A Survey on Behavioral Data Representation Learning

Anonymous authors
Paper under double-blind review

Abstract

Behavioral data, reflecting dynamic and complex interactions among entities, are pivotal for advancing multidisciplinary research and practical applications. Effective modeling and representation of behavioral data facilitate enhanced understanding, predictive analytics, and informed decision-making across diverse domains. This paper presents a comprehensive taxonomy of behavioral data representation learning methods, categorized by data modalities: tabular data, event sequences, dynamic graphs, and natural language. Within each category, we further dissect methods based on distinct modeling strategies and capabilities, and provide detailed reviews of their developments. Additionally, we extensively discuss significant downstream applications, datasets, and benchmarks, highlighting their roles in guiding methodological development and evaluating performance. To support further exploration in behavioral data representation learning, we release a continuously maintained repository at Anonymous GitHub that curates the methods and papers covered in this survey.

1 Introduction

Behavioral data capture intricate and dynamic interactions between individuals and their environments, serving as foundational evidence across multiple disciplines such as psychology, neuroscience, social sciences, economics, and artificial intelligence. They are inherently rich, relational, and dynamic, often described in various data modalities such as tabular data, event sequences, dynamic graphs, and textual data.

Accurate modeling of behavioral data is crucial as it enables a deeper understanding of underlying behavioral mechanisms, facilitates predictive analytics, and informs practical interventions in diverse areas such as marketing, policy-making, financial risk control, and social network analysis. Given its complexity, behavioral data often exhibit high dimensionality, temporal dependencies, and relational structures, necessitating sophisticated representation learning methods.

Representation learning of behavioral data involves transforming raw behavioral inputs into structured, compact representations, i.e., embeddings, that capture essential behavioral patterns suitable for computational models. Unlike traditional feature extraction methods that rely heavily on manual intervention, representation learning leverages advanced machine learning and deep learning techniques to automatically distill informative features directly from complex datasets.

To provide a systematic understanding of behavioral data representation learning, we propose a taxonomy structured around the different forms in which behavioral data can be represented: **(1) Tabular Data**: This form represents behaviors as structured tables, characterized by static or aggregated features such as demographic attributes, user preferences, transaction frequencies, or sensor readings. Tabular data often enables efficient processing and straightforward interpretability, making it ideal for applications like intelligent monitoring, risk assessment, and market analytics. Tabular data modeling has undergone significant evolution, progressing from early tree-based methods to a competitive landscape dominated by both deep learning approaches and tree-based models. More recently, the emergence of large language models has introduced a new paradigm, further advancing the field. **(2) Event Sequences**: Behaviors are frequently recorded as sequential events, each associated with a timestamp, providing insights into temporal dependencies, user trajectories, and action progressions. Common examples include clickstreams from websites, transaction logs, medical events, and sensor data sequences. Representation learning methods for event sequences usually utilize sequential modeling approaches, such as Recurrent Neural Networks (RNNs) (Cho et al., 2014)

and Transformers (Vaswani et al., 2017), to capture inherent sequential dynamics and temporal patterns. **(3) Dynamic Graphs**: Many behavioral phenomena naturally form dynamic relational structures, representing interactions among entities evolving over time, such as social networks, financial transactions, and collaborative systems. Dynamic graphs capture complex relational dynamics, enabling detailed analysis of interactions and structural changes. Representation learning methods for dynamic graphs employ Graph Neural Networks (GNNs) (Scarselli et al., 2009), integrated with recurrent models (Cho et al., 2014; Graves, 2012), attention mechanisms (Vaswani et al., 2017), and memory architectures (Rossi et al., 2020), to accurately encode evolving graph structures and temporal dependencies. **(4) Textual Data**: Behavioral data often manifest through textual interactions and documents, including online reviews, chat logs, social media content, and textual financial disclosures. Textual data inherently contains rich semantic information reflecting sentiments, intents, and contextual interactions. Representation learning in this domain leverages advanced large-scale pre-trained Language Models (LMs), such as BERT, GPT (OpenAI, 2023), and specialized variants like FinBERT (Huang et al., 2023a), to effectively extract contextual and semantic insights, facilitating sentiment analysis, intention recognition, and behavioral prediction tasks.

Within each form, we further subdivide representation learning methods based on specific architectural choices, temporal modeling strategies, or the integration of external domain knowledge and multi-modal fusion capabilities.

Additionally, we extensively explore significant downstream applications, including personalized recommendation systems, financial risk assessment, anomaly and fraud detection, social network analytics, adaptive educational technologies, healthcare monitoring, and behavioral interventions. Each application domain underscores the practical implications, efficacy, and adaptability of behavioral data representation learning methodologies. To facilitate research and practical applications, we further provide an overview of commonly used datasets and benchmarks, outlining their characteristics, scales, domains of applicability, and evaluation standards. This comprehensive survey approach enables clearer alignment between methodological choices, application-specific goals, and available data resources. We summarize our contributions as follows:

- **Novel Taxonomy**: We introduce a new and comprehensive taxonomy for behavioral data representation learning. First, from the perspective of data modalities, we categorize representation methods into distinct classes based on the form of the input behavioral data, including tabular data, event sequences, dynamic graphs, and textual data, and we further provides within-modality classifications. This modality-based taxonomy helps researchers efficiently select appropriate methods according to data characteristics and application scenarios. Second, we classify methods according to various downstream tasks and applications, enabling clear alignment between methodological choices and specific analytical goals.
- **Comprehensive Methodological Reviews**: We provide a thorough and detailed review of the state-of-the-art methods within each category, highlighting their core principles, underlying assumptions, key innovations, and limitations. Our comprehensive analysis includes critical discussions of model architectures, training paradigms, and evaluation metrics, facilitating a deep understanding of existing approaches and guiding future methodological innovations.
- **Integration of Emerging Technologies**: We critically analyze the impact and potential integration of emerging technologies on behavioral data representation learning. Furthermore, we outline promising future directions and open challenges, emphasizing how interdisciplinary advancements can further enhance behavioral representation learning and expand its applicability and effectiveness across various practical and scientific domains.

2 Survey Methodology

We collected the literature covered in this survey through an iterative search and screening process. Our primary sources included peer-reviewed conference and journal papers, complemented by influential preprints when a formal publication was not yet available or when the preprint had already become widely used in the community.

In practice, we searched major scholarly indexes and repositories including Google Scholar, DBLP, Semantic Scholar, and arXiv, and prioritized venues that frequently publish work on behavioral data modeling and representation learning.

Our search process was guided by the taxonomy developed in this survey and combined broad queries with modality- and task-specific keywords. Representative search terms included "behavioral data representation learning", "user behavior modeling", "tabular data representation", "event sequence modeling", "temporal graph", "dynamic graph learning", "temporal interaction graph", "textual behavior modeling", "LLM for user modeling" and etc., and related application-oriented terms such as recommendation, fraud detection, healthcare, and social media analysis and etc.

We then expanded the candidate set by following the references of key papers and adjacent surveys. We included papers that proposed, substantially adapted, or systematically evaluated representation learning methods for behavioral data, as well as benchmark and dataset papers that are central to empirical evaluation.

Because the field has accelerated rapidly in the large-model era, our coverage focuses primarily on work published from 2016 to early 2026, while retaining earlier seminal papers when they are necessary to explain the historical foundations of later methods.

3 Definitions

3.1 Human Behavior

Human behavior refers to the observable actions or reactions of a person in response to its environment or internal state. Behavior is not a static or isolated event; it emerges from a complex interplay of factors and unfolds over time (Gomez-Marin et al., 2014). This highlights that behavior is inherently relational, arising through interactions between an individual (i.e., the behavioral subject) and its surroundings (i.e., target entities, including other people), and is dynamic, progressing as a sequence of events rather than a single event. For example, in social situations multiple individuals may dynamically interact and display species-specific behaviors, with each person's actions influencing the others' in real time. Behavior is also high-dimensional in nature. Even a single behavioral episode can be described by numerous variables (positions of limbs, neural signals, environmental context, etc.) (von Ziegler et al., 2021). This high dimensionality means behavioral data can be extremely rich, but it also poses challenges for analysis and interpretation.

3.2 Behavioral Data

Behavioral data are the recorded or measured manifestations of behavior, referring to data collected to represent what an individual does. Such data are grounded in observable actions or outputs of the subject. Behavioral data can take many forms depending on how the behavior is observed and logged. The key aspect is that the behavior has been externalized into a digital or structured form that can be stored and analyzed. In early behavioral science, data often came from manual observation notes or simple counts of events, whereas today they are frequently complex digital datasets (potentially combining multiple channels) (Wilson & Collins, 2019). Regardless of the format, behavioral data serve as the empirical evidence of behavior, which researchers can analyze to draw inferences about underlying processes or to predict future actions.

3.3 Representation Learning

Representation learning aims to automatically learn effective and generalizable feature representations from raw data through machine learning or deep learning methods. It transforms high-dimensional, complex, and difficult-to-process raw data into more compact, low-dimensional, and semantically rich vector representations. These learned representations can not only significantly improve the performance of downstream tasks, such as classification and relation prediction, but also reduce the reliance on manual feature engineering during analysis. Representation learning has achieved remarkable success across a wide range of domains, including Computer Vision (CV), Speech Processing, and Natural Language Processing (NLP).

In recent years, with the rapid development and maturation of deep learning research, particularly the emergence of Convolutional Neural Networks (CNNs) (Goodfellow et al., 2016), RNNs (Cho et al., 2014), GNNs (Scarselli et al., 2009), and self-attention mechanisms (Vaswani et al., 2017), the expressive power of representation learning for modeling diverse data types has been greatly enhanced. This progress has further provided powerful tools for the analysis of behavioral data.

3.4 Behavioral Data Representation Learning

Behavioral data representation learning is the process of encoding raw behavioral data into structured, compact representations (often low-dimensional vectors or embeddings) that are suitable for computational modeling and analysis. It enables models to automatically learn the salient features or abstract factors of the data directly from the data itself, rather than relying on manual feature selection or domain-specific heuristics (Bengio, 2012; Bengio et al., 2013).

In conventional analyses, human experts must design specific behavioral metrics or features a priori (for example, deciding to measure the frequency of certain actions or the duration of particular events). As behavioral datasets grow more complex and high-dimensional, this manual process becomes a bottleneck: many hand-crafted features do not capture the rich, hierarchical structure of behavior, and creating new features for each problem is labor-intensive and not scalable. Along with the development of machine learning and deep learning, researches have started to use these emerging techniques to discover a latent encoding of behavioral patterns without explicit human supervision, which enables models to automatically capture complex, non-linear temporal dependencies, reduce reliance on handcrafted features, and generalize across diverse behavioral scenarios (Goodfellow et al., 2016; Bengio et al., 2021).

3.5 Forms of Behavioral Data and Their Representations

Behavioral data can be represented or constructed in various forms or modalities depending on how it is measured and what aspects of behavior are of interest. Each modality may require different representation learning techniques. We outline common data modalities for behavior and how they capture behavioral information.

Static Feature Vectors: One straightforward way to represent behavior is as a vector of features summarizing an individual's behavioral tendencies. In this modality, the behavior is reduced to a defined set of attributes or metrics. Such a representation treats behavioral data as a point in a multi-dimensional feature space. Historically, many behavioral studies relied on this kind of summarization – for example, recording the number of online purchases, the average session duration on a platform, or the frequency of logins, which yields numeric descriptors that are relatively easy to analyze statistically. The advantage of static feature vectors is their simplicity and interpretability; however, they may omit the temporal dynamics or context of the behavior.

Event Sequences: Human behavior often unfolds as a sequence of discrete, temporally ordered events, which may occur sporadically or in bursts. This could be a sequence of timestamped events or state changes generated over time by one or more sources. Representing behavior as a sequence preserves the temporal dynamics like the order and timing of events, which are crucial for understanding patterns such as routines, progressions of actions, or cause-and-effect in interactions. Event sequence representations enable the use of sequence-modeling techniques, such as Markov Chain Models (Eddy, 1996), RNNs (Cho et al., 2014; Graves, 2012) or Transformer and its variants (Sun et al., 2019b), to learn temporal patterns in the behavior.

Dynamic Graphs: Some behavioral data are best described not just as a linear sequence but as a set of entities and their interactions evolving over time. Graph-based representations are useful when behavior involves relational structure, such as social interactions or complex event dependencies. In a dynamic graph, nodes usually represent entities, and edges represent interactions or temporal relations between these nodes (Kumar et al., 2019). By applying dynamic graphs, a complex action sequence becomes a structured graph rather than a simple timeline, and analyzing the behavior can involve graph matching or graph neural network algorithms to identify patterns. Dynamic graph representations can naturally capture concurrent events (when multiple things happen at once) and long-range dependencies.

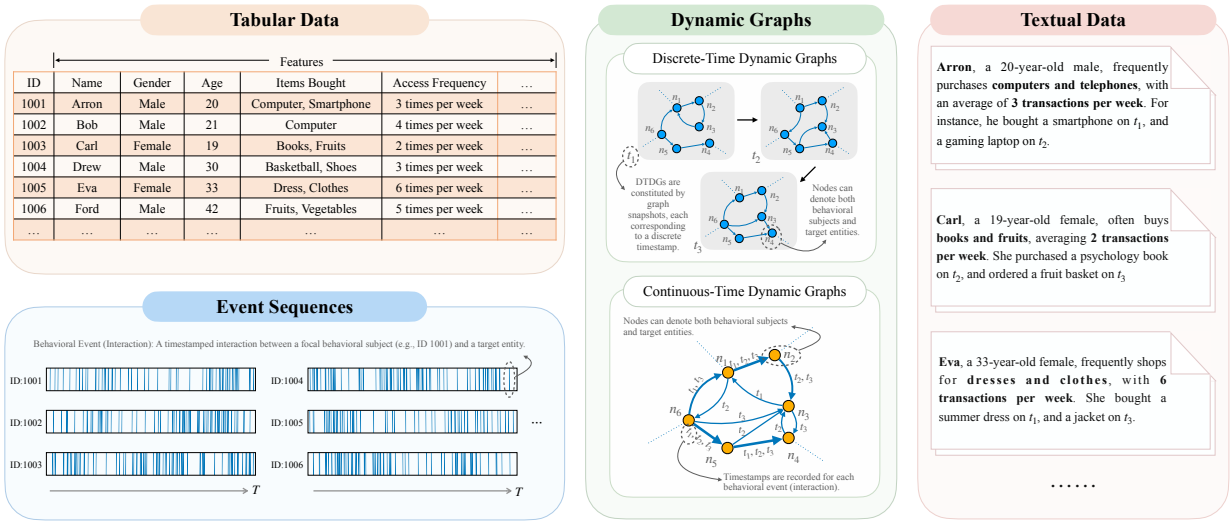


Figure 1: Taxonomy of behavioral data: behavioral data can take various forms, including tabular data, event sequences, dynamic graphs, and textual data.

Multi-modal Representations: Real-world behavior often spans multiple data modalities at once. Many forms of behavioral data are inherently linguistic, such as dialogue transcripts, social media interactions, textual evaluations, or narrative logs of user interactions. These textual behavioral data inherently embed rich semantic information including emotional states, intentions, motivations, and subjective experiences (Feuerriegel et al., 2025). Moreover, recent advancements in large-scale pre-trained language models, such as GPT models (Radford et al., 2019; Brown et al., 2020; OpenAI, 2023), have profoundly changed the landscape (Xiong et al., 2025). These models enable not only direct analysis of inherently textual behavioral data but also allow the transformation of traditionally non-linguistic behavioral datasets into natural language (Feuerriegel et al., 2025).

4 Taxonomy of Behavioral Data Representation Learning

Based on the different modalities of behavioral data introduced in Sec. 3.5, this section presents representation learning methods tailored to each modality and provides a finer-grained taxonomy within each category.

4.1 Tabular Data

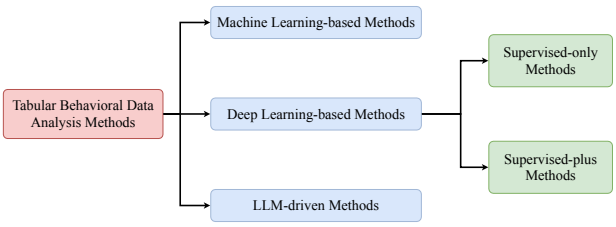


Figure 2: The dissection of the tabular data representation learning methods.

Behavioral data can be instantiated in different forms, one of which is tabular data organized in rows and columns, where each row corresponds to an individual instance and each column represents a specific feature or attribute (Shwartz-Ziv & Armon, 2022; Somvanshi et al., 2024; Ucar et al., 2021; Zhu et al., 2021). In representation learning, such data are typically modeled as static feature vectors, with each instance encoded as a fixed-dimensional vector that aggregates all attribute values. As one of the most prevalent and widely

Table 1: Tabular data representation learning methods.

Method	Training Paradigm	Model Architecture	Year
Tree-based Methods			
XGBoost (Chen & Guestrin, 2016)	Supervised learning	Tree	2016
LightGBM (Ke et al., 2017)	Supervised learning	Tree	2017
CatBoost (Prokhorenkova et al., 2018)	Supervised learning	Tree	2018
Deep Learning-based Methods			
Wide&Deep (Cheng et al., 2016)	Supervised learning	GLM+MLP	2016
DeepFM (Guo et al., 2017)	Supervised learning	FM+MLP	2017
xDeepFM (Lian et al., 2018)	Supervised learning	FM+MLP+CIN	2018
TabNN (Ke et al., 2018)	Supervised learning	GBDT+MLP	2018
RLN (Shavitt & Segal, 2018)	Supervised learning	MLP	2018
NODE (Popov et al., 2020)	Supervised learning	NODE	2020
SuperTML (Sun et al., 2019a)	Supervised learning	CNN	2019
TabNet (Arik & Pfister, 2021)	Supervised + Self-supervised learning	TabNet	2021
DeepGBM (Ke et al., 2019)	Supervised + Online learning	DeepGBM	2019
NON (Luo et al., 2020)	Supervised learning	NON	2020
DNF-Net (Abutbul et al., 2020)	Supervised learning	DNF-Net	2020
VIME (Yoon et al., 2020)	Self-supervised + Semi-supervised learning	VIME	2020
TabTransformer (Huang et al., 2020a)	Supervised + Semi-supervised learning	Transformer	2020
ARM-Net (Cai et al., 2021)	Supervised learning	ARM-Net	2021
NPT (Kossen et al., 2021)	Supervised learning	Transformer	2021
Regularized DNNs (Kadra et al., 2021)	Supervised learning	MLP	2021
Boost-GNN (Ivanov & Prokhorenkova, 2021)	Supervised learning	GBDT+GNN	2021
DNN2LR (Liu et al., 2020)	Supervised learning	DNN2LR	2020
IGTD (Zhu et al., 2021)	Supervised learning	IGTD+CNN	2021
FT-Transformer (Gorishniy et al., 2021)	Supervised learning	Transformer	2021
SAINT (Somepalli et al., 2021)	Supervised + Self-supervised learning	SAINT	2021
SCARF (Bahri et al., 2022)	Self-supervised learning	MLP	2022
GANDALF (Joseph & Raj, 2022)	Supervised learning	GFLU	2022
TabDDPM (Kotelnikov et al., 2023)	Unsupervised learning	Diffusion Model	2023
PTab (Liu et al., 2022b)	Supervised + Self-supervised learning	BERT	2022
TabCBM (Zarlenga et al., 2023)	Supervised learning	MLP	2023
Trompt (Chen et al., 2023a)	Supervised + Prompt learning	MLP	2023
HYTREL (Chen et al., 2023b)	Supervised + Self-supervised learning	HYTREL	2023
ReConTab (Chen et al., 2023c)	Self-supervised	Transformer	2023
XTab (Zhu et al., 2023a)	Supervised + Self-supervised learning	Transformer	2023
IATTN (Thielmann et al., 2024)	Supervised learning	Transformer	2024
MambaTab (Ahamed & Cheng, 2024)	Supervised learning	Mamba	2024
GCondNet (Margeloiu et al., 2024)	Supervised learning	MLP	2024
BiSHop (Xu et al., 2024)	Supervised learning	BiSHop	2024
LF-transformer (Na et al., 2024)	Supervised learning	Transformer	2024
TabR (Gorishniy et al., 2024)	Supervised learning	TabR	2024
TP-BERTa (Yan et al., 2024)	Supervised + Self-supervised learning	BERT	2024
CARTE (Kim et al., 2024)	Supervised + Self-supervised learning	Transformer	2024
SwitchTab (Wu et al., 2024a)	Self-supervised learning	Transformer	2024
DP-2Stage (Afonja et al., 2025)	Supervised learning	Transformer	2025
TARTE (Kim et al., 2025)	Supervised + Self-supervised learning	Transformer	2025
TDColER (Kang et al., 2025)	Supervised + Unsupervised learning	Distillation	2025
LLM-driven Methods			
TAPAS (Herzig et al., 2020)	Self-supervised learning	Bert	2020
TAPEX (Liu et al., 2022c)	Supervised + Self-supervised learning	Transformer	2022
TabLLM (Hegselmann et al., 2023)	Supervised + Self-supervised learning	Transformer	2023
cTBLS (Sundar & Heck, 2023)	Supervised learning	Transformer	2023
TST-LLM (Han et al., 2025)	Self-supervised learning	Transformer	2025
Tabby (Cromp et al., 2026)	Supervised learning	Transformer	2026

used data types in real-world information systems, tabular data are central to a broad range of domains, including medicine (Johnson et al., 2016; Ulmer et al., 2020), finance (Kumar et al., 2016; Clements et al., 2020), e-commerce (Cheng et al., 2016; Guo et al., 2017). Compared to unstructured data such as images or text, tabular data offer clear and interpretable structure, facilitating efficient storage, processing, and analysis. For example, in digital platforms, tabular data are widely used to record user behaviors such as login frequency, browsing histories, purchase transactions, or content interactions; in online education, they capture attendance records, assignment submissions, and test outcomes. With the growing emphasis on data-driven decision-making across industries, tabular behavioral data play an increasingly critical role in tasks such as user profiling, anomaly detection, risk assessment, and personalized recommendation. However, despite its seemingly regular structure, behavioral tabular data often exhibit high heterogeneity, sparsity, and domain-specific characteristics, posing significant challenges for effective modeling and analysis. Therefore, a systematic investigation into the representation, modeling strategies, and practical applications of tabular behavioral data are of both theoretical importance and practical relevance for advancing intelligent decision systems and enabling digital transformation across sectors.

Modeling methods for tabular data have seen a progression —from traditional machine learning to deep learning, and more recently, to large foundation models. Traditional tree-based models provided a solid balance between interpretability and robustness, establishing them as a strong baseline for tabular data learning. With the rise of deep learning, neural networks have been explored for tabular data (Arik & Pfister, 2021; Popov et al., 2020; Abutbul et al., 2020), showing promise in areas like feature learning and multi-modal integration despite inherent challenges. More recently, Large Language Models (LLMs) have introduced a new paradigm, enabling tasks such as reasoning and generation over tabular inputs. Overall, tabular data modeling is evolving toward more generalizable and automated solutions.

In this section, we systematically review representative modeling approaches for tabular behavioral data. Specifically, we categorize existing methods into three major families based on their underlying modeling paradigms: **(1) Machine Learning-based Methods**, **(2) Deep Learning-based Methods**, and **(3) LLM-driven Methods**. For each category, we highlight the core modeling techniques, representative architectures, and recent advancements. All methods are concluded in the Tab. 1.

4.1.1 Machine Learning-based Methods

While classical non-tree methods such as Support Vector Machines (SVMs) and linear models remain relevant in specific niches—particularly for small-sample regimes or high-dimensional sparse data—their sensitivity to data preprocessing and scalability constraints have restricted their broader adoption. Consequently, over the past decade, tree-based ensemble algorithms such as eXtreme Gradient Boosting (XGBoost) have established themselves as the go-to solutions for tabular data modeling in real-world applications (Chen & Guestrin, 2016; Prokhorenkova et al., 2018; Friedman, 2001). Beyond their empirical dominance, researchers have actively investigated the theoretical underpinnings of their superiority on structured data (Shwartz-Ziv & Armon, 2022; Grinsztajn et al., 2022). These tree-based ensembles—most notably Gradient Boosting Decision Trees (GBDT) and their optimized implementations like XGBoost (Chen & Guestrin, 2016), LightGBM (Ke et al., 2017), and CatBoost (Prokhorenkova et al., 2018)—are favored for their exceptional accuracy, training efficiency, and robustness against missing values and unscaled categorical features. Even amidst the rapid proliferation of deep learning, GBDT-based models continue to be regarded as the state-of-the-art for general tabular tasks (Borisov et al., 2023; Gorishniy et al., 2021).

Recent studies have sought to explain the superior performance of tree-based models on tabular data from both theoretical and empirical perspectives (Grinsztajn et al., 2022). Works such as Unmasking Trees (McCarter, 2025) have shifted attention toward understanding and improving the interpretability of tree-based models by analyzing their internal decision mechanisms in tabular settings. On the one hand, tabular datasets often exhibit characteristics such as non-stationarity, feature heterogeneity, and sparsity, which pose significant challenges for traditional neural networks (Somvanshi et al., 2024). Tree-based models, compared to neural networks, naturally adapt to these properties, performing implicit feature selection and capturing complex nonlinear relationships. These models also have demonstrated strong robustness in handling categorical variables, outliers, and missing data, and are less prone to overfitting when trained on

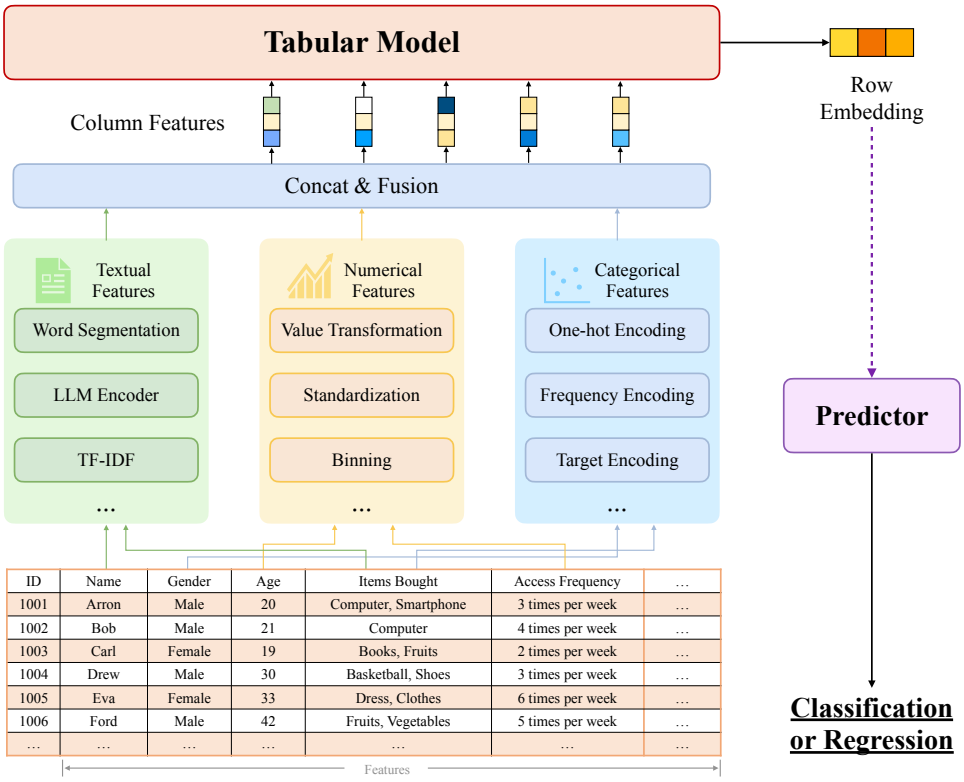


Figure 3: The pipeline of supervised tabular data representation learning methods.

small datasets (Larionov et al., 2025). These advantages have contributed to the widespread adoption and continuous development of tree-based models in both industry and academia.

Summary and takeaways. The machine-learning stage of tabular representation learning is not merely historical background; it explains why tree-based models remain the reference point for much of the later literature. Classical non-tree methods such as linear models and SVMs still retain value in specific niches, especially when data are small, sparse, or strongly preprocessed, but tree ensembles became dominant because they better accommodate heterogeneity, missing values, nonlinear feature interactions, and small-data regimes without heavy feature scaling or architecture tuning.

4.1.2 Deep Learning-based Methods

While traditional machine learning methods continue to demonstrate strong performance on structured data, deep learning models are increasingly gaining traction due to their powerful representation learning capabilities. Motivated by their remarkable success in domains such as CV and NLP, deep learning approaches have been actively explored for tabular data modeling and analysis. In this section, we provide a comprehensive overview of recent advances in applying deep learning to tabular data, with a particular focus on two key methodological paradigms: **(a) Supervised-only Methods** and **(b) Supervised-plus Methods**.

Supervised-only Methods Supervised learning has long been the dominant paradigm in modeling behavioral tabular data. By leveraging manually annotated labels, supervised approaches are capable of learning explicit mappings between input features and target variables, and have been widely applied in tasks such as user behavior prediction, risk assessment, and recommendation systems. We show the pipeline of supervision methods in Fig. 3.

In the early attempts to apply deep learning to tabular data, researchers primarily focused on integrating traditional machine learning methods with deep learning architectures. For instance, Wide & Deep (Cheng et al.,

2016) jointly trains a generalized linear model and a deep neural network, aiming to combine the strengths of memorization and generalization for improved performance in recommendation systems. DeepFM (Guo et al., 2017) integrates the feature interaction modeling capabilities of Factorization-Machine (FM) with the representation learning power of deep neural networks. xDeepFM (Lian et al., 2018) further extends this architecture by incorporating three components: an FM module for capturing low-order feature interactions, a Multilayer Perceptron (MLP) module for implicitly modeling high-order interactions, and a Compressed Interaction Network (CIN) module for explicitly learning high-order feature interactions.

Moreover, TabNN (Ke et al., 2018) leverages GBDT to identify meaningful feature combinations, treating the features used in each decision tree as a group, which are then encoded using an MLP. Experimental results demonstrate that these hybrid approaches achieve strong performance on a variety of tabular data learning tasks. In addition, SuperTML (Sun et al., 2019a) leverages two-dimensional CNNs by transforming tabular data into image-like representations, demonstrating the potential of CNNs for structured data classification. Furthermore, Boost-GNN (Ivanov & Prokhorenkova, 2021) integrates GBDTs with GNNs in an end-to-end framework, where GBDT models heterogeneous tabular features and GNN captures graph structures, achieving superior performance on graph-based tasks with tabular attributes.

Following the phenomenal success of Transformer (Vaswani et al., 2017) and BERT (Devlin et al., 2019) in NLP, researchers have turned their attention to applying these architectures to structured tabular data. Several representative approaches have emerged in this context. NPT (Kossen et al., 2021) jointly models the entire dataset using self-attention to capture inter-sample dependencies and enable cross-sample reasoning. FT-Transformer (Gorishniy et al., 2021) emphasizes the importance of standardized evaluation and proposes a simplified yet effective Transformer variant, highlighting the role of architecture and training strategy. LF-Transformer (Na et al., 2024) incorporates ideas from matrix factorization by introducing a latent attention factor matrix to enhance feature representation. Collectively, these methods demonstrate the growing potential of Transformer-based models as powerful alternatives to traditional approaches in tabular learning. Interpretable Additive Tabular Transformer Networks (IATTN) (Thielmann et al., 2024) combine self-attention with an additive modeling structure, allowing Transformer-based models to provide feature-wise contribution explanations for tabular predictions.

In addition, many works have proposed architectures specifically tailored to the unique characteristics of tabular data. Regularization Learning Networks (RLN) (Shavitt & Segal, 2018) address the challenge that traditional MLPs require extensive hyperparameter tuning to perform well on tabular tasks, by assigning a separate regularization coefficient to each weight, significantly improving performance. Regularized DNNs (Kadra et al., 2021) explore 13 different regularization techniques and jointly optimize both the selection of regularizers and their auxiliary hyperparameters for each dataset. NODE (Popov et al., 2020) introduces differentiable decision trees by employing techniques such as the α-entmax transformation, enabling gradient-based optimization of split decisions and tree routing. It further stacks multiple layers and supports end-to-end training via backpropagation, allowing the model to capture complex data patterns.

NON (Luo et al., 2020) designs a triple-network architecture that explicitly models intra-field and inter-field relationships, effectively boosting classification performance on tabular data. DNF-Net (Abutbul et al., 2020) proposes the Disjunctive Normal Neural Form (DNNF) module, which integrates feature selection and spatial localization components, and can be trained end-to-end using standard gradient-based methods. ARM-Net (Cai et al., 2021) dynamically models feature interactions in the exponential space and incorporates a multi-head gated attention mechanism, capturing complex dependencies in structured data while enhancing both predictive performance and interpretability.

IGTD (Zhu et al., 2021) transforms tabular inputs into image-like representations and leverages CNNs for downstream processing, making it particularly suitable for domains such as biomedical data, where latent spatial relationships among features may exist. DNN2LR (Liu et al., 2020) addresses the trade-off between performance and interpretability by extracting nonlinear feature interactions from deep neural networks and converting them into explicit cross features for a logistic regression model, resulting in a lightweight yet interpretable model suitable for real-world applications. TabCBM (Zarlenga et al., 2023) further introduces a concept-based neural network for tabular data, where human-understandable concepts are learned as an explicit intermediate layer, enabling transparent and concept-level explanations of model

predictions. GANDALF (Joseph & Raj, 2022) relies on a novel table processing unit called the Gated Feature Learning Unit (GFLU), which features a gating mechanism and built-in feature selection for learning feature representations. The superiority of GANDALF is demonstrated through experiments on multiple established public benchmarks. DeepGBM (Ke et al., 2019) introduces a hybrid deep learning framework that combines neural networks and gradient boosting decision trees to efficiently handle sparse categorical and dense numerical features, enabling accurate and adaptive online prediction for tabular data.

In recent studies, researchers have continued to explore similar directions in deep learning for tabular data, while introducing novel architectural innovations. Trompt (Chen et al., 2023a) is motivated by prompt learning in NLP and has been adapted for tabular data. It optimizes column embeddings through prompt embeddings, thereby generating specific feature importance for each sample. MambaTab (Ahamed & Cheng, 2024) presents a new approach based on Structured State Space Models (SSMs), leveraging Mamba (Gu & Dao, 2024), a recent SSM variant, for end-to-end supervised learning on tabular data. Experimental results across a variety of benchmark datasets demonstrate that MambaTab achieves superior performance compared to state-of-the-art baselines, while requiring significantly fewer parameters. GCondNet (Margeloiu et al., 2024) focuses on small, high-dimensional tabular datasets by introducing a tailored conditioning mechanism that improves representation learning and generalization under limited data regimes.

BiSHop (Xu et al., 2024) addresses two fundamental challenges in deep tabular learning: the lack of rotation invariance in data structures and the inherent feature sparsity of tabular data. Inspired by the connection between associative memory and attention mechanisms, BiSHop adopts a dual-component architecture comprising two interlinked directional learning modules that process data sequentially along both column-wise and row-wise dimensions. TabR (Gorishniy et al., 2024) introduces a Feed-forward Neural Network (FNN) architecture that integrates a customized k-Nearest Neighbors (k-NN) retrieval module. This module identifies and retrieves training samples most similar to a given query instance, and incorporates their features and labels to enhance the prediction quality of the FNN. This hybrid design allows TabR to combine the strengths of neural networks and instance-based learning for improved performance on tabular tasks. DP-2Stage (Afonja et al., 2025) adapts large language models for synthetic tabular data generation through a two-stage pipeline with differential privacy guarantees, enabling privacy-preserving data release for supervised learning.

