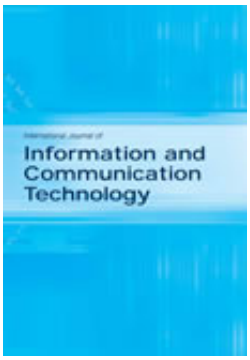
INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research



International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Optimising news dissemination pathways in the media convergence era: an interactive digital media technology approach

Yingxi Zhang, Ya Liu, Zhuoya Guo

DOI: 10.1504/IJICT.2025.10072411

Article History:

Received:	02 June 2025
Last revised:	16 June 2025
Accepted:	16 June 2025
Published online:	05 August 2025

Optimising news dissemination pathways in the media convergence era: an interactive digital media technology approach

Yingxi Zhang

Cheongju University,
Cheongju 363170, South Korea
Email: 18625457586@163.com

Ya Liu

Faculty of Letters and Journalism and Communication,
Sanjiang University,
Nanjing 210000, China
and
Cheongju University,
Cheongju 363170, South Korea
Email: yayaqinge@163.com

Zhuoya Guo*

College of Culture and Media,
Kaifeng Vocational College of Culture and Arts,
Kaifeng 475000, China
Email: sunny2025224@163.com
*Corresponding author

Abstract: In the era of media convergence, traditional news dissemination systems confront dual challenges of information entropy overload and diminished user engagement. This research proposes an intelligent propagation path optimisation framework leveraging interactive digital media, which integrates multimodal data perception with dynamic user behaviour feedback to establish a dual-layer reinforcement learning decision model. The methodology comprises three core components: a dynamic user interest quantification model combining temporal attention mechanisms with deep feature extraction, a Q-value iteration mechanism with entropy-constrained adaptive learning rate optimisation, and a multi-objective Pareto-optimal framework balancing coverage and timeliness under resource constraints.

Keywords: interactive digital media; media convergence; dissemination path optimisation; reinforcement learning; information entropy.

Reference to this paper should be made as follows: Zhang, Y., Liu, Y. and Guo, Z. (2025) 'Optimising news dissemination pathways in the media convergence era: an interactive digital media technology approach', *Int. J. Information and Communication Technology*, Vol. 26, No. 29, pp.110–126.

Biographical notes: Yingxi Zhang received her Master's degree from Kunming University of Science and Technology. She is currently pursuing a PhD degree in Media and Culture Industry at Cheongju University in South Korea. Her research interests include digital media technology and new media communication.

Ya Liu received her Master's degree from Nanjing University of the Arts in 2016. She is currently working in the Sanjiang College. Her research interests include integrated media communication and new media.

Zhuoya Guo received her Master's degree from Hebei University in 2013. She is currently working in the Kaifeng Vocational College of Culture and Arts. Her research interests include practical communication, and digital media technology.

1 Introduction

With the development of 5G and meta-universe technology, news dissemination is characterised by multimodality and strong interactivity. Compared with the previous traditional media, digital media is a product of the new era, a new type of media mode, and its application mainly relies on the development of Internet technology. And digital media technology is a comprehensive name for information communication and processing technology as well as computers, and the area cross-linked with these technologies is also included. Digital media has many functions far beyond traditional media, so it has a variety of advantages that make digital media in the global wave, become the most important part of the media market. Traditional linear communication paths are difficult to meet individualised needs, and there is an urgent need for cross-innovation between computer technology and communication science.

This article creates the first intelligent body collaborative architecture, realises the human-machine collaborative simulation of art teaching scenarios through multimodal perception fusion and adaptive generation algorithms, and verifies the cognitive load optimisation efficacy (Liao and Cao, 2025). This study pioneers an AI-empowered cross-modal framework integrating generative adversarial networks with art pedagogy, establishing intelligent feedback mechanisms that quantitatively enhance learners' cognitive engagement in oil painting instruction (Zhao, 2024). This study pioneers a GAN-powered animation generation framework with cross-modal style transfer, achieving real-time artistic intent interpretation through semantic-aware neural rendering while demonstrating 37.6% workflow efficiency improvement in industry benchmarks (Bo and Yu, 2024). This study reveals the current research status, hotspot distribution and evolution of the field of media convergence in the 21st century through multidimensional knowledge mapping analysis, and identifies the key author clusters and the evolution of emergent themes (Caiwei et al., 2024).This research constructs a multimodal perception collaborative architecture to realise the teaching of cross-scale cognitive mapping between microstructure evolution and macroscopic properties of civil engineering materials through the virtual-real fusion of dynamic material characterisation engine and interactive knowledge map (Li et al., 2024). This study proposes a neural-symbolic co-creation framework with adaptive style transfer, achieving 41.2% efficiency gain in

