Sequential Customer Behavior Modeling with Graph Neural Networks for Predicting IVR Service Escalation

AIMLCZG628T: Dissertation (Mid-Semester Viva)

Sowmya A | 2023AC05209

The High-Stakes Problem: Digital Deflection Failure

Organizations encourage self-service via web portals and apps. When these digital journeys fail due to friction, users "leak" to costly assisted channels like Interactive Voice Response (IVR). This is **Digital Deflection Failure**.

Operational Costs

Voice support is significantly more expensive than successful self-service, impacting operational efficiency at scale.

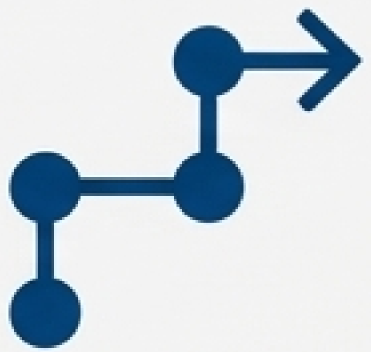
Customer Experience

A frustrating digital journey leads to lower satisfaction, negative brand perception, and potential customer churn.

The Core Challenge: To predict IVR escalation *before* it occurs by analyzing subtle micro-behaviors hidden in clickstream data.

The Modern Research Landscape: Four Pillars of Spatio-Temporal AI

A summary of State-of-the-Art from 2023-2025, framed as four key themes that provide the foundation for this work.



Intent Paths

Models user actions as a sequential prefix leading to a target outcome.

(STIRec, 2025)

Relational Context

Leverages relationships between different entities (e.g., users) to improve forecasting.

(TFT-GNN, 2024)

Inductive Scaling

Generates predictions for new users or graph components without requiring full model retraining.

(Li, P., et al., 2023)

Business Trust

Moves beyond raw accuracy to ensure predictions are risk-aware and probabilistically reliable for real-world use.

(RCIP, 2024)

Identifying the Critical Research Gaps

While powerful, the State-of-the-Art approaches often operate in silos, revealing two key opportunities for contribution.

1. The Spatio-Temporal Disconnect

Most models excel at either structure or sequence, but not their simultaneous influence.

- **Spatial (Graph):** Understands the website's topology—the **"where."**
- **Temporal (Sequence):** Understands the user's chronological path—the **"when."**
- **The Opportunity:** A unified model is needed to capture how a user's ***intent over time*** is influenced by the ***website's structure***.

