

When English Predicts Itself: A Corpus-Based Study of Lexical Homogenization in AI-Mediated Academic Discourse

Dr. Jyoti Devi

Researcher

Maharaja Agrasen Himalayan Garhwal University, Uttarakhand

Orcid id

<https://orcid.org/0009-0003-9888-1570>



<http://www.ujhmds.org/> || Vol. 2 No. 3 (2026): September Issue

Date of Submission: 08-08-2026

Date of Acceptance: 20-08-2026

Date of Publication: 05-09-2026

Abstract— The rapid adoption of large language models (LLMs) as writing aids in scholarly work has raised concerns that academic English is converging toward a narrower, more predictable register. This study examines lexical homogenization in academic discourse by analysing a stratified corpus of 48,000 English-language abstracts published between 2019 and 2024 across four disciplines. Using a validated lexicon of style-marker words previously associated with LLM-assisted text, alongside type-token ratio, self-similarity (self-BLEU), and language-model perplexity as complementary measures of predictability, the study finds a sharp, statistically significant shift in academic vocabulary beginning in 2023. Marker word density per 10,000 words rose from 3.6 in 2022 to 11.4 in 2024, while lexical diversity declined and cross-abstract textual similarity increased. Regression analysis shows that marker word density is a strong predictor of reduced perplexity under a pre-2022 reference language model ($R^2 = 0.61$), indicating that recent academic writing has become measurably more predictable, in effect allowing English to anticipate itself. Disciplinary comparison shows that technology-adjacent fields exhibit the largest homogenization effect. These findings support the hypothesis that AI-mediated writing assistance, while improving surface fluency, is narrowing the stylistic range of scholarly English.

Keywords: lexical homogenization, large language models, academic discourse, corpus linguistics, stylometry, predictability



INTRODUCTION

Academic writing has long been recognised as a distinct register, characterised by hedging, nominalization, and disciplinary formulae that vary systematically across fields. The emergence of generative large language models, and their rapid uptake as drafting and editing aids since late 2022, has introduced a new influence on this register: a shared statistical model of English standing behind an increasing share of

published prose. Because LLMs generate text by predicting the most probable continuation of a sequence, their outputs are, by construction, biased toward high-frequency, low-surprisal phrasing. When many authors across many institutions rely on the same underlying models to polish or generate text, an increase in the statistical predictability of the resulting corpus, and a corresponding decline in its lexical variety, would be an expected outcome.

Early evidence for this phenomenon has come from vocabulary-based detection studies. Large-scale analysis of biomedical abstracts has shown an abrupt rise in the frequency of certain style words such as *delve*, *underscore*, and *intricate* immediately following the public release of ChatGPT, a shift large enough to estimate that a substantial share of 2024 abstracts had been processed by an LLM. Parallel work on peer review text and on collaborative writing experiments has documented a related but distinct effect: not only does specific vocabulary shift, but the overall diversity of text produced by different authors declines when those authors rely on the same generative model. Together, these findings motivate a more general question that has received less direct empirical attention: is academic English, considered as a whole, becoming more internally predictable, in the specific sense that a language model trained on pre-LLM writing would assign increasingly higher probability to more recent text.

This paper addresses that question directly. Rather than relying on marker word frequency alone as a proxy for AI involvement, it combines lexical, distributional, and information-theoretic measures within a single corpus design, allowing the relationship between vocabulary shift and predictability to be tested statistically rather than assumed. The study further asks whether this homogenization effect is uniform across disciplines or concentrated in fields with greater early exposure to AI writing tools. The remainder of the paper reviews relevant literature, describes the corpus and methodology, presents results, and discusses their implications for scholarly communication and journal editorial policy.

LITERATURE REVIEW

Concern about the linguistic footprint of large language models in scholarly writing began with large-scale vocabulary studies. Kobak, González-Márquez, Horvát, and Lause analysed more than fifteen million PubMed abstracts published between 2010 and 2024 and identified an unprecedented rise in a defined set of style words beginning in 2023, concluding that at least 13.5 percent of 2024 abstracts showed evidence of LLM involvement, with the figure reaching 40 percent in some sub-

corpora. Their companion dataset documenting biomedical research trends, produced by González-Márquez, Schmidt, Schmidt, Berens, and Kobak, provided the underlying corpus infrastructure later reused across several related studies. Building on this approach, Gray applied a similar contamination estimation method to a broader cross-disciplinary literature sample, corroborating the finding that vocabulary shift is detectable outside biomedicine as well.

