

Positioning Natural Language Processing Use for Emotion Detection from a Humanistic Perspective: A Bibliometric Study

Rasika Shende¹, Jaipal², Junali Sahoo³, Hemanta Kumar Mishra⁴

¹*Research Scholar, Department of Humanities and Social Sciences, Visvesvaraya National Institute of Technology, India*

²*Assistant Professor, Department of Humanities and Social Sciences, Visvesvaraya National Institute of Technology, India*

³*Research Scholar, School of Development Studies, IIHMR University, India*

⁴*Assistant Professor, School of Development Studies, IIHMR University, India*

The burgeoning use of Natural Language Processing (NLP) across diverse disciplines has significantly broadened the scope of this analysis technique. Integrating NLP into complex domains such as emotion detection (ED) brings forth the potential of NLP beyond traditional realms of sciences, extending its influence into humanistic fields like arts, humanities, and social sciences. Although several bibliometric studies have explored various implications of NLP, no significant work has focused specifically on trends in NLP within the context of ED. Thus, this study aims to investigate the global evolution and publishing trends of the intersections of NLP and ED through the Scopus database. Employing visualization tools such as RStudio and VOSviewer, this comprehensive analysis of 984 research articles, reveals publication trends from 2000 to 2023. The findings of the paper indicate the primary subject areas and leading countries with scholarly contributions. The findings also identified key themes that provide the scope for future researchers to understand the main concepts that define NLP and ED. Thus, helping readers and future scholars navigate organizing data is crucial for building strong theoretical models and practices in this rapidly growing area.

Keywords: Natural language processing, Emotion detection, Sentiment analysis, Bibliometric analysis, Scopus database.

1. Introduction

Natural Language Processing (NLP) encompasses a variety of computational methods designed to automatically analyze and represent human language that has garnered substantial acclaim and widespread recognition in recent years (Holdsworth, 2024). This surge in popularity can be majorly attributed to the sophistication of machine learning which improved the accuracy (Kotsiantis et al., 2006), the availability of large datasets (Kotsiantis et al., 2006; Machová et al., 2023), and the practical and commercial implications of this field (Stewart et al., 2024). Consequently, NLP aims to achieve language comprehension akin to humans across various applications and tasks (Chen et al., 2018).

Most of the studies in NLP are in the areas of Masoumi et al., (2024) explored the trend of using NLP in medical research and established the potential of language models to dictate the review process. Gooljar et al., (2024) examined the combination of predictive modelling and sentiment analysis in the commercial context focusing on the significance of AI in Society 5.0. Studies on NLP largely concentrated on sentiment analysis for predicting market trends, sustainability studies, and medical applications of NLP with a limited focus on arts, humanities, and social sciences domains. The exploration in this direction is limited to curriculum design and some use in language research leaving a gap for psycho-social positioning of NLP application (Malgaroli et al., 2023).

Likewise, Emotion detection (ED) involves the computational study of natural language phrases to identify different emotions from text, including joy, trust, fear, surprise, and sadness (Cabrera, 2022; Chu et al., 2022). It includes the automatic identification and classification of emotions expressed in textual data. It is widely accepted that emotion is a potent motivator and predictor of human motivation and perception (Balakrishnan et al., 2019). Moreover, ED is essential not only for interpreting market sentiment (Borg and Boldt, 2020; Gooljar et al., 2024) and addressing mental health concerns in academia (Newman and Joyner, 2018) and professional settings but also in clinical contexts (Vermeir et al., 2015). A rise of studies on ED using NLP has been observed in recent times due to its profound applications (Keramatfar and Amirkhani, 2019; Kumar et al., 2023). Hence, a bibliometric analysis on NLP and ED from humanistic perspective will help ascertain the possibilities of future scope of research drawing on the prominent trends, authors, journals, keywords, and themes.

Further, the bibliometric analysis encapsulates using quantitative techniques, such as keyword analysis, and citation analysis, on bibliometric data, including publication and citation units (Broadus, 1987). Most of the bibliometric studies highlighted significant advancements across diverse interdisciplinary research fields (Donthu et al., 2021), particularly in the realm of the application of NLP (Chen, Ding, et al., 2018). Among many advantages of bibliometric analysis, assessing prominent scientific researchers or publications (Chen et al., 2017), analyzing the network structure of a scientific field, and identifying primary topics (Wallace et al., 2012), uncovering novel advancements (Boudry and Mouriaux, 2015) are significant applications.

This paper conducts a bibliometric analysis of NLP and ED sourced from the Scopus database spanning from 2000 to 2023. The following research questions address the aim of this study:

RQ 1: What has been the historical trajectory of scholarly publications on Natural Language Processing (NLP) and Emotion Detection (ED), including the factors driving this growth, and which are the predominant journals and the prominent authors contributing to the evolution of these areas?

