it is not surprising to see researchers citing more prior work as background information. However, the decrease in methodology-use citations needs further examination to see whether it indicates a decline in resource transfer, since the growing resource gap may prevent low-resource teams from adopting certain methods due to limitations in computing power, data access, and funding. If this is true, we should expect the increase in background citations and the decrease in methodology-use citations to be more pronounced in citations from low-resource teams to high-resource teams, especially among trending topics, since high-resource teams are more likely to work on trending topics.

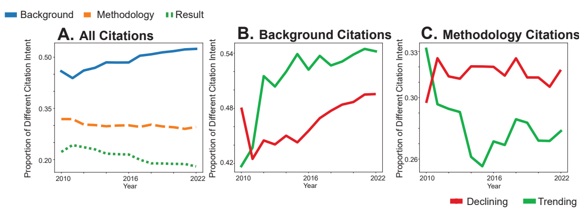

Figure 5: Citation intent patterns for papers on trending and declining topics. The proportion of citation intent for citations to all ACL-OCL papers published after 2010 (A); the proportion of background (B), methodology (C) citations to declining topics and trending topics.

To compare methodology-use citations in trending and declining topics, we selected papers from the top five most trending topics (*Language Model, Computational Social Science and Cultural Analytics, Generation, Question Answering, Multimodality and Language Grounding to Vision, Robotics and Beyond*), and the top five most declining topics (*Grammar Correction, Resources and Evaluation, Syntax: Tagging, Chunking and Parsing / ML, Machine Translation, Speech recognition*).

Figure 5B shows that background citations have increased over time for both trending and declining topics. This trend is supported by the Mann-Kendall test results: for trending topics, $\tau = 0.769$, $p < 0.001$, with a slope of 0.0046 and intercept of 0.50; for declining topics, $\tau = 0.718$, $p < 0.001$, with a slope of 0.0061 and intercept of 0.43. While both trends are significant, the higher intercept for trending topics suggests that they generally require more background citations than declining topics. This is consistent with expectations: researchers working in fast-evolving areas may need to cite a broader base of prior work to contextualize and support their arguments.

Figure 5C shows the proportion of methodology-use citations over time for both trending and declining topics. For trending topics, the Mann-Kendall test reveals a decreasing trend in Methodology citation proportion ($slope = -0.0022, intercept = 0.2911$), whereas declining topics show a more stable and higher baseline level ($slope = -0.0001, intercept = 0.3150$). This suggests that in fast-moving areas, researchers are less likely to cite existing models, datasets, or tools.

3.2.1 Citations from Low-resource Teams to High-resource Teams

We further examined whether low-resource teams experienced a more pronounced decline in methodology citations to work produced by high-resource teams.

Figure 6 clearly shows a declining trend of methodology-use citations from low-resource teams to high-resource teams, with the downward trend accelerating after 2016 ((Mann Kendall test: $\tau = -0.846, p < 0.0001, slope = -0.0058, intercept = 0.3310$). In comparison, the overall trend across all ACL papers shows a much more gradual decline (Mann Kendall test: $\tau = -0.718, p < 0.001, slope = -0.0014, intercept = 0.3082$).

The strong and accelerating decrease in methodology-use citations from low-resource teams to high-resource teams suggests that it is increasingly more challenging for low-resource teams to engage with or build upon the methodologies developed by high-resource teams.

3.2.2 Citations from Low-resource Teams to High-resource Teams between Trending Topics and Declining Topics

Combining the resource and topic factors, we further compared methodology-use citations from low-resource teams to high-resource teams among trending and declining topics.

Figure 7C shows the same trend that methodology-use citations have been declining for both trending topics (Mann Kendall test: $\tau = -0.600, p < 0.05, slope = -0.0067, intercept = 0.2983$) and declining topics (Mann Kendall test: $\tau = -0.527, p < 0.05, slope = -0.0047, intercept = 0.3448$). Additionally, methodology-use citations are