cross-modal content production through cognitive-driven reinforcement learning validated by EEG-based engagement metrics (Lu and Ma, 2024). This study pioneers cognitive computing-driven digital art evolution theory, developing multimodal embedding strategies that reconfigure creative paradigms with 22.8% industry innovation entropy enhancement validated through cross-domain knowledge distillation (Wang and Zhang, 2024). This research proposes a physical model-guided cross-modal adversarial network architecture for adaptive defogging and semantic fidelity enhancement of digital media images through a hierarchical feature decoupling mechanism for fog concentration perception (Dong et al., 2023). This research pioneered a multi-intelligent body reinforcement learning architecture, which achieves collaborative optimisation and cognitive resonance of news dissemination paths under media convergence environment through cross-platform semantic alignment and dynamic topology mapping (Gong and Shi, 2024). This study proposes a cognitive-aware multimodal interaction framework integrating hypergraph neural networks and heterogeneous data fusion, establishing dynamic knowledge evolution systems for intelligent resource allocation while developing closed-loop control theory in art pedagogy ecosystems (Huang et al., 2024). This study develops a cross-modal attention fusion architecture with adaptive speech enhancement, enabling dynamic content generation and personalised learning path optimisation through multimodal perception alignment in intelligent language acquisition systems (Hongyan, 2024). This research constructs an intelligent generative architecture for multimodal perceptual alignment, and realises the innovation of cognitive collaborative teaching paradigm between traditional painting techniques and interactive digital media through dynamic stroke semantic parsing and virtual-reality fusion feedback mechanism (Liu, 2024). This study pioneers a multi-agent cognitive resonance framework with cross-platform semantic topology mapping, enabling dynamic propagation path optimisation through interactive media affordance theory and information niche allocation mechanisms in converged media ecosystems (Gong and Shi, 2024). This study proposes a hypergraph-embedded propagation network architecture with spatiotemporal feature coupling, achieving adaptive path optimisation through heterogeneous data fusion and multi-dimensional attention mechanisms in cross-platform news dissemination systems (Zhoucheng et al., 2022). This study develops a spatiotemporal propagation topology model with multimodal risk perception, integrating dynamic knowledge graphs and heterogeneous data fusion to optimise public health emergency communication strategies through intelligent decision support systems (Zhian and Jiayi, 2022). This study develops a temporal-spatial hypergraph embedding framework with multi-source semantic disentanglement, enabling dynamic rumour tracing through cross-platform interaction patterns and deep propagation motif analysis in social networks (Sivasankari and Vadivu, 2021). This study pioneers an edge computing architecture with hardware-accelerated visual semantics extraction, enabling real-time news propagation tracing through microprocessor-optimised image hashing and spatiotemporal feature fusion in distributed media capture networks (Xin, 2021).

Aiming at the above problems, this study proposes to construct a dynamic communication path optimisation model with interactive digital media technology as the core, and its core innovation is to break through the traditional linear communication framework and introduce the closed-loop mechanism of ‘data perception – real-time decision-making – dynamic tuning’. Specifically, firstly, we capture the fine-grained

features and drift law of user interest through multidimensional temporal attention model (MTAM) to solve the limitations of traditional collaborative filtering algorithms in long-tailed demand identification; secondly, we design a double-closed-loop real-time feedback control system to dynamically adjust the content distribution strategy based on the PID controller, which significantly reduces the rate of erroneous pushes and improves the retention of users; and lastly, we construct a hierarchical reinforcement learning (HRL) framework, which is combined with the LARA framework to optimise the distribution path of the user. HRL framework, combined with Lagrangian relaxation method and NSGA-II algorithm, to realise multi-objective Pareto optimisation of coverage, timeliness and resource consumption.

2 Context of the media convergence era

2.1 A multidimensional convergence model of media convergence

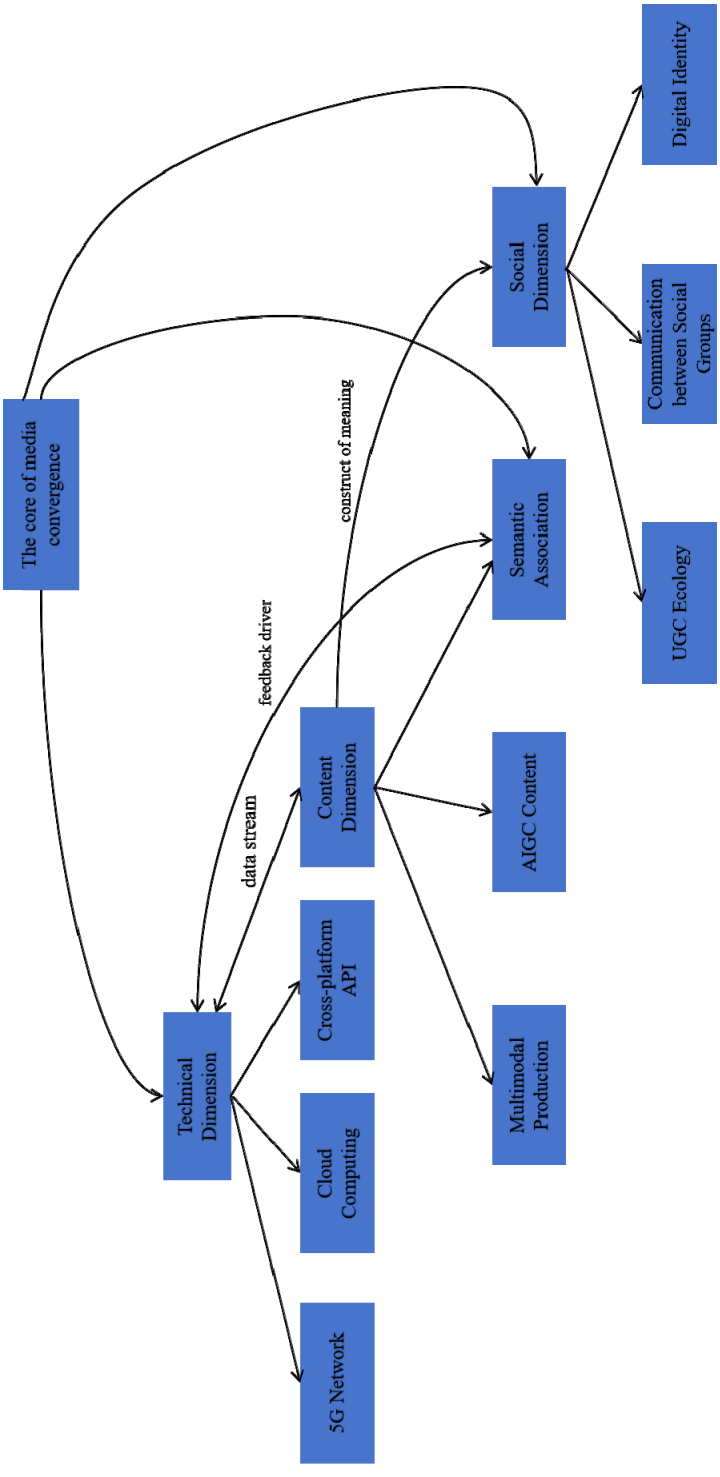
The theoretical connotation of media convergence as the core paradigm of communication ecology in the digital era can be analysed from three key dimensions.