Supervised-plus Methods Supervised-plus methods integrate self-supervised pre-training with supervised fine-tuning, combining the representational strengths of unlabeled-data learning with the task-specific precision of supervised models. Motivated by the success of self-supervised learning in domains such as vision and language, recent research has applied this paradigm to tabular and behavioral data, where labeled samples are often scarce or noisy. By first learning generalizable representations from large-scale unlabeled data and then refining them through supervised adaptation, hybrid approaches enhance model generalization, robustness, and data efficiency in complex real-world scenarios.

Researchers have explored the potential of unsupervised feature representation from multiple perspectives, including graph structure modeling, feature reconstruction, embedding learning, and cross-modal conversion. One prominent line of work focuses on modeling the structural relationships among features. Given that tabular features often exhibit complex interdependencies, several studies have proposed representing tables as graphs to capture these interactions. For instance, CARTE (Kim et al., 2024) introduces an innovative approach by modeling each row of a table as a star-like graph, where individual cell values are treated as nodes and column headers serve as edge labels. By integrating the strengths of Transformers and Graph Attention Networks (GAT), CARTE leverages self-attention mechanisms to capture contextual dependencies within tabular data. Its self-supervised learning strategy is designed within a contrastive learning framework, enabling the model to extract informative representations from large-scale background data, which facilitates effective transfer learning for downstream tasks.

In contrast, HYTREL (Chen et al., 2023b) models tabular data using a hypergraph structure to capture its inherent relational characteristics. Specifically, each cell is represented as a node, while rows, columns, and the entire table are treated as hyperedges. A structure-aware Transformer is employed to encode this hypergraph representation, and the model is trained through a self-supervised objective to enhance its understanding of table structure.

Another widely adopted strategy for self-supervised learning on tabular data is based on masking and reconstruction, inspired by masked language modeling in NLP. VIME (Yoon et al., 2020) facilitates the learning of informative feature representations from unlabeled tabular data by jointly estimating feature vectors and corresponding mask vectors. Tabbie (Iida et al., 2021) generates corrupted cell candidates by swapping two cells within the same table and trains the model to predict whether each cell has been swapped. SCARF (Bahri et al., 2022), on the other hand, randomly selects a subset of features and replaces them with values sampled from their empirical marginal distributions to create corrupted inputs.

SAINT (Somepalli et al., 2021) and TabNet (Arik & Pfister, 2021), adopt Transformer-based architectures. SAINT employs contrastive learning to encourage the model to capture similarities and differences across various "views" of the same data point. It also introduces a denoising task to help the model recover original data from corrupted inputs, thereby enhancing its ability to leverage unlabeled data. TabNet (Arik & Pfister, 2021) randomly masks a portion of the input features and attempts to reconstruct the masked features from the unmasked ones. This task encourages the model to learn the underlying structure of the data and the relationships among features. Building upon the masking-based approach, TabTransformer (Huang et al., 2020a) introduces a replacement token detection task, in which a subset of feature values is randomly replaced, and a binary classifier is trained to predict whether a given feature value has been replaced.

Recontab (Chen et al., 2023c) and XTab (Zhu et al., 2023a) also utilize reconstruction loss as a self-supervised learning signal. Notably, ReConTab (Chen et al., 2023c) further incorporates label information to compute a semi-supervised classification loss as well as a contrastive loss, enabling the model to better capture discriminative features in the data. SwitchTab (Wu et al., 2024a) decomposes the feature representation of each data sample into mutual information and salient information, and reconstructs the data by swapping the mutual information components between different samples. This process not only encourages the model to learn more structured embeddings, but also enables it to explicitly distinguish between shared characteristics across samples and sample-specific distinctive features.

In addition, PTab (Liu et al., 2022b) transforms tabular data into textual format and trains the model using BERT-style pre-training objectives, achieving competitive performance across various tasks. TabDDPM (Kotelnikov et al., 2023) explores the applicability of diffusion models to general-purpose tabular problems and introduces TabDDPM—a diffusion-based framework capable of handling arbitrary tabular datasets and feature types. TP-BERTa (Yan et al., 2024) proposes a novel relative magnitude tokenization approach that converts continuous numerical feature values into finely discretized high-dimensional tokens. It further incorporates an intra-feature attention mechanism to integrate feature values with their corresponding feature names. This unique self-supervised objective enhances the model's adaptability to tabular data by explicitly modeling feature semantics and structure. More broadly, TARTE (Kim et al., 2025) advocate large-scale pre-training on diverse tabular corpora to acquire transferable knowledge and generalizable representations for downstream tabular tasks. In addition, TDColER (Kang et al., 2025) investigates tabular data distillation, focusing on transforming complex tabular structures into compact and distillable representations suitable for efficient downstream learning.

In summary, self-supervised learning for tabular data has witnessed a surge of diverse and creative approaches. Whether through graph-based modeling, feature masking and reconstruction, embedding optimization, or modality transformation, these methods aim to extract meaningful supervisory signals from unlabeled data. Moving forward, future research may further integrate these strategies to develop more robust and transferable pre-training frameworks for a wide range of structured data analysis tasks.

Summary and takeaways. A consistent pattern in tabular representation learning is that strong empirical performance does not come from architectural complexity alone. Instead, successful approaches usually make one of three contributions: they model feature interactions more explicitly, they exploit unlabeled data through pre-training, or they combine neural models with stronger tabular inductive biases such as tree ensembles or retrieval. This also explains why deep tabular models do not uniformly displace GBDT-style methods: on small or heterogeneous datasets, optimization difficulty and weak inductive bias often outweigh representational flexibility. In practice, deep tabular models appear most compelling when the task benefits from transfer, multi-modal fusion, or scalable pretraining, whereas tree-based methods remain difficult to beat in low-resource, high-heterogeneity settings.

4.1.3 LLM-driven Methods

Recently, LLMs have attracted significant attention for their impressive performance across a wide range of tasks, demonstrating generalization capabilities far beyond traditional NLP applications (Fu et al., 2022; Wei et al., 2022a). With their powerful language understanding and generation abilities, LLMs are increasingly regarded as a promising pathway toward Artificial General Intelligence (AGI) (Chang et al., 2024; Zhao et al., 2023a). Naturally, efforts have begun to investigate the potential of LLMs in tabular data processing, leveraging their language interfaces to unify the modeling of structured information. These approaches have shown promising results in tasks such as prediction, question answering, reasoning, and data augmentation (Hegselmann et al., 2023; Sui et al., 2024a; Cromp et al., 2026). Recent work has also explored fairness-aware in-context learning for tabular data, where tabular foundation models are adapted to mitigate bias during inference (Kenfack et al., 2026). In this section, we provide a brief overview of LLM-driven techniques for tabular data modeling.

A natural approach to leveraging LLMs for tabular data is to first convert the structured data into a textual format that the model can understand. This can be achieved by representing the table using programming-language-friendly data structures—such as Python's Pandas DataFrame loaders, line-separated JSON files, nested lists for data matrices, or HTML table representations. Alternatively, tables can be serialized into delimiter-separated values, where delimiters such as commas or tabs are commonly used. Some studies further transform tables into human-readable sentences by applying templates that incorporate column headers and corresponding cell values, enabling LLMs to process the data in a more context-aware manner (Singha et al., 2023; Narayan et al., 2022; Yu et al., 2023a). Several studies further analyze the robustness of LLMs for tabular question answering by examining internal attention patterns, providing insights into how models align questions with tabular evidence (Bhandari et al., 2025).

Another key direction for optimizing LLM-based tabular data modeling lies in how to effectively structure and present tabular information to maximize the model's comprehension. One common strategy is context compression, which addresses the input length limitations of LLMs. For instance, researches propose truncation techniques based on maximum sequence length (Liu et al., 2022c; Herzig et al., 2020). Gpt4table (Sui et al., 2024a) introduce predefined constraints to ensure that the input remains within the acceptable range for LLMs.

Another effective approach involves selective retrieval, where only the most relevant tables, rows, columns, or cells are extracted and fed into the model. In this context, Retrieval-Augmented Generation (RAG) has also emerged as a viable solution, combining retrieval mechanisms with generative models to enhance the model's ability to leverage structured information (Sundar & Heck, 2023; Gao et al., 2023b).

Beyond raw table content, several studies have explored enriching prompts with auxiliary tabular information. TAP4LLM (Sui et al., 2024b) demonstrate that incorporating metadata such as table dimensions, measurement units, and semantic field types can significantly improve model performance across six benchmark datasets. Furthermore, statistical features—particularly in datasets with a high proportion of quantitative cells, such as fever records (Aly et al., 2021)—have been shown to enhance task accuracy. Additional contextual signals, such as document references and terminology explanations, help inject semantic meaning into tabular inputs. However, not all augmentations yield positive results: while statistical and semantic cues are beneficial, features like table size offer limited gains, and header hierarchies may introduce unnecessary complexity, even degrading performance.

In addition to structural input optimization, two widely adopted strategies for enhancing LLM performance on tabular data tasks are prompt engineering and fine-tuning. Prompt engineering techniques aim to guide the model's reasoning and output by crafting effective input prompts. A common baseline is to concatenate a task description with a serialized table, allowing the model to interpret and respond accordingly. Research has shown that well-defined and explicitly stated task instructions significantly improve performance (Ggaliwango et al., 2024). In addition, TST-LLM (Han et al., 2025) leverages LLMs to generate task-aware pretext objectives, enabling more effective representation learning without manual labels. Furthermore, incorporating external context, such as relevant questions or statements alongside the table, can enhance the model's understanding and accuracy (Sui et al., 2024a).

Advanced prompting strategies such as Chain-of-Thought (CoT) prompting (Wei et al., 2022b) encourage step-by-step reasoning, enabling the model to decompose complex tasks into intermediate steps. Variants like program-aided CoT (Chen et al., 2023d) and Self-Consistency (SC) (Wang et al., 2023a) further boost performance by sampling multiple reasoning paths and aggregating the most consistent outcomes. These methods are particularly effective in numerical reasoning, table-based QA, and fact verification tasks.

In contrast to prompt-based approaches, fine-tuning adapts LLMs to specific tabular tasks by updating model parameters with labeled data, often resulting in improved performance and tighter integration with external tools such as SQL and Python. Recent studies have actively explored how LLMs can interact with these tools to enhance reasoning over structured data. For instance, an iterative framework is proposed in (Zhang et al., 2023a), where execution errors from SQL queries are fed back into the LLM to enable successive refinements that significantly improve query generation accuracy. In the realm of analytical platforms, the work of (Liu et al., 2023c) introduces a no-code solution that leverages LLMs to generate data summaries, formulate relevant analytical questions, and produce structured queries.

Furthermore, a survey by (Zhang et al., 2024f) reviews natural language interfaces for tabular data, highlighting recent advances in Text-to-SQL and Text-to-Vis techniques. Together, these fine-tuning and tool-integrated approaches showcase the growing potential of LLMs in enabling more intuitive and effective interaction with tabular data. A recent survey (Fang et al., 2024) also reviews how LLMs are adapted for tabular prediction, data generation, and reasoning, highlighting emerging trends and open challenges.

Summary and takeaways. The main lesson from LLM-driven tabular modeling is that success depends less on using a LLM and more on controlling how structured information is presented to it. Methods tend to improve when they reduce irrelevant context, preserve semantic cues such as field types and units, and complement the model with retrieval or external tools. At the same time, the literature suggests that LLMs are currently strongest when used for table reasoning, augmentation, or tool-mediated interaction, rather than as drop-in replacements for specialized predictive tabular models. The central bottleneck remains representation mismatch: numerical precision, token efficiency, and table structure are only imperfectly preserved after serialization.

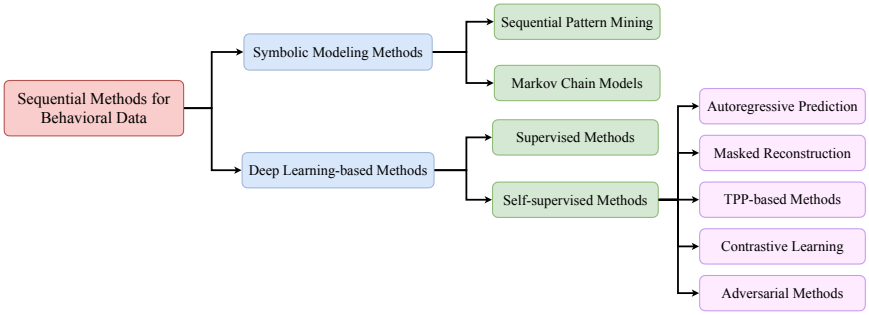


Figure 4: The dissection of the event sequence representation learning methods.

4.2 Event Sequences

Rooted in the foundational principles of sequential decision-making (Sutton & Barto, 1998), human behavior can be naturally modeled as a sequential process. Unlike traditional feature-based analysis, sequence-based structural modeling captures temporal dynamics, enabling more accurate behavior representation by leveraging historical patterns to inform present decision-making. Typically, a behavior sequence consists of a continuous series of ordered interactions, each associated with a timestamp and relevant contextual features. This formulation allows for the application of powerful sequential representation learning models to behavioral data analysis. Broadly, sequential modeling approaches can be categorized into **(1) Symbolic Modeling Methods**, and **(2) Deep Learning-based Methods**.

Table 2: Event sequence representation learning methods.

Method	Training Paradigm	Model Architecture	Year
Symbolic Modeling Methods			
FPMC (Rendle et al., 2010)	Self-supervised learning	Markov Chain	2010
PSPM (Yap et al., 2012)	Self-supervised learning	Sequential pattern mining	2012
PRME (Feng et al., 2015)	Self-supervised learning	Markov Chain	2015
Deep Learning-based Methods			
GRU4Rec (Hidasi et al., 2016)	Supervised learning	GRU	2016
RMTPP (Du et al., 2016)	Self-supervised learning	RNN	2016
NHP (Mei & Eisner, 2017)	Self-supervised learning	LSTM	2017
Event2Vec (Hong et al., 2017)	Supervised learning	Word2Vec	2017
IRGAN (Wang et al., 2017)	Semi-supervised learning	GAN	2017
HRNN (Quadrana et al., 2017)	Supervised learning	RNN	2017
Caser (Tang & Wang, 2018)	Supervised learning	CNN	2018
AttRec (Zhang et al., 2018)	Supervised learning	Transformer	2018
MANN (Chen et al., 2018)	Supervised learning	Memory network	2018
RecGAN (Bharadhwaj et al., 2018)	Semi-supervised learning	GAN+RNN	2018
RPPN (Xiao et al., 2019)	Self-supervised learning	RNN	2019
BERT4Rec (Sun et al., 2019b)	Self-supervised learning	BERT	2019
DTCDR (Zhu et al., 2019a)	Supervised learning	MLP	2019
FDSA (Zhang et al., 2019b)	Supervised learning	Transformer	2019
NextItNet (Yuan et al., 2019)	Supervised learning	CNN	2019
SAHP (Zhang et al., 2020)	Self-supervised learning	Transformer	2020
THP (Zuo et al., 2020)	Self-supervised learning	Transformer	2020
BEHRT (Li et al., 2020c)	Self-supervised learning	BERT	2020
TiSASRec (Li et al., 2020a)	Supervised learning	Transformer	2020
RAPT (Ren et al., 2021)	Self-supervised learning	Transformer	2021
CoSeRec (Liu et al., 2021a)	Self-supervised learning	GAN+CL	2021
ASReP (Liu et al., 2021b)	Supervised learning	Transformer	2021
UniSRec (Hou et al., 2022)	Self-supervised learning	BERT+Transformer	2022
RecGURU (Li et al., 2022a)	Self-supervised learning	Transformer	2022
promptTPP (Xue et al., 2023)	Continual learning	Transformer+ Prompts	2023
Meta TPP (Bae et al., 2023)	Meta learning	Transformer	2023
BERT4ETH (Hu et al., 2023b)	Self-supervised learning	BERT	2023
PrimeNet (Chowdhury et al., 2023)	Self-supervised learning	Transformer	2023
ECGAN-Rec (Ni et al., 2023)	Semi-supervised learning	GAN	2023
SeqLink (Abushaqra et al., 2024)	Supervised learning	Neural ODE	2024
Chronos (Ansari et al., 2024)	Self-supervised learning	Transformer	2024
player2vec (Wang et al., 2024d)	Self-supervised learning	BERT	2024
TOTEM (Talukder et al., 2024)	Self-supervised learning	VQ-VAE	2024
PhASER (Mohapatra et al., 2025)	Supervised learning	Multiple	2025
Residual TPP (Yuan & Fang, 2025)	Self-supervised learning	Hawkes + Neural TPP	2025
IOCLRec (Wang et al., 2025)	Self-supervised learning	Transformer+CL	2025
HORAE (Hu et al., 2025)	Self-supervised learning	Transformer	2025

4.2.1 Symbolic Modeling Methods

Symbolic modeling methods include **(a) Sequential Pattern Mining** and **(b) Markov Chain Models**.

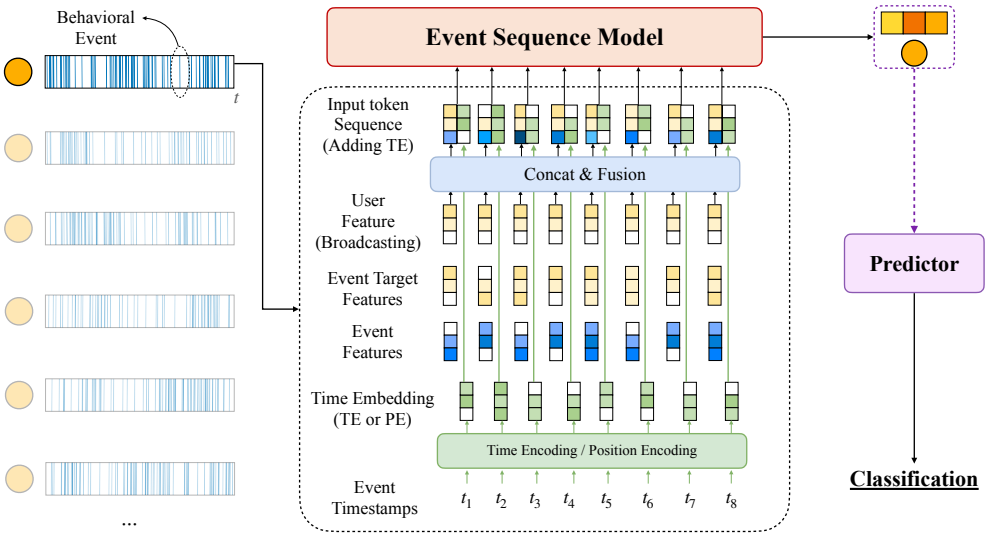


Figure 5: The pipeline of supervised event sequence representation learning methods.

Sequential Pattern Mining Sequential Pattern Mining follows a paradigm that first extracts frequent patterns from sequences and then utilizes them to predict the next action. Several studies have adopted this simple yet effective approach for behavioral data analysis (Yap et al., 2012; Wang & Cao, 2020). Although this solution is feasible, it not only incurs high computational costs but also predominantly relies on frequent patterns for prediction, thereby overlooking the informative value of infrequent patterns (Wang et al., 2019).

Markov Chain Models Markov Chain have been widely employed to model behavior sequence transitions. Existing approaches can be categorized into two main paradigms: (1) direct computation of transition probabilities from historical sequences (Cheng et al., 2013; Zhang et al., 2014), and (2) probability estimation through Euclidean distance between action embeddings (Moore et al., 2013; Feng et al., 2015). However, due to the fundamental Markov assumption, where the next state depends solely on the current state, these methods exhibit two critical limitations: they struggle to capture long-term dependencies and fail to model joint probabilities where multiple actions collectively influence the next action (Feng et al., 2015).

Summary and takeaways. Symbolic sequence modeling methods remain useful because they make the assumptions of early behavioral sequence analysis explicit. Sequential pattern mining emphasizes frequent co-occurrence structures, while Markov models focus on local transition dynamics. Their practical value lies in interpretability and conceptual clarity. However, these approaches also reveal the limitations that motivated later developments: frequent-pattern pipelines are computationally expensive and often biased toward common motifs, whereas Markov assumptions struggle to capture long-range dependencies and complex multi-action interactions.

4.2.2 Deep Learning-based Methods

In recent years, deep learning has become the dominant approach for sequential modeling in behavioral data analysis. These methods can be broadly categorized into **(a) Supervised** and **(b) Self-supervised** learning paradigms, each with distinct frameworks that we will examine separately.

Supervised Methods Supervised learning methods for behavior sequence representation utilize annotated datasets to optimize task-specific objectives such as classification or regression through direct minimization of supervised loss functions (e.g., cross-entropy or mean squared error). Although dependent on manual annotation, these methods achieve strong empirical performance when sufficient labeled data are available, with architectural flexibility across domains. The pipeline of supervised methods is shown in Fig. 5.

To illustrate the progression of this field, we structure the review chronologically, tracing the evolution of sequential supervised models from early RNN-based encoders to the latest LLM-driven architectures for behavior modeling.

The inherent sequential nature of event data found a natural match in RNNs. The core idea of an RNN is to maintain a hidden state, h_t, which acts as a compressed memory of the entire history seen up to event e_t. This state is updated recursively with each new event:

$$h_t = f(h_{t-1}, x_t)$$

where x_t is the embedding of the current event e_t, and $f(\cdot)$ denotes the state update function that integrates the previous hidden state and the current input. To predict the next event, the final hidden state is passed through a prediction layer (e.g., a softmax layer over the item vocabulary). Beyond RNN-based approaches, Memory-Augmented Neural Networks (MANN) (Chen et al., 2018) enhanced collaborative filtering through external memory modules, while neuromorphic event modeling techniques like Event2Vec (Hong et al., 2017) adapted Word2Vec's (Church, 2017) embedding paradigm for supervised classification of benchmark datasets such as the ASL-DVS gesture recognition dataset (Bi et al., 2019). SeqLink (Abushaqra et al., 2024) addresses sparsely and irregularly observed behavior sequences by modeling continuous-time latent dynamics with Neural Ordinary Differential Equations (ODEs), aggregating multiple linked latent trajectories to obtain robust representations under long observation gaps.

RNN-based models are often criticized for their computational inefficiency, as their autoregressive nature requires sequential processing and prevents parallelization. In contrast, CNNs overcome this limitation by processing entire sequences in parallel. Beyond efficiency gains, CNN-based models excel at capturing local temporal patterns in user behavior. Caser (Tang & Wang, 2018) adopted CNN as encoder to capture local temporal dynamics in user-item interactions. Another notable approach is NextItNet (Yuan et al., 2019), which employs stacked holed convolutional layers to efficiently expand receptive fields for sequential modeling.

More recently, Transformers have gained widespread adoption for sequential behavior modeling. SASRec (Kang & McAuley, 2018) was the first to explore Transformer architectures for sequential recommendation tasks, inspiring numerous follow-up improvements. Many subsequent Transformer-based methods are developed. TiSASRec (Li et al., 2020a) incorporated personalized time interval processing into attention mechanism to better model temporal patterns in recommendation sequences. AttRec (Zhang et al., 2018) integrated self-attention with metric learning for next-item prediction and ASReP (Liu et al., 2021b) proposed reverse-trained Transformers to augment short behavior sequences, further advancing recommendation accuracy. PhASER (Mohapatra et al., 2025) further improves supervised classification by explicitly modeling non-stationarity through phase-aware representations, decomposing sequences into magnitude and phase components and applying phase-driven augmentation to enhance robustness under distribution shifts.

Recent advancements have also leveraged LLMs, given their remarkable capabilities in sequence understanding. For example, AlmostRec (Wu et al., 2025c) enhanced predictive modeling in sequential recommendation by aligning multi-modal user behavior (e.g., IDs, text, images) using a large multi-modal model and an ID prediction objective for controllable and accurate recommendations.

The effectiveness of each architecture hinges on its alignment with the domain's inductive biases: RNNs and their gated variants (e.g., Long Short-Term Memory, LSTM (Graves, 2012) and Gated Recurrent Unit, GRU (Chung et al., 2014)) are well-suited for modeling sequential dependencies through recursive state updates, enabling effective modeling of short- to medium-range user behavior; memory networks facilitate iterative refinement of user state via external storage; CNNs excel at extracting shift-invariant local patterns in short-term interactions; Transformers capture global contextual dependencies through fully parallelizable computation; and LLMs generalize across symbolic and multi-modal behavior via large-scale pre-training. Beyond architectural choices, recent work also studies training principles for supervised sequence representations. In particular, Butera et al. (Butera et al., 2025) show that properly regularizing sequence-specific embeddings are crucial to prevent co-adaptation with shared layers, thereby improving representation quality and transferability.

Self-supervised Methods Self-supervised learning for behavior sequence modeling typically learns sequence representations through five key paradigms: Autoregressive Prediction, Masked Reconstruction,

TPP-based Methods, Contrastive Learning, and Adversarial Methods. In the following sections, we categorize and summarize existing work according to these paradigms.

i) Autoregressive Prediction: Autoregressive Prediction is the framework that learns sequence representations by predicting the future values based on observed historical data, thus encouraging the model to learn the underlying temporal structure and behavior patterns. Such models treat the event sequence as a unidirectional process, where the future is generated step-by-step conditioned on the past. This formulation aligns naturally with many real-world tasks such as next-click prediction (Yuan et al., 2019; Cheng et al., 2013), stock price forecasting (Hu et al., 2023b), and sequential decision-making (Wang et al., 2024d). For instance, inspired by the parallels between generative modeling and financial tasks such as churn prediction, credit default, and expenditure forecasting, NPPR (Skalski et al., 2023) employs next event prediction adapted from language modeling to handle multivariate transaction events. Chronos (Ansari et al., 2024) further formulates time series modeling as a language-modeling-like problem by discretizing continuous values into tokens and training transformer models with a next-token prediction objective. By autoregressively modeling token sequences via cross-entropy loss, Chronos learns general-purpose time series representations that transfer across domains without task-specific architectures.

ii) Masked Reconstruction: Masked Reconstruction learns sequence representations by reconstructing masked portions of historical records (e.g., actions or features). During training, the model processes partially masked input sequences and learns to predict the original data using the remaining unmasked tokens as context. This reconstruction objective forces the model to capture both temporal patterns and structural dependencies in the data, which enhances its effectiveness for downstream classification tasks.

This paradigm closely resembles the successful BERT model in NLP. A substantial body of research directly adapts BERT's encoder architecture and Masked Language Modeling (MLM) objective for behavior sequences, repurposing its bidirectional attention mechanism to capture complex temporal dependencies in actions and features. For example, BERT4Rec (Sun et al., 2019b) leverages a bidirectional Transformer encoder for masked item prediction, enabling robust information aggregation across entire historical sequences. Similarly, BEHRT (Li et al., 2020c) builds upon BERT's framework, adopting a two-stage pre-training and fine-tuning approach while integrating multi-modal data from five heterogeneous data sources to enhance depression prediction accuracy. Furthermore, BERT4ETH (Hu et al., 2023b) improves Ethereum transaction pattern recognition using a Transformer architecture with masked address modeling and domain-specific strategies to address data repetitiveness, skewness, and heterogeneity. Meanwhile, player2vec (Wang et al., 2024d) adapts long-context Transformer architectures to model player behavior from sessionized interaction logs, producing enriched behavior representations.

Beyond BERT-like methods, some works adopt reconstruction-based objectives without explicit masking to learn general-purpose behavioral representations. For instance, TOTEM (Talukder et al., 2024) learns discrete time series representations by training a Vector-Quantized Autoencoding Reconstruction (VQ-VAE) tokenizer to reconstruct unmasked historical sequences in a self-supervised manner, producing a fixed codebook of temporal tokens shared across domains and tasks.

Conversely, an alternative research direction argues for architectural modifications to address inherent limitations of vanilla Transformers in modeling behavioral data. Innovations such as RAPT (Ren et al., 2021) augment self-attention with explicit temporal encoding to handle irregular inter-event intervals, while SAGE (Wang et al., 2023b) introduces noisy reconstruction, injecting deletions, substitutions, and insertions into the masked sequence, leading to notable performance improvements.

This methodological divergence highlights distinct design philosophies: while BERT-like models provide powerful general-purpose sequence representations, specialized Transformer variants offer tailored solutions for temporal dynamics and data quality challenges in behavior analytics.

iii) TPP-based Methods: Temporal Point Processes (TPPs) are a class of probabilistic models designed to characterize the timing of events over continuous time by explicitly preserving the original timestamp of each occurrence. Unlike traditional sequence models which predict the next token in a discrete sequence, TPPs model the conditional intensity function to estimate the probability of an event occurring at any time t, given its event history. This unique formulation makes TPPs particularly effective for handling irregular,

asynchronous event sequences as commonly found in behavioral data streams such as click logs, social media actions, or medical visit records.

While traditionally formulated in the context of generative modeling, TPPs can also be viewed as a specialized form of self-supervised learning. They optimize a likelihood-based objective by predicting the timing of future events solely based on past observations, without requiring any external labels.

Early deep learning approaches for TPPs naturally leveraged RNNs (Cho et al., 2014) to encode the event history. For instance, RMTPP (Du et al., 2016) uses an RNN to learn a representation of the event history, embedding the sequence of past event times and their associated markers into a hidden state vector which then parameterizes the intensity function. Similarly, NHP (Mei & Eisner, 2017) introduces a novel continuous-time LSTM architecture where the hidden state evolves continuously between events, allowing it to capture more complex, non-linear patterns of excitation and inhibition than traditional Hawkes models. RPPN (Xiao et al., 2019) jointly model asynchronous event sequences and synchronous time series by parameterizing the endogenous and exogenous intensity functions with separate RNNs.

However, RNN-based TPPs struggle to effectively capture long-term dependencies, subsequent research drew inspiration from the success of attention mechanisms in NLP to address this issue. For example, SAHP (Zhang et al., 2020) and THP (Zuo et al., 2020) employ self-attention to summarize the influence of history events and compute the probability of the next event, which not only models long-range dependencies more effectively but also allows for parallel computation over the event sequence, breaking the sequential processing bottleneck of RNNs (Cho et al., 2014). There has also been research in creative ways of training temporal point processes, such as in continual learning framework (Xue et al., 2023) whose basis is a continuous-time retrieval prompt pool for modeling streaming event sequences and meta learning framework (Bae et al., 2023) to quickly adapt to new, heterogeneous sequences from few examples.

While attention-based TPPs improve expressiveness, they often incur high computational overhead and face difficulty in training on long sequences. To address this, the residual TPP (Yuan & Fang, 2025) proposes a hybrid framework that first employs a lightweight Hawkes process to capture dominant temporal patterns, and then delegates only the residual, irregular dynamics to a neural TPP, significantly reducing training complexity while preserving accuracy.

iv) Contrastive Learning: Contrastive Learning approaches aim at maximizing similarity between semantically related sequences while minimizing similarity between unrelated ones, thus improving generalization across tasks. Contrastive learning has been widely used in many fields including computer vision and sequential modeling. To make it more powerful in modeling behavioral data, many works have proposed modifications and focuses mainly on three aspects: *sample selection, loss function, and data augmentation.*

- *Sample Selection*: A critical challenge in modeling behavioral data lies in the irregular timing of events, which requires specialized sampling strategies to preserve temporal dynamics. For example, PrimeNet (Chowdhury et al., 2023) tackles this issue via time-sensitive stratified sampling and fixed-time masking, explicitly capturing irregular patterns in multivariate time series during augmentation. Beyond temporal irregularity, another line of research focuses on mitigating semantic bias in sampled sequences. In recommendation systems, IOCLRec (Wang et al., 2025) reduces false negatives in contrastive pairs by aligning user behavior sequences with their latent intents, thereby improving both robustness and recommendation accuracy.

- *Loss Function*: Another line of research focuses on enhancing contrastive learning through novel loss function designs to improve adaptability to specific scenarios. For instance, HORAE (Hu et al., 2025) explicitly incorporates temporal dynamics into its contrastive learning framework, enabling multi-interest pre-training by aligning user behaviors with their evolving time-aware representations. Similarly, S3-Rec (Zhou et al., 2020b) adapts contrastive learning to recommendation systems by introducing four auxiliary self-supervised objectives, which jointly capture multi-view semantic correlations (e.g., item attributes, sequential dependencies) to improve representation learning. These approaches demonstrate that tailored loss functions can effectively disentangle complex behavioral semantics.

- *Data Augmentation*: Contrastive learning can also be enhanced through innovative data augmentation strategies. CL4SRec (Xie et al., 2022), a pioneering work in sequential recommendation, proposes three atomic augmentation operations (crop, mask, reorder) to generate self-supervised signals from interaction sequences. Subsequent studies further develop this paradigm: CoSeRec (Liu et al., 2021a) introduces substitute and insert operations to create more informative augmented views, while DuoRec (Qiu et al., 2022) proposes model-level augmentation via Dropout and a supervised positive sampling strategy to address representation degeneration in sequential modeling.

v) Adversarial Methods: Adversarial Methods employ a discriminator to distinguish between generated fake samples and real samples, while the generator improves the quality of its outputs to deceive the discriminator. This framework has been effectively adapted for behavioral data processing. For example, RecGAN (Bharadhwaj et al., 2018) integrates RNNs and GANs through adversarial training to model temporal preferences for recommendations. Similarly, IRGAN (Wang et al., 2017) unifies generative and discriminative retrieval paradigms via a minimax adversarial game, where the generative model produces challenging samples and the discriminative model refines its ranking, achieving significant improvements in search, recommendation, and QA tasks. Further advancing this approach, ECGAN-Rec (Ni et al., 2023) combines contrastive learning with adversarial training to enhance sequential recommendation by balancing data sparsity and noise, while CoSeRec (Liu et al., 2021a) also leverages adversarial training alongside contrastive learning to boost performance of downstream tasks.

Summary and takeaways. The evolution of event sequence modeling shows a clear shift from local, short-horizon transition modeling toward architectures that better preserve long-range temporal context. In supervised settings, Transformers often outperform recurrent baselines when sufficient history and scale are available, largely because they model global dependencies and parallelize training more effectively. However, this does not make recurrent or continuous-time methods obsolete: when event timing is highly irregular, labels are scarce, or the task depends on fine-grained temporal calibration, architectures with stronger temporal inductive bias remain competitive. In self-supervised settings, masked reconstruction has emerged as the most reusable pretraining signal, while contrastive methods are effective only when augmentation preserves behavioral semantics. The main empirical bottlenecks are therefore not only model choice, but also how time irregularity, sequence length, and augmentation validity are handled.

4.3 Dynamic Graphs

Human behaviors frequently manifest as evolving interactions among individuals or entities over time, naturally lending themselves to representation as dynamic graphs. Specifically, a dynamic graph captures temporal variations in connections between entities (e.g., people, items), thereby providing a structured framework to analyze complex, time-dependent relational patterns.

Formally, behavioral data can be modeled as a dynamic graph $G = (V, E, T)$, where V is a set of nodes representing individuals or entities, E denotes the set of interactions or relationships (edges) among nodes, and T explicitly encodes temporal information. Based on how the temporal dimension T is defined, dynamic graphs are typically classified into two distinct types:

- **Discrete-Time Dynamic Graphs**: In Discrete-Time Dynamic Graphs (DTDGs), temporal dynamics are captured via a series of graph snapshots at discrete time intervals (Chen et al., 2024d): $\mathcal{G} = \{G_1, G_2, \ldots, G_T\}, G_t = (V_t, E_t), t = 1, 2, \ldots, T$, where each snapshot G_t describes the structure of interactions within a fixed time interval, representing aggregated relationships.

- **Continuous-Time Dynamic Graphs**: Continuous-Time Dynamic Graphs (CTDGs) record each interaction as individual events at precise timestamps, thus offering finer granularity (Rossi et al., 2020): $G = (V, \mathcal{E}), \mathcal{E} = \{(u, v, t) \mid u, v \in V, t \in \mathbb{R}^+\}$, where each edge (u, v, t) indicates a direct interaction between nodes u and v occurring at a continuous timestamp t.

