Innovation Humanities and Social Sciences Research, Volume 21, Issue 12, 2025, ISSN: 2949-1282
Proceedings of The 12th International Conference on Education, Language, Art and Inter-cultural Communication (ICELAIC 2025)

Research on Talent Development Pathways for Data Journalism in China: An Empirical Analysis Based on Content Analysis

Honglei Zhai[1] Zhe Fu[2]

[1,2] *School of Law and Humanities & Social Sciences, Wuhan University of Technology, Wuhan, China*
[1]*Corresponding author.*

ABSTRACT

Data journalism, characterized by its presentation through visualization techniques, has emerged as a novel form of news reporting. Increasingly embraced by media organizations, it has become a burgeoning field within the journalism industry. Since 2015, the scale of the China Data Journalism Competition has continuously expanded. The award-winning works from this competition represent the latest compendium of data journalism practice in China and reflect the cutting-edge developments in this field domestically. Accordingly, this study takes the China Data Journalism Competition as its research subject, conducting empirical research on the content of its entries and their developmental trends. On one hand, it examines the evolving competition mechanisms to understand the shifting focus and dynamic progression of data journalism practices in China. On the other hand, it performs a content analysis of 381 award-winning works from successive competitions to summarize their content characteristics and changing trends. Based on the findings from these two dimensions, the study identifies the essential elements and directions for cultivating data journalism talent. Furthermore, it explores pathways for optimizing the development of data journalism talent cultivation in China. The aim is to provide valuable reference and theoretical guidance for advancing data journalism practice within the Chinese context.

Keywords: *Data journalism, China Data Journalism Competition, Talent cultivation, Optimization pathways.*

1. INTRODUCTION

Centered on "big data," data journalism is rapidly evolving into an integral part of contemporary news reporting. Worldwide, news media are actively engaging in its development. Data journalism competitions serve as both a showcase and a competitive arena for the field, reflecting its highest standards and signaling its trends. The establishment of prestigious international awards, such as the Global Data Journalism Awards, underscores the global emphasis on advancing this field.

In China, mainstream media and universities have actively embraced data journalism. To recognize outstanding domestic works, Professor Chen Jiyin from Xi ' an Jiaotong University launched the inaugural China Data Journalism Competition in Lanzhou in 2015. The Competition aims to advance data journalism education and foster practical exchange between academia and the industry.

As a pioneering national contest, it brings together outstanding achievements from both sectors, showcasing the distinctive characteristics and cutting-edge developments of data journalism in China. The data-driven shift in the media market highlights the urgency of cultivating interdisciplinary talent equipped with both journalistic literacy and technical skills. By providing a key platform for innovating teaching concepts and aligning talent development with industry transformation, the Competition plays a crucial role. Therefore, systematically examining its award mechanisms and winning entries can offer clear guidance for cultivating data journalism talent in China. It serves as an important reference for universities to optimize their programs, ensuring

that talent supply better aligns with the profession's evolving demands.

2. CONTENT ANALYSIS OF AWARD-WINNING ENTRIES IN THE COMPETITION

2.1 Research Sample and Coding Procedure Description

The research samples for this study were drawn from the winning entries of the 1st to 7th competitions. After screening, a total of 381 valid samples were obtained. (“Table 1”)

Table 1. Sample selection scope

Sample Source	Date	Work Category	Number of Works	Selection Method
The 1st China Data Journalism Competition	2015	Award-winning works	8	Full sample selection
The 2nd China Data Journalism Competition	2016		22	
The 3rd China Data Journalism Competition	2018		30	
The 4th China Data Journalism Competition	2019		75	
The 5th China Data Journalism Competition	2020		84	
The 6th China Data Journalism Competition	2021		61	
The 7th China Data Journalism Competition	2022		61	
Total	2015—2022		381	

This paper adopts a chronological approach, examining the development of seven editions of the competition from 2015 to 2022 as a cross-sectional analysis. It selected 381 valid award-winning data journalism works as samples. Subsequently, the author will code these 381 pieces across dimensions including creators, topic types, data sources, technological applications, presentation formats, high-frequency word analysis, and emotional tones. A coding table will be developed, and the samples will undergo evaluation and categorization. Based on macro-level statistical data and micro-level case observations, the analysis will examine the characteristics and evolution of award-winning entries across successive competitions. (“Table 2”)