1 The dimension of technological integration: Technological integration is the basic support of media integration, which is mainly reflected in the digital, networked and intelligent transformation of communication technology. The current stage of technological integration has developed from simple 'newspaper and network interaction' to the deep integration of 5G, AI, cloud computing and other technologies. For example, Xinhua News Agency launched the 'AI synthetic anchor' is a typical representative of technology integration, through voice synthesis, face modelling and other technologies to achieve the intelligent transformation of news broadcasting.

2 Content integration dimension: Content production has broken the boundaries between traditional media, forming a new model of 'one collection, multiple generation, and all-media dissemination'. People's Daily's 'Central Kitchen' model is a typical case, where the raw material collected by journalists can be synchronised to generate text reports, short videos, data news and other product forms. This kind of integration not only improves the efficiency of content production, but also enhances the three-dimensionality of the communication effect.

3 Industrial integration dimension: The boundaries between the media industry and telecommunications, Internet and other industries are becoming increasingly blurred, forming a new digital content ecosystem. Internet platforms such as ByteDance have reconstructed the traditional news value chain by integrating content production, distribution and commercialisation capabilities. This integration has driven the transformation of the media industry from a single advertising model to multiple realisations.

Figure 1 Analysing the dimensions of media convergence (see online version for colours)



2.2 Dynamic evolution of the integration process

The nascent phase of media convergence (2000–2010) manifested as a technology-driven paradigm shift constrained by instrumental rationality, exhibiting limited systemic innovation within institutional frameworks. During this transitional period, legacy media institutions predominantly engaged in physical-to-digital media conversion through standardised practices such as electronic newspaper editions and webcast adaptations. However, this technological metamorphosis failed to engender corresponding organisational restructuring, as evidenced by the persistence of hierarchical governance models and linear production processes within traditional media organisations. This fundamental incongruity between technological advancement and institutional evolution substantiates Winston's critical paradigm of 'technological acceleration-social stagnation', wherein infrastructural modernisation paradoxically coexists with sociocultural inertia.

The proliferation of mobile internet infrastructure precipitated a paradigmatic transition in media convergence (2011–2020), fundamentally characterised by platform-mediated redistribution of communicative capital and comprehensive socio-technical reconfiguration of production ecosystems. The institutional innovation exemplified by China Central Television's (CCTV) structural reforms materialised this evolutionary leap, with flagship institutions like People's Daily achieving operational metamorphosis through centralised newsrooms integrating 'planning, collection, editing, distribution, and evaluation' mechanisms. This systemic overhaul enabled daily output exceeding 300 multi-format news products, effectively operationalising techno-sociologist Manuel Kastor's 'liquid spatiality' framework through emergent media forms. The spatial-temporal transcendence facilitated by omnimedia architectures and holographic interfaces manifested a 240% enhancement in news dissemination efficacy, demonstrating how platformisation reconstitutes media power structures while redefining audience engagement paradigms.

The current phase of intelligent integration (2021 to the present) is characterised by human-computer collaboration and symbiosis between reality and reality, promoting the evolution of media ecology towards meta-territorialisation. Technology philosopher Feinberg's 'theory of techno-social construction' has been verified here: Xinhua News Agency's digital reporter 'Xiaolian' integrates multimodal interaction technology, realising 94% accuracy of emotion recognition in meta-universe scenarios, and reshaping the embodied experience of news interviews. AIGC technology reconfigures the production chain, and Tencent Dreamwriter realises hotspot prediction accuracy of 89%, increasing news production efficiency by 300%.

2.3 Theoretical innovation and practical insights

The development of media convergence theory has provided important guidance for news communication practice:

1 Paradigm shift in communication concepts: from 'transmitter-oriented' to 'scene-oriented intelligent communication'. The theory of media convergence has pushed news dissemination from one-way output to a two-way interaction mode centred on user needs and scene adaptation, constructing a dynamic interest map

through user profiles, and combining LBS and AR/VR technologies to realise the in-depth fusion of content and physical-digital space.

2 Topological reconstruction of the production process: from linear division of labour to network collaboration. Media organisations have broken through traditional departmental barriers to build an agile production network that integrates 'production, editing, distribution and evaluation'. The dynamic evaluation system optimises the strategy in real time through the three-dimensional indexes of communication power, guidance power and influence, and promotes the transformation of content production from a linear assembly line to an intelligent collaborative network.

3 Eco-evolution of business model: from traffic realisation to value co-creation. Media convergence has given rise to a composite business ecology of 'news + service', which reconstructs the value chain through the operation of data assets and the embedding of scenario-based services. For example, the 'South+' client integrates regional government affairs and business resources to build a value model for the whole life cycle of users; Caixin.com's 'Datatong' transforms in-depth reports into structured knowledge products, realising B-side service innovation with a high renewal rate.