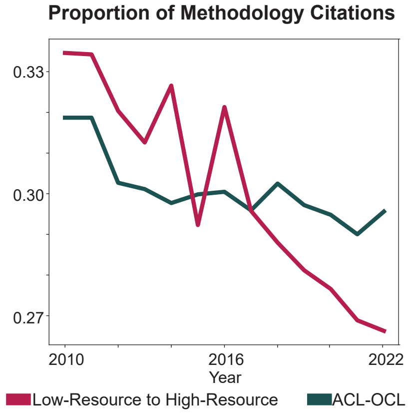

Figure 6: The proportion of methodology citations for citations from low-resource teams to high-resource teams, compared with the proportion of methodology citations to all ACL anthology papers.

consistently less prevalent in trending topics (average proportion for trending topics: 0.27; declining topics: 0.32). Both patterns support the interpretation that low-resource teams face greater challenges in adopting methods from high-resource teams, particularly in trending research areas.

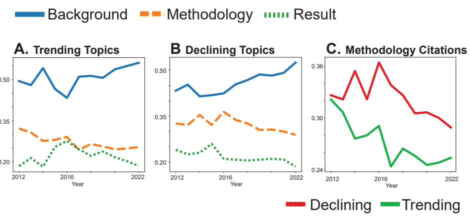

Figure 7: Citation intent from low-resource teams to high-resource teams for citations to papers with trending and declining topics. (A) and (B) show the proportions of different citation intents for citations from low-resource teams to high-resource teams for papers with trending topics (A) and declining topics (B), respectively. (C) further compares the proportion of methodology citations in trending topics and declining topics.

To further validate our findings, we conducted a linear regression analysis, using the proportion of methodology citations received by each cited paper as the dependent variable. We aggregated methodology-use citations at the paper level and included three key predictors: (1) the maximum predicted research expenditure among the cited paper's co-author affiliations as a proxy for institutional resource level, (2) the publication year, and (3) the normalized topic popularity of the cited paper based on the topic popularity ranking from Figure 2.

Table 2: Linear Regression Results

	Dependent variable:		
	Proportion of Methodology Citations		
	(1)	(2)	(3)
Topic Popularity	-0.097*** (0.002)	-0.072*** (0.007)	
Topic Popularity Gap			-0.118*** (0.013)
Year	-0.010*** (0.002)	-0.020*** (0.007)	0.020** (0.008)
Resource	-0.016*** (0.003)		
Resource Gap		-0.029** (0.013)	-0.033** (0.015)
Constant	0.329*** (0.002)	0.324*** (0.008)	0.438*** (0.009)
Observations	99,481	19,915	13,290
R^2	0.023	0.007	0.007
Adjusted R^2	0.022	0.007	0.007
Residual Std. Error	0.197 (df = 99477)	0.277 (df = 19911)	0.251 (df = 13286)
F Statistic	763.412*** (df = 3; 99477)	50.091*** (df = 3; 19911)	31.145*** (df = 3; 13286)
Note:			*p<0.1; **p<0.05; ***p<0.01

As shown in Table 2, the results indicate that the proportion of methodology-use citations is significantly and negatively associated with cited team's resource level, the popularity of the cited paper's topic, and publication year. In additional models, we find that larger gaps in resource levels and topic popularity between citing and cited teams are also significantly associated with fewer methodology citations. These findings reinforce the interpretation that institutional and topical asymmetries may increasingly constrain methodological reuse, particularly disadvantaging low-resource teams when citing high-resource work in trending areas.

The results combined provide more evidence that resource barriers limit the adoption of methods proposed by high-resource teams, and such a phenomenon is more serious for publications related to trending topics. According to (Jurgens et al., 2018), such a trend could indicate a decrease in reusable technologies such as models and datasets, and evaluations of tools. Such a decrease could be related to the increasing resource gap in the field of AI. We also visualized the proportion of non-methodology citations. Latour (Latour, 1987) suggests that non-methodological citations are important for defending proposed ideas. We can interpret from the increase in non-methodology citations that there is an increased need to defend newly proposed ideas, and increasingly less consensus in the ACL anthology community. Such interpretation makes sense as AI is fast-growing, and new ideas need to be introduced to an increasingly more interdisciplinary community.