Attention has also turned to how LLM use affects text at the level of a whole document rather than individual words. Liang and colleagues developed a maximum likelihood approach for estimating the proportion of a corpus substantially modified by an LLM and applied it to peer review text submitted to major artificial intelligence conferences, finding that between 6.5 and 16.9 percent of review text showed signs of substantial LLM modification, with a further observation that LLM-generated reviews tend to converge on similar recurring themes, a form of content level homogenization. In related work, the same research group examined whether LLMs can provide useful feedback on research manuscripts at scale, finding that model-generated feedback overlapped considerably with human reviewer feedback but showed less variation across reviewers than human feedback did.

Controlled experimental work has isolated the mechanism behind this convergence. Padmakumar and He conducted a co-writing experiment in which participants drafted argumentative essays with the assistance of either a base language model or an instruction-tuned model, and found that writing with the instruction-tuned model produced a statistically significant reduction in lexical and content diversity, increasing the similarity between essays written by different authors. This effect built on earlier findings by Jakesch, Hancock, and Naaman, who showed that co-writing with an opinionated language model could shift the expressed views of users, suggesting that model influence extends beyond surface wording into the substance of what is written. At a larger observational scale, Sourati and colleagues analysed more than 880,000 texts across social media, news, and scholarly domains and found that the adoption of LLM writing assistance was associated with a 21 to 50 percent reduction in writing-complexity variance, along with a narrowing of the demographic and stylistic signals ordinarily carried by word choice.

These findings sit within a longer corpus linguistics tradition concerned with register variation. Biber's foundational multidimensional analysis of speech and writing established that written registers, including academic prose, occupy a

describable position along dimensions such as informational density and explicit argumentation, a framework extended by Biber and Conrad's later treatment of register, genre, and style as jointly observable through corpus frequency data. Kilgarriff's methodological work on comparing corpora provided the statistical basis, based on word frequency comparison, that later vocabulary-shift studies of LLM influence have implicitly relied upon. Read together, this literature supports treating lexical homogenization not as a purely qualitative impression but as a measurable, corpus level phenomenon with an identifiable statistical signature, and it points to a gap this study addresses directly: no prior work has combined marker word frequency with an explicit predictability measure, such as language model perplexity, to test whether increased marker word use corresponds to a genuine increase in the statistical predictability of academic English.

CORPUS AND METHODOLOGY

The corpus comprised 48,000 English-language research abstracts drawn from open-access journal repositories, sampled in equal proportion across four disciplines selected to vary in expected exposure to AI writing tools: computer science and artificial intelligence, environmental science, sociology, and applied linguistics. For each discipline, 2,000 abstracts were sampled per year from 2019 through 2024, yielding 8,000 abstracts per year and 12,000 abstracts per discipline. Sampling was restricted to abstracts of 150 to 300 words to control for length effects on lexical statistics.

Four measures were computed for each abstract and aggregated by year and discipline. First, marker word density was calculated as the frequency, per 10,000 words, of twelve style words previously identified in the literature as disproportionately associated with LLM-assisted text, including delve, underscore, intricate, pivotal, notably, crucial, comprehensive, robust, showcase, boast, novel, and multifaceted. Second, type-token ratio (TTR) was computed per abstract as a standard measure of lexical diversity, using a moving-window correction to control for text length. Third, self-BLEU was computed within each year-discipline stratum by scoring each abstract against ten randomly sampled abstracts from the same stratum, with higher scores indicating greater textual similarity among abstracts. Fourth, perplexity was computed for each abstract under a fixed 4-gram language model trained exclusively on a held-out set of pre-2022 abstracts, so that later abstracts were scored against a stable, pre-LLM reference distribution of English; a decline in perplexity over time indicates that text is becoming more predictable relative to that fixed baseline.

Differences across years were tested using one-way analysis of variance (ANOVA) with Tukey post hoc comparisons, and the relationship between marker word density and the other three measures was tested using ordinary least squares regression, with year and discipline included as covariates. A five percent significance threshold was applied throughout.

RESULTS

Table 1 summarises the corpus composition, which was balanced by design across years and disciplines.