Dynamic graphs typically facilitate two primary downstream analytical tasks: Dynamic Link Prediction and Node Classification. While a variety of other downstream tasks have also been explored, such as anomalous

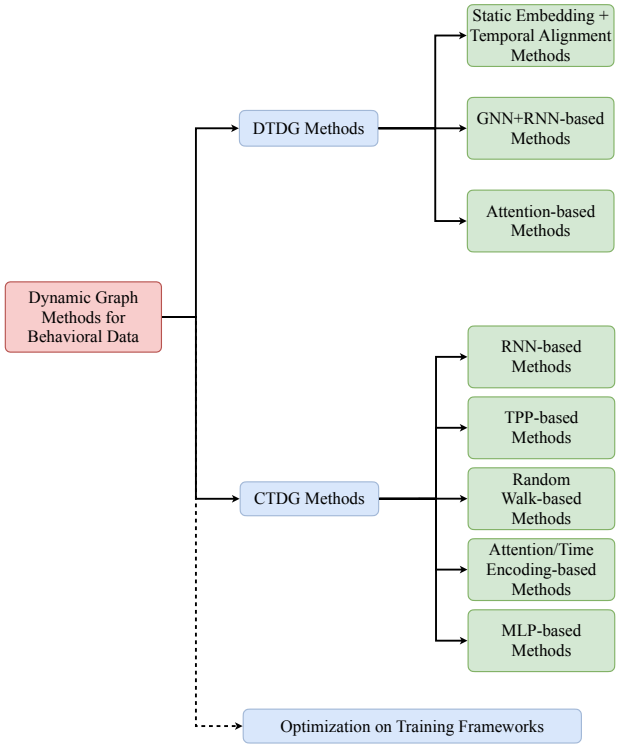


Figure 6: The dissection of dynamic graph representation learning methods.

link detection (Postuvan et al., 2024), they are comparatively less studied and often application-specific; therefore, we focus primarily on the two dominant tasks in this paper. The primary output of representation learning methods for both DTDG and CTDG is typically dynamic node embeddings, denoted as $\mathbf{z}_v(t)$ for a node v at time t. These embeddings encode both structural and temporal information accumulated up to time t, and serve as the basis for various downstream tasks.

- **Dynamic Link Prediction**: Predicting future interactions between nodes based on historical graph structures. Dynamic Link prediction is typically formulated as a self-supervised task, where the model is trained to predict future interactions based on historical data. Specifically, the future edges serve as implicit supervision signals, allowing the model to learn temporal patterns without requiring manual labels. Given a node pair (u, v) and a query time t, the objective is to estimate the likelihood of a future edge:

$$\hat{y}_{uv}(t) = \sigma\left(f_{\text{link}}(\mathbf{z}_u(t), \mathbf{z}_v(t))\right), \tag{1}$$

  where f_{link} denotes a scoring function (e.g., dot product), and $\sigma(\cdot)$ is a sigmoid function. For behavioral data, link prediction corresponds to real-world tasks such as forecasting social connections, recommending potential friends, or predicting future collaborative partnerships.

- **Node Classification**: Identifying node attributes or categories based on their temporal relational context. Node classification is usually treated as a supervised task, where a set of nodes is associated with ground-truth labels. In practice, the node embeddings $\mathbf{z}_v(t)$ learned from the link prediction task are used, and an additional classification head is trained on top of these embeddings using the available label data:

$$\hat{y}_v = \text{softmax}\left(f_{\text{cls}}(\mathbf{z}_v(t))\right), \tag{2}$$

  where f_{cls} is typically a Multi-Layer Perceptron (MLP) or linear classifier. Practical tasks for behavioral data include detecting influential users in social networks, inferring user demographics, or identifying community roles based on interaction patterns.

Table 3: Discrete-time dynamic graph representation learning methods.

Method	Structural Encoding	Temporal Encoding	Year
Static Embedding + Temporal Alignment Methods			
Chakrabarti et al. (Chakrabarti et al., 2006), Chi et al. (Chi et al., 2009), Kim & Han (Kim & Han, 2009), Gupta et al. (Gupta et al., 2011), Yao et al. (Yao et al., 2016), Zhou et al. (Zhou et al., 2018)	Matrix factorization	Smoothness regularization or alignment	2006-2018
Hisano (Hisano, 2018)	Matrix factorization	Time window aggregation	2018
Sharan & Neville (Sharan & Neville, 2008)	Matrix factorization	Time-weighted adjacency matrices	2008
Ibrahim et al. (Ibrahim & Chen, 2015)	Matrix factorization	Exponential decay	2015
Ahmed et al. (Ahmed et al., 2016)	Low-rank adjacency	Temporal sampling strategies	2016
Singer et al. (Singer et al., 2019)	Random walk	Init. from previous step + fine-tuning	2019
DynGEM (Goyal et al., 2018)	Deep autoencoder	Regularization across snapshots	2018
DynamicTriad (Zhou et al., 2018)	Triadic closure	Temporal smoothness	2018
GNN+RNN-based Methods			
GCRN (Seo et al., 2018)	GCN	LSTM	2018
Narayan & Roe (Narayan & O'N Roe, 2018)	GraphSAGE	LSTM	2018
TGCN (Zhao et al., 2020)	GCN	GRU	2020
TNA (Bonner et al., 2019)	GCN	GRU	2019
VGRNN (Hajiramezanali et al., 2019)	VGAE	LSTM	2019
LRGCN (Li et al., 2019)	R-GCN	LSTM	2019
E-LSTM-D (Chen et al., 2021)	Autoencoder	LSTM	2021
EvolveGCN (Pareja et al., 2020)	GCN	GRU	2020
dyngraph2vec (Goyal et al., 2020)	graph2vec	LSTM/GRU	2020
TeMP (Wu et al., 2020)	GCN	GRU or Attention	2020
WD-GCN/CD-GCN (Manessi et al., 2020)	GCN	Modified LSTM	2020
HDGNN (Zhou et al., 2020a)	Heterogeneous random walk	Bi-RNN	2020
HTGN (Yang et al., 2021)	Hyperbolic attention-based GCN	Hyperbolic GRU	2021
GC-LSTM (Chen et al., 2022)	GCN	LSTM	2022
ROLAND (You et al., 2022)	GCN	Adaptive RNN	2022
RPC (Liang et al., 2023)	GNN	GRU	2023
SEIGN (Qin et al., 2023)	GCN-like message passing	GRU parameter adjustments	2023
RETIA (Liu et al., 2023b)	GCN	GRU + LSTM	2023
MegaCRN (Jiang et al., 2023)	Meta-graph learner	Custom GRU	2023
DEFT (Bastos et al., 2023)	GNN	RNN-based parameter evolution +Wavelet	2023
STGNPP (Jin et al., 2023)	GCN +Transformer	Continuous GRU	2023
WinGNN (Zhu et al., 2023b)	GNN	Sliding window	2023
SpikeNet (Li et al., 2023a)	GNN	SSN	2023
TTGCN (Li et al., 2024b)	Truss-based GCN	GRU	2024
Attention-based Methods			
DySAT (Sankar et al., 2020)	Graph attention	Graph attention	2020
TEDIC (Wang et al., 2021d)	Graph diffusion	Temporal convolutional network	2021
DyHATR (Xue et al., 2021)	Hierarchical attention	Temporal attentive RNN	2021
DREAM (Zheng et al., 2023)	Attention	Attention + Reinforcement learning	2023
STGNP (Hu et al., 2023a)	Dilated causal convolution	Cross-set graph convolution	2023
DTFormer (Chen et al., 2024d)	Transformer	Transformer	2024

In the following subsections, we systematically review and analyze representation learning methods designed for discrete-time and continuous-time dynamic graphs, respectively, followed by a discussion of the **Optimization on Training Frameworks** that are applicable across dynamic graph models. We highlight the core methodologies, recent advancements, and existing limitations of each category. It is important to note that this section focuses on introducing how the backbone models modeling behavioral data, while their associated downstream applications will be discussed in detail in Sec. 5.

4.3.1 Discrete-Time Dynamic Graphs

DTDGs capture evolving networks by recording the graph at successive fixed time intervals as a sequence of static snapshots (Feng et al., 2025). This snapshot-based view makes it easy to leverage traditional static graph methods (by treating each snapshot as independent static graph), while still preserving the coarse temporal ordering across snapshots.

Most representation learning methods for DTDG follow a two-step modeling paradigm that first encodes the structural information within each snapshot and then captures temporal dependencies across snapshots. Given that the core distinction between DTDGs and static graphs lies in their temporal dynamics, we classify existing methods primarily based on how they model temporal dependencies, while also taking their

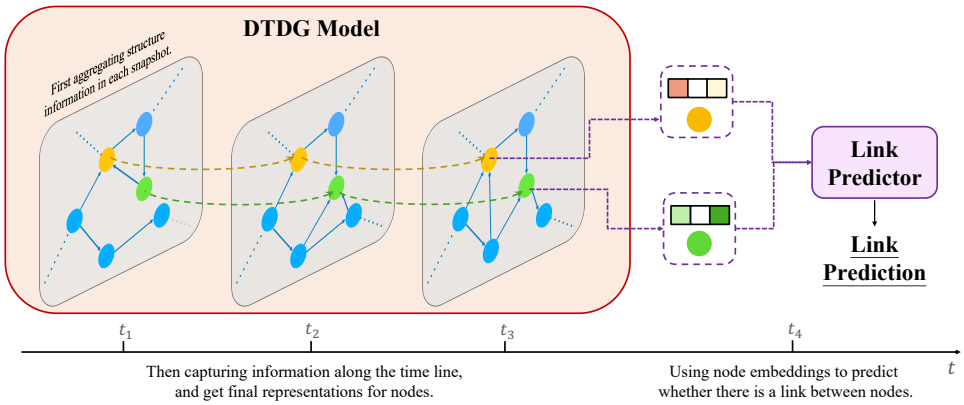
structural modeling strategies into account. Therefore, we propose to classify existing DTDG representation learning methods into three-way taxonomy. **(a) Static Embedding + Temporal Alignment Methods**: These methods adapt static graph algorithms to each snapshot and then fuse or smooth the results over time. Often, an explicit smoothness or alignment constraint is added so that the embedding of each node changes gradually from one snapshot to the next. This "shallow" approach thus incorporates temporal smoothing but does not learn an explicit temporal model. **(b) GNN+RNN-based Methods**: These hybrid models use a GNN to encode the structure of each snapshot and an RNN to propagate information through time. Concretely, a GNN produces a node-embedding vector from each graph snapshot, yielding a time-ordered sequence of embeddings for each node. An RNN is then run on that sequence to capture temporal dependencies, so that each node's final representation reflects both its local topology (via the GNN) and its evolution history (via the RNN). This recurrent approach effectively learns dynamic updates as the graph changes. **(c) Attention-based Methods**: These models leverage self-attention mechanisms to capture both temporal and structural dependencies without relying on recurrent updates. By applying attention mechanism these models enable flexible modeling of long-range temporal interactions, addressing common limitations of RNNs such as vanishing gradients and sequential bottlenecks. Almost all existing methods follow a two-step pipeline: first, they employ GNNs, Transformers, or other encoders to capture structural information; then, they apply a sequence model to capture temporal dynamics along the timeline and produce the final embedding. This pipeline is illustrated in Fig. 7a.

Note that this taxonomy is defined in a broad sense. For example, methods categorized under the GNN+RNN-based Architecture may not strictly adopt a GNN followed by a conventional RNN module; rather, we group together architectures that share a similar design principle. A comparison among DTDG representation learning methods is shown in Tab 3.

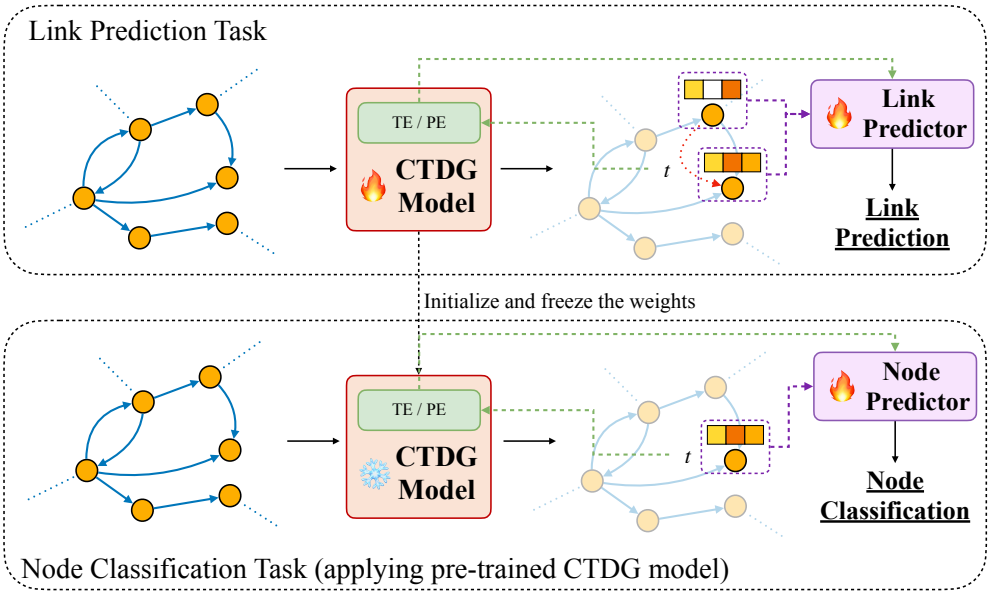
Static Embedding + Temporal Alignment Methods Early approaches to DTDG representation learning built upon techniques for static graphs, extending them to handle temporal evolution, as DTDGs are formed with a series of consecutive snapshots (Chen et al., 2024d). A common strategy is to apply matrix factorization or latent space models to each snapshot of the graph and impose temporal smoothness constraints between snapshots (Chakrabarti et al., 2006; Chi et al., 2009; Kim & Han, 2009; Gupta et al., 2011; Yao et al., 2016; Zhou et al., 2018; Kazemi et al., 2020). This effectively regularized the embedding trajectory of each node, assuming that node characteristics vary smoothly over time. Some methods aggregated multiple snapshots into a single decomposition: e.g., summing or averaging adjacency matrices over a time window (Liben-Nowell & Kleinberg, 2003; Hisano, 2018). However, such aggregation can lose temporal information (Kazemi et al., 2020), so later refinements introduced time-weighted sums (giving higher weight to recent snapshots) or exponential decay factors to emphasize current graph structure (Sharan & Neville, 2008; Ibrahim & Chen, 2015; Ahmed et al., 2016; Ibrahim & Chen, 2016).

Beyond matrix factorization, random walk-based embeddings (e.g., DeepWalk (Perozzi et al., 2014) and Node2Vec (Grover & Leskovec, 2016)) are also adapted: one could generate embeddings for each snapshot via truncated random walks and then align or smooth these embeddings over time. For instance, some methods initialized the embedding at time t with the result from $t-1$ and fine-tuned it, thereby leveraging past computations to speed up training (Chi et al., 2009; Singer et al., 2019). Along with the development of deep learning, DynGEM (Goyal et al., 2018) is one of the pioneering methods which utilizes deep autoencoders to learn node embeddings on each snapshot, while incorporating a regularization term to encourage the embeddings of adjacent snapshots to remain similar, thus capturing temporal smoothness. DynamicTriad (Zhou et al., 2018) further incorporated triadic closure process into structural modeling. Overall, these pre-GNN methods laid the groundwork by combining static graph embeddings with temporal smoothing or forecasting. They captured evolving structure to some extent, but often in a two-step or loosely coupled manner (separate static embedding and temporal alignment), without end-to-end learning of dynamic patterns.

GNN+RNN-based Methods The emergence of deep learning on graphs introduced end-to-end models for DTDG representation learning that integrate GNNs (Scarselli et al., 2009; Yao et al., 2019; Zhang et al., 2019a) with sequence models (i.e., RNNs (Cho et al., 2014)). A dominant early paradigm for DTDGs is the combination of GNNs for structural encoding and RNNs for temporal encoding. We refer to this model design as the GNN+RNN-based architecture. Without loss of generality, and assuming a fixed node set



(a) The general pipeline of DTDG representation learning methods.



(b) The training pipeline of CTDG representation learning methods.

Figure 7: The pipelines of DTDG and CTDG methods.

V across time steps for notational simplicity. Given a sequence of graph snapshots $\mathcal{G} = \{G_1, G_2, \ldots, G_T\}$, each snapshot is defined as $G_t = (V, E_t, X_t, F_t)$, where $X_t \in \mathbb{R}^{|V| \times d_x}$ is the matrix of node features, and $F_t : E_t \to \mathbb{R}^{d_e}$ provides edge features such as interaction types and weights. At each time step t, a GNN is applied to capture spatial dependencies within the snapshot. For each node $v \in V$, the GNN aggregates information from its neighborhood $\mathcal{N}_t(v)$, considering both node features and associated edge features:

$$h_v^{(t)} = \mathrm{GNN}\left(v, \mathcal{N}_t(v); X_t, F_t\right). \tag{3}$$

The resulting structural embeddings $\{h_v^{(t)}\} v \in V$ serve as input to a RNN, which models temporal evolution by tracking each node's state across time:

$$z_v^{(t)} = \mathrm{RNN}(z_v^{(t-1)}, h_v^{(t)}). \tag{4}$$

Here, $z_v^{(t)}$ denotes the final representation of node v at time t, encoding both current structural context and historical dynamics. This two-stage design, which uses a GNN for intra-snapshot relational modeling and an RNN for inter-snapshot temporal modeling, enables end-to-end learning of evolving node representations in dynamic graphs.

In this architecture, each graph snapshot (at a discrete time step) is first processed by a GNN (such as a graph convolutional network (Yao et al., 2019)) to produce node embeddings that capture the structural neighborhood information at that time (Chen et al., 2024d). Then, an RNN (typically an LSTM (Graves, 2012) or GRU (Chung et al., 2014)) takes these node embeddings from each snapshot as sequential inputs to produce updated node states, thereby propagating information through time (Chen et al., 2024d).

The GNN+RNN-based architecture allows the model to learn how a node's representation should evolve based on its current neighbors and its past states.

Early examples include the model GCRN (Seo et al., 2018), which applies a GCN followed by an LSTM to capture structured sequences, and the work proposed by Narayan and Roe (Narayan & O'N Roe, 2018) using GraphSAGE (Hamilton et al., 2017) with an LSTM for representation learning. Subsequent efforts expand on this GNN+RNN paradigm with diverse architectural innovations. TGCN (Zhao et al., 2020) and TNA (Bonner et al., 2019) tightly integrate GCNs with recurrent units to learn temporally-aware node embeddings, while VGRNN (Hajiramezanali et al., 2019) brings in variational inference for modeling temporal uncertainty. LRGCN (Li et al., 2019), E-LSTM-D (Chen et al., 2021), and GC-LSTM (Chen et al., 2022) further enhance temporal modeling through stacked LSTMs or by embedding GCNs into recurrent gating mechanisms.

More recent advances emphasize structural adaptability and computational efficiency: EvolveGCN (Pareja et al., 2020) evolves GCN parameters over time via LSTM; dyngraph2vec (Goyal et al., 2020) combines static embeddings with RNNs for sequential modeling; ROLAND (You et al., 2022) adapts node embeddings via dynamic recurrent controllers. Methods such as TeMP (Wu et al., 2020), WD-GCN/CD-GCN (Manessi et al., 2020), and HDGNN (Zhou et al., 2020a) incorporate attention, hierarchical aggregation, and heterogeneity-aware modules to enhance long-term dynamics modeling.

Additionally, models like HTGN (Yang et al., 2021), MegaCRN (Jiang et al., 2023), and DEFT (Bastos et al., 2023) explore hyperbolic spaces, meta-structure learning, and spectral wavelets respectively. Others target specific application contexts: RPC (Liang et al., 2023), RETIA (Liu et al., 2023b), and SEIGN (Qin et al., 2023) focus on temporal knowledge graphs; STGNPP (Jin et al., 2023) and WinGNN (Zhu et al., 2023b) address spatio-temporal prediction via Transformers and sliding windows; SpikeNet (Li et al., 2023a) leverages spiking neural dynamics; and TTGCN (Li et al., 2024b) exploits multi-scale truss structure for hierarchical temporal modeling.

These models effectively treat dynamic graph learning as a synchronous sequence modeling problem, where the RNN absorbs the node's history across snapshots. Despite architectural differences, all such GNN+RNN-based models address temporal dynamics by maintaining a state that carries information from prior snapshots. However, they also inherit limitations of RNNs: difficulty with very long-term dependencies and high memory usage for long sequences (Sherstinsky, 2020; Zhang et al., 2023d). Likewise, stacking many GNN layers per time step can cause over-smoothing of node features (Chen et al., 2020; Yang et al., 2020a). These challenges motivated the next generation of methods that seek to handle time without a standard RNN.

Attention-based Methods Motivated by the success of attention mechanisms in capturing long-range dependencies across various sequence modeling tasks (Vaswani et al., 2017), researchers have started to explore their application to DTDG representation learning. This has led to the emergence of attention-based methods.

DySAT (Sankar et al., 2020) is a representative two-stage model that replaces GNNs and RNNs with structural and temporal self-attention, enabling long-range dependency modeling while retaining architectural interpretability. TEDIC (Wang et al., 2021d) uses graph diffusion and Temporal Convolutions (TCN) to model interaction dynamics, supporting permutation invariance and efficient global context capture. DyHATR (Xue et al., 2021) combines hierarchical attention over heterogeneous relations with temporal attentive RNNs for evolving pattern modeling.

Later models push further toward non-recurrent and scalable designs: DREAM (Zheng et al., 2023), focusing on temporal knowledge graphs, incorporates multi-faceted attention with imitation learning to enhance reward modeling in multi-hop reasoning tasks. STGNP (Hu et al., 2023a) models extrapolation via a

neural latent variable framework with causal temporal convolutions and Bayesian graph aggregation to capture uncertainty and spatial dependencies. DTFormer (Chen et al., 2024d) adopts a fully attention-based architecture, processing structural and temporal patterns jointly with multi-patch encoders and position encodings, achieving strong scalability and predictive performance.

These models collectively shift from recurrent temporal modeling to attention-based or convolutional paradigms, emphasizing scalability, long-term dependencies, and integration of complex spatial-temporal semantics.

Summary and takeaways. The DTDG literature suggests that snapshot-based dynamic graph learning is appealing largely because it reuses the machinery of static graph representation learning while introducing only a lightweight temporal layer. This design makes DTDGs comparatively accessible for benchmarking and system development. However, it also introduces a clear trade-off: temporal evolution is captured coarsely through alignment, recurrence, or attention across snapshots rather than through event-level updates. Consequently, existing methods differ less in whether they model time than in how tightly structural encoding is coupled with temporal propagation. Early GNN+RNN architectures improved end-to-end temporal modeling, but long-horizon dependencies, sequential bottlenecks, and over-smoothing limited their scalability, motivating the development of attention-based alternatives.

4.3.2 Continuous-Time Dynamic Graphs

As DTDGs discretize time into a sequence of graph snapshots, each capturing the network state within a fixed interval, this formulation simplifies modeling by leveraging established techniques for static graphs, but it may overlook fine-grained temporal information and introduce artificial boundaries between interactions. In contrast, CTDGs model data as sequences of timestamped interaction events, allowing for more precise capture of asynchronous, real-world dynamics. This finer granularity enables CTDGs to better represent time-sensitive behaviors in domains such as financial transactions (Kazemi et al., 2020; Lezmi & Xu, 2023) and e-commerce (Kang & McAuley, 2018; Rossi et al., 2020; Skarding et al., 2021). As a result, CTDGs have attracted growing attention as a complementary approach to snapshot-based modeling, particularly in scenarios where modeling temporal precision and event order is critical (Rossi et al., 2020).

Moreover, CTDGs are also commonly referred to as Temporal Interaction Graphs (TIGs) or event-based temporal graphs. In this paper, we regard these terms as equivalent and collectively refer to them as CTDGs. DTDGs can also be transformed into CTDGs (Jiao et al., 2024; Huang et al., 2025b; Chen et al., 2024d) by assigning timestamp information from each snapshot to the corresponding edges. Intuitively, this conversion often results in multiple edges sharing the same timestamp, reflecting the coarser temporal granularity inherent in DTDGs.

In CTDGs, structural and temporal information are often intricately intertwined, necessitating joint modeling to accurately capture the evolution of interactions over time. As temporal dependencies critically influence both node behavior and structural dynamics (Feng et al., 2025; Jiao et al., 2025), we adopt a temporal modeling perspective as the primary taxonomy dimension, as it highlights the fundamental differences in how models encode, propagate, and reason about time-dependent interactions. Based on this criterion, we categorize existing CTDG methods into five principal classes. **(a) RNN-based Methods**: Use recurrent neural networks (e.g. LSTM/GRU) to sequentially integrate historical interactions, updating node embeddings over time. **(b) TPP-based Methods**: Employ temporal point processes (e.g., Hawkes processes) with conditional intensity functions to model event timing, capturing both historical decay and spontaneous interactions. **(c) Random Walk-based Methods**: Generate temporally constrained random walks that respect chronological order, aggregating spatio-temporal patterns from sequences of timestamped node transitions. **(d) Attention/Time Encoding-based Methods**: Apply self-attention mechanisms to sequences of event, enabling the model to capture long-range temporal dependencies without strict recurrence. However, since the attention mechanism is inherently position-invariant, time encoding is commonly introduced to explicitly incorporate temporal or positional information. In this category, we group together both methods that use attention-based architectures and those that, while not using attention mechanism, but still incorporate explicit time encoding to model temporal information. **(e) MLP-based Methods**: The necessity of complex temporal architectures has been challenged by adopting streamlined designs for tempo-

Table 4: Continuous-time dynamic graph representation learning methods.

Method	Structure Encoding	Temporal Encoding	Memory-based	Year
RNN-based Methods				
DeepCoevolve (Dai et al., 2016)	Implicit (via sequential interactions)	RNN	No	2016
JODIE (Kumar et al., 2019)	Implicit (via sequential interactions)	RNN + Projection	No	2019
Know-Evolve (Trivedi et al., 2017)	Implicit (via sequential interactions)	RNN	No	2017
RE-Net (Jin et al., 2020)	GCN	RNN	No	2020
HierTCN (You et al., 2019)	Implicit (via sequential interactions)	GRU + TCN	No	2019
DynGESN (Micheli & Tortorella, 2022)	Implicit (via sequential interactions)	Echo state network	No	2022
DyGNN (Ma et al., 2020)	GNN	LSTM	No	2020
AER-AD (Fang et al., 2023)	Local anonymous subgraph	GRU	No	2023
RTRGN (Chen et al., 2023f)	Implicit (via sequential interactions)	RNN	No	2023
TGN (Rossi et al., 2020)	Implicit (via message passing)	RNN + Memory	Yes	2020
NAT (Luo & Li, 2022)	Implicit (via sequential interactions)	RNN	Yes	2022
GDCF (Han et al., 2023)	Spatiotemporal GNN	RNN + Memory	Yes	2023
CDGP (Zhang et al., 2023f)	Community message passing	Time-aware aggregation	Yes	2023
TIGER (Zhang et al., 2023e)	Implicit (via message passing)	RNN + Dual memory	Yes	2023
RDGSL (Zhang et al., 2023c)	Implicit (via message passing)	RNN + Memory	Yes	2023
PRES (Su et al., 2024b)	Implicit (via message passing)	GMM-guided memory correction	Yes	2024
Ada-DyGNN (Li et al., 2024a)	Reinforced neighbor update	Time-based Policy	Yes	2024
SEAN (Zhang et al., 2024d)	Representative neighbor selector	RNN + Temporal-aware aggregation	Yes	2024
MemMap (Ji et al., 2024)	Latent memory-cell grid	Systematic memory routing	Yes	2024
MSPipe (Sheng et al., 2024)	Implicit (via message passing)	Staleness-aware update	Yes	2024
TPP-based Methods				
HTNE (Zuo et al., 2018)	Historical neighbor modeling	Hawkes process	No	2018
M^2DNE (Lu et al., 2019)	Micro/Macro temporal co-occurrence	Hierarchical TPP	No	2019
GHN (Han et al., 2020)	Entity-level structure modeling	Hawkes process	No	2020
DyRep (Trivedi et al., 2019)	Attentive structural encoding	Multi-scale TPP	No	2019
LDG (Knyazev et al., 2021)	Edge-quality structure adaptive	Adaptive TPP	No	2021
TREND (Wen & Fang, 2022)	Implicit (via sequential interactions)	Hawkes process + Transfer function	No	2022
DynShare (Zhao et al., 2023c)	Implicit (via sequential interactions)	Personalized TPP	No	2023
EasyDGL (Chen et al., 2024a)	GAT	TPP + Correlation masking	No	2024
Random Walk-based Methods				
CTDNE (Nguyen et al., 2018)	Timestamp-respecting walk	Skip-Gram over walks	No	2018
HNIP (Qiu et al., 2020)	Temporal random walk	Time-decay in walk sequence	No	2020
CAW (Wang et al., 2021c)	Causal Anonymous Walk	Hitting-count encoding	No	2021
NeurTWs (Jin et al., 2022)	Motif-guided random walk + ODE	ODE over walk path	No	2022
PINT (Souza et al., 2022)	Implicit (via message passing)	Provable temporal message passing	No	2022
TPNet (Lu et al., 2024)	Time-decayed walk matrix	Temporal relative encoding	No	2024
Attention/Time encoding-based Methods				
TGAT (Xu et al., 2020)	Temporal self-attention	Functional time encoding	No	2020
TCL (Wang et al., 2021a)	Transformer	Functional time encoding	No	2021
OTGNet (Feng et al., 2023)	Open graph attention	Extended time encoding	No	2023
TGRank (Suresh et al., 2023)	Temporal attention ranking	Enhanced time encoding	No	2023
DHGAS (Ji et al., 2023)	Heterogeneous GNN + Attention	Time encoding	No	2023
DyG2Vec (Alomrani et al., 2023)	Temporal edge encoding	Time encoding	No	2023
SimpleDyG (Wu et al., 2024b)	Transformer	Time and position encoding	No	2024
Todyformer (Biparva et al., 2024)	GNN+Transformer	Position encoding	No	2024
DyGFormer (Yu et al., 2023b)	1-hop neighbor + Co-occurrence	Time and position encoding	No	2023
APAN (Wang et al., 2021b)	Mailbox + Attention	Time encoding	Yes	2021
iLoRE (Zhang et al., 2023d)	Re-occurrence + Identity attention	Time encoding	Yes	2023
TDGNN (Qu et al., 2020)	GNN + Time-decay weighting	Exponential decay kernel	No	2020
DGEL (Tang et al., 2023)	Recent interactions	Time-aware normalization	No	2023
SUPA (Wu et al., 2023a)	Implicit (via sequential interactions)	Time modeling mechanisms	No	2023
FreeDyG (Tian et al., 2024)	Fourier-enhanced GNN	Functional time encoding	No	2024
CNE-N (Cheng et al., 2024)	Hash table-based memory	Temporal-diverse memory	Yes	2024
TG-Mixer (Zhang et al., 2024c)	Clustering patterns	Time encoding	No	2024
DyGMamba (Ding et al., 2025)	Mamba	Time encoding	No	2025
MLP-based Methods				
GraphMixer (Cong et al., 2023)	MLP + Mean pooling	Fixed time encoding	No	2023
RepeatMixer (Zou et al., 2024)	MLP + Repeat-aware sampling	Time-aware aggregation	No	2024
BandRank (Li et al., 2025b)	Frequency-band MLP	Band-pass time filters	No	2025

ral modeling. These methods leverage MLPs as the core building blocks for encoding temporal interactions, often combined with simple aggregation mechanisms and fixed or lightweight time encoding schemes.

In addition to the primary temporal modeling paradigm, another important dimension in CTDG modeling is the use of memory mechanisms (ENNADIR et al., 2025). The memory mechanism introduces the idea of maintaining persistent, continuously updated representations of nodes to track their evolving states over time. Rather than recomputing node representations from scratch, memory-based models incrementally update these state vectors, enabling efficient, asynchronous updates and capturing long-term temporal dependencies. Such memory modules often incorporate temporal encoding or attention to regulate the influence of past events, making them well-suited for modeling evolving behaviors in real-world systems. Importantly, the use of memory is orthogonal to the choice of temporal modeling paradigm and can be viewed as a secondary classification dimension. Thus, by organizing models first by their temporal modeling strategy and then by their use (or absence) of memory, we obtain a two-level taxonomy that more precisely characterizes how CTDG models manage evolving temporal and structural information.

A comparison among CTDG representation learning methods is shown in Tab. 4. Nevertheless, regardless of taxonomy, most methods adopt a similar training paradigm (Chen et al., 2024e): the backbone model is first "pre-trained" via a self-supervised link prediction task, and the resulting node embeddings are then used for downstream tasks, where a projection head is trained. The pipeline is illustrated in Fig. 7b.

RNN-based Methods A seminal work that initiated RNN-based representation learning in CTDGs is DeepCoevolve (Dai et al., 2016). DeepCoevolve (Dai et al., 2016) introduced a co-evolutionary framework with mutually-recursive RNNs and temporal point processes, laying the groundwork for continuous-time event modeling. JODIE (Kumar et al., 2019) extends this by introducing a projection mechanism to estimate future node states given elapsed time, enabling real-time, time-aware predictions.

Subsequent models further enhance this paradigm in various ways: Know-Evolve (Trivedi et al., 2017) and RE-Net (Jin et al., 2020) focus on temporal knowledge graphs, combining RNNs with multivariate point processes or structural aggregation for autoregressive link prediction. HierTCN (You et al., 2019) replaces sequential updates with a hybrid RNN-TCN architecture to capture both short- and long-term user dynamics. DynGESN (Micheli & Tortorella, 2022) replaces trainable RNNs with an Echo State Network, offering a lightweight alternative for sequential modeling.

Incorporating structural information, DyGNN (Ma et al., 2020) integrates LSTM updates with GNN-based context under an event-driven scheme. AER-AD (Fang et al., 2023) focuses on inductive anomaly detection by learning anonymous edge representations and modeling temporal edge sequences via GRUs. Finally, RTRGN (Chen et al., 2023f) enhances temporal neighbor aggregation through recurrent state updates and a heterogeneous revision module, improving the completeness and accuracy of dynamic representation.

The memory mechanism in CTDG models is first proposed in the paper of TGN (Rossi et al., 2020). TGN (Rossi et al., 2020) maintains a timestamped memory vector $\mathbf{s}_i(t)$ for each node, which is updated on each interaction event. Upon an event between nodes i and j at time t, TGN computes messages $m_i(t)$ and $m_j(t)$ from their current memory states, the edge attributes, and time intervals. These messages are optionally aggregated (e.g., mean or last event) and passed through an RNN-based memory-updater (e.g., GRU) to produce updated node memories at time t. A node's embedding $\mathbf{z}_i(t)$ is then computed from its memory (and possibly neighbor context) for downstream tasks. This architecture enables long-term historical information retention and online embedding updates with event-level granularity.

The following memory-based methods largely follow the architectural framework established by TGN. These subsequent models retain this core pipeline—maintaining persistent node memories updated upon each interaction—but introduce various enhancements.

NAT (Luo & Li, 2022) introduces a neighborhood cache (N-cache) that stores multi-hop historical context per node, enabling query-time construction of temporally-aware embeddings. GDCF (Han et al., 2023) and CDGP (Zhang et al., 2023f) adopt memory modules for spatio-temporal prediction and community-aware dynamics, respectively, by maintaining evolving node or community states.

Several works tackle memory staleness and adaptability: TIGER (Zhang et al., 2023e) employs dual memory (pre/post-event) and a restarter module to support parallel and context-aware updates; PRES (Su et al., 2024b) uses a GMM-based correction mechanism guided by gradient trends to adjust memory for robust

long-term retention; Ada-DyGNN (Li et al., 2024a) introduces a learned policy for selective neighbor memory updates, enhancing efficiency and relevance.

