

A Bibliometric Analysis of the Influence of Textile Industrialization During the Industrial Revolution on the Subject Matter of Western Music Composition

Yue Fan

School of Music, Shanxi College of Applied Science and Technology, Taiyuan 030062, Shanxi, China

Corresponding author: Yue Fan (e-mail: fanyue424@163.com)

ABSTRACT Whether textile industrialization during the Industrial Revolution transformed the subject matter of Western music composition has long depended mainly on qualitative interpretation and requires systematic empirical verification. This paper combines bibliometrics, historical text mining, and topic modeling to construct a corpus of music literature and a vocabulary of industrialization features from 1760 to 1900. A dynamic time warping algorithm is first used to identify temporal relationships in term diffusion. A Latent Dirichlet Allocation topic model is then applied to measure the development of themes such as machine imagery and factory scenes. Social network analysis is further integrated to identify key creative groups and the diffusion structure of industrialized vocabulary. The findings show that the penetration rate of industrialization terms increased by 55.6% in the 1840s, while themes related to machine imagery and factory scenes increased by more than 52% in the same period. Composers such as Liszt show increased network centrality, and terms such as factory and worker spread more widely than specialized technical terms. The study provides a replicable quantitative paradigm for analyzing technology-art interactions.

INDEX TERMS industrial revolution, textile industrialization, musical composition themes, bibliometric analysis, topic modeling

I. INTRODUCTION

GIVEN its profound significance in human history, the question of what impact the technological changes and social reorganization of the Industrial Revolution had on the arts has been a central topic in cultural history research [1,2]. The technological advancements (e.g., spinning jenny, mule loom, and power loom), as well as the organizational innovations of the textile industry, as the first industry of the Industrial Revolution, not only changed the economic structure, but also influenced the aesthetic concepts and narrative styles of modern composers [3]. However, this technological-art impact spanning from the Industrial Revolution has long been described qualitatively and studied through case analyses by music historians—either studying the development of the life history and compositional styles of individual composers or confined to the textual interpretation of “machine imagery” in a certain piece—without providing systematic empirical analyses based on documentary data on a large scale.

Existing research mainly revolves around the socio-cultural impact of the Industrial Revolution, often employing art historical case studies or sociological qualitative

descriptions. For example, Zhou took the Zhongshan suit as an example to explore the changes in the aesthetics of modern Chinese clothing and the significance of its “straight line” design language in the transition of modernity, and provided inspiration for the innovative transformation of traditional clothing elements in contemporary design [4]. Mokhtar et al. took Malaysia as the background to explore the dynamic framework of art development in the era of Industry 4.0, analyzed the collaborative creation of artists, cross-genre production and social attributes, and proposed an art development approach that is compatible with Society 5.0 [5]. Yi et al. explored the inheritance and development of Hebei folk music in the digital age from the perspective of ethnomusicology, and proposed a music revitalization framework centered on people [6]. Xia sorted out the development stages of digital humanities at home and abroad from the perspective of paradigm shift, and analyzed the theoretical evolution and methodological changes of its transformation from humanistic computing to digital humanities [7]. Underwood explored the application potential of machine learning in literary studies by training a genre model on the relevant book corpus of different

historical participants, in order to rethink the value of digital methods in literary interpretation [8]. However, these studies are limited by sample size and subjective interpretation, making it difficult to reveal the systematic quantitative relationship between industrialization and music creation, and even more difficult to capture the temporal characteristics and group diffusion patterns of subject matter evolution.

Bibliometrics, historical text mining, and topic modeling have demonstrated the ability to process large-scale text data in the fields of literary history and history of science, effectively identifying the evolutionary trajectory of specific concepts over long periods. For example, Joo et al. used text mining methods to conduct topic analysis on digital humanities literature, revealing the evolutionary trend of research topics and knowledge structures in this field [9]; Horváth and De Weerd, starting from the development of digital humanities research methods, pointed out that technologies such as natural language processing and topic modeling are driving the transformation of historical and cultural research towards large-scale text analysis [10]. In contrast, traditional music history research remains at the level of close reading of a few classic works, unable to exhaust the vast number of music documents of the 19th century, nor can it objectively measure the degree of penetration and diffusion path of industrialization elements in the creative group. In response to this limitation, this paper introduces an interdisciplinary framework that integrates bibliometrics, LDA topic modeling, and social network analysis, and systematically examines the impact of the textile industrialization process on Western music composition themes from three dimensions: lexical penetration, topic evolution, and network diffusion.

