systems

MDPI

Article

Reconfiguring the Media–Public Discourse System After ChatGPT: Agenda-Melding and Experiential Accessibility in South Korea

Hyungkun Hahm [1], Sungbok Chang [2,*] and Jungho Suh [3,*]

[1] Department of Media Communication, College of Social Sciences, Hanyang University, Seoul 04763, Republic of Korea; hkhahm61@gmail.com
[2] Department of Digital Media Design, College of Arts, Cheongju University, Cheongju 28496, Republic of Korea
[3] Department of Media Communication, College of Social Science, Gachon University, Seongnam 13120, Republic of Korea
* Correspondence: sbchang@cju.ac.kr (S.C.); hoseo@gachon.ac.kr (J.S.)

Abstract

This study develops a computational framework for quantifying how technological disruptions are associated with shifts in media–public discourse alignment. Using the public release of ChatGPT (30 November 2022) as a temporal breakpoint, we examine whether the broad public availability of generative AI was associated with structural changes in the relationship between media agendas and online public discourse in South Korea. The framework combines PPMI-weighted semantic network construction with QAP correlation, MRQAP regression and Cohen's q effect-size analysis, applied to 181,081 Korean-language texts encompassing media agendas (national newspapers, economic dailies, regional newspapers, and broadcast news) and public agendas (online communities) over six years (2019–2025). Results reveal that national newspapers lost their dominant agenda-setting position, with public alignment declining sharply (r: 0.678 → 0.430, Cohen's q = 0.369, large effect), while economic papers rose from lowest to second-highest alignment (r: 0.352 → 0.567) by addressing market and industry dimensions that direct AI experience could not supply. Broadcasting emerged as the dominant structural anchor of public discourse (unique coefficient β: 0.263 → 0.569; its removal alone lowers model fit from R^2 = 0.517 to 0.347), while national newspapers' unique contribution reversed in sign. Collective media explanatory power itself remained essentially stable (R^2 = 0.537 → 0.517), indicating a structural reconfiguration of which media align with public discourse rather than a wholesale weakening of media–public alignment—consistent with agenda-melding, in which publics integrate media coverage with firsthand technological experience. A residual analysis further shows that the variance media agendas leave unexplained is not noise but a structured, public-specific layer of discourse organized around the hands-on use of generative-AI tools (e.g., ChatGPT, image generation)—direct evidence of a bounded public autonomy in which public discourse is distinct from, and not reducible to, media agendas. These findings demonstrate the framework's utility for detecting how publicly accessible AI adoption—exemplified by ChatGPT—is associated with shifts in media–public structural dynamics within discourse ecosystems and carry implications for computational social science, technology communication, and applied network analysis.

Keywords: experiential accessibility AI; QAP; MRQAP; agenda-melding; ChatGPT

Check for updates

Academic Editor: Zhen Zhang

Received: 13 April 2026
Revised: 28 June 2026
Accepted: 9 July 2026
Published: 16 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Understanding how public discourse structures evolve in response to technological disruptions is a critical challenge for computational social science. The rapid proliferation of generative artificial intelligence (AI) systems—exemplified by ChatGPT, which achieved 100 million monthly active users within two months of its November 2022 release [1]—provides a natural setting for examining how technological accessibility is associated with structural shifts in discourse ecosystems.

Traditional approaches to studying media–public discourse relationships have relied primarily on survey-based methods, content analysis with manual coding, or keyword frequency comparisons [2,3]. While these methods have yielded important theoretical insights, they face scalability limitations when applied to large-scale, multi-domain discourse ecosystems. Recent advances in computational text analysis and network science offer methodological alternatives capable of capturing structural patterns across hundreds of thousands of texts [4,5].

This study addresses a dual gap. First, from a methodological standpoint, few studies have combined semantic network construction with permutation-based matrix regression (QAP/MRQAP) to quantify multi-domain discourse alignment at scale. While QAP has been applied in organizational network research [6] and intermedia agenda studies [7], its integration with PPMI-weighted semantic networks and systematic effect-size reporting remains underdeveloped. Second, from an applied standpoint, the emergence of generative AI creates unprecedented conditions for studying discourse transformation: a technology that simultaneously serves as an object of discourse and a medium through which publics generate discourse, thereby altering the information dependency structures that underlie traditional media influence [8,9].

South Korea provides an ideal empirical context for this investigation. The country ranks second globally in paid ChatGPT subscriptions [10], maintains over 95% smartphone penetration, and features a diverse media landscape encompassing national newspapers, economic dailies, regional papers, and broadcasting networks alongside vibrant online communities [11]. These conditions yield sufficient textual volume and domain diversity to stress-test computational frameworks for discourse network comparison.

Conceptually, we approach the media–public discourse environment as an interconnected system in which structural perturbations—such as the sudden public availability of generative AI—propagate across domains, reshaping alignment patterns among system components (media types, online publics) in ways that cannot be captured by examining any single domain in isolation.

The contributions of this study are threefold: (1) we formalize a reproducible computational pipeline for constructing, comparing, and regressing multi-domain discourse networks; (2) we introduce systematic effect-size quantification (Cohen's q) alongside permutation-based significance testing to enable substantive interpretation of structural shifts; and (3) we apply this framework to a large-scale empirical case demonstrating how the broad public availability of generative AI was associated with the reconfiguration of media–public discourse alignment in South Korea.

To contextualize the methodological scale of this study, Table 1 compares the present work with prior NAS studies employing QAP-based network comparison.

As shown in Table 1, the present study surpasses prior NAS-QAP research by an order of magnitude in corpus size, employs multi-period comparison with a temporal breakpoint, and introduces effect-size quantification—addressing key methodological gaps in the existing literature.

Table 1. Comparison of the present study with prior NAS-QAP studies in terms of data scale and analytical scope.

Study	Corpus Size	Domains	Time Span	Periods	Method	Effect Size
Liu et al. (2022) [7]	~42,500 texts	3	1 year	1	QAP	No
Cheng (2016) [12]	~2500 texts	3	6 months	1	QAP	No
Wu and Guo (2020) [13]	~1200 texts	3	3 months	1	QAP	No
Guo and Vargo (2015) [14]	~6000 texts	2	1 year	1	QAP	No
Present study	181,081 texts	5	6 years	2	QAP + MRQAP	Cohen's q

The remainder of this paper is organized as follows. Section 2 reviews the theoretical background and related computational methods. Section 3 formalizes the proposed analytical framework. Section 4 describes data collection and preprocessing. Section 5 presents results. Section 6 discusses implications for computational discourse analysis and communication research. Section 7 concludes with limitations and future directions.

2. Theoretical Background and Related Work

2.1. Agenda-Setting Theory and Network Extensions

Agenda-setting theory posits that media coverage patterns shape the salience hierarchies through which publics prioritize issues [2]. Five decades of empirical research have validated this effect across political, social, and international domains [3,15]. The theory's core mechanism assumes information dependency: when publics lack direct experiential access to issues, they rely on media for issue awareness and interpretation [16].

Network Agenda-Setting (NAS) extended this framework from issue salience to structural relationships, conceptualizing agenda influence as the transfer of relational patterns—how issues and attributes are interconnected—between media and audiences [17,18]. NAS draws on cognitive network models [19], arguing that people store information as associative webs rather than isolated facts. When media emphasize specific relationships among concepts, audiences reproduce similar cognitive structures. Studies have validated NAS correlations in diverse contexts: U.S. elections [14], environmental issues in China [12], and Korean media–public discourse [20].

However, agenda-setting research has focused predominantly on domains characterized by high information asymmetry—political campaigns, foreign policy, abstract social issues—where publics lack direct experiential access [16,21]. The emergence of experiential technologies such as generative AI disrupts these conditions: publics can directly access, evaluate, and discuss AI capabilities, potentially reducing media dependency and enabling autonomous discourse construction. We refer to this property—the degree to which a public can directly use and evaluate a technology rather than encountering it only through mediated coverage—as *experiential accessibility*, and it is the central condition this study examines. Shaw et al. [22] proposed agenda-melding theory to capture this dynamic, suggesting that individuals actively construct personal agendas by integrating information from diverse sources—mass media, social media, and personal experience—rather than passively receiving media-defined salience hierarchies.

Digital environments amplify this process. Online platforms enable access to expert commentary, peer discussions, and alternative frames while weakening traditional gatekeeping [23,24]. Information networks have become horizontal, where media, experts, and publics jointly shape discourse [25]. Within this setting, agenda alignment may emerge not from media primacy but from dynamic negotiation among multiple communicators. Re-

cent studies document agenda divergence when social media construct independent issue networks [26], reflecting shifting power relations in hybrid communication environments.

2.2. AI Discourse and Framing Dynamics

Generative AI represents a transformative paradigm that extends beyond automation and prediction toward creative cognition and knowledge production [8,9]. Wessel et al. [27] emphasize that generative AI redefines value creation on digital platforms through intelligent automation, democratization, hyper-personalization, and collaborative innovation. Korinek and Suh [28] characterize transformative AI as capable of performing most cognitive tasks at or above human level, with far-reaching economic implications. Despite its disruptive potential, assessments of AI's actual productivity or social effects remain mixed. Acemoglu [29] reports limited macro-level productivity gains, and Gimbel et al. [30] find little immediate labor-market disruption at the macro scale.

Media framing studies reveal consistent optimism bias: international coverage often prioritizes innovation, business, and growth narratives over ethics or regulation [31–33]. Research in science communication has long emphasized that public engagement with emerging science and technology is shaped as much by how issues are framed as by the underlying facts [34]. Moreover, Korea's AI discourse carries distinctive historical and policy dimensions. The 2016 match in which DeepMind's AlphaGo defeated Lee Sedol in Seoul—a landmark demonstration of contemporary AI capabilities [35]—was watched by millions nationwide and constituted a formative cultural moment that established public awareness of AI well before the generative AI era. Subsequently, the Korean government launched a series of national AI strategies—including the AI National Strategy (2019) [36] and the Digital New Deal (2020) [37]—that positioned AI as a central pillar of economic policy and generated sustained media attention across all outlet types. These contextual factors suggest that Korean publics entered the ChatGPT era with pre-existing AI awareness shaped by both cultural events and policy discourse, making Korea a particularly informative context for testing how direct technological experience reshapes already-established media–public agenda relationships.

2.3. Computational Methods for Discourse Network Analysis

Semantic network analysis represents discourse as graph structures where nodes are terms and edges represent co-occurrence or associative relationships [4,38,39]. This approach enables quantitative comparison of discourse structures across domains and time periods, moving beyond keyword frequency analysis to capture relational patterns.

