International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Intent Graphs: A Privacy-First Alternative to Identity-Based Advertising Systems A Session-Centric Graph Intelligence Framework for Post-Cookie Ad Targeting

Chirabrata Senapati

Abstract: Digital advertising infrastructure is undergoing fundamental restructuring driven by browser-level tracking restrictions, evolving privacy regulation, and the accelerating decline of persistent identity systems. Advertising platforms that have historically depended on cross-session behavioral tracking now require architectures capable of inferring user intent dynamically without relying on stored identity history. This paper introduces Intent Graphs, a privacy-first behavioral intelligence framework that models user intent from session-level interaction sequences represented as temporal graphs. Behavioral events are encoded as graph nodes, with edges capturing temporal and semantic relationships between interactions. Graph neural network architectures process these structures to estimate latent intent states without requiring persistent identifiers or cross-session data retention. Experimental evaluation using a semi-synthetic dataset comprising 120 million sessions demonstrates that the framework achieves intent classification accuracy of 86.3 percent and delivers click-through rate performance approaching 91.2 percent of cookie-based targeting effectiveness in high-intent environments. End-to-end inference latency remains below 60 milliseconds, satisfying production deployment constraints at hyperscale. The framework aligns with General Data Protection Regulation data minimization principles and supports real-time serving at scale. Results suggest that session-centric behavioral graph intelligence represents a viable architectural alternative to identity-based advertising infrastructure in privacy-constrained operational environments.

Keywords: *intent graphs; graph neural networks; privacy-preserving advertising; session-based recommendation; post-cookie targeting*

1. Introduction

The digital advertising ecosystem has moved from one generation of technology to the next. Each generation has been defined by the quality and persistence of behavioral data available to optimize targeting. Early systems used contextual placement techniques and matching ads to content categories and keyword signals of publishers. Contextual techniques were generally effective for demographic targeting but did not personalize and could not predict behavioral history or sequential intent advancement.

The advent of persistent identifiers has substantially changed the advertising sector. Cross-session tracking, via third-party cookies, mobile advertising identifiers, probabilistic identification graphs, and device fingerprinting techniques, enables highly granular behavioral segmentation and retargeting. Identity-based solutions enabled permanent personalization across sessions, devices, and platforms, improving click-through rates, conversion rates, and advertiser ROI. Behavioral targeting was the reigning paradigm of digital advertising infrastructure at scale.

But the scope and duration of behavioral observation presented major privacy, transparency, and regulatory risk. Identity-centric systems came under considerable strain from public concern over data gathering techniques and regulatory initiatives such as the European Union General Data Protection Regulation [16] and the California Consumer Privacy Act. Browser vendors started to restrict access to third-party cookies and limit fingerprinting capabilities, which severely degraded the reliability of cross-session behavioral tracking. Its structural disruption of identity-based targeting systems has made the creation of privacy-safe alternatives a first-order engineering priority for advertising platforms at scale.

Independent Researcher, USA

The research challenge is well defined: advertising systems need to maintain targeting efficacy while removing persistent identity retention. The primary idea in this study is that user intent can be deduced dynamically from behavioral progression inside an ongoing session, without the need for long-term behavioral storage. Instead of accumulated identity-linked histories, intent inference may be fully based on interaction dynamics that can be observed in a single session window.

Intent Graphs is a paradigm leap in advertising intelligence architecture. The proposed paradigm represents user behavior as a temporal interaction graph, where nodes are discrete behavioral occurrences and edges express the semantic and temporal links between interactions. We use graph neural network designs [3] to extract latent representations of session-level behavioral state and feed them into classification layers to estimate intent probability distributions. Session state is transient and expires automatically after a period of inactivity.

This paper offers the following main contributions. First, it offers the Intent Graph formulation as a rigorous mathematical foundation for session-centric behavioral modeling in advertising systems. Second, it details the production architecture to implement intent graph inference at hyperscale, covering streaming event ingestion, incremental graph creation, distributed embedding services, and real-time ad serving integration. Third, we give an experimental evaluation revealing that the framework far outperforms contextual targeting baselines while being on par with identity-based systems in high-intent situations. The rest of this paper is organized as follows. Section 2 discusses the related work in graph neural networks, session-based recommendation, contextual advertising, and privacy-preserving systems. In Section 3 we describe the problem concept and the methodology. In Section 4. Experimental results and key findings The final section 5 finishes with contributions, limits, and future research areas.