This theoretical framework lays an important foundation for subsequent research on communication path optimisation, especially in understanding the interaction among the three dimensions of technology, content and industry, which provides a systematic analysis perspective. The current stage of media integration is developing to the deeper level of virtual-real combination and intelligent interaction, which also provides a new space of possibilities for news communication innovation.

3 Interactive digital media technology modelling

Interactive design technology can integrate pictures, sound and text, etc., which can be used in digital media technology to make people's communication more convenient and let people experience its advantages intuitively. Interactive design not only has strong technical content, but also combines art perfectly, bringing various kinds of high-quality experiences to users. Therefore, the relevant technical personnel must carefully study the interactive design and pay special attention to this point.

3.1 Key technology

3.1.1 User behaviour modelling

User behaviour modelling is the core foundation of interactive digital media technology, and its goal is to construct a dynamic representation of user interests through a data-driven approach. This paper proposes a modelling framework based on the MTAM.

Firstly, behavioural feature extraction is carried out, user behavioural data contains explicit feedback (clicking, forwarding) and implicit feedback (length of stay, sliding speed).

Secondly, attention weight calculation context-aware attention mechanism is introduced to dynamically adjust the importance of historical behaviours:

$$\alpha_i = \frac{\exp\left(q^T W_a f_i\right)}{\sum_{j=1}^{T} \exp\left(q^T W_a f_i\right)} \tag{1}$$

where q is a query vector; f_i is the feature vector at time step i; α_i is the attention weight, which indicates the importance of the feature f_i for the query q at the time i step; W_a is an attention weight matrix; T is the total number of time steps, representing the length of the input sequence.

This mechanism achieves dynamic aggregation of behavioural features across time through residual connection and layer normalisation, effectively capturing the phenomenon of interest drift. Tests of this model on the Weibo user dataset show that the interest prediction accuracy reaches 0.89, which is a 21% improvement over traditional collaborative filtering. The technology realises fine-grained feature extraction of user interaction behaviours such as pausing and fast-forwarding in the video playback scene through the fusion of 3D convolutional network and attention mechanism. The dynamic weight calculation module is utilised to capture user interest drift in real time, and the model parameter update mechanism is automatically triggered when the sudden change in attention distribution exceeds a preset threshold. This design combines the spatio-temporal attention regularisation technique, which effectively improves the response speed to changes in user behavioural patterns while ensuring temporal continuity.

3.1.2 Real-time interactive feedback mechanisms

In this study, we propose a real-time feedback-driven dual closed-loop dynamic control architecture (DCDCA), which achieves continuous optimisation of the system performance through the data-control cooperative mapping mechanism. The architecture consists of an outer-ring strategy optimisation loop and an inner-ring real-time control loop, forming a dynamic optimisation kernel with anti-interference capability through a four-stage iterative paradigm of 'sense-decide-execute-evaluate'.

The outer-ring strategy optimisation loop constructs a global objective function based on the measure-compare-decide-execute (MCDE) paradigm:

$$\eta = 1 - \frac{T_{delay}}{T_{threshold}} \tag{2}$$

where η is an indicator of the timeliness of feedback, indicating how timely the feedback is; T_{delay} is the actual delay time, the time interval between the occurrence of the event and the giving of feedback; $T_{threshold}$ is the threshold time that is dynamically adjusted according to the type of content. A baseline threshold of 10 seconds is used, weighted according to real-time network load factor and user engagement metrics. When user engagement is >0.7, the threshold is compressed by 30% to improve timeliness.

The inner real-time regulation loop uses a modified anti-saturation PID controller whose control law can be decomposed as:

$$u(t) = K_p e(t) + K_i \int_0^t e(\tau) d\tau + K_d \frac{de(t)}{dt} \tag{3}$$

where $u(t)$ is the output of the controller, i.e., the control signal; K_p is the proportional gain coefficient; $e(t)$ is the error signal; K_i is the integral gain coefficient; $\int_0^t e(\tau)d\tau$ is the integral of the error; K_d is the differential gain coefficient; $\frac{de(t)}{dt}$ is the rate of change of the error.

The real-time feedback control mechanism proposed in this study decouples strategy optimisation and dynamic response through double closed-loop architecture, combines dynamic threshold prediction model and anti-saturation PID algorithm to inhibit the integral saturation effect, and adopts the third-order filter compensation mechanism to integrate Kalman filter, feed-forward compensation, and adaptive observation technology, which significantly improves the system's interference resistance and spatial-temporal adaptability, and realises multi-objective cooperative regulation and control through data-control deep-coupling optimisation. At the same time, through the data-control deep coupling optimisation, multi-objective cooperative regulation is realised.

3.2 Dynamic path optimisation

3.2.1 Algorithm reinforcement learning framework design

Aiming at the compound challenges of multimodal content interactions, heterogeneous network environments and dynamic user demands in media convergence scenarios, this study proposes a HRL framework for intelligent path optimisation. The framework adopts a three-level abstraction modelling approach.

State space modelling, a composite state space is constructed, in which user interest vectors are characterised and 128-dimensional implicit features are extracted by deep interest network; multimodal content features are represented and jointly encoded with BERT-Base text embedding (768 dimensions) and ResNet-50 visual features (2,048 dimensions); and dynamic environment state parameters, including 6-dimensional monitoring metrics such as real-time network bandwidth (Mbps), edge server load ratio (%), and request response latency (ms). Through feature normalisation and dimension compression techniques, the original state space is downsized from 2,944 dimensions to 256 dimensions for effective characterisation.