Table 2. Sample awarded works categorization table

Category	Specific Category	Code Information
Text Analysis	(1) Work Creator	1=Academic Group; 2=Industry Group
	(2) Topic Selection for Works	1=Politics; 2=Economy; 3=Livelihood; 4=Crime; 5=Culture; 6=Technology; 7=Sports; 8=Environment; 9=Health; 10=Other
	(3) Emotional Tone of the Work	1 = Anger; 2 = Fear; 3 = Sadness; 4 = No Clear Emotional Bias; 5 = Hope; 6 = Courage; 7 = Emotional Resonance
	(4) Data Sources for Works	1=Media 2=Government Organizations 3=Non-Governmental Organizations; Enterprises; 4=Research Institutions; Academic Institutions; Scholars; 5=Self-collected
	Technical Application in the Work	1=Excel； 2=SPSS； 3=Python； 4=SQL； 5=R； 6=Tableau； 7=Echart； 8=openrefine
	(6) Work Presentation	Format 1 = Static image; 2 = Dynamic image (including video)

During the coding phase, the author randomly selected 40 award-winning entries (approximately 10% of the total sample) for reliability testing and independently coded them along with another coder,

who also holds a background in journalism and communication. Holsti's formula was used to calculate inter-coder consistency, which is expressed as: Reliability = (Number of Identical Codes) / (Number of Identical Codes + Number of Different Codes).

In this reliability test, both coders coded 360 units each. The number of identical codes between them was 336, yielding a reliability coefficient of 0.93. This meets the reliability test requirements.

2.2 Analysis of Awarded Works Creators

This section coded the authors of 381 valid award-winning entries from the competition, categorizing them into “academic group” and “industry group” based on the organizations of the creators. (“Figure 1”)

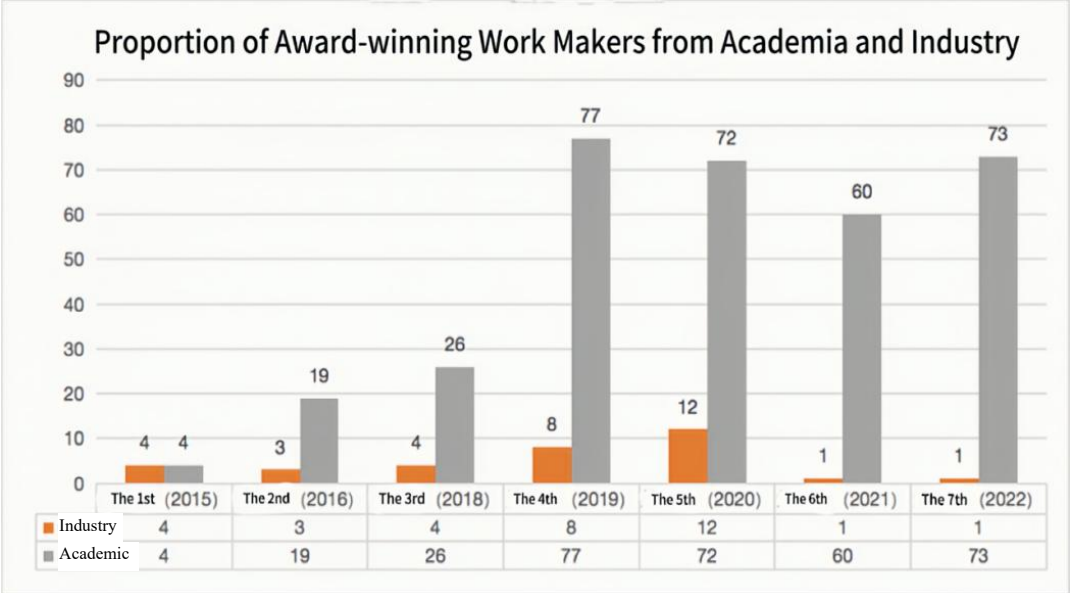

Figure 1 The ratio of academic versus industry creators among award-winning entries in previous competitions.

