Information Research - Vol. 30 No. iConf (2025)

Review articles as windows into Knowledge accumulation: the case of AI research

Siyuan Peng, Lei Hu, Jingrui Hou, Youqing Xia, and Ping Wang

DOI: https://doi.org/10.47989/ir30iConf47224

Abstract

Introduction. Review articles are essential in evolving scholarly information systems but have been underexplored in scientometrics. This paper aims to expand scientometric research on review articles, focusing on their role in understanding knowledge accumulation within specific domains.

Method. This study collected 4,315 review articles on artificial intelligence (AI). Using keyword frequency analysis and the Task-Technology Fit (TTF) model, the articles were classified into three categories: task-oriented, technology-oriented, and application-oriented.

Analysis. The temporal distribution of the review articles, the age distribution of their cited references, and the updating characteristics of references cited in review articles were analysed to provide preliminary insights into the evolution, dynamism, and updating patterns of knowledge in AI.

Results. The results show a marked increase in the publication frequency of review articles, especially over the past five years, with the application domain exhibiting the highest growth rate. Over half of the references cited in review articles across all domains are from the past five years. Additionally, older references are cited more frequently in newer reviews than more recent ones.

Conclusion. This study can be seen as an expansion of scientometric research based on review articles and highlights several intriguing research questions for exploration within this field.

Introduction

Recent decades have witnessed exponential growth in the volume of new scientific knowledge (Park et al., 2023). This incredible rate of scientific productivity leads to information overload, making it hard to keep up with state-of-the-art and assess the collective evidence in a particular research area (Snyder, 2019). By integrating diverse knowledge streams in a particular area, the genre of the review article could provide a solution. Review articles are critical evaluations of prior studies that have already been published (Palmatier et al., 2018). They are imperative for consolidating knowledge, synthesizing empirical findings, developing novel theories, comprehending the depth and breadth of research phenomena, and identifying unexplored areas for future investigation (Farrukh & Sajjad, 2023). Consequently, review articles have gradually gained recognition as important objects of scholarly research in the field of scientometrics (Blümel & Schniedermann, 2020). Most studies have focused on the structural features of review articles, such as the number of co-authors, pages, and patterns of institutional collaboration (Barrios et al., 2013). Other studies have explored the impact of review articles, including their influence on journal rankings and metrics (Ketcham & Crawford, 2007; Fassin, 2021), subsequent citations of cited papers (Lachance et al., 2014), citation-based analysis (Ho et al., 2017), and the measurement of individual research output (Miranda & Garcia-Carpintero, 2018). However, scientometric studies on review articles have generally been less often executed for more exploratory purposes (Blümel & Schniedermann, 2020), particularly when it comes to understanding the phenomenon of knowledge accumulation.

Review articles hold significant potential as a genre for exploring knowledge accumulation, defined as the collective body of knowledge gathered over time (Chandra & Dong, 2018). This potential arises from the fact that scientific knowledge in a particular field is often published in research articles, while review articles serve as a means of integrating and synthesizing these research articles (Praus, 2023; Snyder, 2019). Consequently, this paper aims to broaden the research scope regarding review articles in scientometrics by examining their role in elucidating the phenomenon of knowledge accumulation. We identify three potential research directions that warrant further exploration: The first is ***temporal distribution of review articles***. As review articles synthesize and present the current state of knowledge, their temporal distribution can serve as an indicator of the field's evolutionary pace and the periods of significant theoretical or empirical advancements (Patriotta, 2020). Analyzing the frequency and timing of review articles could reveal patterns of knowledge maturation and periods of heightened scholarly interest or innovation. The second is ***age distribution of references cited in review articles***, which investigates the age profile of references cited in review articles. Since the age of cited references is indicative of the rate of growth of scientific knowledge (Stacey, 2021), this paper proposes that the age distribution of references cited in review articles can, to some extent, reveal the knowledge accumulation and dynamism in the field. Finally, the third is ***updating characteristics of references cited in review articles***, which explores the patterns of reference updating in review articles. Review articles are regarded as an expansion and update of a previously published article on this topic (Torraco, 2016). Based on this viewpoint, this paper reckons that examining the updating patterns of references in review articles can further help reveal the knowledge accumulation speed and pattern.