4 Discussion

Our finding shows that high-resource research teams have been focusing on trending AI topics, while low-resource teams focus on declining AI topics. This might suggest that access to resources such as high-performance computing infrastructure and large-scale datasets could influence which research directions are feasible to pursue. Low-resource institutions may be constrained to explore research directions that do not require extensive computational power, even if those areas are becoming less relevant to the AI research community (Togelius and Yannakakis, 2023). Our study also shows that research produced by high-resource teams is becoming increasingly difficult for other researchers, especially those from low-resource institutions, to utilize or build upon.

This phenomenon has long-term consequences for the diversity of AI research. If certain topics and methods become inaccessible due to resource constraints, researchers with fewer resources may struggle to contribute meaningfully to emerging research areas. This could lead to an increased concentration of influence within a small group of well-resourced institutions, limiting the diversity of perspectives and innovation in AI (Abdalla et al., 2023; Togelius and Yannakakis, 2023).

This raises important questions about the role of high-resource organizations, such as industry companies, in shaping the AI research landscape. Industry research teams have played a critical role by developing large-scale computational infrastructure, curating extensive datasets, and publishing impactful AI research. Industry-academia collaborations have traditionally been seen as a means of technology transfer—where academic discoveries are translated into practical applications. The current dynamics suggest that industry research may be operating in a way that restricts, rather than facilitates, knowledge dissemination. If left unaddressed, this could exacerbate disparities in AI research and limit opportunities for researchers outside of well-funded institutions (Ahmed et al., 2023).

Given these findings, this study highlights the need for actions from researchers, universities, companies, policymakers, funding agencies, to bridge the gap between high-resource and low-resource institutions. Our result could better inform strategies to ensure that the field remains accessible, diverse, and open to innovation from a broader range of contributors.

5 Conclusion

In this work, by analyzing the research topics and citation intent, we investigate the disparities between low-resource and high-resource institutions in the natural language processing research community. Our findings indicate that high-resource teams have been focusing on research topics that are gaining popularity, whereas low-resource teams have been more likely to work on topics that are becoming less prominent. This suggests that access to resources such as computational power and large datasets plays a significant role in determining what research topic a team can study. Furthermore, our result reveals that research produced by high-resource teams is becoming increasingly difficult for other researchers to build upon. Such results suggest a growing divide in AI research, where advancements driven by high-resource industry corporations and universities may inadvertently limit the accessibility of cutting-edge research to those with fewer resources. These findings indicate the need for more inclusive research practices and collaborative efforts to ensure that AI innovation remains accessible to a broader research community. Future work should explore potential strategies for bridging this gap.

Limitations

We are only focusing on research work before 2023 due to the limitations of the ACL-OCL dataset. However, we would argue that with the newer advancements such as GPT-4 and ChatGPT (Ma et al.; OpenAI, 2023), the trend we observed in this study should remain valid if not more severe. While we compared low-resource and high-resource teams, researchers with different affiliations, such as industry companies and academia universities, do research for different reasons, and the funding mechanisms for different institutions are different. Our findings do not account for such differences. On the other hand, impact from high-resource teams is not only present in publications, but also in other means, such as software packages. In our work, using only citations, we cannot account for the use of software packages. Lastly, most findings from this work are correlational rather than causal, yet we believe our results could provide evidence future work could build upon.

References

Mohamed Abdalla, Jan Philip Wahle, Terry Ruas, Aurélie Névéol, Fanny Ducel, Saif Mohammad, and Karen Fort. 2023. The Elephant in the Room: Analyzing the Presence of Big Tech in Natural Language Processing Research. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 13141–13160, Toronto, Canada. Association for Computational Linguistics.

Nur Ahmed, Muntasir Wahed, and Neil C. Thompson. 2023. The growing influence of industry in AI research. *Science*, 379(6635):884–886. Publisher: American Association for the Advancement of Science.

Arman Cohan, Waleed Ammar, Madeleine van Zuylen, and Field Cady. 2019. Structural Scaffolds for Citation Intent Classification in Scientific Publications. *arXiv preprint*.

Arman Cohan, Sergey Feldman, Iz Beltagy, Doug Downey, and Daniel Weld. 2020. SPECTER: Document-level representation learning using citation-informed transformers. pages 2270–2282, Online.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for*

Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Santo Fortunato, Carl T. Bergstrom, Katy Börner, James A. Evans, Dirk Helbing, Staša Milojević, Alexander M. Petersen, Filippo Radicchi, Roberta Sinatra, Brian Uzzi, Alessandro Vespignani, Ludo Waltman, Dashun Wang, and Albert-László Barabási. 2018. Science of science. *Science*, 359(6379):eaao0185.