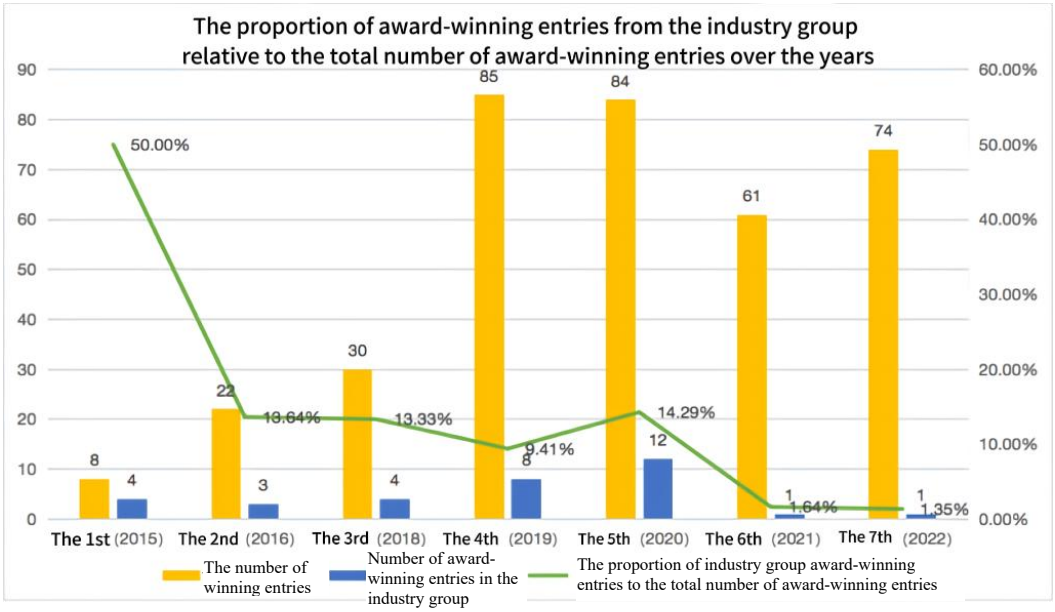

Figure 2 The proportion of award-winning entries from the industry group relative to the total number of award-winning entries over the years.

As shown in the "Figure 2"， the proportion of awards won by the industry group has generally declined across successive competitions. The inaugural competition established separate categories with equal awards, acknowledging the initial capability gap between academia and industry—where industry teams excel in practical experience and academic teams in analytical depth. From the second competition onward, the award structure clearly shifted toward academia, peaking in the seventh edition, where only one of the 74 winning entries came from industry.

The competition consistently bridges academia and industry while increasingly focusing on university-based practices, embodying the mission of "promoting learning and teaching through competition". It supports national media convergence and journalism education transformation, facilitating interdisciplinary integration. Notably, substantial participation from students with STEM backgrounds, such as big data and computer science, fosters journalism talent that integrates humanities, sciences, and engineering.

2.3 Analysis of Awarded Work Topics

2.3.1 Distribution of Topics

In terms of topic categories, this section coded the themes of the 381 valid award-winning entries from the competition, classifying them into ten categories: politics, economy, livelihood, crime, culture, science and technology, sports, environment, health, and others.

Since the second competition had a fixed theme of "Belt and Road," the related entries were excluded from the thematic distribution analysis. The actual sample consisted of 359 works from the remaining six competitions. To present trends clearly, the sample was divided into three periods: 2015–2018 (1st–3rd competitions), 2019–2020 (4th–5th competitions), and 2021–2022 (6th–7th competitions). These works cover a wide range of themes, focus predominantly on social hot topics, and often adopt a down-to-earth, everyday perspective. ("Figure 3")

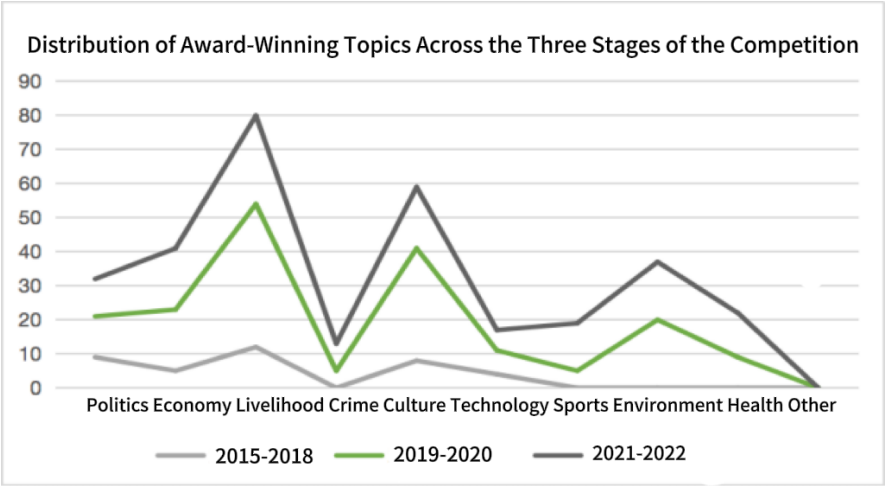

Figure 3 Distribution of Award-Winning Topics Across the Three Stages of the Competition.