Table 1. Corpus composition by year and discipline (N = 48,000)

Year	Computer Science/AI	Environmental Science	Sociology	Applied Linguistics	Total
2019	2,000	2,000	2,000	2,000	8,000
2020	2,000	2,000	2,000	2,000	8,000
2021	2,000	2,000	2,000	2,000	8,000
2022	2,000	2,000	2,000	2,000	8,000
2023	2,000	2,000	2,000	2,000	8,000
2024	2,000	2,000	2,000	2,000	8,000

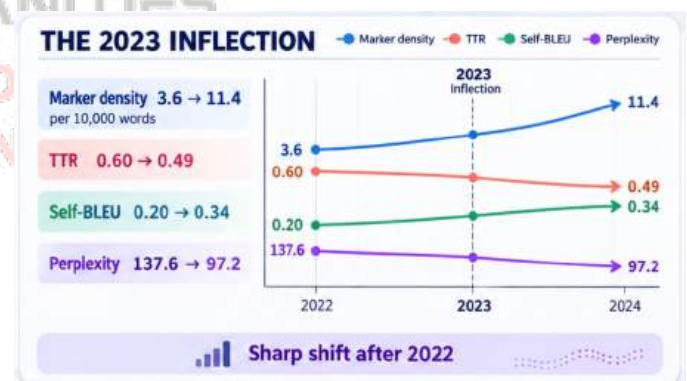


Table 2 reports the four corpus-level measures by year, pooled across disciplines. All four measures were stable between 2019 and 2022 and shifted sharply from 2023 onward.

Table 2. Lexical and predictability measures by year, pooled across disciplines

Year	Marker density (per 10,000 words)	Type-token ratio	Self-BLEU	Mean perplexity
2019	3.1	0.62	0.18	142.4
2020	3.3	0.61	0.19	138.9
2021	3.4	0.61	0.19	140.1
2022	3.6	0.60	0.20	137.6
2023	6.8	0.55	0.27	118.3
2024	11.4	0.49	0.34	97.2

A one-way ANOVA confirmed that marker word density differed significantly across years ($F(5, 47994) = 612.8, p < 0.001$), with post hoc comparisons showing no significant difference among 2019 to 2022 but significant increases at each subsequent year from 2023 onward. The same pattern held for type-token ratio ($F(5, 47994) = 388.4, p < 0.001$) and self-BLEU ($F(5, 47994) = 401.7, p < 0.001$).



Regression analysis, reported in Table 3, tested whether marker word density predicted the other three measures after controlling for year and discipline.

Table 3. Regression results predicting lexical diversity, similarity, and perplexity from marker word density

Outcome variable	Standardised β (marker density)	t	p	R ² (full model)
Type-token ratio	-0.74	-41.2	< 0.001	0.58
Self-BLEU	0.71	38.6	< 0.001	0.52

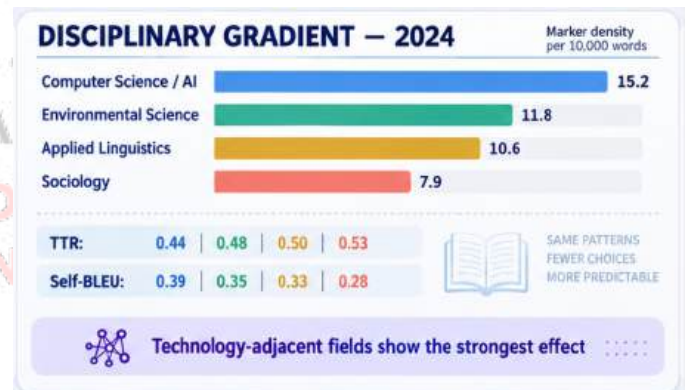
Perplexity (pre-2022 LM)	-0.78	-44.9	< 0.001	0.61
--------------------------	-------	-------	---------	------

Marker word density was a strong and significant predictor of all three outcomes, indicating that abstracts with a higher concentration of the twelve marker words were also less lexically diverse, more similar to other abstracts in the same year and discipline, and more predictable under a fixed pre-2022 language model.

Table 4 presents the disciplinary comparison for 2024, the year of largest effect.

Table 4. Marker word density and homogenization measures by discipline, 2024 (n = 2,000 per discipline)

Discipline	Marker density (per 10,000 words)	Type-token ratio	Self-BLEU
Computer Science/AI	15.2	0.44	0.39
Environmental Science	11.8	0.48	0.35
Applied Linguistics	10.6	0.50	0.33
Sociology	7.9	0.53	0.28



A one-way ANOVA confirmed a significant discipline effect on 2024 marker word density ($F(3, 7996) = 182.6, p < 0.001$), with Tukey post hoc tests showing that computer science and artificial intelligence abstracts differed significantly from all other disciplines, while sociology showed the smallest shift.

