Πρακτικά του Διεθνούς Συνεδρίου για τη Διεθνή Επιχειρηματικότητα (ICIB)

Τόμ. 1, Αρ. 1 (2025)

Διεθνές Συνέδριο για τη Διεθνή Επιχειρηματικότητα (ICIB) - Πρακτικά Συνεδρίων 2023–2024

Machine learning in Social Sciences and Humanities research: A structured literature review

Aristidis Bitzenis, Nikos Koutsoupias, Marios Nosios

doi: 10.12681/icib.8050

Βιβλιογραφική αναφορά:

Bitzenis, A., Koutsoupias, N., & Nosios, M. (2025). Machine learning in Social Sciences and Humanities research: A structured literature review. *Πρακτικά του Διεθνούς Συνεδρίου για τη Διεθνή Επιχειρηματικότητα (ICIB)*, *1*(1), 1–17. https://doi.org/10.12681/icib.8050

Machine learning in Social Sciences and Humanities research: A structured literature review

Aristidis Bitzenis, Nikos Koutsoupias, Marios Nosios

Department of International and European Studies, University of Macedonia (UoM)

Thessaloniki, Greece

bitzenis@uom.edu.gr, nk@uom.edu.gr, mnosios@uom.edu.gr

Abstract

This research explores the growing role of machine learning in social sciences and humanities research, aiming to provide a comprehensive bibliometric review of its applications, trends, and impact. A systematic approach was employed, combining bibliometric analysis and visualization techniques to analyze a dataset extracted from the Scopus database, focusing on scholarly works related to machine learning in disciplines such as sociology, philosophy, history, and economics. The analysis highlights the increasing volume of publications and the evolving research landscape, with notable growth in recent years, particularly after 2005. Key findings indicate that machine learning is primarily applied in areas such as decision-making, sustainability, social media analysis, and natural language processing, reflecting its diverse potential in addressing complex societal issues. The study also reveals a strong collaborative nature, with a significant percentage of publications involving international co-authorship. Geographic analysis shows that countries like the United Kingdom, India, and the United States are leading contributors, while emerging nations also make noteworthy contributions. The findings suggest that machine learning is becoming an essential tool for social and cultural research, offering new insights into behavioral patterns, governance, and global challenges. In conclusion, the interdisciplinary nature of this approach is emphasized, along with the need for continued methodological innovation and cross-border collaboration to enhance its impact in these fields.

Introduction

The integration of Machine Learning (ML) techniques and methodologies in social and humanities sciences research has emerged as a transformative approach, offering innovative tools for analyzing complex datasets, uncovering hidden patterns and enhancing traditional research methodologies. The convergence of computational advancements and the increasing availability of large datasets have significantly heightened interest in applying ML to address multidimensional inquiries across various disciplines (Lundberg et al., 2022; Rahal et al., 2022).

The use of ML methods in the social sciences and humanities can be broadly categorized into four primary types. Firstly, supervised machine learning algorithms, such as Support Vector Machines (SVMs), logistic regression, and neural networks, are widely utilized for classification tasks. These include sentiment analysis, which assesses public sentiment through social media and predictive modeling, used to forecast voting behavior, economic trends or public health outcomes (Sarker, 2021). For instance, supervised learning models have been applied to analyze social media data during elections, providing insights into voter sentiment and predicting outcomes (Chauhan et al., 2020). The second method, unsupervised machine learning, includes clustering methods like k-means and hierarchical clustering, help researchers identify latent patterns and groupings within data (Alloghani et al., 2019). These techniques are extensively used in social network analysis, where they identify community structures, map the diffusion of information, or trace misinformation in online platforms (Usama et al., 2019). In the social sciences and humanities disciplines, clustering methods have been employed to categorize large collections of historical texts, artworks or cultural artifacts based on thematic or stylistic similarities, revealing patterns that were previously inaccessible through manual analysis (Chapinal-Heras & Díaz-Sánchez, 2023; Zhang & Peng, 2022; Grimmer et al., 2021).

Third, Natural Language Processing (NLP) has become a cornerstone of ML applications in the social sciences and humanities, enabling tasks such as text classification, Named Entity Recognition (NER), and semantic analysis (Chowdhary, 2020; Nadkarni et al., 2011). For instance, topic modeling techniques like Latent Dirichlet Allocation (LDA) have been used to analyze large corpora of documents, extracting thematic trends from archival materials or public speeches (Hagen, 2018; Quinn et al., 2009). Lastly, Deep Learning (DL) approaches, particularly models such as BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer), have revolutionized the analysis of unstructured data, such as text and images. These advanced models have been applied to tasks such as machine translation, text summarization, and image recognition (Yenduri et al., 2024; Sarker, 2021; Janiesch et al., 2021). In the humanities, DL models have been employed to classify and authenticate visual

artworks, aiding art history and cultural heritage preservation (Santos et al., 2021; Castellano & Vessio, 2021; Polak et al., 2017).

In addition to the above, there is also semi-supervised learning, a hybrid approach between supervised and unsupervised learning, which is increasingly gaining traction in the field. This method leverages a small amount of labeled data alongside a large volume of unlabeled data to build models that can generalize effectively, addressing challenges posed by datasets where acquiring labeled data is costly, time-consuming, or difficult (Van Engelen & Hoos, 2019; Zhu & Goldberg, 2009).

