Has ACL Lost Its Crown? A Decade-Long Quantitative Analysis of Scale and Impact Across Leading AI Conferences

Anonymous ACL submission

Abstract

The recent surge of language models (LMs) has rapidly expanded NLP/AI research, driving an exponential rise in submissions and acceptances at major conferences. Yet this growth has been shadowed by escalating concerns over conference quality, e.g., plagiarism, reviewer inexperience and collusive bidding. However, existing studies rely largely on qualitative accounts (e.g., expert interviews, social media discussions, etc.), lacking longitudinal empirical evidence. To fill this gap, we conduct a ten-year empirical study (2014–2024) spanning seven leading conferences. We build a four-dimensional bibliometric framework covering conference scale, core citation statistics, impact dispersion, cross-venue and journal influence, etc. Notably, we further propose a metric *Quality–Quantity Elasticity* (QQE), which measures the elasticity of citation growth relative to acceptance growth. We highlight two key findings: (1) conference expansion does not lead to proportional growth in scholarly impact, as QQE consistently declines over time across all venues; (2) ACL has not lost its crown, continuing to outperform other NLP conferences in median citations, milestone contributions, and citation coverage. This study provides the first decade-long, cross-venue empirical evidence on the evolution of major NLP/AI conferences[1].

1 Introduction

The evolution from neural language models to large language models (LLMs) has propelled NLP research to global prominence, leading to an exponential surge in conference submissions and acceptances (Naveed et al., 2025). For example, the ACL main conference grew from 692 long-paper submissions and 173 acceptances in 2015 to 4,407 submissions and 865 acceptances in 2024, with an increase of 536% and 400%, respectively (Ku et al., 2024), as shown in Fig. 1. While this expansion reflects the vitality of the NLP research, it also introduces new challenges for academic governance and quality assurance.

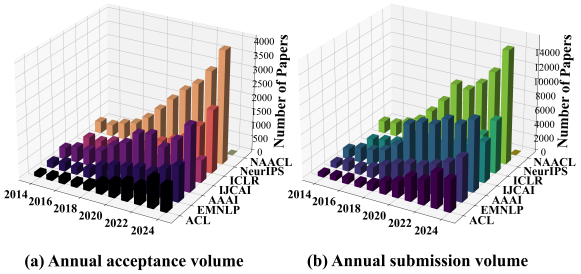

Figure 1: Annual submissions and acceptances across major NLP/AI conferences from 2014 to 2024.

Recent years have seen increasing reports of plagiarism, excessive emphasis on storytelling over science. At the same time, the peer-review system shows signs of strain, including collusive bidding, reviewer overload and the rising involvement of inexperienced reviewers. Well-known controversies, such as the "*Who's Adam?*"[2], *the AAAI bid-group coordination incident*[3], and the community's satirical nickname for ICLR (*I Can pay the biLl foR you*)[4], have further amplified skepticism toward the robustness of peer review. An increasing number of researchers are beginning to question whether such major conferences still represent the genuine frontier of technical and theoretical progress.

Existing studies approach this issue through two limited perspectives: (1) analytical work on peer-review mechanisms (Rogers and Augenstein, 2020;

[1] Our codes: https://anonymous.4open.science/r/acl-crown-analysis-38D5.

[2] A NeurIPS 2025 review mistakenly treats the Adam optimizer as a person's name.

[3] A private group of authors and reviewers allegedly coordinated their bid selections.

[4] The ICLR incident involving a score-0 review and the author publicly accusing the reviewer of being "mentally ill" and offering to pay for treatment.

Wei et al., 2025), and (2) publication-ranking analyses based on citation counts (Croft and Sack, 2022). In addition, qualitative investigations, such as expert interviews, case analyses, and social media discourse[5] also provide valuable insights. However, they neither conduct longitudinal, cross-venue quantitative analyses nor develop a unified metric for linking scale expansion with shifts in scholarly impact.

To fill this gap, we compile the complete long-paper dataset from seven leading NLP/AI conferences, i.e., ACL, EMNLP, NAACL, AAAI, IJCAI, ICLR and NeurIPS, covering the years 2014 to 2024 and totaling 60,406 long papers. We develop a four-dimensional bibliometric framework comprising: *conference scale* (e.g., annual paper growth rate), *scholarly influence* (e.g., median citations, H-index), *cross-venue/journal citation patterns* (e.g., top-conference citation share) and *Quality–Quantity Elasticity* (QQE), yielding 14 indicators in total. The QQE is proposed to quantify the elasticity of scholarly-impact growth relative to acceptance growth, enabling the detection of potential "inflated prosperity" in conference expansion.

Empirical results show that: (1) conference expansion does not lead to proportional growth in scholarly impact, as QQE consistently declines over time across all venues; (2) conference differences are no longer driven by size, but by the ability to sustain impact, with ML venues most stable, AI venues weaker, and NLP venues increasingly differentiated; (3) ACL has not lost its crown, continuing to outperform other NLP conferences in median citations, milestone contributions, and citation coverage. Our work makes the following main contributions:

- We provide the first ten-year quantitative analysis across seven major NLP/AI conferences.
- We propose *QQE*, an elasticity-based metric linking acceptance growth to citation-impact growth.
- We directly test whether ACL has lost its crown and show that it remains the leading NLP venue.

2 Related Work

Conference quality assessment has long been recognized as a critical issue in computer science, dating back to the CRA *Best Practices Memo* (1999). Subsequent studies largely adopt journal-style bibliometric methods, relying on citation-based indicators such as citation counts, H-index, CiteScore, Y-factor, and machine-learning–based ranking models (Vrettas and Sanderson, 2015; Mohammad, 2020; Meho, 2019; Iqbal et al., 2021). However, they primarily provide static rankings and rarely examine the long-term, cross-year evolution of individual conferences, particularly under the rapid expansion and structural volatility of NLP/AI research.

Complementary work evaluates conference quality through expert- and community-based mechanisms, including curated ranking systems such as CORE, MSAR, and the CCF recommendations. While expert judgments provide domain-informed coarse-grained categorizations, they are susceptible to subjectivity, regional bias, and limited sensitivity to intra-tier variation or year-to-year change (Rogers and Augenstein, 2020; Kuznetsov et al., 2019; Russo et al., 2025). However, such perspectives remain largely anecdotal and lack a unified, data-driven framework for assessing whether conference growth is accompanied by sustained scholarly impact.

Conference	No. of Papers	Citations
ACL	4,965	488,499
EMNLP	7,106	573,104
NAACL	2,540	414,779
AAAI	12,786	713,125
IJCAI	6,617	245,391
ICLR	7,691	1,771,388
NeurIPS	18,701	2,375,612
Total	**60,406**	**6,581,898**

Table 1: Paper and citation statistics.

3 Methodology

3.1 Data Acquisition

(1) Venue Selection. We focus on seven widely recognized leading conferences: the core NLP venues *ACL*, *EMNLP*, and *NAACL*; and the core AI venues *AAAI*, *IJCAI*, *NeurIPS*, and *ICLR*. We restrict our analysis to main-track long papers. Findings papers, short papers, and workshop papers are excluded.

(2) Temporal Scope. The study spans 2014–2024. The year 2014 serves as a pre-expansion baseline, preceding the rapid growth of

[5] https://yoshuabengio.org/2020/02/26/time-to-rethink-the-publication-process-in-machine-learning/

ACL submissions, while 2024 is chosen as the latest year with complete and comparable proceedings across all venues[6].

(3) Data Sources. Because Google Scholar does not provide an official API due to licensing restrictions, we collect paper metadata and citation statistics from established open sources. Paper titles are obtained from the official conference proceedings. Citation data are primarily sourced from Semantic Scholar, which successfully matches 99.3% of all papers in our dataset, and OpenAlex is used only as a supplemental source to cover the remaining unmatched entries (0.7%).

(4) Data Cleaning and Consistency Check. The data-cleaning process includes deduplication, removal of incomplete metadata and consolidation of citation statistics for papers appearing across multiple databases. No author- or institution-level data were collected, as the study does not involve analyses at the individual level. These procedures are aligned with best practices in bibliometric research (Waltman and van Eck, 2012).