The main contribution of this paper is that it is the first to systematically introduce quantitative text analysis methods into the study of music social history, constructs a corpus of industrialization-themed music documents covering the period from 1760 to 1900, and reveals the correlation between textile industrialization and music composition themes through multi-dimensional empirical analysis, providing a repeatable and verifiable quantitative paradigm for understanding the interaction mechanism between technology and art.

II. DIACHRONIC CONSTRUCTION OF A THESAURUS OF TEXTILE INDUSTRIALIZATION CHARACTERISTICS

We first systematically extracted and dated industrialization-related terms to construct a thesaurus for subsequent quantitative analysis. To account for linguistic evolution and avoid under-counting early-stage concepts, the search parameters were expanded beyond modern terms like ‘factory system’ to include historical variants and period-specific descriptors (e.g., ‘manufactory,’ ‘mill-work,’ or ‘spinning engine’) identified through the British Patent Office Archives and the Parliamentary Blue Books. The sources of the terms were primarily based on three types of literature: first, specifications of textile machinery patents registered between 1760

and 1900 in the *History of Technology* (Textile Volume) and the *British Patent Office Archives*, used to extract core technical terms; second, engineering manuals and factory survey reports published during the same period, used to supplement vocabulary related to production processes and organizational forms; and third, survey reports on the factory system in the British Parliamentary Blue Book, used to collect keywords related to social change.

Somewhere in the term selection process we set forth three inclusion criteria: The term has to be relevant for the process of textile industrialization, it has to have stable technical connotations or social references in recent literature, and the time of its first commercial use or social relevance must be identifiable. These criteria yielded the following three groups of three core terms: First group: technical instruments: names for particular machines such as “spinning jenny,” “water-powered spinning machine,” “mule,” and “powered loom”;

Second group: production organization: institutional terms such as “factory system,” “factory structure,” and “workshop”; Third group: social groups: social roles that emerged in connection with the industrialization process such as “factory worker,” “hand weaver,” and “Luddite.”

The chronological division of the entries is based on the date of their first commercial application. Technical terms are based on the year of patent authorization or the date of the first practical machine put into production as recorded in historical materials, while terms related to social change are based on the year in which the term began to be widely used in public literature. After this periodization, we divided the period from

1760 to 1900 into four stages: the nascent stage of mechanization (1760s-1780s), the power revolution stage (1790s-1810s), the factory popularization stage (1820s-1840s), and the social adjustment stage (1850s-1900s). This periodization framework lays the foundation for subsequent analysis of the correspondence between the temporal penetration of industrialization terms in musical texts and the rhythm of textile technology diffusion.

III. MULTI-SOURCE COLLECTION AND STRUCTURED PROCESSING OF A CORPUS OF MUSICAL THEMES

In order to reveal the distribution trend of industrial elements in musical texts in an orderly manner, we built a thematic corpus of Western music literature from 1760 to 1900. The corpus was mainly based on three kinds of digital resources: first, the librettos of opera, lyrics of art song and titles and descriptions of symphonic poems and other instrumental works were collected from the IMSLP International Music Library project; Second, descriptive texts of works from contemporary music reviews, composer’s letters and journal articles were obtained from the RILM International Music Archive; Third, biographical entries and works catalogue from Grove Music Online were used to supplement and correct the composition years and composer attributions. The collection range was first confined to the works that were

first composed or premiered from 1760 to 1900, mainly from Britain, Germany, France and Italy Western European countries, and some representative works of Eastern and Northern European composers. We collected 340 opera librettos, more than 1200 lyrics of art song, titles and descriptions of 280 symphonic poems and other instrumental works, as well as more than 600 music reviews from the same period. The original texts were mainly in PDF or image files. We used the pdfplumber library of Python based on the Tesseract OCR engine to extract texts. Finally, we manually sampled and verified the extraction results, and the text recognition accuracy reached more than 95%. After digitizing the text, we carried out systematic cleaning and structuring. First, we standardized the encoding format and removed the noise information such as typesetting symbols and page number markers. Then, we segmented the text according to the work ID and constructed a collection of documents based on the same work. Finally, we labeled each work with the year of composition, composer's name, nationality, genre (Classical, Early Romantic, Middle Romantic, etc) and type (opera, art song, symphonic poem, etc). To facilitate the diachronic analysis in the later stage, we aggregated all the texts into 14 time slices (1760-1769 to 1890-1900) based on the ten-year period. Each slice contains a collection of all the works in that slice. The corpus built in this project contains about 1.45 million words, which provides initial data support for the subsequent frequency statistics and topic modeling of industrialization terms.