In conclusion, the breadth of ML applications reflects its transformative potential in advancing research within the social sciences and humanities. By leveraging cutting-edge methodologies such as supervised learning, unsupervised learning, semi-supervised learning, NLP, and deep learning, researchers can explore complex phenomena and gain insights previously unattainable through traditional approaches.

Methodology

To investigate the role of ML in social sciences and humanities research, this study employed a systematic approach combining bibliometric analysis and visualization techniques (Thomos et al., 2023; Bitzenis & Koutsoupias, 2023B). The research process began with the construction of a comprehensive dataset focusing on scholarly works related to ML applications in these disciplines. The dataset was extracted from the Scopus database using a well-defined search strategy to ensure the inclusion of a broad range of relevant academic publications (Baas et al., 2020).

In addition to the above and for the needs of this analysis, the PRISMA (2020) guidelines (Preferred Reporting Items for Systematic Reviews and Mesa-Analysis) were adhered for the optimal management and reporting of manuscripts, as demonstrated in Figure 1 (Page et al., 2021).

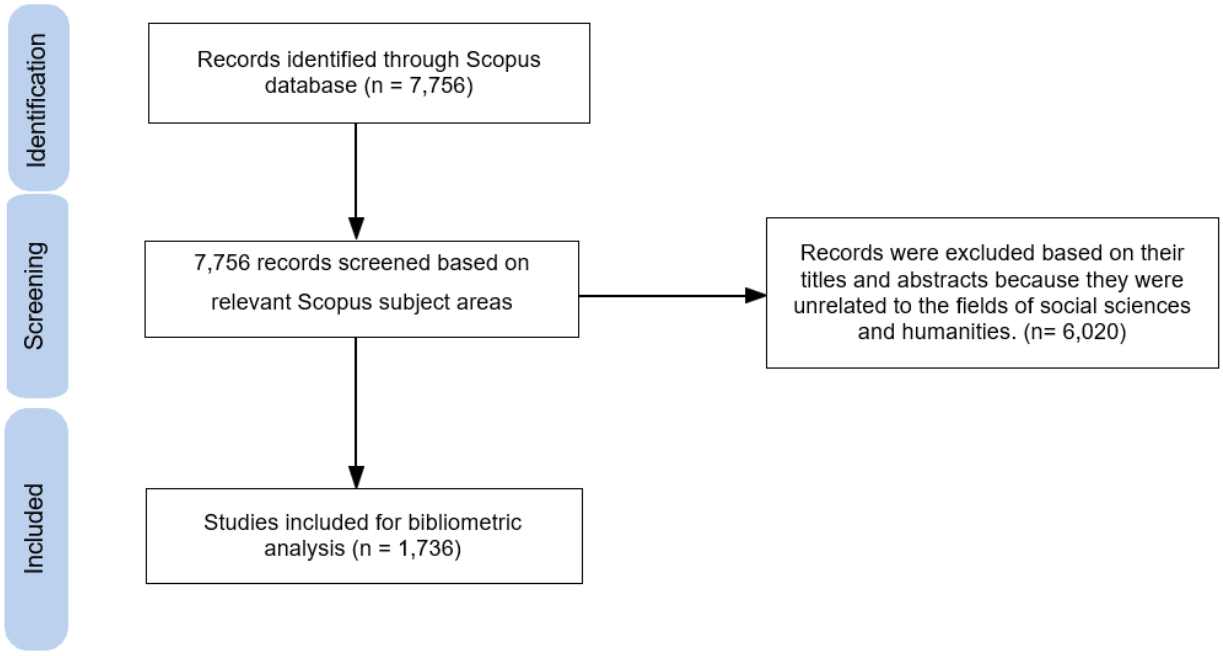

Figure 1: The PRISMA 2020 statement

This query targeted English-language journal articles within subject areas such as sociology, arts, psychology, and economics to align with the study's social sciences and humanities focus area. The extracted dataset included bibliographic metadata such as publication years, keywords, abstracts, citation counts, and authorship details, ensuring a rich foundation for analysis. The query used is the following:

TITLE-ABS-KEY (("machine learning" OR "ML" OR "artificial intelligence" OR "AI") AND ("social sciences" OR "humanities" OR "sociology" OR "philosophy" OR "history" OR "literature" OR "arts" OR "anthropology" OR "political science" OR "education" OR "humanities" OR "social sciences") AND ("literature review" OR "systematic review" OR "extensive review" OR "bibliometric analysis")) AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (SUBJAREA , "SOCI") OR LIMIT-TO (SUBJAREA , "ARTS") OR LIMIT-TO (SUBJAREA , "PSYC") OR LIMIT-TO (SUBJAREA , "ECON"))

The extracted dataset was analyzed using the Bibliometrix R package, a robust tool for bibliometric research that facilitated detailed exploration of publication trends, collaboration networks, and research themes (Aria & Cuccurullo, 2017; Koutsoupias, 2024; Koutsoupias & Papadimitriou, 2020; Bitzenis & Koutsoupias, 2023A). Descriptive statistics, such as annual publication growth rates, average citations per document, and international collaboration levels, were calculated to characterize the field’s evolution. Keyword co-occurrence networks were constructed to identify core research themes and emerging topics, while authorship and co-authorship analyses provided insights into collaboration dynamics. The geographic

analysis underscored the contributions of leading nations and institutions to the field, offering insights into global patterns of engagement and revealing disparities in regional involvement.