Finally, the dataset contains 60,406 long papers across the seven conferences, with more than 6M citation records, as shown in Tab. 1 (Details please see App. A and B).

3.2 Multi-Dimensional Bibliometric Framework

We develop a four-dimensional bibliometric framework that enables a more holistic characterization of conference evolution, as shown in Fig. 2.

(1) Scale Dimension. This dimension quantifies the growth of each conference as a scholarly venue. It includes three indicators: *(i) publication count (PC)*, the total number of accepted long papers, reflecting the venue's output capacity; and *(ii) annual paper growth rate (APGR)*, the year-over-year percentage increase in accepted papers, which captures how rapidly a conference expands in size. *(iii) annual citation growth rate (ACGR)*, the year-over-year percentage increase in total citations accumulated by papers from each conference.

(2) Influence Dimension. To capture the magnitude and distribution of scholarly influence, we use six indicators: *(i) average citations (AC)* and *(ii) median citations (MC)*, representing central tendencies of citation performance; *(iii) highly cited ratio (HCR)* (share of papers with more than 100 citations), measuring the upper tail of impact; and

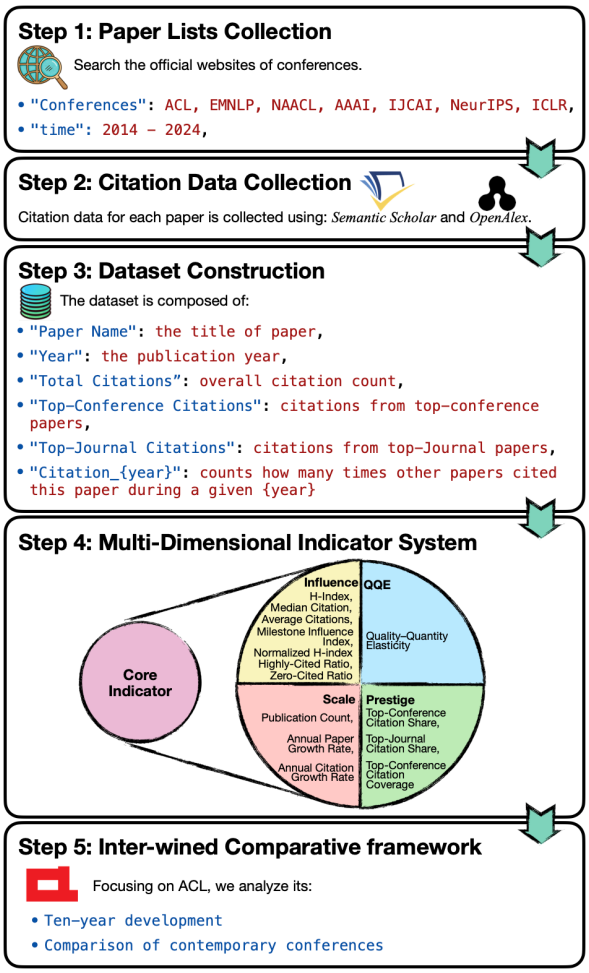

Figure 2: The proposed bibliometric framework.

(iv) zero-cited ratio (ZCR), quantifying the share of papers that never attracted citations; *(v) H-index*, which reflect the concentration of highly influential papers; *(vi) milestone influence index (MII)*, defined as the proportion of papers exceeding both an absolute citation threshold ($\geq$1000 citations)[7] (Wang et al., 2021) and a relative threshold (top 0.1% of the annual citation distribution) (Wang et al., 2023).

(3) Prestige Dimension. Beyond citation volume, we analyze citation sources to situate a venue in the scholarly network. We compute: *(i) top-conference citation share (TCS)*, the fraction of citations coming from leading peer conferences; *(ii) top-journal citation share (TJS)*, the fraction coming from elite journals, reflecting broader uptake; and *(iii) top-conference citation coverage (TCC)*, the fraction of the venue's papers that receive at least one citation from a top conference, reflecting penetration breadth. Top conferences are defined

[6] Up to October 2025.

[7] Scientific milestones lack a unified definition and are commonly proxied using citation-based indicators in scientometric studies.

as the union of CCF A-level and CORE A* venues, and top journals are represented by *Nature* and *Science* (Ioannidis, 2006).

(4) Quality–Quantity Elasticity (QQE). While static bibliometric indicators capture snapshots of scale or impact, they fail to track whether expansion is supported by a proportional diffusion of knowledge. To quantify this dynamic coupling, we propose *Quality–Quantity Elasticity (QQE)*, grounded in scientific growth theory (Aghion et al., 1998). This framework posits that an efficient venue should exhibit proportional or superlinear growth in citation-based influence relative to its paper volume, as a larger knowledge stock should generate greater spillover effects.We define C_t as the citation mass and N_t as the number of accepted long papers in year t. The annual elasticity is calculated as:

$$QQE_t = \frac{C_t/C_{t-1}}{N_t/N_{t-1}}. \tag{1}$$

This metric captures how scholarly influence scales relative to conference size by comparing their respective multiplicative growth rates. To distinguish between operational modes, we define the log-growth of scale as $s_t = \ln(N_t/N_{t-1})$, where $s_t > 0$ denotes *expansion* and $s_t < 0$ denotes *contraction*. This unified framework allows for a consistent interpretation of growth efficiency across diverse institutional regimes.

Impact-Scale Elasticity Regime. The relationship between citation growth and scale expansion is characterized by QQE_t: *(i)* $|QQE_t| > 1$ *(Effective Elasticity).* Indicates that citation-based influence changes faster than scale. When $s_t > 0$, this corresponds to *efficient expansion*; when $s_t < 0$, it reflects *quality-concentrating contraction*, where impact density increases despite shrinking volume. *(ii)* $|QQE_t| = 1$ *(Unitary Elasticity).* Indicates proportional change, where citation mass evolves in direct alignment with conference scale. *(iii)* $|QQE_t| < 1$ *(Ineffective Elasticity).* Indicates lagging impact gains, corresponding to *impact dilution* during expansion ($s_t > 0$) or *malignant contraction* during scale reduction ($s_t < 0$).

3.3 Analytical Method

We adopt a combined cross-conference and longitudinal analytical design. At the cross-sectional level, we compute ten-year averages of thirteen bibliometric indicators for ACL, EMNLP, NAACL, AAAI, IJCAI, ICLR, and NeurIPS to characterize each venue's overall scale, citation performance, impact distribution, citation composition, and growth efficiency (QQE). At the longitudinal level, we use a one-year sliding window to track temporal dynamics based on six core indicators—average and median citations (AC, MC), highly cited and zero-cited paper ratios (HCR, ZCR), QQE, and milestone influence (MII)—with additional within-NLP comparisons among ACL, EMNLP, and NAACL, enabling an analysis of how citation impact and growth trajectories evolve over time (Details please see App. D).

3.4 Normalization Controls

Because the selected conferences differ substantially in scale, raw impact indicators can systematically favor larger venues. To make fair comparisons, we apply two normalization controls to adjust for differences in publication volume.

(1) Normalized H-index. The conventional H-index increases with the number of publications, which means that larger conferences tend to obtain higher H-indices simply due to their scale. To mitigate this bias, we compute a *normalized H-index*:

$$\text{Norm-H} = \frac{H\text{-index}}{\text{total publications}} \tag{2}$$

This ratio reflects the density of highly influential papers per accepted paper.

(2) Gini Coefficient of Citation Distribution. Citations in NLP/AI conferences typically follow a long-tail pattern. As conference volume increases, this tail may expand, potentially inflating total citation counts without improving median quality. To measure this concentration, we compute the annual *Gini coefficient* for each venue. Let citation counts be sorted as $c_1 \leq c_2 \leq \cdots \leq c_n$, where n is the paper count and μ is the mean citation count. The Gini coefficient is computed as:

$$\text{Gini} = \frac{1}{n\mu}\sum_{i=1}^{n}(2i - n - 1), c_i \tag{3}$$

A value of 0 indicates a perfectly uniform distribution, while values approaching 1 signify extreme concentration in a small subset of "blockbuster" papers.