IV. SLIDING WINDOW-BASED WORD FREQUENCY-TIME SERIES CORRELATION ANALYSIS

To study the temporal correlation relationship between the frequency of appearance of industrialization terms in musical texts and the real diffusion process of textile technology, we design a sliding window analysis method. Firstly, we take the ten-year width of the sliding window and five-year sliding steps, and divide the period from 1760 to 1900 into multiple periods. Then, in each period, we statistically analyze the standardized word frequency (times/10,000 words) of industrialization characteristic words and the number of textile technology patents granted in UK in the same period. The patent data comes from the compilation of Chronology of Textile Patents compiled by UK Patent Office, and can be used as a substitute indicator to reflect the intensity of technology diffusion.

After that, we calculate the Pearson correlation coefficient between the two time series in each window located, and save the position of the window and the lag order with the maximum correlation coefficient. Then, through the continuous calculation of sliding window, we can observe the change trend of the correlation coefficient with time, and find the node where the correlation of the two changes from weak to strong. In addition, we use cross-correlation function to further explore the response lag of the word frequency sequence to the patent sequence, that is, there will be a certain delay of several years in the entry of

industrialization terms into musical texts? And whether there is a periodic change in the delay period in different historical stages. The method can dynamically reflect the rhythmic characteristics of the transmission of technological impact to the art field, and controls the cumulative effect over time, which provides reference for the subsequent regression analysis.

V. CO-OCCURRENCE MODELING OF COMPOSER-TECHNICAL TERMINOLOGY BIMODAL NETWORK

To examine the distribution differences and dissemination paths of industrialization themes among different composer groups, we constructed a bimodal network of composers and industrialization technical terms.

Network nodes are divided into two categories: one category consists of 40 representative composers active between 1760 and 1900 whose works are available in the corpus, covering different schools including Classical, Early Romantic, Middle Romantic, and Nationalist; the other category consists of 15 core terms from the previously constructed industrialization feature vocabulary, including terms related to technological objects, production organizations, and social groups.

Edges in the network denote co-occurrence relationships between terms, and edge weights reflect the frequency of the term in a composer's works. That is, we examined all of the works by each composer (including opera librettos, art song lyrics, and music reviews) and counted the number of times a given industrialization term was encountered to create a 40×15 co-occurrence frequency matrix. Because the size of the body of works for each composer can influence term frequency, we standardized each frequency count using the total number of words in a composer's works and then rounded the number to obtain the final weights of the edges in the network.

After the network was constructed, we visualized the network using Gephi software. We also calculated the network centrality indices for two types of nodes in the network: composer nodes and terminology nodes. For composer nodes, we calculated weighted degree centrality to reflect the overall degree of acceptance of industrialization terminology and calculated betweenness centrality to identify composers whose creative output most frequently synthesizes diverse industrial descriptors. In this historical context, high betweenness centrality (notably in Liszt) suggests a prolific and diverse thematic range that bridges various social and technical lexicons, rather than a literal 'transmission' of terms between individuals as seen in modern social networks. For terminology nodes, we calculated weighted degree centrality in a similar way to reflect the diffusion breadth of a given term among different composers and calculated eigenvector centrality to identify the relative influence of a terminology's associated composer. In addition, we examined the correlation between the industrialization terminology co-occurrence network and the composer so-

cial networks constructed from the contemporaneous letters and travelogues using QAP correlation analysis to assess whether the diffusion of industrialization themes was facilitated by actual interactions among composers. Through this modeling process, we sought to uncover the diffusion structure of industrialization imagery in music composition research at the group level of dissemination nodes and core terminology clusters.

VI. A QUANTITATIVE MEASUREMENT SYSTEM FOR THE SHIFT IN MUSICAL THEMES UNDER THE IMPACT OF INDUSTRIALIZATION

A. FREQUENCY OF INDUSTRIALIZATION TERMS

Industrialization terms emerged from a corpus of musical themes. The frequency of occurrence of these terms in librettos of opera, lyrics of art song and music reviews were statistically analyzed in a ten-year span by calculating the standardized lexical density (occurrences per 10,000 words) for the respective decades. The trend test was used to analyze the trend of the penetration rate time series from 1760 to 1900. The growth rate between each decade was calculated to analyze the temporal features and acceleration inflection points of industrialization terms entering into the musical vocabulary.