Ultimately, through this integrative bibliometric approach, the study offers systematic and data-driven exploration of ML's role in social sciences and humanities research (Bitzenis et al., 2025; Bitzenis & Koutsoupias, 2024; Bitzenis et al., 2023). The findings provide critical insights into publication patterns, thematic priorities and global collaboration, contributing to the growing body of knowledge on the intersection of technology and social sciences. This methodology illuminates key trends and opportunities for future exploration in this rapidly evolving field.

Results

This section presents the key findings from the analysis, based on data extracted from bibliometric and keyword analysis. The results are illustrated through a combination of tables and figures, providing a detailed overview of the dominant themes, trends, and geographic distribution of research, as well as the relationships between key terms in the field. These visualizations help to contextualize the evolution of ML methodologies and their growing impact on interdisciplinary research.

Table 1: Descriptive Statistics

Description	Results
Timespan	1986:2025
Sources (Journals, Books, etc.)	794
Documents	1,736
Annual Growth Rate %	2.86%
Document Average Age	1.86 years
Average Citations per Doc	19.99
References	0
Keywords Plus (ID)	4,682
Author's Keywords (DE)	5,186
Authors	5,263
Authors of Single-authored Docs	207
Single-authored Docs	222
Co-Authors per Doc	3.38
International Co-Authorships %	31.6%
Document Types	Article (1,736)

Table 1 provides a comprehensive overview of the dataset's structure and insights into the research trajectory within this interdisciplinary field. The dataset spans nearly four decades (1986–2025),

demonstrating the long-standing interest in this interdisciplinary domain. It encompasses 796 diverse sources (journals, books, etc.), reflecting its wide-ranging nature and the multidisciplinary appeal of the topic. The 1,736 documents show steady academic activity in this field, with an annual growth rate of 2.86%, indicating moderate yet consistent growth in scholarly contributions.

The average document age of 1.86 years suggests that the field is relatively new and dynamic, with most documents published in the recent past. This aligns with the emergence of advanced ML methods and their applications in the social sciences and humanities. The average of 19.99 citations per document underline the significant academic impact of these works, suggesting that the research resonates well with the broader academic community. However, the lack of references in the dataset indicates that specific citation metrics are unavailable for deeper bibliometric analysis. Furthermore, the document content category reveals a rich repository of information, with 4,682 Keywords Plus and 5,186 author-provided keywords. These extensive keywords highlight the thematic diversity of the field and the importance of understanding ML in diverse social, cultural, and ethical contexts. The higher number of author-provided keywords indicates an emphasis on personalized and specific thematic areas selected by researchers.

In terms of authors, a total of 5,263 contributors underscore the significant level of interest and collaboration in this interdisciplinary field. Interestingly, only 207 authors contributed single-authored documents, while the majority engaged in collaborative efforts, which is further supported by the average of 3.38 co-authors per document. The relatively high 31.6% rate of international co-authorship reflects the global nature of this research, indicating that addressing social science and humanities questions with the use of ML requires diverse perspectives and cross-border collaborations. Lastly, the document types section reveals that all 1,736 entries have been categorized as articles. This suggests that peer-reviewed journal articles are the dominant medium for disseminating knowledge in this domain, further emphasizing the academic rigor of the field.

Ultimately, the analysis highlights the collaborative and expanding nature of research within this dynamic and evolving domain. The increasing number of publications, significant citation impact, and strong international collaborations underscore the field's interdisciplinary appeal. The relatively recent emergence of many documents points to a rapidly evolving research landscape, while the diverse range of keywords reflects the broad thematic scope. As the field matures, it is likely to benefit from deeper bibliometric analyses and greater integration across disciplines.