4 Experiment

4.1 Main Results

We first summarize the major empirical patterns observed across all seven conferences and all four evaluation dimensions, as shown in Tab. 2.

	Scale			Influence							Prestige			QQE
Conference	PC	APGR	ACGR	AC	MC	HCR	ZCR	H	Norm-H	MII	TCS	TJS	TCC	QQE
ACL	4965	22.06	87.89	98.39	32.00	20.75	0.89	299	6.02↑	1.09	31.26	2.04	**94.34**	1.46
EMNLP	7106	20.68	151.00	80.65	20.00	13.72	2.03	284	4.00↑	0.73	23.22	1.56	89.39	2.06
NAACL	2540	20.53	**231.63**	163.30	26.00	15.83	1.61	187	**7.36**↑	0.75	26.63	1.90	91.54	**2.34**
AAAI	12786	21.54	87.11	55.77	18.00	10.88	2.00	315	2.46↓	0.48	12.89	1.71	85.53	1.55
IJCAI	6617	5.08	129.11	37.08	14.00	7.42	**3.73**	196	2.96–	0.17	8.29	1.07	80.94	1.77
ICLR	7691	**54.63**	92.60	**230.32**	**35.00**	**26.90**	1.52	494	6.42–	**2.94**	**45.77**	**3.91**	91.91	1.61
NeurIPS	**18701**	26.44	117.74	127.03	22.00	16.47	2.34	**564**	3.02↓	1.51	26.06	2.68	85.60	1.83

Table 2: Comparison of major NLP/AI conferences across four dimensions. **Bold** indicates the highest value, and underlined indicate the lowest value.

Scale Dimension. NeurIPS ($PC = 18,701$) and AAAI ($12,786$) dominate in publication volume, followed by ICLR ($7,691$), EMNLP ($7,106$), and ACL ($4,965$), with NAACL being the smallest ($2,540$). Growth trajectories vary substantially across venues: ICLR demonstrates the most aggressive scaling ($APGR = 54.63$), whereas IJCAI remains the most stagnant (5.08). Annual citation growth (ACGR) follows a distinct hierarchy, with NAACL leading by a large margin (231.63), followed by EMNLP (151.00) and IJCAI (129.11). In contrast, AAAI exhibits the weakest citation growth (87.11), with ACL also showing relatively modest increases (87.89).

Influence Dimension. ICLR leads per-paper impact, achieving the highest average citations ($AC = 230.32$), median citations ($MC = 35.0$), and highly-cited ratio ($HCR = 26.9\%$). Although NeurIPS attains the largest absolute H-index ($H = 564$), normalized measures reveal a different hierarchy: NAACL ranks first in impact density ($Norm\text{-}H = 7.36$), followed by ICLR (6.42) and ACL (6.02), whereas volume-driven venues such as AAAI and NeurIPS exhibit substantially lower normalized impact. At the distribution tail, ACL shows the lowest zero-cited ratio (0.89%), in sharp contrast to IJCAI's highest proportion of uncited papers (**3.73%**). A similar pattern emerges in milestone influence, where ICLR (2.94) clearly dominates and NeurIPS (1.51) follows, while IJCAI generates the fewest breakthrough contributions relative to its scale (0.17).

Prestige Dimension. ICLR exhibits the strongest structural integration within elite citation networks, achieving the highest top-conference citation share ($TCS = 45.77$), followed by ACL (31.26) and NAACL (26.63), whereas AAAI (12.89) and IJCAI (8.29) show markedly lower concentration. In terms of top-conference citation coverage (TCC), ACL reaches a pronounced peak (**94.34**), indicating exceptionally broad penetration into top venues, followed by ICLR (91.91) and NAACL (91.54); IJCAI (80.94) again lags behind. A similar hierarchy is observed for top-journal citation share (TJS), led by ICLR (**3.91**), NeurIPS (2.68), and ACL (2.04). Notably, NAACL (1.90) surpasses larger venues such as AAAI (1.71) and EMNLP (1.56), while IJCAI (1.07) remains the weakest. Overall, ICLR and ACL are most deeply embedded in elite citation channels, where broader AI venues such as AAAI and IJCAI exhibit more limited prestige integration.

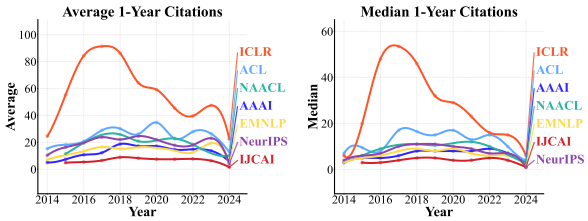

Figure 3: Temporal trends of average and median citations across major Conferences (2014–2024).

QQE Dimension. The QQE index evaluates whether citation growth outpaces publication expansion. All venues exhibit $QQE > 1.0$, indicating that scale growth has not diluted impact density. NAACL (2.34) and EMNLP (2.06) demonstrate the most elastic, high-yield growth, achieving more than proportional citation gains per unit expansion. NeurIPS (1.83), IJCAI (1.77), ICLR (1.61), and AAAI (1.55) show strong to moderate elasticity, while ACL (1.46) records the lowest QQE, reflecting a more linear coupling between scale and impact. Taken together, the results suggest that sustained prestige among top venues is driven by structural integration and efficient growth strategies rather than publication volume alone.

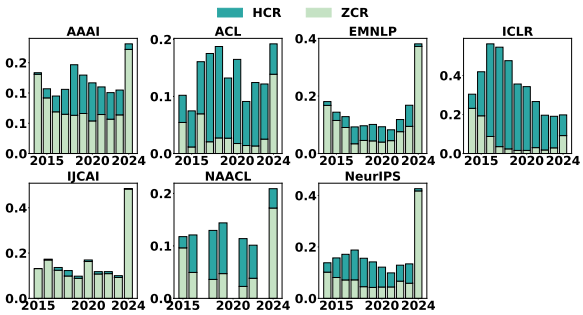

Figure 4: The head and tail of citation impact result.

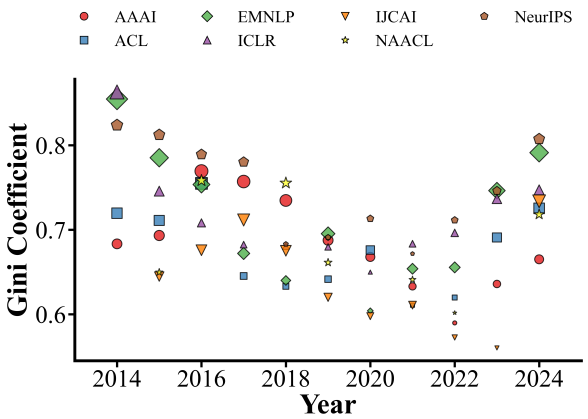

Figure 5: Gini coefficient trends (2014–2024).

4.2 Comprehensive Empirical Decomposition of Conference Trends (RQ1–RQ5)

RQ1. How Have Average and Median Citations Evolved Over Time?

Using a one-year citation window (Fig. 3), we observe clear differences in both impact level and temporal volatility across venues. ICLR consistently achieves the highest average and median citations, peaking during 2016–2018. ACL exhibits the strongest fluctuations, with pronounced peaks in 2017–2020 driven by Transformers and early pre-trained models (e.g., BERT/GPT), a slowdown in 2021, and a renewed surge in 2022–2023 following the rise of LLMs (e.g., ChatGPT and GPT-4), indicating high sensitivity to paradigm-shifting trends. NeurIPS remains comparatively stable with modest peaks around 2019 and 2023, while EMNLP shows sustained stability until a single peak in 2023 and NAACL consistently outperforms EMNLP. AAAI displays flat citation dynamics throughout, and IJCAI persistently records the lowest and most stable average and median citations. Overall, both AC and MC suggest a long-term ordering of ICLR $>$ ACL $>$ NeurIPS $\approx$ NAACL $>$ AAAI $\approx$ EMNLP $>$ IJCAI.