Based on statistical analysis of entries from the six editions of the competition, topics related to politics, economics, and people's livelihoods have consistently been key areas of focus in data journalism. Among these, entries addressing people's livelihoods have been the most numerous and accounted for the largest proportion of award-winning works in each edition. Topics covered include the trend of singledom, and women's rights, highlighting the humanistic concern reflected in the competition entries.

In the cultural domain, as shown in Figure 3, cultural topics have gradually increased. This aligns with the national strategic emphasis on cultural confidence and revitalization. Participants have utilized digital methods to showcase China's outstanding traditional culture, demonstrating contemporary relevance.

Regarding environmental topics, China has reinforced environmental governance reforms in recent years, with clear goals set for carbon peaking and carbon neutrality. The sixth and seventh editions of the competition have kept pace with these developments, featuring award-winning works focused on key national issues such as

"carbon neutrality," highlighting the competition's alignment with policy directions.

2.3.2 High-Frequency Word Analysis

This author compiled the titles of 381 award-winning entries from the competition. To avoid bias from the mandated "Belt and Road" theme in the second competition, its 22 entries were excluded, leaving 359 titles from the other six editions for analysis. After text preprocessing, word-frequency analysis revealed the top keywords: "data, city, era, development, education, student, women, behind, Beijing, predicament." Among them, "city" appeared 13 times, "development" 10 times, "era" and "education" 9 times each, "student" and "women" 7 times each, and "Beijing" and "predicament" 6 times each. The results indicate that contestants focused heavily on themes such as urban governance, era development, education, youth employment, and women's welfare—topics closely linked to people's livelihoods and politics ("Figure 4").

Figure 4 High-frequency word cloud in award-winning work titles.

2.4 Emotional Analysis of the Winning Work

To analyze the attitudes and emotions conveyed by communication entities in award-winning works, this chapter focuses on emotional expression as a key dimension, statistically examining the emotional expressions across all winning entries from previous competitions. Based on Arnold's Emotion Classification Theory, with adaptive adjustments, seven emotion categories suitable for award-winning works were ultimately established: anger, fear, sadness, neutral emotion, hope, courage, and inspiration. These seven categories were used to code 381 award-winning works. Due to the limited number of award-winning works from the first to third competitions, to ensure statistical rigor, this section selected only award-winning works from the fourth to seventh competitions as the sample for coding.

As shown in "Figure 5", statistical analysis indicates that over time, the proportion of award-winning entries conveying positive emotions such as "hope, courage, and love" has gradually increased, while the share focusing on negative emotions like "anger, fear, and sadness" has consistently declined. This reflects a growing tendency among authors to share uplifting content and highlight solutions to social issues, demonstrating the cohesive and inspiring power of emotion in journalistic works.

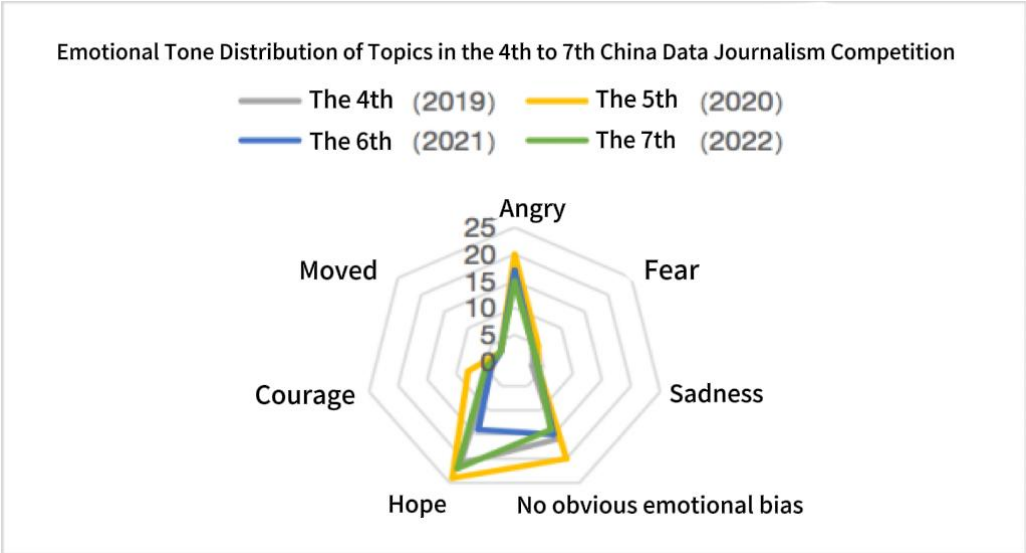

Figure 5 Emotional tone distribution of selected award-winning entries from the 4th to 7th competitions.