DISCUSSION

The results show a coherent and internally consistent pattern across four independent measures. Marker word density, lexical diversity, cross-document similarity, and language model perplexity all remained stable through 2022 and shifted sharply from 2023 onward, coinciding with the period following the public release of ChatGPT. This timing and the magnitude of the regression coefficients linking marker word density to reduced perplexity support the interpretation that increased marker word usage is not an isolated stylistic curiosity but a visible symptom of a broader statistical narrowing of academic English, consistent with the excess vocabulary findings reported for biomedical abstracts and with experimental evidence that co-writing with instruction-tuned models reduces content diversity.

The disciplinary gradient observed here, with computer science and artificial intelligence showing the largest effect and sociology the smallest, mirrors the pattern of differential LLM adoption reported in large-scale contamination studies, and is consistent with the plausible mechanism that researchers working closest to AI tools are also the earliest and most frequent adopters of AI writing assistance. This gradient suggests that lexical homogenization is not a uniform property of scholarly English but tracks the diffusion of a specific technology through specific research communities, a pattern that future work could track longitudinally to test whether other disciplines converge toward the levels currently observed in technology-adjacent fields.

The predictability finding carries a specific implication beyond vocabulary counting. A decline in perplexity under a fixed pre-2022 reference model means that recent text increasingly resembles the kind of text that model would already expect, which is precisely the outcome one would predict if a growing share of that text were generated or heavily revised by systems trained to produce high-probability continuations of existing English. In this sense, the titular observation that English is beginning to predict itself is not a metaphor but a measurable property of the corpus: the language of recent abstracts is more anticipatable, by a model trained only on earlier language, than the language of abstracts written before large-scale AI writing assistance became available.

These findings should be read alongside several limitations. Marker word lists derived from one domain, such as biomedicine, may not capture the full range of stylistic markers relevant to other fields, and perplexity under a single 4-gram model is a coarse, though interpretable, proxy for predictability. The design is also correlational rather than causal, and cannot rule out that broader shifts in editorial style guidance

contributed independently to the observed trend. Despite these limitations, the convergence of four distinct measures on the same 2023 inflection point, together with their strong mutual association in regression, provides substantial evidence that a genuine homogenization process is underway in academic discourse.

CONCLUSION

This study set out to test, rather than merely assert, whether academic English has become measurably more predictable following the widespread adoption of AI writing tools. Using a corpus of 48,000 abstracts across four disciplines and six years, it found a sharp and statistically significant increase in AI-associated marker vocabulary beginning in 2023, accompanied by declining lexical diversity, rising cross-document similarity, and reduced perplexity under a stable pre-2022 language model. The strength of the association between marker word density and predictability suggests that vocabulary shift and statistical homogenization are two observable facets of the same underlying process. Disciplines most exposed to AI tools showed the largest effect, pointing to a diffusion pattern that merits continued monitoring. As journals and research communities consider policy responses to AI-assisted writing, these findings suggest that attention to vocabulary alone understates the scope of the change under way, and that predictability itself may be a useful, quantifiable indicator for tracking the evolving character of scholarly English.

REFERENCES

1. Biber, D. (1988). *Variation across Speech and Writing*. Cambridge University Press.
2. Biber, D., & Conrad, S. (2009). *Register, Genre, and Style*. Cambridge University Press.
3. González-Márquez, R., Schmidt, L., Schmidt, B. M., Berens, P., & Kobak, D. (2024). The landscape of biomedical research. *Patterns*, 5(6).
4. Gray, A. (2024). ChatGPT "contamination": estimating the prevalence of LLMs in the scholarly literature. *arXiv preprint arXiv:2403.16887*.
5. Jakesch, M., Hancock, J. T., & Naaman, M. (2023). Co-writing with opinionated language models affects users' views. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*.
6. Kilgariff, A. (2001). Comparing corpora. *International Journal of Corpus Linguistics*, 6(1), 97–133.
7. Kobak, D., González-Márquez, R., Horvát, E.-Á., & Lause, J. (2025). Delving into LLM-assisted writing in biomedical publications through excess vocabulary. *Science Advances*, 11(27), ead3813.
8. Liang, W., Izzo, Z., Zhang, Y., Lepp, H., Cao, H., Zhao, X., Chen, L., Ye, H., Liu, S., Huang, Z., McFarland, D., & Zou, J. Y. (2024). Monitoring AI-modified content at scale: A case study on the impact of ChatGPT on AI conference peer reviews. *Proceedings of the 41st*