RQ2. How Have the Head and Tail of Citation Impact Shifted?

Using a one-year citation window to characterize impact density (Fig. 4 and Fig. 9), we observe pronounced differences in the evolution of citation heads and tails across venues. ICLR stands out with an exceptionally strong head, with its highly-cited ratio peaking at 51.01% in 2017 while maintaining persistently low zero-cited ratios (e.g., $< 2\%$ during 2018–2020). ACL also exhibits a clear rise in high-impact papers starting in 2017, reaching 15.46%, alongside consistently low tail mass. In contrast, NeurIPS shows a more moderate and stable head, without the sharp shifts observed at ICLR or ACL. EMNLP remains largely stable over time, though its tail expands noticeably in the most recent year. IJCAI consistently records the weakest head (peaking at only 2.40%) and a persistently heavy tail, with zero-cited ratios rising to 16.27% in 2020 and nearly 50% by 2024. Overall, the citation head–tail structure suggests an approximate ordering of ICLR $>$ ACL $>$ NeurIPS $>$ NAACL $>$ EMNLP $\approx$ AAAI $>$ IJCAI.

RQ3. Is Citation Inequality Increasing Over Time?

Across all venues, Gini coefficients (Fig. 5) consistently fall within the 0.6–0.8 range, indicating that moderately high citation inequality is a common feature of conference-based scholarly communication. Differences across venues reflect distinct impact profiles: IJCAI consistently exhibits the lowest Gini values, which, together with its weak citation performance in earlier analyses, suggests uniformly low citation levels. In contrast, NeurIPS frequently ranks among the most unequal venues, reflecting stronger concentration of citations on a small subset of highly influential papers. ICLR, EMNLP, and AAAI typically occupy the upper-middle range, while ACL and NAACL remain comparatively lower and more stable over time. Temporally, most venues become most balanced around 2019–2021 during expansion, followed by renewed concentration. Overall, citation inequality does not increase monotonically.

RQ4. When Does Scale–Impact Decoupling Emerge?

Fig. 6 shows a clear downward trend across all venues over time, indicating that as conferences expand, citation growth becomes increasingly difficult to sustain at a pace faster than publication growth. ICLR and NeurIPS exhibit the most stable

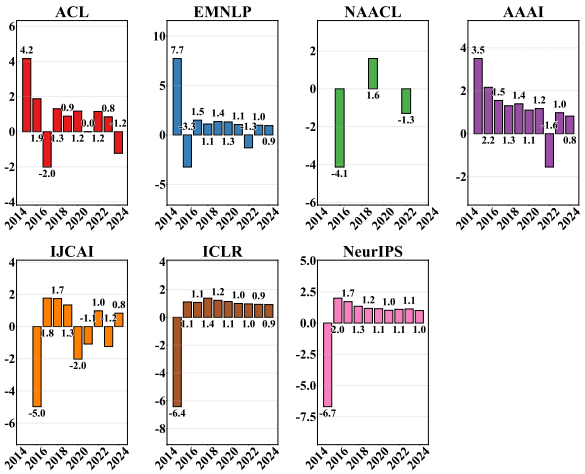

Figure 6: Annual QQE of major NLP/AI conferences (2014–2024).

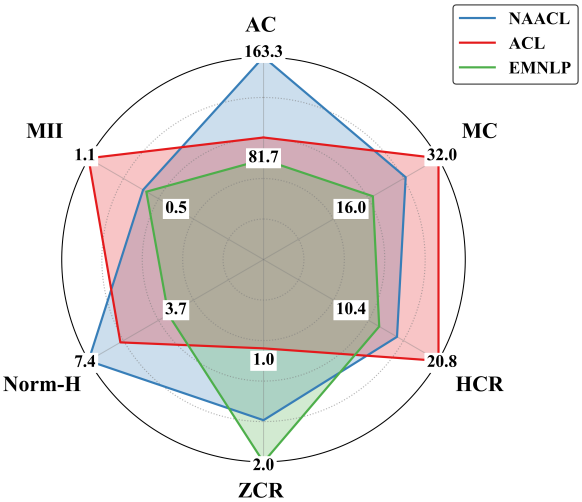

Figure 7: Internal impact ecology of ACL, EMNLP, and NAACL.

elasticity profiles. Their QQE remains close to 1 from 2016, suggesting that citation impact generally scales in line with conference size. In contrast, NLP venues (ACL, EMNLP, NAACL) and IJCAI show much higher volatility, frequently alternating between values above and below 1 and repeatedly exhibiting large negative QQE values (e.g., ACL: -2.0 in 2017 and -1.2 in 2024; EMNLP: -3.2 in 2016; IJCAI: -5.0 in 2016), reflecting sharp mismatches between impact change and scale adjustment in certain years. AAAI maintains relatively high QQE during its early growth phase (2015–2019, mostly above 1.3) but weakens noticeably after 2022, falling below 1 and indicating reduced growth efficiency in later stages.

RQ5. How Has the Internal Ecology of NLP Conferences Evolved?

Category	r	p
Representation Learning & Optimization	-0.915	0.000
Multimodal Learning	-0.818	0.004
Probabilistic Statistics & Inference	-0.842	0.002
Machine Learning (incl. Deep Learning)	-0.939	0.000
AI Foundations & Theory	-0.455	0.187
Computer Vision (CV)	-0.673	0.033
Interpretability, Fairness & Applications	-0.867	0.001
Reinforcement Learning & Decision Making	-0.948	0.000
Natural Language Processing	-0.879	0.001

Table 3: Spearman rank correlations between annual publication counts and the QQE across categories. **r** represents **Spearman r** and **p** represents **Spearman p**.

Fig. 7 compares major NLP venues. NAACL achieves the highest average citations (AC = 163.30), exceeding ACL (98.39) and EMNLP (80.65), indicating strong per-paper influence despite its smaller scale. ACL records the highest median citations (MC = 32) and the largest share of highly cited papers (HCR = 20.75%), reflecting a strong and stable impact baseline. EMNLP exhibits the heaviest tail, with the highest zero-cited ratio (ZCR = 2.03%), whereas ACL maintains the lowest (0.89%). In terms of impact density, NAACL leads in normalized H-index (Norm-H = 7.36), while ACL achieves the highest milestone influence (MII = 1.09), reinforcing its central role in producing landmark NLP work. Overall, ACL combines a strong baseline, a robust head, and a minimal tail; NAACL delivers the highest impact density at smaller scale; and EMNLP shows a more diffuse impact profile following expansion, yielding an internal ranking of ACL > NAACL > EMNLP.

5 Publication Growth and Diminishing Impact: A Field-Level Extension

Our main experiment shows that growth in publication volume is accompanied by a systematic decline in QQE. However, this overall trend may be influenced by changes in the composition of research fields. To address this, we conduct a field-level analysis. Specifically, we first classify all papers into one of nine major research fields (Fig. 8, left) and compute the annual average of the absolute value of QQE for each field (Fig. 8, right). We then examine, within each field, the Spearman rank correlation between annual publication volume and absolute QQE over the period 2015–2024(Tab. 3).

