International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

ADAPTIVE AD PLACEMENT USING MULTIMODAL CONTEXTUAL AI FOR LIVE STREAMING

Viharika Bhimanapati

Southern University and A&M College

viharikareddy.b@gmail.com

Abstract

The growth rate of the usage of digital platforms to stream live content has raised an urgent demand to have context and adaptive placement of advertisements. The classic systems of delivering the advertisements are not dynamic enough to be able to reflect the dynamic nature of the live environment, in which the focus of the user, the semantics of the content, and the tone of emotion may change in the very moment. The present paper examines how the contextual Artificial Intelligence (AI) of multimodal, i.e., audio, video, textual, and behavioral data, can be used to personalize and adapt to ad placements in live streaming conditions. Multimodal AI systems dynamically evaluate the context of the streaming content by combining machine learning, deep learning, and real-time decision engines to present the most appropriate ad content at the most appropriate time. The study explores the design of architecture, strategies of models, data streams, assessment models, and the future. It also manages such challenges as data fusion, real-time limitations, and ethical matters. This paper highlights the revolutionary nature of multimodal AI to transform online advertising and increase the enrichment of the viewer experience in live media.

Keywords: Multimodal AI, Contextual Advertising, Live Streaming, Adaptive Ad Placement, Deep Learning

1. Introduction

The fast emergence of live-streamed content on different media, like YouTube Live, Twitch, and Facebook Live, has upset the traditional models of broadcasting and content delivery. These settings require interactive dynamism, real-time responsiveness, and user experience, more so monetization via advertising. Compared to pre-recorded videos, live streaming presents such complexities as temporal uncertainty, unscripted content, and uneven attention of viewers, posing a challenge to the validity of traditional advertising paradigms [1][2]. In conventional ways of placing advertisements, time slots are important; there are assumptions of certain demographics and certain content that must be placed. These strategies do not have the dynamism to adapt to the constantly changing content and do not leverage the contextual cues capable of having a great impact on ad relevance and user receptivity. The resultant effect is the lack of alignment between the content of the ads and the interest of the viewer, which most often leads to ad fatigue, a lower rate of click-through, and decreased advertiser ROI [3][4]. The recent developments in Artificial Intelligence (AI), specifically in the field of multimodal learning, have preconditioned the new paradigm of ad delivery. Multimodal contextual AI is the combination and processing of different data streams, including video

Received: August 08, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

frames, audio speech, subtitle text, sentiment, biometric feedback, and user interaction data into a multifaceted, real-time perception of the live material and state of the viewer. The models can then be applied in making decisions as to when and what ad to place, maximizing impact, and minimizing disruption [5][6]. This is the direction of the further tendency of customization of media intake, in which the viewer not only needs media content aimed at them and their tastes but also advertising that is considered to be appropriate and non-obtrusive. Through the combination of multimodal contextual AI, real-time delivery of adaptive advertising can be provided, being sensitive to emotional tone, scene context, viewer behavior, and engagement metrics [7][8].

Such systems require strong technical foundations such as high-throughput data pipelines, low-latency inference engines, edge computing, and new machine learning models that have been trained using multimodal data. Besides, it requires a system of balancing personalization with privacy, transparency, and ethical targeting of adverts, particularly sensitive or emotive content contexts [9][10]. In this paper, the author discusses the end-to-end pipeline of adaptive ad placement based on multimodal contextual AI with live streaming. It starts with an analysis of architectural characteristics of managing multimodal inputs in real-time and moves on to the model strategies of performing fusion, inference, and ad selection. It then addresses deployment issues, performance measurements, present challenges, and future trends. Gluing these parts together through a constant narrative, the study will attempt to give a complete picture of this developing field.

2. Multimodal Contextual Understanding in Live Streaming

Having stated the necessity of adaptive and intelligent ad systems in live-streaming systems, we need to discuss how multimodal contextual AI is the cognitive core of real-time ad decision-making, as depicted in Figure 1. Multimodal AI implies the inclusion of multiple kinds of input data, of which each mode or modality is expressed in a different representation of the environment. Applied to live streaming, these modalities are video imagery, audio cues, transcription speech, user interaction patterns, chat data, biometric reactions (e.g., eye-tracking or facial expression), and even network metadata [11][12].