The Quadratic Assignment Procedure (QAP) addresses the fundamental statistical challenge of network comparison: observations are not independent because network ties involve dyadic relationships, violating standard parametric assumptions [6]. QAP employs permutation-based significance testing to assess whether observed structural similarity exceeds chance expectations [40]. Multiple Regression QAP (MRQAP) extends this to multivariate settings, enabling simultaneous assessment of multiple predictor networks while controlling for intercorrelations. The double semi-partialling (DSP) variant improves robustness against autocorrelation and multicollinearity common in network data [40]. Co-occurrence weighting schemes significantly affect network construction quality. Positive Pointwise Mutual Information (PPMI) has emerged as a preferred approach, emphasizing semantically meaningful rather than merely frequent co-occurrences by comparing observed co-occurrence probabilities against expected frequencies under independence assumptions [41,42].

2.4. Research Questions

Drawing on the theoretical and methodological foundations outlined above, this study addresses three research questions:

RQ1. How did the structural correlation between media discourse networks and public discourse networks regarding AI change before and after the diffusion of generative AI technologies?

RQ2. How did the relative (unique) structural association of different media types—national newspapers, economic papers, regional papers, and broadcasters—with public AI discourse networks vary across the two periods?

RQ3. How did the collective explanatory power of media discourse networks over public AI discourse networks shift following the spread of generative AI?

3. Analytical Framework

Quantifying how a technological disruption reconfigures media–public discourse relationships requires a framework that moves progressively from raw text to structural comparison. No single analytical technique can simultaneously capture semantic associations within a discourse domain, measure pairwise structural similarity between domains, and decompose the relative (unique) structural association of multiple media types while controlling for their intercorrelations. We therefore designed a four-stage computational pipeline in which each stage generates the specific data product required by the next: text preprocessing yields a common vocabulary, network construction transforms that vocabulary into comparable adjacency matrices, QAP correlation quantifies pairwise structural alignment, and MRQAP regression decomposes multivariate structural association.

Figure 1 provides an overview of this sequential workflow. The pipeline's design reflects three guiding principles. First, comparability: all discourse domains must be represented within an identical term space so that structural differences reflect genuine divergence rather than vocabulary mismatch. This motivates the common vocabulary construction in Stage 1 and the uniform PPMI weighting in Stage 2. Second, dyadic completeness: because agenda alignment is fundamentally a relational property—defined over pairs of terms, not individual terms—the framework operates on full adjacency matrices rather than reduced summary statistics, preserving the relational granularity required for permutation-based inference. Third, analytical layering: bivariate correlation (QAP) answers whether two networks are structurally similar, but cannot distinguish a media type's unique contribution from shared variance with other media types. MRQAP regression in Stage 4 addresses this limitation by estimating each media type's independent (unique) structural association with public discourse while controlling for collinearity among media networks. Together, these stages yield a progressively richer account: Stage 3 reveals which media types align with public discourse, and Stage 4 reveals which media types are uniquely associated with that alignment.

The addition of Cohen's q effect-size analysis (Equation (3)) at Stage 3 addresses a recurrent limitation in QAP-based studies: with thousands of dyadic observations, virtually all correlations achieve statistical significance, rendering p-values uninformative for distinguishing substantively meaningful shifts from trivial ones. By converting correlation changes to a standardized effect-size metric, the framework enables substantive interpretation of the magnitude—not merely the existence—of discourse realignment.

The subsections below formalize each stage in turn. Section 3.1 justifies the key methodological decisions, Sections 3.2 and 3.3 detail Stages 1–2 (preprocessing and network construction), Sections 3.4 and 3.5 formalize Stages 3 and 4 (QAP and MRQAP), and Section 3.6 describes supplementary network structural metrics.

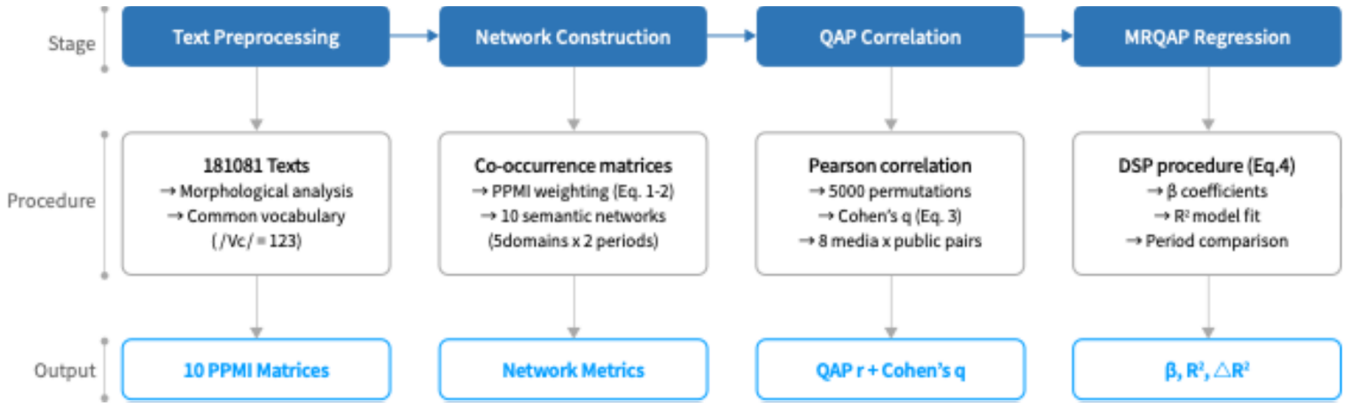


Figure 1. Overview of the four-stage computational pipeline for measuring structural shifts in the media–public discourse system. The Stage layer identifies each analytical phase, the Procedure layer specifies the operations performed, and the Output layer indicates the resulting data products. Arrows denote the sequential flow from text preprocessing through MRQAP regression.

3.1. Rationale for Methodological Choices

Before formalizing each stage, we briefly justify the key methodological decisions underlying this framework.

First, we employ co-occurrence-based semantic networks rather than word-embedding-based approaches (e.g., Word2Vec [43], BERT). While embeddings capture distributional semantics with high dimensionality, co-occurrence networks offer two advantages for this study: (a) they produce interpretable, directly comparable adjacency matrices suitable for QAP/MRQAP permutation testing, and (b) they preserve explicit relational structures between term pairs, which aligns with NAS theory's emphasis on associative linkages rather than latent semantic similarity [17,18]. Embedding-based similarity would require additional dimensionality reduction and matrix conversion steps that introduce analytic choices without clear theoretical motivation for agenda-network comparison.

Second, PPMI weighting is preferred over raw co-occurrence counts or TF-IDF-based alternatives because PPMI corrects for marginal frequency effects while emphasizing semantically meaningful associations [41,42]. Unlike TF-IDF, which is designed for document-level term importance, PPMI directly models pairwise associative strength—the relevant unit of analysis for semantic network construction.

Third, we use QAP and MRQAP rather than standard parametric tests because network data violate the independence assumption underlying conventional regression. Permutation-based inference provides valid significance estimates without distributional assumptions [6,40].

3.2. Text Preprocessing and Common Vocabulary Construction

Let $D = \{d_1, d_2, \ldots, d_N\}$ denote a corpus of N documents. Each document is segmented into sentences, and morphological analysis extracts content-bearing tokens (nouns, proper nouns, and compound nouns). From each discourse domain, the top-K most frequent terms are extracted ($K = 50$), and the common vocabulary is defined as the union of the domain- and period-specific top-K term sets (i.e., terms ranking among the top K in at least one of the ten domain–period sub-corpora), yielding $|V_c| = 123$ terms that define a shared analytical term space for cross-domain comparison; because this space is a union rather than an intersection, an individual term need not appear in every domain (see Section 4.3).

3.3. PPMI-Weighted Co-Occurrence Network Construction

For each discourse domain, we construct a co-occurrence matrix C where each element C_{ij} represents the frequency with which terms v_i and v_j co-occur within the same sentence. A minimum co-occurrence threshold ($\tau = 3$) is applied to reduce noise. To correct for frequency bias, raw counts are transformed into Positive Pointwise Mutual Information (PPMI) weights [41,44]. Let $p(i, j)$ denote the joint probability of terms i and j co-occurring, and $p(i)$, $p(j)$ their marginal probabilities:

$$p(i,j) = \frac{C_{ij}}{\sum_{i'} \sum_{j'} C_{i'j'}}, \quad p(i) = \frac{\sum_j C_{ij}}{\sum_{i'} \sum_{j'} C_{i'j'}} \tag{1}$$

The PPMI weight is then:

$$\mathrm{PPMI}(i,j) = \max\left(\log_2 \frac{p(i,j)}{p(i) \cdot p(j)}, 0\right) \tag{2}$$

The resulting PPMI matrix serves as the weighted adjacency matrix of the semantic network. To maintain interpretable network density, each node retains edges to its top-L most strongly associated terms (L = 15), producing a sparse, comparable network structure across all domains and periods.

3.4. QAP Correlation and Effect-Size Analysis

To assess structural similarity between two semantic networks represented by their adjacency matrices X and Y, we compute the Pearson correlation over corresponding off-diagonal cells using the Quadratic Assignment Procedure [6]. Significance is assessed through 5000 random permutations of the row and column indices of one matrix; the observed correlation is compared to the resulting permutation distribution to determine whether structural similarity exceeds chance expectations [40,45].

To quantify the substantive magnitude of correlation changes between periods—beyond statistical significance, which is virtually guaranteed with large numbers of dyadic observations—we compute Cohen's q [46]:

$$q = \frac{1}{2}\ln\frac{1+r_1}{1-r_1} - \frac{1}{2}\ln\frac{1+r_2}{1-r_2} \tag{3}$$

where r_1 and r_2 are QAP correlations from Period 1 and Period 2, respectively. Following Cohen's benchmarks [46], $|q| < 0.1$ is classified as small, $0.1 \leq |q| < 0.3$ as medium, and $|q| \geq 0.3$ as large.

3.5. MRQAP Regression Analysis

To partition the relative (unique) structural association of multiple media types with public discourse while controlling for intercorrelations among media networks, we employ MRQAP with the double semi-partialling (DSP) procedure [40]. The model specifies the public discourse network Y as a function of p media network predictors:

$$Y_{ij} = \beta_0 + \beta_1 X_{ij}^{(1)} + \beta_2 X_{ij}^{(2)} + \cdots + \beta_p X_{ij}^{(p)} + \epsilon_{ij} \tag{4}$$

where each observation (i, j) represents a dyadic co-occurrence relationship between terms i and j, X ^ (k) is the adjacency matrix for media type k, and ε_ij is the error term. The DSP procedure controls for collinearity among predictor matrices by semi-partialling each predictor and the dependent variable before computing test statistics; significance is assessed through 5000 permutations [40].

The model yields unstandardized coefficients (b), standardized coefficients ($*\beta**$), and the coefficient of determination (R^2) indicating the proportion of variance in public discourse structure collectively explained by all media networks. Changes in R^2 between periods quantify shifts in collective media explanatory power.