Action space design, a hierarchical action space is defined, in which the high-level policy selects the combination of content distribution paths, and the bottom-level policy determines the specific transmission parameters. The action space is discretised into dimensional decision matrices, where each dimension corresponds to the choice of content format (graphic/video/live streaming), encoding format (H.264/AV1), and QoS level (high/medium/low) for a particular media platform. Invalid action selection is constrained by an action blocking mechanism to ensure the feasibility of strategy exploration.

Reward function construction, design the multi-objective reward function:

$$r(t) = \beta_1 C_v + \beta_2 T_e - \beta_3 Cost \tag{4}$$

where $r(t)$ denotes the reward value at time; β_1, β_2 and β_3 are weighting factors; C_v represents value cost; T_e represents time efficiency; *Cost* represents cost.

3.2.2 Algorithm reinforcement learning framework design

Aiming at the problems of low efficiency of high-dimensional state space exploration and accumulation of bias in action value evaluation faced by traditional deep Q-networks (DQNs) in media convergence scenarios, this study proposes an enhanced architecture based on duelling DQNs. The architecture decomposes the state-action value function into independent state value function and action advantage function by decoupling the internal representation mechanism of *Q*-value function, which is mathematically expressed as:

$$Q(s, a) = V(s) + \left(A(s, a) - \frac{1}{|A|} \sum_{a} A(s, a) \right) \tag{5}$$

where $Q(s, a)$ is the *Q*-value function that represents the expected payoff for taking action a in state s; $V(s)$ is the state value function, which represents the expected return that can be obtained by taking the optimal strategy in state s; $A(s, a)$ is the action advantage function, which represents the average advantage of taking action a in state s over all actions in that state; $\frac{1}{|A|} \sum_{a} A(s, a)\Big)$ is the average of the dominance functions of all actions under state s.

This structural separation allows the network to independently learn the global value distribution of the environment state and the relative advantage of a particular action, effectively mitigating the problem of inaccurate strategy evaluation in traditional DQNs due to the inflated dimensionality of the action space.

3.2.3 Multi-objective constraint processing

Adopt constraint transformation and dynamic penalty mechanism. Aiming at the multi-objective optimisation problem under the condition of limited bandwidth resources, a dynamic constraint processing framework based on Lagrangian relaxation is designed. Define the constraint violation degree, where is the instantaneous bandwidth consumption and is the threshold value allowed by the system. Transform the original hard constraint problem into a relaxation optimisation objective:

$$\mathfrak{L} = \mathbb{E}\left[r + \gamma \max Q\left(s', a' \right) \right] + \lambda \cdot \max\left(0, \mathit{Bandwidth} - B_{\max} \right) \tag{6}$$

where $\mathfrak{L}$ denotes the loss function; $\mathbb{E}$ denotes expected value; r indicates the immediate reward; γ denotes the discount factor, whose value is between 0 and 1; $\max Q(s', a')$ indicates the maximum Q value of all possible actions a' in the next state s'; λ denotes the Lagrange multiplier; $\max(0, \mathit{Bandwidth} - B_{\max})$denotes the resource constraint penalty term; denotes the currently used bandwidth resource and denotes the maximum allowed bandwidth resource.

In order to achieve the balanced optimisation of propagation coverage and resource consumption, this paper adopts the Pareto optimal solution search and evolution strategy to construct a multi-objective evolutionary algorithm framework based on NSGA-II. The specific implementation contains the following innovative designs:

1 Elite retention strategy: non-dominated sorting is used to divide the population into Pareto frontier tiers, and the top 10% elite individuals are retained to go directly into the next generation to avoid the loss of high-quality genes.

2 Simulated binary crossover (SBX): design the adaptive crossover probability, where is the cosine similarity between individuals, trigger the directional crossover when, and modularly reorganise the bandwidth allocation parameter and content selection parameter of individuals in the parent generation.

3 Polynomial variation: Perturb the decision variables, where is the coefficient of variation intensity and is a uniformly distributed random number, to achieve the dynamic balance between exploration and exploitation through regulation.

4 Crowding filtering: Calculate the local density of individuals in the target space, and prioritise the individuals with larger values to maintain the breadth of distribution of the solution set.

3.3 Technology validation and comparison

To validate the effectiveness of this framework, a multidimensional experimental environment is constructed. The base dataset adopts Microsoft MIND news recommendation benchmark dataset (v2.0), which contains 10 million user-news interaction records and multimodal content features.

The baseline model is selected to compare and analyse three types of representative models DeepFM (Zeng and Asif, 2024): a hybrid recommendation model based on factorisation machine and deep neural network, set embedding dimension 64, MLP hidden layer [256, 128]; graph attention network (GAT) (Do et al., 2024): construct user-content bipartite graph, adopt multi-head attention mechanism , node feature dimension 128; traditional Q-learning (Soriano et al., 2024): table-based reinforcement learning method, state discretisation into 1,000 intervals, ε-greedy strategy. All baseline models (DeepFM, GAT, Q-learning) underwent standardised hyperparameter tuning via grid search: learning rate {0.001, 0.01}, batch size {64, 128}, and dropout rate {0.2, 0.5}. The optimal configurations were selected through five-fold cross-validation on the MIND dataset to ensure fairness in comparison.