Video feeds provide spatial and semantic information that can aid in identifying the genre, tone, or type of content that is being streamed. As an example, a fast switch of scenes, movement recognition, or expression on a face can show the intensity of emotion or change of the subject. There is another layer of context added by audio inputs such as tone of speech, background music, and volume levels, which can reflect mood or even urgency. Speech-to-text engines are real-time engines that transform spoken words into text that can be analyzed with NLP models to gain topic, sentiment, and keywords [13][14]. Live chat activity is an unofficial feedback mechanism that records the responses of the audience in real-time. The use of sentiment analysis on chat posts and the response of viewers, including likes, emotes, or spikes in engagement, can give an idea of viewer emotion, attention, and resonance of the content. Such cues play a primary role in determining whether an advertisement placement would be received positively or viewed as obtrusive [15][16]. This is a challenge of multimodal data fusion. Early fusion strategies entail the combination of crude or low-level features of several modalities,

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

and then they are passed on to a common model. Although this methodology is a good way of getting deep correlations, it is also computationally expensive and may have issues of data sparsity. In late fusion, in contrast, the modality predictions or embeddings are independently processed, and then later the processes are combined. This provides flexibility and real-time responsiveness, but can miss intermodal interactions. Hybrid fusion, through attention mechanisms or transformer-based architectures, offers a compromise in that a dynamic weight to the contribution of each modality is provided depending on the relevance [17][18].

In order to process these data streams, the real-time multimodal AI systems normally incorporate a combination of convolutional neural networks (CNNs) to prepare image frames, recurrent neural networks (RNNs) or transformers to prepare text and audio sequences, and reinforcement learning to make adaptive decisions. These models are implemented on a streaming architecture, which allows time-windowed processing, buffering, and latency-constrained inference [19][20]. In addition to this, context windows are built in such a way that they contain not only the immediate scene but also the recent past content, allowing the model to comprehend narrative flow. As an illustration, a scene with a sports win, celebrations of fans, and music may be perfect to position a commercial with high energy. The AI system should be able to identify such instances and take action on them in a few milliseconds, which requires the model to perform well and the infrastructure to be high-performing [21][22].

Multimodal input is very rich, and this is a potent prism with which to appraise the suitability and timing of advertisements. However, to decode these clues correctly, one would need a lot of practice in a marked-up multimodal sample of multimedia that covers a variety of contents, audience groups, and cultures. Pre-trained vision-language models, including CLIP or VideoBERT, are typically transferred to a domain-specific task as a form of bootstrapping [23][24]. It is based on this basis of multimodal contextual knowledge that we can now look at the construction of adaptive ad placement engines, such that intelligent and personalized decision making can occur in real time. These engines use input of multimodal analysis to identify ad type, content, timing, and channel of delivery to be most effective.

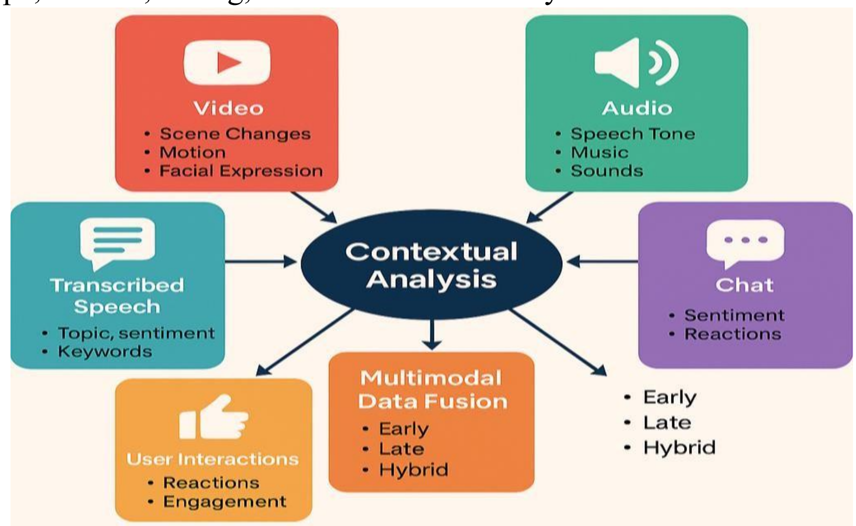

Figure 1: Diagram illustrating multimodal contextual understanding in live streaming, integrating video, audio, speech, chat, and user interactions for real-time contextual analysis and decision-making using early, late, or hybrid fusion techniques.