As shown in Table 1, the penetration rate of industrialization terms in musical texts generally showed an upward trend from 1760 to 1900, increasing from 0.025 times/10,000 words in the 1760s to 0.241 times/10,000 words in the 1890s. The data sequence exhibits distinct phased characteristics: the first surge occurred in the 1780s (growth rate of 172.7%), possibly related to the diffusion of technology after the expiration of Arkwright's water-powered spinning machine patent; subsequently, a plateau period lasting over a decade followed, with the penetration rate fluctuating narrowly between 0.055 and 0.081; a significant inflection point occurred again in the 1840s, with the growth rate reaching 55.6%, and the penetration rate exceeding 0.12 times per 10,000 words for the first time; after continuing its growth trend in the 1850s, the growth rate slowed to 1.8% in the 1860s, before rebounding to 19.5% in the 1870s. This fluctuation pattern coincides with the phased characteristics of technological revolutions in the textile industry—each major technological breakthrough (such as the improvement of the power loom in the 1780s and the popularization of the steam engine in the 1840s) was often followed by a creative response period of several years, followed by a stage of thematic sedimentation and digestion. This reflects that the acceptance of social change by artistic creation is not a linear process, but rather exhibits a cyclical characteristic of “impact—response—sedimentation”.

B. TRACKING THE EVOLUTION OF INDUSTRIALIZATION THEMES BASED ON LDA TOPIC MODEL

We applied unsupervised learning with a Latent Dirichlet Allocation (LDA) topic model on musical texts from 1760–

1900. The selection of $K=15$ topics was determined by calculating the coherence score (C_v) and perplexity across a range of K values (from 5 to 50). The value of 15 was found to be the optimal ‘elbow’ point, maximizing semantic interpretability while minimizing redundant clustering, thereby ensuring that industrial themes were captured without losing nuanced sub-themes. I extracted the two industrialization-related topics “machine imagery” and “factory scene” from the clustered results. To ensure the LDA model effectively distinguished between traditional folk imagery (e.g., the hand-operated ‘spinning wheel’) and industrial tropes (e.g., the ‘spinning machine’), the keyword filtering process excluded terms specifically associated with pre-industrial craftsmanship, focusing instead on the mechanical and rhythmic descriptors unique to factory-based production. I computed the posterior probability of each topic in the annual corpus as an intensity value and plotted the curve of its growth and decline over time for identifying the prevalence of industrialization-related themes in musical composition.

As shown in Figure 1, the intensity of the two industrialization themes, “machine imagery” and “factory scene,” generally showed an upward trend from 1760 to 1900. The intensity of the machine imagery theme rose from 0.12 in the 1760s to 0.58 in the 1890s, while the intensity of the factory scene theme rose from 0.08 to 0.54. A significant turning point occurred in the 1840s: previously, the two themes showed slow growth, with intensities of 0.25 and 0.21 respectively in the 1830s; between 1840 and 1849, they jumped to 0.38 and 0.32, representing increases of 52% and 52.4% respectively, after which they continued to climb, but the growth rate gradually slowed. This inflection point coincides with the historical context of the full maturation of textile industrial mechanization technology in the 1840s and its spread across continental Europe, indicating that the substantial penetration of the industrialization wave into musical composition themes began in the mid-19th century.

C. SOCIAL NETWORK CENTRALITY ANALYSIS OF THE DIFFUSION OF INDUSTRIALIZATION THEMES

According to the co-occurrence relationship between representative composers and industrialization terms between 1760 and 1900, two-mode network matrix was constructed, and the element of matrix represented the occurrence frequency of terms in the works of composer. Then, weighted degree centrality and intermediary centrality of each node were calculated to explore the key nodes and core terms in the spread of industrialization theme, and then diffusion path and structural features of industrialization image in the community of music composition were revealed.

Figure 2(a) shows the weighted degree centrality distribution of ten representative composers in the co-occurrence network of industrialization terms, and Figure 2(b) shows the weighted degree centrality distribution of five core industrialization terms in the composer-terminology binary network. Liszt, Wagner, and Verdi exhibit weighted degree

TABLE 1. Comparison of the Frequency of Industrialization Terms

Period	Corpus Size (10,000 words)	Frequency of Industrial Terms	Standardized Penetration Rate (per 10,000 words)	Decadal Growth Rate (%)
1760–1769	40	1	0.025	—
1770–1779	45	1	0.022	-12
1780–1789	50	3	0.06	172.7
1790–1799	55	3	0.055	-8.3
1800–1809	62	4	0.065	18.2
1810–1819	70	5	0.071	9.2
1820–1829	78	6	0.077	8.5
1830–1839	86	7	0.081	5.2
1840–1849	95	12	0.126	55.6
1850–1859	105	18	0.171	35.7
1860–1869	115	20	0.174	1.8
1870–1879	125	26	0.208	19.5
1880–1889	135	31	0.23	10.6
1890–1900	145	35	0.241	4.8