Other models focus on addressing robustness and personalization: RDGSL (Zhang et al., 2023c) integrates dynamic noise filtering with attention-based edge selection; SEAN (Zhang et al., 2024d) proposes a plug-and-play neighborhood encoder with personalized neighbor selection and temporal aggregation. MemMap (Ji et al., 2024) maps nodes into latent semantic memory cells to encode high-level evolving patterns, while MSPipe (Sheng et al., 2024) decouples memory access from event order using a minimal staleness pipeline and resource-aware scheduling.

These methods highlight the growing trend of memory-centric architectures that move beyond traditional message passing to enable robust, asynchronous, and temporally rich dynamic graph representations.

TPP-based Methods The TPP-based methods directly model event occurrence intensities based on the history of event timestamps and use learned intensity functions or personalized projection operators that condition on past events. The representation of temporal dependencies is encoded statistically through the intensity functions, thus, they do not require extra memory modules.

HTNE (Zuo et al., 2018) and GHN (Han et al., 2020) employ Hawkes processes to model the excitation effects of historical neighbors, capturing inter-event influence in dynamic graphs and TKGs. M^2DNE (Lu et al., 2019) introduces a hierarchical TPP framework for modeling both edge-level (micro) and network-level (macro) dynamics. DyRep (Trivedi et al., 2019) unifies communication and association dynamics into a multiscale TPP model with attention-based structural encoding.

Later models extend expressiveness and interpretability: LDG (Knyazev et al., 2021) focuses on edge durability over time, while TREND (Wen & Fang, 2022) and DynShare (Zhao et al., 2023c) combine individual and collective temporal dynamics via learnable intensity functions or personalized projections. EasyDGL (Chen et al., 2024a) integrates TPP-modulated attention, a principled masking-likelihood training scheme, and graph spectral perturbations to support scalability and global interpretability.

These models highlight the strength of TPP-based frameworks in modeling continuous-time dynamics with high temporal precision, offering strong support for event prediction and temporal reasoning in CTDGs.

Random Walk-based Methods In Random Walk-based techniques, the model constructs representations by sampling time-respecting paths (walks) and aggregating information along these trajectories. This path-centric process captures spatio-temporal motifs without storing explicit node states. The node's representation is computed from sampled walks. As such, these methods avoid the complexity and overhead of maintaining per-node memory.

CTDNE (Nguyen et al., 2018) is a pioneering approach that generates timestamp-consistent walks and applies Skip-Gram for embedding learning. HNIP (Qiu et al., 2020) enhances this with high-order temporal modeling and a guided autoencoder for supervised walk reconstruction.

CAW (Wang et al., 2021c) introduces causal walks and anonymous encoding to support inductive learning and generalization to unseen nodes, achieving a good balance between expressiveness and scalability. NeurTWs (Jin et al., 2022) encodes temporal motifs from biased walks using neural ODEs and contrastive learning, capturing fine-grained temporal regularities.

PINT (Souza et al., 2022) and TPNet (Lu et al., 2024) formalize temporal walk aggregation via principled message-passing frameworks, enabling provable temporal expressivity and scalable pairwise representation learning through time-decayed walk matrices or random feature propagation.

These methods highlight the path-centric perspective in CTDGs, offering strong generalization, interpretability, and compatibility with transductive and inductive settings by leveraging temporal ordering and walk-based encoding.

Attention/Time Encoding-based Methods TGAT (Xu et al., 2020) introduces temporal self-attention with functional time encoding based on Bochner's theorem (Xu et al., 2019), enabling translation-invariant

modeling of time intervals and supporting inductive learning. TCL (Wang et al., 2021a) extends this with a two-stream Transformer and co-attention fusion, trained via contrastive learning to enhance robustness. OTGNet (Feng et al., 2023) and TGRank (Suresh et al., 2023) further extend temporal attention to handle open dynamic graphs and dynamic link ranking, respectively.

Beyond node-level temporal modeling, DHGAS (Ji et al., 2023) applies neural architecture search to discover optimal attention-based architectures for dynamic heterogeneous graphs. DyG2Vec (Alomrani et al., 2023) introduces a window-based attention-driven encoder that captures temporal motifs via relative time encoding and temporal edge features. It also adopts a non-contrastive self-supervised joint-embedding framework to learn task-agnostic representations. SimpleDyG (Wu et al., 2024b) explores the feasibility of directly applying vanilla Transformers by encoding node interaction sequences with temporal alignment tokens. Todyformer (Biparva et al., 2024) introduces a structure-aware tokenization mechanism that preserves holistic structural and temporal dependencies when converting dynamic graphs into Transformer-compatible sequences. DyGFormer (Yu et al., 2023b) enhances long-history modeling through neighbor co-occurrence encoding and a patching mechanism, effectively adapting standard Transformers for dynamic link predictions.

While several attention-based approaches directly model temporal interaction graphs through self-attention layers—effectively capturing temporal dependencies via time encodings and long-range attention—they typically treat each interaction independently, without retaining historical context beyond attention windows. Several works propose to integrate explicit memory mechanisms into Transformer architectures.

APAN (Wang et al., 2021b) introduces a mailbox-based mechanism, where each node asynchronously receives and queues messages triggered by neighbor events, and performs attention-based summarization only when queried. iLoRE (Zhang et al., 2023d) further enhances temporal modeling via a hybrid design: an adaptive short-term updater filters noise, an attention-based long-term updater employs identity-aware attention, and a re-occurrence graph module captures repeated interactions.

Although, attention-based models often employ time encodings to capture temporal ordering, time encoding is not exclusive to the Transformer architecture. A number of methods adopt explicit time encoding strategies within non-Transformer architectures. These models aim to integrate temporal signals into their representations through different time related embeddings.

TDGNN (Qu et al., 2020) applies exponential decay in its temporal aggregator to weight edges by recency. DGEL (Tang et al., 2023) updates embeddings in real time for recommendation, aligning structural evolution with normalization constraints. SUPA (Wu et al., 2023a) introduces time-aware short-term decay, time-sensitive propagation, and incremental learning tailored for multiplex heterogeneous graphs.

To capture temporal patterns more expressively, FreeDyG (Tian et al., 2024) integrates functional time encoding with Fourier-based frequency enhancement for periodicity modeling. CNE-N (Cheng et al., 2024) compresses neighborhoods into hash-based memory and introduces temporal-diverse mechanisms for multi-scale dynamics. TG-Mixer (Zhang et al., 2024c) models interaction burstiness via silence decay and clustering-aware sampling of historical links. DyGMamba (Ding et al., 2025) leverages SSMs to capture long-term temporal dependencies by maintaining compact latent states, thereby enabling efficient and scalable temporal reasoning over long event sequences.

These methods demonstrate the versatility of temporal modeling in dynamic graphs, incorporating decay, frequency, memory compression, and burstiness to address complex, real-world temporal dynamics.

MLP-based Methods Recent years have witnessed a surge in models leveraging complex architectures such as recurrent neural networks (RNNs) and Transformers to capture fine-grained temporal dependencies. While these approaches have achieved strong empirical performance, they often come at the cost of increased model complexity, slower training, and limited interpretability. As a response to this growing complexity, an emerging line of research has begun to question whether such sophisticated mechanisms are truly necessary. This has led to a re-examination of simpler alternatives—such as MLPs—as potentially sufficient for effective temporal representation learning, thereby inspiring methods like GraphMixer that demonstrate competitive performance through conceptually and technically streamlined designs.

GraphMixer (Cong et al., 2023) replaces RNNs and attention with MLP-based link encoders, mean-pooling aggregation, and fixed time encoding, offering a fast and effective alternative. RepeatMixer (Zou et al., 2024) emphasizes repeat behavior modeling by introducing repeat-aware sampling and time-sensitive aggregation across different orders. BandRank (Li et al., 2025b) adopts a frequency-disentangled approach, decomposing temporal signals into multiple bands using adaptive filters, and employs frequency-enhanced MLPs along with a novel Harmonic Ranking loss to stabilize multi-scale supervision.

These models reflect a growing trend toward simpler yet targeted architectures that leverage signal decomposition, behavioral priors, and lightweight designs to achieve efficient and robust dynamic graph representation.

Summary and takeaways. The central contribution of CTDG modeling lies not merely in providing a finer taxonomy of dynamic graph methods, but in adopting a fundamentally different view of temporal behavior. In CTDG modeling, behavior are treated as timestamped events whose order and temporal spacing must be modeled with much greater precision than snapshot-based approaches allow. Consequently, the key design axes of CTDG models shift from backbone architecture alone to temporal modeling strategies and memory mechanisms. Memory-based architectures have become influential due to their ability to support asynchronous updates and long-term state tracking. Meanwhile, TPP-based, walk-based, attention-based, and more recent MLP-based approaches illustrate different trade-offs among temporal precision, scalability, and architectural complexity. Together, these developments position CTDG models as the methodological core of dynamic graph representation learning.

4.3.3 Optimization on Training Frameworks

In contrast to DTDG models, which are typically trained in an end-to-end manner for link prediction task, CTDG models commonly adopt a "pre-train, then fine-tune" training paradigm (Chen et al., 2024e), where a backbone encoder is first trained on temporal interaction data and task-specific predictors are subsequently optimized for different downstream tasks (as illustrated in Fig. 7b). Owing to the fine-grained temporal dependencies and event-driven modeling mechanisms in CTDGs, large-scale training and deployment of CTDG models are generally more complex than their DTDG counterparts. Therefore, in the following, we primarily focus on recent advances in training frameworks tailored for CTDG models.

Beyond fundamental CTDG representation learning methods, new frameworks have emerged to boost both efficiency and effectiveness. These frameworks are generally motivated by two key needs. (1) Scalability across large, evolving graphs: As CTDGs grow and update rapidly, traditional single-GPU or static training pipelines struggle to keep up. Frameworks designed for multi-GPU acceleration address this by parallelizing event processing, optimizing storage structures, and enabling distributed training. (2) Adaptability to downstream tasks under semantic and temporal shifts: CTDGs often feature complex, time-evolving relational patterns that standard pre-training setups do not fully capture. Prompt-based learning frameworks step in here: by freezing backbone parameters and injecting lightweight, task/time-conditioned prompts, they enable models to adapt dynamically across temporal domains and tasks—without heavy retraining.

Frameworks like TGL (Zhou et al., 2022), DistTGL (Zhou et al., 2023), SPEED (Chen et al., 2023e), GNNFlow (Zhong et al., 2023), and Deep-Graph-Sprints (DGS) (Eddin et al., 2024) address the challenge of training on large-scale CTDGs by optimizing system-level bottlenecks. For instance, TGL introduces a time-sorted Temporal-CSR structure, parallel temporal sampling, and random-chunk scheduling to enable efficient neighbor retrieval and mini-batch updates, achieving large speedups on multi-GPU setups. DistTGL extends this to distributed environments, synchronizing node memory across workers to maintain throughput and convergence. SPEED contributes by offering Streaming Edge Partitioning and Parallel Acceleration Component, while GNNFlow layers on adaptive, time-indexed blocks, GPU/CPU hybrid caching, and distributed training support. DGS introduces a recurrent, sampling-free embedding architecture for CTDGs that updates node states online using vectorized forgetting factors, learnable feature embeddings, and mixed-mode automatic differentiation for efficient end-to-end learning. Collectively, these frameworks tackle storage, sampling, scheduling, caching, and parallelism to scale CTDG training reliably.

TIGPrompt (Chen et al., 2024e) innovatively brings the "pre-train, prompt" training paradigm to CTDG by injecting lightweight, time- and task-aware prompts, which is designed to enhance the adaptability of CTDG models by bridging the temporal and semantic gaps between pre-training and downstream tasks. It

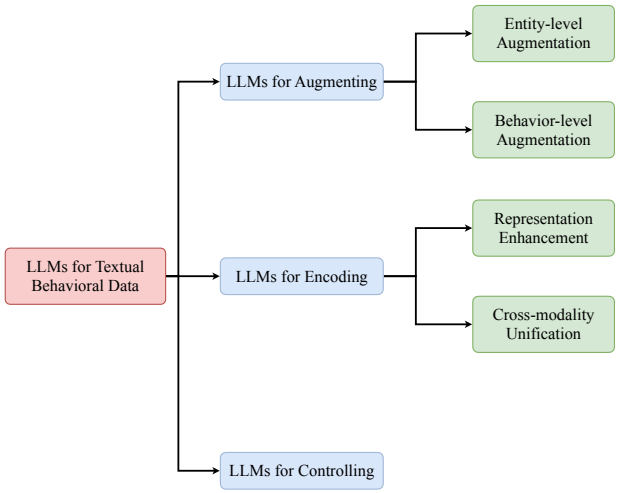


Figure 8: The dissection of the LLM methods for textual behavioral data.

introduces three different temporal prompt generator that produces personalized, time-aware prompts for each node, enabling efficient adaptation without modifying the pre-trained model.

Summary and takeaways. Compared with DTDG models, CTDG learning places stronger demands on both scalable training infrastructure and flexible adaptation strategies. Recent frameworks therefore evolve along two complementary directions: system-level solutions that accelerate event-driven training on large dynamic graphs, and prompt-based approaches that enable efficient task and temporal adaptation of pre-trained CTDG encoders.

4.4 Textual Data

Behavioral data are often expressed in textual form, arising from user-generated content and textual interactions, such as online reviews, chat logs, social media posts, and conversational records. Such textual behavioral data naturally carry rich semantic information about user intent, preferences, and contextual interactions.

LLMs, such as GPT-4o (OpenAI, 2024), Gemini (Gemini Team, 2024), PaLM (Chowdhery et al., 2023), and LLaMA (Touvron et al., 2023), have demonstrated remarkable capabilities in understanding and generating human-readable text, making them particularly well-suited for modeling textual behavioral data. Beyond traditional NLP tasks, their influence has extended to a wide range of deep learning applications, including education (Li et al., 2024c), healthcare (Nov et al., 2023), finance (Wu et al., 2023c), and recommendation systems (Bao et al., 2023).

Consequently, recent research has begun to explore the application of LLMs to behavioral data. By integrating extensive open-world knowledge and emergent abilities such as reasoning, LLMs can effectively analyze entity preferences from textual behavioral traces and enrich the semantic understanding of both entities and behaviors, thereby significantly improving the quality of behavioral data representations. Moreover, LLMs enable modeling more complex behavioral scenarios, such as conversational behaviors (Gao et al., 2023a), explainable behaviors (Chen, 2023), and real-time behaviors (Marripudugala, 2024; Gao et al., 2024a).

Based on the role of LLMs within the pipeline, existing LLM-related methods for behavioral data can be broadly categorized into three groups: **(1) LLMs for Augmenting**, **(2) LLMs for Encoding**, and **(3) LLMs for Controlling**. We will introduce these categories in the following subsections.

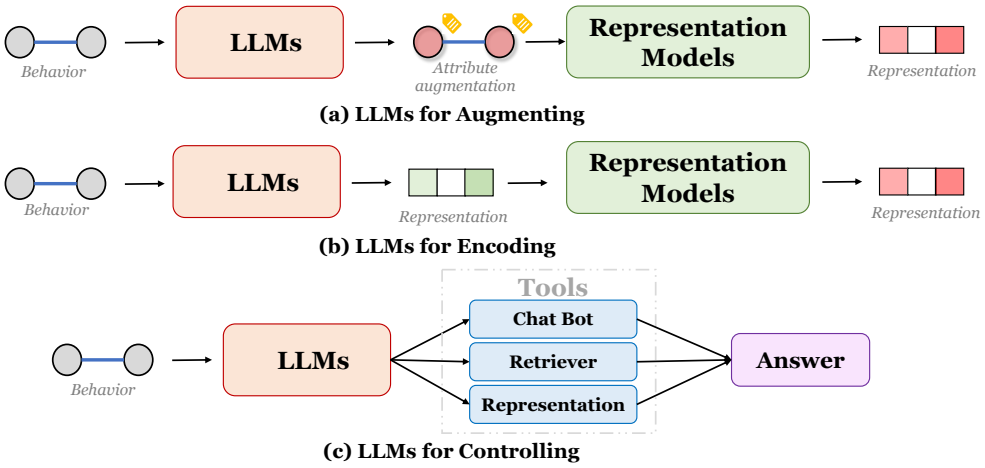


Figure 9: The pipelines of the LLM-related methods for behavioral data in textual form.

4.4.1 LLMs for Augmenting

LLMs for Augmenting refers to the process of selecting, transforming, and enriching raw behavioral data to create more expressive features/behaviors that are more readily understood by deep learning methods like time series models and dynamic graph approaches. In this process, LLMs take the original data (*e.g.*, entity descriptions, profiles, behavioral features, timestamps) as input and generate auxiliary textual data for augmentation, with various objectives such as enriching the features of entities/behaviors or synthesizing more informative behaviors for both training and evaluation. Based on the ways of augmentation mentioned above, it is easy to further classify these methods into two categories: **(a) Feature Augmentation** and **(b) Behavior Augmentation**.

Feature Augmentation Integrated with exceptional reasoning capabilities and vast parameterized knowledge, LLMs can serve as experts to provide external knowledge during training. This enables them to generate auxiliary features that could promote the understanding of entities and their behaviors. For instance, KAR (Xi et al., 2024) leverages LLMs to produce user-side behavioral insights and item-side factual information in recommender systems, which are then incorporated as supplementary features into traditional recommendation methods. SAGCN (Liu et al., 2025) proposes a prompting mechanism to discover semantic, aspect-aware behavioral interactions, providing more fine-grained interpretations of behaviors at the semantic level. In the CUP (Torbati et al., 2023), ChatGPT is used to extract and summarize each user's behavioral patterns into a set of concise keywords derived from their review history. By condensing user behavioral profiles into just 128 tokens, this method enables efficient encoding with smaller language models that have restricted context windows. In addition to the above methods that rely solely on LLM inference for feature augmentation, some methods (Cao et al., 2024a; Ren et al., 2025; Pandey & Singh, 2025) employ instruction fine-tuning with the foundational language models for various tasks across different domains, such as behavior categorization or intent prediction in e-commerce systems. Other methods also leverage LLMs for feature augmentation, including tasks such as text augmentation (Qin et al., 2025), graph completion (Ghanem & Cruz, 2024; 2025), and attribute generation (Regino et al., 2024; Gao et al., 2024c).

Behavior Augmentation Besides feature-level augmentation, LLMs are also good at generating synthetic behaviors, which can be used to empower the training datasets and subsequently boost the quality of representation generation. For example, in Precious2GPT (Sidorenko et al., 2024), generative models can be fine-tuned to produce realistic synthetic behavior logs, enriching the training datasets and improving model robustness. In (Carranza et al., 2024), privacy concerns are also being addressed by training LLMs under differential privacy constraints, enabling the generation of synthetic behavioral traces that protect identities while preserving data utility. In (Chen et al., 2024c), prompt engineering techniques are increasingly used to extract richer behavioral summaries, generate user profiles, and even simulate hypothetical interactions,

thereby providing more nuanced signals for downstream models. In addition, Trisum (Jiang et al., 2024) facilitates the integration of heterogeneous behavioral datasets by aligning features across domains, making it possible to leverage external sources for improvements. In TF-DCon (Wu et al., 2025a), data condensation strategies have also emerged, where LLMs generate compact yet informative behavioral samples, reducing the volume of training data required without sacrificing representativeness. In e-commerce, LLMs are fine-tuned for tasks such as query rewriting (KAG (Liang et al., 2025)), helping to bridge semantic gaps for less common or ambiguous behaviors.

Summary and takeaways. LLM-based augmentation expands behavioral representation learning by enriching both features and interaction signals. Feature-level methods leverage LLMs to inject external knowledge and semantic interpretations into entity and behavioral attributes, while behavior-level methods generate synthetic or summarized interaction traces to enhance data coverage and robustness. Together, these approaches demonstrate how LLMs can act as powerful data transformation interfaces, bridging raw behavioral logs and downstream representation learning models.

4.4.2 LLMs for Encoding

In traditional deep learning methods, behavioral data are often formulated as the digitized features through techniques like one-hot encoding, which subsequently allows for an embedding layer to summarize and transform these features into dense vector representations. On the other hand, LLMs are built upon Transformer-based architectures, which excel at modeling contextual dependencies in textual data within behaviors. Such alignment between LLMs and embedding layers makes it intuitive to utilize LLMs as advanced feature encoders for more effective and context-aware representation generation.

Depending on the encoding strategy, existing methods that use LLMs for encoding can generally be categorized into two types: **(a) Representation Enhancement** methods, which utilize semantic information to further elevate representations; and **(b) Cross-modality Unification** methods, which exploit the generalization capabilities of LLMs to unify behavioral data from different modalities [1], thereby synthesizing more expressive and universal representations. The following sections will describe these two categories in detail.

Representation Enhancement In this group of methods, LLMs are utilized as feature encoders for the available textual information within behavioral data, such as user profiles or item descriptions in recommendation systems (Bao et al., 2023), abstracts or introductions in citation networks (Jiang & Ferrara, 2025), and financial documents in financial systems (Srivastava et al., 2024). These approaches typically leverage LLM-based feature encoding in various applications, including news recommendation (Gao et al., 2024b), document ranking (Lu et al., 2025), tweet search (Chavinda & Thayasivam, 2025), and tag selection (Nazary et al., 2025). Additionally, beyond static textual features above, there also exist methods that utilize LLMs to capture dynamic and evolving semantic information, thereby modeling temporal dynamics and adapting to shifting preferences over time, especially in user-item behavioral data among recommendations. For instance, U-BERT (Qiu et al., 2021) enhances user representations by transforming review texts into sequences of dense vectors using BERT (Devlin et al., 2019), and then applies tailored attention mechanisms to effectively model user interests. Similarly, LLM4ARec (Li et al., 2023b) leverages GPT-2 to extract personalized aspect terms and latent features from user profiles and reviews, thereby improving the quality of recommendations. Unlike the aforementioned methods that directly use the textual features as LLM input, some methods leverage encoded embedding tokens to represent behavioral semantics. For example, LC-Re (Zheng et al., 2024) proposes to integrate collaborative semantic representation information to generate items from the entire item set for recommendation. Similarly, EAGER (Wang et al., 2024h) introduces a two-stream behavior-collaborative architecture that extracts semantic information into embedding tokens, thereby enhancing overall performance in recommendation systems.

Cross-modality Unification In addition to representation improvement methods, the generalization capability of LLMs offers a promising solution for transfer learning and cross-modality unification, where natural language serves as a bridge to align distribution differences across various modalities. For instance,

[1] Different modalities refer to behavioral data with different distributions, such as series/graphs, different scenarios, different datasets, different systems, etc.

OneLLM (Han et al., 2024) proposes a multi-modal LLM that aligns eight types of behavioral data to language within a unified framework, utilizing a multi-modal encoder and a progressive multi-modal alignment pipeline. RUNSRec (Zhang et al., 2025a) comprehends universal user behavioral patterns across different domains but also captures their inherent preferences to make recommendations. SocialMind (Yang et al., 2025a) employs human-like perception leveraging multi-modal sensors to extract both verbal and nonverbal behaviors, social factors, and implicit personas, incorporating these social behaviors into LLM reasoning for social suggestion generation. Brooks et al. (Brooks et al., 2023) present two models available from an API-based suite of emotional expression models that measure nuanced facial and vocal signals, providing rich, high-dimensional emotional expression estimates among human-computer interactions. Uni-CTR (Fu et al., 2025) utilizes a shared LLM to extract semantic representations at multiple layers, effectively capturing shared features across diverse domains. This approach enhances the model's ability to perform multi-domain recommendation by leveraging the underlying commonalities among different domains. Some works in recommendation systems (Tian et al., 2025; Lin et al., 2025) leverage unified cross-domain textual embeddings from a fixed LLM (e.g., Sheared-LLaMA (Touvron et al., 2023)) to tackle scenarios with cold-start users/items or low-frequency long-tail features. CROSS (Zhang et al., 2025b) further explores layerwise mixer to obtain better representations over textual features between the LLM-generated texts and behavior structures.

Summary and takeaways. LLM-based encoding enhances behavioral representation learning by introducing a powerful semantic encoding layer. Representation enhancement methods leverage LLMs to extract contextual and semantic information from textual features, improving entity and interaction embeddings. Cross-modality unification methods further use language as a shared representation space to align heterogeneous behavioral data across domains and modalities. Together, these approaches position LLMs as universal semantic encoders that bridge diverse behavioral signals and downstream representation learning models.

4.4.3 LLMs for Controlling

As the number of model parameters continues to grow, LLMs have demonstrated remarkable *emergent abilities* that smaller models cannot achieve, such as in-context learning, instruction following, self-reflection, and effective tool use. With these capabilities, LLMs can move beyond serving as components within a traditional pipeline. They can take on the role of pipeline controllers or decision-makers across modules in behavioral data representation learning, thereby enhancing interactivity and interpretability of models.

In recommendation systems, conversational agents like Chat-REC (Gao et al., 2023a) and RecLLM (Friedman et al., 2023) leverage LLMs to interact with users in a dialogue format. These models can infer user preferences from purchase histories, determine when to invoke recommendation APIs, and dynamically filter or rerank candidate items, ultimately delivering more accurate and personalized recommendations. Rather than relying on LLMs to manage the entire recommendation workflow, a more granular approach involves breaking down the recommendation task into smaller, manageable components. RecMind (Wang et al., 2024g) processes behavioral data using a self-inspiring prompting strategy and multi-step reasoning, incorporating tools like expert models, SQL databases, and search engines to enrich the analysis and interpretation of user behaviors.

In finance systems, FinCon (Yu et al., 2024) mimics an investment firm's structure with "Manager" and specialist "Analyst" agents collaborating via natural language. A Manager LLM consolidates insights from news, reports, and quantitative analysts, while dual-level risk-control agents monitor market risk and update investment beliefs through self-reflection. As for multi-modal behavioral data, FinAgent (Zhang et al., 2024g) integrates a GPT-based core with specialized tools: it processes news text, stock charts (vision), and APIs for pricing and analytics. Its pipeline has modules for market intelligence (news summarization), trading behavior generation (via Chain-of-Thought), and portfolio execution. These finance-domain systems (Xiao et al., 2024; Ye et al., 2024c; Cao et al., 2024b) position LLMs not as stand-alone predictors but as decision-making cores in modular architectures. They integrate chain-of-thought reasoning, retrieval or vision tools, and multiple specialized agents to tackle complex tasks like trading, forecasting, QA, and fraud detection.

In social networks, LLMs are central agents or modules in simulations of human behavior. They exhibit multi-turn reasoning and context-awareness, utilize tools, and even self-reflect. For example, AgentSociety (Piao et al., 2025) integrates environmental data with LLM agents that plan activities, interact, and communicate,

Table 5: LLM-related methods for behavioral data.

Methods	LLM Function	Modality Transformation	LLM Fine-tuning	Year
LLMs for Augmenting				
Carranza et al. (Carranza et al., 2024)	Behavior augmentation	Language	No	2024
TF-DCon (Wu et al., 2025a)	Behavior augmentation	Language	No	2025
CUP (Torbati et al., 2023)	Feature augmentation	Language	No	2023
Precious2GPT (Sidorenko et al., 2024)	Behavior augmentation	Multi-omics → Language	Yes	2024
TriSum (Jiang et al., 2024)	Behavior augmentation	Sequence → Language	Yes	2024
Chen et al. (Chen et al., 2024c)	Behavior augmentation	Sequence → Language	No	2024
KAR (Xi et al., 2024)	Feature augmentation	Language	No	2024
Ghanem et al. (Ghanem & Cruz, 2024)	Feature augmentation	Graph → Language	Yes	2024
Ghanem et al. (Ghanem & Cruz, 2025)	Feature augmentation	Graph → Language	No	2025
APLe (Cao et al., 2024a)	Feature augmentation	Sequence (multi-modal) → Language	Yes	2024
KAG (Liang et al., 2025)	Behavior augmentation	Language	Yes	2025
SAGCN (Liu et al., 2025)	Feature augmentation	Sequence → Language	No	2025
Pandey et al. (Pandey & Singh, 2025)	Feature augmentation	Sequence → Language	Yes	2025
DyLas (Ren et al., 2025)	Feature augmentation	Language	Yes	2025
LLMs for Encoding				
U-BERT (Qiu et al., 2021)	Representation enhancement	Language	Yes	2021
Social-LLM (Jiang & Ferrara, 2025)	Representation enhancement	Graph → Language	No	2025
Brooks et al. (Brooks et al., 2023)	Cross-modality unification	Sequence → Language	No	2023
UFIN (Tian et al., 2025)	Cross-modality unification	Sequence → Language	No	2025
Uni-CTR (Fu et al., 2025)	Cross-modality unification	Sequence → Language	No	2025
GNR (Gao et al., 2024b)	Representation enhancement	Language	Yes	2024
EAGER (Wang et al., 2024h)	Representation enhancement	Sequence → Language	Yes	2024
LC-Re (Zheng et al., 2024)	Representation enhancement	Language	Yes	2024
OneLLM (Han et al., 2024)	Cross-modality unification	Graph/Sequence → Language	Yes	2024
Lu et al. (Lu et al., 2025)	Representation enhancement	Language	Yes	2025
Chavinda et al. (Chavinda & Thayasivam, 2025)	Representation enhancement	Language	Yes	2025
Poison-RAG (Nazary et al., 2025)	Representation enhancement	Sequence → Language	No	2025
RUNSRec (Zhang et al., 2025a)	Cross-modality unification	Sequence → Language	Yes	2025
SocialMind (Yang et al., 2025a)	Cross-modality unification	Graph/Sequence → Language	No	2025
CROSS (Zhang et al., 2025b)	Cross-modality unification	Graph → Language	Yes	2025
Lin et al. (Lin et al., 2025)	Cross-modality unification	Sequence → Language	No	2025
LLMs for Controlling				
RecLLM (Friedman et al., 2023)	Pipeline control	Language	No	2023
RecMind (Wang et al., 2024g)	Pipeline control	Language	No	2024
FinCon (Yu et al., 2024)	Pipeline control	Language	Yes	2024
FinAgent (Zhang et al., 2024g)	Pipeline control	Language	Yes	2024
TradingAgents (Xiao et al., 2024)	Pipeline control	Language	Yes	2024
TS-Reasoner (Ye et al., 2024c)	Pipeline control	Language	Yes	2024
RiskLabs (Cao et al., 2024b)	Pipeline control	Language	Yes	2024
AgentSociety (Piao et al., 2025)	Pipeline control	Language	Yes	2025
ProSim (Zhou et al., 2025c)	Pipeline control	Language	Yes	2025

where LLMs function as autonomous social actors whose emergent behaviors shed light on community phenomena. ProSim (Zhou et al., 2025c) is a modular framework for simulating prosocial behavior, where LLM agents autonomously decide whether to cooperate, punish unfairness, etc. It shows agents naturally exhibit human-like prosocial actions and adjust to policy changes. This illustrates LLMs as modular social actors capable of nuanced, multi-step social reasoning. The research suggests that LLMs can mirror collective dynamics across modules through techniques such as cooperation, tool usage, or reasoning capabilities, making them powerful for controlling the model pipeline in behavioral data processing.

Summary and takeaways. LLM-based controlling frameworks elevate language models from individual modeling components to orchestration layers in behavioral learning systems. By leveraging reasoning, tool usage, and multi-agent coordination, LLMs can dynamically manage task decomposition, module interaction, and decision-making across complex pipelines. This paradigm demonstrates how LLMs can integrate heterogeneous models and behavioral signals into interactive and interpretable systems that support more flexible behavioral analysis and decision processes.

5 Application Scenarios

Having reviewed the foundational principles and empirical characteristics of major behavioral data representation methods, we now shift focus to their practical applicability across real-world tasks. This section examines the application of various data representations—including tabular, event sequence, dynamic graph, and textual—to key application scenarios across five prominent domains: **E-commerce**, **Healthcare**, **Social Media**, **Finance**, and **Gaming**.

5.1 E-commerce

In e-commerce, user interactions generate large-scale behavioral data, including clickstreams, search queries, and purchase records. Effectively learning representations from these data is crucial for powering intelligent downstream applications. We will now explore the application of behavioral data representations in two pivotal application scenarios in e-commerce: **(1) Personalized Search and Recommendation**, **(2) Enhanced User Understanding and Profiling**.

5.1.1 Personalized Search and Recommendation

As the core of e-commerce platforms, personalized search and recommendation aim to match users with the most relevant products by leveraging their behavioral data and contextual signals. Tabular data has long served as the backbone in e-commerce. Early studies leveraged clickstream logs and transaction tables to build aggregate user profiles for personalization (Mobasher et al., 2000). With the growth of large-scale e-commerce platforms, structured tabular features have become central to collaborative filtering (e.g., NCF (He et al., 2017)) and hybrid deep learning models for recommendation (e.g., DeepFM (Guo et al., 2017), xDeepFM (Lian et al., 2018), AutoInt (Song et al., 2019)).

By modeling a user's behavior as a time-stamped list of events, event sequence modeling has become a cornerstone of personalized recommendation. The foundation was laid by methods combining traditional matrix factorization with sequential signals, such as FPMC (Rendle et al., 2010), which integrated first-order Markov chains to capture dependencies on the user's last action. The paradigm shifted decisively with the introduction of deep learning. GRU4Rec (Hidasi et al., 2016) was a seminal work that first applied RNNs to model session-based clickstreams, demonstrating the power of learning complex sequential patterns.

To overcome the limitations of RNNs in capturing long-range dependencies, the field then embraced attention mechanisms, leading to the rise of attention-based models. SASRec (Kang & McAuley, 2018), for instance, pioneered the use of a unidirectional Transformer, allowing the model to weigh the importance of all past items to make more context-aware predictions. Building on this architectural foundation, further research has explored innovations in other dimensions. One key direction is enhancing item representations. MARN (Li et al., 2020b), for example, disentangles item features into modality-specific and modality-invariant components to create a richer representation before feeding it into a sequential model. Similarly, Moreira et al. (Moreira et al., 2021) combine tabular data with textual and image vectors for multi-modal feature learning within a Transformer ensemble. Another direction focuses on optimizing the sequence structure itself. SAGE (Wang et al., 2023b), inspired by DNA expression, organizes user behavior into critical "snapshots" to preserve essential signals while reducing sequence length and computational overhead. Recently, the fusion of tabular data with event sequences has become a popular paradigm in modern e-commerce recommendation. For example, IU4Rec (Wu et al., 2025b) leverages tabular product attributes to construct stable "Interest Units" (IUs) and jointly models fine-grained item sequences with higher-level IU sequences for improved recommendation.

However, representing user-item interactions purely as sequences can overlook latent structural connections that evolve over time. To address this, dynamic graph modeling has been introduced to capture both temporal and structural patterns. There are two key applications of dynamic graph modeling in personalized search and recommendation: First, edge prediction. DGSR (Zhang et al., 2023b) connects different user sequences through a dynamic graph structure, and converts the next-item prediction task in sequential recommendation into a link prediction between the user node and the item node. While existing Temporal Graph Neural Networks (T-GNNs) often suffer from scalability and efficiency challenges due to high compu-

tational overhead, to address this limitation, a lightweight framework is proposed, named EAGLE (Li et al., 2025a), which integrates short-term temporal recency and long-term global structural patterns. Second, recommendation and ranking. SRJGraph (Zhao et al., 2022) jointly models search and recommendation with a unified graph neural network to improve ranking performance by alleviating user behavior sparsity.