As a pioneering study, this study provides some preliminary phenomena and trend disclosures within the three potential research directions to enhance our understanding of the accumulation of domain knowledge, and to expand research on review articles in scientometrics. Moreover, given the accelerating pace of knowledge accumulation within the field of artificial intelligence (AI) research and the growing number of review articles providing a comprehensive overview of this field (Leeming, 2021), we have chosen the review articles in the AI domain to conduct our analysis.

Data and preliminary investigation

Data preparation

Our research focused on the '*computer science, artificial intelligence*' category from the Web of Science (WoS), considering literature published up to the end of 2023. We initially obtained 5,991 results based on the review document type assigned by WoS. However, due to potential misclassification of review articles in WoS (Blümel & Schniedermann, 2020), this study underwent automatic preprocessing and manual review to exclude incomplete or non-review articles. As a result, we compiled a final set of 4,315 AI review articles.

To identify major research themes within this body of literature, we performed a keyword frequency analysis, allowing us to systematically uncover prevailing topics and trends discussed in the AI review articles. These findings were then structured with the task-technology fit (TTF) model (Goodhue & Thompson, 1995), which categorizes research topics based on their focus on either tasks, technology, or applications (i.e., the utilization of TTF). Accordingly, we classified the identified topics into three broad categories:

- **Task-oriented reviews:** these reviews emphasize AI tasks that are central to various domains. Examples include tasks such as '*classification*', '*pattern recognition*', and '*clustering*'. These reviews often concentrate on algorithmic methodologies and performance evaluation specific to task-related challenges.
- **Technology-oriented reviews:** in this category, the reviews focus on specific AI technologies, such as '*feature extraction*', '*convolutional neural networks (CNNs)*', and '*reinforcement learning*'. These reviews aim to provide a comprehensive understanding of cutting-edge AI methodologies and their theoretical underpinnings.
- **Application-oriented reviews:** this category explores the application of AI techniques to real-world challenges, often across diverse sectors, such as '*COVID-19 diagnostics*', '*medical imaging*' and '*autonomous systems*.' These reviews assess how AI technology, when aligned with the TTF model, can drive innovation, and provide solutions to complex, domain-specific problems.

Temporal distribution characteristics of review articles

This study mapped the temporal trends of AI review articles across three domains (see Figure 1), observing a general increase in publication frequency over time, especially notable since 2019. This aligns with the rapid growth of scientific literature noted by Bornmann et al. (2021). Then, we introduced '*density*' as a measure for the frequency of articles per year, identifying the '*COVID-19*' topic within the application domain as having the highest density (19.5) despite its brief existence. Other topics like '*internet of things (IoT)*', '*generative adversarial networks (GAN)*', '*blockchain,*' and '*federated learning*' have also shown high densities and an upward knowledge accumulation trend. Moreover, traditional task domain topics like '*classification*' and '*object detection*' demonstrate high densities with stable development.

As shown in Table 1, the comparative analysis of the past five years and five years prior revealed that the task domain typically showed higher article densities than the technology domain, which outpaced the application domain. The last five years witnessed significant density increases across all domains. Annual growth rates across these domains have converged and stabilized post-2019, with the application domain experiencing a marked growth rate increase after 2019, while the task and technology domains maintained consistent growth rates before and after 2019, stabilizing thereafter.