3.6. Network Structural Metrics

To characterize discourse network topology, we compute weighted degree centrality (sum of edge weights per node), network density (ratio of actual to maximum possible edges), weighted clustering coefficient [47], and network centralization (degree to which the network concentrates around a central hub) [48]. These metrics enable comparison of topological properties across domains and periods, independent of the QAP/MRQAP structural comparison.

4. Data Collection and Preprocessing

4.1. Corpus Description

We collected 181,081 Korean-language texts discussing artificial intelligence over six years (30 November 2019–20 August 2025), divided into two periods using ChatGPT's public release date (30 November 2022; based on GPT-3.5) as the temporal breakpoint:

- Period 1 (Pre-ChatGPT): 30 November 2019–29 November 2022 (N_1 = 44,486)
- Period 2 (Post-ChatGPT): 30 November 2022–20 August 2025 (N_2 = 136,595)

Media texts were collected from Big Kinds, Korea's comprehensive news database, covering 75 outlets across four media categories (Table 2). Online public discourse was collected from four major platforms—Naver Cafe, Naver Blog, Kakao Tistory, and Facebook (via Google Search API)—using the Textom web crawler. Posts from online communities were analyzed not as individual attitudes but as proxy indicators of collective public discourse reflecting agenda formation. Throughout this study, the term "public discourse" refers operationally to this online, platform-based discourse—digital traces generated on a small set of indexed platforms—rather than to the Korean public in a general or representative sense; such traces are shaped by platform-specific user populations, activity levels, and indexing and crawling criteria [49], and the corresponding limitations are addressed in Section 7.

The asymmetric growth rates—32-fold increase in online community discourse versus 1.28× to 3.5× growth across media categories—are consistent with disproportionate public engagement relative to media coverage expansion following generative AI diffusion, although they may also partly reflect platform-, indexing-, and data-availability factors rather than engagement alone (see Sections 6.2 and 7).

Table 2. Overview of data sources, collection methods, and volume changes.

Category	Period	Collection Channels	Collection Tool	Sample Size	Change (Post vs. Pre)
Online Communities (4 platforms)	Period 1	Naver Cafe, Naver Blog, Kakao Tistory Blog, Facebook (Google Search)	Custom crawling	2338	+3112% (×32)
	Period 2			75,079	
Broadcasting (5 networks)	Period 1	KBS, MBC, OBS, SBS, YTN	BigKinds database	1543	+254% (×3.5)
	Period 2			5476	
Economic Newspapers (13 outlets)	Period 1	Daehan Economic Daily, Maeil Business, Money Today, Metro Economy, Bridge Economy, Seoul Economy, Asia Economy, Aju Business Daily, Edaily, E-Today, Financial News, Korea Economic Daily, Herald Economy		21,286	+27.6% (×1.28)
	Period 2			27,163	

Table 2. *Cont.*

Category	Period	Collection Channels	Collection Tool	Sample Size	Change (Post vs. Pre)
Regional Newspapers (45 outlets)	Period 1	Gangwon Domin Ilbo, Gangwon Ilbo, Gyeonggi Shinmun, Gyeonggi Ilbo, Gyeongnam Domin Ilbo, Gyeongnam Shinmun, Gyeongnam Ilbo, Gyeongbuk Domin Ilbo, Gyeongbuk Maeil, Gyeongbuk Ilbo, Gyeongsang Ilbo, Gyeongin Ilbo, Gwangnam Ilbo, Gwangju Maeil, Gwangju Ilbo, Kookje Shinmun, Gumgang Ilbo, Kiho Ilbo, Namdo Ilbo, Daegu Shinmun, Daegu Ilbo, Daejeon Ilbo, Dongyang Ilbo, Maeil Shinmun, Mudeung Ilbo, Busan Ilbo, Saejeonbuk Shinmun, Yeongnam Ilbo, Ulsan Maeil, Ulsan Shinmun, Incheon Ilbo, Jeonnam Ilbo, Jeonla Ilbo, Jeonbuk Domin Ilbo, Jeonbuk Ilbo, Jemin Ilbo, Jeju Ilbo, Jungdo Ilbo, Jungbu Maeil, Jungbo Ilbo, Chungbuk Ilbo, Chungcheong Ilbo, Chungcheong Times, Chungcheong Today, Halla Ilbo	BigKinds database	10,823	+53% (×1.53)
	Period 2			16,563	
National Newspapers (12 outlets)	Period 1	Kyunghyang Shinmun, Kookmin Ilbo, Naeil Shinmun, Dong-A Ilbo, Munhwa Ilbo, Seoul Shinmun, Segye Ilbo, Asia Today, Chosun Ilbo, JoongAng Ilbo, Hankyoreh, Hankook Ilbo		8496	+44.9% (×1.45)
	Period 2			12,314	
Subtotal (Sample Size)	Period 1	44,486			+207.2% (×3.07)
	Period 2	136,595			
Total		181,081			

4.2. Preprocessing Pipeline

Text preprocessing followed a 12-stage pipeline: (1) encoding normalization (UTF-8), (2) HTML/URL removal, (3) morphological analysis with Korean POS tagging and noun extraction, (4) stopword removal, (5) synonym consolidation, (6) compound noun detection via statistical collocations, (7) document-frequency filtering (minimum DF ≥ 2), (8) per-domain frequency ranking (top $K = 50$), (9) common vocabulary construction ($|V_c| = 123$), (10) sentence-level co-occurrence matrix generation, (11) minimum co-occurrence thresholding ($\tau = 3$), and (12) PPMI weighting (Equations (1) and (2)).

All computations were implemented in Python 3.11 using NumPy 1.26 and SciPy 1.12 for sparse matrix operations. Vectorized probability transformations were applied in computing PPMI weights to enhance efficiency and ensure reproducibility. QAP correlation analyses and network visualizations were performed in UCINET 6.772 and NetDraw 2.182 [50]. MRQAP regression was implemented separately in Python (NumPy/SciPy), with coefficient significance assessed by Dekker double-semi-partialling permutation (5000 permutations); this MRQAP code is released with the reproducibility materials (see the Data Availability Statement).

4.3. Network Construction

Applying the pipeline described above, we generated 10 semantic networks: five discourse domains (community, broadcasting, economic, regional, national) × two temporal periods. Each network shares the common vocabulary of 123 nodes, with edges weighted by PPMI values and pruned to retain each node's top 15 strongest connections.

Table 3 presents the 123 common terms with their document frequency (DF, the number of the ten domain–period networks in which the term ranks among the top-50; DF therefore ranges from 1 to 10) and global term frequency (TF).

Table 3. The 123 terms comprising the unified analytical vocabulary (the union of the domain–period top-50 term sets). DF = number of the ten domain–period networks in which the term ranks among the top-50; Global TF = total term frequency across the corpus.

Rank	Term	DF	Global TF	Rank	Term	DF	Global TF	Rank	Term	DF	Global TF
1	Technology	10	73,646	42	COVID-19	5	4376	83	Cyber	2	55
2	Education	10	45,018	43	Online	5	3552	84	R&D	2	52
3	Data	10	40,579	44	Revolution	5	2403	85	Automatic	1	1003
4	Development	10	39,726	45	Telecom	5	1996	86	COVID	1	973
5	Industry	10	29,140	46	Kakao	5	1887	87	Homepage	1	495
6	Future	10	27,011	47	NVIDIA	4	5719	88	Deep	1	459
7	Research	10	25,621	48	Chat	4	4915	89	Internet	1	377
8	Project	10	22,206	49	Transformation	4	2707	90	Infrastructure	1	369
9	Science	10	21,961	50	Media	4	1841	91	Portal	1	338
10	Digital	10	21,714	51	Hub	4	1558	92	Blockchain	1	319
11	Information	10	19,450	52	Communication	4	1487	93	Commercialization	1	258
12	Samsung	10	13,605	53	Virtual	4	650	94	Recruitment	1	223
13	Semiconductor	10	12,989	54	Vision	3	1377	95	Deepfake	1	170
14	Smart	10	12,949	55	Agency	3	1317	96	Network	1	149
15	Startup	10	12,850	56	Venture	3	918	97	OpenAI	1	148
16	Utilization	9	16,913	57	Mobility	3	791	98	Engineering	1	142
17	Innovation	9	16,729	58	Participating	3	330	99	Ecosystem	1	127
18	System	9	14,016	59	Bot	3	326	100	Fusion	1	123
19	Platform	9	10,155	60	Pilot	3	310	101	User	1	119
20	Robot	9	8858	61	Voice	3	221	102	Job	1	116
21	Electronics	9	8550	62	Hynix	3	218	103	GenerativeAI	1	109
22	Naver	9	7300	63	Lab	3	204	104	Smartphone	1	107
23	ICT	9	6500	64	Generator	3	203	105	Electric	1	101
24	Analysis	8	7030	65	SK	3	189	106	NewDeal	1	100
25	Chatbot	8	6941	66	Meta	3	125	107	Automobile	1	97
26	Power	8	6543	67	Coding	2	3251	108	Operation	1	88
27	Market	7	15,793	68	ChatGPT	2	2473	109	Creative	1	73
28	Convergence	7	10,394	69	Computer	2	2573	110	POSCO	1	68
29	Video	7	8932	70	Next-generation	2	304	111	Mobile	1	66
30	Software	7	4684	71	LG	2	274	112	Galaxy	1	62
31	Application	6	11,022	72	Bigdata	2	207	113	Creation	1	57
32	Investment	6	10,375	73	Training	2	171	114	Strategy	1	45
33	Program	6	10,208	74	Demonstration	2	157	115	Prediction	1	40
34	Google	6	8991	75	Implementation	2	137	116	Leesedol	1	37
35	Hyundai	6	4071	76	YouTube	2	116	117	Net	1	37
36	Metaverse	6	1416	77	Tesla	2	83	118	Investors	1	36
37	Generative	5	12,086	78	Evaluation	2	79	119	Verification	1	32
38	Model	5	7457	79	Microsoft	2	69	120	Design	1	30

Table 3. *Cont.*

Rank	Term	DF	Global TF	Rank	Term	DF	Global TF	Rank	Term	DF	Global TF
39	Content	5	6751	80	iPhone	2	68	121	Powerhouse	1	30
40	Cloud	5	5557	81	Apple	2	66	122	Transition	1	30
41	Solution	5	5231	82	DeepSeek	2	64	123	AlphaGo	1	26

5. Results

5.1. Network Structural Properties

Several patterns emerge from the structural comparison (Table 4). First, all networks share the same underlying vocabulary of 123 terms, of which approximately 50 are active (non-isolated) in each domain—reflecting the common vocabulary construction procedure. Second, weighted centralization exhibits meaningful variation across domains and periods. National newspapers showed the largest decline in centralization (0.803 → 0.741), indicating a shift from concentrated hub-and-spoke discourse toward more distributed semantic structures. This is consistent with the declining QAP correlations reported below: as national newspapers' discourse became less centralized, their structural alignment with public discourse weakened. Third, the economic papers' hub node shifted from "data" in Period 1 to "technology" in Period 2, suggesting a thematic reorientation from data-centric to broader technology discourse. Fourth, total edge weight remained relatively stable across periods (0.474–0.484), indicating that the overall volume of semantic associations was preserved even as their structural distribution changed.