Nowadays, representing behavioral data with LLMs has initiated a paradigm shift, reconceptualizing user-item interaction sequences as a semantic language of consumer intent. This approach enables a more nuanced, context-aware understanding of user journeys, moving beyond traditional ID-based methods. This paradigm has evolved significantly: Initial work, such as BERT4Rec (Sun et al., 2019b), adapted the masked language model objective from BERT to behavioral data, demonstrating the power of deep bidirectional context. The modern approach, however, directly leverages generative LLMs. A seminal work, P5 (Geng et al., 2022), established a text-to-text framework using prompts to unify diverse recommendation tasks (e.g., sequential prediction, rating, explanation) within a single T5-based model. Following this, the field of LLM4Rec (Wu et al., 2023b) has burgeoned, exploring the use of powerful foundation models. Chat-REC (Gao et al., 2023a), which demonstrates how models like ChatGPT can be effectively prompted and guided to act as high-quality conversational recommender agents with minimal fine-tuning. Furthermore, methods like M6-Rec (Cui et al., 2022) and V-RECS (Podo et al., 2024) integrate multi-modal information (textual descriptions, images) into the sequence, creating richer item representations that allow the LLM to capture complex semantic and visual preferences, thus mitigating cold-start issues and improving recommendation quality.

5.1.2 Enhanced User Understanding and Profiling

In e-commerce, enhancing user understanding through their data is key to personalization. Initially, this was achieved using tabular data processed through extensive feature engineering. This paradigm was dominant in critical industrial tasks such as Click-Through Rate (CTR) prediction and customer modeling, often employing models like logistic regression or factorization machines on massive, sparse feature sets (McMahan et al., 2013; Rendle, 2010; He et al., 2014). For example, the Repeat Buyer Prediction model (Liu et al., 2016) perfectly illustrates this approach by constructing static profiles for users and merchants from transactional statistics to identify potential loyal customers.

To overcome the limitations of static profiles, later research began leveraging deep learning to model user behavior sequences as event sequences. DUPN (Ni et al., 2018) demonstrated that a universal user representation could be learned from these sequences and shared across multiple tasks like search and recommendation. Further advancing this direction, HUP (Gu et al., 2020) introduced a hierarchical profiling method, which captures users' multi-granularity interests by modeling not only item-level sequences but also fine-grained micro-behaviors.

In dynamic graph modeling, enhanced user understanding and profiling is commonly formulated as a node classification problem. For instance, by constructing a tripartite graph connecting users, products, and personas, TriPer (Gupta et al., 2025) reformulates the multi-label persona identification task into a link prediction problem, allowing the model to generalize to new personas not seen during training.

Besides, LLMs excel at generating deep, qualitative insights into user behavior, creating rich, human-readable user profiles from raw interaction logs. This moves analysis from "what" a user will do next to "who" the user is. This paradigm is exemplified by architectures like the Language-based Factorization Model (LFM) (Zhou et al., 2024), which employs an LLM encoder-decoder to distill a user's interaction history into a compact, natural-language summary, thereby enhancing interpretability and addressing cold-start challenges. To refine this process, the GPG (Zhang, 2024) method demonstrates that adding a crucial intermediate summarization step is vital; this guides the LLM to effectively parse sparse data and extract salient personal features, significantly boosting personalization accuracy. Furthermore, this approach has evolved to capture the dynamic nature of user preferences. For instance, the framework from Sabouri et al. (Sabouri et al., 2025), explicitly models temporal dynamics by generating distinct profiles for a user's short-term versus long-term interests, offering enhanced explainability. Complementing this, the PURE (Bang & Song, 2025) framework focuses on the continuous evolution of profiles by systematically extracting information from new user-generated text, such as reviews, to maintain an up-to-date understanding of the user. Collectively, these

innovations in generating, refining, and dynamically managing textual profiles are paving the way for e-commerce systems that are not only more accurate but also fundamentally more transparent and adaptive.

5.2 Healthcare

Hospitals and clinics record a continuous flow of patient information, such as admissions, diagnoses, lab results, and treatments. These clinical events, each characterized by a timestamp and metadata, form a patient's Electronic Health Record (EHR). Therefore, this section is constructed around two primary categories: **(1) Disease Prediction and Risk Stratification**, **(2) Personalized Treatment Recommendation**.

5.2.1 Disease Prediction and Risk Stratification

Disease prediction and risk stratification aim to forecast future health events—such as the onset of sepsis, ICU transfer, or hospital readmission—based on a patient's historical clinical behaviors. Behavioral representation learning, particularly from event sequences and tabular EHR data, has enabled significant advancements in this field. For example, models like RETAIN (Choi et al., 2016a) and more recent attention-based frameworks (e.g., BEHRT (Li et al., 2020c)) learn temporal dependencies from sequences of visits, diagnoses, and medications to produce informative patient embeddings. ProtoEHR (Cai et al., 2025) further advances event sequence modeling by introducing a hierarchical prototype-based representation framework that captures relationships across medical codes, visits, and patients.

In addition, dynamic graph-based representations are increasingly used to model evolving patient states. MHGRL (Liu et al., 2024b) utilizes heterogeneous graphs to model patient data, integrating multi-modal attributes and structural relationships for enhanced clinical prediction tasks such as disease prediction and EHR clustering.

Recently, the application of LLMs has expanded this field beyond direct sequence modeling into more interactive and reasoning-driven frameworks. EHRAgent (Shi et al., 2024) is an LLM-powered agent designed for natural language interaction with EHR systems, capable of multi-tabular reasoning and autonomous code execution to retrieve complex patient information.

5.2.2 Personalized Treatment Recommendation

Treatment recommendation systems rely heavily on accurately modeling patient behaviors and responses to previous interventions.

Tabular EHR data, which typically include demographic profiles, laboratory test results, and vital signs, remain a key source for personalized treatment modeling. For example, DeepSurv (Katzman et al., 2018) extends the classical Cox proportional hazards model with deep neural networks to capture nonlinear interactions between patient covariates and treatment effectiveness, thereby enabling individualized survival-based treatment recommendations.

Patient care histories are naturally represented as clinical event sequences, capturing diagnoses, medications, and procedures over time. To address the patient-specific variability inherent in such sequences, Lee and Hauskrecht (Lee & Hauskrecht, 2023) proposed adaptive event sequence prediction models that refine population-wide models into subpopulation- and patient-specific dynamics, enabling more accurate and individualized clinical predictions. Building on this line, SRL-RNN (Wang et al., 2018) introduces a supervised reinforcement learning framework with recurrent neural networks that leverages both clinician prescription signals and long-term survival rewards from EHR sequences, achieving more effective and safer dynamic treatment recommendations.

Besides, knowledge graphs have also become a critical component in enhancing treatment reasoning. KARE (Jiang et al., 2025) integrates LLMs with community-level knowledge graph retrieval to provide reasoning-enhanced predictions in clinical settings. GAME (Zhou et al., 2025a) is a graph-augmented representation learning framework that harmonizes multi-institutional EHR data through knowledge graph integration and federated learning, enabling robust cross-site clinical studies while preserving data privacy.

Finally, LLMs are increasingly applied in an end-to-end fashion. Recent advancements in LLMs, such as Med-PaLM 2 (Singhal et al., 2025), have significantly improved medical question answering, outperforming human physicians in certain clinical tasks and demonstrating potential for real-world medical applications.

5.3 Social Media

The social media domain contains multi-modal, high-velocity behavioral traces (i.e., posts, likes, shares, comments, follows, clicks). To make these signals actionable, researchers have applied diverse representation paradigms. Below we summarize how they are used across key application scenarios and categorize them into four classes: **(1) User Profiling and Interest Modeling**, **(2) Content Recommendation and Feed Ranking**, **(3) Misinformation and Rumor Detection**, and **(4)Toxicity and Safety Moderation**.

5.3.1 User Profiling and Interest Modeling

In social media platforms, user profiling and interest modeling aim to infer latent characteristics—such as personality traits, preferences, and evolving interests—from rich behavioral traces. Natural language–based approaches focus on extracting signals directly from user-generated text. For instance, Peters demonstrated that GPT-3.5/4 can predict Big Five personality traits from Facebook status updates in a zero-shot setting, achieving moderate correlations with self-reports (Peters & Matz, 2024). More recently, Sabouri et al. (Sabouri et al., 2025) introduced LLM-TUP, which summarizes user histories into natural-language descriptions and encodes them with pre-trained language models, effectively capturing both short- and long-term preferences. Graph-based representations, by contrast, leverage the relational context of users. GREENER (Panayotov et al., 2022) modeled news outlets as nodes in a user–media bipartite graph, applying GNNs to predict factuality and bias, while HL-MRFs (Farnadi et al., 2020) integrated textual content, profile images, and social ties through a structured relational model to jointly infer user demographics and traits. Sequential modeling provides yet another perspective by emphasizing temporal dynamics of user behavior. Liu et al. (Liu et al., 2019) proposed SA-LSTM to jointly model user interaction sequences and the influence of friends, while Yang et al. (Yang et al., 2025b) used Bi-LSTMs with a specialized optimizer to capture temporal dependencies in ad click streams for interest prediction.

5.3.2 Content Recommendation and Feed Ranking

In social platforms, content recommendation and feed ranking aim to surface the most relevant items from massive content streams by learning expressive user and item representations. Natural language–based methods leverage the semantic richness of text to enhance recommendation. For example, SPAR (Zhang et al., 2024a) uses pre-trained language models to encode user histories and candidate item text, introducing sparse poly-attention to effectively handle long engagement sequences and improve personalization. Sequential modeling approaches emphasize temporal dynamics in user behavior. Yang (Yang et al., 2024a) augments user interaction sequences with latent item relations mined from language models, thereby enriching sequential dependencies beyond simple co-occurrence. Meanwhile, Mamba4Rec (Liu et al., 2024a) replaces attention-heavy Transformers with selective state space models to capture long-term dependencies more efficiently, achieving both scalability and accuracy on large datasets. Graph-based approaches bring a structural perspective. ContextGNN (Yuan et al., 2025) integrates pair-wise subgraph representations with global two-tower embeddings, addressing the pair-agnostic limitation of traditional architectures. Similarly, CGAT (Yang et al., 2020b) exploits item knowledge graphs with attention over local and non-local neighbors, enhancing item embeddings and ultimately boosting recommendation accuracy.

5.3.3 Misinformation and Rumor Detection

Detecting misinformation and rumors on social media is a critical task, requiring models that can capture both the semantic content of claims and the structural and temporal dynamics of their spread. Natural language–based approaches exploit textual cues to directly judge veracity. Chen et al. (Chen et al., 2024b) conducted a broad empirical study showing that large language models can match strong baselines in content-only classification, while Ruiz et al. (Ruiz-Dolz & Lawrence, 2025) went further by integrating argumentation schemes and critical question answering to provide interpretable justifications for misinformation detection.

Graph-based methods instead emphasize propagation and relational signals. For example, Liu et al. (Liu et al., 2024c) jointly modeled user correlations and diffusion trees, while GSMA (Ma et al., 2024a) improved GraphSAGE by weighting neighbors according to their structural positions in propagation graphs, achieving stronger accuracy on Weibo datasets. Sequential modeling brings temporal order into focus: Dong et al. (Dong & Qian, 2022) captured word- and sentence-level dependencies to improve detection with limited labels, and GETAE (Truică et al., 2024) combined Bi-RNN text encoders with propagation dynamics for robust ensemble predictions. Hybrid approaches such as MAGIC (Xu, 2024) further integrate multi-modal content and adaptive graph attention to capture both textual semantics and relational context.

5.3.4 Toxicity and Safety Moderation

Ensuring safe and respectful online communities requires robust moderation systems capable of detecting toxic or harmful behavior across diverse contexts. Natural language–based approaches remain the backbone of this task: lightweight yet effective transformer models such as Tiny-toxic-detector (Kamphuis, 2024) achieve competitive accuracy with minimal resources, while recent survey (Shahi & Majchrzak, 2025) synthesize progress across multilingual datasets, highlighting the challenges of implicit language and cultural variation. Graph-enhanced models enrich textual detection with structured knowledge or relational context. MetaTox (Zhao et al., 2025) integrates a meta-toxic knowledge graph with LLMs to provide external toxic concept grounding, while ShieldVLM (Cui et al., 2025) extends moderation to multi-modal implicit toxicity, using deliberative reasoning across vision–language graphs to capture harmful meanings hidden in text–image combinations. Sequential methods bring temporal and contextual signals into play: Cheng et al. (Cheng et al., 2022) formulates moderation as a sequential decision problem to reduce bias in detection over comment streams, and Oyetayo et al. (Oyetayo et al., 2024) leverages dialogue context with RNN-based encoders to classify multiple types of toxicity simultaneously. Hybrid multi-modal frameworks further expand the scope, such as MHSDF (Prabhu & Seethalakshmi, 2025), which fuses textual, visual, and audio/video modalities through attention mechanisms to detect complex hate speech embedded in memes or multi-modal content.

5.4 Finance

The data in the financial sector are inherently heterogeneous, temporally structured, and often high-stakes. Different behavioral data representations are prominent in specific financial applications. For risk management and fraud detection, event sequences, dynamic graphs, and LLMs are the primary modalities. In contrast, stock forecasting and portfolio optimization frequently rely on tabular data and LLMs. Consequently, we structure our review of these applications around two critical domains: **(1) Risk Management and Fraud Detection**, **(2) Stock Forecasting and Portfolio Optimization**.

5.4.1 Risk Management and Fraud Detection

In financial services, user behavioral event sequences have gained increasing attention in recent years for applications in risk management and fraud detection. These sequences effectively capture the distinctive digital behavioral signature of individual users, which is inherently more difficult to forge or manipulate compared to static profile data. By learning expressive representations from these event streams, models can capture subtle temporal patterns and behavioral deviations, thereby offering enhanced capabilities for proactive risk assessment and anomaly detection. Since SimpleRNN networks are prone to the vanishing gradient problem, Xavier et al. (Xavier et al., 2024) find that GRU outperforms both SimpleRNN and LSTM models. Branco et al. (Branco et al., 2020) treats user behavior as a sequence composed of many interleaved, unbounded sub-sequences, and use GRUs to build fraud detection models. FinDeepBehaviorCluster (Min et al., 2021) combines RNNs with intuitive features generated by risk experts to form a hybrid feature representation, which models the sequence of user clicks for fraud detection.UB-PTM (Liu et al., 2022a) designs three agent tasks at different granularities, such as action, intention, and sequence levels, to detect fraud activities from unlabeled data. CLeAR (Huang et al., 2024b) devise an Intensity-Aware Transformer to capture dramatic changes in the intensity of transaction amounts and timing, while design two representation learning tasks that utilizes varying levels of contrastive learning to improve representation robustness.

Node representation learning on discrete dynamic graphs divides data into static snapshots at discrete time intervals, then each snapshot is processed with graph network models. Researches use time-related models to get the user (node) representations. DyHDGE (Wang et al., 2024f) jointly captures temporal transaction dependencies and dynamic heterogeneous graph structures to enable accurate real-time fraud detection in financial scenarios via integrated risk assessment and contrastive learning. EvolveGCN (Pareja et al., 2020) adapts the GCN model over time without relying on node embeddings. Instead, it captures the evolving nature of graph sequences by using a RNN to update the GCN parameters dynamically. FiFrauD (Khodabandehlou & Golpayegani, 2024) turns these real-time transactions into a stream of graphs, utilizing density signals in graphs in unsupervised training paradigm. Unlike discrete dynamic graphs, continuous dynamic graphs capture changes in nodes and edges as they occur continuously over time, rather than at discrete intervals. GADY (Lou et al., 2023) proposes a continuous dynamic graph model to capture complex time information, and uses a message-passing framework integrated with positional features to generate edge embeddings, which are subsequently decoded to detect anomalies. EL^2-DGAD (Chen et al., 2025a) integrates continuous-time embeddings into an attention-based graph encoder model. SSH-T^3 (Gu et al., 2025) then uses a self-supervised hierarchical two-tower Transformer framework that models long, multi-scenario payment behavior sequences via robust day-level pre-training and amount-aware scenario modeling to significantly improve financial risk assessment under scarce labels and complex real-world settings.

LLMs are fundamentally reshaping risk management and fraud detection, moving beyond static rules to analyze dynamic, context-rich data. They are notably applied to detect scams and fraud by analyzing textual and conversational data, where models like GPT-3.5 have proven effective at identifying phishing and other deceptive language (Jiang, 2024), leading to higher precision and recall than traditional methods (Korkanti, 2024). This capability is exemplified in advanced real-time systems, such as a RAG framework that analyzes phone calls to prevent impersonation and solicitation fraud as they occur (Singh et al., 2025). Beyond immediate threats, LLMs enable more comprehensive financial risk assessment by integrating diverse, multimodal data sources. The RiskLabs framework pioneers this approach by fusing textual and vocal data from earnings calls with market time-series data to forecast financial volatility and variance, demonstrating a more holistic risk prediction capability (Cao et al., 2024b). Collectively, these applications showcase a shift towards a more predictive and integrative approach to managing financial risk.

5.4.2 Stock Forecasting and Portfolio Optimization

In financial modeling, tabular data are advancing stock forecasting and portfolio optimization by leveraging heterogeneous numerical and categorical features to support accurate predictions and interpretable investment strategies.

For stock price forecasting, a common approach is to treat tabular data as a time series. Within this paradigm, the LSTM network has been widely adopted due to its ability to capture long-range dependencies and address the vanishing gradient problem (Ta et al., 2020; Han & Fu, 2023; Wang et al., 2024a). Building on this, recent works have explored the Transformer model, which leverages the attention mechanism not only for long-range dependencies but also to explicitly model complex interactions across multiple assets (Lezmi & Xu, 2023). In contrast to these sequence-focused methods, another line of research modeling the problem by leveraging the inherent attribute of tabular data. For instance, Zhang et al. (Zhang et al., 2024e) propose a hybrid model combining a deep neural network with TabNet, a model specifically designed for tabular data, which achieves instance-wise feature selection through its unique attention mechanism. In the context of portfolio optimization, the dominant paradigm is a two-stage, "predict-then-optimize" framework. In this approach, the primary challenge is framed as a time-series forecasting problem as mentioned above, these forecasts are fed into classical optimization algorithms, most commonly the Mean-Variance Optimization (MVO) model, to determine optimal asset allocation (Ta et al., 2020; Wysocki & Sakowski, 2022; Singh et al., 2023). In addition, some end-to-end methods have also been proposed. SARL (Ye et al., 2020) employs a reinforcement learning agent to make end-to-end portfolio decisions, it explicitly addresses the challenge of information-rich environments by augmenting the agent's state representation to tackle the challenges of data heterogeneity and market non-stationarity.

At the same time, LLMs are revolutionizing stock forecasting and portfolio optimization by synthesizing diverse data sources for sophisticated investment strategies. A key innovation is their ability to enhance

predictive models by integrating unstructured text with traditional market data. For instance, the Stock-Time (Wang et al., 2024c) architecture treats stock price series as a language sequence for prediction, while LED-GNN (Xu et al., 2025) uses an LLM to extract inter-stock relationships from news to build dynamic graphs, improving trading volume forecasts. Beyond pure prediction, LLMs are enabling more transparent and customized investment frameworks. The SEP (Koa et al., 2024) framework trains an LLM to generate human-readable explanations for its stock predictions through a self-reflective process, addressing the "black box" problem. Concurrently, research on Persona-Based Ensembles (Abe et al., 2024) leverages LLMs to simulate different institutional investor personas, creating adaptive strategies that outperform traditional methods in specific market conditions. These advancements collectively push towards more accurate, explainable, and personalized portfolio management.

5.5 Gaming

The gaming domain provides an exceptionally rich environment for behavioral data analysis due to the high volume, velocity, and complexity of player interactions. To decipher player intent, skill, and engagement from these intricate data streams, researchers have developed diverse representation learning approaches, which have been applied to a range of application scenarios. In this section, we review representative tasks and highlight how different representation paradigms—such as event-sequence modeling and natural language–based approaches—are leveraged within them. We organize the discussion from three perspectives: **(1) Player Goal and Strategy Recognition**, **(2) Non-Player Character (NPC) Generation and Control**, and **(3) Toxicity Detection and Community Management**.

5.5.1 Player Goal and Strategy Recognition

Understanding player intent and strategy is a foundational task in game analytics. Sequence-based models have shown strong effectiveness here: early work used RNNs and LSTMs to frame goal recognition as sequence labeling, significantly improving prediction performance (Min et al., 2016). In the Real-Time Strategy (RTS) game domain, unsupervised replay embeddings learned from event sequences enabled automatic discovery and labeling of player strategies without manual annotation (Kantharaju & Ontañón, 2020). More recently, attention-based methods such as player2vec (Wang et al., 2024d) have tokenized in-game events and used self-supervised learning to capture long-range dependencies, offering robust player representations for intent recognition. Moreover, innovations like Masking by Token Confidence (MTC) further refined pre-training by adaptively masking task-relevant actions, improving recognition accuracy in Massively Multiplayer Online Role-Playing Game (MMORPGs) (Wang et al., 2024e).

5.5.2 NPC Generation and Control

A second key task is generating believable and adaptive NPCs. Sequence learning approaches demonstrated early promise by using LSTMs with attention mechanisms to generate human-like action sequences, enhancing NPC realism and engagement (Zhao et al., 2021). With the advent of LLMs, this direction has expanded substantially: LLMs now serve as the reasoning "brains" of game agents. For example, MineDojo provides an internet-scale knowledge base to support embodied agents trained via LLMs (Fan et al., 2022), and Voyager showcases lifelong skill acquisition through automatic curriculum generation (Wang et al., 2024b). Other frameworks integrate LLMs with reinforcement learning, where the LLM constructs a skill graph and plan that the RL agent executes to solve long-horizon tasks (Yuan et al., 2023). Beyond planning, code-tuned LLMs empower NPCs to not only generate free-form dialogue but also produce executable, game-state-altering code, enabling deeply interactive experiences (Volum et al., 2022).

5.5.3 Toxicity Detection and Community Management

Ensuring healthy in-game communities is another important application scenario. Early linguistic analyses characterized the communication patterns of toxic players (Kwak & Blackburn, 2015), measured the prevalence of disruptive behaviors (Aguerri et al., 2023), and established links between toxicity and game outcomes (Märtens et al., 2015). Modern approaches extend this line by tackling real-time, multi-modal toxicity detection challenges (Ng et al., 2024). Here, LLM-based language representations have been especially

valuable in understanding complex chat dynamics. Furthermore, reinforcement learning techniques, such as contextual bandits, have been introduced to allocate monitoring resources efficiently toward sessions with the highest predicted toxicity, improving both detection efficiency and player safety (Morrier et al., 2025).

5.6 Summary and Takeaways

A recurring pattern across applications is that the most effective representation is usually determined by which aspect of behavior is most decision-critical. Tabular representations remain competitive when the signal is largely aggregated and the prediction target is stable; event sequences dominate when order and recency matter; dynamic graphs become advantageous once interactions among entities are themselves predictive; and textual or LLM-based representations are most useful when latent intent, semantics, or explanation are central. This suggests that modality choice should be treated as a modeling decision rather than a data-format convenience: the same behavioral system may reasonably be represented in multiple ways depending on whether the task is prediction, retrieval, reasoning, or intervention.

6 Cross-Modal Synthesis and Comparative Analysis

The preceding sections reviewed behavioral data representation learning methods organized by data modality and their downstream applications. While a modality-based taxonomy enables focused comparisons within each paradigm, it can also obscure insights that emerge only when methods are juxtaposed across modalities. In this section we provide an integrated cross-modal analysis. We **(1) examine shared technical building blocks**, **(2) compare the core assumptions and trade-offs of each modality**, and **(3) offer practical guidance on when to prefer one method over another given specific application requirements**. Table 6 provides a consolidated overview.

6.1 Shared Technical Building Blocks

Despite apparent differences in data structure, representation learning methods across modalities draw on a surprisingly convergent set of technical components. Identifying these shared building blocks clarifies both the deep connections among modalities and the modality-specific adaptations that differentiate them.

6.1.1 Attention Mechanisms and Transformers

Self-attention has become a near-universal backbone. In tabular learning, FT-Transformer (Gorishniy et al., 2021) and SAINT (Somepalli et al., 2021) treat columns as tokens and model inter-feature dependencies via self-attention. In event sequences, SASRec (Kang & McAuley, 2018) and BERT4Rec (Sun et al., 2019b) apply unidirectional or bidirectional attention over historical item interactions. In dynamic graphs, TGAT (Xu et al., 2020) and DyGFormer (Yu et al., 2023b) use temporal self-attention to aggregate neighbor interactions, while DySAT (Sankar et al., 2020) and DTFormer (Chen et al., 2024d) employ attention jointly over structural and temporal dimensions. In textual behavioral modeling, LLMs such as GPT (OpenAI, 2023) and BERT (Devlin et al., 2019) are inherently attention-based. The common motivation is that attention can capture long-range dependencies without the sequential bottleneck of recurrent models; however, each modality adapts the attention formulation differently: tabular methods attend across features within a single sample, sequential methods attend over time-ordered events, graph methods attend over temporally filtered neighborhoods, and textual methods attend over token sequences enriched with behavioral semantics.

6.1.2 Self-supervised Pre-training

Pre-training with self-supervised objectives is another cross-cutting theme that spans all four modalities, reflecting a shared philosophical commitment: learning general-purpose behavioral representations from the structure of the data itself before adapting to specific downstream tasks.

The dominant strategy varies by modality but follows analogous logic. Masking and reconstruction (inspired by BERT's MLM) appears in tabular learning (TabTransformer (Huang et al., 2020a), VIME (Yoon et al., 2020)), event sequences (BERT4Rec (Sun et al., 2019b), BEHRT (Li et al., 2020c)), and dynamic graphs

Table 6: Cross-modal comparison of behavioral data representation learning paradigms.

Dimension	Tabular Data	Event Sequences	Dynamic Graphs	Textual Data
Entity representation	Constructed from aggregated static features across behavioral attributes	Derived from encoding a chronologically ordered sequence of timestamped interactions	Learned from evolving relational structure and temporal interaction context	Derived from semantic encoding of behavioral text via pre-trained language models
Core assumption	Behavior is captured by aggregated features; temporal and relational structure is secondary	Chronological order and timing of events are the primary behavioral signals	Inter-entity relationships and their temporal evolution are the dominant predictive signal	Behavioral semantics can be expressed or enriched through natural language
Information preserved	Feature values, cross-feature interactions	Event order, inter-event intervals, sequential dependencies	Structural topology, relational dynamics, multi-hop context	Semantic meaning, intent, sentiment, world knowledge
Information lost	Temporal ordering, relational structure, fine-grained dynamics	Multi-entity relational context, global structural patterns	Non-relational entity attributes, aggregated statistics	Numerical precision, exact temporal intervals, structural topology
Typical supervision	Supervised classification/ regression; growing self-supervised pre-training	Supervised (classification); self-supervised (masking, autoregressive, TPP)	Self-supervised link prediction as pre-training; supervised node classification	Self-supervised LM pre-training; supervised fine-tuning; zero/few-shot prompting
Key strengths	Simplicity; interpretability (tree models); efficiency; robust to missing values	Captures temporal dynamics; mature pre-training ecosystem; scalable	Models relational context and structural evolution; strong for interaction-heavy data	Flexibility; zero-shot generalization; rich semantic understanding; world knowledge
Key limitations	Loses temporal/structural information; deep models often underperform trees on small data	Cannot model multi-entity structure; sensitive to sequence length; cold-start for new users	High computational cost; complex training pipelines; poor for isolated nodes	Token inefficiency for structured data; numerical imprecision; high inference cost; opacity
Best-fit applications	Risk scoring, user profiling, CTR prediction, tabular classification	Sequential recommendation, session prediction, clinical event forecasting, fraud sequence detection	Social network analysis, financial transaction monitoring, collaborative filtering, link prediction	Conversational agents, semantic user profiling, cross-domain transfer, explanation generation

(DyG2Vec (Alomrani et al., 2023)). Contrastive learning is used in tabular settings (SCARF (Bahri et al., 2022), CARTE (Kim et al., 2024)), sequential modeling (CL4SRec (Xie et al., 2022), CoSeRec (Liu et al., 2021a)), and graph methods (TCL (Wang et al., 2021a)). In dynamic graphs, the most prevalent form of self-supervised pre-training is temporal link prediction: the model learns to predict future interactions based on historical ones, using the natural temporal ordering of edges as an implicit supervision signal. This "future-supervises-past" formulation, adopted by nearly all CTDG methods (e.g., TGN, TGAT, DyGFormer) as the default pre-training task (Rossi et al., 2020; Xu et al., 2020; Yu et al., 2023b), is conceptually parallel to autoregressive prediction in sequences—both exploit the arrow of time to generate labels without manual annotation. For LLMs, self-supervised pre-training is foundational rather than optional: models like GPT are trained via next-token prediction and BERT via masked token prediction on massive text corpora. When applied to behavioral data, this pre-trained knowledge provides a semantic and reasoning substrate that can be adapted through prompting or fine-tuning, effectively making every LLM-based behavioral method a beneficiary of self-supervised pre-training, even when the downstream adaptation itself is supervised.

Despite these parallels, the augmentation and pretext task strategies required to generate effective training signals diverge significantly across modalities: tabular methods corrupt feature values, sequential methods crop or reorder sub-sequences, graph methods leverage temporal edge ordering, and LLMs rely on token-level masking or autoregressive generation over linguistic corpora. This suggests that while the pre-training philosophy, i.e., exploit the intrinsic structure of unlabeled data to learn transferable representations, is genuinely universal across behavioral data modalities, effective implementation remains modality-specific and constitutes a key engineering challenge in each domain.

6.1.3 Time and Position Encoding

Encoding "when" an event occurs is critical for all modalities except static tabular data. Functional time encodings (Xu et al., 2019; Kazemi et al., 2019) are shared between sequential models and CTDG methods, with recent learnable variants (e.g., LeTE (Chen et al., 2025b)) extending their expressiveness. Positional encodings from the Transformer literature (sinusoidal, RoPE (Su et al., 2024a), ALiBi (Press et al., 2022)) are applied in both sequential recommendation and dynamic graph Transformers. The key distinction is that event sequences typically encode a single timeline per user, whereas dynamic graphs must encode time for every edge in a multi-entity interaction history, making scalability a more pressing concern in graph settings. More detailed discussions about time and position encoding are provided in Sec. 7.

6.1.4 Memory and State-tracking Mechanisms

Maintaining persistent node or user states that evolve with each new interaction is a shared design pattern. In sequential modeling, RNN hidden states and memory-augmented networks (MANN (Chen et al., 2018)) serve this purpose. In CTDG methods, TGN (Rossi et al., 2020) formalized the memory module for graph nodes, spawning a large family of memory-based models. In LLM-based behavioral analysis, context windows and RAG play an analogous role by bringing past behavioral context into the model's reasoning scope. These mechanisms share the goal of long-term information retention but differ in update granularity: sequential models update per event, graph models update per interaction involving a node, and LLMs re-read or retrieve entire behavioral histories.

6.2 Core Assumptions and Trade-offs

Each modality encodes a different set of assumptions about which aspects of behavior are most important for downstream tasks. These assumptions introduce distinct trade-offs.

6.2.1 Tabular Data: Aggregation as Abstraction

Tabular methods assume that behavior can be meaningfully summarized into a fixed-dimensional feature vector. This abstraction discards temporal ordering and relational structure but gains simplicity, interpretability, and computational efficiency. Tree-based models further benefit from natural handling of heterogeneous feature types and missing values. The main limitation is that aggregated features cannot distinguish behavioral patterns (e.g., binge purchasing vs. steady purchasing) from behavioral volumes, leading to information loss when temporal dynamics or interaction structures are decision-relevant.

6.2.2 Event Sequences: Order as the Primary Signal

Sequential methods assume that the chronological order of events (and, in more recent models, their exact timestamps) is the most informative behavioral signal. This makes them powerful for next-action prediction but introduces challenges when events are sparse, irregularly timed, or involve complex multi-party interactions that a linear timeline cannot capture. Another practical limitation is sensitivity to sequence length: very short histories provide insufficient context, while very long histories increase computational cost and may introduce noise from outdated behavior.

6.2.3 Dynamic Graphs: Structure as Context

Graph-based methods assume that behavior is best understood through the evolving relational context among entities. This is advantageous when inter-entity dependencies (e.g., social influence, collaborative patterns, cascading interactions) are predictive. However, graph methods introduce higher computational overhead, more complex training pipelines (especially for CTDGs), and greater sensitivity to graph construction choices (e.g., how to define edges, what temporal granularity to use for snapshots). The reliance on graph structure also means that isolated entities with few connections receive poor representations.

6.2.4 Textual Data: Semantics as a Bridge

LLM-based methods assume that behavioral data can be meaningfully translated into or enriched through natural language. This assumption is strongest for inherently textual behaviors (reviews, conversations) and weakest for purely numerical or structural interactions. The main advantages are flexibility, zero-shot generalization, and the ability to inject world knowledge. The main limitations are token efficiency (serializing large tables or long sequences is expensive), numerical precision (LLMs struggle with exact quantitative reasoning), and the opacity of how the model weighs behavioral versus linguistic signals.

6.2.5 Cross-cutting Trade-offs

Beyond the modality-specific assumptions discussed above, several fundamental trade-offs cut across all four modalities and recur whenever practitioners must choose among representation learning paradigms. These

tensions are not fully resolvable as they reflect inherent design compromises, but making them explicit helps frame modality selection as a deliberate engineering decision rather than a default choice driven by data format alone.

- **Expressiveness vs. Efficiency.** Richer representations (graphs, long sequences) capture more behavioral nuance but require more computation and data.
- **Temporal Precision vs. Scalability.** CTDG and TPP-based methods offer fine-grained temporal modeling but are harder to train at scale than snapshot-based or tabular approaches.
- **Generalization vs. Inductive Bias.** LLMs generalize broadly but lack task-specific inductive biases; tree models and GNNs embed strong domain priors but transfer less easily.
- **Interpretability vs. Performance.** Tabular (especially tree-based) and symbolic sequential methods offer interpretable decisions, whereas deep graph and LLM methods often trade interpretability for predictive power.

6.3 Method Selection Guidance

A central practical question for researchers and practitioners is: given a behavioral dataset and a target application, which modality and method family should one choose? While no universal rule exists, the following guidelines synthesize the patterns observed across the literature reviewed in this survey.

Task's Decision-critical Signal As noted in Sec. 5, the most effective representation is usually determined by which aspect of behavior is most decision-critical:

- If the task depends primarily on **aggregated user/entity profiles** (e.g., risk scoring, demographic classification), tabular representations are often sufficient and provide the best efficiency-accuracy trade-off, especially when tree-based models can be applied.
- If the task requires understanding **temporal order and recency** (e.g., next-item prediction, session-based recommendation, clinical event forecasting), event sequence modeling is the natural choice.
- If the task is sensitive to **inter-entity interactions and structural evolution** (e.g., fraud detection in transaction networks, social influence modeling, link prediction), dynamic graph methods are most appropriate.
- If the task involves **semantic reasoning, intent understanding, or explanation generation** (e.g., conversational recommendation, user profiling from text, financial report analysis), LLM-based approaches offer the richest representational capacity.

Data Characteristics Beyond the task, several data properties influence method and modality choice:

- **Label scarcity:** Self-supervised pre-training is most mature for sequential (BERT4Rec-style masking) and graph (link prediction as pretext) settings. LLMs offer zero-shot or few-shot alternatives. Tabular self-supervised methods exist but remain less established.
- **Data scale:** Tabular models and shallow sequential models scale well. Deep graph methods, especially CTDGs, face scalability challenges on billion-edge graphs despite recent system-level optimizations (TGL (Zhou et al., 2022), SPEED (Chen et al., 2023e)). LLMs are constrained by context length and inference cost.
- **Feature heterogeneity:** When features include a mix of numerical, categorical, and textual attributes, tabular deep learning (e.g., FT-Transformer (Gorishniy et al., 2021)) or LLM serialization can handle heterogeneity naturally, whereas pure sequential or graph models may require careful feature engineering.

- **Temporal irregularity:** For highly irregular event timing, continuous-time models (Neural TPPs, CTDG methods) are preferable to discrete-time methods.