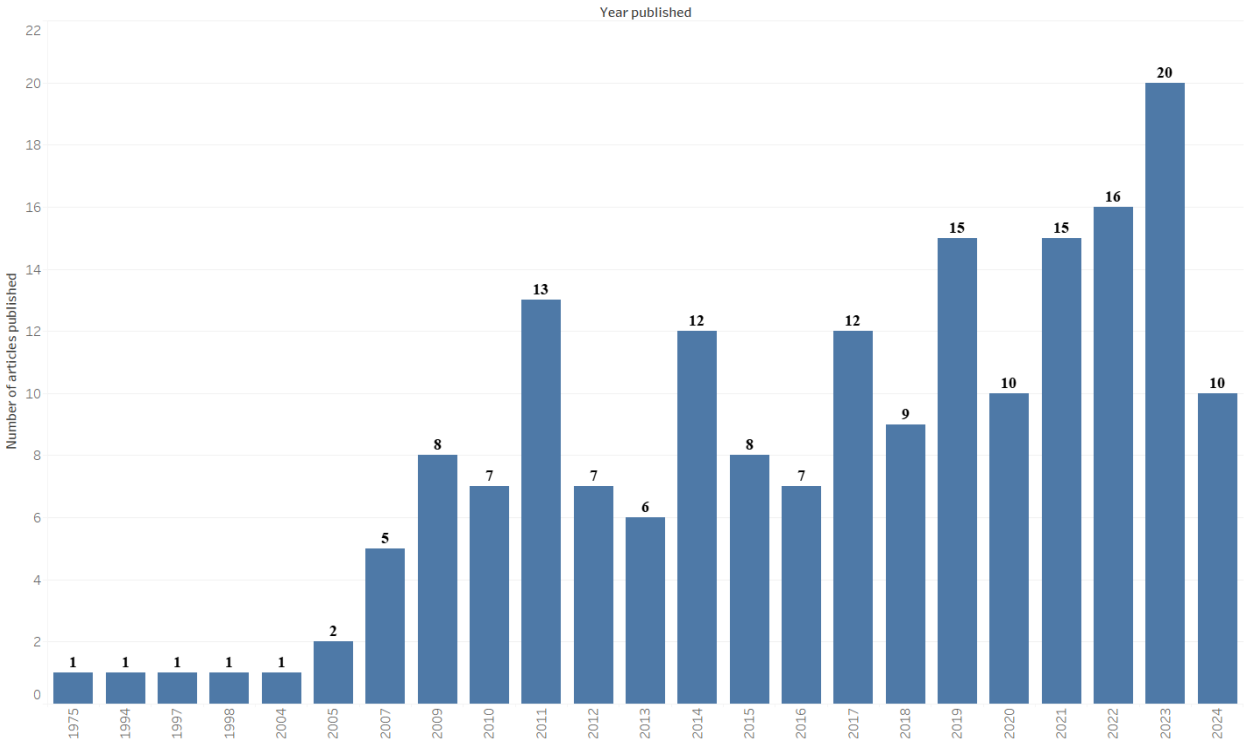

Figure 2: Annual scientific production

Figure 2 presents the annual progression of published articles in the research domain from 1975 to 2024, illustrating a significant shift in research activity over time, with a marked increase in publications reflecting the expanding relevance and integration across various academic sectors. As shown in Figure 2, the period from 1975 to 2004 was characterized by consistently low publication rates, with no more than one or two articles published annually. A significant shift occurred in 2005, marking the onset of a growth phase. This period saw a steady increase in publications, culminating in a peak of 13 articles in 2011. Between 2012 and 2019, the trend stabilized, with annual publication numbers fluctuating between 6 and 12 articles. The rise in output during this time indicates growing interest in and adoption of ML methodologies for addressing complex problems in social sciences and humanities. From 2020 onwards a second wave of growth emerged, with a rapid increase in the number of published articles. Peaks of 16 publications in 2021 and 20 in 2023 underscore a renewed and intensified focus on this area of research. This surge might be attributed to advancements in the field, heightened relevance of the research, or external factors driving increased interest, such as global challenges or societal demands. While 2024 shows a decline to 10 articles, this could be due to incomplete data for the year or an early indication of slowing growth. Overall, the chart illustrates two distinct periods of significant growth, one starting around 2005 and another following 2020. These phases mark crucial milestones in the development of ML applications in these fields, highlighting

shifts in research priorities, technological advancements, and the growing importance of computational methods in addressing complex, interdisciplinary research challenges. (see Figure 2)

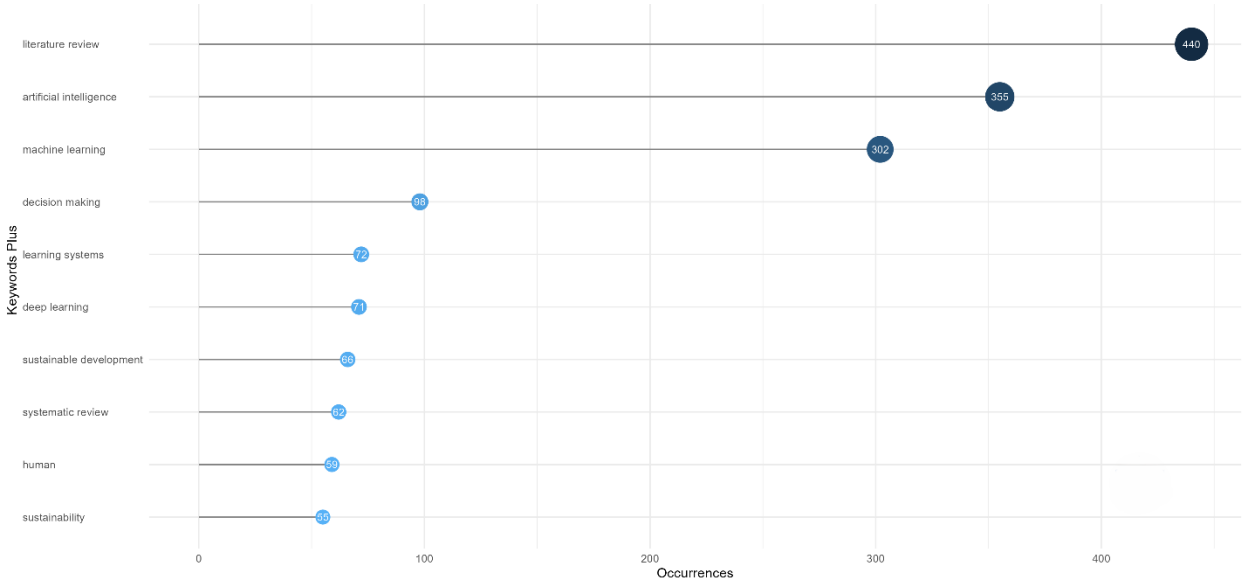

Figure 3: Keyword appearance frequency

Figure 3 presents a frequency analysis of keywords, shedding light on the dominant themes shaping research in this field of research. "Literature review," with its exceptionally high frequency (440 occurrences), indicates a mature body of work consolidating insights into the applications of ML in social sciences and humanities. The strong presence of "artificial intelligence" (355 occurrences) and "machine learning" (302 occurrences) reinforces the technological backbone of the research. Notably, "decision-making" emerges as a key focus (98 occurrences), suggesting that one of the primary applications of ML lies in providing tools to enhance decision-making processes in complex social systems. The appearance of terms like "learning systems" and "deep learning" underscore a focus on advancing sophisticated computational techniques, while "sustainable development" highlights the growing recognition of applying ML to address global challenges, particularly in sustainability and environmental contexts. This distribution of keywords demonstrates a balanced interest in both the technological progression and the societal applications of ML, aligning directly with the interdisciplinary goals of social sciences and humanities research.