RQ 2: Which countries have made substantial contributions to research at the intersection of Natural Language Processing (NLP) and Emotion Detection (ED), and how do researchers typically frame their studies in these areas using keywords?

2. Methodology

The increased use of NLP in the fields of ‘Social science’, ‘Arts & Humanities’, and ‘Psychology’ depicts the rising interest of authors in this domain. To emphasize the human and social components of NLP and ED, the paper employed the Scopus database using the following keywords ("Natural Language Processing" OR "NLP" OR "Computational Linguistics" OR "Language Modelling") AND ("Emotion Detection" OR "Emotion Recognition" OR "Mood Recognition" OR "Sentiment Detection" OR "Emotion Identification" OR "Mood Detection" OR "Sentiment Recognition") in the ‘All fields’ setting that provided a clear understanding of the intellectual contribution in the form of studies published in this area. Scopus is considered a pivotal data source for affective computing (Chen, Xie, et al., 2018b), with an extensive amount of bibliometric data that can be used by tools like MS Excel, VOSviewer, and RStudio.

All records matching the original search string in Scopus are included in this phase of the initial search to collect every material pertinent to NLP and ED. The document types were filtered to include Articles only to narrow down the research. Language was selected as English to ensure understanding and uniformity in analysis. Journals were selected to ensure rigorous peer review and the final stage of publication was opted to exclude manuscripts that are presently under consideration or those that have preliminary access status. The publications were taken that were published between 2000-2023 to emphasize more current research, which captures the growth of NLP and ED over the last twenty years. Maintaining a reasonable dataset, this range probably encompasses major developments in the field.

Furthermore, this paper's performance analysis and science mapping techniques render a descriptive account of constituents including year-wise publication trends, prominent authors, journals, and countries in this field. The relationship between the constituent studies depicted is based on using bibliographic coupling and co-occurrence analysis in this paper (Donthu et al., 2021). Table 1 provides the main information about the entire data. Figure 1 illustrates the PRISMA diagram to understand the search strategy with inclusion and exclusion criteria, and the screening strategy followed by bibliometric methodology after formulating RQs.

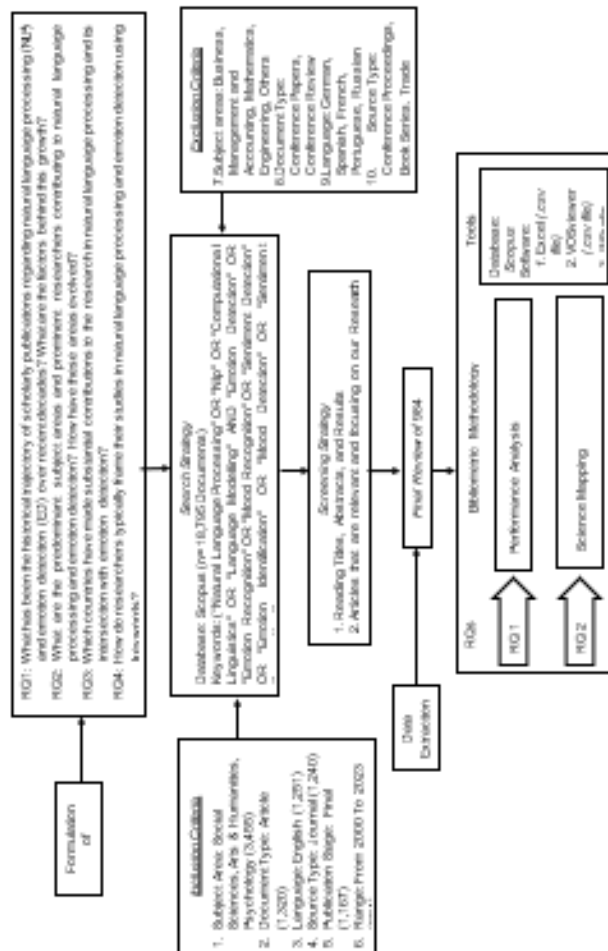


Figure 1. PRISMA Flow diagram of Data Extraction and Methodology of the study

Table 1 Information About the Data

Description	Results
Timespan	2000-2023
Sources (Journals, Books, etc)	327
Documents	984
Annual Growth Rate %	27.13
Document Average Age	4.16
Average citations per doc	32.94
References	65268
DOCUMENT CONTENTS	
Keywords Plus (ID)	4101
Author's Keywords (DE)	2838

AUTHORS	
Authors	3216
Authors of single-authored docs	80
AUTHORS COLLABORATION	
Single-authored docs	82
Co-Authors per Doc	3.55
International co-authorships %	27.44
DOCUMENT TYPES	
Article	984

Source: Scopus 2024

3. Results

3.1 Performance Analysis

Performance analysis aspires to assess the scholarly contributions and impact of scientific entities through the meticulous examination of bibliographic data (Noyons et al., 1999). Performance analysis was integrated into the study to assesses the activity of scientific constituents and their impact.