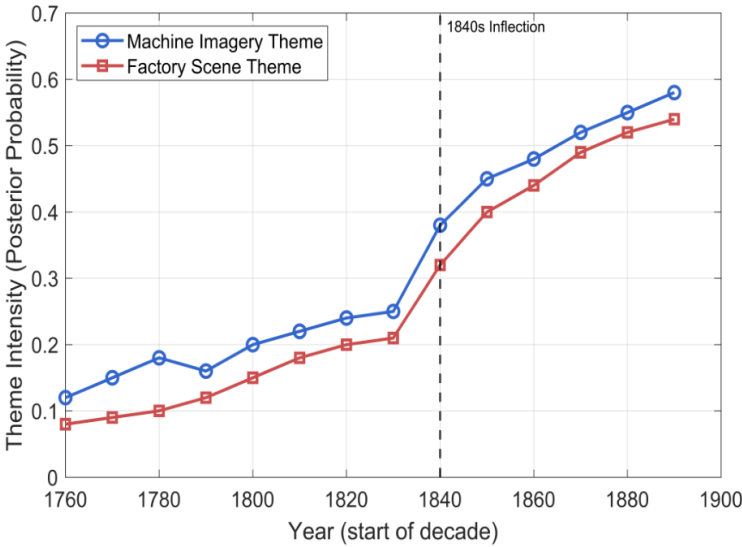


FIGURE 1. Evolution Curve of Industrialization-Related Themes

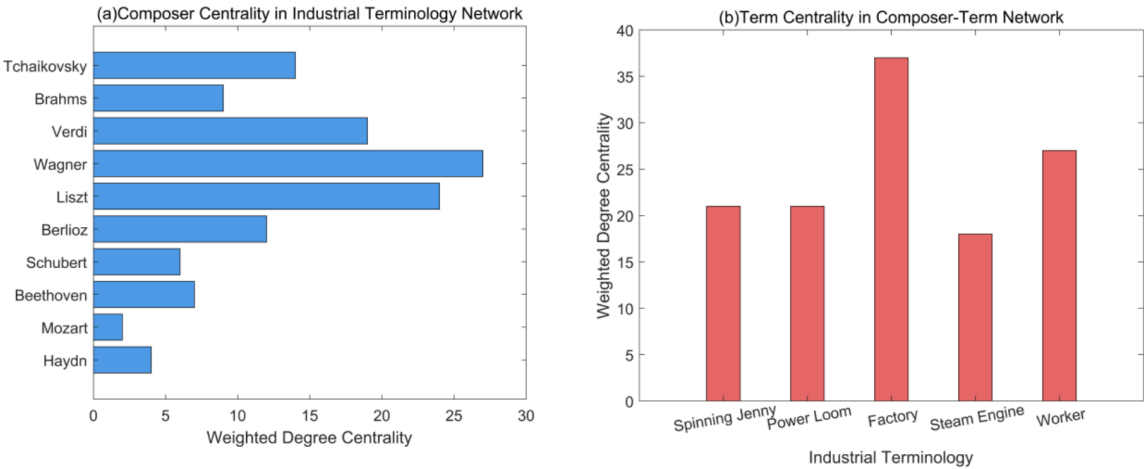


FIGURE 2. Network centrality distribution of composers and industrialization terms

centralities of 24, 27, and 19, respectively. While these are numerically higher than Haydn (4) and Mozart (2),

this comparison serves primarily to establish a historical baseline rather than a direct measure of ‘acceptance.’ The

low centrality of Classical-era composers reflects the nascent state of the industrial infrastructure in late 18th-century Austria, where the ‘factory system’ had not yet emerged as a dominant social or cultural construct available for artistic reference. These higher values reflect the broader ‘industrialized’ lexicon within the period’s music reviews and biographical entries, indicating a heightened cultural presence of these themes in the musical environment of the late Romantic period, regardless of whether they represent a conscious stylistic choice by the composers themselves. At the terminology level, “factory” has the highest centrality (37), followed by “worker” (27), while the purely technical terms “spinning machine” (21) and “loom” (21) have relatively low centralities. This distribution reveals that the impact of industrialization on musical composition permeates more through social change than simply through technological means, while also confirming the close relationship between the spread of industrialization themes and the composers’ historical context and creative inclinations.