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

3. Adaptive Ad Placement Engines

As an extension of the contextual knowledge presented by the multimodal AI, the essence of an intelligent ad delivery system in live streaming is the design of adaptive ad placement engines. These engines are in charge of dynamically choosing, timing, and personalizing out on the dynamically changing analysis of the context of the live stream and the state of the viewer. The responsiveness of these engines enables them to react in milliseconds to the variation of scene content, viewers' interaction, and the mood of the ad, which will enhance the relevance and effectiveness of the advert. The core of adaptive ad placement is that of a real-time decision-making loop, which is fed with signals by the multimodal context detection system. This loop also assesses different contextual variables, including present content feeling, viewer interaction variables and ad fatigue variables, and previous response variables. As a result of these inputs, it does forecasting regarding when to place an ad and choosing among a set of creatives based on reinforcement learning or multi-armed bandit algorithms [25][26]. The main objectives of the ad placement engine are to enhance the contextual relevance and ensure the user is not interrupted in the process. To give an example, in case the AI notices that the stream is in a high-tension narrative scene, say, a live football game penalty kick, it does not show an ad even though the schedule says so. Rather, it leaves for a less turbulent time, when attention drop-off is less dangerous. Such dynamically planned deferrals and insertions are planned with the predictive attention models based on past streaming engagement data [27][28].

The advertisement choice module is based on a database of advertisements with metadata, including target audience, theme of the content, emotionality, and compliance tags. The real-time scoring of each ad creative is done against the current multimodal context vector to identify the most suitable match. As an example, in the case of audio and text analysis, the keywords such as celebration, victory, partying, etc., are included, and the music that is identified as the upbeat one, an advertisement of a drink or music streaming service with an equally cheerful mood may be prioritized [29][30].

To make it even more personal, user profile information, such as viewing history, behavioral segments, geographic location, and inferences, is also included in the system. This information enables the ad placement engine to modify the content and the format of ads, even as it ensures that it does not violate the privacy regulations. As an example, the ad can be shown to mobile users as an overlay, whereas to desktop users, the ad can be shown as a skippable video depending on the historical patterns of interaction [30]. The mechanism of real-time feedback is very important in adapting the system. These are click-through rates (CTR), dwell time, user behavior (likes, chat mentions, emoji reactions), and post ad dropout rates. Such measures are sent back to the placement engine to revise prediction models and re-rank ad creatives continuously. The agents of reinforcement learning maximize the reward functions that aim at balancing the advertiser's ROI and the user retention and satisfaction of the system, and therefore, the system becomes more effective with time [3][14].

Notably, contextual safety filters are incorporated into the engine to prevent inappropriate or non-compliant placements. As an illustration, in case of any tragedy or sensitive event, which

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

is identified in the stream (e.g., a news broadcast about a natural calamity), the system blocks an advertisement that might seem insensitive or irrelevant, regardless of the timing. They are determined by NLP-based topic detection and sentiment moderation models that are trained using a wide variety of sources of content [15][16]. Finally, the adaptive ad placement engines are the smart coordinators of delivering ads in live-streamed content. They are much more efficient compared to the non-reactive systems because of their responsiveness to the content and user dynamics. The following section goes further to outline the system architecture and data pipelines that can allow these engines to be used in real-time, at scale.

In order to describe the complexity of the decision-making in adaptive ad placement engines, the table below provides a comparative description of some of the major contextual signals employed in real-time ad decisioning, their likely data sources, qualities, and usage in ad selection reasoning.

Table 1. Contextual Signals for Adaptive Ad Decisioning in Live Streaming

Signal Type	**Typical Data Source**	**Characteristics**	**Use in Ad Placement**
Emotional Tone	Audio (speech, background music), facial cues	Dynamic, subjective, culturally sensitive	Avoids mismatch (e.g., sad tone → no upbeat ad)
Viewer Engagement	Clicks, emotes, chat frequency, likes	Fast-changing, high variance	Determines ad timing and insertion windows
Scene Semantics	Video frames, scene changes, object detection	Visual context requires deep learning	Aligns ad theme with content (e.g., sports)
Speech Content	ASR-transcribed speech, NLP pipelines	Topic-specific, interpretable	Selects ads with topical relevance
Viewer Device Context	Browser metadata, app usage, screen resolution	Technical, hardware-dependent	Chooses format (overlay, skippable, interactive)
Behavioral History	Watch history, skip patterns, dwell time	Persistent, personal	Enables personalized ad recommendations

The ability to interpret and act upon these heterogeneous signals in real time gives adaptive engines a significant edge over rule-based systems, particularly in live-streamed environments where both content and viewer states evolve rapidly.