2.5 *Analysis of Data Sources for Award-Winning Works*

The author categorized the data sources of shortlisted entries from previous competitions into the following five types and conducted coded statistical analysis: First, media outlets, including data published by news agencies or other publications; Second, governmental organizations, encompassing domestic institutions such as the State Council and the Supreme People's Procuratorate, as well as international governmental bodies like the United Nations and the World Health Organization; Third, non-governmental organizations and enterprises; Fourth, research institutions, academic institutions, and scholars; Fifth, self-collected data, such as firsthand information gathered directly through field investigations, interviews, internet platforms, and other channels.

As shown in “Figure 6”, among the 381 award-winning entries from past competitions, data sources demonstrate considerable diversity. Works using two categories of data sources were most common (43%), followed by those relying on one category (30%) and three or more categories (27%), indicating that the vast majority integrated multiple data types. In terms of source types, government materials were most frequently used (72%), followed by non-governmental organizations and enterprises (46%), media sources (26%), research institutions and scholars (24%), and self-collected data (14%). Overall, the diversity of data sources significantly enriches the content and types of award-winning works.

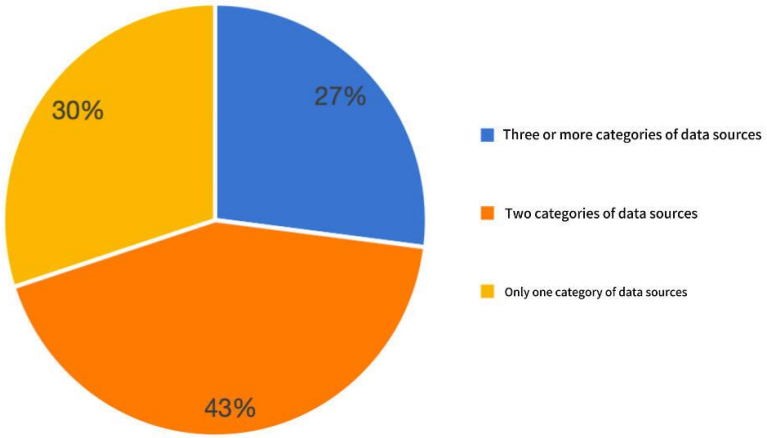

Figure 6 Data sources used in award-winning entries from previous competitions.

2.6 *Technical Application Analysis of Award-Winning Works*

Effective data journalism requires the seamless integration of data, with the key lying in leveraging tools for mining and analysis to uncover valuable news angles.

By analyzing the tool descriptions provided on the award-winning pages of 381 entries from past competitions, this paper lists tools that appeared twice or more across all contests: Excel, SPSS, Python, SQL, R, Tableau, Echart, and OpenRefine.

Statistics show that among the 381 award-winning entries, Excel was the most frequently used tool (86%), followed by Python (43%), SPSS (28%), Tableau (20%), SQL (14%), R (12%), and Echart (9%). Data journalists employ diverse technical approaches. For instance, some entries integrated Python for data collection, Tableau for visualization, and Excel for analysis; others utilized SQL for data mining alongside Python for text analysis. However, it was also observed that some projects relied on limited data processing tools, which constrained the depth of data analysis.(“Figure 7”)