VII. CONCLUSION

This paper has been a systematic study of the connection between the process of industrialization of the textile industry and the themes of Western musical composition between 1760 and 1900 through the use of bibliometrics techniques. The research determined that the penetration rate of industrialization terminology within musical texts has increased significantly since the 1840s. Furthermore, the prevalence of machine imagery and factory scenes grew by more than 52 percent during this period, demonstrating the substantial impact of the industrial wave on musical composition starting in the mid-19th century. The social network analysis additionally demonstrates that the influence on the acceptance of the industrialization themes was the most extensive among the late Romantic composers (Liszt, Wagner), and the influence of social terms (factories, workers) was more widespread than the influence of the technical ones. It must be noted that the corpus of this study primarily deals with mainstream Western European music literature. The exclusion of a broader sample from Eastern Europe—particularly Russia, which underwent a unique, rapid industrialization—is a limitation. Such late-stage industrial aesthetics, including the ‘Iron Foundry’ style of the late 19th and early 20th centuries, suggest that including these regions might significantly amplify the penetration data observed for the 1890–1900 period. The study might also be further developed and expanded to a larger geographical region and, in association with the letters of composers, diaries and other personal documents, the particular psychological motivation and social conditions of industrialization topics penetrating the musical composition might be addressed in more detail.

AUTHOR CONTRIBUTIONS

Yue Fan designed, collected and analyzed the data, and drafted the manuscript. Yue Fan conducted the study, criti-

cally revised the manuscript for important intellectual content, and gave final approval of the version to be published. Yue Fan participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

- [1] S. Wei, “A Reinterpretation of the Pompidou Centre Beyond Formal Expression,” *Urban Architecture*, vol. 20, no. 6, pp. 97-99, 2023, doi: 10.19892/j.cnki.csjz.2023.06.27.
- [2] M. Rampley, F. Meng, and C. Sun, “Reflections and Critiques of Classical Art History on Evolutionary Aesthetics, Neuro-Art History and Systems Aesthetics,” *Art Exploration*, vol. 35, no. 4, pp. 84-99, 2021, doi: 10.13574/j.cnki.artsexp.2021.04.008.
- [3] K. Dong, “Some Thoughts on the Transformation and Upgrading of Traditional Cotton Textile Manufacturing Industry,” *Cotton Textile Technology*, vol. 53, no. 1, pp. 2-5, 2025, doi: 10.3969/j.issn.1001-7415.2025.01.001.
- [4] S. Zhou, “A Study on the Cultural Autonomy and Aesthetic Change of Zhongshan Suit under the Perspective of Design Aesthetics,” *Art and Technology*, vol. 38, no. 22, pp. 1-3, 2025, doi: 10.3969/j.issn.1004-9436.2025.22.001.
- [5] M. Mokhtar, A. WMD S, and R. Abd Rahim, “Art throughout the Industrial Revolution and Art Framework for the Industrial Revolution 4.0,” *Environment-Behaviour Proceedings Journal*, vol. 7, no. SI7, pp. 149-154, 2022, doi: 10.21834/ebpj.v7iSI7.3776.
- [6] Z. Yi, S. Jiawei, and L. Yuhui, “A study on the strategies for cross-cultural innovation and development of Hebei folk music from the perspective of ethnomusicology,” *Digital Scholarship in the Humanities*, no. 4, p. 4, 2025, doi: 10.1093/lilc/fqaf115.
- [7] C. Xia, “Theoretical Evolution and Paradigm Shift of Digital Humanities: From “Digital Methods in Humanities Research” to “Humanities Research in the Post-Digital Era”,” *Journal of Library Science in China*, vol. 50, no. 5, pp. 83-101, 2024, doi: 10.13530/j.cnki.jlis.2024040.
- [8] T. Underwood, “Machine learning and human perspective,” *PMLA*, vol. 135, no. 1, pp. 92-109, 2020, doi: 10.1632/pmla.2020.135.1.92.
- [9] S. Joo, J. Hootman, and M. Katsurai, “Exploring the digital humanities research agenda: a text mining approach,” *Journal of Documentation*, vol. 78, no. 4, pp. 853-870, 2022, doi: 10.1108/JD-03-2021-0066.
- [10] A. Horváth and H. De Weerd, “Special issue on digital humanities and East Asian studies,” *International Journal of Digital Humanities*, vol. 4, no. 1, pp. 1-4, 2023, doi: 10.1007/s42803-022-00064-5.