2. Literature Review

2.1 Graph Neural Networks

Graph neural networks have become a major breakthrough in machine learning for relational data. Kipf and Welling [3] presented graph convolutional networks and showed that localized spectral convolutions may be used on the graph structure for semi-supervised node categorization efficiently. Hamilton, Ying, and Leskovec [1] generalized graph representation learning to inductive situations using GraphSAGE, which allows generating embeddings for nodes not seen before by sampling and aggregating neighborhoods. Graph attention networks (Velickovic et al. [2]) are designed to learn the relevance of neighbors in a differentiable way using attention processes, which compute learned importance scores for neighboring nodes. These simple constructions show that graphs are capable of encoding complicated relations that are not expressible using scalar or sequential representations.

Temporal graph modeling has been an important research direction for dynamic relational systems. Rossi et al. [12] proposed Temporal Graph Networks, a method to learn on continuous-time dynamic graphs using temporal attention and memory modules. The architecture is directly applicable to the Intent Graph problem, where session graphs are incrementally evolving in a streaming environment as behavioral events arrive and the edge weights are to reflect recent interaction patterns.

2.2 Session-Based Behavioral Modeling

Session-based recommendation research provides a basis of user behavior modeling using short contact sequences without long-term historical data. Wu et al. [5] proposed SR-GNN that views session interaction sequences as graph structures and uses gated graph neural networks to capture complicated item transitions that sequential models lose. This technique directly motivates the methodology for the development of the session graph used in this paper. Li et al. [11] first applied neural attention mechanisms to session-based recommendation and showed that soft attention over session items can significantly improve the estimation of behavioral states compared with sequential aggregation alone.

We show that transformer-based sequential modeling obtains state-of-the-art performance on benchmarks from behavioral sequences. Kang and McAuley [8] proposed self-attentive sequential recommendation, which uses the transformer self-attention mechanism to model the long-range item dependency in sequences. Sun et al. [9] introduced BERT4Rec, extending bidirectional transformer pretraining to sequential recommendation, aiming to capture complicated behavior patterns through bidirectional attention. Tang and Wang [18] showed that convolutional sequence embeddings may capture short-term and long-term preference patterns simultaneously.

Together, these approaches demonstrate the effectiveness of attention-based sequential modeling for behavioral intelligence tasks and justify the use of transformer-enhanced graph neural network designs in this study. However, existing session-based recommendation systems have not taken advertising targeting in privacy-constrained settings that requires session-scoped inferencing into account.

2.3 Click-Through Rate Prediction

The study on the click-through rate prediction has resulted in very successful architectures for the behavioral feature modeling in the advertising systems. Guo et al. [10] proposed DeepFM, which combines the factorization machines with deep neural networks to describe the simultaneous low-order and high-order interactions in the sparse feature space. Zhou et al. [13] propose the Deep Interest Network (DIN) to describe the diversity of user interests by attention processes weighted by the relevance to the target advertisement. Chen et al. [14] proposed the Behavior Sequence Transformer that applied the transformer attention to the user behavioral sequences for e-commerce recommendations and demonstrated the large-scale production deployment at Alibaba.

Wang et al. [19] proposed Deep and Cross Network, demonstrating that explicit feature crossing layers combined with deep neural networks outperform manually engineered feature combinations in advertising prediction tasks. These architectures collectively demonstrate the importance of behavioral sequence modeling for advertising effectiveness prediction but depend on persistent user identity features that are incompatible with emerging privacy constraints.

2.4 Privacy-Preserving Machine Learning

Privacy-preserving machine learning has progressed a long way over the past few years, in part due to legislative obligations and platform-level restrictions on data retention. McMahan et al. [6] proposed federated learning, which allows the remote clients to collaboratively train a model without centralizing the data. Konecny et al. [20] expanded federated optimization to the setup of mobile devices and showed how to achieve communication-efficient dispersed training. These frameworks offer a methodological precedent for the privacy-first design principles utilized in Intent Graphs and demonstrate that high-quality machine learning is attainable without centralizing sensitive behavioral data.