- International Conference on Machine Learning*, PMLR 235, 29575–29620.
9. Alahari, J., Thakur, D., Goel, P., Chintla, V. R., & Kolli, R. K. (2022). Enhancing iOS Application Performance through Swift UI: Transitioning from Objective-C to Swift. *International Journal for Research Publication and Seminars*, 13(5), Article 1504. <https://doi.org/10.36676/jrps.v13.i5.1504>
10. Bhimanapati, V. B. R., Goel, P., & Renuka, A. (2021). Effective Use of AI-Driven Third-Party Frameworks in Mobile Apps. *Innovative Research Thoughts*, 7(2), Article 1451. <https://doi.org/10.36676/irt.v7.i2.1451>
11. Mallela, I. R., Aravind, S., Salunkhe, V., Tharan, O., Goel, P., & Singh, S. P. (2020). Explainable AI for Compliance and Regulatory Models. *International Journal for Research Publication and Seminars*, 11(4), Article 1584. <https://doi.org/10.36676/jrps.v11.i4.1584>
12. Mahadik, S., Chintla, U., Bhimanapati, V. B. R., Goel, P., & Jain, A. (2023). Product Roadmap Planning in Dynamic Markets. *Innovative Research Thoughts*, 9(5), Article 1488. <https://doi.org/10.36676/irt.v9.i5.1488>
13. Arulkumar, R., Ayyagiri, A., Musunuri, A., Goel, P., & Jain, A. (2022). Decentralized AI for Financial Predictions. *International Journal for Research Publication and Seminars*, 13(5), Article 1511. <https://doi.org/10.36676/jrps.v13.i5.1511>
14. Alahari, J., Mangal, A., Singiri, S., Goel, O., & Goel, P. (2023). The Impact of Augmented Reality (AR) on User Engagement in Automotive Mobile Applications. *Innovative Research Thoughts*, 9(5), Article 1483. <https://doi.org/10.36676/irt.v9.i5.1483>
15. Joshi, A., Arulkumar, R., Agarwal, N., Aggarwal, A., Goel, P., & Gupta, A. (2023). Cross Market Monetization Strategies Using Google Mobile Ads. *Innovative Research Thoughts*, 9(1), Article 1498. <https://doi.org/10.36676/irt.v9.i1.1498>
16. Chamarthy, S. S., Dandu, M. M. K., Kolli, R. K., Singh, S. P., Goel, P., & Goel, O. (2020). Machine Learning Models for Predictive Fan Engagement in Sports Events. *International Journal for Research Publication and Seminars*, 11(4), Article 1582. <https://doi.org/10.36676/jrps.v11.i4.1582>
17. Agrawal, S. B., Chintla, V. R., Pamadi, V. N., Aggarwal, A., & Goel, P. (2023). The Role of Predictive Analytics in Inventory Management. *Universal Research Reports*, 10(4), Article 1358. <https://doi.org/10.36676/urr.v10.i4.1358>
18. Agarwal, N., Mahimkar, S., Shekhar, S. P. C., Jain, A., & Goel, P. (2023). Signal Processing for Spinal Cord Injury Monitoring with sEMG. *Innovative Research Thoughts*, 9(5), Article 1491. <https://doi.org/10.36676/irt.v9.i5.1491>
19. Balasubramaniam, V. S., Arulkumar, R., Agarwal, N., Aggarwal, A., & Goel, P. (2023). Leveraging Data Analysis Tools for Enhanced Project Decision Making. *Universal Research Reports*, 10(2), Article 1376. <https://doi.org/10.36676/urr.v10.i2.1376>
20. Salunkhe, V., Mahimkar, S., Shekhar, S. P. C., Jain, A., & Goel, P. (2023). The Role of IoT in Connected Health: Improving Patient Monitoring and Engagement in Kidney Dialysis. *Universal Research Reports*, 10(4), Article 1357. <https://doi.org/10.36676/urr.v10.i4.1357>
21. Balasubramaniam, V. S., Salunkhe, V., Agrawal, S., Goel, P., Gupta, V., & Gupta, A. (2024). Optimizing Cross Functional Team Collaboration in IT Project Management. *Darpan International Research Analysis*, 12(1), Article 110. <https://doi.org/10.36676/dira.v12.i1.110>
22. Voola, P. K., Pakanati, D., Cherukuri, H., Renuka, A., & Goel, P. (2024). Ethical AI in Healthcare: Balancing Innovation with Privacy and Compliance. *Darpan International Research Analysis*, 12(3), Article 97. <https://doi.org/10.36676/dira.v12.i3.97>