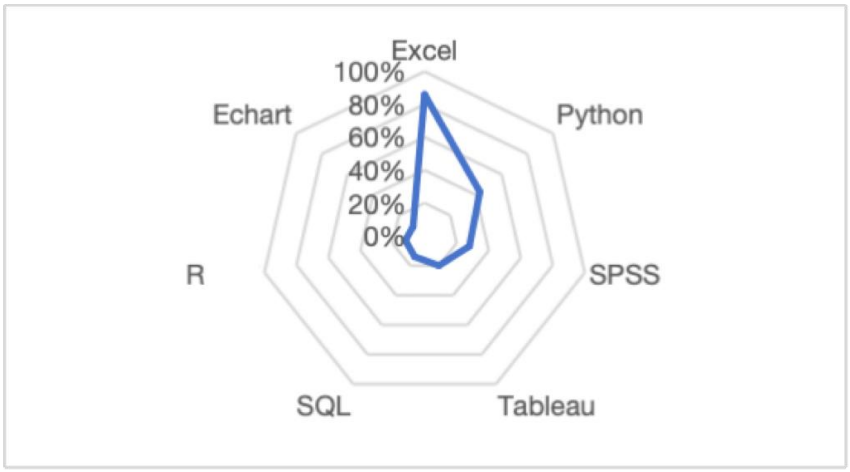

Figure 7 Data analysis tools used in the competition award-winning entries.

2.7 Analysis of the Form of Award-Winning Works

In the context of digital reading, visual expression is key to evaluating data journalism. This study distinguishes between static and dynamic visuals. Static graphics, comprising 73% of award-winning works, offer compatibility and speed but limited interactivity. Interactive dynamic visuals enhance both information richness and engagement, and their growing use reflects technological advances in the field.

Award-winning entries show proficiency in web-based presentation, characterized by "minimal text, rich graphics," and often employ social-media-friendly formats like H5 and galleries. These works skillfully blend narrative and visual elements to communicate effectively.

2.8 Chapter Summary

A systematic analysis of award-winning entries across creators, topics, data sources, technology, and presentation formats outlines their core characteristics and developmental path. The competition shows a clear shift from academia-industry balance toward academic predominance, underscoring its role as a practical platform for journalism education. Winning works focus on positive, contemporary, and increasingly international topics, responding to social and global issues. Official data dominate, highlighting an emphasis on authority but also revealing challenges in depth and transparency. While static infographics remain mainstream, interactive visuals are growing, supported by expanding technical tools.

These trends depict the evolution of data journalism training in China: from balanced to academia-led participation, from domestic to international perspectives, and from limited tools to diverse, interactive visual forms. The competition has become a key arena for advancing data literacy, visualization skills, and interdisciplinary collaboration. Its progress in thematic breadth, data use, and technical expression reflects data journalism's ongoing role in serving national priorities, innovating narratives, and adapting to global communication. This accumulation of experience not only informs the cultivation of data journalism talent but also charts a clear direction for its future development.

3. PROBLEM IDENTIFICATION BASED ON EMPIRICAL RESEARCH FINDINGS

3.1 Some Works Have Limited Data Sources and Insufficient Analytical Depth

Analysis of the 381 valid samples reveals that a significant 72% of award-winning entries relied on government data sources, while only 14% utilized self-collected data. This reflects a widespread preference for official channels among participants. However, limitations in the type, breadth, and depth of official data, combined with the generally weak independent data mining capabilities of university teams, resulted in insufficient depth in the submissions. Furthermore, while most award-winning works cite the names of data-providing institutions, they often fail to provide specifically traceable datasets, making it difficult for readers to

verify sources or conduct further exploration, thereby undermining the depth and credibility of the works.

For instance, the entry "The Loneliness of Living Alone: Data on the Challenges Facing Elderly Individuals Living Alone" while citing authoritative sources such as the National Bureau of Statistics, was ultimately constrained by the student author's limited ability to acquire data and conduct empirical analysis. It relied primarily on existing literature to present the issue instead of employing more scientific primary data for in-depth investigation, illustrating the limitations described above.

3.2 Some Works Feature Limited Visual Presentation Formats and Lack Interactivity

Data visualization enhances journalism by making content tangible, clarifying complexity, and transforming news into knowledge. Award-winning works show that both academia and industry employ diverse tools, primarily delivering outputs as web pages supplemented by short videos, WeChat articles, H5, long graphics, and animations—with static images still prevalent. However, some entries adopt formulaic designs or prioritize form over substance, merely stacking charts and text or using exaggerated visuals. Such works often remain descriptive, lacking deeper analysis of data and context, which limits their journalistic value.

For example, the fourth-competition entry "How Well Have You Slept These Past Years?" used a long static graphic to present national sleep-quality data from 2017–2019. Though timely, its rigid layout and lack of interactivity reduced engagement, while an ornate starry-sky background distracted from the core message.

3.3 Some Works Suffer from a Technocentric Approach, Neglecting Narrative Elements

Professor Chen Jiyin, founder of the competition, observed that data journalism often "neglects news value and social function in pursuit of visual effects." Statistical analysis confirms that some works still prioritize data accumulation and visual spectacle, weakening narrative depth and meaning. Data should enhance clarity and insight, not become an end in itself; narrative integrity must remain central.