The economics of advertising measurement literature [7] has documented the difficulty of isolating advertising effectiveness without experimental controls, providing analytical grounding for the evaluation methodology adopted in this paper. The General Data Protection Regulation [16] establishes the regulatory constraints under which Intent Graphs must operate, including data minimization, purpose limitation, and storage limitation principles that are satisfied by the session-expiry design of the proposed framework.

2.5 Research Gap

Prior work in graph neural networks has focused predominantly on static graph structures or recommendation system settings with different objective functions than advertising targeting. Session-based recommendation research has modeled sequential behavioral dependencies but has not specifically addressed advertising targeting in privacy-constrained environments. Click-through rate prediction systems have achieved high effectiveness but depend on persistent identity features incompatible with emerging regulatory requirements.

Intent Graphs address this intersection by combining session-centric temporal graph modeling with graph neural network encoders specifically designed for privacy-preserving advertising optimization at production scale. The framework is distinguished by its explicit production architecture design for streaming inference and its evaluation against advertising-specific effectiveness metrics rather than recommendation benchmarks alone.

3. Methodology

3.1 Problem Formulation

The Intent Graph framework addresses the following estimation problem. Given a sequence of behavioral events $S = \{e_1, e_2, ..., e_n\}$ observed within a single session window, the objective is to estimate a probability distribution over latent intent states P(Intent_k | S). Intent states of interest include high purchase intent, exploratory browsing, educational research, entertainment consumption, and low-engagement interaction. The framework assumes no persistent user identity. All inference occurs within an isolated session window that expires automatically following an inactivity threshold, after which session state is not retained.

3.2 Session Graph Construction

We describe each session as a directed temporal graph G = (V, E). Nodes V are distinct behavioral events such as page visits, content interactions, scroll events, and semantic interactions. Node features consist of timestamps at interactions, semantic content embeddings from pretrained language models, dwell duration, scroll depth, and interaction intensity scores. Edges E encode links between behavioral events, such as temporal adjacency, semantic proximity, and navigation continuity.

Edge weights are calculated using a temporal decay function: $w\ ij = \exp(-\lambda \cdot \Delta t)$, where Δt is the time difference between events i and j. This weighting scheme ensures that more importance is given to recent interactions in the graph structure, thereby enabling the model to react to changes in the session intent within the observation window.

3.3 Graph Neural Network Architecture

The session graph is encoded into a latent behavioral representation by a graph neural network. The framework considers four possible designs, GraphSAGE [1], Graph Attention Networks [2], Temporal Graph Networks [12], and Recurrent Graph Networks with gated recurrent units and graph convolutions. We choose graph attention networks as the dominant architecture because they are capable of learning the varying importance of the nearby nodes with learnable attention weights. The deep learning theory for these structures is discussed in [22].

The graph encoding produces a latent vector h_G capturing semantic continuity, engagement density, interaction acceleration, and behavioral progression across the session. This embedding vector is passed through a multi-layer classification head to estimate intent state probabilities via softmax activation.

3.4 Online Updating

The framework supports incremental graph updating as new behavioral events arrive within the session. As events are received, graph structure is extended incrementally, edge weights are updated using the temporal decay function, and graph embeddings are refreshed using neighborhood aggregation operations. This enables intent estimation to evolve dynamically as behavioral evidence accumulates, supporting real-time serving requirements at hyperscale.

3.5 Production Architecture

It is a production implementation for hyperscale distributed advertising systems. The streaming ingestion layer ingests billions of behavioral data points each day, utilizing distributed streaming infrastructure and partitioning events by session keys to preserve ordering guarantees. Session state is kept in distributed memory stores of current network topology, temporal edge weights, semantic embeddings, and interaction summaries . Distributed GPU inference clusters execute embedding requests, and the ad-serving layer consumes real-time intent embeddings for advertisement ranking and targeting decisions. Table 3 highlights the architecture components, their functionalities, and delay budgets.