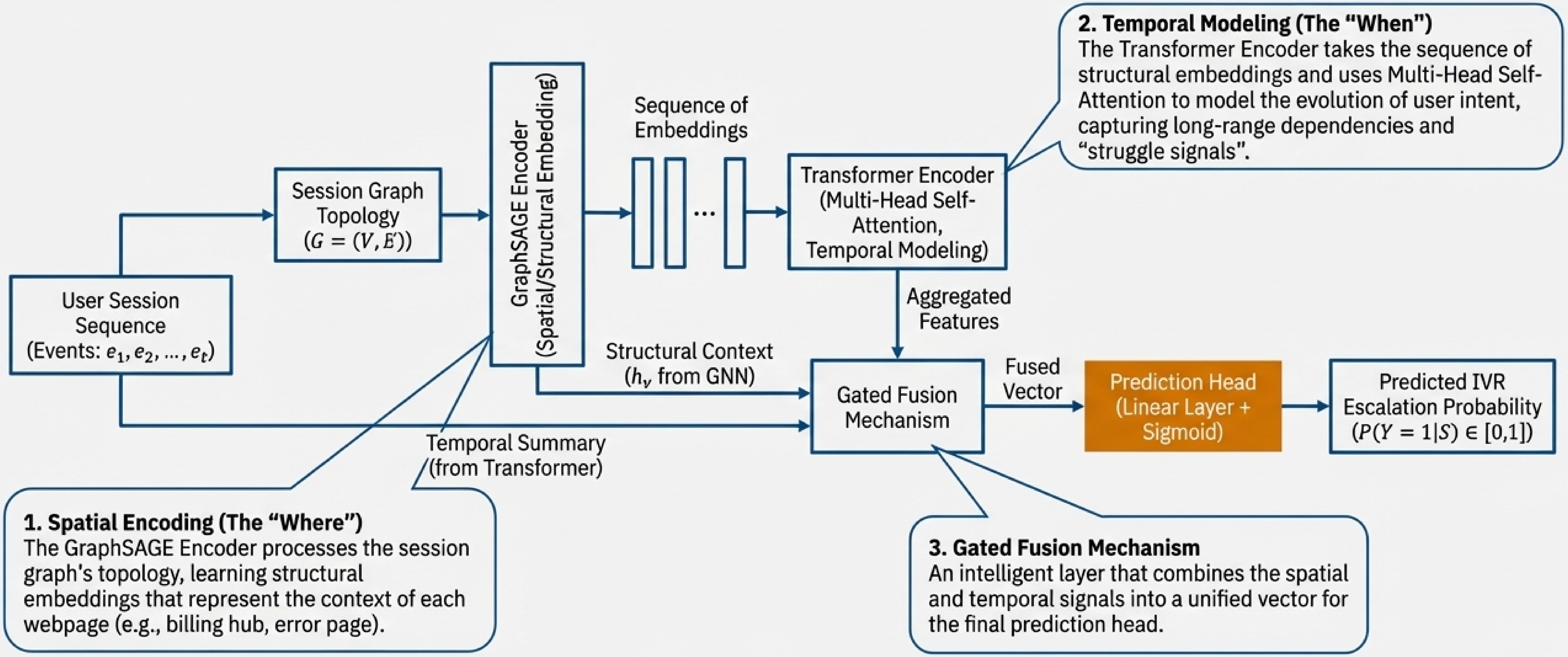
2. The Calibration Imperative

For enterprise automation, the *confidence* of a prediction is as important as its accuracy.

- **SOTA Focus:** Optimization for ranking or classification metrics (AUC, F1-Score, NDCG, HR).
- **The Opportunity:** A lack of focus on **probabilistic calibration**. An uncalibrated model is unusable for automated interventions, yet metrics like the Brier Brier Score are rarely prioritized.

Our Solution: A Hybrid Architecture to Bridge the Gap

We propose a dual-stream GNN-Transformer architecture that jointly models the "where" and the "when" of user behavior.



Methodology: From Raw Clicks to a Hybrid Brain

1. Data Synthesis & Graph Construction

- Simulated 10,000 customer journeys to mirror telecommunications portal topology.
- Engineered a realistic class imbalance with 23.44% callers.
- Transformed sessions into a directed graph, capturing non-linear behaviors like “navigation loops” characteristic of user struggle.