We find that in most research fields, annual publication volume is negatively correlated with the

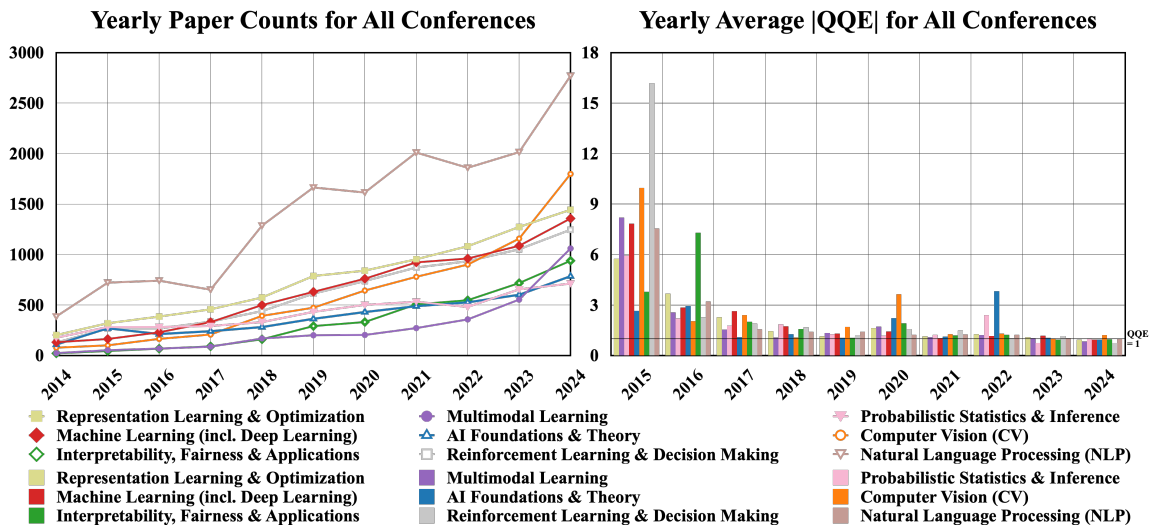

Figure 8: Historical trends in Paper Counts and QQE across different categories.

absolute value of QQE: eight of the nine fields show statistically significant correlations ($p \leq 0.05$). The only exception is the **AI Foundations & Theory** field, where the correlation is weaker and not statistically significant ($p > 0.05$), possibly indicating that paper growth in foundational research does not dilute scholarly impact. Our results demonstrate that the decline in absolute QQE is a robust cross-field phenomenon, suggesting diminishing marginal returns to scholarly impact as publication volumes expand in these ML, AI, and NLP conferences. For more details, please refer to App. H.

6 Discussion

In machine learning venues (ICLR and NeurIPS), continued growth does not undermine impact: average and median citations remain high, the share of highly cited papers is stable, tails are well controlled, and citation inequality does not worsen persistently after expansion. Their QQE trajectories stay close to 1 over many years, indicating that changes in impact largely track changes in scale. These venues sustain high visibility with stable head concentration under growth.

In broader AI venues (AAAI and IJCAI), expansion is not accompanied by comparable impact concentration. Both venues show lower high-impact ratios and heavier tails, with IJCAI in particular exhibiting uniformly low citation levels rather than balanced influence. Their QQE trajectories fluctuate early and converge more quickly toward values below 1 in recent years, indicating that impact growth increasingly lags behind scale adjustment.

Within the NLP domain (ACL, NAACL, and EMNLP), expansion leads to differentiation rather than convergence. ACL consistently maintains higher medians, a larger share of highly cited papers, tighter tails, and the strongest contribution to milestone work, confirming its role as the central NLP venue. NAACL achieves the highest impact density relative to its smaller scale, while EMNLP absorbs expansion at the cost of greater dispersion, with longer tails, rising citation inequality, and QQE converging below 1 in recent years.

7 Conclusion

This work presents a decade-long empirical comparison of seven major NLP and AI conferences (2014–2024) using a four-dimensional bibliometric framework and the proposed Quality–Quantity Elasticity (QQE). We examine how conference scale, citation impact, impact dispersion, and cross-venue influence evolve under sustained expansion. Our results show that ML venues (ICLR and NeurIPS) maintain consistently high and stable influence with tight scale–impact alignment. Within NLP, expansion leads to clear internal differentiation: ACL remains the central venue for high-impact and milestone work, NAACL achieves high impact density under limited scale, and EMNLP exhibits increasing dispersion as it absorbs growth. Broader AI venues (AAAI and IJCAI) show weaker impact concentration and lower scale–impact alignment. This work provides the first systematic longitudinal evidence that post-expansion conference standing is determined not by size, but by how effectively scholarly impact is concentrated and sustained as scale grows.

8 Limitations

Our study has several limitations. First, although we aggregate data from multiple high-quality sources (ACL Anthology, Semantic Scholar, and OpenAlex), citation records may still contain minor inconsistencies due to update delays or incomplete indexing. These issues are unlikely to affect cross-venue comparisons but may influence fine-grained year-to-year values. Second, our analysis focuses exclusively on long papers; short papers, findings papers, and workshop papers are excluded due to inconsistent archival standards across venues. While this choice improves comparability, it may omit certain forms of early-stage or exploratory research. Third, citation-based indicators inherently lag behind real-time research influence and may partially reflect field-specific citation practices or short-term topic popularity rather than long-term scientific contribution. Finally, while our framework captures aggregate and distributional patterns, it does not explicitly model venue-specific policy changes (e.g., review procedures, submission formats, or acceptance strategies), which may also shape conference trajectories.

These limitations suggest several directions for future work, including incorporating acceptance-rate dynamics, policy changes, and field-normalized impact measures to achieve a more fine-grained understanding of conference evolution.

Potential risks. A potential risk is that reliance on citation-based indicators may overemphasize short-term visibility and popular research topics, potentially underrepresenting slower-burning or practice-oriented contributions.

References

Philippe Aghion, Peter Howitt, Maxine Brant-Collett, and Cecilia García-Peñalosa. 1998. *Endogenous growth theory*. MIT press.

William L Croft and Jörg-Rüdiger Sack. 2022. Predicting the citation count and citescore of journals one year in advance. *Journal of Informetrics*, 16(4):101349.

John PA Ioannidis. 2006. Concentration of the most-cited papers in the scientific literature: analysis of journal ecosystems. *PloS one*, 1(1):e5.

Sehrish Iqbal, Saeed-Ul Hassan, Naif Radi Aljohani, Salem Alelyani, Raheel Nawaz, and Lutz Bornmann. 2021. A decade of in-text citation analysis based on natural language processing and machine learning techniques: an overview of empirical studies. *Scientometrics*, 126(8):6551–6599.

Lun-Wei Ku, André FT Martins, and Vivek Srikumar. 2024. Proceedings of the 62nd annual meeting of the association for computational linguistics (volume 1: Long papers). In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*.

Ilia Kuznetsov, Iryna Gurevych, Yusuke Miyao, Yang Gao, and Steffen Eger. 2019. Does my rebuttal matter? insights from a major. In *Proceedings of the 2019 Conference of the North*. Association for Computational Linguistics (ACL).

Lokman I Meho. 2019. Using scopus's citescore for assessing the quality of computer science conferences. *Journal of Informetrics*, 13(1):419–433.

Saif Mohammad. 2020. Examining citations of natural language processing literature. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5199–5209.

Humza Naveed, Asad Ullah Khan, Shi Qiu, Muhammad Saqib, Saeed Anwar, Muhammad Usman, Naveed Akhtar, Nick Barnes, and Ajmal Mian. 2025. A comprehensive overview of large language models. *ACM Transactions on Intelligent Systems and Technology*, 16(5):1–72.

Anna Rogers and Isabelle Augenstein. 2020. What can we do to improve peer review in NLP? In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 1256–1262, Online. Association for Computational Linguistics.

Giuseppe Russo, Manoel Horta Ribeiro, Tim Ruben Davidson, Veniamin Veselovsky, and Robert West. 2025. The ai review lottery: Widespread ai-assisted peer reviews boost paper scores and acceptance rates. *Proceedings of the ACM on Human-Computer Interaction*, 9(7):1–28.

George Vrettas and Mark Sanderson. 2015. Conferences versus journals in computer science. *Journal of the Association for Information Science and Technology*, 66(12):2674–2684.

Ludo Waltman and Nees Jan van Eck. 2012. A new methodology for constructing a publication-level classification system of science. *Journal of the American Society for Information Science and Technology*, 63(12):2378–2392.

Shiyun Wang, Yaxue Ma, Jin Mao, Yun Bai, Zhentao Liang, and Gang Li. 2023. Quantifying scientific breakthroughs by a novel disruption indicator based on knowledge entities. *Journal of the Association for Information Science and Technology*, 74(2):150–167.

Xue Wang, Xuemei Yang, Jian Du, Xuwen Wang, Jiao Li, and Xiaoli Tang. 2021. A deep learning approach for identifying biomedical breakthrough discoveries using context analysis. *Scientometrics*, 126(7):5531–5549.