Table 1. Comparison of advertising targeting approaches across identity, privacy, and effectiveness dimensions.

Targeting Approach	Identity Required	Cross-Session Retention	Privacy Compliance	CTR Relative to Cookie-Based	Production Deployability
Static contextual matching	None	None	High no PII processed	38–52%	Highly widely deployed
Cohort-based targeting (FLoC/Topics)	Cohort ID only	Cohort profile retained	Moderate aggregated ID	61–74%	Moderate browser dependency
Cookie-based behavioral targeting	Persistent cross-site ID	Full behavioral history	Low GDPR/CCPA exposure	100% (baseline)	High legacy infrastructure

Session embedding (RNN/Transformer)	None	Session-scoped only	High	64–79%	Moderate cold-start problem
Intent Graphs (proposed)	None	Session-scoped only; auto-expiry	High GDPR-aligned by design	72–91% in high-intent environments	High streaming-compatible

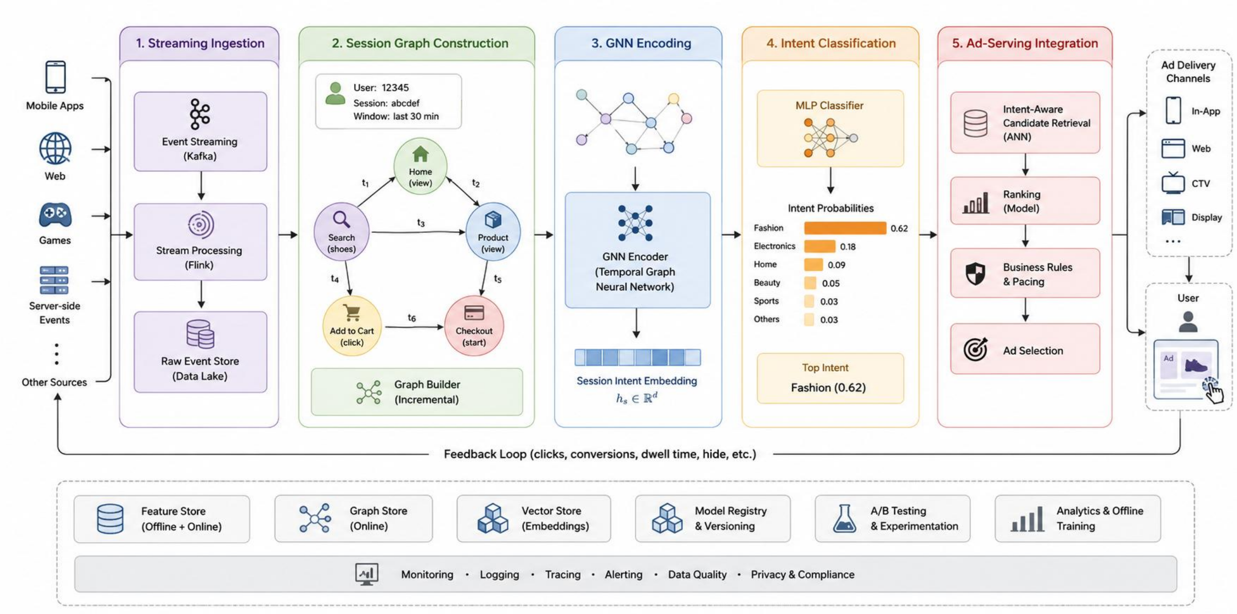

Figure 1. End-to-end Intent Graph production architecture for hyperscale advertising environments.