4. System Architecture and Deployment Pipelines

After going through the adaptive ad engines, it is important to learn how the real-time placement of ads is done through the underlying system architecture, which supports

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

multimodal contextual AI. The architecture also has to be able to handle high-throughput, low-latency data ingestion, do inference on multimodal signals continuously, and have smooth integration with ad serving services, all without violating the strict time constraints of live broadcasting, as illustrated in Figure 2. The system is generally designed with four fundamental layers, which are data ingestion and pre-processing, contextual AI inference, ad decision and placement, and feedback and retraining. The layers sit atop a streaming data infrastructure that uses Apache Kafka as a message queue, Apache Flink or Apache Spark Streaming as a real-time processing system, and cloud-native services (e.g., AWS Kinesis or Google Dataflow) as a scalable and fault-tolerant system [17][18].

Ingestion layer captures live video feeds, audio streams, speech transcripts, chat logs, viewer interactions, and telemetry data of client devices. This multimodal data is synchronized by using timestamping and stored in memory buffers to be processed in the short term. At this step, compression, noise removal, and tokenization are done to prepare inputs to the inference layer [9]. The machine learning models that operate on individual modalities are deployed in the contextual AI inference layer. Video CNNs use pre-trained models to extract scene embeddings, audio recurrent or transformer models can understand tone and content, and text (speech and chat) models can do sentiment and entity recognition using BERT-style models. The outputs are combined into a context vector that summarizes the semantic and emotional condition of the stream at every point [10][21].

This context vector is then combined with user segmentation data and sent to the ad decision engine to decide if ad creatives are appropriate or not. This module communicates with the ad inventories and demand side platforms (DSPs) through API. it returns a plan of ad placement, such as timing, format, and creative ID, and this is now pushed to the ad rendering module on the client side. This handoff is time-constrained and has to be synchronized with the video playback timeline in order to create seamless transitions [12]. There is a difference in the deployment of edge-based and cloud-based systems. On platforms where mobile users are large, edge computing is used in tasks that require latency, like scene change detection and simple context inference. The fundamental AIs and decision logic are regularly executed in the cloud setting to take advantage of GPU acceleration, centralized data access, and distributed training infrastructure [13][24]. To provide reliability, the architecture has resilience features built into it, like hot failover instances, orchestration of containers through Kubernetes, and model versioning with rollback. CI/CD pipelines are used to automate retraining and deployment of models, which include continuous integration with new datasets and performance metrics. Last but not least is the feedback and retraining loop, which is used to optimize model parameters using post-ad telemetry, user interactions, and engagement data. This loop serves online learning structures that do not need to be retrained to adapt to dynamically changing viewer behavior and content type. Both batch data and streaming data are in use, and the retraining cycle is done weekly, daily, or even hourly, depending on the volume of data and the critical nature of the system [5][6]. Having such solid architecture, adaptive ad placement systems will be able to support millions of simultaneous users and make decisions in a split second, at a high level of confidence. This operational capability leads to

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

the second key element, namely the way the performance of the system is measured, and this aspect will define the effectiveness of the system and how it can be optimized.