3.1a Publication Trends

Figure 2 shows a substantial increase in scholarly publication output related to NLP and ED throughout the selected years. The expansion of machine learning across various disciplines with the rise in interdisciplinarity fields contributed to the rise in production. Various other factors like accessibility to advanced training, funding, computational power, and the rise in demand due to numerous applications of ED are reflected in the increased popularity of the domain.

At its inception, the early 2000s saw minimal research activity in this field, with only one publication in 2000 and 2002, four publications in 2003 with a decline in publications in 2005 which can be attributed to the beginning of the internet broadband boom and the start of the penetration of mobile technologies which remained less explored for NLP and ED initially (Coffman & Odlyzko, 2002). Since, 2009 there has been a consistent upward trend in the amount of published works, with a slight decrease in the years 2012, 2014 and 2018, with a more pronounced and remarkable surge.

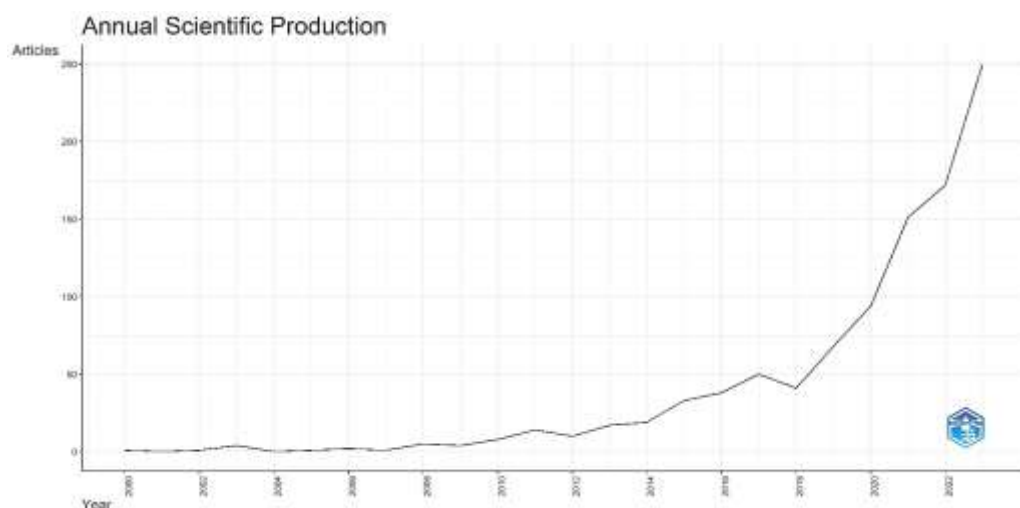


Figure 2 Annual publication trend on NLP and ED between 2000-2023

Alongside the burgeoning publication output, the Total Citations (TC) of these scholarly works has experienced one publication in 2001. And showing a sharp decline through years 2002 and 2003 as can be seen in Table 2. There is a significant drop, reaching a low point in the year 2006, and then a gradual increase from the year 2007 to 2008. A notable increase in citations peaked in the year 2011. The trend continues to fluctuate, peaking again around the year 2018, followed by a steady decline towards the year 2022.

Table 2 Annual Citations Per Year

Year	MeanTCperArt	N	MeanTCperYear	CitableYears
2000	157	1	6.28	25
2002	72	1	3.13	23
2003	98.75	4	4.49	22
2005	1	1	0.05	20
2006	16.5	2	0.87	19
2007	16	1	0.89	18
2008	94.4	5	5.55	17
2009	163.25	4	10.2	16
2010	141.88	8	9.46	15
2011	102.43	14	7.32	14
2012	18.8	10	1.45	13
2013	32.41	17	2.7	12
2014	53.37	19	4.85	11
2015	65.24	33	6.52	10
2016	43.21	38	4.8	9
2017	76.56	50	9.57	8
2018	41.66	41	5.95	7
2019	34.59	68	5.77	6
2020	65.3	94	13.06	5
2021	25.67	151	6.42	4
2022	17.12	172	5.71	3
2023	6.6	250	3.3	2