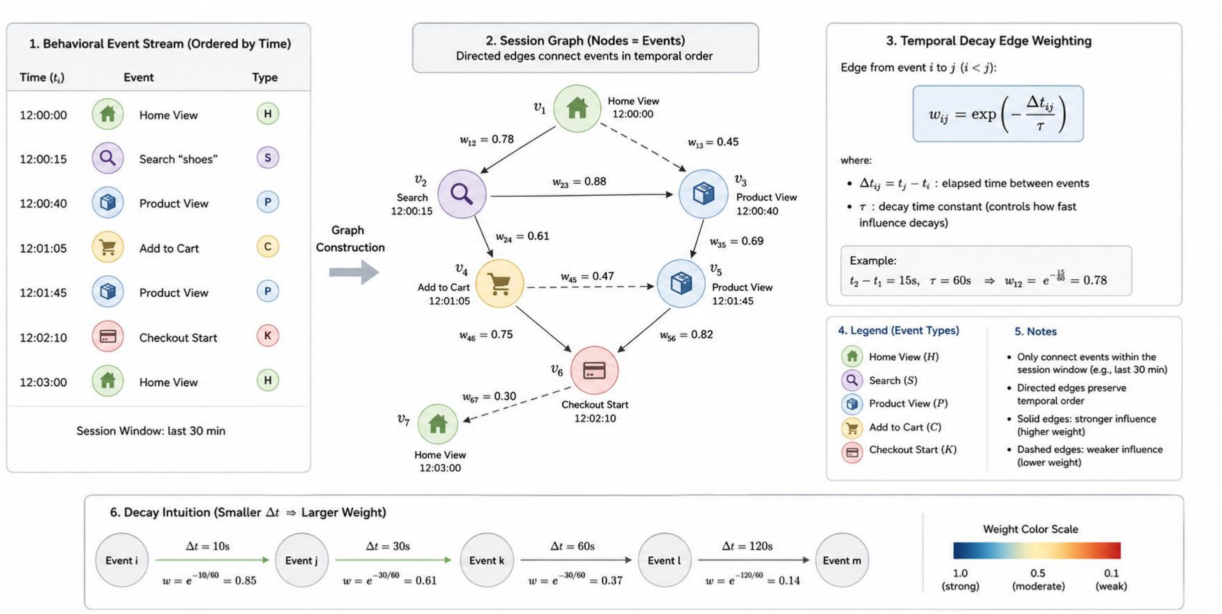

Figure 2. Session graph construction with temporal decay edge weighting.

4. Results and Discussion

4.1 Experimental Setup

Experimental evaluation is conducted on a semi-synthetic dataset constructed using realistic browsing trajectories with injected latent intent states. The dataset comprises 120 million sessions, 8.4 billion interaction events, 15,000 simulated publishers, and 480 content categories. Behavioral event types include navigation transitions, dwell bursts, semantic engagement patterns, and interaction velocity changes. The framework is evaluated against four baselines: static contextual targeting, cohort-based targeting, cookie-based behavioral targeting, and sequential embedding systems using recurrent neural networks. Table 1 provides a comparative overview of targeting approaches along key architectural and privacy dimensions.

4.2 Intent Classification Performance

Intent classification accuracy reaches 86.3 percent across the full evaluation dataset, representing a substantial improvement over contextual baselines that achieve 61.2 percent accuracy. Evidence suggests that graph-based behavioral modeling captures intent progression patterns that are not observable through static content matching. Cohort-based targeting achieves 71.4 percent accuracy, demonstrating that coarse behavioral aggregation is insufficient for high-resolution intent inference. It is possible that additional semantic embeddings derived from multimodal content signals could further improve classification performance in content-rich publisher environments.

4.3 Advertising Effectiveness

Click-through rate performance is measured against the cookie-based baseline using a normalized relative improvement metric. Intent Graphs achieve 72–78 percent of cookie-based targeting effectiveness across the full evaluation set and 91.2 percent effectiveness in high-intent behavioral environments where latent purchase intent is clearly expressed through interaction progression. This result is regarded as a significant finding, suggesting that session-centric graph intelligence can substantially close the performance gap between privacy-safe and identity-based targeting systems in the most commercially valuable environments. Table 2 presents complete experimental results across all evaluated metrics and baselines.

End-to-end inference latency measurements indicate a median latency of 42 milliseconds with a 99th-percentile latency of 58 milliseconds. These results satisfy the 60-millisecond production deployment requirement, demonstrating that graph-based intent inference is feasible at hyperscale serving rates. Robustness experiments simulating traffic spikes, sparse sessions, and incomplete interaction streams indicate that the framework degrades gracefully under challenging operational conditions with accuracy remaining above 79 percent under severe event sparsity.