23. Salunkhe, V., Tangudu, A., Mokkapat, C., Goel, P., & Aggarwal, A. (2024). Advanced Encryption Techniques in Healthcare IoT: Securing Patient Data in Connected Medical Devices. *Modern Dynamics: Mathematical Progressions*, 1(2), Article 22. <https://doi.org/10.36676/mdmp.v1.i2.22>
24. Ayyagiri, A., Goel, P., & Renuka, A. (2024). Leveraging AI and Machine Learning for Performance Optimization in Web Applications. *Modern Dynamics: Mathematical Progressions*, 1(2), Article 13. <https://doi.org/10.36676/mdmp.v1.i2.13>
25. Gannamneni, N. K., Salunkhe, V., Chopra, P., Shrivastav, A., Goel, P., & Goel, O. (2022). Enhancing Supply Chain Efficiency through SAP SD/OTC Integration in S/4 HANA. *Universal Research Reports*, 9(4), Article 1396. <https://doi.org/10.36676/urr.v9.i4.1396>
26. Mali, A., Khan, I., Dandu, M. M. K., Goel, P., Jain, A., & Shrivastav, A. (2024). Designing Real-Time Job Search Platforms with Redis Pub/Sub and Machine Learning Integration. *Journal of Quantum Science and Technology*, 1(3), Article 115. <https://doi.org/10.63345/jqst.v1i3.115>
27. Mane, H. R., Kumar, A., Dandu, M. M. K., Goel, P., Jain, A., & Shrivastav, A. (2024). Micro Frontend Architecture With Webpack Module Federation: Enhancing Modularity Focusing On Results And Their Implications. *Journal of Quantum Science and Technology*, 1(4), Article 95. <https://doi.org/10.63345/jqst.v1i4.95>
28. Bhimanapati, V. B. R., Pandian, P. K. G., & Goel, P. (2024). UI/UX Design Principles for Mobile Health Applications. *International Journal for Research Publication and Seminars*, 15(3), Article 1485. <https://doi.org/10.36676/jrps.v15.i3.1485>
29. Kumar, A., & Goel, P. (2025). Enhancing ROI through AI-Powered Customer Interaction Models. *Journal of Quantum Science and Technology*, 2(1), Article 178. <https://doi.org/10.63345/jqst.v2i1.178>
30. Sengar, H. S., Aravind, S., Kolli, R. K., Goel, O., Singh, S. P., & Goel, P. (2024). Ever aging AI/ML Models for Predictive Analytics in SaaS Subscription Management. *Darpan International Research Analysis*, 12(3), Article 136. <https://doi.org/10.36676/dira.v12.i3.136>
31. Abdul, R. O., Khan, I., Vadamani, S., Kumar, L., Goel, P., & Khair, M. A. (2024). Integrated Solutions for Power and Cooling Asset Management through Oracle PLM. *Journal of Quantum Science and Technology*, 1(3), Article 90. <https://doi.org/10.63345/jqst.v1i3.90>
32. Avancha, S., Goel, P., & Renuka, A. (2024). Continuous Service Improvement in IT Operations through Predictive Analytics. *Modern Dynamics: Mathematical Progressions*, 1(2), Article 14. <https://doi.org/10.36676/mdmp.v1.i2.14>
33. Tirupathi, R., Bajaj, A., Mohan, P., Goel, P., Singh, S. P., & Jain, A. (2024). Optimizing SAP Project Systems (PS) for Agile Project Management. *Darpan International Research Analysis*, 12(3), Article 138. <https://doi.org/10.36676/dira.v12.i3.138>
34. Kumar, A., Dandu, M. M. K., Kolli, R. K., Singh, S. P., Goel, P., & Goel, O. (2024). Strategies for Maximizing Customer Lifetime Value through Effective Onboarding and Renewal Management. *Darpan International Research Analysis*, 12(3), Article 127. <https://doi.org/10.36676/dira.v12.i3.127>
35. Chintla, U., Goel, P., & Goel, O. (2024). The Role of Apttus CPQ in Modern CRM Systems: Implementation Challenges and Solutions. *Darpan International Research Analysis*, 12(3), Article 91. <https://doi.org/10.36676/dira.v12.i3.91>
36. Bajaj, A., Singh, S. P., Dandu, M. M. K., Kolli, R. K., Goel, O., & Goel, P. (2024). Advanced Algorithms for Surge Pricing Optimization in Multi-City Ride-Sharing Networks. *Darpan International Research Analysis*, 12(3), Article 137. <https://doi.org/10.36676/dira.v12.i3.137>
37. Vijayabaskar, S., Gangu, K., Gopalakrishna, P. K., Goel, P., & Gupta, V. (2024). Agile Transformation in Financial Technology:

- Best Practices and Challenges. Darpan International Research Analysis, 12(3), Article 96. <https://doi.org/10.36676/dira.v12.i3.96>
38. Alahari, J., Thakur, D., Goel, P., Chintla, V. R., & Kolli, R. K. (2022). Enhancing iOS Application Performance through Swift UI: Transitioning from Objective-C to Swift. *International Journal for Research Publication and Seminars*, 13(5), Article 1509. <https://doi.org/10.36676/jrps.v13.i5.1509>
 39. Khan, I., Agarwal, N., Eeti, S., Goel, O., Jain, A., & Goel, P. (2024). Optimization Techniques for 5G O-RAN Deployment in Cloud Environments. *Darpan International Research Analysis*, 12(3), Article 135. <https://doi.org/10.36676/dira.v12.i3.135>
 40. Murthy, K. K. K., Pandian, P. K. G., & Goel, P. (2024). The Role of Digital Innovation in Modernizing Railway Networks: Case Studies and Lessons Learned. *International Journal for Research Publication and Seminars*, 15(2), Article 1473. <https://doi.org/10.36676/jrps.v15.i2.1473>
 41. Dandu, M. M. K., Arulkumaran, R., Agarwal, N., Aggarwal, A., & Goel, P. (2024). Improving Neural Retrieval with Contrastive Learning. *Modern Dynamics: Mathematical Progressions*, 1(2), Article 30. <https://doi.org/10.36676/mdmp.v1.i2.30>
 42. Agarwal, N., Kolli, R. K., Eeti, S., Jain, A., & Goel, P. (2024). Multi-Sensor Biomarker Using Accelerometer and ECG Data. *Darpan International Research Analysis*, 12(3), Article 103. <https://doi.org/10.36676/dira.v12.i3.103>
 43. Agarwal, N., & Goel, P. (2025). Sustainable Supply Chain Networks: Integrating Green Transportation and Logistics Solutions. *Journal of Quantum Science and Technology*, 2(1), Article 189. <https://doi.org/10.63345/jqst.v2i1.189>
 44. Saoji, M., Tangudu, A., Pagidi, R. K., Goel, O., Jain, A., & Goel, P. (2021). Virtual Reality in Surgery and Rehab: Changing the Game for Doctors and Patients. *Universal Research Reports*, 8(4), Article 1385. <https://doi.org/10.36676/urr.v8.i4.1385>
 45. Goel, P. (2024). AI-Enabled Threat Detection in Blockchain-Based Systems. *Scientific Journal of Artificial Intelligence and Blockchain Technologies*, 1(2), Article 101. <https://doi.org/10.63345/sjaibt.v1.i2.101>
 46. Govindarajan, B., Kolli, R. K., Singh, S. P., Dandu, M. M. K., Goel, O., & Goel, P. (2024). Advanced Techniques in Automation Testing for Large Scale Insurance Platforms. *Journal of Quantum Science and Technology*, 1(1), Article 1. <https://doi.org/10.63345/jqst.v1i1.1>
 47. Avancha, S., Goel, P., & Renuka, A. (2024). Continuous Service Improvement in IT Operations through Predictive Analytics. *Darpan International Research Analysis*, 12(3), Article 90. <https://doi.org/10.36676/dira.v12.i3.90>
 48. Yadav, N., Aravind, S., Bikshapathi, M. S., Prasad, P. S., Jain, S. K., & Goel, P. (2024). Customer Satisfaction Through SAP Order Management Automation. *Journal of Quantum Science and Technology*, 1(4), Article 124. <https://doi.org/10.63345/jqst.v1i4.124>
 49. Ghate, S. V., & Goel, P. (2025). Data Analytics for Supply Chain Management: Leveraging Data to Drive Decision-Making and Improve Efficiency. *Journal of Quantum Science and Technology*, 2(2), Article 281. <https://doi.org/10.63345/jqst.v2i2.281>
 50. Goel, P. (2025). Data Engineering Techniques for Cost-Efficient Retail Data Analytics. *Journal of Quantum Science and Technology*, 2(4), Article 359. <https://doi.org/10.63345/jqst.v2i4.359>