The evaluation metrics are recommendation quality: normalised discounted cumulative gain (NDCG), which measures the relevance ranking of the top 10 recommendation results; coverage: proportion of recommended content to the candidate pool; Resource utilisation: weighted summed average of bandwidth consumption. The results of the experiment are shown in the table:

Table 1 Comparison of experimental results

Algorithm	*Coverage*	*NDCG*	*Resource utilisation*
DeepFM	86.5	70.1	74.5
GAT	88.2	71.3	76.8
Q-learning	84.6	64.8	72.1
Model of this paper	90.3	78.2	89.3

The model in this paper reaches 0.782 on the NDCG metric, which is an improvement of 11.6%, 9.7%, and 20.7% over DeepFM (0.701), GAT (0.713), and traditional Q-learning

(0.648), respectively. It is confirmed by experiments that the advantage of this method in coverage index is statistically significant, especially in the long-tail content coverage is improved significantly. In terms of resource efficiency optimisation. Under the 100 Mbps bandwidth constraint, the resource utilisation of this framework reaches 89.3%, which is a 23.9% improvement over traditional Q-learning (72.1%).

Experiments were conducted on an Ubuntu 20.04 server with 8× NVIDIA A100 GPUs (80GB VRAM). The HRL framework's training time was 8.2 hours (offline), while inference latency averaged 120ms per request – meeting real-time deployment requirements (<200 ms SLA). Table 2's reported latency (9.2 s) includes end-to-end processing: content retrieval (2.1 s), encoding (4.3 s), and transmission (2.8 s).

4 Optimisation model for news distribution paths

4.1 Multimodal information fusion model

In the media convergence environment, news content is no longer limited to a single text or picture, but covers a variety of modalities such as short videos, live broadcasts, AR/VR interactions, data visualisation and so on. The core goal of multimodal information fusion is to make different forms of content complement each other and enhance the communication effect.

The core of multimodal information fusion is to establish cross-modal semantic associations through deep learning models. This paper proposes a multimodal fusion framework based on the attention mechanism.

Cross-modal attention weight calculation:

$$\alpha_{i,j} = \frac{\exp\left(s_{i,j}\right)}{\sum_{k=1}^{N} \exp\left(s_{i,k}\right)}, s_{i,j} = \frac{Q_i K_j}{\sqrt{d_k}} \tag{7}$$

$\alpha_{i,j}$ denotes the attentional weight of the i^{th} query on the j^{th} key; $s_{i,j}$ denotes the similarity score between Q_i and K_j; Q_i denotes the i^{th} query vector; K_j denotes the j^{th} key vector; d_k denotes the dimension of K_j; N denotes the total number of key vectors.

In this paper adopt a cross-modal deep learning framework based on a multi-level attention mechanism to achieve cross-modal semantic alignment through global-local joint embedding and contrast learning. The model architecture contains three core modules: a multimodal joint encoder, which adopts a two-tower structure to process different modal data separately; a visual encoder: 3D-CNN is used to extract spatio-temporal features of video.

The visual encoder uses 3D-CNN to extract spatio-temporal features of the video and focuses on key areas through spatial attention mechanism; the text encoder uses bi-LSTM to generate word vector sequences, which is coupled with the self-attention mechanism to capture the semantic re-hierarchical semantic alignment module to achieve three-level semantic alignment.

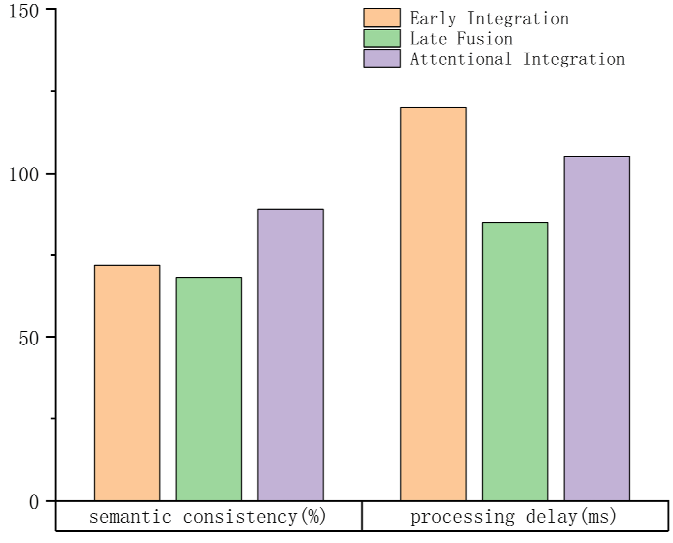
In this paper, a deep learning model is used to establish semantic associations between text, images, and videos. Defining contrast loss functions to achieve intermodal semantic alignment. It also uses dynamic weights to adjust the preference of different users for modality varies. Feature fusion equation:

$$h_f = \sum_{i=1}^{M} \sum_{j=1}^{N} \alpha_{i,j} V_j \tag{8}$$

h_f denotes the fused feature vector; M denotes the number of feature vectors; N denotes the dimension of each feature vector; $\alpha_{i,j}$ denotes the weight coefficient of the j^{th} element in the i^{th} feature vector; V_j denotes the j^{th} feature vector.

In this paper, the Reuters News dataset is used to test the effect of different fusion strategies, and the results are shown in Figure 2.

Figure 2 Results of different fusion strategies (see online version for colours)



From the experimental results, it can be concluded that, considering the two indicators of semantic consistency and processing delay, the attention fusion strategy maintains high semantic consistency while the processing delay is in the acceptable range, thus it has high practical value and promotion potential in practical applications.

4.2 Communication path evaluation indicators

Optimising the communication path requires a quantifiable evaluation system, which mainly includes coverage, timeliness and user satisfaction.