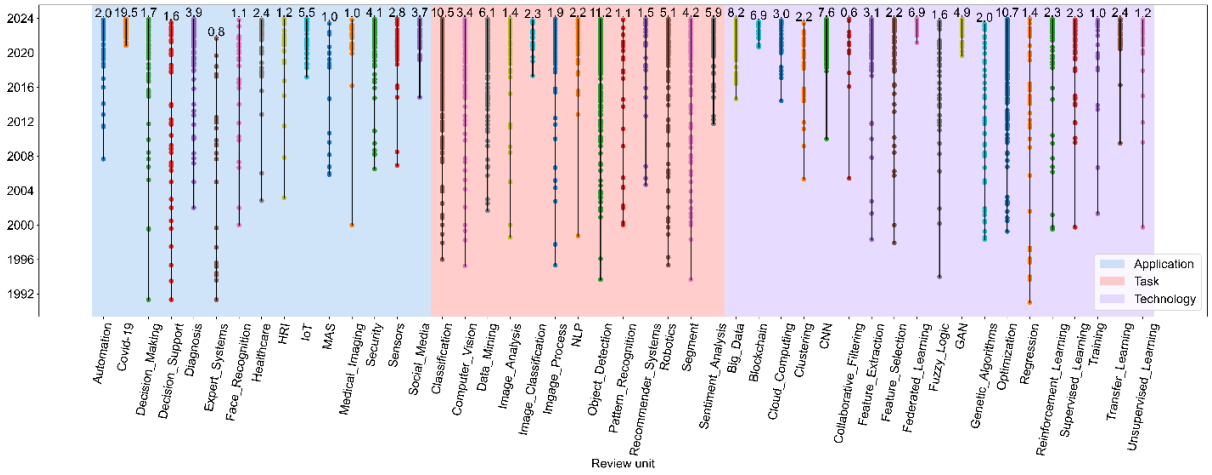

Figure 1. The temporal distribution of review article publications. The vertical span of each line represents the time span of publication for each review unit. Each point on the line segment represents a review paper emerging at a particular time point.

Review domain	Density (Overall)	Density (After 2019)	Density (Before 2019)	AAGR (After 2019)	AAGR (Before 2019)
Application	19.67	97.52	6.18	0.30±0.04	0.17±0.15
Task	47.21	201.85	17.27	0.27±0.03	0.27±0.30
Technology	32.27	156.68	10.46	0.29±0.03	0.25±0.24

Table 1. Density and average annual growth rate (AAGR) in different review domains

Age distribution characteristics of references cited in review articles

This study delved into the age distribution of references cited in AI review articles across three domains, defining '*reference lag years*' as the time difference between the publication year of the review and its cited literature. Figure 2 revealed a predominant lag of 2 to 3 years in task and technology domains, with the application domain occasionally extending to 4 years. Post-peak, all domains showed a rapid decline in citation numbers with increasing lag time.

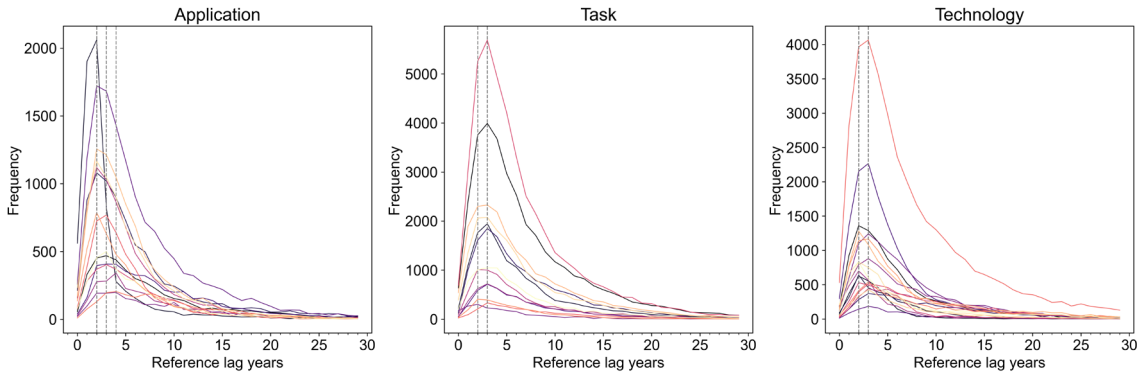

Figure 2. The distribution pattern of reference lag years in different review domains

A closer examination of reference lag years within a five-year span showed distinct patterns as shown in Table 2, the application domain's citations peak in the second year (0.136), the task domain in the third year (0.123), and the technology domain equally in the second and third years (0.123). The application domain demonstrated notably higher citation frequencies in the first two years compared to task and technology, suggesting quicker incorporation of recent literature.

Across all three domains, citations with a lag of up to five years constituted over half of the total, with the highest proportion in the application domain.

Review domain	Year 1	Year 2	Year 3	Year 4	Year 5	Sum
Application	0.125	0.136	0.124	0.106	0.084	0.575
Task	0.086	0.114	0.123	0.108	0.090	0.521
Technology	0.095	0.123	0.123	0.105	0.085	0.531

Table 2. Proportion of citations within recent five years for each review domain

In addition to analysing the relative reference lag years, this paper also examined the overall citation time distribution of review articles across different years. As shown in Figure 3, recent review articles span longer historical periods than older reviews. Specifically, literature time spans starting years shifted significantly backward from 1971, 1962, and 1958 in the year 2000 to before 1950 by 2022 and 2023 for the application, task, and technology domains respectively. Early 2000s and recent reviews showed more concentrated interquartile ranges, with wider ranges and noticeable fluctuations in other periods.