Maarten Grootendorst. 2022. Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*.

Sireesh Gururaja, Amanda Bertsch, Clara Na, David Widder, and Emma Strubell. 2023. To build our future, we must know our past: Contextualizing paradigm shifts in natural language processing. pages 13310–13325, Singapore.

Heng Huang, Donghua Zhu, and Xuefeng Wang. 2022. Evaluating scientific impact of publications: combining citation polarity and purpose. *Scientometrics*, 127(9):5257–5281.

Oana Ignat, Zhijing Jin, Artem Abzaliev, Laura Biester, Santiago Castro, Naihao Deng, Xinyi Gao, Aylin Ece Gunal, Jacky He, Ashkan Kazemi, Muhammad Khalifa, Namho Koh, Andrew Lee, Siyang Liu, Do June Min, Shinka Mori, Joan C. Nwatu, Veronica Perez-Rosas, Siqi Shen, and 3 others. 2024. Has It All Been Solved? Open NLP Research Questions Not Solved by Large Language Models. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 8050–8094, Torino, Italia. ELRA and ICCL.

Xiaorui Jiang and Junjun Liu. 2023. Extracting the evolutionary backbone of scientific domains: The semantic main path network analysis approach based on citation context analysis. *Journal of the Association for Information Science and Technology*, 74(5):546–569.

Karen Sparck Jones. 1994. Natural Language Processing: A Historical Review. In *Current Issues in Computational Linguistics: In Honour of Don Walker*, pages 3–16. Springer, Dordrecht.

David Jurgens, Srijan Kumar, Raine Hoover, Dan McFarland, and Dan Jurafsky. 2018. Measuring the Evolution of a Scientific Field through Citation Frames. *Transactions of the Association for Computational Linguistics*, 6:391–406.

Bruno Latour. 1987. *Science in action: how to follow scientists and engineers through society*. Harvard University Press, Cambridge, Mass.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2021. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *arXiv preprint*. ArXiv:2005.11401 [cs].

Loet Leydesdorff and Staša Milojević. 2012. Scientometrics. *arXiv preprint arXiv:1208.4566*.

Michael L. Littman, Ifeoma Ajunwa, Guy Berger, Craig Boutilier, Morgan Currie, Finale Doshi-Velez, Gillian Hadfield, Michael C. Horowitz, Charles Isbell, Hiroaki Kitano, Karen Levy, Terah Lyons, Melanie Mitchell, Julie Shah, Steven Sloman, Shannon Vallor, and Toby Walsh. 2022. Gathering Strength, Gathering Storms: The One Hundred Year Study on Artificial Intelligence (AI100) 2021 Study Panel Report. Publisher: arXiv Version Number: 1.

Wei Ma, Shangqing Liu, Wenhan Wang, Q. Hu, Yefei Liu, Cen Zhang, Liming Nie, and Yang Liu. ChatGPT: Understanding Code Syntax and Semantics.

Rajiv Movva, Sidhika Balachandar, Kenny Peng, Gabriel Agostini, Nikhil Garg, and Emma Pierson. 2023. Topics, Authors, and Networks in Large Language Model Research: Trends from a Survey of 17K arXiv Papers. Publisher: arXiv Version Number: 2.

Kai Nishikawa. 2023. How and why are citations between disciplines made? A citation context analysis focusing on natural sciences and social sciences and humanities. *Scientometrics*, 128(5):2975–2997.

OpenAI. 2023. GPT-4 Technical Report.

Lorenzo Paolini, Sahar Vahdati, Angelo Di Iorio, Robert Wardenga, Ivan Heibi, and Silvio Peroni. 2024. Why do you cite? an investigation on citation intents and decision-making classification processes. *arXiv preprint arXiv:2407.13329*.

Dylan Patel. 2023. Google Gemini Eats The World – Gemini Smashes GPT-4 By 5X, The GPU-Poors.