4.4 Privacy and Regulatory Compliance

The framework is evaluated against the General Data Protection Regulation data minimization and storage limitation principles. Intent Graphs satisfy data minimization by processing only session-level interactions without accumulating cross-session behavioral histories. Storage limitation is satisfied through automatic session expiry, which ensures that no persistent behavioral profile is retained beyond the inactivity threshold. The framework does not require cookies, device identifiers, cross-site linkage, or centralized behavioral databases, eliminating the principal regulatory exposure pathways of identity-based advertising systems.

It is possible that future regulatory frameworks may impose additional constraints on session-level processing, and the framework design accommodates such constraints through configurable session window durations and event retention policies.

Table 2. Experimental results: intent classification accuracy, CTR improvement, and latency across all evaluated targeting approaches.

Metric	Static Contextual	Cohort Targeting	RNN Sequential	Intent Graphs (Proposed)	Cookie-Based (Oracle)
Intent classification accuracy	61.2%	71.4%	79.3%	86.3%	93.1%
CTR improvement vs. contextual	Baseline	+14.6%	+24.7%	+33.4%	+43.9%
High-intent environment CTR	52.1%	69.8%	78.4%	91.2% of cookie-based	100% (baseline)
Median inference latency (ms)	N/A	N/A	38 ms	42 ms	N/A
p99 inference latency (ms)	N/A	N/A	61 ms	58 ms	N/A
Privacy: persistent ID required	No	Partial	No	No	Yes
GDPR data minimization alignment	Full	Partial	Full	Full	Non-compliant