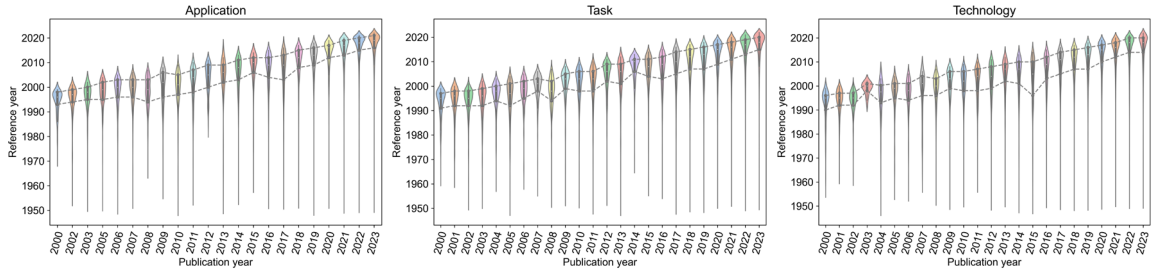

Figure 3. The distribution of publication years of references in review articles for different publication years

Updating characteristics of references cited in review articles

In examining the dynamics of references cited in AI review articles, we define references as '*knowledge units*' with temporal attributes, together forming a '*knowledge set*' over time. Our assessment of knowledge accumulation for each year's new references is threefold: (1) **accumulated references**: the cumulative total of references cited in all review articles up to a given year, indicating the knowledge set's size to that date. (2) **New references:** references cited for the first time in that year, expanding the existing knowledge set. (3) Types of new references: distinguished as either **recent** (published within the past year) or **previous** (published more than a year ago).

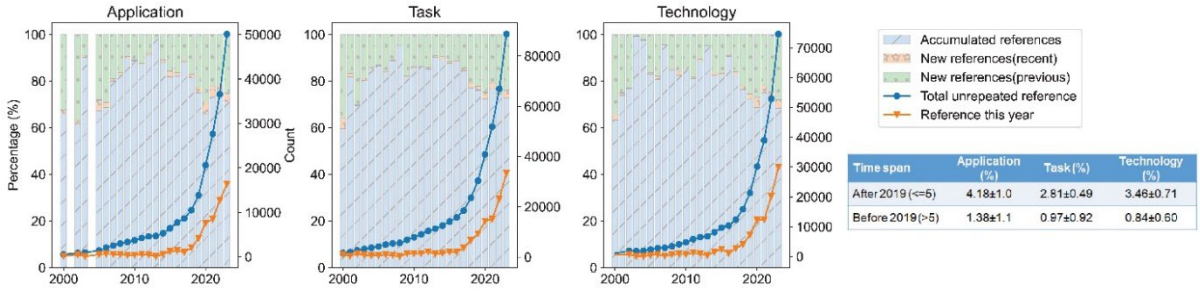

Figure 4. The annual changes in accumulated references and new references and proportion of new references (recent) in each review domain

As shown in Figure 4, the ratio of accumulated to new references indicates a predominance of accumulated references annually. Within new references, those published over a year ago consistently outnumber recent publications. Nevertheless, the share of recent references within

new citations has been rising lately. Quantitative analysis shows that the proportion of recent references in new citations post-2019 has increased across all domains.

Discussion and conclusion

Key findings

This study investigates the dynamics of knowledge accumulation in scientific fields by analysing review articles across multiple dimensions. Using AI research as a case study, we uncover patterns that are likely applicable to other evolving domains. Our analysis reveals several noteworthy findings and insights. Firstly, with regard to the temporal distribution characteristics, there has been a significant surge in their publication frequency, particularly in the past five years. This surge indicates that the pace of knowledge accumulation in AI is rapidly accelerating, aligning with the explosive growth of AI publications (Leeming, 2021). The domain of application has experienced the highest growth rate in terms of review articles, which can be attributed to the emergence of AI as a powerful tool for addressing complex interdisciplinary problems across various research fields (Zhuang et al., 2020). Secondly, in terms of the age distribution characteristics of references, more than half of the references cited in review articles across all domains fall within a five-year timeframe, with the application domain exhibiting the highest proportion. This trend signifies a notable acceleration in knowledge accumulation across the domains, especially in the application domain over the past five years, which aligns with the observed temporal distribution patterns. Furthermore, the analysis demonstrates that recent review articles encompass a broader timeframe of references. This finding is consistent with the nature of scientific knowledge accumulation, which involves the growth of the available stock of knowledge over time (Petrovich, 2018). Thirdly, in relation to the updating patterns of cited references, the number of older references cited in newer reviews consistently exceeds that of more recent references. This observation confirms the existence of a citation lag and suggests a pattern of delayed citation across review articles. Moreover, the increasing proportion of recent references among new citations in recent years indicates a heightened pace of knowledge accumulation in the three domains under examination.