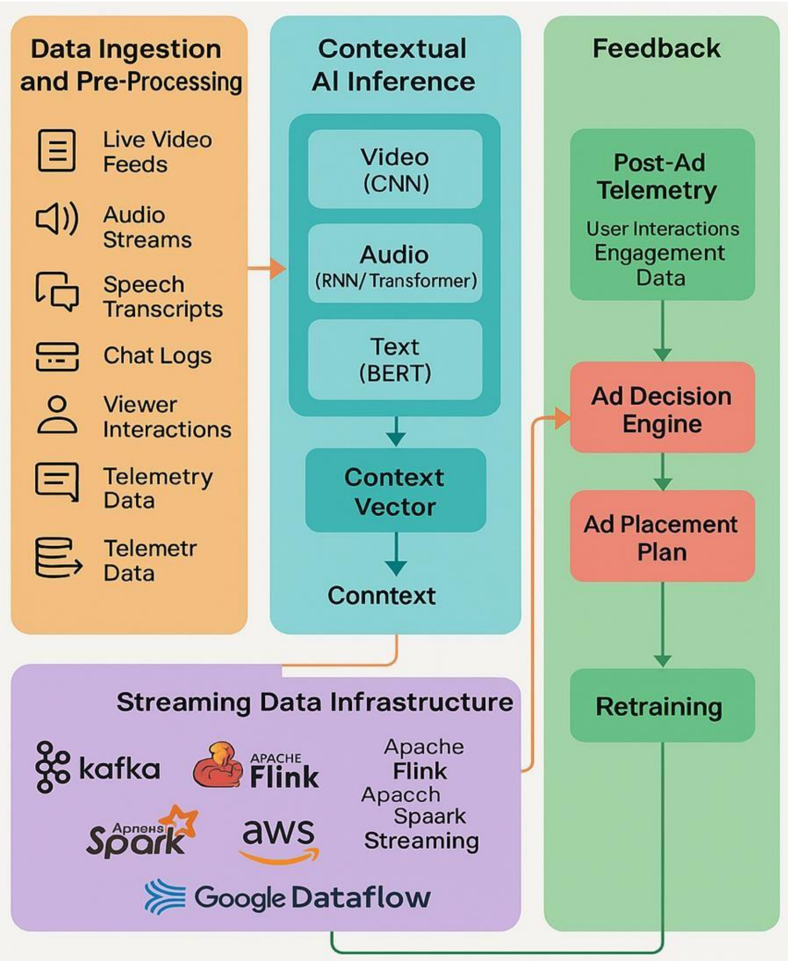

Figure 2: Layered architecture for real-time ad placement using multimodal contextual AI, showing the flow from data ingestion and contextual inference to ad decisioning and feedback, all supported by scalable streaming infrastructure.

5. Evaluation Metrics and Performance Analysis

Having dug into the system architecture and the operational pipeline of adaptive ad placement engines, it becomes clear that it is necessary to create powerful assessment structures that will be able to measure the performance of the system in various dimensions. In the live streaming dynamic and high-stakes industry, where an advert position will affect both audience experience and advertiser income, performance measurement accuracy is not only desirable but rather necessary.

Click-through rate (CTR) is considered to be one of the fundamental metrics applied to the effectiveness of adaptive ad placement systems. This metric determines the percentage of users who have clicked on an advertisement compared to the total number of users who saw the ad. Increased CTRs mean that contextually relevant and interesting ads are being displayed in the system. Nevertheless, CTR does not provide enough information, and it lacks user satisfaction or content interruption [1][2]. Engagement Retention Rate is used to deal with user-centric

Received: August 08, 2025

viewpoints. This is a percentage of the audience that does not leave the stream immediately after the placement of the ads. When the viewers watch full commercials every time they tune in, then that could be a sign that this timing is not good or that the context does not fit the video, despite the high CTR. Therefore, the systems would have to balance both monetization and viewer loyalty by maximizing retention and direct conversions [3][4].

Another important measure is Ad Relevance Score. This can be calculated based on feedback, such as user rating, implicit feedback like mouse hovers or dwell time, and qualitative feedback of A/B test groups. The score displays the level of compatibility of the ad with the streaming environment as defined by multimodal AI. It is an indirect measure of the contextual comprehension and the decision-making of the AI model [5][6]. Technically, Model Latency and Inference Throughput are the vital operation parameters. Latency is the duration of time required by the system to handle multimodal inputs to give an ad decision. This is normally calculated in milliseconds. In the case of live-streaming, latency limits can be modified to less than 500ms in order to make it real-time. Throughput is the number of ad decisions per second, or of the stream instance, and is particularly significant when serving millions of simultaneous users [7][8]. It can also be used with Precision, Recall, and F1-Score, particularly in assessing the content classification, sentiment detection, and ad-context matching accuracy. Precision assesses the percentage of ads that were placed and that were actually relevant, and Recall assesses the percentage of the relevant opportunities that have been used. F1-Score is a harmonic mean of both and gives an objective perspective of predictive quality of the model [9][10]. Revenue Per Thousand Impressions (RPM) and Return on Ad Spend (ROAS) are business-oriented sources of information about the financial efficiency of the adaptive system in the broader performance evaluation. Advertisers and operators of the platform use these metrics to evaluate whether the intelligent ad placement strategy can achieve a higher monetization than traditional approaches [11][12]. A/B testing is one of the main aspects of performance evaluation, during which alternative iterations of the ad placement engine (e.g., rule-based vs. AI-driven) are rolled out to random groups of users. The statistical analysis of results is conducted to define the significant improvement of CTR, retention, and engagement. It can also be used to optimize hyperparameters of the AI models and the fusion strategies applied to the multimodal input processing [13][14].

In addition to the quantitative measurements, the qualitative feedback is being increasingly integrated into the performance measurement. Post-ad survey, real-time chat log, and social media response viewer sentiment analysis provide fine-grained data on how the customers perceive the advertisement relevance and intrusiveness. Such indications can be used to identify more cultural/emotional sensitivity failures that would not be identified by quantitative models [15][16]. Lastly, the elements of drift detection are employed to track the changes in the effectiveness of the models over time. The metrics detect the changes in the behavior of the viewers, the popularity of the content genres, or the distribution of emotional tones. This could be an indication of concept drift since a decline in model accuracy or ad relevance score over time, leading to the need to retrain or reparametrize the model [17][18]. With the application of this multifaceted assessment system, platforms will be in a position to sustain high performance standards and make sure that user demands and expectations are met through

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

adaptive ad systems. However, despite these powerful tools, a number of issues still exist in this direction, especially the issues concerning scalability, privacy of data, and longevity of the models. These shall be discussed in the next section.