For instance, the fourth-competition entry "Why Do They Endure?..." used only basic icons to show rankings, lacking interpretive depth. Its ornate diagrams poorly matched the text, and over-reliance on citations led to superficial analysis. Therefore, tools and technology should serve the story—not the reverse. Technical display must not compromise the intrinsic value of data journalism.

3.4 Some Works Exhibit Compatibility Issues, and Their Dissemination Potential Remains To Be Enhanced

Compatibility issues have persisted across past winning entries. Some web-based submissions experience problems on mobile devices, such as failure to display content or loss of original page layout, which hinders normal viewing and information access. Such issues not only interfere with online voting but also constrain the external dissemination of the competition. Therefore, data journalism should prioritize data compatibility and adaptability in its presentation methods.

4. OPTIMIZING PATHWAYS FOR CULTIVATING DATA JOURNALISM TALENT IN HIGHER EDUCATION INSTITUTIONS

4.1 Leading the Way in Showcasing Data Journalism's Distinctive Features: Mastering the Essentials of Data Journalism

The primary task of data journalism education lies in anchoring its fundamental attributes. The training process must adhere to the new liberal arts philosophy of "cultivating virtue and talent" and "integrating humanities and sciences," implementing programs to cultivate high-quality professionals. Upholding authenticity as an inviolable baseline, the following principles must be followed in practice: first, establishing a tight connection between data and events; second, leveraging visual storytelling and interactive presentation to enhance audience comprehension; third, maintaining an audience-centric approach by delivering personalized content and analyzing news consumption trends for optimization; fourth, respecting the professional norms of data journalism. Within the context of new liberal arts, innovative practices in data journalism education

manifest as the integration of theory and technology and the cultivation of comprehensive competencies.

4.2 Optimizing Data Journalism Teaching Models: Innovating Blended Learning Approaches

To meet the interdisciplinary demands of data journalism, journalism programs must innovate teaching models, restructure curricula, and integrate online and offline resources to deliver personalized, cross-disciplinary instruction that enhances students' practical problem - solving skills. Specific approaches include: first, restructuring curricula through phased course design to progressively cultivate students' data literacy; second, establishing an online resource repository by integrating high-quality digital course materials into daily teaching, enriching case libraries to foster blended learning, and strengthening topic selection and visualization skills through practical exercises; third, innovating blended teaching models by combining self-directed online learning, classroom seminars, and laboratory practice—incorporating video - based pre - class learning, inquiry - based discussions on humanities topics during class, and progressively challenging lab sessions to cultivate versatile professionals.

4.3 The "Multimedia" Transformation of Data Journalism: Diversifying the Development of Journalism Studies

Technological progress has accelerated the digital transformation of journalism disciplines. Data journalism education should therefore integrate theory and technology: cultivating students' humanistic literacy and data processing skills to uncover the meanings behind data, while enhancing their practical abilities in data storytelling, visualization, and multimedia expression. This approach improves communication effectiveness and enables professionals to adapt to and lead industry change.

4.4 Breaking Down Disciplinary Boundaries: Cross-Disciplinary Collaboration

Mobile internet has broadened channels for information access, compelling news media to innovate formats to capture user attention and cultivate versatile, application-oriented talent. As a product of multidisciplinary convergence, data journalism education centers on cross-disciplinary training. Rapid development in Europe and the United States stems from established interdisciplinary programs at institutions like Columbia University and the University of Missouri. Chinese universities are now exploring collaborative teaching models where journalism faculty lead theoretical instruction while computer science instructors handle practical skills like data mining and visualization. This joint effort builds curricula and gradually establishes localized educational pathways.

4.5 Establishing an Instructional Matrix: Innovative Practice-Oriented Teaching Through School-Enterprise Collaboration

Data lies at the core of modern journalism, making the cultivation of data-savvy communication professionals a pivotal direction for media education reform. Training has evolved from theoretical case studies to hands-on skill development, aiming to nurture versatile talents proficient in data reporting, analysis, and programming, equipped with critical data thinking.

Such interdisciplinary cultivation requires collaboration across sectors. In recent years, there has been a concerted push for industry-academia integration, establishing joint training bases and developing market-oriented talent through cooperative models that align with industry needs.