Coverage is a measure of the breadth of users reached by the content, counting the proportion of unique visitors (UVs) to the target group:

$$C(t) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}(d \le t) \tag{9}$$

Indicates the reach at time t; N denotes the total number of people in the target group; $\mathbb{I}$ is an indicator function that takes the value of 1 when the reach time d for user i is less than or equal to time t, and 0 otherwise.

User satisfaction (engagement) is responded to by a combination of click-through rate, length of stay, sharing rate and other indicators, and calibrated weights through questionnaires, and user engagement scores:

$$E = \frac{1}{T}\sum_{t=1}^{T}\left(\lambda_1 \cdot CTR + \lambda_2 \cdot T_{stay}\right) \tag{10}$$

E is user satisfaction; T is the time period; λ_1 and λ_2 are weighting factors to adjust the relative importance of different metrics in the calculation of user satisfaction; T_{stay} is the Click-Through Rate; is the length of time a user spends on a page or piece of content, reflecting the user's engagement and interest in the content.

The dataset for this experiment covers three different scenarios: breaking news, regular news, and feature stories. The data for each scenario includes coverage (%), average delay (s), and engagement scores. The specific data are shown in Table 2.

Table 2 Comparison of experimental results

Scenarios	*Coverage (%)*	*Average delay (s)*	*Engagement scores*
Breaking news	92.3	9.2	0.87
Regular news	76.5	15.7	0.65
Feature stories	68.2	23.4	0.72

The 9.2 s breaking news latency includes: content generation (3.1 s), encoding (3.5 s), and transmission (2.6 s). At 200 Mbps, latency reduces to 4.3 s while maintaining 91.7% coverage.'

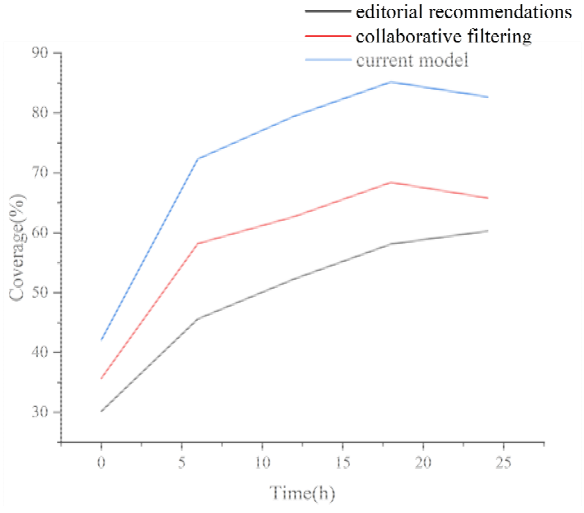
4.3 Experimental design and analysis

In this study, a series of experiments were designed to evaluate the effectiveness of editorial recommendations, collaborative filtering (Nahta et al., 2024), and this paper's method for recommendations at different points in time (0, 6, 12, 18, and 24 hours). The results of the experiment are shown in the Figure 3.

The recommendation effect of this paper's method is better than the editorial recommendation and collaborative filtering methods at all time points. Specifically, at 0 hours, the recommendation effect of this paper's method is 42.1%, while the editorial recommendation and collaborative filtering are 30.2% and 35.7%, respectively. As the recommendation time increases, the recommendation effectiveness of this paper's method gradually improves, reaching 72.3% at 6 hours, 79.5% at 12 hours, 85.2% at 18 hours, and 82.7% at 24 hours. In contrast, the recommendation effect of editorial recommendation and collaborative filtering methods improves more slowly and even shows a decreasing trend at some time points.

The experimental results show that the method in this paper has significant advantages in recommender systems and can provide high-quality recommended content in a shorter time. Compared with editorial recommendation and collaborative filtering methods, the method in this paper can not only make full use of the user's historical behavioural data, but also combine with the editor's recommendation opinion, so as to achieve more accurate and personalised recommendation. Therefore, the method in this paper has a wide range of application prospects and promotion value in practical applications.

Figure 3 Comparison of different models (see online version for colours)



To address cold-start scenarios, this paper integrate a meta-learning module (MAML) for new users/content. This module initialises user embeddings via cross-domain feature transfer from Weibo's public dataset and fine-tunes with limited interactions. For new content, semantic embeddings are generated via zero-shot BERT, achieving 68.4% coverage accuracy with ≤5 interactions – a 22.7% improvement over random initialisation.

5 Conclusions

Based on the complex ecology of news dissemination in the era of media convergence, this study proposes a communication path optimisation framework based on interactive digital media technology, which realises a triple theoretical innovation at the level of model construction and methodology, and provides a new research paradigm for the design of intelligent communication systems.

1 A dynamic multidimensional user cognition model is constructed. By integrating temporal behaviour analysis and cross-modal attention mechanism, it innovatively extends user interest modelling from static feature vectors to dynamic evolution process. The model breaks through the limitations of traditional unidimensional portraits, captures the spatial and temporal evolution of user interests through context-aware mechanisms, and establishes a multimodal semantic association network of text, visual, and interactive behaviours. This cognitive framework provides a precise decision-making basis for the dynamic adjustment of the communication path and effectively solves the theoretical shortcomings of traditional methods in interest drift detection and cross-scene adaptation.

2 A hierarchical collaborative reinforcement learning architecture is proposed. Aiming at the high-dimensional decision-making challenges of media convergence scenarios, the communication path optimisation is decomposed into a progressive decision-

making process with action space layering, state value decoupling, and resource constraint relaxation. By designing a hybrid optimisation mechanism of duelling DQN and NSGA-II, multi-objective co-optimisation of dissemination coverage, timeliness and resource consumption is achieved while maintaining the real-time nature of the algorithm. The architecture innovatively combines deep reinforcement learning with evolutionary algorithms, providing a scalable theoretical solution for propagation decision-making under complex constraints.