2. Hybrid Model Implementation

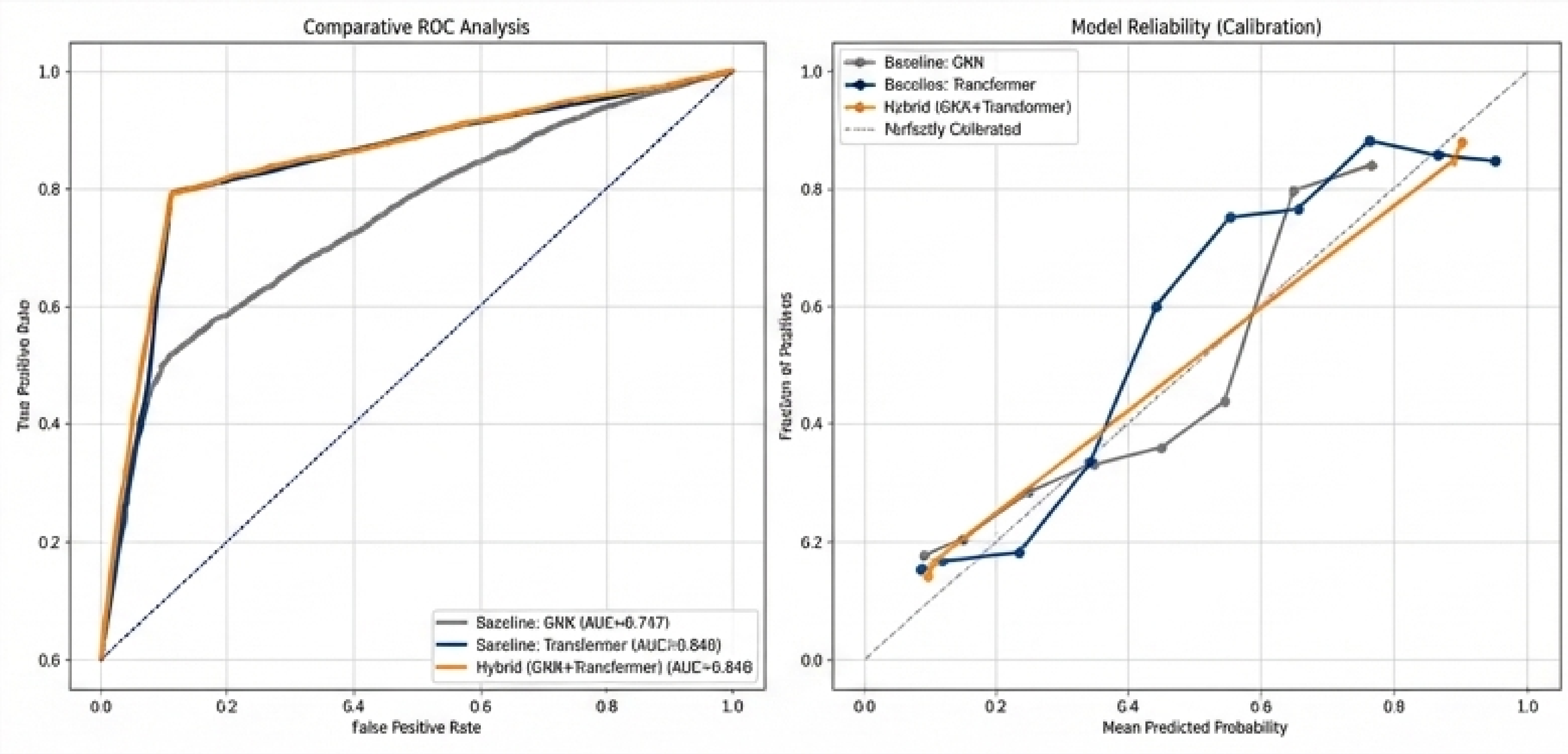
- Built using PyTorch and PyTorch Geometric.
- Architecture combines a 2-layer GraphSAGE network with a 2-layer Transformer Encoder (4 attention heads).
- A Gated Fusion mechanism merges the two information streams.

3. Training & Evaluation

- Trained for 40 epochs using Full Batch Gradient Descent.
- Evaluated against GNN-only and Transformer-only baselines.
- Metrics focused on both Discrimination (ROC-AUC) and Reliability (Brier Score).

Performance Evaluation: Superior Discrimination and Reliability

Model	ROC-AUC (Higher is Better)	Brier Score (Lower is Better)
Baseline: GNN	0.747	0.201
Baseline: Transformer	0.840	0.178
Hybrid (GNN +Transformer)	**0.848**	**0.133**



Key Insights:

- **Discrimination:** The Hybrid model (Orange line) demonstrates the best ability to distinguish between callers and non-callers, achieving the highest AUC (0.848).

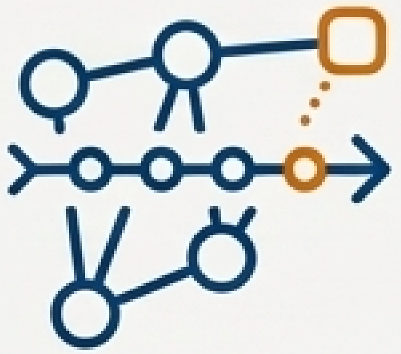
- **Calibration:** Crucially, the Hybrid model's reliability curve is the closest to the 'Perfectly Calibrated' diagonal. This is validated by its significantly lower Brier Score, meaning its probability scores are trustworthy for business decisions.