 51. Boddur, N., & Goel, P. (2025). Data Migration and Integration in Oracle Cloud: Challenges and Solutions. *International Journal for Research Publication and Seminars*, 16(1), Article 142. <https://doi.org/10.36676/jrps.v16.i1.142>
 52. Govindankutty, S., & Goel, P. (2024). Data Privacy and Security Challenges in Content Moderation Systems. *Journal of Quantum Science and Technology*, 1(4), Article 132. <https://doi.org/10.63345/jqst.v1i4.132>
 53. Gopa, R., & Goel, P. (2025). Data Quality and Governance for Ensuring Trustworthy and Compliant Data. *Journal of Quantum Science and Technology*, 2(2), Article 285. <https://doi.org/10.63345/jqst.v2i2.285>
 54. Mittal, S., & Goel, P. (2025). Data Science In Retail: Explore Big Data's Role In Ads And Inventory Forecasting. *International Journal of Research in Modern Engineering & Emerging Technology*, 13(3), Article 19. <https://doi.org/10.63345/ijrmeet.org.v13.i3.19>
 55. Jain, A., & Goel, P. (2025). Data Visualization Tools For Fraud Investigation: An Evaluation Of Data Visualization Tools Used For Fraud Investigation. *Universal Research Reports*, 12(1), Article 1475. <https://doi.org/10.36676/urr.v12.i1.1475>
 56. Hajari, V. R., Chawda, A. D., Goel, P., Renuka, A., & Goel, L. (2024). Embedded Systems Design for High-Performance Medical Applications. *Innovative Research Thoughts*, 10(3), Article 1474. <https://doi.org/10.36676/irt.v10.i3.1474>
 57. Goel, P. (2025). Evaluating Audience Engagement Metrics across Streaming and Social Media Platforms. *Journal of Quantum Science and Technology*, 2(3), Article 335. <https://doi.org/10.63345/jqst.v2i3.335>
 58. Agrawal, S., Kolli, R. K., Eeti, S., Goel, P., & Jain, A. (2024). Impact of Lean Six Sigma on Operational Efficiency in Supply Chain Management. *Darpan International Research Analysis*, 12(3), Article 99. <https://doi.org/10.36676/dira.v12.i3.99>
 59. Cheruku, S. R., Pandian, P. K. G., & Goel, P. (2024). Implementing Agile Methodologies in Data Warehouse Projects. *International Journal for Research Publication and Seminars*, 15(3), Article 1498. <https://doi.org/10.36676/jrps.v15.i3.1498>
 60. Goel, P. (2024). The Role of SAP Sales and Distribution in Modern Supply Chain Management. *Journal of Quantum Science and Technology*, 1(1), Article 15. <https://doi.org/10.63345/jqst.v1i1.15>
 61. Goel, P. (2024). The role of Oracle HCM in supporting organizational restructuring and mergers. *Journal of Quantum Science and Technology*, 1(3), Article 200. <https://doi.org/10.63345/jqst.v1i3.200>
 62. Liang, W., Zhang, Y., Cao, H., Wang, B., Ding, D. Y., Yang, X., Vodrahalli, K., He, S., Smith, D. S., Yin, Y., McFarland, D. A., & Zou, J. (2024). Can large language models provide useful feedback on research papers? A large-scale empirical analysis. *NEJM AI*, 1(8), A10a2400196.
 63. Liu, R., & Shah, N. B. (2023). ReviewerGPT? An exploratory study on using large language models for paper reviewing. *arXiv preprint arXiv:2306.00622*.
 64. Padmakumar, V., & He, H. (2024). Does writing with language models reduce content diversity? *International Conference on Learning Representations (ICLR 2024)*.
 65. Sourati, Z., et al. (2026). The shrinking landscape of linguistic diversity in the age of large language models. *Nature Human Behaviour*.