Table 4. Network structural metrics by domain and period. Active Nodes = number of terms with at least one non-zero PPMI connection (out of 123 common vocabulary terms). Mean/Max Node Strength = mean/maximum sum of PPMI edge weights per node. Wtd. Centralization = weighted Freeman centralization based on node strength. Hub Node = the node with the highest weighted degree in each network.

Domain	Period	Active Nodes	Total Weight	Mean Node Strength	Max Node Strength	Wtd. Centralization	Hub Node
Community	1	50	0.478	0.0191	0.0930	0.811	technology
Community	2	50	0.483	0.0193	0.0933	0.809	technology
Broadcasting	1	50	0.479	0.0192	0.1038	0.832	technology
Broadcasting	2	50	0.481	0.0192	0.1020	0.828	technology
Economic	1	50	0.477	0.0191	0.1309	0.872	data
Economic	2	49	0.482	0.0197	0.0917	0.802	technology
Regional	1	50	0.479	0.0192	0.0850	0.790	education
Regional	2	50	0.474	0.0190	0.1184	0.857	education
National	1	50	0.482	0.0193	0.0904	0.803	technology
National	2	50	0.484	0.0194	0.0707	0.741	technology

5.2. RQ1: Structural Correlation Shifts (QAP Analysis)

QAP correlation analysis revealed significant restructuring of media–public discourse alignment following ChatGPT's release (Table 5).

Table 5. QAP correlation coefficients and effect sizes.

Media Type	Period 1 r	Period 2 r	Δr	Cohen's q	Effect Size
National Papers	0.678 ***	0.430 ***	−0.248	0.369	Large
Economic Papers	0.352 ***	0.567 ***	+0.215	−0.275	Medium
Regional Papers	0.652 ***	0.499 ***	−0.153	0.238	Medium
Broadcasting	0.622 ***	0.665 ***	+0.043	−0.075	Small

Note. *** $p < 0.001$ based on 5000 permutations. $N = 7503$ dyadic observations per comparison. Cohen's q computed via Equation (3).

In Period 1, national newspapers exhibited the strongest correlation with public discourse ($r = 0.678$), followed by regional papers ($r = 0.652$), broadcasting ($r = 0.622$), and economic papers ($r = 0.352$). This hierarchy is consistent with traditional agenda-setting dynamics where broad-reach legacy media show stronger public alignment than specialized outlets.

Period 2 demonstrated substantial reconfiguration. Broadcasting emerged with the strongest correlation ($r = 0.665$), displacing national newspapers, which fell to fourth position ($r = 0.430$). This decline represents a large effect (Cohen's $q = 0.369$). Economic papers rose from fourth to second position ($r = 0.567$), a medium-sized increase ($q = -0.275$). Regional papers declined moderately ($q = 0.238$), while broadcasting showed negligible change ($q = -0.075$).

All correlations remained statistically significant ($p < 0.001$), indicating persistent structural relationships between media and public agendas. However, the effect-size analysis differentiates substantively meaningful shifts (national newspapers: large effect) from negligible changes (broadcasting: small effect)—a distinction obscured by significance testing alone given the large number of dyadic observations.

5.3. RQ2: Shifts in Unique Structural Association (MRQAP Analysis)

MRQAP regression analysis controlling for intercorrelations among media types revealed convergent evidence of restructuring in unique structural associations (Table 6).

Table 6. MRQAP regression coefficients by period.

Media Type	Period 1		Period 2		
	b (SE)	β	b (SE)	β	Δβ
National Papers	0.326 *** (0.018)	0.262	−0.214 *** (0.013)	−0.202	−0.464
Economic Papers	0.054 * (0.009)	0.052	0.173 *** (0.011)	0.191	+0.139
Regional Papers	0.282 *** (0.013)	0.273	0.156 *** (0.008)	0.234	−0.039
Broadcasting	0.281 *** (0.012)	0.263	0.512 *** (0.010)	0.569	+0.306

Note. * $p < 0.05$, *** $p < 0.001$; 5000 permutations. b = unstandardized coefficient, β = standardized coefficient, SE = permutation standard error.

In Period 1, the unique structural associations of the three legacy domains were relatively balanced—regional papers ($\beta = 0.273$), broadcasting ($\beta = 0.263$), and national papers ($\beta = 0.262$) each contributed comparable unique variance—whereas economic papers added little beyond the others ($\beta = 0.052$). In Period 2 the configuration became sharply concentrated: broadcasting dominated ($\beta = 0.569$), followed by regional ($\beta = 0.234$) and economic papers ($\beta = 0.191$), while national newspapers' unique coefficient reversed to a negative value ($\beta = -0.202$). All coefficients were significant under double-semi-partialling permutation tests (economic papers in Period 1 at $p < 0.05$; all other coefficients at $p < 0.001$);

node-bootstrap resampling (Section 5.6), however, showed the Period 2 coefficients to be far more robust than those of Period 1: in Period 2 the intervals excluded zero for broadcasting, economic, and national papers (regional being marginally positive), whereas in Period 1 only broadcasting's interval excluded zero—a contrast consistent with the much smaller Period 1 corpus.

National newspapers' simultaneous decline in bivariate correlation (Table 5) and the reversal of their unique regression coefficient to a negative value (Table 6) is consistent with a pronounced weakening of traditional elite-media agenda-setting in technological discourse following the diffusion of publicly accessible AI. The negative Period 2 coefficient is best read as a suppression effect rather than substantive opposition: national coverage overlaps heavily with broadcasting (r = 0.636), so once broadcasting is included national's unique contribution is absorbed—indeed, the coefficient returns to approximately zero when broadcasting is removed from the model (Section 5.6).

Taken together, the Period 2 coefficients describe a reconfiguration rather than a uniform attenuation: broadcasting's unique structural association strengthened markedly and economic papers' rose, regional papers' held roughly steady, and only national newspapers' moved against the others by reversing in sign. This redistribution—rather than an across-the-board weakening—is the structural signature of agenda-melding, in which publics blend multiple media framings with firsthand experience instead of tracking any single dominant agenda.

5.4. RQ3: Collective Explanatory Power (Model Fit)

The collective explanatory power of the media discourse networks remained essentially stable across periods ($R^2 = 0.537 \rightarrow 0.517$; $\Delta R^2 = -0.020$). The node-bootstrap intervals for the two values overlap almost entirely (Period 1: [0.27, 0.74]; Period 2: [0.26, 0.71]), so the small numerical difference is not statistically distinguishable; roughly half of the variance in public discourse structure remained unexplained by media agendas in both periods (46.3% in Period 1, 48.3% in Period 2). What changed was therefore not the overall magnitude of media–public alignment but its internal structure. The leave-one-media-out jackknife makes this concentration explicit: in Period 2, removing broadcasting alone reduces model fit from $R^2 = 0.517$ to 0.347—an approximately one-third drop—whereas removing any other single medium leaves R^2 between 0.49 and 0.50, so broadcasting's unique contribution is roughly six to eleven times that of any other domain. No medium was comparably dominant in Period 1, where omitting any single domain leaves R^2 between 0.50 and 0.54. Collective explanatory power was thus not weakened after ChatGPT but concentrated onto broadcasting—a structural reconfiguration of the media–public alignment system rather than a collapse of its overall coherence. Table 7 presents MRQAP model fit statistics for both periods. We return to this substantial unexplained variance, and characterize its public-specific structure, in Section 5.7.

Table 7. MRQAP model fit statistics.

Period	R^2	Adjusted R^2	*p*-Value	Dyadic Obs.	Unexplained Variance
Period 1 (Pre-ChatGPT)	0.537	0.537	<0.001	7503	46.3%
Period 2 (Post-ChatGPT)	0.517	0.517	<0.001	7503	48.3%
Change (ΔR^2)	−0.020	−0.020			+2.0 pp

Note. Significance based on 5000 permutations (MRQAP). R^2 = coefficient of determination; Adjusted R^2 = R^2 adjusted for the number of predictors; Dyadic Obs. = number of dyadic observations included in each model; Unexplained Variance = 1 − R^2; ΔR^2 = change in R^2 from Period 1 to Period 2; pp = percentage points.

The consistency of findings across correlation analysis (reordered alignments, with broadcasting displacing national newspapers) and regression coefficients (a shift from balanced unique structural associations toward broadcasting-centered concentration, with national newspapers reversing in sign), without any decline in collective explanatory power, provides robust, multi-method evidence that the post-ChatGPT change was a structural reconfiguration of the media–public system rather than a wholesale weakening of alignment.

5.5. Thematic Divergence Analysis

Examination of top-30 term rankings across discourse domains reveals thematic differentiation accompanying structural realignment. In Period 1, all five domains shared core terms—"Technology," "Education," "Data," "Research," "Development"—at top-10 positions, reflecting convergent discourse structures. The network visualizations (Figure 2) show concentrated interconnections among core nodes, with high centralization as quantified in Table 4. This convergence is consistent with classical agenda-setting dynamics: when publics lacked direct AI experience, media coverage served as the primary interpretive resource, producing structurally homogeneous discourse across domains.