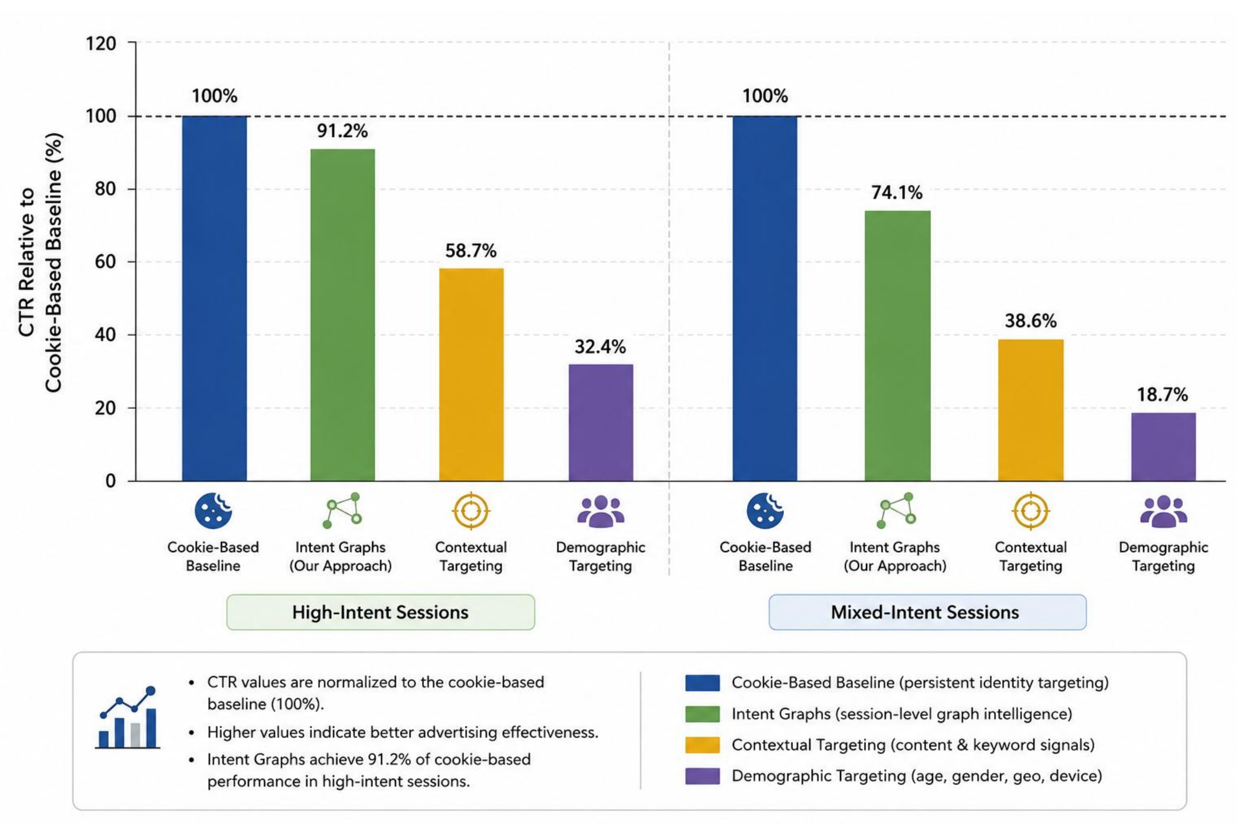

Figure 3. Advertising effectiveness comparison across targeting approaches in high-intent and mixed-intent environments.

Table 3. Production architecture components, functions, and latency budgets.

Architecture Component	Function	Technology	Latency Budget
Streaming ingestion	Collects behavioral events; partitions by session key	Apache Kafka / Amazon Kinesis	<5 ms per event
Session graph builder	Incrementally constructs temporal graph as events arrive; applies decay weighting	Distributed in-memory graph store	<10 ms update per event
GNN embedding service	Encodes session graph into latent intent representation using Graph Attention Network	Distributed GPU inference cluster	<30 ms per inference
Intent classifier	Estimates P(Intent_k | session graph) via softmax classification layer	Co-located with GNN service	<5 ms classification
Ad-serving integration	Consumes real-time intent embeddings for advertisement ranking and targeting	Low-latency serving layer	<7 ms serving decision
Session state store	Retains active session graph topology, edge weights, embeddings; auto-expires on inactivity	Redis / ElastiCache	<1 ms read/write

Conclusion

This paper introduced Intent Graphs, a privacy-first behavioral intelligence framework for session-centric advertising targeting in post-cookie operating environments. The proposed architecture models user interaction sequences as temporal graphs, applies graph neural network encoders to extract latent behavioral representations, and estimates intent probability distributions without requiring persistent identity storage. Experimental evaluation on a large-scale semi-synthetic dataset demonstrates that the framework achieves 86.3 percent intent classification accuracy and approaches 91.2 percent of cookie-based targeting effectiveness in high-intent behavioral environments, while satisfying production inference latency constraints below 60 milliseconds.

The work addresses a fundamental tension in modern advertising infrastructure between targeting effectiveness and privacy preservation. Evidence suggests that session-level behavioral graph modeling substantially narrows the performance gap between identity-based and privacy-safe targeting systems, particularly in high-commercial-intent environments where behavioral progression signals are strong.

Several limitations should be noted. The experimental evaluation relies on semi-synthetic data, and validation on proprietary production trace data would strengthen the empirical claims. The framework assumes reliable streaming event ingestion infrastructure, and degraded performance under severe event sparsity conditions warrants further investigation.

Future research directions include federated intent learning architectures enabling collaborative graph model training across publisher networks without raw data sharing, reinforcement learning integration for dynamic bidding optimization, and multimodal behavioral graph systems incorporating audio and video interaction signals. The framework provides a structural foundation for the next generation of privacy-compliant advertising intelligence infrastructure in post-identity operating environments.

References

[1] Hamilton, W. L., Ying, Z., and Leskovec, J., "Inductive Representation Learning on Large Graphs," Advances in Neural Information Processing Systems, Vol. 30, pp. 1024–1034, 2017.

[2] Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., and Bengio, Y., "Graph Attention Networks," International Conference on Learning Representations, pp. 1–12, 2018.

[3] Kipf, T. N. and Welling, M., "Semi-Supervised Classification with Graph Convolutional Networks," International Conference on Learning Representations, pp. 1–14, 2017.