*TC total citation Source: Scopus 2024

3.1b Most Prominent Journals

The publication surge in NLP-ED is attributable not solely to the advent and widespread
Nanotechnology Perceptions Vol. 20 No.6 (2024)

adoption of ‘advanced technologies’ (Kurtz et al., 2024) but also to various other contributing factors. The integration of sophisticated technologies into myriad fields is now commonplace (Aldoseri et al., 2024). Rising educational levels (Stupples, 2015), increasing the gross domestic product (GDP) of nations (Callen, n.d.), and individuals' growing spending capacities collectively propel this burgeoning market. A primary application of NLP for ED lies in analyzing consumer sentiment (Borg and Boldt, 2020), which has garnered significant popularity. Figure 3 focuses on the top ten most pertinent journals publishing in area of NLP and ED. The preeminent position is occupied by the Information Processing and Management, published articles (n=74) in the scrutinized domain, closely trailed by Artificial Intelligence Review (n=52) documents. The journal Information Processing and Management demarcated an impact factor of 7.4 and h index of 31, g index of 56, and m index of 2.214 with a focus on publishing in the intersection of computer and information science including social computing which supports the high number of publications and citations in this journal in the area of NLP and ED. A list of the journal’s local impact variables is given in Table 3.

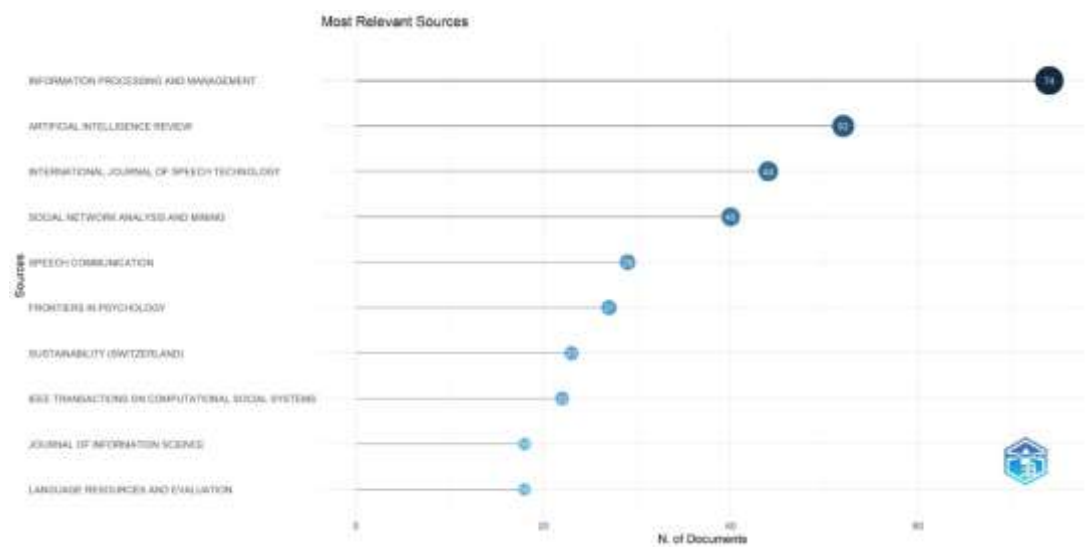


Figure 3. Top 10 most relevant sources in NLP and ED

Table 3 Journal’s Local Impact

Source	h_index	g_index	m_index	TC	NP	PY_start
Information processing and management	31	56	2.214	3299	74	2011
Artificial intelligence review	26	52	2.6	5973	52	2015
International journal of speech technology	17	28	0.773	822	44	2003
Speech communication	16	29	0.727	1556	29	2003
Social network analysis and mining	13	28	1.083	807	40	2013
Journal of the Association for Information Science and Technology	12	14	1.091	566	14	2014
Sustainability (Switzerland)	12	19	2	373	23	2019
Journal of Information Science	11	18	0.917	553	18	2013
Computers in human behavior	9	12	0.9	654	12	2015
Proceedings of the ACM on human-computer interaction	9	14	1.125	201	16	2017
IEEE transactions on computational social Systems	8	18	1.6	334	22	2020
Natural language engineering	8	11	0.571	692	11	2011

Online social networks and media	8	11	1	283	11	2017
User Modeling and user-adapted interaction	8	9	0.348	900	9	2002
Decision support systems	7	9	0.636	721	9	2014
Frontiers in psychology	7	18	0.5	333	27	2011
International journal of information management data insights	7	7	1.75	641	7	2021
Language resources and evaluation	7	14	0.583	224	18	2013
Scientometrics	7	12	0.778	152	12	2016
Computational linguistics	6	6	0.75	172	6	2017

Source: Scopus 2024

3.1c Most Prominent Authors

NLP in ED research has captivated researchers' interest for the past decade as illustrated in Figure 4. Moreover, the deployment of NLP in educational settings for assessment and scoring purposes can significantly alleviate teachers' workloads (Jin, 2022; Watkins et al., 2020). It can also potentially serve as an early warning system for identifying 'concerning student essays' that may warrant intervention. The academic realm offers numerous applications for ED via NLP, for instance, the maintenance of student mental well-being (Newman and Joyner, 2018). The utility of NLP extends to various sensitive professions and vocational education fields requiring high emotional intelligence, such as nursing, engineering, and law (Margaroli et al., 2023).