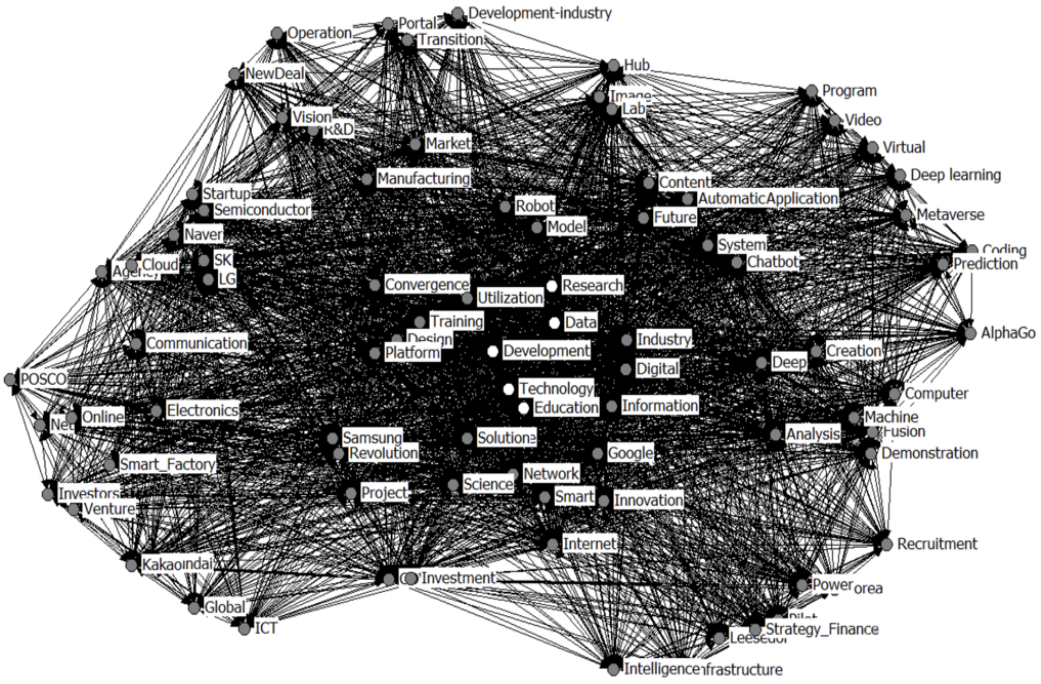


Figure 2. Semantic network visualization of the top 30 co-occurring terms across all five discourse domains during Period 1.

In Period 2 (Figure 3), discourse domains diverged thematically along functionally distinct axes. Community discourse uniquely elevated experiential terms ("ChatGPT," "OpenAI," "Generative," "Job") into top-10 positions, reflecting direct user engagement with AI systems. Notably, "ChatGPT" and "OpenAI"—terms directly referencing the technology and its developer—appeared exclusively in community top-10 rankings but not in any media domain's top-10, underscoring the experiential gap between media framing and public engagement. Economic papers emphasized industry-specific terms ("Semiconductor," "Investment," "Market," "Samsung"), constructing discourse around commercial implications and supply chain dynamics that required specialized journalistic expertise beyond individual user experience. Broadcasting highlighted accessibility terms ("NVIDIA," "Chatbot," "Google"), consistent with its role in translating technical developments for gen-

eral audiences. National newspapers increased policy-oriented terms ("Digital," "Project," "Platform"), yet these abstract governance frames showed weaker resonance with publics whose concerns had shifted toward practical, experiential dimensions. Regional papers maintained education-centric discourse ("Education," "University," "Student"), reflecting localized institutional concerns.

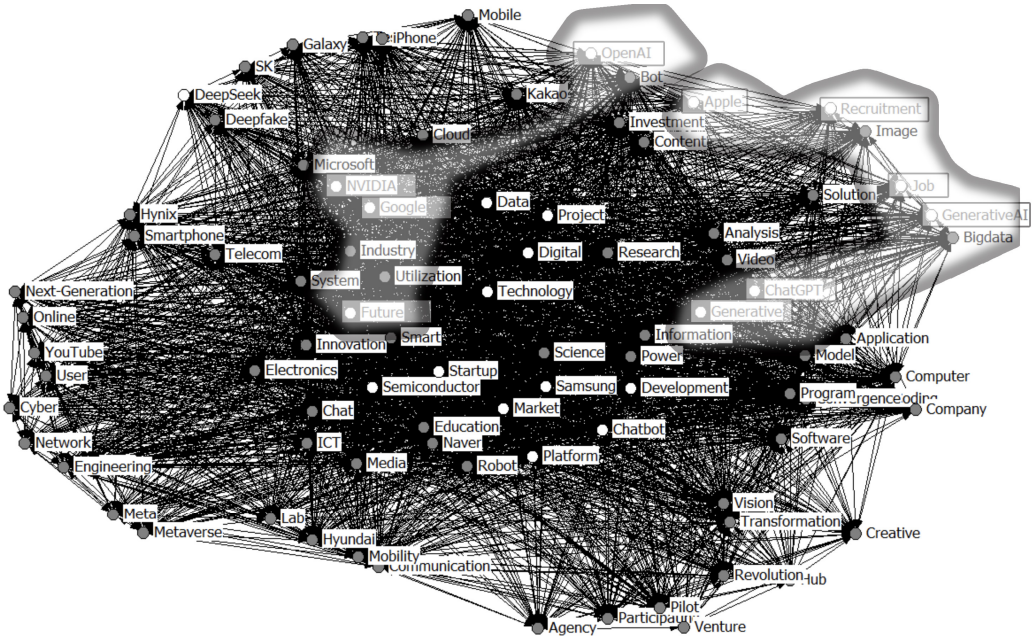


Figure 3. Semantic network visualization of the top 30 co-occurring terms across all five discourse domains during Period 2.

This thematic divergence maps onto the QAP findings reported in Section 5.2. The domains whose top-30 vocabularies diverged most from community discourse—national newspapers with policy abstraction and regional papers with educational focus—exhibited the largest declines in structural correlation (Cohen's $q = 0.369$ and 0.238, respectively). Conversely, economic papers, whose industry-oriented vocabulary addressed informational needs complementary to direct experience, showed increasing alignment ($\Delta r = +0.215$). Broadcasting's thematic stability—maintaining accessible framing across both periods—corresponds to its negligible correlation change ($q = -0.075$).

This pattern suggests that thematic heterogeneity and structural realignment are not independent phenomena but mutually reinforcing dimensions of discourse ecosystem reconfiguration. As publics developed autonomous experiential frameworks, media domains that adapted by providing orthogonal expertise maintained or increased alignment, while those that persisted in abstract or general framing lost structural resonance.

5.6. Sensitivity and Robustness Considerations

To assess the robustness of the findings, we examined the sensitivity of QAP correlations to the edge pruning parameter L (the number of top-weighted edges retained per node). The main analyses used $L = 15$ (top 15 strongest connections per node); we re-examined QAP correlations with stricter pruning at $L = 10$ and more permissive pruning at $L = 20$ (Table 8).

The rank ordering of media–public correlations was perfectly preserved across all L values: in Period 2, broadcasting maintained the highest correlation, followed by economic papers, regional papers, and national newspapers (broadcasting > economic > regional > national), regardless of pruning intensity. With $L = 10$, absolute correlation values

decreased slightly (by 0.01–0.04), but the direction and relative magnitude of all Period 1–to–Period 2 changes remained consistent. With $L = 20$, results were virtually identical to the original analysis.

Table 8. Sensitivity of QAP correlations (media × community) to edge pruning parameter L.

Media Type	Period	Original	$L = 10$	$L = 20$
Broadcasting	1	0.622	0.612	0.622
Broadcasting	2	0.665	0.631	0.653
Economic	1	0.352	0.336	0.352
Economic	2	0.567	0.540	0.556
Regional	1	0.652	0.656	0.652
Regional	2	0.499	0.472	0.495
National	1	0.678	0.684	0.676
National	2	0.430	0.389	0.429

These robustness checks confirm that the core findings—declining national newspaper alignment, rising economic paper alignment, and stable broadcasting correlation—are not artifacts of the specific edge pruning parameter.

To assess whether the QAP and MRQAP results depend on the particular composition of the 123-term analytical vocabulary or on any single media predictor—two robustness questions that can be addressed directly from the post-PPMI adjacency matrices—we conducted two further resampling-based checks. First, to evaluate sensitivity to the specific term set, we applied node-level bootstrap resampling: on each of 1000 iterations we drew a sample of the 123 vocabulary nodes with replacement, restricted every domain–period network to the resampled nodes, and recomputed both the media–public QAP correlations and the MRQAP regression coefficients, yielding percentile-based stability intervals and the proportion of replicates that preserved the observed Period 1-to-Period 2 ordering. Second, to evaluate whether the MRQAP coefficient pattern is driven by any one predictor, we performed leave-one-media-out jackknife re-estimation, refitting the multiple-regression QAP model with each media network omitted in turn and recording the change in the retained coefficients. Across the 1000 bootstrap replicates, all four media–public QAP correlations had 95% percentile intervals that excluded zero in both periods, and the direction of every Period 1-to-Period 2 change was preserved in the large majority of replicates (national decline in 99.8%, economic rise in 96.0%, regional decline in 91.2%); the broadcasting change, being small ($\Delta r = +0.043$), was directionally stable in 60.5% of replicates, and the full four-way correlation ranking—reflecting overlapping intervals—was preserved in 46.8%, so the QAP evidence is reported at the level of these directional changes rather than the exact rank order. For the MRQAP coefficients, the node-bootstrap 95% intervals in Period 2 excluded zero for broadcasting ([0.31, 0.71]), economic papers ([0.02, 0.63]), and national papers ([−0.37, −0.04]); for regional papers the interval was almost entirely positive but its lower bound fell marginally below zero ([−0.00, 0.40], with 97% of replicates positive). In Period 1, by contrast, only broadcasting's interval excluded zero ([0.05, 0.69]), with the regional, national, and economic intervals all spanning zero—an asymmetry consistent with the substantially smaller Period 1 corpus. The double-semi-partialling permutation tests that provide the significance reported in Table 6 indicate all four Period 2 coefficients are significant; the node bootstrap is a deliberately conservative, high-variance check, and the marginal regional interval should be read in that light. The leave-one-media-out jackknife confirmed broadcasting's centrality in Period 2: omitting

broadcasting lowered the model R^2 from 0.517 to 0.347, whereas omitting any other single medium left R^2 between 0.49 and 0.50; in Period 1 no medium was comparably influential (R^2 between 0.50 and 0.54 across omissions). National's negative Period 2 coefficient was contingent on broadcasting's inclusion, returning to approximately zero when broadcasting was removed, consistent with a suppression rather than an oppositional interpretation. The procedures and code for both checks are available with the other reproducibility materials (see the Data Availability Statement).

The present sensitivity analysis deliberately focuses on the edge-pruning parameter L, together with the node-level bootstrap and leave-one-media-out jackknife checks reported above, because this is the analytical stage at which the available data product—the post-PPMI, domain–period adjacency matrices—can be perturbed without reconstructing the upstream pipeline. Robustness to other modeling choices that act earlier in the pipeline—the vocabulary cutoff K, the PPMI association threshold τ, the co-occurrence window width, alternative vocabulary-construction rules, and sample-size-balancing procedures—cannot be assessed from the post-PPMI matrices alone, as each requires re-deriving the co-occurrence statistics from the raw corpus. We therefore flag these additional checks, together with temporally and platform-disaggregated subsample analyses, as important directions for future validation rather than as evidence already in hand (see Section 7). The convergence of QAP, MRQAP, and Cohen's q evidence across the structural patterns reported above should accordingly be read as internally consistent under a fixed construction procedure, not as having been stress-tested against every reasonable alternative specification.