Qiyao Wei, Samuel Holt, Jing Yang, Markus Wulfmeier, and Mihaela van der Schaar. 2025. The ai imperative: Scaling high-quality peer review in machine learning. *arXiv preprint arXiv:2506.08134*.

A Data Collection and Processing

To ensure the comprehensiveness and accuracy of our longitudinal analysis, we curated a large-scale dataset of publications and their corresponding citation metrics from seven premier AI and NLP conferences. We restrict our analysis to main-track long papers. Findings papers, short papers, and workshop papers are excluded to ensure cross-year and cross-venue comparability. The data collection process was conducted in two primary stages: metadata extraction and citation retrieval.

Temporal Scope: 2014–2024. Our analysis covers 2014–2024. Since ACL submission volumes were essentially stable from 2009–2014, e.g., 634 (2011), 571 (2012), 664 (2013), and 572 (2014), but began accelerating sharply in 2015 (692), 2014 provides a natural pre-expansion baseline. The year 2024 is chosen as the endpoint to ensure complete and comparable proceedings across all venues[8].

In the first stage, we extracted the official metadata for all published papers, specifically focusing on publication years and paper titles. The source repositories were selected based on the official hosting platforms of each venue:

- ACL, EMNLP, and NAACL: Data were retrieved from the ACL Anthology, the comprehensive digital archive for conference proceedings in computational linguistics.
- IJCAI: Data were collected from the IJCAI Proceedings archive.
- AAAI: Metadata were sourced from the AAAI Digital Library.
- NeurIPS: Data were extracted from the NeurIPS Proceedings.
- ICLR: Given its open-reviewing model, data were retrieved from OpenReview.net.

Due to the absence of an official Google Scholar API, we rely on Semantic Scholar and OpenAlex to obtain high-quality citation data. In our pipeline, paper titles were first harvested directly from the official websites of each conference and subsequently matched with external databases using exact string matching—augmented by venue and year constraints—to minimize ambiguity. Semantic Scholar served as the primary data source due to its superior coverage and its provision of time-sliced citation records essential for calculating dynamic metrics like Quality-Quantity Elasticity (QQE). This approach achieved a 99.3% successful match rate, with the remaining 0.7% (approximately 430 entries) supplemented via OpenAlex. Together, these sources provided granular, programmatically retrieved metadata, including annual and total citation counts, necessary for our longitudinal analysis.Due to indexing limitations in the source databases, the number of collected papers is slightly lower than the total number of published papers. However, the missing data accounts for less than 1% of the total volume and has a negligible impact on the overall analysis.

[8] Up to October 2025.

Conference	No. of Papers	Citations
ACL	4,965	488,499
EMNLP	7,106	573,104
NAACL	2,540	414,779
AAAI	12,786	713,125
IJCAI	6,617	245,391
ICLR	7,691	1,771,388
NeurIPS	18,701	2,375,612
Total	**60,406**	**6,581,898**

Table 4: Paper and citation statistics.

The final dataset underwent a cleaning process to resolve potential title mismatches and to ensure that only main conference research papers (excluding workshops and tutorials) were included in the longitudinal study.

B Raw Dataset Structure and Schema

The initial dataset constructed for this study comprises raw records for publications from the seven major conferences spanning 2014 to 2024. Each record represents an individual paper and includes the following fields, designed to support multi-dimensional and longitudinal analysis of academic impact:

- **Core Metadata Fields**:
  - *paperId*: The unique identifier assigned to each publication (sourced from Semantic Scholar).
  - *title*: The full official title of the paper.
  - *year*: The formal year of publication.
  - *ai_category*: The specific sub-field or category classification of the paper within the AI domain.

- *notes*: Supplementary information regarding paper types or special session inclusions.

- **Aggregated Citation Fields**:
  - *citationCount*: The cumulative citation count at the time of data collection.
  - *top_conf_citations*: The total number of citations received from the peer top-tier conferences defined in this study.
  - *top_journal_citations*: The total number of citations received from high-impact top-tier journals.

- **Annual Time-Series Citation Fields**:
  - *citations_2014* to *citations_2024*: A series of annual fields recording the incremental citations received by the paper within each specific calendar year from 2014 to 2024. This granular time-series data serves as the fundamental basis for calculating dynamic metrics such as Quality-Quantity Elasticity (QQE), annual highly-cited ratios, and Gini coefficients.

Table 4 summarizes the total volume of data collected from the various meetings.

C List of Identified Top-Tier Venues

The following lists define the venues used to calculate the top-tier citation metrics . These categories ensure that the cross-disciplinary impact of the research is accurately captured.

Top Conferences

- **Artificial Intelligence & ML**: AAAI, IJCAI, ICML, NeurIPS (including NIPS).
- **Natural Language Processing**: ACL, EMNLP.
- **Computer Vision**: CVPR, ICCV, ECCV.
- **Data Mining & Web**: KDD, SIGIR, WWW.
- **Software Engineering**: ICSE, FSE, ASE.
- **Systems & Networking**: SOSP, OSDI, NSDI, SIGCOMM, INFOCOM, MOBICOM.
- **Security**: CCS, S&P (Oakland), USENIX Security.
- **Databases**: SIGMOD, VLDB, ICDE.
- **Theory**: STOC, FOCS, SODA.
- **Computer Graphics**: SIGGRAPH.
- **Human-Computer Interaction**: CHI.

Top Journals

- **General Science**: Nature, Science.

D Analytical Method

We adopt a combined cross-conference and longitudinal analytical design.

(1) Cross-Conference Comparison. We compute the ten-year averages of thirteen bibliometric indicators for ACL, EMNLP, NAACL, AAAI, IJCAI, ICLR, and NeurIPS. This provides a macro-level snapshot of each venue's overall scale, citation performance, impact distribution, citation composition, and growth efficiency (QQE). The analysis characterizes the long-term static landscape of major NLP/AI conferences and their relative positions.

(2) Longitudinal Evolution Analysis. We then track year-to-year changes using a one-year computation window, focusing on six core dynamic indicators: average and median citations (AC, MC), the proportions of highly cited and zero-cited papers (HCR, ZCR), QQE, milestone influence (MII), and within-NLP comparisons among ACL, EMNLP, and NAACL. This longitudinal component examines the temporal evolution of citation levels, impact distribution, and growth efficiency across the seven venues over the past decade, revealing the shifts in their developmental trajectories.

E Experimental Setup and Hyperparameters

Our study does not involve model training or learning-based optimization; therefore, no conventional hyperparameters are required. All experiments are conducted within a descriptive and longitudinal bibliometric analysis framework. The experimental setup consists of (1) selecting seven major NLP/AI conferences, (2) restricting analysis to main-track long papers, (3) defining a fixed temporal window (2014–2024), and (4) computing citation-based indicators using consistent rules across venues and years.

Key analytical settings, including the one-year citation window, the definition of highly cited papers,

the construction of milestone indicators, and the computation of Quality–Quantity Elasticity (QQE), are fixed a priori and applied uniformly throughout all experiments. No parameter tuning or post-hoc threshold adjustment is performed. No stochastic experiments are involved, so error bars are not applicable. This design ensures that all reported results are fully reproducible and not sensitive to hyperparameter choices.

F Is the Change Driven by Head Papers, Tail Papers, or the Full Distribution?

The per-year violin plots (Fig. 9) utilize a logarithmic scale to compress multi-order citation differences, revealing a clear structural divergence between venues. ML conferences (ICLR, NeurIPS) exhibit elevated centers and stable heads, indicating consistently high visibility and tight alignment with community attention. Conversely, NLP venues (notably EMNLP and NAACL) maintain stable centers but show widening tails after 2022, reflecting increasing heterogeneity and a growing share of minimally cited papers during rapid expansion. ICLR experienced the most dramatic shift, moving from an exceptionally long head during its 2014–2017 breakthrough era to post-2020 contraction, while IJCAI shows recent head compression. Overall, these patterns demonstrate that citation evolution is driven by differentiated movements across the distribution's head, body, and tail rather than uniform shifts.