Jason Priem, Heather Piwowar, and Richard Orr. 2022. Openalex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. *arXiv preprint arXiv:2205.01833*.

Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, and 1 others. 2018. Improving language understanding by generative pre-training.

Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. 2021. Zero-Shot Text-to-Image Generation. *arXiv preprint*. ArXiv:2102.12092 [cs].

Shaurya Rohatgi, Yanxia Qin, Benjamin Aw, Niranjana Unnithan, and Min-Yen Kan. 2023. The ACL OCL corpus: Advancing open science in computational linguistics. pages 10348–10361, Singapore.

Michael Röder, Andreas Both, and Alexander Hinneburg. 2015. Exploring the Space of Topic Coherence Measures. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, WSDM ’15, pages 399–408, New York, NY, USA. Association for Computing Machinery.

Zeren Shui, Petros Karypis, Daniel S. Karls, Mingjian Wen, Saurav Manchanda, Ellad B. Tadmor, and George Karypis. 2024. Fine-Tuning Language Models on Multiple Datasets for Citation Intention Classification. *arXiv preprint*. ArXiv:2410.13332 [cs].

Julian Togelius and Georgios N. Yannakakis. 2023. Choose Your Weapon: Survival Strategies for Depressed AI Academics. Publisher: arXiv Version Number: 1.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. LLaMA: Open and Efficient Foundation Language Models. *arXiv preprint*. ArXiv:2302.13971 [cs].

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Alex D Wade. 2022. The semantic scholar academic graph (s2ag). In *Companion Proceedings of the Web Conference 2022*, pages 739–739.

A Appendix

Table 3: ACL submission topics

Topic Name
Machine Translation
Dialogue and Interactive Systems
Sentiment Analysis
Information Extraction
Question Answering
Syntax: Tagging, Chunking and Parsing / ML
Summarization
Semantics: Sentence-level Semantics, Textual Inference and Other areas
NLP for Biomed
Speech recognition, text-to-speech and spoken language understanding
Computational Social Science and Cultural Analytics
Language Model
Semantics: Sentence-level Semantics, Textual Inference and Other areas
Multimodality and Language Grounding to Vision, Robotics and Beyond
Syntax: Tagging, Chunking and Parsing / ML
Discourse and Pragmatics
Phonology, Morphology and Word Segmentation
Grammar Correction
Resources and Evaluation
Generation
Argument Mining

Table 4: Feature importance for the random forest model

Feature Name	**Feature Importance**
Number of citations	0.339028
Number of publications	0.172261
Number of venue-level references	0.083720
Number of venue-level references	0.083224
Number of works referenced	0.050995
Average researcher h-index	0.048643
Number of co-institutions	0.035174
Number of institutions cited	0.033076
Number of institution-level references	0.032882
Maximum researcher h-index	0.028772
Number of researchers	0.027982
Number of institution-level references	0.026424
Median researcher h-index	0.025503
Number of authors cited	0.012300
Minimum researcher h-index	0.000015

Table 5: Bibliometric features used for predicting research expenditures for information and computing science.

Name	**Description**
Number of publications	The number of publications affiliated with the institution of interest
Number of citations	The number of citations received by publications affiliated with the institution of interest
Number of co-institutions	The number of other institutions collaborating with the institution of interest on a publication
Number of researchers	The number of researchers affiliated with the institution of interest
Average researcher h-index	Average h-index of researchers affiliated with the institution of interest
Maximum researcher h-index	Maximum h-index of researchers affiliated with the institution of interest
Minimum researcher h-index	Minimum h-index of researchers affiliated with the institution of interest
Median researcher h-index	Median h-index of researchers affiliated with the institution of interest
Number of works cited	The number of works cited by publications affiliated with the institution of interest
Number of author-level citations	The number of times authors not affiliated with the institution of interest got cited by work affiliated with the institution of interest
Number of authors cited	The number of authors cited by publications affiliated with the institution of interest
Number of venue-level citations	The number of times other venues got cited by work affiliated with the institution of interest
Number of venues cited	The number of venues cited by publications affiliated with the institution of interest
Number of institution-level citations	The number of times other institutions got cited by work affiliated with the institution of interest
Number of institutions cited	The number of institutions cited by publications affiliated with the institution of interest