Although the foregoing discussion has presented the single performance measures, it is important to know how the measures benefit the various stakeholders within the advertising industry. The following table will classify evaluation measures depending on their relevance to the stakeholders and their overall objectives.

Table 2. Stakeholder-Oriented Metrics in Adaptive Ad Placement Evaluation

Metric	Primary Stakeholder	Objective	Measurement Frequency
Click-Through Rate (CTR)	Advertiser	Measure ad appeal and engagement	Per campaign, per session
Engagement Retention	Platform/Content Provider	Assess viewer drop-off and ad interruption effect	Per session, real-time
Revenue Per Mille (RPM)	Advertiser/Platform	Quantify monetization efficiency	Daily, hourly
Ad Relevance Score	Advertiser/User	Ensure context alignment of ads	Per ad impression
Model Latency	Engineering/Operations	Evaluate system responsiveness	Real-time
A/B Test Lift	Product/Analytics Teams	Compare system variants for performance gains	Weekly/experiment-based
Feedback Sentiment	User Experience Teams	Assess subjective response to ad experience	Post-session or post-ad

By aligning performance measurement with stakeholder goals, adaptive ad systems can be optimized holistically, balancing commercial returns with user experience and technical feasibility.

6. Challenges and Future Directions

Even though adaptive ad placement on the basis of multimodal contextual AI has proved to have considerable potential, several technical, operational, and ethical issues still remain on the way to its large-scale use. These problems are vital to the performance, user trust, and advertiser value in the long term.

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Among the greatest issues is multimodal data fusion in real-time. All modalities, such as video, audio, text, and interaction data, have their unique features regarding sampling rates, latency, data quality, and noise. To generate coherent context vectors in real time without adding bottlenecks, a sophisticated architecture design, such as asynchronous processing, buffer management, and late binding mechanisms, is necessary to synchronize these streams [19][20]. Another issue closely related to this is the problem of model scalability. Since live-streamed content is growing in terms of its variety and territories, the system should be able to accommodate a large variety of content and cultural standards. Pre-trained models usually have difficulties in domain generalization and have to be retrained in region-specific or topic-specific data. This increases the cost and complexity of operation in deployment [21][22].

Limits are also provided by data sparsity and bottlenecks in annotation. Although supervised learning methods are effective when you have labeled data, it is costly and time-consuming to gather quality multimodal annotations, and in particular, the ability to capture finer emotional nuance or be contextually sensitive. Alternative methods under investigation include semi-supervised learning, transfer learning, and weak supervision, which are not yet developed enough to bridge the gap fully [23][24]. Privacy and data ethics are another issue. Behavioural data, chat logs, and biometric data (e.g., facial expressions) into an AI system are very concerning when it comes to user consent and the use of data. Such regulatory frameworks as the General Data Protection Regulation (GDPR) provide tight restrictions on data collection and processing. To make sure that the adaptive ad placement systems do not violate these frameworks, a combination of anonymization mechanisms, on-device processing, and transparent data policy is necessary [25][26]. The interpretability and explainability of models have not been developed well in real-time situations. Transformer-based models and deep learning architectures are effective, but they are frequently black boxes, and thus it is hard to examine the process or determine why a specific ad was shown at a specific time. This transparent absence of transparency can lead to loss of trust by the users and advertisers. Explainable AI (XAI) systems are being built, which can give real-time information about the decision paths of models, but efforts to integrate them into systems with low latency remain an active area of development [27][28].

Infrastructurally, edge computing is a challenge and opportunity. Latency can be minimized by placing the units of the adaptive system (e.g., lightweight models to identify the context) on client machines, as well as maintaining privacy. But consistency, synchronization, and version control among distributed edge nodes are more complex in architecture [29]. In the future, research and development are taking a turn towards a number of potentially promising innovations. Federated learning is a new framework that is potentially a strong privacy-preserving model training system across distributed data sources. In this model, models are trained on the user devices and only model updates are exchanged, thus keeping raw data confidential. This may allow personalization on a fine-grained level with no loss of user trust [30].