Application-modality Alignment Drawing on the application analysis in Sec. 5, we observe recurring patterns of modality preference across domains:

- **E-commerce recommendation:** Dominated by event sequence methods (SASRec (Kang & McAuley, 2018), BERT4Rec (Sun et al., 2019b)), with growing adoption of LLM-augmented approaches (P5 (Geng et al., 2022), Chat-REC (Gao et al., 2023a)). Dynamic graphs are used when collaborative signals or user-item-user interactions are important (DGSR (Zhang et al., 2023b), EAGLE (Li et al., 2025a)).
- **Healthcare prediction:** EHR data naturally map to event sequences (BEHRT (Li et al., 2020c), RETAIN (Choi et al., 2016a)) and tabular features (DeepSurv (Katzman et al., 2018)). Knowledge graphs and LLMs (EHRAgent (Shi et al., 2024), KARE (Jiang et al., 2025)) are emerging for treatment reasoning.
- **Financial fraud detection:** Transaction networks favor dynamic graph modeling (EvolveGCN (Pareja et al., 2020), GADY (Lou et al., 2023), DyHDGE (Wang et al., 2024f)) and behavioral sequence analysis (CLeAR (Huang et al., 2024b), UB-PTM (Liu et al., 2022a)). Tabular features remain important for aggregated risk scores.
- **Social media analysis:** User profiling leverages textual/LLM methods (LLM-TUP (Sabouri et al., 2025), GPT-based personality inference) and graph methods (GREENER (Panayotov et al., 2022), HL-MRFs (Farnadi et al., 2020)). Content recommendation blends sequential and graph approaches.
- **Gaming:** Player behavior modeling primarily uses event sequences (player2vec (Wang et al., 2024d), RNN-based goal recognition). LLMs are increasingly adopted for NPC control (Voyager (Wang et al., 2024b), MineDojo (Fan et al., 2022)).

Multi-modal Combinations. In practice, many real-world systems combine multiple modalities. For example, tabular user features are often concatenated with sequential embeddings in recommendation systems (Wu et al., 2025b), and LLM-generated features are fed into graph or sequence models (Xi et al., 2024; Liu et al., 2025). The most robust strategy is often to match the primary modality to the task's decision-critical signal and augment with secondary modalities to address coverage gaps (e.g., using LLMs to enrich cold-start items with textual descriptions, or using graph structure to model collaborative signals that sequences alone cannot capture).

Cross-Modal Comparison Tab. 6 provides a systematic comparison across the four behavioral data modalities along key dimensions including representation type, core assumptions, typical supervision, evaluation protocols, strengths, limitations, and application alignment. We also include Tab. 7 to compare the dominant method families within each modality along practical axes.

Specifically, we evaluate each method family along five practically oriented dimensions that we consider most relevant for guiding modality and model selection. **Scalability** reflects the method's ability to handle large-scale behavioral data, a primary deployment constraint given that real-world systems routinely involve millions of users and billions of interactions. **Label Efficiency** captures how well a method performs when labeled data are scarce—a common situation in behavioral modeling where annotations are expensive or rare (e.g., fraud labels, clinical outcomes)—and directly relates to the self-supervised and semi-supervised paradigms discussed throughout this survey. **Temporal Precision** measures the granularity at which a method models the timing of behavioral events, which is a central axis of differentiation across modalities: tabular methods discard temporal information entirely, event sequences preserve order and intervals, and CTDG methods retain exact timestamps. **Interpretability** indicates whether the method produces decisions that humans can inspect and understand, a critical requirement in high-stakes domains such as finance and healthcare, and a key factor in the expressiveness-vs-transparency trade-off identified in Sec. 6.2. Finally, **Cold-start Handling** assesses the method's robustness when encountering new entities with little

Table 7: Cross-modal comparison of method families along practical axes. Ratings are relative within each row: **H** = High, **M** = Medium, **L** = Low.

Method Family	Scalability	Label Efficiency	Temporal Precision	Interpret-ability	Cold-start Handling
Tabular Data (Sec. 4.1)					
Machine Learning-based	H	M	L	H	L
DL-based: Supervised-only	M	L	L	M	L
DL-based: Supervised-plus	M	M–H	L	M	M
LLM-driven	L	H	L	M	H
Event Sequences (Sec. 4.2)					
Symbolic: Sequential Pattern Mining	M	L	L	H	L
Symbolic: Markov Chain Models	H	M	L	H	L
DL-based: Supervised (RNN/CNN/Transformer)	H	L	M	L	L
DL-based: Self-supervised (Autoregressive)	H	M–H	M	L	L–M
DL-based: Self-supervised (Masked Reconstruction)	H	H	M	L	M
DL-based: Self-supervised (TPP-based)	M	M–H	H	M	L
DL-based: Self-supervised (Contrastive)	M–H	H	M	L	M
DL-based: Self-supervised (Adversarial)	M	M–H	M	L	L–M
Dynamic Graphs (Sec. 4.3)					
DTDG: Static Embed. + Temporal Alignment	M–H	M	L	M	L
DTDG: GNN+RNN-based	M	M	M	L	L
DTDG: Attention-based	M–H	M	M	L–M	L
CTDG: RNN-based	M	M	H	L	L
CTDG: TPP-based	M	M–H	H	M	L
CTDG: Random Walk-based	M–H	M–H	M–H	M	M
CTDG: Attention/Time Encoding-based	M	M–H	H	L–M	L–M
CTDG: MLP-based	H	M	M	M	L
Textual Data (Sec. 4.4)					
LLMs for Augmenting: Feature Augmentation	M	H	L	M–H	H
LLMs for Augmenting: Behavior Augmentation	M	H	L–M	M	H
LLMs for Encoding: Representation Enhancement	L–M	H	L–M	L–M	M–H
LLMs for Encoding: Cross-modality Unification	L–M	H	L–M	L	H
LLMs for Controlling	L	H	L	M–H	H

or no behavioral history—one of the most pervasive practical challenges in behavioral data modeling, where different paradigms exhibit sharply different capabilities (e.g., LLM-based methods can leverage semantic priors, whereas purely ID-based models fail entirely).

7 Key Modules

Although foundational modeling approaches such as tabular modeling, sequential modeling, and dynamic graph modeling have achieved significant progress in behavioral data analysis, the inherent complexity and diversity of behavioral data often render these basic methods insufficient for fully capturing critical factors such as temporal dependencies, structural patterns, and contextual information. To address these limitations, researchers have proposed a variety of enhanced modeling strategies designed to further improve the representational capacity and predictive performance of behavior models. These enhancement methods are not standalone modeling paradigms, but rather modular components or mechanisms that can be flexibly integrated into different types of models. They include, but are not limited to, time encoding, position encoding, context modeling, structure-aware mechanisms, contrastive learning, and pre-training. These strategies have demonstrated strong generalizability and effectiveness across a wide range of behavioral modeling tasks in recent years, becoming key drivers in the advancement of behavioral data modeling.

This section provides a systematic overview of these enhanced modeling methods, analyzing their underlying design principles, applicable scenarios, and integration approaches with primary modeling techniques.

7.1 Time Encoding

In behavioral data modeling, temporal information often contains critical dynamic patterns such as periodicity, time-delay dependencies, and triggering sequences. Whether in user behavior, medical events, or financial transactions, the timing or intervals of actions play a crucial role in understanding their underlying evolution. Early studies typically relied on handcrafted temporal features specifically designed for downstream tasks (Choi et al., 2016b; Baytas et al., 2017; Kwon et al., 2019). These methods involved decomposing timestamps into components (e.g., year, month, day), assigning embeddings to each component, and summing them to form the final time representation. While effective in certain scenarios, such approaches were labor-intensive, domain-specific, and limited in their ability to capture only predefined temporal patterns (Kazemi et al., 2019).

With the rapid development of attention mechanisms, researchers began exploring more flexible ways to model temporal information, particularly in capturing long-term dependencies and dynamic changes. This led to the emergence of Functional Time Encoding (FTE) methods, which aimed to provide time representations compatible with self-attention architectures. Representative works include Functional Time Representation (Xu et al., 2019) and Time2Vec (Kazemi et al., 2019). These methods apply multiple linear transformations to the raw time input, followed by pre-defined nonlinear functions (e.g., trigonometric functions) to generate multi-dimensional time embeddings. Their ability to effectively encode periodic patterns made them widely adopted in dynamic graph representation learning and time series modeling (Yu et al., 2023b; Chen et al., 2025b).

Despite the success of functional time encoding in capturing periodic behaviors, these methods inherently rely on strong inductive biases rooted in predefined periodic assumptions (Li et al., 2018; Xu et al., 2019; Kazemi et al., 2019). As a result, they often struggle to model more complex temporal patterns such as non-periodic and mixed behaviors. To address these limitations, recent studies proposed more expressive approaches like LeTE (Chen et al., 2025b), which aim to capture a broader range of temporal dynamics. LeTE introduces neural transformations that allow the model to automatically learn meaningful temporal features directly from raw time values, without relying on handcrafted components or fixed periodic functions. This enables the model to better adapt to the diverse and complex nature of real-world temporal behaviors. In contrast, Chung et al. propose using simpler linear time encodings and show that self-attention can extract temporal patterns from these features while reducing model complexity (Chung et al., 2025).

The evolution of time encoding techniques has significantly advanced the field of behavioral data analysis. In many real-world scenarios, such as user interactions, clinical events, or financial activities, behaviors are not only determined by what happens, but also by when and in what order they occur. Accurate temporal representations help uncover latent behavioral patterns such as periodicity, recency effects, burstiness, or long-term dependencies. Modern time encoding methods, especially those based on neural transformations, enable models to better capture these complex temporal dynamics without relying on rigid assumptions. This has led to more faithful representations of behavioral sequences, improved interpretability of behavioral trends, and enhanced generalization across diverse behavioral contexts. As research continues, there is a growing need for time encoding techniques that are not only expressive but also tailored to the unique challenges of behavioral data, such as irregular sampling, multi-scale temporal patterns, and context-dependent timing effects.

7.2 Position Encoding

In early sequence modeling, positional information was either implicitly captured through the use of sequential architectures such as RNNs, or explicitly introduced using integer-based position indices. These approaches allowed models to process tokens in a temporal order, thereby encoding positional dependencies through recurrence or positional tags. However, the emergence of attention-based models—particularly the Transformer—marked a paradigm shift, as these models lack an inherent notion of sequence order. This limitation sparked a surge of research interest in position encoding, aiming to explicitly inject positional information into attention mechanisms. Among the proposed solutions, two main categories have emerged: absolute position encoding, which assigns unique representations to fixed positions, and relative position encoding, which focuses on modeling the distance or offset between tokens. The development of more effective

position encoding schemes has since become a central theme in advancing the performance and generalization of attention-based models.

7.2.1 Absolute Position Encoding

Along with the attention mechanism, the sinusoidal position encoding is proposed (Vaswani et al., 2017), a form of absolute position encoding that maps each position to a unique, analytically defined vector. While this encoding is theoretically extrapolatable to arbitrarily long sequences, subsequent studies (Dai et al., 2019; Neishi & Yoshinaga, 2019) revealed practical limitations. Specifically, since the model is trained on sequences of limited length, the attention weights may fail to generalize to longer sequences during inference, leading to unstable outputs. This highlights a fundamental limitation of absolute position encoding in terms of extrapolation capacity in real-world applications.

In response to the extrapolation challenges of absolute position encoding, researchers have proposed various enhancement strategies. Ideally, position encodings should possess specific properties; notably, shift invariance—the property that a function's output remains unchanged when its input is uniformly shifted—was emphasized in (Wang et al., 2020) as a key factor for achieving robust extrapolation. a method involving random positional shifts during training was introduced by (Kiyono et al., 2021), where the original position index pos with $pos + k$, where k is sampled from a discrete uniform distribution $u\,(0, K)$. This approach discourages the model from relying on absolute positions and instead promotes learning of relative positional patterns. Building on this idea, additional perturbations, including global shifts, local jitters, and global scaling, were proposed in (Likhomanenko et al., 2021) to model position encodings as continuous signals with controlled variability to improve generalization. Additionally, representing word embeddings as continuous functions of position has been suggested (Wang et al., 2020), allowing word representations to evolve smoothly with increasing position indices. Finally, a dynamic systems perspective was explored by (Liu et al., 2023a), directly modeling the temporal dynamics between positional representations. Collectively, these efforts represent significant progress in enhancing absolute position encoding, aiming to improve its generalization and robustness in long-sequence modeling tasks.

7.2.2 Relative Position Encoding

While absolute position encodings provide a straightforward way to inject position information into Transformer models, they are inherently limited in their ability to generalize to sequences of unseen lengths or shifted patterns. More importantly, in many sequence modeling tasks, it is not the absolute position of a token that matters most, but rather its relative position with respect to other tokens. For instance, syntactic dependencies or temporal correlations are often defined by how far apart two elements are, rather than their fixed positions in the sequence (Huang et al., 2020b; Sinha et al., 2022).

Motivated by this observation, relative position encoding was first formalized in (Shaw et al., 2018) by explicitly incorporating pairwise positional relationships into the self-attention mechanism. This approach enhanced the model's ability to capture local structures and long-range dependencies.

Building upon this foundation, subsequent works extended relative position encoding in various directions. For instance, Transformer-XL (Dai et al., 2019) introduced introduced a segment-level recurrence mechanism to enable longer context modeling during inference. Similarly, a simplified form of relative position encoding was adopted in T5 (Raffel et al., 2020), where each relative position is mapped to a learnable scalar that is directly added to the attention logits, improving computational efficiency. The concept was further advanced by DeBERTa (He et al., 2021), which decoupled content and positional information and applied the embeddings in a Transformer-XL-style architecture. Additionally, a log-bucket relative position bucketing strategy has been utilized (Chi et al., 2022); this method discriminates nearby positions more finely while coarsely representing distant ones. Such a logarithmic approach reduces the number of parameters and improves extrapolation capacity.

To further enhance length generalization, strategies such as adding bias terms to the attention scores (Wennberg & Henter, 2021) and incorporating actual distance information (Wu et al., 2021) have been explored. Notably, ALiBi (Press et al., 2022) emerged as the first relative position encoding method

Table 8: Behavioral datasets for tabular representation learning

Dataset	# Numeric	# Categories	# Samples	# Tasks
California Housing (CA)	8	0	20,640	Regression
Adult (AD)	6	8	48,842	Classification
Helena (HE)	27	0	65,196	Classification
Jannis (JA)	54	0	83,733	Classification
Higgs (HI)	28	0	1,000,000	Classification
ALOI(AL)	128	0	108,000	Classification
Epsilon (EP)	2,000	0	500,000	Classification
Year (YE)	90	0	515,345	Regression
Covertype(CO)	54	0	518,012	Classification
Bank(BK)	7	9	45,211	Classification
Blastchar (BC)	3	17	7,043	Classification
Shoppers (SH)	4	14	12,330	Classification
Volkert (VO)	180	1	58,310	Classification
Income (IC)	6	8	32,561	Classification
Yahoo (YA)	699	0	709,877	Regression
Microsoft (MI)	136	0	1,200,192	Regression

specifically designed for length extrapolation. It combines simplicity with strong performance by applying fixed, position-dependent biases to the attention weights.

Inspired by sinusoidal encodings, Rotary Position Embedding (RoPE) is introduced (Su et al., 2024a), which encodes relative position by rotating the query and key vectors in a position-dependent manner, rather than adding explicit position vectors. This continuous and extrapolatable encoding has been widely adopted in large-scale language models such as LLaMA and ChatGLM, making it one of the most prominent relative position encoding techniques today.

In conclusion, enhancing temporal and position encoding is essential for effective behavioral data analysis. Behavioral sequences often involve complex temporal patterns where the relative timing and order of actions are more informative than their absolute positions. Advanced encoding strategies—such as relative position embeddings and time-aware attention—enable models to better capture dependencies, generalize across varying sequence lengths, and improve both performance and interpretability. These improvements highlight the critical role of encoding-aware designs in modeling real-world behavioral data.

7.3 Summary and Takeaways

The discussion about key modules also reveals that time and position encoding are not isolated technical details, but recurring design choices that distinguish stronger behavioral representations across modalities. In event sequences, they determine whether irregular intervals and long histories remain usable; in dynamic graphs, they regulate how event order and temporal distance affect message passing; and in textual or LLM-based settings, they shape the model's ability to align semantics with behavioral chronology. The broader lesson is that many empirical gains attributed to larger architectures are in fact mediated by better encoding of temporal and positional structure.

8 Datasets and Benchmarks

High-quality datasets play a crucial role in advancing research on behavioral data representation learning, as they provide a solid foundation for model development and evaluation. Corresponding benchmarks further enable fair and consistent comparisons across different methods. In this section, we discuss commonly used behavioral data datasets along with benchmarks, and present relevant statistical information for each.

Table 9: Behavioral benchmarks for tabular representation learning.

Benchmark	Paper	Repository	Year
OpenML-CC18	(Bischl et al., 2021)	OpenML	2017
Regularized DNNs	(Kadra et al., 2021)	GitHub	2021
TabularBench	(Grinsztajn et al., 2022)	GitHub	2022
TabZilla	(McElfresh et al., 2023)	GitHub	2023
OpenTabs	(Ye et al., 2024a)	GitHub	2024
TALENT	(Ye et al., 2024b)	GitHub	2024
TDBench	(Kang et al., 2025)	GitHub	2025

8.1 Tabular Data

Datasets To support research on behavior modeling in tabular data formats, both academia and industry have developed a wide range of representative datasets that are widely used for evaluating and comparing model performance. Among them, datasets such as California Housing (CA), Adult (AD), Helena (HE), and Jannis (JA) have been repeatedly adopted in numerous studies (Wu et al., 2024a; Na et al., 2024; Ahamed & Cheng, 2024; Xu et al., 2024; Joseph & Raj, 2022), and have gradually become standard benchmarks for tabular behavior modeling tasks. These datasets span a variety of prediction objectives—from regression to multi-class classification—and cover diverse application domains, including financial risk assessment, socioeconomic analysis, and local lifestyle services.

Some of these datasets originate from public data science competitions, such as those hosted on the UCI Machine Learning Repository and Kaggle, ensuring a high degree of standardization and broad generality. This contributes to the reproducibility of research and facilitates fair comparison across methods. Moreover, behaviors in these datasets are typically recorded in a structured form—such as samples composed of user attributes and behavioral outcomes—naturally framing them as classification or regression problems, and providing a solid foundation for building more complex behavior modeling frameworks. We summarize the most commonly used datasets in Tab. 8.

Benchmarks To systematically evaluate the performance of various behavior modeling approaches, researchers have proposed a range of benchmark settings to standardize experimental protocols and enable fair comparisons across models. In the context of behavior modeling with tabular data, most datasets are sourced from the open platform OpenML, and many well-established benchmarks in this domain have been built upon it. Depending on their specific evaluation goals, researchers selectively construct benchmark suites by sampling from the OpenML repository.

One of the most widely adopted benchmarks is OpenML-CC18 (Bischl et al., 2021), which was constructed using eight filtering criteria and includes 72 high-quality classification datasets. This benchmark has become a standard reference for evaluating machine learning models on tabular data. In addition, the influential work (Grinsztajn et al., 2022) selected 45 representative datasets from OpenML based on criteria such as heterogeneity, feature richness, data realism, and task difficulty, creating a challenging and insightful benchmark for comparing deep learning and traditional models.

Building upon this, TaZilla (McElfresh et al., 2023) conducted a comprehensive analysis of 176 classification datasets from OpenML. It introduced additional filtering criteria that consider not only task diversity but also practical constraints such as runtime. Furthermore, it ensured broad coverage of datasets commonly used in popular studies, contributing to a more representative and reproducible benchmark foundation. Alongside TDColER (Kang et al., 2025), the authors also propose TDBench, a tabular distillation benchmark covering 23 datasets, 7 model classes, and 11 distillation strategies. We summarize the collected benchmarks in Tab. 9.

8.2 Event Sequences

Table 10: Sequential behavioral datasets

Dataset	# Users	# Items/Businesses	# Samples	# Domain
Amazon Product Reviews 2023	54.51M	48.19M	571.54M	E-commerce
Amazon Q&A	–	191K	1.48M Q, 4.02M A	E-commerce
ModCloth Marketing Bias	44.78K	1.02K	99.89K	E-commerce
Google Local Reviews 2021	113.64M	4.96M	666.32M	Local services
Google Restaurants	1.01M	65K	1.77M	Local services
Twitch	15.5M	465K	124M	Media content
Food.com Recipe & Review	226.57K	231.64K	1,13M	Media content
EndoMondo Fitness Tracking Data	1.10K	–	253.02K	Healthcare
Behance Community Art Data	63.50K	178.79K	1M	Art
Taobao UserBehavior	987.99K	4.16M	100.15M	E-commerce
MovieLens 32M	200.95K	87.59K	32.00M Ratings	Media content
Steam Video Game and Bundle Data	2.57M	15.47K Items, 615 Bundles	7.79M	Gaming

Table 11: Behavioral sequential representation learning benchmarks.

Benchmark	Paper	Repository	Year
GRU	(Hidasi et al., 2016)	GitHub	2015
BERT	(Devlin et al., 2019)	GitHub	2017
CPC	(van den Oord et al., 2018)	GitHub	2018
Transformer	(Vaswani et al., 2017)	GitHub	2017
PrimeNet	(Chowdhury et al., 2023)	GitHub	2023
RMTPP	(Du et al., 2016)	GitHub	2016

Datasets The datasets utilized in event sequence modeling encompass a broad spectrum of domains, such as E-commerce, Local Services, Media Content, and specialized verticals including Healthcare, Art, and Gaming. In these datasets, an event is typically defined as a timestamped user interaction, such as a purchase, rating, review, view, or content engagement.

A significant portion of research focuses on the E-commerce domain. Prominent examples include the Amazon Product Reviews 2023 (Hou et al., 2024) and Amazon Q&A (Wan & McAuley, 2016; McAuley & Yang, 2016), which capture rich user feedback and interaction histories. In the fashion and retail sector, ModCloth Marketing Bias (Wan et al., 2020) and the widely used Taobao UserBehavior (Zhu et al., 2018; 2019b; Zhuo et al., 2020) are frequently employed to benchmark models on purchasing and browsing sequences.

In the realm of Local Services, event sequences are extensively studied to understand user preferences for physical establishments. Datasets such as Google Local Reviews 2021 (Li et al., 2022b; Yan et al., 2023) and Google Restaurants (Yan et al., 2023) offer valuable insights into location-based user activities and service interactions. For Media Content, large-scale datasets are used to evaluate recommendation and sequence prediction capabilities across different media types. This category includes Twitch (Rappaz et al., 2021) for live streaming interactions, the massive MovieLens 32M (Harper & Konstan, 2015) for movie ratings, and Food.com Recipe & Review (Majumder et al., 2019) for culinary content exploration.

Furthermore, several datasets target specific vertical domains. These include EndoMondo Fitness Tracking Data (Ni et al., 2019) for Healthcare, Behance Community Art Data (He et al., 2016) for Art appreciation, and Steam Video Game and Bundle Data (Kang & McAuley, 2018; Wan & McAuley, 2018; Pathak et al., 2017) for Gaming behavior analysis. We summarize the statistics and characteristics of these event sequence datasets in Tab. 10.

Benchmarks To evaluate the performance of various approaches, a set of classical and widely effective sequential modeling methods have been selected as benchmarks. These methods are representative works within the categories of time series modeling discussed previously. They are general-purpose and applicable to diverse downstream domains, such as item recommendation, financial forecasting, and healthcare predic-

Table 12: Behavioral datasets for dynamic graph representation learning.

	Dataset	# Nodes	# Edges	# Snapshots / Unique Steps	Domain	Time Granularity
DTDG	BSI-ZK	1,744,561	56,194,191	257	Finance	Days
	BSI-SVT	89,564	190,133	49	Finance	Weeks
	Bitcoin-OTC	5,881	35,592	279	Finance	Weeks
	Bitcoin-Alpha	3,783	24,186	274	Finance	Weeks
	Reddit-Title	54,075	571,927	178	Interaction	Weeks
	Reddit-Body	35,776	286,561	178	Interaction	Weeks
	AS-733	7,716	11,965,533	733	Traffic	Days
	UCI-Message	1,899	59,835	49	Social	Weeks
	Flights	13,169	1,927,145	122	Transport	Days
	Can. Parl.	734	74,478	14	Politics	Years
	US Legis.	225	60,396	12	Politics	Congresses
	UN Trade	255	507,497	32	Finance	Years
	UN Vote	201	1,035,742	72	Politics	Years
	Contact	692	2,426,279	8,064	Proximity	5 Minutes
CTDG	Wikipedia	9,227	157,474	152,757	Social	Unix timestamps
	Reddit	10,984	672,447	669,065	Social	Unix timestamps
	MOOC	7,144	411,749	345,600	Interaction	Unix timestamps
	LastFM	1,980	1,293,103	1,283,614	Interaction	Unix timestamps
	Enron	184	125,235	22,632	Social	Unix timestamps
	Social Evo.	74	2,099,519	565,932	Proximity	Unix timestamps
	ML25M	221,588	25,000,095	20,115,267	Interactions	Unix timestamps
	DGraphFin	4,889,537	4,300,999	821	Finance	Unix timestamps
	Taobao	5,149,747	100,135,088	815,859	E-commerce	Unix timestamps
	tgbl-review-v2	352,637	4,873,540	6,865	E-commerce	Unix timestamps
	tgbl-coin	638,486	22,809,486	1,295,720	Finance	Unix timestamps
	tgbl-comment	994,790	44,314,507	30,998,030	Interaction	Unix timestamps
	tgbl-flight	18,143	67,169,570	1,385	Transport	Unix timestamps

tion. While domain-specific methods exist, they often build upon these general paradigms (e.g., GRU4Rec in recommendation is based on GRU, BEHRT in healthcare on BERT, and player2vec in gaming on the Transformer architecture). These general benchmark methods are summarized in Tab. 11.

8.3 Dynamic Graphs

Datasets The datasets used in dynamic graph representation learning span a wide range of domains, including finance, social media, and e-commerce, providing researchers with diverse resources to study behavioral data across different application scenarios. Based on the intrinsic characteristics of the datasets, we categorize them into DTDG and CTDG datasets. As discussed in Sec. 4.3.2, CTDG representation learning models can also handle DTDG data. For example, in DyGFormer (Yu et al., 2023b), the authors evaluate their model using several DTDG datasets, such as UCI-Message, Flights, and Can.Parl.

Commonly used DTDG datasets include BSI-ZK, BSI-SVT, Bitcoin-OTC, Bitcoin-Alpha, Reddit-Title, Reddit-Body, and AS-733 (You et al., 2022). For CTDG datasets, Wikipedia, Reddit, MOOC, and LastFM are among the most widely adopted. These four datasets are originally introduced in the JODIE paper (Kumar et al., 2019) and have since been extensively used in subsequent research. The authors of JODIE continue to maintain and update the Stanford Network Analysis Project (SNAP)[2], which hosts many other dynamic graph datasets.

[2] https://snap.stanford.edu/index.html

Table 13: Benchmarks for dynamic graph representation learning.

Benchmark	Paper	Repository	Specialize	Year
PyG-Temporal	(Rozemberczki et al., 2021)	GitHub	DTDG	2021
TGL	(Zhou et al., 2022)	GitHub	Large-scale CTDG	2022
SPEED	(Chen et al., 2023e)	GitHub	Large-scale CTDG	2023
DYGL	(Ma et al., 2024b)	GitHub	DTDG and CTDG	2023
DyGLib	(Yu et al., 2023b)	GitHub	CTDG	2024
TGB	(Huang et al., 2023b; Gastinger et al., 2024)	GitHub	CTDG	2024
BenchTemp	(Huang et al., 2024a)	GitHub	CTDG	2024
DGB	(Gravina & Bacciu, 2024)	GitHub	DTDG and CTDG	2024
BenchTGNN	(Yang et al., 2024b)	GitHub	CTDG	2024
TGX	(Shirzadkhani et al., 2024)	GitHub	CTDG	2024
DGNN	(Feng et al., 2025)	GitHub	DTDG and CTDG	2024
UTG	(Huang et al., 2025b)	GitHub	DTDG and CTDG	2024

Since the publication of DyGFormer, broader evaluations on CTDG representation learning models have emerged. For instance, in addition to several DTDG datasets, DyGFormer also uses Enron and Social Evo. datasets for evaluation. Large-scale datasets such as DGraphFin (Huang et al., 2022), ML25M (Harper & Konstan, 2015), and Taobao (Zhuo et al., 2020) are often used to benchmark scalable training frameworks. Recently, the release of the Temporal Graph Benchmark (TGB) (Huang et al., 2023b; Gastinger et al., 2024) has introduced a series of larger-scale datasets, including tgbl-review-v2, tgbl-coin, tgbl-comment and tgbl-flight, which have further enriched the evaluation landscape. We summarize the commonly used dynamic graph datasets in Tab. 12.

Benchmarks The early stage of dynamic graph benchmark development, exemplified by PyTorch Geometric Temporal (PyG-Temporal) (Rozemberczki et al., 2021), focuses primarily on DTDGs. This early framework provides essential support for snapshot-based modeling but remains limited in scalability and task diversity.

With the increasing interest in modeling fine-grained temporal dynamics, the next phase sees the emergence of CTDG benchmarks. TGL (Zhou et al., 2022) and SPEED (Chen et al., 2023e) emphasize large-scale learning and efficient training on billion-edge graphs, marking a shift toward practical deployment scenarios and industrial applicability. Recent years have witnessed rapid expansion of the ecosystem. The Dynamic Graph Library (DyGLib) (Yu et al., 2023b), released alongside DyGFormer, as well as the Temporal Graph Benchmark (TGB) (Huang et al., 2023b) and its extended version TGB 2.0 (Gastinger et al., 2024), set new standards by offering large-scale, domain-diverse CTDG datasets alongside automated evaluation pipelines and leaderboards, significantly improving reproducibility and comparability. Parallel efforts—such as BenchTemp (Huang et al., 2024a), BenchTGNN (Yang et al., 2024b), Temporal Graph Analysis with TGX (TGX) (Shirzadkhani et al., 2024), DGNN (Feng et al., 2025) have further enriched the landscape with modular toolkits that cover an even wider spectrum of models.

In addition to CTDG-focused efforts, several benchmark libraries aim to bridge the DTDG and CTDG paradigms. DYGL-library (DYGL) (Ma et al., 2024b) and the Dynamic Graph Benchmark (DGB) (Gravina & Bacciu, 2024) support both DTDG and CTDG models and tasks under a unified framework. Similarly, Unifying Temporal Graph (UTG) (Huang et al., 2025b) attempts to provide a general infrastructure that accommodates both DTDG and CTDG representation learning methods, further promoting methodological consistency across different temporal settings. We summarize representative benchmarks in Tab. 13.

8.4 Textual Data

Datasets Finally, we introduce the *language-based behavioral* datasets. These datasets typically consist of entities and the behavioral interactions occurring among them. Such interactions are often timestamped, and both entities and interactions are associated with rich textual attributes. We summarize these datasets

Table 14: Datasets for textual behavioral data.

Datasets	# Entity	# Behavior	# Behavior Category	# Timestamp	Domain
Enron	42,711	797,907	10	1,006	E-mail
GDELT	6,786	1,339,245	237	2,591	Knowledge graph
ICEWS1819	31,796	1,100,071	266	730	Knowledge graph
Stack elec	397,702	1,262,225	2	5,224	Multi-round dialogue
Stack ubuntu	674,248	1,497,006	2	4,972	Multi-round dialogue
Googlemap_CT	111,168	1,380,623	5	55,521	E-commerce
Amazon movies	293,566	3,217,324	5	7,287	E-commerce
Yelp	2,138,242	6,990,189	5	6,036	E-commerce

in Tab. 14, which collectively span a broad spectrum of domains, ranging from *E-mail communication* and *Knowledge graphs* to *Multi-round dialogues* and large-scale *E-commerce platforms*. For example, the Enron dataset models organizational email exchanges: entities correspond to individual users annotated with descriptive metadata such as email addresses, while interactions represent message transmissions that are further accompanied by the full textual content of the emails. These datasets not only mirror real-world communication and decision-making processes, but also provide fertile ground for advancing research at the nexus of natural language processing and behavioral modeling. They offer a robust empirical foundation for harnessing LLMs to capture fine-grained behavioral regularities and to drive performance gains across a diverse set of downstream applications.

Benchmarks The benchmarking practice of applying LLMs to behavioral data is still in its early stages. For instance, DTGB (Zhang et al., 2024b) has recently introduced a benchmark that models user behaviors through dynamic text-attributed graphs, thereby providing one of the first systematic attempts to bridge behavioral dynamics with textual semantics in an LLM-friendly manner. Nevertheless, beyond this effort, we find relatively few benchmarks or frameworks that explicitly focus on modeling language-based behavioral data with LLMs. One possible reason is the inherent complexity of such data: interactions are not only temporally ordered but also linguistically rich, which poses challenges in aligning heterogeneous modalities, designing scalable representations, and ensuring efficient training. Moreover, the absence of standardized datasets and evaluation protocols further hinders the community from conducting consistent and reproducible studies. This gap highlights a promising research direction for future work. Establishing comprehensive benchmarks, unified evaluation metrics, and diverse datasets would significantly accelerate progress, allowing LLMs to better capture the interplay between textual semantics and behavioral regularities.

8.5 Benchmark maturity and evaluation gaps

The benchmark landscape is highly uneven across modalities. Tabular learning already benefits from relatively mature suites built on OpenML-style protocols, and dynamic graph learning has recently converged toward more standardized toolkits and leaderboards such as TGB and DyGLib. By contrast, event-sequence evaluation remains more fragmented and application-specific, while textual behavioral modeling still lacks broadly adopted benchmark infrastructure. A more fundamental limitation is that current benchmarks are almost entirely modality-specific: they rarely allow the same underlying behavioral data to be modeled and evaluated as tabular features, sequences, graphs, or text under a unified protocol. As a result, many reported improvements are convincing within a modality, but much less informative for cross-modality model selection.

9 Comparison with Existing Surveys

Tab. 15 summarizes representative surveys adjacent to our work. Existing surveys generally remain limited along at least one important dimension. Some are modality-specific, focusing on only one form of behavioral data, such as tabular representation learning (Jiang et al., 2026), sequential modeling (Shchur et al., 2021; Zhou et al., 2025b), dynamic graph representation learning (Jiao et al., 2025; Feng et al., 2025), or textual

Table 15: Comparison with representative adjacent surveys. ∘ denotes partial coverage.