Beginning with a single article published by Dr. Pushpak Bhattacharyya, there has been a notable and consistent rise in the number of publications annually. A significant surge was observed from 2019 to 2021, during which authors increasingly focused on this topic, resulting in the publication of more than one article per year. The rise can potentially be attributed to several factors, including the increased awareness and importance of mental health, and the proliferation of social media and digital communication.

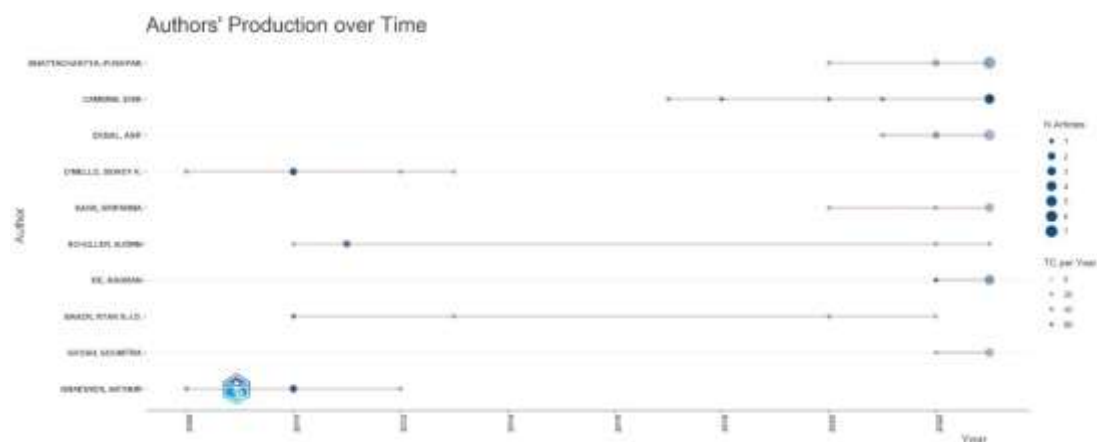


Figure 4. Top 10 Authors' production over time

3.1d Countries' Collaborations on NLP and ED

The interdisciplinary nature of NLP and ED provides opportunities for collaboration with science and arts departments resulting in innovative research at the intersection of social sciences and technology (Margaroli et al., 2023). Further, the increase in funding and the

real-time applications of NLP is reflected by its rise in arts, humanities and social sciences research. Furthermore, investigations on NLP and ED require the cooperation of authors from multiple areas to generate new knowledge. Table 4 depicts the top 10 countries making significant contributions to research in NLP and ED based on the total number of publications. The United States is foremost with n=622 published documents. With this quantity of publication production, the United States showcases its commitment to shaping knowledge in this domain. China follows closely with n= 566 publications, exemplifying the country’s burgeoning interest and active engagement in the area of critical study. India showcases its resolve by holding the third position with n= 349 published documents, reaffirming its substantial contribution to global research on NLP and ED.

Table 4 Countries’ Production of NLP and ED Research

Country	Frequency
USA	622
China	566
India	349
UK	171
Germany	128
Spain	117
Italy	102
Australia	90
South Korea	82
Canada	79

Source: Scopus 2024

Countries collaborating over the domain of NLP and ED provide a global perspective of the plethora of work being done with a rich exchange of information. Australia and Pakistan worked together on 4 articles, followed by Canada and France who also collaborated on 4 publications. Australia-Malaysia; Australia-Singapore collaborated on 3 articles and Australia collaborated with Hong Kong on 2 articles. The dense collaboration network is evident through the Australian efforts in this area. Canada followed by Australia further showcased its commitment towards NLP and ED. The increased funding, technological support, and government initiatives to encourage technology use are some contributors to the rise in production along with the promotion of the collaboration culture in Australia and Canada. The high number of diasporas could also be one of the reasons for cross-country collaborations of these select countries. The collaboration network of countries is depicted in Figure 5.