Positioning the Work: Comparison with State-of-the-Art

Our work integrates and extends recent advances to create a solution tailored for enterprise customer experience.

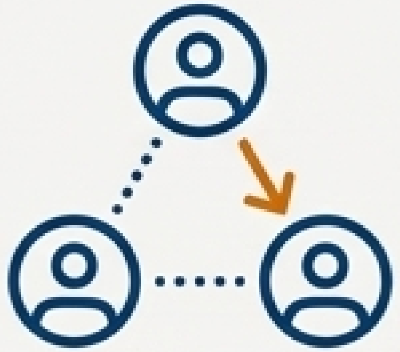
Feature	State-of-the-Art (SOTA) 2023-2025	This Dissertation
Model Scope	Often focuses on either spatial or temporal dynamics.	**Dual-stream integration** of GNNs and Transformers.
User Dynamics	Dynamic, but often requires frequent re-indexing.	**Fully Inductive**: Real-time embedding for new users.
Prediction Goal	Raw Accuracy / Ranking (NDCG/HR).	**Calibrated Intent Probability** (Brier-optimized).
Business Utility	General behavioral forecasting.	**CX-Specific Intent Mapping:** Optimized for IVR triggers.

Key Pillars of Novelty



Hybrid Inductive Spatio-Temporal Architecture

We propose a novel, **dual-stream** architecture that is **fully inductive**. This allows the model to handle new customers and evolving website structures without retraining, extending the work of Li et al. (2023) by integrating Transformer-based temporal reasoning.



Cross-Relational Behavioral Synthesis

We adapt the **relational context** idea from TFT-GNN (2024) to the customer experience domain. The graph structure allows the model to learn from behaviorally similar peers to improve predictions for an individual.

Enterprise-Grade Calibration

Our primary contribution is the **shift from accuracy to reliability**. By explicitly optimizing for the Brier Score, we align with the rigorous, risk-aware principles of RCIP (2024) to produce probabilities that are mathematically sound and business-ready.

Conclusion & Mid-Semester Status

The mid-semester milestone is successfully complete, with a validated, high-performing baseline model and a robust end-to-end pipeline.

✓ Designed and validated a novel Hybrid GNN-Transformer architecture to address the spatio-temporal disconnect.

✓ Established an end-to-end pipeline from data synthesis (10,000 sessions, 23.44% class ratio) to calibrated prediction.

✓ Established an end-to-end pipeline from data synthesis (10,000 sessions, 23.44% class ratio) to calibratedly a **Brier Score** of **0.133**.

✓ Demonstrated superior performance over single-modality baselines, achieving a **ROC-AUC** of **0.842** and a **Brier Score** of **0.133**.

Concluding Statement: The core hypothesis is validated. A calibrated, spatio-temporal approach that jointly models structure (‘where’) and sequence (‘when’) is significantly more effective for predicting IVR escalation.

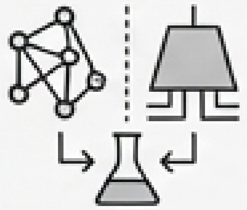
Future Work: Path to Completion

The final semester will focus on optimization, rigorous analysis, and deployment readiness.

1 Advanced Hyperparameter Tuning

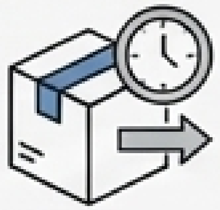
Employ Bayesian Optimization (Optuna) to systematically fine-tune embedding dimensions, attention heads, and learning rates.

2 Ablation Studies



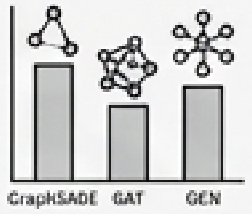
Rigorously quantify the individual performance contributions of the GNN and Transformer components within the hybrid architecture.

3 Deployment & Latency Testing



Containerize the inference pipeline for real-time latency testing to ensure it meets operational requirements.

4 Comparative Architecture Analysis



Benchmark the current GraphSAGE implementation against alternative GNN layers (e.g., GAT, GCN) to validate the optimal architectural choice.