5.7. Characterizing the Unexplained Variance: Public-Specific Discourse Structure

The model-fit analysis in Section 5.4 left roughly half of the structure of the public discourse network unexplained by the four media networks (R^2 = 0.517 in Period 2; 48.3% of variance unexplained). A natural question is whether this unexplained portion is merely noise or whether it has a structure of its own—a public-specific layer of discourse not reducible to media co-occurrence. To examine this, we mapped the residuals of the MRQAP regression (the part of each community dyad not predicted by the media networks) back onto the 123 terms, summing each term's incident residuals to obtain a net public-specific emphasis score, and we compared each term's strength-centrality in the public network with its centrality in the media networks. A node-label permutation test (500 permutations) assessed whether the residual was more concentrated on particular terms than chance would produce.

The residual was significantly structured in both periods (permutation p = 0.002), indicating that the variance the media do not explain is not diffuse noise but is organized around a specific set of terms. Moreover, the terms that anchor the public network differ systematically from those that anchor the media networks (Table 9; Figure 4). In Period 2, the most central terms specific to public discourse were the experiential objects of generative AI itself—ChatGPT (the third most central term in the public network, but only 27th in broadcasting and not active among the top-ranked terms of the economic and national networks), generative, OpenAI, application, and utilization; the term image (image generation) was active only in the public network. By contrast, the terms most central to the media networks but peripheral in public discourse were corporate and hardware referents—semiconductor (third in broadcasting, seventh in the national press), Samsung, NVIDIA, and Apple. The public and the media thus emphasized different facets of the same phenomenon: the media foregrounded the AI industry, while the public foregrounded the hands-on use of AI tools.

Table 9. Strength-centrality rank of selected terms in the public network versus the four media networks (Period 2). Rank 1 denotes the most central term; a dash (—) denotes a term not active among the top-ranked nodes of that network. Terms are grouped by whether public or media discourse emphasizes them more.

Term	Public	Broadcasting	Economic	Regional	National
Public-specific terms (more central in public than in media discourse)					
ChatGPT	3	27	—	38	—
Utilization	8	22	20	16	33
Generative	9	28	34	26	50
Application	10	—	42	35	—
Image	11	—	—	—	—
OpenAI	13	30	—	—	—
Chatbot	31	—	40	36	22
Media-specific terms (more central in media than in public discourse)					
Semiconductor	21	3	18	24	7
Samsung	20	5	21	40	10
NVIDIA	30	10	48	—	30
Apple	42	21	—	—	—

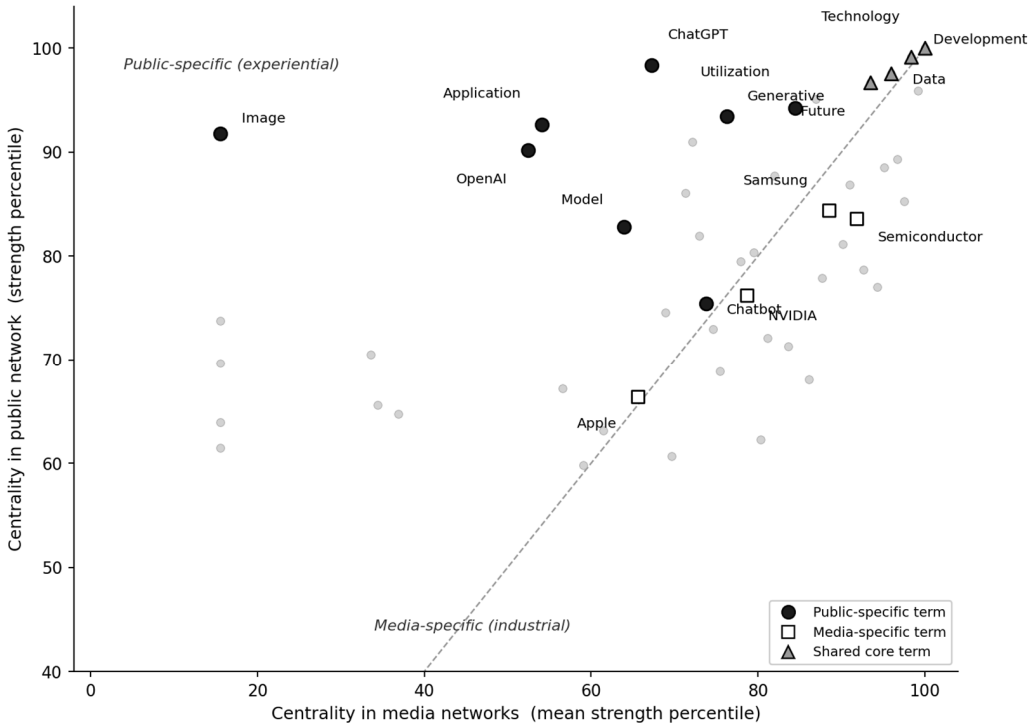


Figure 4. Centrality of terms in the public discourse network versus the media networks (Period 2). Each point is a term; the horizontal axis is its mean strength-centrality percentile across the four media networks and the vertical axis its percentile in the public network. Points above the diagonal are more central in public discourse than in the media. The experiential objects of generative AI (ChatGPT, generative, OpenAI, image, application, utilization) cluster above the diagonal, whereas corporate and hardware referents (semiconductor, Samsung, NVIDIA) fall on or below it.

This public-specific layer also changed in character across the two periods. In Period 1, the terms the public emphasized beyond the media were oriented toward learning and

labor—program, coding, education, and recruitment—consistent with AI being encountered largely as a topic to be studied or a skill to acquire. In Period 2, the same layer reorganized around direct tool use (ChatGPT, generative, chatbot, image, utilization). The distinctive content of public discourse shifted, in other words, from learning about AI to using it.

These results give direct, measured support to the study's central proposition. The shift toward experiential terms in the public-specific layer operationalizes experiential accessibility not as an assumption but as an observable property of the discourse: once the public could interact with generative AI directly, it developed a coherent body of discourse—organized around the tools and their use—that the media did not equally articulate. This is also consistent with a functional respecialization of media and public roles rather than a wholesale weakening of media influence, with the public and the media occupying complementary, not competing, regions of the same semantic space. The interpretive limits should be stated plainly. Greater centrality of a term in the public network shows that public discourse is structurally distinct from, and not reducible to, the media agendas—a bounded form of public autonomy—but it does not establish that the public originated these terms or led the media in time; the cross-sectional design measures distinctiveness, not temporal precedence. Establishing the latter would require time-resolved, lead–lag or interrupted time-series analysis, which we identify among the priorities for future work (Section 7).

6. Discussion

6.1. Framework Validity and Methodological Contributions

The results demonstrate the proposed computational framework's capacity to detect and quantify multi-dimensional discourse reconfigurations. Three levels of analysis—bivariate structural correlation (QAP), multivariate decomposition of unique structural associations (MRQAP), and aggregate explanatory power (R^2)—converge on consistent findings, providing triangulated evidence for discourse ecosystem transformation.

The introduction of systematic effect-size reporting via Cohen's q (Equation (3)) adds substantive interpretability beyond statistical significance. While all QAP correlations remained significant ($p < 0.001$)—unsurprising given 7503 dyadic observations—effect-size analysis differentiates substantively meaningful shifts (national newspapers: $q = 0.369$, large) from negligible changes (broadcasting: $q = -0.075$, small). This distinction is critical for applied discourse analysis where significance testing alone provides insufficient guidance for interpretation.

The 12-stage preprocessing pipeline, PPMI-weighted network construction, and permutation-based inference collectively provide a reproducible toolkit applicable to discourse ecosystems beyond the specific case of AI communication. The framework's scalability—demonstrated here with 181,081 texts across five domains—positions it as a viable approach for monitoring large-scale discourse dynamics in computational social science. The sensitivity analyses reported in Section 5.6 indicate that the core findings are stable across reasonable variations in the edge pruning parameter (L), although robustness to the other construction parameters discussed in Section 7—including the vocabulary cutoff K, the PPMI threshold τ, and the co-occurrence window—remains to be established in future validation; the node-level bootstrap and leave-one-media-out jackknife analyses reported in Section 5.6 do, however, address robustness to the composition of the term set and to the removal of any individual media predictor.

6.2. Discourse Ecosystem Reconfiguration

The structural changes documented above are consistent not with a dissolution of the media–public hierarchy but with its reconfiguration. The collective explanatory power of media remained essentially stable (Table 7); what changed was the apex of alignment, which shifted from the national press to broadcasting (Tables 5 and 6). Authority was therefore relocated rather than eliminated: a small number of sources continued to structure public understanding, but the identity of the structuring medium changed, while the remaining media contributed partial, non-redundant perspectives that publics integrated selectively.

Economic papers strengthened on both measures: their bivariate alignment with public discourse rose markedly (r: 0.352 → 0.567) and their unique structural association rose in parallel (β: 0.052 → 0.191), moving from a negligible to a significant independent contributor. This convergent increase is consistent with the growing salience of the industrial and market dimensions of AI in public discussion after ChatGPT—commercial implications, investment, and applications—an informational terrain that direct, individual AI experience does not by itself supply.

Broadcasting moved to the center of the media–public configuration: it held the strongest bivariate correlation in Period 2 (r = 0.665) and simultaneously became the dominant unique predictor (β: 0.263 → 0.569)—the only medium whose removal substantially reduces model fit, with R^2 falling from 0.517 to 0.347 (Section 5.6). This convergence of high correlation and high unique association indicates that, in the post-ChatGPT period, broadcasting discourse was the structural anchor most closely tracking public discourse. A plausible contextual mechanism is broadcasting's migration into the platform environment where public discourse itself unfolds: South Korea has the highest rate of news consumption via YouTube among major markets, and broadcasters' news channels dominate that space [51]. Because the cross-sectional design measures structural alignment rather than temporal precedence, this convergence is best interpreted as broadcasting and public discourse increasingly occupying the same communicative space, rather than as unidirectional influence in either direction.

The 32-fold increase in online-community texts from Period 1 to Period 2 (Table 2) has both substantive and methodological implications. Substantively, the larger and more heterogeneous Period 2 corpus—spanning AI professionals, casual users, educators, and job seekers—is consistent with a broadening of public discourse and with the integration of multiple framings rather than a single media agenda. Methodologically, the same asymmetry bears on estimate stability: because the Period 1 network is built from far fewer texts, its edge weights are estimated less precisely, which is reflected in the wider node-bootstrap intervals for the Period 1 coefficients (Section 5.6). The Period 1 individual coefficients should therefore be interpreted at the level of the overall media configuration rather than fine distinctions among media. Several features of the analytical design partially buffer the structural results against raw corpus-size effects: PPMI weighting normalizes co-occurrence counts against the marginal frequencies of individual terms, an identical analytical vocabulary is imposed on every domain–period network, and top-L edge pruning fixes the number of retained edges per node so that a larger corpus does not mechanically inflate network density. These safeguards attenuate but do not eliminate compositional effects, and we accordingly identify rarefaction, balanced month- or source-level sampling, and document-level bootstrap resampling of the community corpus as concrete robustness procedures for future work (Section 7).