Despite these insightful findings, the limitations should be acknowledged. Our analysis focuses on the AI research field. The findings may not generalize to other disciplines. Future research can explore different fields to help further clarify the role of review articles in understanding the knowledge accumulation phenomenon.

Conclusion and future work

Overall, this study presents some findings on knowledge accumulation in the field of AI based on a quantitative analysis of review articles. Based on these findings, this paper identifies several intriguing research questions that can be explored in the context of review-based scientometric studies.

The convergent growth of review articles and research articles. This question arises from the intuitive notion that review articles serve as summaries and evaluations of research articles (Praus, 2023) and the observed increase in the publication frequency of review articles, alongside a similar trend in research articles (Leeming, 2021). A promising avenue for further investigation is to delineate the quantitative relationship between the expansion of review articles and the overall growth of research articles.

When will a new review article emerge in a particular knowledge domain? Prior research suggests that the presence of review articles reflects the maturity of scientific fields (Bastide, 1989) and may manifest even in the early stages of their formation (Blümel, 2016), a notion supported by our observations (see Figure 1). Therefore, examining the predictors of when a new review paper will surface, based on the existing structure of research articles, could yield valuable insights.

Rethinking knowledge accumulation via both the citation pattern and semantic evolution of review articles. Leveraging review articles to explore the evolution of domain knowledge can provide valuable insights into the dynamic changes in knowledge development. While current analyses rely primarily on references and citation lags, future research should incorporate textual data and semantic evolution for deeper insights into patterns of knowledge accumulation.

In conclusion, this study not only enhances our understanding of knowledge accumulation in AI but also raises broader questions that are central to scientometrics and information studies. By advancing methods for tracking and analyzing review articles, this work establishes a foundation for future research into how scientific knowledge evolves across diverse disciplines. It suggests that review article analysis can be further developed to capture the dynamics of research development more effectively. Moreover, expanding the scope to include additional disciplines will enable future studies to refine existing models of knowledge diffusion and scholarly communication, offering deeper insights into the processes of knowledge creation and dissemination across scientific domains.

Acknowledgements

This work was supported by the National Social Science Fund of China (Grants No. 24&ZD186).

About the authors

Siyuan Peng is a Doctoral student at the School of Information Management, Wuhan University. Her research interests lie in scientometrics and knowledge mining. She can be contacted at siyuanpeng@whu.edu.cn.

Lei Hu is a master's student at the School of Microelectronics, Hubei University. His current research interests include data mining, deep learning, intelligent document understanding, and large language models. He can be contacted at hulei org@126.com.

Jingrui Hou is the corresponding author of this paper. He is a Postdoctoral researcher at the School of Information Management, Wuhan University, with research interests in natural language processing and knowledge evolution analysis. He can be contacted at jhou.research@outlook.com.

Youqing Xia received his master's degree from The University of Queensland, Australia. His research interests include the interdisciplinary application of archival studies and the processing of complex data. He can be contacted at xiayq519@gmail.com.

Ping Wang is a Professor at the School of Information Management, Wuhan University. He received his Ph.D. from Wuhan University, and his research interests include human-computer interaction and natural language processing. He can be contacted at wangping@whu.edu.cn.