One more direction of work is the use of multimodal transformer architectures, which expect different types of data to be processed with a single model, like VideoBERT, MMF, and

Received: August 08, 2025

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Perceiver IO. These models have demonstrated excellent results in the offline case and are being translated to real-time inference cases. They can be made possible in even latency-sensitive applications, paired with techniques like model distillation and quantization. Lastly, the future of adaptive ad placement is its combination with larger content intelligence systems that do not just deal with ads but also with recommendations, content moderation, and viewer interaction streams. The next frontier is to create a unified, smart system of content delivery that meets the needs of the users and maximizes business goals. To summarize, adaptive ad placement based on multimodal contextual AI is a radical transformation of digital advertising, as it makes it not only reactive but also perceptive and responsive. As the model architecture, data privacy, real-time architecture, and user-friendly design continue to evolve, the industry will see live-streamed advertising cease to be a disruption to the viewer experience and instead will become a natural and uninterrupted continuation.

7. Conclusion

The changing nature of the environment of digital media consumption, as a result of the influx of live-streamed content, has led to the need to change the paradigm of delivering, contextualizing, and experiencing advertisements. The conventional, unchanging modes of advertising placement are becoming less and less sufficient to mirror the adaptable and subtle quality of live content and real-time audience interaction. To address this change, the addition of multimodal contextual AI offers a radical potential to provide adaptive, personalized, and contextually relevant ads on live streaming platforms. The paper has discussed how the adaptive ad placement system based on AI models that consider video, audio, text, and user interaction data, and intelligent assessment of the streaming context in real time can be accessed. Using an attentional fusion technique, sentiment analysis, and reinforcement learning, such systems are capable of delivering high-impact ad decisions that can improve user experience and improve advertising performance. The additional real-time feedback mechanisms allow constant learning, system tuning, and optimization, so that the models are kept up-to-date with the changing viewer behaviour and semantics of the content.

The architecture that allows such systems is naturally a complex one and consists of scalable streaming data pipelines, latency-sensitive inference engines, and user data management in a way that is privacy-compliant. The issues of data heterogeneity, concept drift, lack of labeled data, and privacy requirements highlight the importance of cautious engineering and ethical design. However, developments of federated learning, multimodal models based on transformers, and explainable AI are also leading to stronger, more efficient, and transparent systems. In the future, the further development of multimodal contextual AI, along with edge computing and real-time analytics, will not only enhance the performance of ads but also transform the interaction of content creators, advertisers, and viewers in the context of live digital spaces. The emerging trend in which entertainment, business, and live interaction become increasingly indistinct, adaptation and placement will form an inseparable part of intelligent delivery systems that will bring a day when advertising is not only perceived but actually felt in context.

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

References

1. Wang, S., Hu, C., & Jia, G. (2024). Deep Learning-Based Saliency Assessment Model for Product Placement in Video Advertisements. *Journal of Advanced Computing Systems*, *4*(5), 27-41.
2. Yuan, Y., Wang, F., Li, J., & Qin, R. (2014, October). A survey on real-time bidding advertising. In *Proceedings of 2014 IEEE International Conference on Service Operations and Logistics, and Informatics* (pp. 418-423). IEEE.
3. Rajaram, P. (2021). *Modeling Viewer and Influencer Behavior on Streaming Platforms* (Doctoral dissertation).
4. He, L., Chen, H., Wang, D., Jameel, S., Yu, P., & Xu, G. (2021, October). Click-through rate prediction with multi-modal hypergraphs. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management* (pp. 690-699).
5. Soleymani, M., Garcia, D., Jou, B., Schuller, B., Chang, S. F., & Pantic, M. (2017). A survey of multimodal sentiment analysis. *Image and Vision Computing*, *65*, 3-14.
6. Moreno-Armendáriz, M. A., Calvo, H., Duchanoy, C. A., Lara-Cázares, A., Ramos-Díaz, E., & Morales-Flores, V. L. (2021). Deep-learning-based adaptive advertising with augmented reality. *Sensors*, *22*(1), 63.
7. Alam, M. A., Nabil, A. R., Mintoo, A. A., & Islam, A. (2024). Real-time analytics in streaming big data: techniques and applications. *Journal of Science and Engineering Research*, *1*(01), 104-122.
8. Wu, J., Wang, H., Qian, K., & Feng, E. (2023). Optimizing latency-sensitive AI applications through edge-cloud collaboration. *Journal of Advanced Computing Systems*, *3*(3), 19-33.
9. Kumar, P. (2024). AI-Powered Fraud Prevention in Digital Payment Ecosystems: Leveraging Machine Learning for Real-Time Anomaly Detection and Risk Mitigation. Journal of Information Systems Engineering and Management 2024, 9(4) e-ISSN: 2468-4376
10. Jin, Y., Wanvarie, D., & Le, P. (2017, November). Combining lightly supervised text classification models for accurate contextual advertising. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)* (pp. 545-554).
11. Almestarihi, R., Ahmad, A. Y., Frangieh, R. H., Abualsondos, I. A., Nser, K. K., & Ziani, A. (2024). Measuring the ROI of paid advertising campaigns in digital marketing and its effect on business profitability.
12. Kumar, P. (2025). Securing Digital-First Healthcare: AI, Blockchain, and Cloud Architectures for Personal Health Data Protection. International Journal of Applied Mathematics, 38(7s).
13. Kim, J., Maathuis, H., & Sent, D. (2024). Human-centered evaluation of explainable AI applications: a systematic review. *Frontiers in Artificial Intelligence*, *7*, 1456486.