Survey	Year	Main organizing idea	Coverage						
			Tab.	Event Seq.	Dyn. Graph	Text/LLM	Cross-modal view	Tasks	Data/Bench.
Neural Temporal Point Processes: A Review (Shchur et al., 2021)	2021	Intensity / encoder / objective design	×	✓	×	×	×	∘	∘
A Survey on User Behavior Modeling in Recommender Systems (He et al., 2023)	2023	Behavior modeling paradigms in RecSys	×	∘	×	×	×	✓	∘
Embedding in Recommender Systems: A Survey (Zhao et al., 2023b)	2023	Matrix / sequential / graph embeddings	∘	✓	∘	∘	×	✓	∘
User Modeling and User Profiling: A Comprehensive Survey (Purificato et al., 2024)	2024	Historical evolution of user modeling	×	×	∘	∘	×	✓	×
A Survey on Deep Tabular Learning (Somvanshi et al., 2024)	2024	Historical evolution of tabular deep learning	✓	×	×	×	×	✓	∘
A Survey on Temporal Interaction Graph Representation Learning (Jiao et al., 2025)	2025	Structural / temporal / application taxonomy	×	∘	✓	×	×	✓	✓
Using NLP to Analyse Text Data in Behavioural Science (Feuerriegel et al., 2025)	2025	NLP methods for textual behavioral data	×	×	×	✓	×	✓	∘
A Survey of Foundation Model-Powered Recommender Systems (Huang et al., 2025a)	2025	Feature-based / generative / agentic FM RecSys	∘	∘	×	✓	×	✓	∘
Advances in Temporal Point Processes: Bayesian, Neural, and LLM Approaches (Zhou et al., 2025b)	2025	Bayesian → neural → LLM TPP taxonomy	×	✓	×	∘	×	∘	✓
A Comprehensive Survey of Dynamic Graph Neural Networks (Feng et al., 2025)	2025	DTDG / CTDG models and benchmarks	×	∘	✓	×	×	✓	✓
Representation Learning for Tabular Data (Jiang et al., 2026)	2026	Specialized / transferable / general tabular RL	✓	×	×	∘	∘	∘	✓
A Survey on Sequential Recommendation (Pan et al., 2026)	2026	Model families in sequential recommendation	×	✓	×	∘	∘	✓	∘
Ours: A Survey on Behavioral Data Representation Learning	2026	Unified modality taxonomy + within-modality sub-taxonomies	✓	✓	✓	✓	✓	✓	✓

behavioral data analysis (Somvanshi et al., 2024; Feuerriegel et al., 2025). Others are organized around a particular application domain, especially recommender systems, including user behavior modeling (He et al., 2023), sequential recommendation (Pan et al., 2026), foundation-model-powered recommendation (Huang et al., 2025a), and recommendation embeddings (Zhao et al., 2023b). More broadly scoped surveys such as user modeling and profiling (Purificato et al., 2024) cover related themes, but do not organize the literature through a representation-learning-centered, cross-modality taxonomy. As a result, prior surveys typically provide either deep coverage of a single modality or broad discussion within a single application domain, but not a unified comparison across major behavioral data forms together with their associated tasks, datasets, and benchmarks.

In comparison, our survey adopts a unified perspective that covers four major behavioral data forms within a single framework. This organization reflects the observation that the same underlying behavioral phenomenon can be represented in different data forms depending on modeling choices, and that this representational choice fundamentally determines the appropriate learning paradigm. Beyond the top-level categorization, we further introduce finer-grained sub-taxonomies within each modality. The survey also systematically connects these method taxonomies to downstream application scenarios, shared modeling modules, and modality-specific datasets and benchmarks.

10 Conclusion and Future Directions

In conclusion, behavioral data representation learning plays a critical role in extracting meaningful patterns from complex, dynamic behaviors, supporting diverse real-world applications. This survey systematically categorizes existing methods, clarifies their functionalities and application scenarios, and provides an insightful framework to guide future developments.

Looking forward, several critical areas demand further research: (1) development of integrated multi-modal models for richer and more comprehensive behavioral representations; (2) scaling representation learning techniques to handle real-time data streams and large-scale datasets; (3) enhancing interpretability and explainability to facilitate user trust and regulatory compliance; (4) leveraging the potential of foundational models such as large language models and pre-trained graph neural networks for advanced transfer learning

and domain adaptation; and (5) ensuring fairness, transparency, and ethical considerations in behavioral modeling practices. Addressing these areas will substantially advance the practical utility and theoretical robustness of behavioral data representation methodologies.

References

Yoshia Abe, Shuhei Matsuo, Ryoma Kondo, and Ryohei Hisano. Leveraging large language models for institutional portfolio management: Persona-based ensembles. *arXiv preprint arXiv:2411.19515*, 2024.

Futoon M. Abushaqra, Hao Xue, Yongli Ren, and Flora D. Salim. Seqlink: A robust neural-ode architecture for modelling partially observed time series. *Transactions on Machine Learning Research*, 2024.

Ami Abutbul, Gal Elidan, Liran Katzir, and Ran El-Yaniv. DNF-net: A neural architecture for tabular data. *arXiv preprint arXiv:2006.06465*, 2020.

Tejumade Afonja, Hui-Po Wang, Raouf Kerkouche, and Mario Fritz. Dp-2stage: Adapting language models as differentially private tabular data generators. *Transactions on Machine Learning Research*, 2025.

Jesús C. Aguerri, Mario Santisteban, and Fernando Miró-Llinares. The enemy hates best? toxicity in League of Legends and its content moderation implications. *European Journal on Criminal Policy and Research*, 29(3):437–456, 2023.

Md. Atik Ahamed and Qiang Shawn Cheng. MambaTab: A plug-and-play model for learning tabular data. In *Proceedings of the 2024 IEEE 7th International Conference on Multimedia Information Processing and Retrieval*, pp. 369–375, 2024.

Nahla Mohamed Ahmed, Ling Chen, Yulong Wang, Bin Li, Yun Li, and Wei Liu. Sampling-based algorithm for link prediction in temporal networks. *Information Sciences*, 374:1–14, 2016.

Mohammad Alomrani, Mahdi Biparva, Yingxue Zhang, and Mark Coates. Dyg2vec: Efficient representation learning for dynamic graphs. *Transactions on Machine Learning Research*, 2023.

Rami Aly, Zhijiang Guo, Michael Sejr Schlichtkrull, James Thorne, Andreas Vlachos, Christos Christodoulopoulos, Oana Cocarascu, and Arpit Mittal. FEVEROUS: Fact extraction and verification over unstructured and structured information. In *Advances in Neural Information Processing Systems*, 2021.

Abdul Fatir Ansari, Lorenzo Stella, Ali Caner Turkmen, Xiyuan Zhang, Pedro Mercado, Huibin Shen, Oleksandr Shchur, Syama Sundar Rangapuram, Sebastian Pineda Arango, Shubham Kapoor, Jasper Zschiegner, Danielle C. Maddix, Hao Wang, Michael W. Mahoney, Kari Torkkola, Andrew Gordon Wilson, Michael Bohlke-Schneider, and Bernie Wang. Chronos: Learning the language of time series. *Transactions on Machine Learning Research*, 2024.

Sercan Ö. Arik and Tomas Pfister. TabNet: Attentive interpretable tabular learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 6679–6687, 2021.

Wonho Bae, Mohamed Osama Ahmed, Frederick Tung, and Gabriel L. Oliveira. Meta temporal point processes. In *International Conference on Learning Representations*, pp. 1–14, 2023.

Dara Bahri, Heinrich Jiang, Yi Tay, and Donald Metzler. SCARF: Self-supervised contrastive learning using random feature corruption. In *Proceedings of the Tenth International Conference on Learning Representations*, pp. 1–14, 2022.

Seunghwan Bang and Hwanjun Song. LLM-based user profile management for recommender system. *arXiv preprint arXiv:2502.14541*, 2025.

Keqin Bao, Jizhi Zhang, Yang Zhang, Wenjie Wang, Fuli Feng, and Xiangnan He. TALLRec: An effective and efficient tuning framework to align large language model with recommendation. In *Proceedings of the 17th ACM Conference on Recommender Systems*, pp. 1007–1014, 2023.

Anson Bastos, Abhishek Nadgeri, Kuldeep Singh, Toyotaro Suzumura, and Manish Singh. Learnable spectral wavelets on dynamic graphs to capture global interactions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 6779–6787, 2023.

Inci M. Baytas, Cao Xiao, Xi Zhang, Fei Wang, Anil K. Jain, and Jiayu Zhou. Patient subtyping via time-aware LSTM networks. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 65–74, 2017.

Yoshua Bengio. Deep learning of representations for unsupervised and transfer learning. In *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*, volume 27, pp. 17–36, 2012.

Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828, 2013.

Yoshua Bengio, Yann LeCun, and Geoffrey Hinton. Deep learning for AI. *Communications of the ACM*, 64 (7):58–65, 2021.

Kushal Raj Bhandari, Sixue Xing, Soham Dan, and Jianxi Gao. Exploring the robustness of language models for tabular question answering via attention analysis. *Transactions on Machine Learning Research*, 2025.

Homanga Bharadhwaj, Homin Park, and Brian Y. Lim. RecGAN: Recurrent generative adversarial networks for recommendation systems. In *Proceedings of the 12th ACM Conference on Recommender Systems*, pp. 372–376, 2018.

Y Bi, A Chadha, A Abbas, E Bourtsoulatze, and Y Andreopoulos. Graph-based object classification for neuromorphic vision sensing. In *2019 IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2019.

Mahdi Biparva, Raika Karimi, Faezeh Faez, and Yingxue Zhang. Todyformer: Towards holistic dynamic graph transformers with structure-aware tokenization. *Transactions on Machine Learning Research*, 2024.

Bernd Bischl, Giuseppe Casalicchio, Matthias Feurer, Pieter Gijsbers, Frank Hutter, Michel Lang, Rafael G. Mantovani, Jan N. van Rijn, and Joaquin Vanschoren. OpenML benchmarking suites. In *Advances in Neural Information Processing Systems*, 2021.

Stephen Bonner, Amir Atapour-Abarghouei, Philip T. Jackson, John Brennan, Ibad Kureshi, Georgios Theodoropoulos, Andrew Stephen McGough, and Boguslaw Obara. Temporal neighbourhood aggregation: Predicting future links in temporal graphs via recurrent variational graph convolutions. In *Proceedings of the 2019 IEEE International Conference on Big Data*, pp. 5336–5345, 2019.

Vadim Borisov, Kathrin Seßler, Tobias Leemann, Martin Pawelczyk, and Gjergji Kasneci. Language models are realistic tabular data generators. In *International Conference on Learning Representations*, 2023.

Bernardo Branco, Pedro Abreu, Ana Sofia Gomes, Mariana S. C. Almeida, João Tiago Ascensão, and Pedro Bizarro. Interleaved sequence RNNs for fraud detection. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 3101–3109, 2020.

Jeffrey A. Brooks, Vineet Tiruvadi, Alice Baird, Panagiotis Tzirakis, Haoqi Li, Chris Gagne, Moses Oh, and Alan Cowen. Emotion expression estimates to measure and improve multimodal social-affective interactions. In *Companion Publication of the 25th International Conference on Multimodal Interaction*, pp. 353–358, 2023.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, volume 33, pp. 1877–1901, 2020.

Luca Butera, Giovanni De Felice, Andrea Cini, and Cesare Alippi. On the regularization of learnable embeddings for time series forecasting. *Transactions on Machine Learning Research*, 2025.

Shaofeng Cai, Kaiping Zheng, Gang Chen, H. V. Jagadish, Beng Chin Ooi, and Meihui Zhang. ARM-net: Adaptive relation modeling network for structured data. In *Proceedings of the 2021 International Conference on Management of Data*, pp. 207–220, 2021.

Zi Cai, Yu Liu, Zhiyao Luo, and Tingting Zhu. ProtoEHR: Hierarchical prototype learning for EHR-based healthcare predictions. *arXiv preprint arXiv:2508.18313*, 2025.

Guiming Cao, Kaize Shi, Hong Fu, Huaiwen Zhang, and Guandong Xu. APLe: Token-wise adaptive for multi-modal prompt learning. *arXiv preprint arXiv:2401.06827*, 2024a.

Yupeng Cao, Zhi Chen, Prashant Kumar, Qingyun Pei, Yangyang Yu, Haohang Li, Fabrizio Dimino, Lorenzo Ausiello, K. P. Subbalakshmi, and Papa Momar Ndiaye. RiskLabs: Predicting financial risk using large language model based on multimodal and multi-sources data. *arXiv preprint arXiv:2404.07452*, 2024b.

Aldo Gael Carranza, Rezsa Farahani, Natalia Ponomareva, Alexey Kurakin, Matthew Jagielski, and Milad Nasr. Synthetic query generation for privacy-preserving deep retrieval systems using differentially private language models. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pp. 3920–3930, 2024.

Deepayan Chakrabarti, Ravi Kumar, and Andrew Tomkins. Evolutionary clustering. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 554–560, 2006.

Yupeng Chang, Xu Wang, Jindong Wang, Yuan Wu, Linyi Yang, Kaijie Zhu, Hao Chen, Xiaoyuan Yi, Cunxiang Wang, Yidong Wang, Wei Ye, Yue Zhang, Yi Chang, Philip S. Yu, Qiang Yang, and Xing Xie. A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology*, 15(3):39:1–39:45, 2024.

Krishan Chavinda and Uthayasanker Thayasivam. A dual contrastive learning framework for enhanced hate speech detection in low-resource languages. In *Proceedings of the First Workshop on Challenges in Processing South Asian Languages*, pp. 115–123, 2025.

Chao Chen, Haoyu Geng, Nianzu Yang, Xiaokang Yang, and Junchi Yan. EasyDGL: Encode, train and interpret for continuous-time dynamic graph learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024a.

Deli Chen, Yankai Lin, Wei Li, Peng Li, Jie Zhou, and Xu Sun. Measuring and relieving the over-smoothing problem for graph neural networks from the topological view. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 3438–3445, 2020.

Jiazhen Chen, Sichao Fu, Zheng Ma, Mingbin Feng, Tony S. Wirjanto, and Qinmu Peng. Semi-supervised anomaly detection with extremely limited labels in dynamic graphs. In *Proceedings of the 30th International Conference on Database Systems for Advanced Applications*, volume 15987, pp. 373–382, 2025a.

Jinyin Chen, Jian Zhang, Xuanheng Xu, Chenbo Fu, Dan Zhang, Qingpeng Zhang, and Qi Xuan. E-LSTM-D: A deep learning framework for dynamic network link prediction. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 51(6):3699–3712, 2021.

Jinyin Chen, Xueke Wang, and Xuanheng Xu. GC-LSTM: graph convolution embedded LSTM for dynamic network link prediction. *Applied Intelligence*, 52(7):7513–7528, 2022.

Junyi Chen. A survey on large language models for personalized and explainable recommendations. *arXiv preprint arXiv:2311.12338*, 2023.

Kuan-Yu Chen, Ping-Han Chiang, Hsin-Rung Chou, Ting-Wei Chen, and Tien-Hao Chang. Trompt: Towards a better deep neural network for tabular data. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202, pp. 4392–4434, 2023a.

Mengyang Chen, Lingwei Wei, Han Cao, Wei Zhou, and Songlin Hu. Explore the potential of LLMs in misinformation detection: An empirical study. *arXiv preprint arXiv:2311.12699*, 2024b.

Pei Chen, Soumajyoti Sarkar, Leonard Lausen, Balasubramaniam Srinivasan, Sheng Zha, Ruihong Huang, and George Karypis. HyTrel: Hypergraph-enhanced tabular data representation learning. In *Advances in Neural Information Processing Systems*, volume 36, pp. 32173–32193, 2023b.

Qing Chen, Wei Shuai, Jiyao Zhang, Zhida Sun, and Nan Cao. Beyond numbers: Creating analogies to enhance data comprehension and communication with generative AI. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2024c.

Suiyao Chen, Jing Wu, Naira Hovakimyan, and Handong Yao. ReConTab: Regularized contrastive representation learning for tabular data. In *Table Representation Learning Workshop, Advances in Neural Information Processing Systems 2023*, 2023c.

Tianqi Chen and Carlos Guestrin. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016.

Wenhu Chen, Xueguang Ma, Xinyi Wang, and William W. Cohen. Program of thoughts prompting: Disentangling computation from reasoning for numerical reasoning tasks. *Transactions on Machine Learning Research*, 2023d.

Xi Chen, Yongxiang Liao, Yun Xiong, Yao Zhang, Siwei Zhang, Jiawei Zhang, and Yiheng Sun. SPEED: Streaming partition and parallel acceleration for temporal interaction graph embedding. *arXiv preprint arXiv:2308.14129*, 2023e.

Xi Chen, Yun Xiong, Siwei Zhang, Jiawei Zhang, Yao Zhang, Shiyang Zhou, Xixi Wu, Mingyang Zhang, Tengfei Liu, and Weiqiang Wang. DTFormer: A transformer-based method for discrete-time dynamic graph representation learning. *arXiv preprint arXiv:2407.18523*, 2024d.

Xi Chen, Siwei Zhang, Yun Xiong, Xixi Wu, Jiawei Zhang, Xiangguo Sun, Yao Zhang, Feng Zhao, and Yulin Kang. Prompt learning on temporal interaction graphs. *arXiv preprint arXiv:2402.06326*, 2024e.

Xi Chen, Yateng Tang, Jiarong Xu, Jiawei Zhang, Siwei Zhang, Sijia Peng, Xuehao Zheng, and Yun Xiong. Rethinking time encoding via learnable transformation functions. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267, pp. 9073–9104, 2025b.

Xu Chen, Hongteng Xu, Yongfeng Zhang, Jiaxi Tang, Yixin Cao, Zheng Qin, and Hongyuan Zha. Sequential recommendation with user memory networks. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pp. 108–116, 2018.

Yizhou Chen, Anxiang Zeng, Qingtao Yu, Kerui Zhang, Yuanpeng Cao, Kangle Wu, Guangda Huzhang, Han Yu, and Zhiming Zhou. Recurrent temporal revision graph networks. In *Advances in Neural Information Processing Systems*, volume 36, pp. 69348–69360, 2023f.

Chen Cheng, Haiqin Yang, Michael R. Lyu, and Irwin King. Where you like to go next: Successive point-of-interest recommendation. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*, pp. 2605–2611, 2013.

Heng-Tze Cheng, Levent Koc, Jeremiah Harmsen, Tal Shaked, Tushar Chandra, Hrishi Aradhye, Glen Anderson, Greg Corrado, Wei Chai, Mustafa Ispir, Rohan Anil, Zakaria Haque, Lichan Hong, Vihan Jain, Xiaobing Liu, and Hemal Shah. Wide & deep learning for recommender systems. In *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems*, pp. 7–10, 2016.

Ke Cheng, Linzhi Peng, Junchen Ye, Leilei Sun, and Bowen Du. Co-neighbor encoding schema: A light-cost structure encoding method for dynamic link prediction. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 421–432, 2024.

Lu Cheng, Ahmadreza Mosallanezhad, Yasin N. Silva, Deborah L. Hall, and Huan Liu. Bias mitigation for toxicity detection via sequential decisions. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 1750–1760, 2022.

Ta-Chung Chi, Ting-Han Fan, Peter J. Ramadge, and Alexander I. Rudnicky. KERPLE: Kernelized relative positional embedding for length extrapolation. In *Advances in Neural Information Processing Systems*, volume 35, pp. 8386–8399, 2022.

Yun Chi, Xiaodan Song, Dengyong Zhou, Koji Hino, and Belle L. Tseng. On evolutionary spectral clustering. *ACM Transactions on Knowledge Discovery from Data*, 3(4):1–30, 2009.

Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pp. 1724–1734, 2014.

Edward Choi, Mohammad Taha Bahadori, Joshua A. Kulas, Andy Schuetz, Walter F. Stewart, and Jimeng Sun. RETAIN: An interpretable predictive model for healthcare using reverse time attention mechanism. In *Advances in Neural Information Processing Systems*, volume 29, pp. 3512–3520, 2016a.

Edward Choi, Mohammad Taha Bahadori, Andy Schuetz, Walter F. Stewart, and Jimeng Sun. Doctor AI: Predicting clinical events via recurrent neural networks. In *Proceedings of the 1st Machine Learning for Healthcare Conference*, volume 56, pp. 301–318, 2016b.

Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, Parker Schuh, Kensen Shi, Sasha Tsvyashchenko, Joshua Maynez, Abhishek Rao, Parker Barnes, Yi Tay, Noam Shazeer, Vinodkumar Prabhakaran, Emily Reif, Nan Du, Ben Hutchinson, Reiner Pope, James Bradbury, Jacob Austin, Michael Isard, Guy Gur-Ari, Pengcheng Yin, Toju Duke, Anselm Levskaya, Sanjay Ghemawat, Sunipa Dev, Henryk Michalewski, Xavier Garcia, Vedant Misra, Kevin Robinson, Liam Fedus, Denny Zhou, Daphne Ippolito, David Luan, Hyeontaek Lim, Barret Zoph, Alexander Spiridonov, Ryan Sepassi, David Dohan, Shivani Agrawal, Mark Omernick, Andrew M. Dai, Thanumalayan Sankaranarayana Pillai, Marie Pellat, Aitor Lewkowycz, Erica Moreira, Rewon Child, Oleksandr Polozov, Katherine Lee, Zongwei Zhou, Xuezhi Wang, Brennan Saeta, Mark Diaz, Orhan Firat, Michele Catasta, Jason Wei, Kathy Meier-Hellstern, Douglas Eck, Jeff Dean, Slav Petrov, and Noah Fiedel. PaLM: Scaling language modeling with pathways. *Journal of Machine Learning Research*, 24(240):1–113, 2023.

Ranak Roy Chowdhury, Jiacheng Li, Xiyuan Zhang, Dezhi Hong, Rajesh K. Gupta, and Jingbo Shang. PrimeNet: Pre-training for irregular multivariate time series. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 7184–7192, 2023.

Hsing-Huan Chung, Shravan S Chaudhari, Xing Han, Yoav Wald, Suchi Saria, and Joydeep Ghosh. Between linear and sinusoidal: Rethinking the time encoder in dynamic graph learning. *Transactions on Machine Learning Research*, 2025.

Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*, 2014.

Kenneth Ward Church. Word2vec. *Natural Language Engineering*, 23(1):155–162, 2017.

Jillian M. Clements, Di Xu, Nooshin Yousefi, and Dmitry Efimov. Sequential deep learning for credit risk monitoring with tabular financial data. *arXiv preprint arXiv:2012.15330*, 2020.

Weilin Cong, Si Zhang, Jian Kang, Baichuan Yuan, Hao Wu, Xin Zhou, Hanghang Tong, and Mehrdad Mahdavi. Do we really need complicated model architectures for temporal networks? In *Proceedings of the Eleventh International Conference on Learning Representations*, pp. 1–14, 2023.

Sonia Cromp, Satya Sai Srinath Namburi GNVV, Mohammed Alkhudhayri, Catherine Cao, Samuel Guo, Nicholas Roberts, and Frederic Sala. Tabby: A language model architecture for tabular and structured data synthesis. *Transactions on Machine Learning Research*, 2026.

Shiyao Cui, Qinglin Zhang, Xuan Ouyang, Renmiao Chen, Zhexin Zhang, Yida Lu, Hongning Wang, Han Qiu, and Minlie Huang. ShieldVLM: Safeguarding the multimodal implicit toxicity via deliberative reasoning with LVLMs. *arXiv preprint arXiv:2505.14035*, 2025.

Zeyu Cui, Jianxin Ma, Chang Zhou, Jingren Zhou, and Hongxia Yang. M6-Rec: Generative pretrained language models are open-ended recommender systems. *arXiv preprint arXiv:2205.08084*, 2022.

Hanjun Dai, Yichen Wang, Rakshit Trivedi, and Le Song. Recurrent coevolutionary latent feature processes for continuous-time recommendation. In *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems*, pp. 29–34, 2016.

Zihang Dai, Zhilin Yang, Yiming Yang, Jaime Carbonell, Quoc Le, and Ruslan Salakhutdinov. Transformer-XL: Attentive language models beyond a fixed-length context. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 2978–2988, 2019.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171–4186, 2019.

Zifeng Ding, Yifeng Li, Yuan He, Antonio Norelli, Jingcheng Wu, Volker Tresp, Michael M. Bronstein, and Yunpu Ma. Dygmamba: Efficiently modeling long-term temporal dependency on continuous-time dynamic graphs with state space models. *Transactions on Machine Learning Research*, 2025.

Xishuang Dong and Lijun Qian. Semi-supervised bidirectional RNN for misinformation detection. *Machine Learning with Applications*, 10:100428, 2022.

Nan Du, Hanjun Dai, Rakshit Trivedi, Utkarsh Upadhyay, Manuel Gomez-Rodriguez, and Le Song. Recurrent marked temporal point processes: Embedding event history to vector. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1555–1564, 2016.

Ahmad Naser Eddin, Jacopo Bono, David Oliveira Aparicio, Hugo Ferreira, Pedro Manuel Pinto Ribeiro, and Pedro Bizarro. Deep-graph-sprints: Accelerated representation learning in continuous-time dynamic graphs. *Transactions on Machine Learning Research*, 2024.

Sean R Eddy. Hidden Markov models. *Current Opinion in Structural Biology*, 6(3):361–365, 1996.

Sofiane ENNADIR, Gabriela Zarzar Gandler, Filip Cornell, Lele Cao, Oleg Smirnov, Tianze Wang, Levente Zólyomi, Björn Brinne, and Sahar Asadi. Expressivity of representation learning on continuous-time dynamic graphs: An information-flow centric review. *Transactions on Machine Learning Research*, 2025.

Linxi Fan, Guanzhi Wang, Yunfan Jiang, Ajay Mandlekar, Yuncong Yang, Haoyi Zhu, Andrew Tang, De-An Huang, Yuke Zhu, and Anima Anandkumar. Minedojo: Building open-ended embodied agents with Internet-scale knowledge. *arXiv preprint arXiv:2206.08853*, 2022.

Lanting Fang, Kaiyu Feng, Jie Gui, Shanshan Feng, and Aiqun Hu. Anonymous edge representation for inductive anomaly detection in dynamic bipartite graphs. *Proceedings of the VLDB Endowment*, 16(5): 1154–1167, 2023.

Xi Fang, Weijie Xu, Fiona Anting Tan, Ziqing Hu, Jiani Zhang, Yanjun Qi, Srinivasan H. Sengamedu, and Christos Faloutsos. Large language models (llms) on tabular data: Prediction, generation, and understanding - a survey. *Transactions on Machine Learning Research*, 2024.

Golnoosh Farnadi, Lise Getoor, Marie-Francine Moens, and Martine De Cock. User profiling using hinge-loss markov random fields. *arXiv preprint arXiv:2001.01177*, 2020.

Kaituo Feng, Changsheng Li, Xiaolu Zhang, and Jun Zhou. Towards open temporal graph neural networks. In *The Eleventh International Conference on Learning Representations*, pp. 1–14, 2023.

Shanshan Feng, Xutao Li, Yifeng Zeng, Gao Cong, Yeow Meng Chee, and Quan Yuan. Personalized ranking metric embedding for next new POI recommendation. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence*, pp. 2069–2075, 2015.

ZhengZhao Feng, Rui Wang, TianXing Wang, Mingli Song, Sai Wu, and Shuibing He. A comprehensive survey of dynamic graph neural networks: Models, frameworks, benchmarks, experiments and challenges. *IEEE Transactions on Knowledge and Data Engineering*, 38(1):26–46, 2025.

Stefan Feuerriegel, Abdurahman Maarouf, Dominik Bär, Dominique Geissler, Jonas Schweisthal, Nicolas Pröllochs, Claire E. Robertson, Steve Rathje, Jochen Hartmann, Saif M. Mohammad, Oded Netzer, Alexandra A. Siegel, Barbara Plank, and Jay J. Van Bavel. Using natural language processing to analyse text data in behavioural science. *Nature Reviews Psychology*, 4(2):96–111, 2025.

Jerome H. Friedman. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5):1189–1232, 2001.

Luke Friedman, Sameer Ahuja, David Allen, Zhenning Tan, Hakim Sidahmed, Changbo Long, Jun Xie, Gabriel Schubiner, Ajay Patel, Harsh Lara, Brian Chu, Zexi Chen, and Manoj Tiwari. Leveraging large language models in conversational recommender systems. *arXiv preprint arXiv:2305.07961*, 2023.

Yao Fu, Hao Peng, and Tushar Khot. How does GPT obtain its ability? tracing emergent abilities of language models to their sources, 2022. Accessed: 2026-01-04.

Zichuan Fu, Xiangyang Li, Chuhan Wu, Yichao Wang, Kuicai Dong, Xiangyu Zhao, Mengchen Zhao, Huifeng Guo, and Ruiming Tang. A unified framework for multi-domain CTR prediction via large language models. *ACM Transactions on Information Systems*, 43(5):117:1–117:33, 2025.

Nan Gao, Zhuolei Yu, Yue Xu, Chun Yu, Yuntao Wang, Flora D. Salim, and Yuanchun Shi. Leveraging large language models for generating mobile sensing strategies in human behavior modeling. In *Companion of the 2024 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 729–735, 2024a.

Shen Gao, Jiabao Fang, Quan Tu, Zhitao Yao, Zhumin Chen, Pengjie Ren, and Zhaochun Ren. Generative news recommendation. In *Proceedings of the ACM Web Conference 2024*, pp. 3444–3453, 2024b.

Yunfan Gao, Tao Sheng, Youlin Xiang, Yun Xiong, Haofen Wang, and Jiawei Zhang. Chat-rec: Towards interactive and explainable LLMs-augmented recommender system. *arXiv preprint arXiv:2303.14524*, 2023a.

Yunfan Gao, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yi Dai, Jiawei Sun, Meng Wang, and Haofen Wang. Retrieval-augmented generation for large language models: A survey. *arXiv preprint arXiv:2312.10997*, 2023b.

Zhenxiang Gao, Lingyao Li, Siyuan Ma, Qinyong Wang, Libby Hemphill, and Rong Xu. Examining the potential of ChatGPT on biomedical information retrieval: Fact-checking drug-disease associations. *Annals of Biomedical Engineering*, 52(8):1919–1927, 2024c.

Julia Gastinger, Shenyang Huang, Mikhail Galkin, Erfan Loghmani, Ali Parviz, Farimah Poursafaei, Jacob Danovitch, Emanuele Rossi, Ioannis Koutis, Heiner Stuckenschmidt, Reihaneh Rabbany, and Guillaume Rabusseau. TGB 2.0: A benchmark for learning on temporal knowledge graphs and heterogeneous graphs. In *Advances in Neural Information Processing Systems*, volume 37, pp. 140199–140229, 2024.

Gemini Team. Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. *arXiv preprint arXiv:2403.05530*, 2024.

Shijie Geng, Shuchang Liu, Zuohui Fu, Yingqiang Ge, and Yongfeng Zhang. Recommendation as language processing (RLP): A unified pretrain, personalized prompt & predict paradigm (P5). In *Proceedings of the Sixteenth ACM Conference on Recommender Systems*, 2022.

Marvin Ggaliwango, Hellen Nakayiza, Daudi Jjingo, and Joyce Nakatumba-Nabende. Prompt engineering in large language models. In *Proceedings of the International Conference on Data Intelligence and Cognitive Informatics*, pp. 387–402, 2024.

Hussam Ghanem and Christophe Cruz. Fine-tuning vs. prompting: Evaluating the knowledge graph construction with LLMs. In *Proceedings of the 3rd International Workshop on Knowledge Graph Generation from Text, co-located with the Extended Semantic Web Conference*, volume 3747, 2024.

Hussam Ghanem and Christophe Cruz. Enhancing knowledge graph construction: Evaluating with emphasis on hallucination, omission, and graph similarity metrics. In *Knowledge Graphs and Semantic Web: 6th International Conference, KGSWC 2024, Paris, France, December 11–13, 2024, Proceedings*, volume 15459, pp. 32–46, 2025.

Alex Gomez-Marin, Joseph J. Paton, Adam R. Kampff, Rui M. Costa, and Zachary F. Mainen. Big behavioral data: psychology, ethology and the foundations of neuroscience. *Nature Neuroscience*, 17(11):1455–1462, 2014.

Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.

Yury Gorishniy, Ivan Rubachev, Valentin Khrulkov, and Artem Babenko. Revisiting deep learning models for tabular data. In *Advances in Neural Information Processing Systems*, volume 34, pp. 18932–18943, 2021.

Yury Gorishniy, Ivan Rubachev, Nikolay Kartashev, Daniil Shlenskii, Akim Kotelnikov, and Artem Babenko. TabR: Tabular deep learning meets nearest neighbors. In *The Twelfth International Conference on Learning Representations*, pp. 1–14, 2024.

Palash Goyal, Nitin Kamra, Xinran He, and Yan Liu. DynGEM: Deep embedding method for dynamic graphs. *arXiv preprint arXiv:1805.11273*, 2018.

Palash Goyal, Sujit Rokka Chhetri, and Arquimedes Canedo. dyngraph2vec: Capturing network dynamics using dynamic graph representation learning. *Knowledge-Based Systems*, 187:104816, 2020.

Alex Graves. Long short-term memory. In *Supervised Sequence Labelling with Recurrent Neural Networks*, volume 385, pp. 37–45. 2012.

Alessio Gravina and Davide Bacciu. Deep learning for dynamic graphs: Models and benchmarks. *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–14, 2024.

Léo Grinsztajn, Edouard Oyallon, and Gaël Varoquaux. Why do tree-based models still outperform deep learning on typical tabular data? In *Advances in Neural Information Processing Systems*, volume 35, pp. 507–520, 2022.

Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 855–864, 2016.

Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. In *Proceedings of the Conference on Language Modeling*, 2024.

Yulong Gu, Zhuoye Ding, Shuaiqiang Wang, and Dawei Yin. Hierarchical user profiling for E-commerce recommender systems. In *Proceedings of the Thirteenth ACM International Conference on Web Search and Data Mining*, pp. 223–231, 2020.

Zehao Gu, Yateng Tang, Jiarong Xu, Siwei Zhang, Xuehao Zheng, Xi Chen, and Yun Xiong. SSH-T3: A hierarchical pre-training framework for multi-scenario financial risk assessment. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management*, pp. 5674–5682, 2025.

Huifeng Guo, Ruiming Tang, Yunming Ye, Zhenguo Li, and Xiuqiang He. DeepFM: A factorization-machine based neural network for CTR prediction. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, pp. 1725–1731, 2017.

Anjali Gupta, Prashant Kumar, Aniket Mishra, Abhishek Singh, Surender Kumar, Muthusamy Chelliah, Abhijnan Chakraborty, and Sayan Ranu. Persona identification in E-Commerce with scarce labels and in-context graph learning. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, volume 2, pp. 778–789, 2025.

Manish Gupta, Charu C. Aggarwal, Jiawei Han, and Yizhou Sun. Evolutionary clustering and analysis of bibliographic networks. In *Proceedings of the 2011 International Conference on Advances in Social Networks Analysis and Mining*, pp. 63–70, 2011.

Ehsan Hajiramezanali, Arman Hasanzadeh, Krishna Narayanan, Nick Duffield, Mingyuan Zhou, and Xiaoning Qian. Variational graph recurrent neural networks. *Advances in Neural Information Processing Systems*, 32:10700–10710, 2019.

William L. Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems*, volume 30, pp. 1024–1034, 2017.

Chenyu Han and Xiaoyu Fu. Challenge and opportunity: Deep learning-based stock price prediction by using bi-directional LSTM model. *Frontiers in Business, Economics and Management*, 8(2):51–54, 2023.

Jiaming Han, Kaixiong Gong, Yiyuan Zhang, Jiaqi Wang, Kaipeng Zhang, Dahua Lin, Yu Qiao, Peng Gao, and Xiangyu Yue. OneLLM: One framework to align all modalities with language. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 26584–26595, 2024.

Liangzhe Han, Ruixing Zhang, Leilei Sun, Bowen Du, Yanjie Fu, and Tongyu Zhu. Generic and dynamic graph representation learning for crowd flow modeling. In *Proceedings of the 37th AAAI Conference on Artificial Intelligence*, volume 37, pp. 4293–4301, 2023.

Sungwon Han, Seungeon Lee, Meeyoung Cha, Sercan O Arik, and Jinsung Yoon. Llm-guided self-supervised tabular learning with task-specific pre-text tasks. *Transactions on Machine Learning Research*, 2025.

Zhen Han, Yunpu Ma, Yuyi Wang, Stephan Günnemann, and Volker Tresp. Graph hawkes neural network for forecasting on temporal knowledge graphs. In *Automated Knowledge Base Construction*, 2020.

F. Maxwell Harper and Joseph A. Konstan. The MovieLens datasets: History and context. *ACM Transactions on Interactive Intelligent Systems*, 5(4):19:1–19:19, 2015.

Pengcheng He, Xiaodong Liu, Jianfeng Gao, and Weizhu Chen. DeBERTa: Decoding-enhanced BERT with disentangled attention. In *International Conference on Learning Representations*, pp. 1–20, 2021.