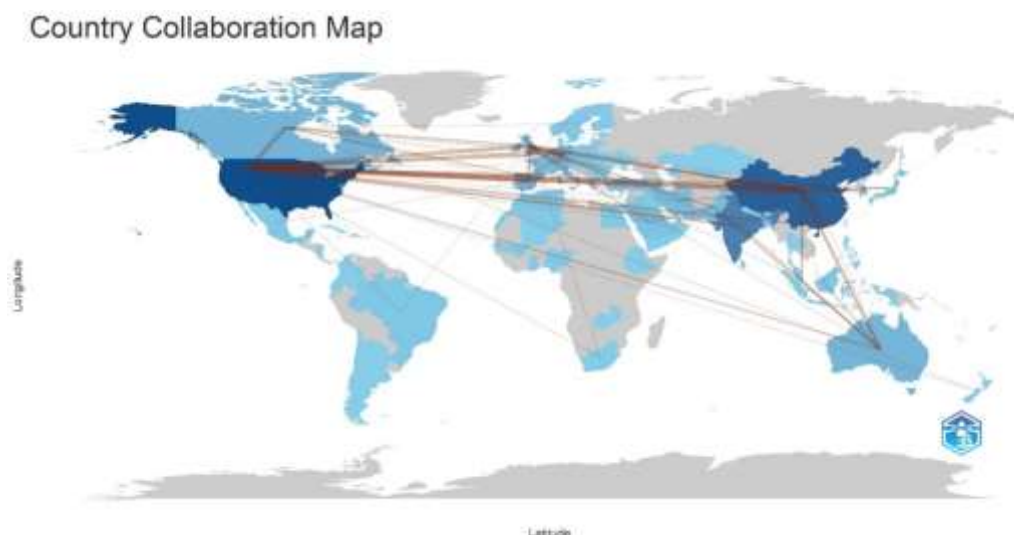


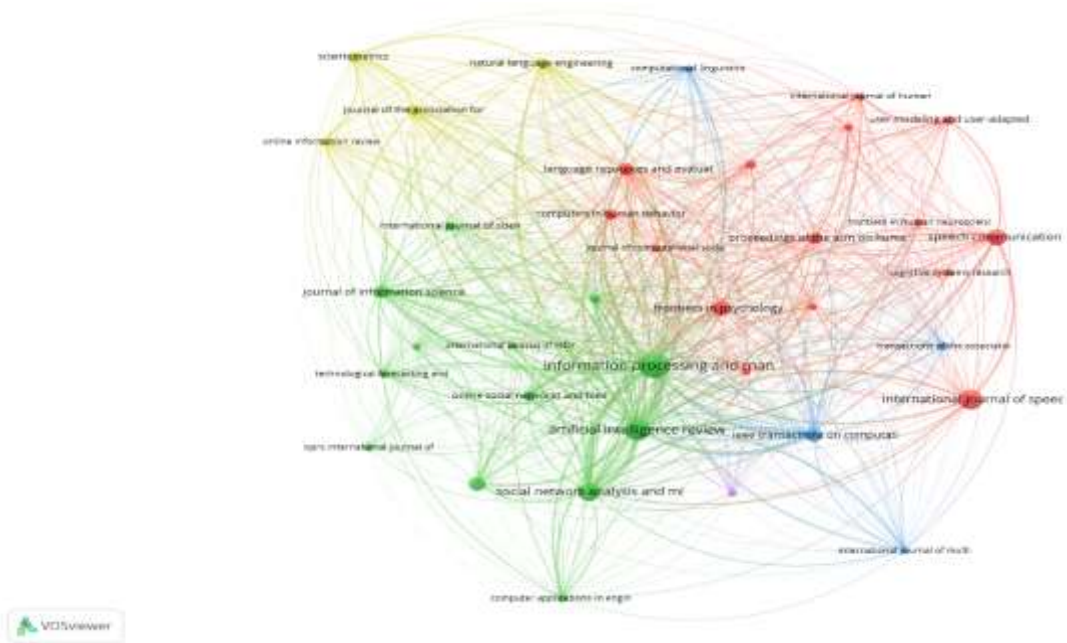
Figure 5. Top collaborating countries to research in NLP and ED

3.2 Science Mapping

Science mapping primarily focuses on monitoring a scientific field to ascertain its structure, evolution, and key contributors. A standard science mapping study explores the relationship between the constituent studies that are depicted using bibliographic coupling and co-occurrence analysis in this study to answer RQ 2 (Donthu et al., 2021).

3.2a Bibliographic Coupling

In Figure 6, the clusters include journals primarily related to psychology, human behavior, and speech communication. Key journals include *Frontiers in Psychology*, *Computers in Human Behavior*, *Speech Communication*, and *Cognitive Systems Research* which are represented in red color. The green-colored cluster consists of *Information Processing & Management*, *Social Network Analysis and Mining*, *Journal of Information Science*, *Artificial Intelligence Review*. *Information Processing & Management* is central in the visualization, indicating strong bibliographic coupling with journals from multiple clusters, which suggests it is widely cited across various related fields. Journals represented through Yellow cluster are related to metrics, *Online Information Review* etc., and are named *Scientometrics*, *Online Information Review*. Journals indicated by Blue clusters include *Natural Language Engineering*, *User Modelling and User-Adapted Interaction*, and *International Journal of Human-Computer Studies*, however, Purple is Smaller than the others and has *Computer Applications in Engineering* journal.