Received: August 08, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

14. Rahman, H., Chowdhury, M. T., Sumon, M. I., Mim, F. M., Rashid, M. R. A., Islam, R. U., & Hasan, M. (2024, September). Analyzing A/B Testing Results in Digital Marketing. In *2024 2nd International Conference on Artificial Intelligence, Blockchain, and Internet of Things (AIBThings)* (pp. 1-5). IEEE.
15. Silverman, G., Te'eni, D., Schwartz, D. G., Mann, Y., Cohen, D., & Lewinsky, D. (2025). A hybrid mixed methods design of qualitative enhancement and reciprocal feedback loop for augmented text classification. *Quality & Quantity*, 1-22.
16. Kumar, P. (2025) Next-generation secure authentication and access control architectures: advanced techniques for securing distributed systems in modern enterprises. International Journal of Computational and Experimental Science and Engineering (IJCESEN) Vol. 11-No.3, pp. 4966-4995
17. Kraus, A., & van der Aa, H. (2025). Machine learning-based detection of concept drift in business processes. *Process Science*, *2*(1), 1-26.
18. Dounas, L., Mazo, R., Salinesi, C., & El Beqqali, O. (2015, November). Continuous monitoring of adaptive e-learning systems requirements. In *2015 IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA)* (pp. 1-8). IEEE.
19. Lahat, D., Adali, T., & Jutten, C. (2015). Multimodal data fusion: an overview of methods, challenges, and prospects. *Proceedings of the IEEE*, *103*(9), 1449-1477.
20. Yusupov, A., Park, S., & Kim, J. (2024). Synchronized Delay Measurement of Multi-Stream Analysis over Data Concentrator Units. *Electronics*, *14*(1), 81.
21. Xia, Y., Huang, H., Zhu, J., & Zhao, Z. (2023). Achieving cross modal generalization with multimodal unified representation. *Advances in Neural Information Processing Systems*, *36*, 63529-63541.
22. Xu, Y., Cao, H., Xie, L., Li, X. L., Chen, Z., & Yang, J. (2024). Video unsupervised domain adaptation with deep learning: A comprehensive survey. *ACM Computing Surveys*, *56*(12), 1-36.
23. Daunhawer, I. (2023). *Multimodal Representation Learning under Weak Supervision* (Doctoral dissertation, ETH Zurich).
24. Zhang, L., Gao, G., & Zhang, H. (2022, November). Towards data-efficient continuous learning for edge video analytics via smart caching. In *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems* (pp. 1136-1140).
25. Sanchez, D., & Viejo, A. (2018). Privacy-preserving and advertising-friendly web surfing. *Computer Communications*, *130*, 113-123.
26. Smith, M., Torres-Agüero, A., Grossman, R., Sen, P., Chen, Y., & Borcea, C. (2024, May). A study of gdpr compliance under the transparency and consent framework. In *Proceedings of the ACM Web Conference 2024* (pp. 1227-1236).

27. Khan, W., Ahmad, J., Alasbali, N., Al Mazroa, A., Alshehri, M. S., & Khan, M. S. (2025). A novel transformer-based explainable AI approach using SHAP for intrusion detection in vehicular ad hoc networks. *Computer Networks*, 111575.

28. Prince Kumar. (2025). Agentic ai-driven enterprise architecture: a foundational framework for scalable, secure, and resilient systems. *International Journal of Computational and Experimental Science and Engineering*, *11*(4).

29. Murshed, M. S., Murphy, C., Hou, D., Khan, N., Ananthanarayanan, G., & Hussain, F. (2021). Machine learning at the network edge: A survey. *ACM Computing Surveys (CSUR)*, *54*(8), 1-37.

30. Lo, S. K., Lu, Q., Zhu, L., Paik, H. Y., Xu, X., & Wang, C. (2022). Architectural patterns for the design of federated learning systems. *Journal of Systems and Software*, *191*, 111357.