5. CONCLUSION

This study employs a multidimensional content analysis of award-winning works from the China Data Journalism Competition, systematically mapping the evolution and defining features of data journalism practice in China. Findings indicate a shift in the competition from "balance between academia and industry" to "academic predominance," underscoring its role in "cultivating talent and improving teaching through competition." Winning entries actively engage with contemporary and global issues, rely mainly yet increasingly diversely on official data sources, utilize a broadening set of technological tools, and exhibit a move toward interactive and dynamic visual forms. The study also identifies persistent challenges, including superficial data analysis, imbalance between narrative and technology, and limitations in cross-platform compatibility and dissemination.

Theoretically, this research offers an analytical framework and trend evaluation based on extensive empirical evidence for data journalism education in China. Practically, the proposed pathways — centered on curriculum redesign, interdisciplinary collaboration, industry-academia partnerships, and technical upskilling—provide actionable references for reforming journalism education and training interdisciplinary talent. Moving forward, data journalism education should uphold journalistic authenticity and humanistic values while further fostering cross - disciplinary integration, technological innovation, and industry - academia synergy. This will contribute to building a localized talent - development system adapted to the intelligent media era, thereby better serving national strategic objectives and the evolving global communication landscape.

ACKNOWLEDGMENTS

Funding Projects:

1.National Social Science Fund General Project: “Audience Perception and Adoption Behavior of Data Journalism” (20BXW029);

2.Wuhan University of Technology 2024 University-Level Graduate Education Teaching Reform Research Project: “Exploratory Research on Interdisciplinary Models for Cultivating Innovative Journalism and Communication Talent” (y2024017).

REFERENCES

[1] Zhao Hongxiang. Exploring Pathways for Cultivating Data Journalism Talent in China [J]. Media, 2019, (10): 83-85.

[2] Qi Yinmo. Forging New Glories in Socialist Culture Through Cultural Confidence and Self-Strengthening: A Summary of the “Theoretical Symposium on Cultural Development in Implementing the Spirit of the 20th CPC National Congress” [J]. Library Forum, 2022, 42(11): 1-7.

[3] Lu Mengmeng. Research on Interactive Narrative Strategies in Data Journalism: An Analysis of Award-Winning Works from the Global Data Journalism Awards (2014-2018) [J]. Publishing Science, 2019, 27(01): 92-98.

[4] Zhao Yang. Visual Presentation and Visualization in Data Journalism: A Case Study of Award-Winning Works from Two Editions of the “Competition” [J]. China Packaging, 2017, 37(02): 33-38.

[5] Chen Jiyin, Feng Jiao. An Empirical Study Based on Finalists of the Third Competition [J]. Education Media Research, 2019, (01): 61-64.

[6] Feng Fan, Zhao Ruojin. Exploring Pathways for Learning Through Data Journalism Competitions [J]. Audiovisual Theory and Practice, 2023,(02):65-70.

[7] Liu Hengyu. Chen Jiyin: Strengthening New Liberal Arts Thinking for Data Journalism Development [J]. Journalism Forum, 2021, 35(01): 35-36.

[8] Chen Jiyin, Yang Lian. Current Status, Challenges, and Countermeasures for Data Journalism Development in China [J]. Journalist, 2016, (11): 64-70.

[9] Chen Jiyin, Liu Yingqi. Practice and Prospects of Data Journalism [J]. Journalist, 2015(08).

[10] Chen Jiyin. The Frontier of Data Journalism Practice in China: An Empirical Study Based on Entries from the First China Data Journalism Competition [J]. Journal of Northwest Normal University (Social Sciences Edition), 2016, 53(05): 140-144.

[11] Fang Yihua, Yang Huihan. On the Narrative Paradigm of Data Journalism [J]. Modern Communication (Journal of Communication University of China), 2018, 40(12): 45-49.

[12] Li Yaping. Exploring Pathways for Reforming Journalism and Communication Practice Teaching in the Context of New Liberal Arts [J]. Journal of Journalism Research, 2022, 13(15): 61-63.

[13] Arnold M B. Emotion and personality: Volume 1, psychological aspects[M]. New York: Columbia University Press, 1960:203-218

[14] Ministry of Education, Publicity Department of the CPC Central Committee. Opinions on Enhancing the Capacity for Cultivating Journalism and Communication Talent in Higher Education Institutions and Implementing the Excellence Program for Journalism and Communication Talent Education 2.0 [Z]. Jiao Gao [2018] No. 7, 2018.