3.2b Co-occurrence Analysis for Author's Keywords

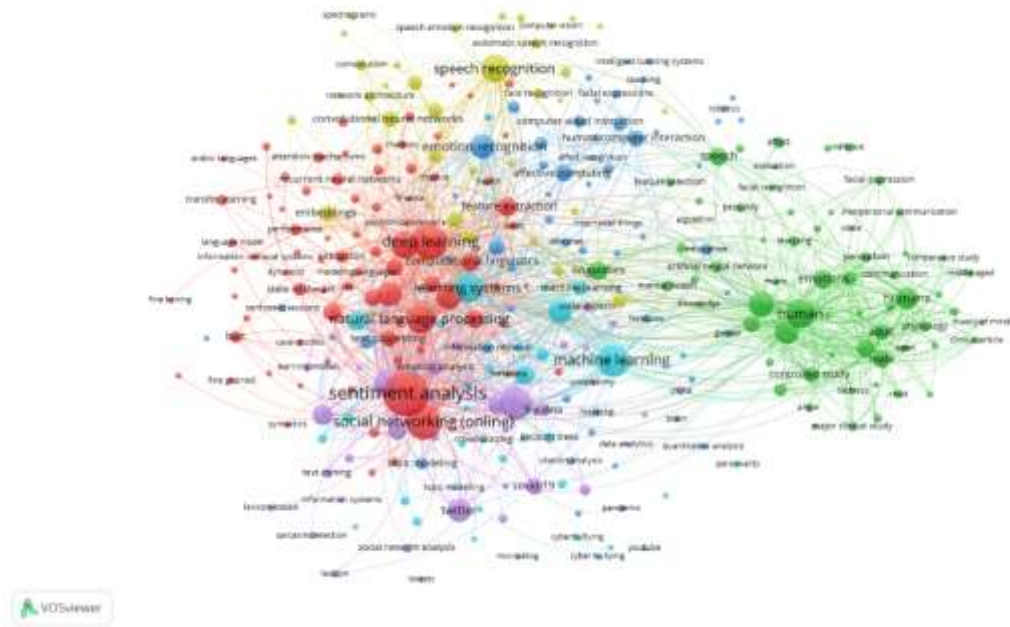


Figure 7 Visualization based on the author's keywords

4. Conclusion

The comprehensive bibliometric analysis provided insights into the intricate relationship between NLP and ED. The rising interest in NLP for ED, as evidenced by the increasing number of publications, reflects the growing recognition of its potential applications, particularly in educational technology beyond a conventionally demarcated science and arts domain. The identification of key publication subjects and prominent authors, journals, and collaborating countries provides insights into the interdisciplinary nature of NLP and the growing global collaboration network. The prevalence of the major keywords like ‘sentiment analysis’, ‘human’, ‘emotion recognition’, ‘speech recognition’, and ‘social media’ highlights significance in the research landscape, indicating the heightened interest among authors in this domain and carving scope for future study. The results emphasize the noteworthy application of NLP for ED in academic, psychological, economic, political, and social relevance and confirm the necessity of further exploring this subject. This research demonstrates its importance as a data source for understanding the landscape of NLP and ED research. The insights gained from this study can guide future research directions, foster collaboration, and ultimately enhance the development and application of NLP technologies in various domains.