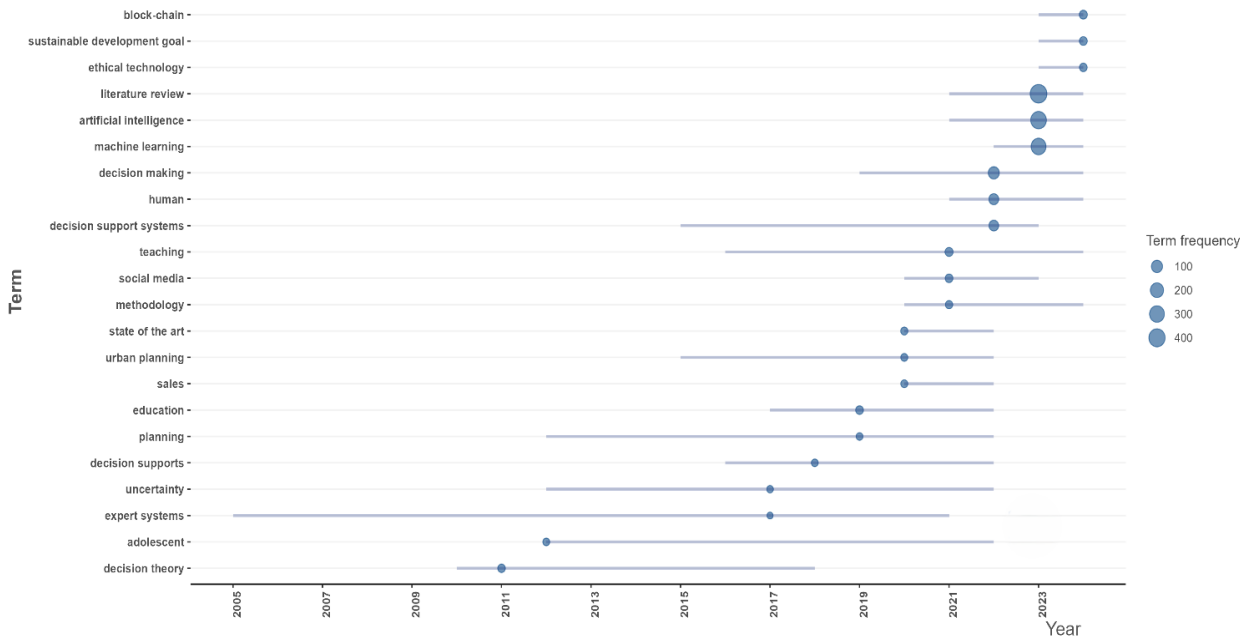

Figure 4: Trend topics

Figure 4 highlights the temporal trends of keywords illustrating the evolution of the research field and its responsiveness to emerging societal and technological challenges. The size of the bubble represents the term's frequency and its location in which the term appeared most, whilst the horizontal lines indicate the span of years during which each term has been prominent in the literature. Over time, there is a clear growth in the prominence of core terms such as *"machine learning"*, *"artificial intelligence"* and *"decision-making"*, reflecting their continued relevance in addressing complex social phenomena. In recent years, new themes such as "blockchain" and "sustainable development goals" have gained traction, suggesting an expanding focus on integrating ML with cutting-edge technologies and ethical considerations to solve modern global issues. Earlier trends, such as *"expert systems"* and *"decision theory"* which were prevalent in the early 2000s, have declined, signaling a shift away from traditional methods toward more versatile approaches like deep learning. Additionally, the rise of topics like *"social media"*, *"teaching"* and *"urban planning"*, indicates that ML is penetrating broader domains within the humanities, highlighting its versatility and interdisciplinary utility. This progression underscores the field's adaptability, as it evolves to address both technical innovation and its applications in shaping social and cultural systems (see Figure 4).