G Annual QQE Metrics and Milestone-Level Papers

In this appendix, we further address two research questions(RQ6-RQ7) that complement the comprehensive development trends of the surveyed conferences. We present the annual trajectories of the Quality-Quantity Elasticity (QQE) index along with the detailed data used for its calculation. Additionally, we visualize the distribution of milestone-level papers across the various venues. *RQ5. When Does Scale–Impact Decoupling Emerge?*

Annual QQE trajectories (Fig. 10) reveal a strategic divergence: AI venues (ICLR and NeurIPS) maintain stable elastic expansion, while NLP venues and IJCAI exhibit high volatility. ICLR and NeurIPS consistently achieve QQE values mostly at or above 1.0 (ICLR: 0.92–1.38; NeurIPS: 1.00–1.99), indicating sustainable growth where impact effectively outpaces scale. In contrast, NLP venues (ACL, EMNLP, NAACL) and IJCAI show frequent oscillations and negative QQE phases (e.g., IJCAI: -4.98 in 2016; ACL: -1.23 in 2024), reflecting sensitivity to shifting acceptance controls and short-term disruptions. Overall, while ML conferences demonstrate a steady growth path, the development trajectories of NLP and broader AI venues are characterized by irregular cycles of rapid expansion and scale contraction.

RQ7. Which Venues Produce the Most Milestone-Level Papers in the LLM Era?

Fig. 11 shows substantial variation across conferences in their production of milestone-level research. Although NeurIPS has the highest absolute number of $\geq$ 1000-citation papers (283), it also has the largest overall volume, resulting in a proportion of 1.51%. In contrast, ICLR, despite having far fewer total papers, achieves the highest proportion of milestone papers at 2.94%. Within the NLP community, ACL leads with a proportion of 1.09%, while NAACL (0.75%) and EMNLP (0.73%) exhibit similar levels, indicating that high-impact work remains a significant component of these language-focused venues. AAAI and IJCAI show much lower proportions (0.48% and 0.17%, respectively), suggesting that they produce fewer ultra-high-impact papers relative to their total volume in the LLM era.

H Field-Level Decomposition of Publication Growth and QQE Dynamics

Our main experiment considers all papers in aggregate and reveals a systematic decline in the absolute value of QQE as publication volume expands. However, these aggregate trends may potentially be influenced by changes in the mix of fields, as different research fields exhibit different growth rates and citation dynamics.

To separate these effects, we conduct a field-level analysis by first categorizing papers into major research areas using DeepSeek-V3, with the annual publication counts and the absolute values of QQE shown in Tab. 5. We then examined the relationship between annual publication counts and the absolute values of QQE over the period 2015–2024 within each field. Specifically, for each field, we compute the Spearman rank correlation between annual publication volume and the absolute values of QQE, allowing us to assess whether the observed negative association reflects a field-specific

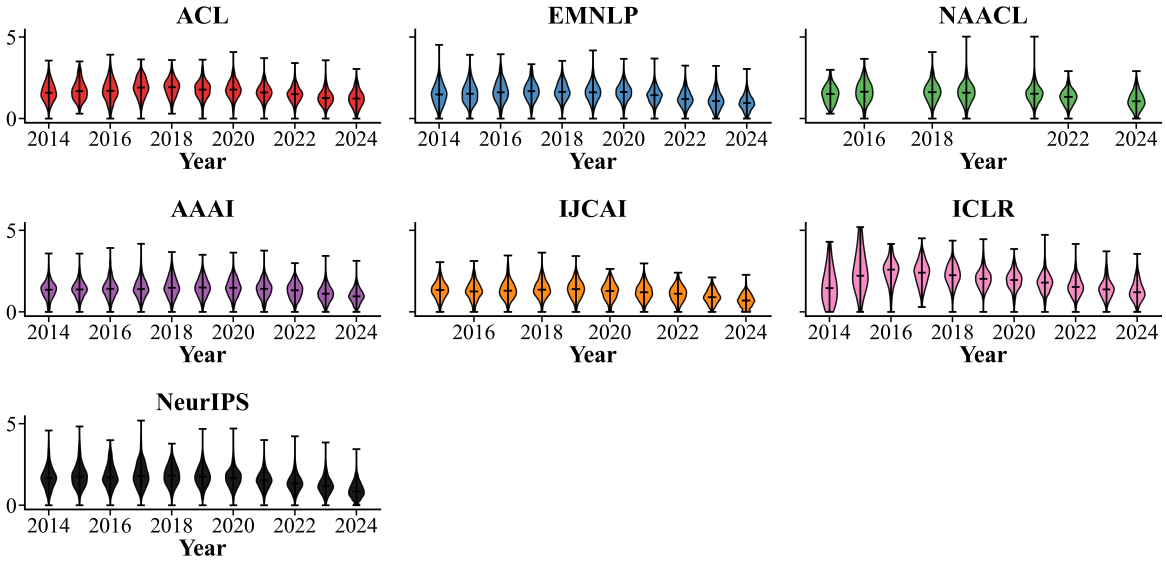

Figure 9: Evolution of citation distribution in major NLP/AI conferences.

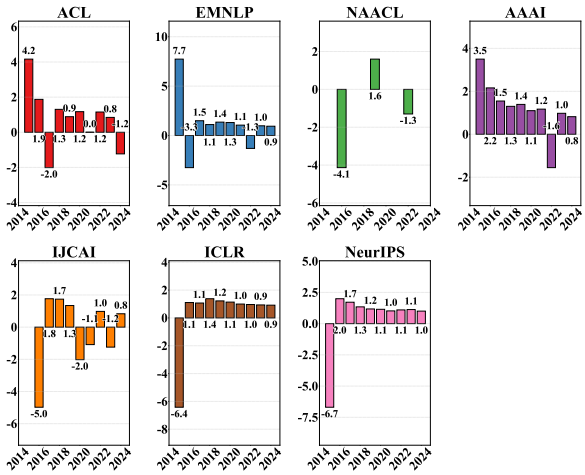

Figure 10: Annual QQE of major NLP/AI conferences (2014–2024).

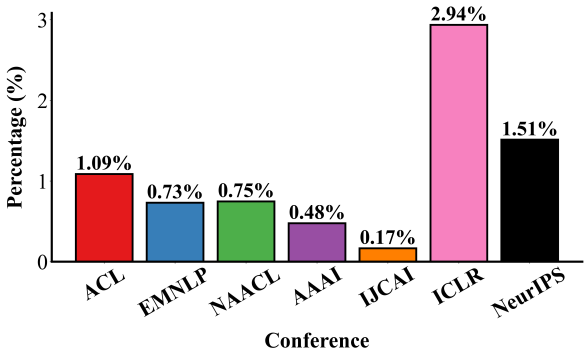

Figure 11: Percentage of Milestone Papers.

phenomenon or a more general structural pattern.

As shown in Tab. 3, we observe predominantly strong and statistically significant negative correlations across most research fields. In particular, Representation Learning & Optimization, Machine Learning (incl. Deep Learning), Interpretability, Fairness & Applications, Reinforcement Learning & Decision Making, and Natural Language Processing exhibit Spearman correlation coefficients below -0.85 ($p \leq 0.01$), indicating that sustained increases in publication volume are systematically associated with declining absolute QQE. These results suggest that as these fields expand, marginal increases in publication output yield diminishing returns in scholarly impact.

An exception is AI Foundations & Theory area, where the correlation is weaker and statistically insignificant ($p > 0.05$), possibly reflecting the fact that publication growth in foundational research does not necessarily dilute research impact. Overall, the persistence of negative correlations within individual fields demonstrates that the decline in the absolute values of QQE represents a robust, cross-fields phenomenon in these ML, AI, and NLP conferences. While spillover effects from accumulated knowledge remain present, their marginal contribution to scholarly impact appears to weaken as research fields scale up.

I Statement on AI Assistance

ChatGPT was used for grammar checking and language polishing only. Additionally, in the supplementary analysis of Section 5, DeepSeek-V3 (a 671B parameter Mixture-of-Experts model) was

employed to assist in the automated classification of paper topics. All AI-generated classifications were manually audited and verified by the authors. All core analyses, results, and interpretations were conducted by the authors, who maintain full responsibility for the contents of this paper.