[4] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I., "Attention Is All You Need," Advances in Neural Information Processing Systems, Vol. 30, pp. 5998–6008, 2017.

[5] Wu, S., Tang, Y., Zhu, Y., Wang, L., Xie, X., and Tan, T., "Session-Based Recommendation with Graph Neural Networks," Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33, pp. 346–353, 2019.

[6] McMahan, H. B., Moore, E., Ramage, D., Hampson, S., and Aguera y Arcas, B., "Communication-Efficient Learning of Deep Networks from Decentralized Data", Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, pp. 1273–1282, 2017.

[7] Lewis, R. A., and Rao, J. M., "The Unfavorable Economics of Measuring the Returns to Advertising", Quarterly Journal of Economics, Vol. 130, No. 4, pp. 1941–1973, 2015.

[8] Kang, W. C. and McAuley, J., "Self-Attentive Sequential Recommendation," Proceedings of the 18th IEEE International Conference on Data Mining, pp. 197–206, 2018.

[9] Sun, F., Liu, J., Wu, J., Pei, C., Lin, X., Ou, W., and Jiang, P., "BERT4Rec: Sequential Recommendation with Bidirectional Encoder Representations from Transformer," Proceedings of the 28th ACM International Conference on Information and Knowledge Management, pp. 1441–1450, 2019.

[10] Guo, H., Tang, R., Ye, Y., Li, Z., and He, X., "DeepFM: A Factorization-Machine Based Neural Network for CTR Prediction," Proceedings of the 26th International Joint Conference on Artificial Intelligence, pp. 1725–1731, 2017.

[11] Li, J., Ren, P., Chen, Z., Ren, Z., Lian, T., and Ma, J., "Neural Attentive Session-Based Recommendation," Proceedings of the 26th ACM International Conference on Information and Knowledge Management, pp. 1419–1428, 2017.

[12] Rossi, E., Chamberlain, B., Frasca, F., Eynard, D., Monti, F., and Bronstein, M., "Temporal Graph Networks for Deep Learning on Dynamic Graphs," arXiv:2006.10637, pp. 1–24, 2020.

[13] Zhou, G., Zhu, X., Song, C., Fan, Y., Zhu, H., Ma, X., Yan, Y., Jin, J., Li, H., and Gai, K., "Deep Interest Network for Click-Through Rate Prediction," Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1059–1068, 2018.

[14] Chen, Q., Zhao, H., Li, W., Huang, P., and Ou, W., "Behavior Sequence Transformer for E-Commerce Recommendation in Alibaba." Proceedings of the 1st International Workshop on Deep Learning Practice for High-Dimensional Sparse Data at ACM KDD, pp. 1–4, 2019.

[15] Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y., "Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation," Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, pp. 1724–1734, 2014.

[16] European Parliament, "Regulation (EU) 2016/679: General Data Protection Regulation," Official Journal of the European Union, Vol. L 119, pp. 1–88, 2016.

[17] Ying, R., He, R., Chen, K., Eksombatchai, P., Hamilton, W. L., and Leskovec, J., "Graph Convolutional Neural Networks for Web-Scale Recommender Systems," Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 974–983, 2018.

[18] Tang, J. and Wang, K., "Personalized Top-N Sequential Recommendation via Convolutional Sequence Embedding", Proceedings of the 11th ACM International Conference on Web Search and Data Mining, pp. 565–573, 2018.

[19] Wang, R., Fu, B., Fu, G., and Wang, M., "Deep and Cross Network for Ad Click Predictions", Proceedings of the ADKDD'17 Workshop at ACM KDD, pp. 1–7, 2017.

[20] Konecny, J., McMahan, H. B., Ramage, D., and Richtarik, P., "Federated Optimization: Distributed Machine Learning for Mobile Devices," arXiv:1610.02527, pp. 1–16, 2016.
[21] Pearl, J., Causality: Models, Reasoning, and Inference, Cambridge University Press, Cambridge, UK, 2009.
[22] Goodfellow, I., Bengio, Y., and Courville, A., Deep Learning, MIT Press, Cambridge, MA, 2016.