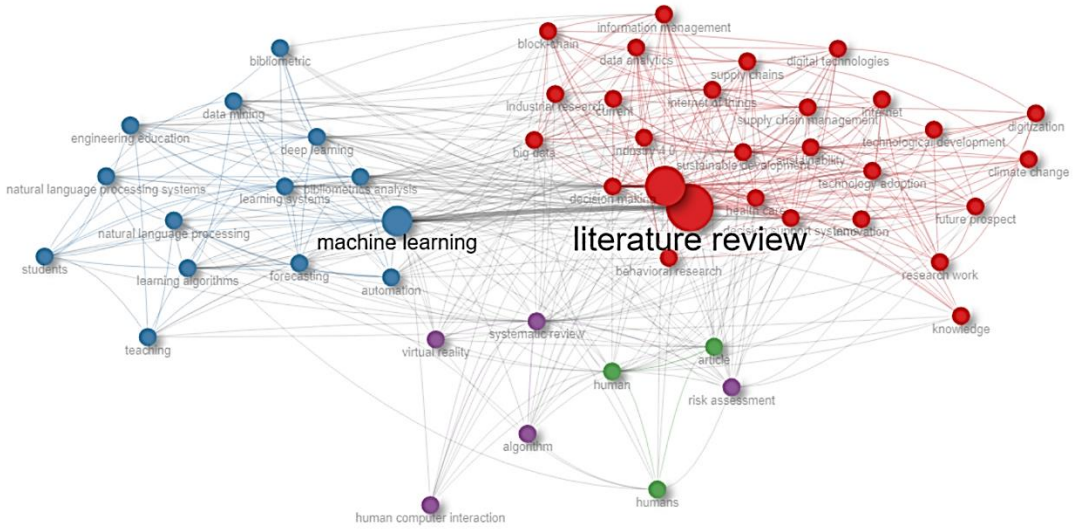

Figure 5: Co-occurrence network

Figure 5 is a co-occurrence network, mapping out the relationships between key terms associated with the research topic. The two largest nodes, *"literature review"* and *"machine learning"*, dominate the network, reflecting their centrality to this field. The network structure highlights two primary thematic clusters. The red cluster emphasizes the application of ML to societal and managerial challenges, including *"decision-making", "supply chain management"*, *"sustainability"* and *"technology adoption"*. This indicates a strong focus on using ML to address practical issues in governance, technology, and environmental systems. The blue cluster, on the other hand, is rooted in technical advancements, showcasing terms such as *"deep learning"*, *"natural language processing"* and *"forecasting"*. This suggests that the field also prioritizes developing and refining computational methods. Smaller, specialized clusters link ML to emerging topics such as *"human-computer interaction"* and *"risk assessment"*, illustrating its role in behavioral and interactive systems. The dense connectivity across clusters demonstrates a deep interdisciplinary synergy, with technical advancements feeding directly into societal and applied contexts, underscoring the value of ML as a bridge between computational sciences and humanities research.

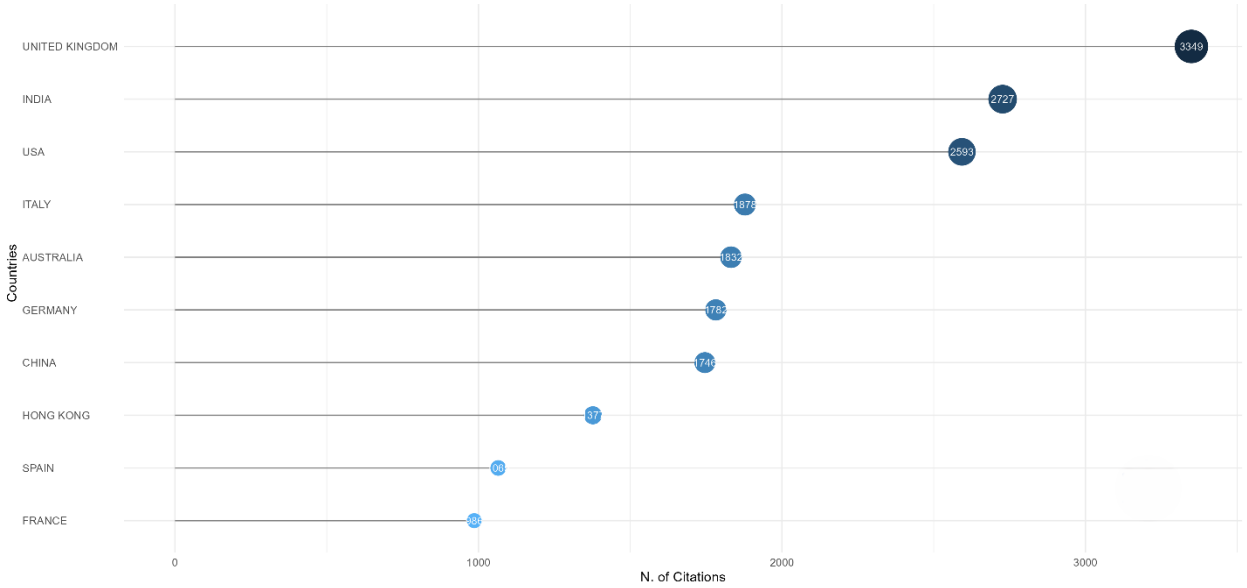

Figure 6: Most cited countries

Figure 6 is a graph that provides an overview of the global influence of research on ML in social sciences and humanities, by highlighting the countries whose research output has garnered the highest number of citations. As seen in Figure 6, the United Kingdom emerges as the leading contributor, with a significant margin of 3,349 citations, underscoring its prominent role in advancing interdisciplinary research that bridges ML with social science and humanities topics. The UK is followed by India, which is ranked second with 2,727 citations, highlighting its increasing impact in this domain. This may reflect its expanding research ecosystem, emphasizing education, natural language processing for local languages, and applications in governance and societal challenges. The USA follows closely with 2,593 citations, reflecting its technological leadership, particularly in the development of ML models and their application to diverse fields, including decision-making and behavioral studies in the humanities.

Italy, Australia, and Germany occupy the middle tier, with citations ranging from 1,700 to 1,800. Their contributions likely focus on specialized subfields, such as decision-support systems, ethical technology, and education-focused ML. These countries benefit from collaborative research networks and strong cross-disciplinary integration within their academic institutions. China and Hong Kong contribute substantially, with citations reflecting their growing investment in AI research and application in social sciences contexts, such as urban planning and behavioral analytics. However, the gap between these regions and the top contributors indicate a need for greater focus on interdisciplinary and global collaboration. Finally, countries such as Spain, France, and other European nations show a modest but noteworthy presence, highlighting their contributions to niche areas within this research field.