Paper Count	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**	**2024**
Representation Learning & Optimization	202	319	385	457	574	787	840	953	1082	1273	1443
Multimodal Learning	23	52	68	87	168	200	203	271	357	553	1060
Probabilistic Statistics & Inference	176	281	282	292	330	432	501	532	480	655	712
Machine Learning (incl. Deep Learning)	131	163	229	330	499	630	760	922	960	1086	1357
AI Foundations & Theory	108	269	212	240	281	363	430	488	526	601	785
Computer Vision (CV)	77	100	164	207	392	473	643	779	900	1157	1798
Interpretability, Fairness & Applications	21	44	67	90	160	291	331	506	547	717	938
Reinforcement Learning & Decision Making	128	267	267	335	439	612	737	871	933	1052	1245
Natural Language Processing (NLP)	386	721	741	651	1286	1664	1614	2010	1859	2014	2770
Avg. of the Absolute Values of QQE	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**2022**	**2023**	**2024**
Representation Learning & Optimization	/	5.74	3.68	2.28	1.44	1.15	1.63	1.14	1.26	1.09	0.93
Multimodal Learning	/	8.20	2.57	1.53	1.07	1.33	1.72	1.08	1.22	1.04	0.83
Probabilistic Statistics & Inference	/	5.95	2.20	1.76	1.86	1.26	1.18	1.22	2.40	0.74	1.02
Machine Learning (incl. Deep Learning)	/	7.84	2.84	2.63	1.72	1.30	1.42	1.03	1.16	1.18	0.93
AI Foundations & Theory	/	2.64	2.94	1.09	1.27	1.05	2.21	1.12	3.83	1.04	0.94
Computer Vision (CV)	/	9.96	2.03	2.39	1.06	1.69	3.64	1.27	1.30	1.01	1.21
Interpretability, Fairness & Applications	/	3.78	7.30	2.00	1.58	1.05	1.90	1.19	1.24	0.93	0.96
Reinforcement Learning & Decision Making	/	16.19	2.27	1.90	1.69	1.17	1.56	1.47	1.02	1.16	0.74
Natural Language Processing (NLP)	/	7.55	3.21	1.55	1.40	1.40	1.22	1.25	1.23	1.04	0.98

Table 5: Historical trends in Paper Counts and the absolute values of QQE across different categories.

Conference	Year	N_t	C_t	P_t	g_t	QQE	$\ln(P_t)$
ACL	2014	147	495	-	-	-	-
ACL	2015	174	2440	1.184	4.929	4.164	0.169
ACL	2016	230	6047	1.322	2.478	1.875	0.279
ACL	2017	194	10263	0.843	1.697	-2.012	-0.170
ACL	2018	256	17678	1.320	1.722	1.305	0.277
ACL	2019	446	27312	1.742	1.545	0.887	0.555
ACL	2020	571	40957	1.278	1.500	1.173	0.245
ACL	2021	577	51227	1.000	1.251	0.001	0.000
ACL	2022	603	62525	1.058	1.221	1.154	0.056
ACL	2023	911	80032	1.511	1.280	0.847	0.413
ACL	2024	864	93298	0.948	1.166	-1.229	-0.053
EMNLP	2014	227	187	-	-	-	-
EMNLP	2015	313	1994	1.379	10.663	7.733	0.321
EMNLP	2016	265	5490	0.847	2.753	-3.252	-0.166
EMNLP	2017	324	10041	1.223	1.829	1.496	0.201
EMNLP	2018	549	18849	1.694	1.877	1.108	0.527
EMNLP	2019	682	31715	1.242	1.683	1.354	0.217
EMNLP	2020	752	45938	1.103	1.448	1.314	0.098
EMNLP	2021	848	54477	1.128	1.186	1.052	0.120
EMNLP	2022	829	69510	0.978	1.276	-1.305	-0.023
EMNLP	2023	1048	86360	1.264	1.242	0.983	0.234
EMNLP	2024	1269	98426	1.211	1.140	0.941	0.191
NAACL	2014	0	36	-	-	-	-
NAACL	2015	187	642	-	17.833	-	-
NAACL	2016	182	2583	0.973	4.023	-4.134	-0.027
NAACL	2017	0	4452	0.000	1.724	-	-
NAACL	2018	332	7787	-	1.749	-	-
NAACL	2019	424	15893	1.277	2.041	1.598	0.245
NAACL	2020	0	21370	0.000	1.345	-	-
NAACL	2021	483	23456	-	1.098	-	-
NAACL	2022	444	28024	0.919	1.195	-1.300	-0.084
NAACL	2023	0	32186	0.000	1.149	-	-
NAACL	2024	488	32419	-	1.007	-	-
AAAI	2014	397	664	-	-	-	-
AAAI	2015	533	3120	1.343	4.699	3.500	0.295
AAAI	2016	537	6788	1.008	2.176	2.159	0.007
AAAI	2017	632	12367	1.177	1.822	1.548	0.163
AAAI	2018	934	23770	1.478	1.922	1.301	0.391
AAAI	2019	1119	39603	1.198	1.666	1.391	0.181
AAAI	2020	1559	60618	1.393	1.531	1.099	0.332
AAAI	2021	1645	74868	1.055	1.235	1.171	0.054
AAAI	2022	1322	93505	0.804	1.249	-1.554	-0.219
AAAI	2023	1707	117888	1.291	1.261	0.976	0.256
AAAI	2024	2401	135756	1.407	1.152	0.819	0.341
IJCAI	2014	0	93	-	-	-	-
IJCAI	2015	571	624	-	6.710	-	-
IJCAI	2016	549	2987	0.961	4.787	-4.979	-0.039
IJCAI	2017	660	6345	1.202	2.124	1.767	0.184
IJCAI	2018	709	11857	1.074	1.869	1.740	0.072
IJCAI	2019	847	19031	1.195	1.605	1.344	0.178
IJCAI	2020	590	26804	0.697	1.408	-2.022	-0.362
IJCAI	2021	585	28839	0.992	1.076	-1.085	-0.009
IJCAI	2022	678	32726	1.159	1.135	0.979	0.148
IJCAI	2023	639	38303	0.942	1.170	-1.242	-0.059
IJCAI	2024	789	39351	1.235	1.027	0.832	0.211
ICLR	2014	69	710	-	-	-	-
ICLR	2015	31	2048	0.449	2.885	-6.420	-0.800
ICLR	2016	80	5848	2.581	2.855	1.106	0.948
ICLR	2017	198	15512	2.475	2.653	1.072	0.906
ICLR	2018	336	36339	1.697	2.343	1.380	0.529
ICLR	2019	502	66467	1.494	1.829	1.224	0.401
ICLR	2020	685	103262	1.365	1.554	1.139	0.311
ICLR	2021	859	129222	1.254	1.251	0.998	0.226
ICLR	2022	1094	158795	1.274	1.229	0.965	0.242
ICLR	2023	1576	213685	1.441	1.346	0.934	0.365
ICLR	2024	2261	281333	1.435	1.317	0.918	0.361
NeurIPS	2014	412	758	-	-	-	-

Continued on next page

Conference	Year	N_t	C_t	P_t	g_t	QQE	$\ln(P_t)$
NeurIPS	2015	407	5012	0.988	6.612	-6.693	-0.012
NeurIPS	2016	572	14055	1.405	2.804	1.995	0.340
NeurIPS	2017	681	28705	1.191	2.042	1.715	0.174
NeurIPS	2018	1013	57378	1.488	1.999	1.344	0.397
NeurIPS	2019	1432	95382	1.414	1.662	1.176	0.346
NeurIPS	2020	1903	145529	1.329	1.526	1.148	0.284
NeurIPS	2021	2342	182743	1.231	1.256	1.020	0.208
NeurIPS	2022	2674	229410	1.142	1.255	1.100	0.133
NeurIPS	2023	3229	312565	1.208	1.362	1.128	0.189
NeurIPS	2024	4036	392189	1.250	1.255	1.004	0.223

Table 7: Conference growth rates and QQE