A complementary perspective on this reconfiguration comes from the discourse that the media do not explain. As shown in Section 5.7, the variance unaccounted for by the media networks is not residual noise but a structured, public-specific layer organized around the experiential objects of generative AI—ChatGPT, image generation, and the

hands-on use of AI tools—whereas the media networks remain anchored by corporate and hardware referents such as semiconductors and chip makers. The public and the media thus came to occupy complementary, rather than competing, regions of the same semantic space: the media specialized in the industrial and policy dimensions of AI, while the public specialized in its direct use. This functional respecialization, rather than a simple weakening of media influence, is what the stable but internally redistributed explanatory power ultimately reflects.

6.3. Theoretical Implications

These findings extend agenda-setting theory by empirically demonstrating how technological accessibility conditions the media–public discourse relationship. Classical agenda-setting assumes information dependency [2,12]: when publics lack direct access to issues, media coverage determines salience and framing. Publicly accessible AI adoption—exemplified by ChatGPT's conversational interface reaching 100 million users—altered these dependency conditions by shifting AI from a primarily media-mediated topic toward a directly experiential technology.

The resulting pattern is consistent with agenda-melding theory [29], where publics actively integrate multiple information sources—media coverage, direct technological experience, peer discussion—to construct personalized interpretive frameworks. The stability of overall explanatory power alongside the internal redistribution of unique associations ($R^2 = 0.537 \rightarrow 0.517$, but with broadcasting ascendant and national newspapers reversing in sign) indicates not a wholesale retreat of media but a restructuring of which media align with public discourse; the continued significance of the media predictors confirms media's persistent role as one set of voices among multiple discourse sources. Media's function has shifted from dominant agenda-setter determining what people think about to agenda contributor offering one voice among many.

This has broader implications for the study of technology communication: agenda-setting effects may be conditional on what we term "experiential accessibility"—the degree to which publics can directly access and evaluate the phenomena under discussion. Publicly accessible AI adoption represents a paradigmatic case of high experiential accessibility, where natural-language interfaces eliminate technical barriers to direct engagement. Technologies that remain experientially inaccessible (e.g., quantum computing, nuclear fusion) may continue to exhibit strong traditional agenda-setting effects, while accessible technologies (e.g., conversational AI, social media platforms) may show attenuated media influence. Section 5.7 gives this concept direct empirical grounding: the terms unique to public discourse were precisely the experiential objects of the technology, indicating that experiential accessibility is not only an interpretive lens but a measurable property of the resulting discourse. Future research should develop typologies of technological experientiality to predict which communication domains will undergo agenda transformation.

6.4. Practical Implications

For media organizations, the findings suggest that maintaining relevance in experiential technology domains requires providing information orthogonal to direct user experience—investigating non-obvious implications, contextualizing within broader trends, analyzing policy and market dynamics, and facilitating public deliberation about contested meanings. Economic journalism's relative success demonstrates the viability of this complementary strategy. National newspapers' decline may reflect a failure to adapt equivalently: traditional opinion journalism and abstract social commentary provided less distinct value when publics could form independent technological assessments.

For computational social scientists, the PPMI–QAP framework offers a scalable, reproducible pipeline for monitoring discourse ecosystem dynamics applicable to diverse domains—health communication, climate discourse, political polarization—wherever multi-source textual data are available.

7. Conclusions

This study developed and validated a PPMI–QAP computational framework for quantifying how publicly accessible AI adoption is associated with structural shifts in media–public discourse networks. By integrating PPMI-weighted semantic network construction with QAP correlation, MRQAP regression, and Cohen's q effect-size analysis, the framework provides a multi-level analytical toolkit capable of detecting structural reconfigurations that conventional content analysis methods cannot capture at scale. Applied to 181,081 Korean-language texts spanning six years across five discourse domains, the framework revealed that ChatGPT's public release was associated with significant and multidimensional shifts in the discourse ecosystem. Four principal findings emerge. First, the traditional hierarchy of media–public alignment was substantially restructured: national newspapers lost their dominant agenda-setting position (Cohen's q = 0.369, large effect), while economic papers rose from lowest to second-highest alignment by providing specialized industry knowledge complementary to direct user experience. Second, the configuration of unique media–public associations was reconfigured rather than uniformly weakened: broadcasting became the dominant structural anchor of public discourse (β: 0.263 $\rightarrow$ 0.569), while national newspapers' unique contribution reversed in sign, even as the collective explanatory power of media remained essentially stable (R^2 = 0.537 $\rightarrow$ 0.517)—a pattern consistent with agenda-melding, in which publics integrate multiple media framings with firsthand technological experience rather than tracking a single dominant agenda. Third, thematic analysis revealed that discourse domains diverged along functionally distinct axes—experiential, commercial, policy-oriented, and educational—suggesting that homogeneous media agendas fragmented into heterogeneous, specialized discourse contributions when publics gained independent evaluative capacity. Fourth, the variance that media agendas left unexplained proved not to be noise but a structured, public-specific layer of discourse: it was organized around the hands-on use of generative-AI tools and evolved from a focus on learning about AI before ChatGPT to direct tool use afterward, providing structural evidence that the public maintained an experience-anchored agenda distinct from, and not reducible to, any single media agenda—a bounded form of public autonomy.

These findings make three contributions to computational social science and communication research. Methodologically, the PPMI–QAP framework offers a scalable, reproducible pipeline for monitoring discourse dynamics across multi-source textual ecosystems, with systematic effect-size reporting that distinguishes substantively meaningful shifts from statistically significant but trivial changes. Theoretically, the results extend agenda-setting theory by demonstrating that media influence is conditional on experiential accessibility: when publics can directly engage with the phenomena under discussion, the media–public relationship shifts from agenda-setting to agenda-melding, with media serving as one contributor among many rather than the dominant frame provider. Empirically, the South Korean case provides granular evidence of how a specific technological development—the public release of ChatGPT—was associated with reconfiguration of discourse structures across an entire media ecosystem within a compressed timeframe.

Several limitations should be noted. First, the two-period design, while leveraging a natural temporal breakpoint, cannot definitively rule out confounding factors such as concurrent policy changes, pandemic effects, or other technological developments during the

observation period. Both windows encompass multiple milestones—new model releases, regulatory debates, and evolving public concerns—that may have produced short-term oscillations in agenda alignment. Future studies could employ time-series or rolling-network approaches to distinguish stable structural trends from transient reactions. In particular, interrupted time-series models, rolling-window estimation, and placebo breakpoint tests at dates prior to the actual ChatGPT release (e.g., arbitrary 2020 or 2021 cut points) would help confirm that the observed structural shifts are specifically aligned with the release rather than reflecting a pre-existing secular trend. Second, PPMI-based networks capture associative co-occurrence relationships but not semantic directionality or sentiment valence; integrating sentiment analysis or causal inference techniques could enrich the framework's interpretive capacity. Third, the study is situated within South Korea's specific media ecology, characterized by high digital connectivity, vibrant online community platforms, and distinctive media consumption patterns; cross-national replication, particularly in societies with different media systems and varying levels of AI adoption, would strengthen generalizability [52,53]. Fourth, the common vocabulary approach, while enabling rigorous cross-domain comparison, may exclude domain-specific terms important for within-domain dynamics. Fifth, the "public" is not monolithic: online ecosystems likely encompass heterogeneous subcommunities—from AI professionals to casual users—each engaging differently with media narratives and direct technological experience. Relatedly, the online discourse analyzed here consists of platform-based digital traces drawn from a small set of indexed services (Naver Cafe, Naver Blog, Kakao Tistory, and Facebook via search-API retrieval); these traces are shaped by platform-specific user populations, algorithmic ranking, and crawling and indexing criteria, and should not be interpreted as a statistically representative sample of the Korean public. Sixth, the online-community corpus grew substantially and shifted in composition across the two periods (Table 2), so part of the observed change in public-side network structure may reflect changes in who participates rather than changes in agenda dependence alone; rarefaction, balanced month- or source-level sampling, and bootstrap resampling are needed to fully separate these effects. The bootstrap resampling referred to here operates at the level of the community corpus and is distinct from the node-level bootstrap and leave-one-media-out jackknife already reported in Section 5.6: the latter establish robustness to the composition of the term set and to the removal of individual media predictors, but they do not address corpus-size or compositional change, which remains the specific concern for future work.

These limitations point toward productive future research directions. Temporal network analysis incorporating rolling windows or event-triggered segmentation could capture within-period dynamics obscured by the binary periodization employed here. Word-embedding-based similarity measures could complement co-occurrence networks by capturing latent semantic relationships beyond surface-level term associations. Community detection algorithms applied to public discourse networks could identify whether agenda-melding concentrates among technically literate subcommunities or operates as a general population phenomenon. Cross-national comparative studies applying this framework to societies with different media systems, digital infrastructure levels, and AI adoption rates would test the generalizability of the experiential accessibility hypothesis. As generative AI systems become increasingly embedded in daily life, the relationship between media discourse and public understanding will continue to evolve. The framework developed here provides computational tools for tracking these transformations systematically, offering both scholars and media practitioners empirical foundations for understanding how publicly accessible technologies reshape the discourse ecosystems through which societies negotiate the meaning and implications of technological change.