The dominance of the UK, India, and the USA points to a regional concentration of impactful research. However, the growing contributions from diverse nations signal an encouraging trend toward global collaboration in exploring how ML can address pressing social and cultural questions. This distribution emphasizes the field's interdisciplinary and international nature, with research output shaped by varying local contexts, priorities, and resources (see Figure 6).

Conclusions

The application of ML methods in social sciences and humanities has seen notable growth and diversification, offering promising avenues for advancing our understanding of complex societal and cultural phenomena. This bibliometric review has highlighted key trends in the integration of ML techniques across various disciplines, such as *"machine learning"*, *"artificial intelligence"* and *"decision-making"*. While the overall research output has steadily increased, particularly in the last two decades, it remains a relatively young field, with a significant portion of publications emerging in recent years. This points to the growing interest and potential of ML in social sciences and humanities, driven by technological advancements and the increasing availability of large, complex datasets. The data presented also suggests a broad thematic scope, with key areas of focus such as decision-making, sustainability, and social media, reflecting the diverse applications of ML in tackling societal challenges.

The bibliometric analysis further reveals a strong collaborative dimension to this research, with a substantial percentage of publications involving international co-authorship. This underscores the interdisciplinary and global nature of the field, where insights from diverse academic and cultural contexts contribute to the development of ML methods within this domain. The geographic distribution of research highlights the leading roles played by countries like the United Kingdom, India, and the United States, though emerging contributions from other regions also point to an expanding global interest in this area.

While these trends highlight the potential of using ML in social sciences and humanities, there are limitations that should be noted. The predominance of English-language publications within the field presents a potential source of linguistic and regional bias, which may result in the underrepresentation or exclusion of significant scholarly contributions published in other languages. This limitation restricts the global inclusivity of the research landscape, particularly in areas where academic output is not primarily produced in English. Additionally, despite the promise of ML techniques, the field faces critical challenges, notably the interpretability of complex models and the quality of underlying data. These issues raise concerns about the transparency and reliability of findings derived from ML applications. Ethical considerations also pose significant barriers, with algorithmic bias, data privacy concerns, and the potential

misuse of AI-driven insights emerging as critical challenges, particularly in socially sensitive or high-stakes contexts. Addressing these barriers is imperative to ensure that ML methodologies are deployed responsibly and equitably within the social sciences and humanities.

Lastly, while the field is still evolving, the trends observed suggest that its integration into these disciplines holds great promise. As ML methodologies continue to advance, they may offer deeper insights into social, cultural, and behavioral phenomena. Future research could further explore emerging themes and refine methods, while also promoting greater cross-disciplinary collaboration. This will help to further advance the role of ML in addressing complex societal and cultural challenges, reinforcing its potential to transform research in these disciplines.

References

Alloghani, M., Al-Jumeily, D., Mustafina, J., Hussain, A., & Aljaaf, A. J. (2019). A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science. *Unsupervised and Semi-supervised Learning*, 3–21. https://doi.org/10.1007/978-3-030-22475-2_1

Aria, M., & Cuccurullo, C. (2017). bibliometrix : An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, *11*(4), 959–975. https://doi.org/10.1016/j.joi.2017.08.007

Baas, J., Schotten, M., Plume, A., Côté, G., & Karimi, R. (2020). Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. *Quantitative Science Studies*, *1*(1), 377–386. https://doi.org/10.1162/qss_a_00019

Bitzenis, A., & Koutsoupias, N. (2023A). Foreign Direct Investment And Sustainability: A Visualized Bibliometric Review. *South-Eastern Europe Journal of Economics*, *1*, 41–52. https://doi.org/10.17605/osf.io/pcaxg

Bitzenis, A., Koutsoupias, N., & Nosios, M. (2025). Artificial intelligence and machine learning in production efficiency enhancement and sustainable development: a comprehensive bibliometric review. *Frontiers in Sustainability*, *5*. https://doi.org/10.3389/frsus.2024.1508647

Bitzenis, A., & Koutsoupias, N. (2023B). Visualizing Economics and Business in the field of Education: A Quantitative Analysis of Scholarly Literature. *Global Business & Economics Anthology*, *Volumes I&II*(December 2023). https://doi.org/10.47341/gbea.23126

Bitzenis, A., & Koutsoupias, N. (2024). Big Data in Economics Research. In N. Tsounis & A. Vlachvei (Eds.), *Applied Economic Research and Trends* (pp. 1063–1072). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-49105-4_6

Bitzenis, A., Koutsoupias, N., & Boutsiouki, S. (2023). Business Research and Data Mining: a Bibliometric Analysis. *2023 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, 1–6. https://doi.org/10.1109/iceccme57830.2023.10252699

Castellano, G., & Vessio, G. (2021). Deep learning approaches to pattern extraction and recognition in paintings and drawings: an overview. *Neural Computing and Applications*, *33*(19), 12263–12282. https://doi.org/10.1007/s00521-021-05893-z

Chapinal-Heras, D., & Díaz-Sánchez, C. (2023). A review of AI applications in Human Sciences research. *Digital Applications in Archaeology and Cultural Heritage*, *30*, e00288. https://doi.org/10.1016/j.daach.2023.e00288