References

Barrios, M., Guilera, G., & Gómez-Benito, J. (2013). Impact and structural features of meta-analytical studies, standard articles, and reviews in psychology: Similarities and differences. Journal of Informetrics, 7(2), 478-486. https://doi.org/10.1016/j.joi.2013.01.012

Bastide, F., Courtial, J., & Callon, M. (1989). The use of review articles in the analysis of a research area. Scientometrics, 15(5-6), 535-562. https://doi.org/10.1007/bf02017070

Blümel, C., & Schniedermann, A. (2020). Studying review articles in scientometrics and beyond: a research agenda. Scientometrics, 124(1), 711-728. https://doi.org/10.1007/s11192-020-03431-7

Blümel, C. (2016). Enrolling the toggle switch: Visionary claims and the capability of modeling objects in the disciplinary formation of synthetic biology. NanoEthics, 10, 269-287. https://doi.org/10.1007/s11569-016-0276-z

Bornmann, L., Haunschild, R., & Mutz, R. (2021). Growth rates of modern science: a latent piecewise growth curve approach to model publication numbers from established and new literature databases. Humanities and Social Sciences Communications, 8(1), 1-15. https://doi.org/10.1057/s41599-021-00903-w

Chandra, P., & Dong, A. (2018). The relation between knowledge accumulation and technical value in interdisciplinary technologies. Technological Forecasting and Social Change, 128, 235-244. https://doi.org/10.1016/j.techfore.2017.12.006

Farrukh, A., & Sajjad, A. (2023). A critical review of literature review methodologies. Advancing Methodologies of Conducting Literature Review in Management Domain, 2, 103-123. https://doi.org/10.1108/S2754-586520230000002006

Fassin, Y. (2021). The impact of review articles in management and economics journal rankings and metrics. Scientometrics, 126(12), 9623-9632. https://doi.org/10.1007/s11192-021-04160-1

Goodhue, D. L., & Thompson, R. L. (1995). Task-technology fit and individual performance. MIS quarterly, 213-236. https://doi.org/10.2307/249689

Ho, M. H. C., Liu, J. S., & Chang, K. C. T. (2017). To include or not: the role of review papers in citation-based analysis. Scientometrics, 110, 65-76. https://doi.org/10.1007/s11192-016-2158-0

Ketcham, C.M., & Crawford, J.M. (2007). The impact of review articles. Laboratory Investigation, 87, 1174-1185. https://doi.org/10.1038/labinvest.3700688

Lachance, C., Poirier, S., & Larivière, V. (2014). The kiss of death? The effect of being cited in a review on subsequent citations. Journal of the Association for Information Science and Technology, 65(7), 1501-1505. https://doi.org/10.1002/asi.23166

Miranda, R., & Garcia-Carpintero, E. (2018). Overcitation and overrepresentation of review papers in the most cited papers. Journal of Informetrics, 12(4), 1015-1030. https://doi.org/10.1016/j.joi.2018.08.006

Leeming, J. (2021). AI and its growing foothold in the natural sciences. Nature, 598, S5. *https://doi.org/10.1038/d41586-021-02762-6*

Palmatier, R. W., Houston, M. B., & Hulland, J. (2018). Review articles: Purpose, process, and structure. Journal of the Academy of Marketing Science, 46, 1-5. https://doi.org/10.1007/s11747-017-0563-4

Park, M., Leahey, E., & Funk, R. J. (2023). Papers and patents are becoming less disruptive over time. Nature, 613(7942), 138-144. https://doi.org/10.1038/s41586-022-05543-x

Patriotta, G. (2020). Writing impactful review articles. Journal of Management Studies, 57(6), 1272-1276. https://doi.org/10.1111/joms.12608

Petrovich, E. (2018). Accumulation of knowledge in para-scientific areas: The case of analytic philosophy. Scientometrics, 116(2), 1123-1151. https://doi.org/10.1007/s11192-018-2796-5

Praus, P. (2023). Empirical relationship between the number of review and research articles. Scientometrics, 128(4), 2201-2209. https://doi.org/10.1007/s11192-023-04654-0

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. Journal of business research, 104, 333-339. https://doi.org/10.1016/j.jbusres.2019.07.039

Stacey, A. G. (2021). Ages of cited references and growth of scientific knowledge: an explication of the gamma distribution in business and management disciplines. Scientometrics, 126(1), 619-640. https://doi.org/10.1007/s11192-020-03761-6

Torraco, R. J. (2016). Writing integrative literature reviews: Using the past and present to explore the future. Human resource development review, 15(4), 404-428. https://doi.org/10.1177/1534484316671606

Zhuang, Y., Cai, M., Li, X., Luo, X., Yang, Q., & Wu, F. (2020). The next breakthroughs of artificial intelligence: The interdisciplinary nature of AI. Engineering, 6(3), 245-247. https://doi.org/10.1016/j.eng.2020.01.009