Author Contributions: Conceptualization, H.H. and J.S.; methodology, J.S.; software, H.H.; validation, S.C. and H.H.; formal analysis, J.S.; investigation, J.S.; resources, S.C., H.H. and J.S.; data curation, J.S.; writing—original draft preparation, H.H.; writing—review and editing, S.C.; visualization, J.S.; supervision, J.S.; project administration. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: All data generated or analyzed in the study are included in the article. Further inquiries can be directed to the corresponding author due to privacy.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Hu, K. ChatGPT Sets Record for Fastest-Growing User Base. *Reuters*, 2 February 2023. Available online: https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/ (accessed on 13 February 2026).
2. McCombs, M.E.; Shaw, D.L. The Agenda-Setting Function of Mass Media. *Public Opin. Q.* **1972**, *36*, 176–187. [CrossRef]
3. McCombs, M. *Setting the Agenda: Mass Media and Public Opinion*, 2nd ed.; Polity Press: Cambridge, UK, 2014.
4. Diesner, J.; Carley, K.M. Revealing Social Structure from Texts: Meta-Matrix Text Analysis as a Novel Method for Network Text Analysis. In *Causal Mapping for Information Systems and Technology Research*; Narayanan, V.K., Armstrong, D.J., Eds.; IGI Global: Hershey, PA, USA, 2005; pp. 81–108. [CrossRef]
5. Guo, L. The Application of Social Network Analysis in Agenda Setting Research: A Methodological Exploration. *J. Broadcast. Electron. Media* **2012**, *56*, 616–631. [CrossRef]
6. Krackhardt, D. Predicting with Networks: Nonparametric Multiple Regression Analysis of Dyadic Data. *Soc. Netw.* **1988**, *10*, 359–381. [CrossRef]
7. Liu, K.; Geng, X.; Liu, X. The Application of Network Agenda Setting Model during the COVID-19 Pandemic Based on Latent Dirichlet Allocation Topic Modeling. *Front. Psychol.* **2022**, *13*, 954576. [CrossRef] [PubMed]
8. Eloundou, T.; Manning, S.; Mishkin, P.; Rock, D. GPTs Are GPTs: Labor Market Impact Potential of LLMs. *Science* **2024**, *384*, 1306–1308. [CrossRef] [PubMed]
9. Trammell, P.; Korinek, A. *Economic Growth Under Transformative AI*; NBER Working Paper No. 31815; NBER: Cambridge, MA, USA, 2023. [CrossRef]
10. Korea Bizwire. South Korea Ranks Second Globally in Paid ChatGPT Subscriptions as AI Adoption Surges. *Korea Bizwire*, 29 May 2025. Available online: http://koreabizwire.com/south-korea-ranks-second-globally-in-paid-chatgpt-subscriptions-as-ai-adoption-surges/319847?ckattempt=1 (accessed on 13 February 2026).
11. Ministry of Science and ICT; National Information Society Agency. *2024 Survey on the Internet Usage*; NIA: Daegu, Republic of Korea, 2025.
12. Cheng, Y. The Third-Level Agenda Setting Study: An Examination of Media, Implicit, and Explicit Public Agendas in China. *Asian J. Commun.* **2016**, *26*, 319–332. [CrossRef]
13. Wu, H.D.; Guo, L. Beyond Salience Transmission: Linking Agenda Networks between Media and Voters. *Commun. Res.* **2020**, *47*, 1010–1033. [CrossRef]
14. Guo, L.; Vargo, C. The Power of Message Networks: A Big-Data Analysis of the Network Agenda Setting Model and Issue Ownership. *Mass Commun. Soc.* **2015**, *18*, 557–576. [CrossRef]
15. Scheufele, D.A.; Tewksbury, D. Framing, Agenda Setting, and Priming: The Evolution of Three Media Effects Models. *J. Commun.* **2007**, *57*, 9–20. [CrossRef]
16. Zucker, H.G. The Variable Nature of News Media Influence. *Commun. Yearb.* **1978**, *2*, 225–240. [CrossRef]
17. Guo, L.; McCombs, M. Network Agenda Setting: A Third Level of Media Effects. In Proceedings of the International Communication Association Annual Conference, Boston, MA, USA, 26–30 May 2011.
18. Guo, L.; Vu, H.T.; McCombs, M. An Expanded Perspective on Agenda-Setting Effects: Exploring the Third Level of Agenda Setting. *Rev. Comun.* **2012**, *11*, 51–68.
19. Anderson, J.R. *The Architecture of Cognition*; Harvard University Press: Cambridge, MA, USA, 1983.
20. Kwon, S.; Cha, H.; Jeon, J. A Study on Intermedia Network Agenda Setting of Fine Dust Issues: Focusing on News Media Coverage and Naver Blogs. *J. Cybercommun. Acad. Soc.* **2020**, *37*, 5–52. (In Korean) [CrossRef]
21. Erbring, L.; Goldenberg, E.N.; Miller, A.H. Front-Page News and Real-World Cues: A New Look at Agenda-Setting by the Media. *Am. J. Polit. Sci.* **1980**, *24*, 16–49. [CrossRef] [PubMed]

22. Shaw, D.L.; McCombs, M.; Weaver, D.H.; Hamm, B.J. Individuals, Groups, and Agenda Melding: A Theory of Social Dissonance. *Int. J. Public Opin. Res.* **1999**, *11*, 2–24. [CrossRef]
23. Bennett, W.L.; Iyengar, S. A New Era of Minimal Effects? The Changing Foundations of Political Communication. *J. Commun.* **2008**, *58*, 707–731. [CrossRef]
24. Shoemaker, P.J.; Vos, T. *Gatekeeping Theory*; Routledge: New York, NY, USA, 2009. [CrossRef]
25. Benkler, Y. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*; Yale University Press: New Haven, CT, USA, 2006.
26. Chen, Z.; Su, C.C.; Chen, A. Top-Down or Bottom-Up? A Network Agenda-Setting Study of Chinese Nationalism on Social Media. *J. Broadcast. Electron. Media* **2019**, *63*, 512–533. [CrossRef]
27. Wessel, M.; Adam, M.; Benlian, A.; Majchrzak, A.; Thies, F. Generative AI and Its Transformative Value for Digital Platforms. *J. Manag. Inf. Syst.* **2025**, *42*, 346–369. [CrossRef]
28. Korinek, A.; Suh, D. *Scenarios for the Transition to AGI*; NBER Working Paper No. 32255; National Bureau of Economic Research: Cambridge, MA, USA, 2024. [CrossRef]
29. Acemoglu, D. *The Simple Macroeconomics of AI*; NBER Working Paper No. 32487; NBER: Cambridge, MA, USA, 2024. [CrossRef]
30. Gimbel, M.; Kinder, M.; Kendall, J.; Lee, M. Evaluating the Impact of AI on the Labor Market: Current State of Affairs. Budget Lab at Yale/Brookings Institution. 1 October 2025. Available online: https://budgetlab.yale.edu/research/evaluating-impact-ai-labor-market-current-state-affairs (accessed on 13 February 2026).
31. Entman, R.M. Framing: Toward Clarification of a Fractured Paradigm. *J. Commun.* **1993**, *43*, 51–58. [CrossRef]
32. Ittefaq, M.; Zain, A.; Arif, R.; Ala-Uddin, M.; Ahmad, T.; Iqbal, A. Global News Media Coverage of Artificial Intelligence (AI): A Comparative Analysis of Frames, Sentiments, and Trends across 12 Countries. *Telemat. Inform.* **2025**, *96*, 102223. [CrossRef]
33. Gamson, W.A.; Modigliani, A. Media Discourse and Public Opinion on Nuclear Power: A Constructionist Approach. *Am. J. Sociol.* **1989**, *95*, 1–37. [CrossRef] [PubMed]
34. Nisbet, M.C.; Scheufele, D.A. What's Next for Science Communication? Promising Directions and Lingering Distractions. *Am. J. Bot.* **2009**, *96*, 1767–1778. [CrossRef] [PubMed]
35. Silver, D.; Huang, A.; Maddison, C.J.; Guez, A.; Sifre, L.; van den Driessche, G.; Schrittwieser, J.; Antonoglou, I.; Panneershelvam, V.; Lanctot, M.; et al. Mastering the Game of Go with Deep Neural Networks and Tree Search. *Nature* **2016**, *529*, 484–489. [CrossRef] [PubMed]
36. Government of the Republic of Korea. *National Strategy for Artificial Intelligence*; Ministry of Science and ICT: Sejong, Republic of Korea, 2019. Available online: https://wp.oecd.ai/app/uploads/2021/12/Korea_National_Strategy_for_Artificial_Intelligence_2019.pdf (accessed on 13 February 2026).
37. Government of the Republic of Korea. *Korean New Deal: National Strategy for a Great Transformation*; Ministry of Economy and Finance: Sejong, Republic of Korea, 2020. Available online: https://english.moef.go.kr/pc/selectTbPressCenterDtl.do?boardCd=N0001&seq=4948 (accessed on 13 February 2026).
38. Manning, C.D.; Raghavan, P.; Schütze, H. *Introduction to Information Retrieval*; Cambridge University Press: Cambridge, UK, 2008. [CrossRef]
39. Salton, G.; Wong, A.; Yang, C.S. A Vector Space Model for Automatic Indexing. *Commun. ACM* **1975**, *18*, 613–620. [CrossRef]
40. Dekker, D.; Krackhardt, D.; Snijders, T.A.B. Sensitivity of MRQAP Tests to Collinearity and Autocorrelation Conditions. *Psychometrika* **2007**, *72*, 563–581. [CrossRef] [PubMed]
41. Church, K.W.; Hanks, P. Word Association Norms, Mutual Information, and Lexicography. *Comput. Linguist.* **1990**, *16*, 22–29.
42. Jurafsky, D.; Martin, J.H. *Speech and Language Processing*, 3rd ed. draft; Stanford University: Stanford, CA, USA, 2023. Available online: https://web.stanford.edu/ jurafsky/slp3/ (accessed on 13 February 2026).
43. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient Estimation of Word Representations in Vector Space. In Proceedings of the ICLR, Scottsdale, AZ, USA, 2–4 May 2013. Available online: https://arxiv.org/abs/1301.3781 (accessed on 21 June 2026).
44. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [CrossRef]
45. Phipson, B.; Smyth, G.K. Permutation P-Values Should Never Be Zero: Calculating Exact P-Values When Permutations Are Randomly Drawn. *Stat. Appl. Genet. Mol. Biol.* **2010**, *9*, 39. [CrossRef] [PubMed]
46. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed.; Lawrence Erlbaum Associates: Hillsdale, NJ, USA, 1988.
47. Barrat, A.; Barthélemy, M.; Pastor-Satorras, R.; Vespignani, A. The Architecture of Complex Weighted Networks. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 3747–3752. [CrossRef] [PubMed]
48. Borgatti, S.P.; Everett, M.G.; Johnson, J.C. *Analyzing Social Networks*, 2nd ed.; SAGE: London, UK, 2018.
49. Freelon, D. On the Interpretation of Digital Trace Data in Communication and Social Computing Research. *J. Broadcast. Electron. Media* **2014**, *58*, 59–75. [CrossRef]
50. Borgatti, S.P.; Everett, M.G.; Freeman, L.C. *UCINET 6 for Windows: Software for Social Network Analysis*; Analytic Technologies: Harvard, MA, USA, 2002.

51. Newman, N.; Fletcher, R.; Eddy, K.; Robertson, C.T.; Nielsen, R.K. *Reuters Institute Digital News Report 2023*; Reuters Institute for the Study of Journalism: Oxford, UK, 2023. Available online: https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2023 (accessed on 21 June 2026).
52. Hallin, D.C.; Mancini, P. *Comparing Media Systems: Three Models of Media and Politics*; Cambridge University Press: Cambridge, UK, 2004. [CrossRef]
53. Christians, C.G.; Glasser, T.L.; McQuail, D.; Nordenstreng, K.; White, R.A. *Normative Theories of the Media: Journalism in Democratic Societies*; University of Illinois Press: Champaign, IL, USA, 2009.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.