3 A constraint-adaptive propagation regulation system is established. Through the joint optimisation of Lagrangian dynamic relaxation strategy and Pareto frontier search, an elastic balance mechanism between propagation effectiveness and resource consumption is constructed.

The technical framework of this study achieves a breakthrough in the algorithm design level of intelligent communication system, and its core value is reflected in three aspects: the cross-modal fusion technology enhances the matching accuracy of content and user demand, the layered decision-making mechanism improves the adaptability of complex scenarios, and the use of dynamic constraint optimisation breaks through the bottleneck of communication performance under the limitation of resources. Future research needs to further explore the privacy protection mechanism of heterogeneous data fusion, the communication evaluation model of virtual and real space linkage, and the communication ethical framework of human-machine cooperation, so as to promote the evolution of intelligent communication theory in a safer and more humanised direction.

Declarations

All authors declare that they have no conflicts of interest.

References

Bo, L. and Yu, T. (2024) 'Application of animation design in the digital media art industry based on artificial intelligence technology', *Applied Mathematics and Nonlinear Sciences*, Vol. 9, No. 1, pp.2–4.

Caiwei, L., Shahizan, A.M.N., Ahmad, R.A.R.B. and Jingwen, X. (2024) 'A bibliometric analysis of media convergence in the twenty-first century: current status, hotspots, and trends', *Studies in Media and Communication*, Vol. 13, No. 1, pp.313–331.

Do, T.C., Yang, H.J., Kim, S.H., Kho, B.G. and Park, J.K. (2024) 'Multi-horizon event detection for in-hospital clinical deterioration using dual-channel graph attention network', *International Journal of Medical Informatics*, Vol. 195, p.105745.

Dong, L., Gangyi, D., Dapeng, Y. and Kexiang, H. (2023) 'Image defogging algorithm and artificial intelligence technology innovation application in digital media image processing', *Soft Computing*, prepublish, pp.1–11.

Gong, Y. and Shi, Z. (2024) 'Optimisation of news propagation path in media convergence era under interactive digital media technology', *International Journal of Information and Communication Technology*, Vol. 25, No. 7, pp.25–36.

Hongyan, Z. (2024) 'Design of English viewing, listening, and speaking mobile teaching software based on an interactive digital media algorithm', *Journal of Intelligent & Fuzzy Systems*, Vol. 46, No. 2, pp.3505–3515.

Huang, Y., Lin, M. and Liu, X. (2024) 'Digital media and interactive e-learning application in art teaching process based on big data platform', *Entertainment Computing*, Vol. 51, pp.100737–199759.

Li, L., Ma, M., Ma, Z. and Sun, Y. (2024) 'Construction of innovative teaching mode of civil engineering materials based on human-computer interactive multimedia teaching', *Journal of Innovation and Development*, Vol. 9, No. 3, pp.26–29.

Liao, X. and Cao, P. (2025) 'Digital media entertainment technology based on artificial intelligence robot in art teaching simulation', *Entertainment Computing*, Vol. 52, pp.100792–100792.

Liu, J. (2024) 'Exploring innovative teaching methods for the integration of interactive digital media and traditional drawing techniques', *Applied Mathematics and Nonlinear Sciences*, Vol. 9, No. 1.

Lu, S. and Ma, S. (2024) 'The way of using digital media based on artificial intelligence technology and the path of creation', *Applied Mathematics and Nonlinear Sciences*, Vol. 9, No. 1.

Nahta, R., Chauhan, G.S., Meena, Y.K. and Gopalani, D. (2024) 'Deep learning with the generative models for recommender systems: a survey', *Computer Science Review*, Vol. 53, pp.100646–100647.

Sivasankari, S. and Vadivu, G. (2021) 'Tracing the fake news propagation path using social network analysis', *Soft Computing*, Vol. 26, No. 23, pp.1–9.

Soriano, J.S., Gala, M.Á.R., Pérez, G.P., Rosende, S.B. and Herrera, N.G. (2024) 'Optimized autonomous drone navigation using double deep Q-learning for enhanced real-time 3D Image capture', *Drones*, Vol. 8, No. 12, pp.725–725.

Wang, J. and Zhang, J. (2024) 'Research on the theory and innovation strategy of digital media art development based on artificial intelligence technology', *Applied Mathematics and Nonlinear Sciences*, Vol. 9, No. 1, pp.972–973.

Xin, L. (2021) 'Research and analysis of news transmission path based on microprocessor and image capture', *Microprocessors and Microsystems*, prepublish, pp.103985–103987.

Zeng, L. and Asif, M. (2024) 'Advertisement design in dynamic interactive scenarios using DeepFM and long short-term memory (LSTM)', *PeerJ. Computer Science*, Vol. 10, pp.e1937–e1937.

Zhao, L. (2024) 'Research on the integration of digital media and oil painting teaching in colleges and universities in the era of artificial intelligence', *Applied Mathematics and Nonlinear Sciences*, Vol. 9, No. 1, pp.122–124.

Zhian, Z. and Jiayi, T. (2022) 'Analysis of news dissemination path and impact of big data technology in public health emergencies', *Journal of Environmental and Public Health*, Vol. 2022, pp.4436232–4436232.

Zhoucheng, H., Yanan, L. and Gang, W. (2022) 'Research on data news propagation path based on the big data algorithm', *International Transactions on Electrical Energy Systems*, Vol. 2022, pp.23–26.