Declarations

Availability of data and materials

Not Applicable

Ethical considerations

Not Applicable

Funding

No Funding

References

1. Aldoseri A, Al-Khalifa KN and Hamouda AM (2024) AI-Powered Innovation in Digital Transformation: Key Pillars and Industry Impact. *Sustainability* (Switzerland) 16(5). Multidisciplinary Digital Publishing Institute (MDPI).
2. Balakrishnan V, Cynthia Martin M, Kaur W, et al. (2019) A comparative analysis of detection mechanisms for emotion detection. In: *Journal of Physics: Conference Series*, 16 December 2019. Institute of Physics Publishing.
3. Borg A and Boldt M (2020) Using VADER sentiment and SVM for predicting customer response sentiment. *Expert Systems with Applications* 162.
4. Boudry C and Mouriaux F (2015) Eye Neoplasms Research: A Bibliometric Analysis from 1966 to 2012. *European Journal of Ophthalmology* 25(4): 357–365.
5. Broadus RN (1987) Toward a definition of “bibliometrics”. *Scientometrics* 12(5): 373–379.
6. Cabrera RG (2022) Title Sentiment Analysis for a Humanist Framework: How Emotions are Recognized and Interpreted in the Age of Social Media Sentiment Analysis for a Humanist Framework: How Emotions are Recognized and Interpreted in the Age of Social Media. *Rupkatha Journal* 14(2).
7. Callen T (n.d.) Gross Domestic Product: An Economy’s All - Back to Basics: Economics Concepts Explained - FINANCE & DEVELOPMENT.
8. Chen X, Boyu C, Zhang C, et al. (2017) Discovering the Recent Research in Natural Language Processing Field Based on a Statistical Approach., pp. 507–517.
9. Chen X, Xie H, Wang FL, et al. (2018a) A bibliometric analysis of natural language processing in medical research. *BMC Medical Informatics and Decision Making* 18. BioMed Central Ltd.
10. Chen X, Xie H, Wang FL, et al. (2018b) A bibliometric analysis of natural language processing in medical research. *BMC Medical Informatics and Decision Making* 18.
11. Chen X, Ding R, Xu K, et al. (2018) A Bibliometric Review of Natural Language Processing Empowered Mobile Computing. Hindawi Limited.
12. Chu KE, Keikhosrokiani P and Asl MP (2022) A Topic Modeling and Sentiment Analysis Model for Detection and Visualization of Themes in Literary Texts. *Pertanika Journal of Science and Technology* 30(4): 2535–2561.
13. Donthu N, Kumar S, Mukherjee D, et al. (2021) How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research* 133. Elsevier Inc.: 285–296.
14. Gooljar V, Issa T, Hardin-Ramanan S, et al. (2024) Sentiment-based predictive models for online purchases in the era of marketing 5.0: a systematic review. *Journal of Big Data* 11(1): 107.
15. Holdsworth J (2024) What is NLP (natural language processing)?
16. Jin Y (2022) Sentiment Analysis in British and American Literature Teaching under Formative Assessment and Machine Learning. *Wireless Communications and Mobile Computing* Hashmi MF (ed.) 2022: 1–9.
17. Keramatfar A and Amirkhani H (2019) Bibliometrics of sentiment analysis literature. *Journal of Information Science* 45(1): 3–15.
18. Kotsiantis SB, Zaharakis ID and Pintelas PE (2006) Machine learning: a review of *Nanotechnology Perceptions* Vol. 20 No.6 (2024)

- classification and combining techniques. *Artificial Intelligence Review* 26(3): 159–190.
19. Kumar A, Bhartee AK, Gupta H, et al. (2023) A Review on Emotion Detection of Contextual Text. In: *Proceedings - IEEE 2023 5th International Conference on Advances in Computing, Communication Control and Networking, ICAC3N 2023, 2023*, pp. 465–471. Institute of Electrical and Electronics Engineers Inc.
 20. Kurtz G, Amzalag M, Shaked N, et al. (2024) Strategies for Integrating Generative AI into Higher Education: Navigating Challenges and Leveraging Opportunities. *Education Sciences* 14(5). Multidisciplinary Digital Publishing Institute (MDPI).
 21. Machová K, Szabóová M, Paralič J, et al. (2023) Detection of emotion by text analysis using machine learning. *Frontiers in Psychology* 14. Frontiers Media SA: 1–14.
 22. Malgaroli M, Hull TD, Zech JM, et al. (2023) Natural language processing for mental health interventions: a systematic review and research framework. *Translational Psychiatry* 13(1): 309.
 23. Masoumi S, Amirkhani H, Sadeghian N, et al. (2024) Natural language processing (NLP) to facilitate abstract review in medical research: the application of BioBERT to exploring the 20-year use of NLP in medical research. *Systematic Reviews* 13(1): 107.
 24. Newman H and Joyner D (2018) Sentiment Analysis of Student Evaluations of Teaching., pp. 246–250. Available at: http://link.springer.com/10.1007/978-3-319-93846-2_45.
 25. Noyons ECM, Moed HF and Raan AFJ Van (1999) INTEGRATING RESEARCH PERFORMANCE ANALYSIS AND SCIENCE MAPPING.
 26. Stewart R, Chaturvedi J and Roberts A (2024) Natural language processing–relevance to patient outcomes and real-world evidence. *Expert Review of Pharmacoeconomics and Outcomes Research*. Taylor and Francis Ltd.
 27. Stupples D (2015) The world is more educated than ever. How?
 28. Vermeir P, Vandijck D, Degroote S, et al. (2015) Communication in healthcare: A narrative review of the literature and practical recommendations. *International Journal of Clinical Practice*.
 29. Wallace M, Larivière V and Gingras Y (2012) A Small World of Citations? The Influence of Collaboration Networks on Citation Practices. *PloS one* 7: e33339.
 30. Watkins J, Fabielli M and Mahmud M (2020) SENSE: a Student Performance Quantifier using Sentiment Analysis. In: *2020 International Joint Conference on Neural Networks (IJCNN)*, July 2020, pp. 1–6. IEEE