Chauhan, P., Sharma, N., & Sikka, G. (2020). The emergence of social media data and sentiment analysis in election prediction. *Journal of Ambient Intelligence and Humanized Computing*, *12*(2), 2601–2627. https://doi.org/10.1007/s12652-020-02423-y

Chowdhary, K. R. (2020). Natural language processing. In *Springer eBooks* (pp. 603–649). https://doi.org/10.1007/978-81-322-3972-7_19

Grimmer, J., Roberts, M. E., & Stewart, B. M. (2021). Machine Learning for Social Science: an agnostic approach. *Annual Review of Political Science*, *24*(1), 395–419. https://doi.org/10.1146/annurev-polisci-053119-015921

Hagen, L. (2018). Content analysis of e-petitions with topic modeling: How to train and evaluate LDA models? *Information Processing & Management*, *54*(6), 1292–1307. https://doi.org/10.1016/j.ipm.2018.05.006

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, *31*(3), 685–695. https://doi.org/10.1007/s12525-021-00475-2

Koutsoupias, N. (2024). Multiple Correspondence Analysis in Social Sciences and Humanities Research: A Longitudinal Mapping. *Data Analysis Bulletin*, *20*(1), 59–77. https://doi.org/10.17605/osf.io/693dr

Koutsoupias, N., & Papadimitriou, I. (2020). Multivariate Data Analysis in R. *Data* Analysis *Bulletin*, *20*(2020), 68–83. (in Greek) https://doi.org/10.5281/zenodo.7557364

Lundberg, I., Brand, J. E., & Jeon, N. (2022). Researcher reasoning meets computational capacity: Machine learning for social science. *Social Science Research*, *108*, 102807. https://doi.org/10.1016/j.ssresearch.2022.102807

Nadkarni, P. M., Ohno-Machado, L., & Chapman, W. W. (2011). Natural language processing: an introduction. *Journal of the American Medical Informatics Association*, *18*(5), 544–551. https://doi.org/10.1136/amiajnl-2011-000464

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., . . . Moher, D. (2021). The PRISMA 2020

statement: an updated guideline for reporting systematic reviews. *BMJ*, n71. https://doi.org/10.1136/bmj.n71

Polak, A., Kelman, T., Murray, P., Marshall, S., Stothard, D. J., Eastaugh, N., & Eastaugh, F. (2017). Hyperspectral imaging combined with data classification techniques as an aid for artwork authentication. *Journal of Cultural Heritage*, *26*, 1–11. https://doi.org/10.1016/j.culher.2017.01.013

Quinn, K. M., Monroe, B. L., Colaresi, M., Crespin, M. H., & Radev, D. R. (2009). How to Analyze Political Attention with Minimal Assumptions and Costs. *American Journal of Political Science*, *54*(1), 209–228. https://doi.org/10.1111/j.1540-5907.2009.00427.x

Rahal, C., Verhagen, M., & Kirk, D. (2022). The rise of machine learning in the academic social sciences. *AI & Society*, *39*(2), 799–801. https://doi.org/10.1007/s00146-022-01540-w

Santos, I., Castro, L., Rodriguez-Fernandez, N., Torrente-Patiño, Á., & Carballal, A. (2021). Artificial Neural Networks and Deep Learning in the Visual Arts: a review. *Neural Computing and Applications*, *33*(1), 121–157. https://doi.org/10.1007/s00521-020-05565-4

Sarker, I. H. (2021). Machine learning: algorithms, Real-World applications and research directions. *SN Computer Science*, *2*(3). https://doi.org/10.1007/s42979-021-00592-x

Thomos, K., Bitzenis, A., & Koutsoupias, N. (2023). Credit Rating in Business and Economics Research: Europe (2000-2022). *Global Business & Economics Anthology*, *Volumes I&II*(December 2023). https://doi.org/10.47341/gbea.23128

Usama, M., Qadir, J., Raza, A., Arif, H., Yau, K. A., Elkhatib, Y., Hussain, A., & Al-Fuqaha, A. (2019). Unsupervised machine learning for networking: techniques, applications and research challenges. *IEEE Access*, *7*, 65579–65615. https://doi.org/10.1109/access.2019.2916648

Van Engelen, J. E., & Hoos, H. H. (2019). A survey on semi-supervised learning. *Machine Learning*, *109*(2), 373–440. https://doi.org/10.1007/s10994-019-05855-6

Yenduri, G., Ramalingam, M., Selvi, G. C., Supriya, Y., Srivastava, G., Maddikunta, P. K. R., Raj, G. D., Jhaveri, R. H., Prabadevi, B., Wang, W., Vasilakos, A. V., & Gadekallu, T. R. (2024). GPT (Generative Pre-Trained Transformer) – a comprehensive review on enabling technologies, potential applications, emerging challenges, and future directions. *IEEE Access*, *12*, 54608–54649. https://doi.org/10.1109/access.2024.3389497

Zhang, H., & Peng, Y. (2022). Image Clustering: an unsupervised approach to categorize visual data in social science research. *Sociological Methods & Research*, 004912412210826. https://doi.org/10.1177/00491241221082603

Zhu, X., & Goldberg, A. B. (2009). Introduction to Semi-Supervised Learning. In *Synthesis lectures on artificial intelligence and machine learning*. https://doi.org/10.1007/978-3-031-01548-9