Ruining He, Chen Fang, Zhaowen Wang, and Julian J. McAuley. Vista: A visually, socially, and temporally-aware model for artistic recommendation. In *Proceedings of the 10th ACM Conference on Recommender Systems*, pp. 309–316, 2016.

Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. Neural collaborative filtering. In *Proceedings of the 26th ACM World Wide Web Conference*, pp. 173–182, 2017.

Xinran He, Junfeng Pan, Ou Jin, Tianbing Xu, Bo Liu, Tao Xu, Yanxin Shi, Antoine Atallah, Ralf Herbrich, Stuart Bowers, and Joaquin Quiñonero Candela. Practical lessons from predicting clicks on ads at Facebook. In *Proceedings of the Eighth International Workshop on Data Mining for Online Advertising*, pp. 1–9, 2014.

Zhicheng He, Weiwen Liu, Wei Guo, Jiarui Qin, Yingxue Zhang, Yaochen Hu, and Ruiming Tang. A survey on user behavior modeling in recommender systems. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, pp. 6656–6664, 2023.

Stefan Hegselmann, Alejandro Buendia, Hunter Lang, Monica Agrawal, Xiaoyi Jiang, and David Sontag. TabLLM: Few-shot classification of tabular data with large language models. In *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, volume 206, pp. 5549–5581, 2023.

Jonathan Herzig, Pawel Krzysztof Nowak, Thomas Müller, Francesco Piccinno, and Julian Eisenschlos. TaPas: Weakly supervised table parsing via pre-training. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 4320–4333, 2020.

Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, and Domonkos Tikk. Session-based recommendations with recurrent neural networks. In *Proceedings of the 4th International Conference on Learning Representations*, 2016.

Ryohei Hisano. Semi-supervised graph embedding approach to dynamic link prediction. In *Complex Networks IX: Proceedings of the 9th Conference on Complex Networks*, pp. 109–121, 2018.

Shenda Hong, Meng Wu, Hongyan Li, and Zhengwu Wu. Event2vec: Learning representations of events on temporal sequences. In *Web and Big Data: First International Joint Conference, APWeb-WAIM 2017, Beijing, China, July 7–9, 2017, Proceedings, Part II*, volume 10367, pp. 33–47, 2017.

Yupeng Hou, Shanlei Mu, Wayne Xin Zhao, Yaliang Li, Bolin Ding, and Ji-Rong Wen. Towards universal sequence representation learning for recommender systems. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 585–593, 2022.

Yupeng Hou, Jiacheng Li, Zhankui He, An Yan, Xiusi Chen, and Julian McAuley. Bridging language and items for retrieval and recommendation. *arXiv preprint arXiv:2403.03952*, 2024.

Junfeng Hu, Yuxuan Liang, Zhencheng Fan, Hongyang Chen, Yu Zheng, and Roger Zimmermann. Graph neural processes for spatio-temporal extrapolation. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 752–763, 2023a.

Shirui Hu, Weichang Wu, Zuoli Tang, Zhaoxin Huan, Lin Wang, Xiaolu Zhang, Jun Zhou, Lixin Zou, and Chenliang Li. HORAE: Temporal multi-interest pre-training for sequential recommendation. *ACM Transactions on Information Systems*, 43(4):88:1–88:29, 2025.

Sihao Hu, Zhen Zhang, Bingqiao Luo, Shengliang Lu, Bingsheng He, and Ling Liu. BERT4ETH: A pre-trained transformer for ethereum fraud detection. In *Proceedings of the ACM Web Conference 2023*, pp. 2189–2197, 2023b.

Allen H. Huang, Hui Wang, and Yi Yang. Finbert: A large language model for extracting information from financial text. *Contemporary Accounting Research*, 40(2):806–841, 2023a.

Chengkai Huang, Hongtao Huang, Tong Yu, Kaige Xie, Junda Wu, Shuai Zhang, Julian Mcauley, Dietmar Jannach, and Lina Yao. A survey of foundation model-powered recommender systems: From feature-based, generative to agentic paradigms. *arXiv preprint arXiv:2504.16420*, 2025a.

Qiang Huang, Xin Wang, Susie Xi Rao, Zhichao Han, Zitao Zhang, Yongjun He, Quanqing Xu, Yang Zhao, Zhigao Zheng, and Jiawei Jiang. Benchtemp: A general benchmark for evaluating temporal graph neural networks. In *Proceedings of the 2024 IEEE 40th International Conference on Data Engineering*, pp. 4044–4057, 2024a.

Shenyang Huang, Farimah Poursafaei, Jacob Danovitch, Matthias Fey, Weihua Hu, Emanuele Rossi, Jure Leskovec, Michael Bronstein, Guillaume Rabusseau, and Reihaneh Rabbany. Temporal graph benchmark for machine learning on temporal graphs. In *Advances in Neural Information Processing Systems*, volume 36, pp. 2056–2073, 2023b.

Shenyang Huang, Farimah Poursafaei, Reihaneh Rabbany, Guillaume Rabusseau, and Emanuele Rossi. UTG: Towards a unified view of snapshot and event based models for temporal graphs. In *Proceedings of the Third Learning on Graphs Conference*, volume 269, pp. 28:1–28:16, 2025b.

Shuaibin Huang, Yun Xiong, Yi Xie, Tianyu Qiu, and Guangzhong Wang. Robust sequence-based self-supervised representation learning for anti-money laundering. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, pp. 4571–4578, 2024b.

Xin Huang, Ashish Khetan, Milan Cvitkovic, and Zohar Karnin. Tabtransformer: Tabular data modeling using contextual embeddings. *arXiv preprint arXiv:2012.06678*, 2020a.

Xuanwen Huang, Yang Yang, Yang Wang, Chunping Wang, Zhisheng Zhang, Jiarong Xu, Lei Chen, and Michalis Vazirgiannis. DGraph: A large-scale financial dataset for graph anomaly detection. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

Zhiheng Huang, Davis Liang, Peng Xu, and Bing Xiang. Improve transformer models with better relative position embeddings. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pp. 3327–3335, 2020b.

Nahla Mohamed Ahmed Ibrahim and Ling Chen. Link prediction in dynamic social networks by integrating different types of information. *Applied Intelligence*, 42(4):738–750, 2015.

Nahla Mohamed Ahmed Ibrahim and Ling Chen. An efficient algorithm for link prediction in temporal uncertain social networks. *Information Sciences*, 331:120–136, 2016.

Hiroshi Iida, Dung Thai, Varun Manjunatha, and Mohit Iyyer. TABBIE: Pretrained representations of tabular data. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 3446–3456, 2021.

Sergei Ivanov and Liudmila Prokhorenkova. Boost then convolve: Gradient boosting meets graph neural networks. In *Proceedings of the 9th International Conference on Learning Representations*, 2021.

Shuo Ji, Xiaodong Lu, Mingzhe Liu, Leilei Sun, Chuanren Liu, Bowen Du, and Hui Xiong. Community-based dynamic graph learning for popularity prediction. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 930–940, 2023.

Shuo Ji, Mingzhe Liu, Leilei Sun, Chuanren Liu, and Tongyu Zhu. MemMap: An adaptive and latent memory structure for dynamic graph learning. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 1257–1268, 2024.

Julie Jiang and Emilio Ferrara. Social-LLM: Modeling user behavior at scale using language models and social network data. *Sci*, 7(4):138, 2025.

Jun-Peng Jiang, Si-Yang Liu, Hao-Run Cai, Qi-Le Zhou, and Han-Jia Ye. Representation learning for tabular data: A comprehensive survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2026.

Liming Jiang. Detecting scams using large language models. *arXiv preprint arXiv:2402.03147*, 2024.

Pengcheng Jiang, Cao Xiao, Zifeng Wang, Parminder Bhatia, Jimeng Sun, and Jiawei Han. TriSum: Learning summarization ability from large language models with structured rationale. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pp. 2805–2819, 2024.

Pengcheng Jiang, Cao Xiao, Minhao Jiang, Parminder Bhatia, Taha Kass-Hout, Jimeng Sun, and Jiawei Han. Reasoning-enhanced healthcare predictions with knowledge graph community retrieval. In *Proceedings of the 13th International Conference on Learning Representations*, pp. 1–46, 2025.

Renhe Jiang, Zhaonan Wang, Jiawei Yong, Puneet Jeph, Quanjun Chen, Yasumasa Kobayashi, Xuan Song, Shintaro Fukushima, and Toyotaro Suzumura. Spatio-temporal meta-graph learning for traffic forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 8078–8086, 2023.

Pengfei Jiao, Hongjiang Chen, Huijun Tang, Qing Bao, Long Zhang, Zhidong Zhao, and Huaming Wu. Contrastive representation learning on dynamic networks. *Neural Networks*, 174:106240, 2024.

Pengfei Jiao, Hongjiang Chen, Xuan Guo, Zhidong Zhao, Dongxiao He, and Di Jin. A survey on temporal interaction graph representation learning: Progress, challenges, and opportunities. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pp. 10499–10507, 2025.

Guangyin Jin, Lingbo Liu, Fuxian Li, and Jincai Huang. Spatio-temporal graph neural point process for traffic congestion event prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 14268–14276, 2023.

Ming Jin, Yuan-Fang Li, and Shirui Pan. Neural temporal walks: Motif-aware representation learning on continuous-time dynamic graphs. In *Advances in Neural Information Processing Systems*, volume 35, pp. 19874–19886, 2022.

Woojeong Jin, Meng Qu, Xisen Jin, and Xiang Ren. Recurrent event network: Autoregressive structure inference over temporal knowledge graphs. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, pp. 6669–6683, 2020.

Alistair E. W. Johnson, Tom J. Pollard, Lu Shen, Li-wei H. Lehman, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G. Mark. MIMIC-III, a freely accessible critical care database. *Scientific Data*, 3:160035, 2016.

Manu Joseph and Harsh Raj. GANDALF: Gated adaptive network for deep automated learning of features. *arXiv preprint arXiv:2207.08548*, 2022.

Arlind Kadra, Marius Lindauer, Frank Hutter, and Josif Grabocka. Well-tuned simple nets excel on tabular datasets. In *Advances in Neural Information Processing Systems*, volume 34, pp. 23928–23941, 2021.

Michiel Kamphuis. Tiny-toxic-detector: A compact transformer-based model for toxic content detection. *arXiv preprint arXiv:2409.02114*, 2024.

Inwon Kang, Parikshit Ram, Yi Zhou, Horst Samulowitz, and Oshani Seneviratne. On learning representations for tabular data distillation. *Transactions on Machine Learning Research*, 2025.

Wang-Cheng Kang and Julian J. McAuley. Self-attentive sequential recommendation. In *Proceedings of the 2018 IEEE International Conference on Data Mining*, pp. 197–206, 2018.

Pavan Kantharaju and Santiago Ontañón. Discovering meaningful labelings for RTS game replays via replay embeddings. In *Proceedings of the 2020 IEEE Conference on Games*, pp. 160–167, 2020.

Jared L. Katzman, Uri Shaham, Alexander Cloninger, Jonathan Bates, Tingting Jiang, and Yuval Kluger. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Medical Research Methodology*, 18:24, 2018.

Seyed Mehran Kazemi, Rishab Goel, Sepehr Eghbali, Janahan Ramanan, Jaspreet Sahota, Sanjay Thakur, Stella Wu, Cathal Smyth, Pascal Poupart, and Marcus Brubaker. Time2Vec: Learning a vector representation of time. *arXiv preprint arXiv:1907.05321*, 2019.

Seyed Mehran Kazemi, Rishab Goel, Kshitij Jain, Ivan Kobyzev, Akshay Sethi, Peter Forsyth, and Pascal Poupart. Representation learning for dynamic graphs: A survey. *Journal of Machine Learning Research*, 21(70):1–73, 2020.

Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. LightGBM: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems*, volume 30, pp. 3146–3154, 2017.

Guolin Ke, Jia Zhang, Zhenhui Xu, Jiang Bian, and Tie-Yan Liu. TabNN: A universal neural network solution for tabular data. ICLR 2019 Conference Blind Submission, OpenReview, 2018.

Guolin Ke, Zhenhui Xu, Jia Zhang, Jiang Bian, and Tie-Yan Liu. Deep-GBM: A deep learning framework distilled by GBDT for online prediction tasks. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 384–394, 2019.

Patrik Kenfack, Samira Ebrahimi Kahou, and Ulrich Aïvodji. Towards fair in-context learning with tabular foundation models. *Transactions on Machine Learning Research*, 2026.

Samira Khodabandehlou and Alireza Hashemi Golpayegani. FiFrauD: unsupervised financial fraud detection in dynamic graph streams. *ACM Transactions on Knowledge Discovery from Data*, 18(5):1–29, 2024.

Min-Soo Kim and Jiawei Han. A particle-and-density based evolutionary clustering method for dynamic networks. *Proceedings of the VLDB Endowment*, 2(1):622–633, 2009.

Myung Jun Kim, Léo Grinsztajn, and Gaël Varoquaux. CARTE: Pretraining and transfer for tabular learning. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235, pp. 23843–23866, 2024.

Myung Jun Kim, Félix Lefebvre, Gaëtan Brison, Alexandre Perez-Lebel, and Gaël Varoquaux. Table foundation models: on knowledge pre-training for tabular learning. *Transactions on Machine Learning Research*, 2025.

Shun Kiyono, Sosuke Kobayashi, Jun Suzuki, and Kentaro Inui. SHAPE: Shifted absolute position embedding for transformers. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 3309–3321, 2021.

Boris Knyazev, Carolyn Augusta, and Graham W. Taylor. Learning temporal attention in dynamic graphs with bilinear interactions. *PLOS ONE*, 16(3):e0247936, 2021.

Kelvin J.L. Koa, Yunshan Ma, Ritchie Ng, and Tat-Seng Chua. Learning to generate explainable stock predictions using self-reflective large language models. In *Proceedings of the ACM Web Conference 2024*, pp. 4304–4315. Association for Computing Machinery, 2024.

Sukanth Korkanti. Enhancing financial fraud detection using LLMs and advanced data analytics. In *Proceedings of the 2024 2nd International Conference on Self Sustainable Artificial Intelligence Systems*, pp. 1328–1334, 2024.

Jannik Kossen, Neil Band, Clare Lyle, Aidan N. Gomez, Thomas Rainforth, and Yarin Gal. Self-attention between datapoints: Going beyond individual input-output pairs in deep learning. In *Advances in Neural Information Processing Systems*, volume 34, pp. 28742–28756, 2021.

Akim Kotelnikov, Dmitry Baranchuk, Ivan Rubachev, and Artem Babenko. TabDDPM: Modelling tabular data with diffusion models. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202, pp. 17564–17579, 2023.

Arun Kumar, Ishan Garg, and Sanmeet Kaur. Loan approval prediction based on machine learning approach. *IOSR Journal of Computer Engineering*, 18(3):79–81, 2016.

Srijan Kumar, Xikun Zhang, and Jure Leskovec. Predicting dynamic embedding trajectory in temporal interaction networks. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1269–1278, 2019.

Haewoon Kwak and Jeremy Blackburn. Linguistic analysis of toxic behavior in an online video game. In *Social Informatics: SocInfo 2014 International Workshops, Barcelona, Spain, November 11, 2014, Revised Selected Papers*, volume 8852, pp. 209–217, 2015.

Bum Chul Kwon, Min-Je Choi, Joanne Taery Kim, Edward Choi, Young Bin Kim, Soonwook Kwon, Jimeng Sun, and Jaegul Choo. Retainvis: Visual analytics with interpretable and interactive recurrent neural networks on electronic medical records. *IEEE Transactions on Visualization and Computer Graphics*, 25 (1):299–309, 2019.

Alexander Larionov, Niall M. Adams, and Kevin N. Webster. Investigating the impact of missing value handling on boosted trees and deep learning for tabular data: A claim reserving case study. *Transactions on Machine Learning Research*, 2025.

Jeong Min Lee and Milos Hauskrecht. Personalized event prediction for Electronic Health Records. *Artificial Intelligence in Medicine*, 143:102620, 2023.

Edmond Lezmi and Jiali Xu. Time series forecasting with transformer models and application to asset management. *Amundi Research Center Working Paper*, (139):1–44, 2023.

Chenglin Li, Mingjun Zhao, Huanming Zhang, Chenyun Yu, Lei Cheng, Guoqiang Shu, Beibei Kong, and Di Niu. RecGURU: Adversarial learning of generalized user representations for cross-domain recommendation. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, pp. 571–581, 2022a.

Hanjie Li, Changsheng Li, Kaituo Feng, Ye Yuan, Guoren Wang, and Hongyuan Zha. Robust knowledge adaptation for dynamic graph neural networks. *IEEE Transactions on Knowledge and Data Engineering*, 36(11):6920–6933, 2024a.

Haoyang Li, Yuming Xu, Yiming Li, Hanmo Liu, Darian Li, Chen Jason Zhang, Lei Chen, and Qing Li. When speed meets accuracy: An efficient and effective graph model for temporal link prediction. *Proceedings of the VLDB Endowment*, 18(10):3396–3405, 2025a.

Hongxi Li, Zuxuan Zhang, Dengzhe Liang, and Yuncheng Jiang. K-truss based temporal graph convolutional network for dynamic graphs. In *Proceedings of the 15th Asian Conference on Machine Learning*, volume 222, pp. 739–754, 2024b.

Jia Li, Zhichao Han, Hong Cheng, Jiao Su, Pengyun Wang, Jianfeng Zhang, and Lujia Pan. Predicting path failure in time-evolving graphs. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1279–1289, 2019.

Jiacheng Li, Yujie Wang, and Julian J. McAuley. Time interval aware self-attention for sequential recommendation. In *Proceedings of the Thirteenth ACM International Conference on Web Search and Data Mining*, pp. 322–330, 2020a.

Jiacheng Li, Jingbo Shang, and Julian McAuley. UCTopic: Unsupervised contrastive learning for phrase representations and topic mining. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 6159–6169, 2022b.

Jintang Li, Zhouxin Yu, Zulun Zhu, Liang Chen, Qi Yu, Zibin Zheng, Sheng Tian, Ruofan Wu, and Changhua Meng. Scaling up dynamic graph representation learning via spiking neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 8588–8596, 2023a.

Pan Li, Yuyan Wang, Ed H. Chi, and Minmin Chen. Prompt tuning large language models on personalized aspect extraction for recommendations. *arXiv preprint arXiv:2306.01475*, 2023b.

Qingyao Li, Lingyue Fu, Weiming Zhang, Xianyu Chen, Jingwei Yu, Wei Xia, Weinan Zhang, Ruiming Tang, and Yong Yu. Adapting large language models for education: Foundational capabilities, potentials, and challenges. *arXiv preprint arXiv:2401.08664*, 2024c.

Xiang Li, Chao Wang, Jiwei Tan, Xiaoyi Zeng, Dan Ou, and Bo Zheng. Adversarial multimodal representation learning for click-through rate prediction. In *Proceedings of The Web Conference 2020*, pp. 827–836, 2020b.

Yang Li, Nan Du, and Samy Bengio. Time-dependent representation for neural event sequence prediction. In *6th International Conference on Learning Representations, Workshop Track*, 2018.

Yikuan Li, Shishir Rao, José Roberto Ayala Solares, Abdelaali Hassaine, Rema Ramakrishnan, Dexter Canoy, Yajie Zhu, Kazem Rahimi, and Gholamreza Salimi-Khorshidi. BEHRT: Transformer for electronic health records. *Scientific Reports*, 10(1):7155, 2020c.

Yingxuan Li, Yuanyuan Xu, Xuemin Lin, Wenjie Zhang, and Ying Zhang. Ranking on dynamic graphs: An effective and robust band-pass disentangled approach. In *Proceedings of the ACM on Web Conference 2025*, pp. 3918–3929, 2025b.

Jianxun Lian, Xiaohuan Zhou, Fuzheng Zhang, Zhongxia Chen, Xing Xie, and Guangzhong Sun. xDeepFM: Combining explicit and implicit feature interactions for recommender systems. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 1754–1763, 2018.

Ke Liang, Lingyuan Meng, Meng Liu, Yue Liu, Wenxuan Tu, Siwei Wang, Sihang Zhou, and Xinwang Liu. Learn from relational correlations and periodic events for temporal knowledge graph reasoning. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 1559–1568, 2023.

Lei Liang, Mengshu Sun, Zhengke Gui, Zhongshu Zhu, Zhouyu Jiang, Ling Zhong, Yuan Qu, Peilong Zhao, Zhongpu Bo, Jin Yang, Huaidong Xiong, Lin Yuan, Jun Xu, Zaoyang Wang, Zhiqiang Zhang, Wen Zhang, Huajun Chen, Wenguang Chen, and Jun Zhou. KAG: Boosting LLMs in professional domains via knowledge augmented generation. In *Companion Proceedings of the ACM on Web Conference 2025*, pp. 334–343, 2025.

David Liben-Nowell and Jon Kleinberg. The link prediction problem for social networks. In *Proceedings of the 12th ACM International Conference on Information and Knowledge Management*, pp. 556–559, 2003.

Tatiana Likhomanenko, Qiantong Xu, Gabriel Synnaeve, Ronan Collobert, and Alex Rogozhnikov. CAPE: Encoding relative positions with continuous augmented positional embeddings. In *Advances in Neural Information Processing Systems*, volume 34, pp. 16079–16092, 2021.

Jianghao Lin, Xinyi Dai, Rong Shan, Bo Chen, Ruiming Tang, Yong Yu, and Weinan Zhang. Large language models make sample-efficient recommender systems. *Frontiers of Computer Science*, 19(4):194328, 2025.

Can Liu, Yuncong Gao, Li Sun, Jinghua Feng, Hao Yang, and Xiang Ao. User behavior pre-training for online fraud detection. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 3357–3365, 2022a.

Chengkai Liu, Jianghao Lin, Jianling Wang, Hanzhou Liu, and James Caverlee. Mamba4Rec: Towards efficient sequential recommendation with selective state space models. *arXiv preprint arXiv:2403.03900*, 2024a.

Chi Harold Liu, Jie Xu, Jian Tang, and Jon Crowcroft. Social-aware sequential modeling of user interests: A deep learning approach. *IEEE Transactions on Knowledge and Data Engineering*, 31(11):2200–2212, 2019.

Fan Liu, Yaqi Liu, Huilin Chen, Zhiyong Cheng, Liqiang Nie, and Mohan S. Kankanhalli. Understanding before recommendation: Semantic aspect-aware review exploitation via large language models. *ACM Transactions on Information Systems*, 43(2):1–26, 2025.

Feiyan Liu, Liangzhi Li, Xiaoli Wang, Feng Luo, Chang Liu, Jinsong Su, and Yiming Qian. MHGRL: An effective representation learning model for electronic health records. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation*, pp. 11272–11282, 2024b.

Guang Liu, Jie Yang, and Ledell Wu. Ptab: Using the pre-trained language model for modeling tabular data. *arXiv preprint arXiv:2209.08060*, 2022b.

Guimei Liu, Tam T. Nguyen, Gang Zhao, Wei Zha, Jianbo Yang, Jianneng Cao, Min Wu, Peilin Zhao, and Wei Chen. Repeat buyer prediction for e-commerce. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 155–164, 2016.

Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. Visual instruction tuning. In *Advances in Neural Information Processing Systems*, volume 36, pp. 34892–34916, 2023a.

Kangzheng Liu, Feng Zhao, Guandong Xu, Xianzhi Wang, and Hai Jin. RETIA: Relation-entity twin-interact aggregation for temporal knowledge graph extrapolation. In *Proceedings of the 39th IEEE International Conference on Data Engineering*, pp. 1761–1774, 2023b.

Qian Liu, Bei Chen, Jiaqi Guo, Morteza Ziyadi, Zeqi Lin, Weizhu Chen, and Jian-Guang Lou. TAPEX: Table pre-training via learning a neural SQL executor. In *International Conference on Learning Representations*, 2022c.

Shang-Ching Liu, ShengKun Wang, Tsungyao Chang, Wenqi Lin, Chung-Wei Hsiung, Yi-Chen Hsieh, Yu-Ping Cheng, Sian-Hong Luo, and Jianwei Zhang. Jarvix: A LLM no code platform for tabular data analysis and optimization. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: Industry Track*, pp. 622–630, 2023c.

Tianrui Liu, Qi Cai, Changxin Xu, Bo Hong, Fanghao Ni, Yuxin Qiao, and Tsungwei Yang. Rumor detection with a novel graph neural network approach. *arXiv preprint arXiv:2403.16206*, 2024c.

Zhaocheng Liu, Qiang Liu, Haoli Zhang, and Yuntian Chen. DNN2LR: Interpretation-inspired feature crossing for real-world tabular data. *arXiv preprint arXiv:2008.09775*, 2020.

Zhiwei Liu, Yongjun Chen, Jia Li, Philip S. Yu, Julian McAuley, and Caiming Xiong. Contrastive self-supervised sequential recommendation with robust augmentation. *arXiv preprint arXiv:2108.06479*, 2021a.

Zhiwei Liu, Ziwei Fan, Yu Wang, and Philip S. Yu. Augmenting sequential recommendation with pseudo-prior items via reversely pre-training transformer. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 1608–1612, 2021b.

Shiqi Lou, Qingyue Zhang, Shujie Yang, Yuyang Tian, Zhaoxuan Tan, and Minnan Luo. GADY: Unsupervised anomaly detection on dynamic graphs. *arXiv preprint arXiv:2310.16376*, 2023.

Tianyu Lu, Zhongwen Zhou, Jiayi Wang, and Yong Wang. A large language model-based approach for personalized search results re-ranking in professional domains. *Academia Nexus Journal*, 4(1), 2025.

Xiaodong Lu, Leilei Sun, Tongyu Zhu, and Weifeng Lv. Improving temporal link prediction via temporal walk matrix projection. In *Advances in Neural Information Processing Systems*, volume 37, pp. 141153–141182, 2024.

Yuanfu Lu, Xiao Wang, Chuan Shi, Philip S. Yu, and Yanfang Ye. Temporal network embedding with micro- and macro-dynamics. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pp. 469–478, 2019.

Yuanfei Luo, Hao Zhou, Wei-Wei Tu, Yuqiang Chen, Wenyuan Dai, and Qiang Yang. Network on network for tabular data classification in real-world applications. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 2317–2326, 2020.

Yuhong Luo and Pan Li. Neighborhood-aware scalable temporal network representation learning. In *Proceedings of the First Learning on Graphs Conference*, volume 198, pp. 1:1–1:18, 2022.

Manfu Ma, Cong Zhang, Yong Li, Jiahao Chen, and Xuegang Wang. Rumor detection model with weighted graphsage focusing on node location. *Scientific Reports*, 14(1):27127, 2024a.

Teng Ma, Bin Shi, Yiming Xu, Zihan Zhao, Siqi Liang, and Bo Dong. DYGL: A unified benchmark and library for dynamic graph. In *Web and Big Data – 7th International Joint Conference, APWeb-WAIM 2023, Wuhan, China, October 6–8, 2023, Proceedings, Part III*, volume 14333, pp. 389–401, 2024b.

Yao Ma, Ziyi Guo, Zhaochun Ren, Jiliang Tang, and Dawei Yin. Streaming graph neural networks. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 719–728, 2020.

Bodhisattwa Prasad Majumder, Shuyang Li, Jianmo Ni, and Julian McAuley. Generating personalized recipes from historical user preferences. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pp. 5976–5982, 2019.

Franco Manessi, Alessandro Rozza, and Mario Manzo. Dynamic graph convolutional networks. *Pattern Recognition*, 97:107000, 2020.

Andrei Margeloiu, Nikola Simidjievski, Pietro Lio, and Mateja Jamnik. Gcondnet: A novel method for improving neural networks on small high-dimensional tabular data. *Transactions on Machine Learning Research*, 2024.

Manideep Marripudugala. Real-time IoT data analytics using advanced large language model techniques. In *Proceedings of the 2024 Global Conference on Communications and Information Technologies*, pp. 1–6, 2024.

Marcus Märtens, Siqi Shen, Alexandru Iosup, and Fernando A. Kuipers. Toxicity detection in multiplayer online games. In *Proceedings of the 2015 International Workshop on Network and Systems Support for Games*, pp. 1–6, 2015.

Julian J. McAuley and Alex Yang. Addressing complex and subjective product-related queries with customer reviews. In *Proceedings of the 25th ACM World Wide Web Conference*, pp. 625–635, 2016.

Calvin McCarter. Unmasking trees for tabular data. *Transactions on Machine Learning Research*, 2025.

Duncan McElfresh, Sujay Khandagale, Jonathan Valverde, Vishak C Prasad, Benjamin Feuer, Chinmay Hegde, Ganesh Ramakrishnan, Micah Goldblum, and Colin White. When do neural nets outperform boosted trees on tabular data? In *Advances in Neural Information Processing Systems*, volume 36, pp. 76336–76369, 2023.

H. Brendan McMahan, Gary Holt, David Sculley, Michael Young, Dietmar Ebner, Julian Grady, Lan Nie, Todd Phillips, Eugene Davydov, Daniel Golovin, Sharat Chikkerur, Dan Liu, Martin Wattenberg, Arnar Mar Hrafnkelsson, Tom Boulos, and Jeremy Kubica. Ad click prediction: a view from the trenches. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1222–1230, 2013.

Hongyuan Mei and Jason M. Eisner. The neural hawkes process: A neurally self-modulating multivariate point process. In *Advances in Neural Information Processing Systems*, volume 30, pp. 6754–6764, 2017.

Alessio Micheli and Domenico Tortorella. Discrete-time dynamic graph echo state networks. *Neurocomputing*, 496:85–95, 2022.

Wei Min, Weiming Liang, Hang Yin, Zhurong Wang, Mei Li, and Alok Lal. Explainable deep behavioral sequence clustering for transaction fraud detection. *arXiv preprint arXiv:2101.04285*, 2021.

Wookhee Min, Bradford Mott, Jonathan Rowe, Barry Liu, and James Lester. Player goal recognition in open-world digital games with long short-term memory networks. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, pp. 2590–2596, 2016.

Bamshad Mobasher, Robert Cooley, and Jaideep Srivastava. Automatic personalization based on web usage mining. *Communications of the ACM*, 43(8):142–151, 2000.

Payal Mohapatra, Lixu Wang, and Qi Zhu. Phase-driven generalizable representation learning for nonstationary time series classification. *Transactions on Machine Learning Research*, 2025.

Joshua L. Moore, Shuo Chen, Douglas R. Turnbull, and Thorsten Joachims. Taste over time: The temporal dynamics of user preferences. In *Proceedings of the 14th International Society for Music Information Retrieval Conference*, pp. 401–406, 2013.

Gabriel de Souza Pereira Moreira, Sara Rabhi, Ronay Ak, Md Yasin Kabir, and Even Oldridge. Transformers with multi-modal features and post-fusion context for e-commerce session-based recommendation. *arXiv preprint arXiv:2107.05124*, 2021.

Jacob Morrier, Rafal Kocielnik, and R. Michael Alvarez. Reinforcement learning for efficient toxicity detection in competitive online video games. *arXiv preprint arXiv:2503.20968*, 2025.

Kwangtek Na, Ju-Hong Lee, and Eunchan Kim. LF-transformer: Latent factorizer transformer for tabular learning. *IEEE Access*, 12:10690–10698, 2024.

Apurva Narayan and Peter H. O'N Roe. Learning graph dynamics using deep neural networks. In *9th Vienna International Conference on Mathematical Modelling*, volume 51, pp. 433–438, 2018.

Avanika Narayan, Ines Chami, Laurel J. Orr, Simran Arora, and Christopher Ré. Can foundation models wrangle your data? *arXiv preprint arXiv:2205.09911*, 2022.

Fatemeh Nazary, Yashar Deldjoo, and Tommaso Di Noia. Poison-RAG: Adversarial data poisoning attacks on retrieval-augmented generation in recommender systems. In *Advances in Information Retrieval: 47th European Conference on Information Retrieval, Lucca, Italy, April 6–10, 2025, Proceedings, Part IV*, volume 15575, pp. 239–251, 2025.

Masato Neishi and Naoki Yoshinaga. On the relation between position information and sentence length in neural machine translation. In *Proceedings of the 23rd Conference on Computational Natural Language Learning*, pp. 328–338, 2019.

Lynnette Hui Xian Ng, Adrian Xuan Wei Lim, and Michael Miller Yoder. Challenges for real-time toxicity detection in online games. *arXiv preprint arXiv:2407.04383*, 2024.

Giang Hoang Nguyen, John Boaz Lee, Ryan A. Rossi, Nesreen K. Ahmed, Eunyee Koh, and Sungchul Kim. Dynamic network embeddings: From random walks to temporal random walks. In *Proceedings of the 2018 IEEE International Conference on Big Data*, pp. 1085–1092, 2018.

Jianmo Ni, Larry Muhlstein, and Julian McAuley. Modeling heart rate and activity data for personalized fitness recommendation. In *Proceedings of the 28th ACM World Wide Web Conference*, pp. 1343–1353, 2019.

Shuang Ni, Wei Zhou, Junhao Wen, Linfeng Hu, and Shutong Qiao. Enhancing sequential recommendation with contrastive generative adversarial network. *Information Processing & Management*, 60(3):103331, 2023.

Yabo Ni, Dan Ou, Shichen Liu, Xiang Li, Wenwu Ou, Anxiang Zeng, and Luo Si. Perceive your users in depth: Learning universal user representations from multiple e-commerce tasks. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 596–605, 2018.

Oded Nov, Nina Singh, and Devin M. Mann. Putting ChatGPT's medical advice to the (turing) test: Survey study. *JMIR Medical Education*, 9:e46939, 2023.

OpenAI. GPT-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.

OpenAI. GPT-4o system card. *arXiv preprint arXiv:2410.21276*, 2024.

Tosin Tunde Oyetayo, John Olalekan Adeniran, Stephen Adebanji, Taiwo Adenike Adekunle, and Olajide Yusuf Adebayo. A neural multi-label model for enhanced toxicity detection in online conversations. *Adeleke University Journal of Science*, 3(1):9–15, 2024.

Li-Wei Pan, Wei-Ke Pan, Mei-Yan Wei, Hong-Zhi Yin, and Zhong Ming. A survey on sequential recommendation. *Frontiers of Computer Science*, 20(3):2003606, 2026.

Panayot Panayotov, Utsav Shukla, Husrev Taha Sencar, Mohamed Nabeel, and Preslav Nakov. GREENER: Graph neural networks for news media profiling. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pp. 7470–7480, 2022.

Prince Pandey and Jyoti Prakash Singh. Generating product reviews from aspect-based ratings using large language models. *Journal of Retailing and Consumer Services*, 84:104244, 2025.

Aldo Pareja, Giacomo Domeniconi, Jie Chen, Tengfei Ma, Toyotaro Suzumura, Hiroki Kanezashi, Tim Kaler, Tao B. Schardl, and Charles E. Leiserson. Evolvegcn: Evolving graph convolutional networks for dynamic graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 5363–5370, 2020.

Apurva Pathak, Kshitiz Gupta, and Julian McAuley. Generating and personalizing bundle recommendations on Steam. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 1073–1076, 2017.

Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 701–710, 2014.

Heinrich Peters and Sandra C. Matz. Large language models can infer psychological dispositions of social media users. *PNAS Nexus*, 3(6):pgae231, 2024.

Jinghua Piao, Yuwei Yan, Jun Zhang, Nian Li, Junbo Yan, Xiaochong Lan, Zhihong Lu, Zhiheng Zheng, Jing Yi Wang, Di Zhou, Chen Gao, Fengli Xu, Fang Zhang, Ke Rong, Jun Su, and Yong Li. Agentsociety: Large-scale simulation of LLM-driven generative agents advances understanding of human behaviors and society. *arXiv preprint arXiv:2502.08691*, 2025.

Luca Podo, Marco Angelini, and Paola Velardi. V-RECS, a low-cost LLM4VIS recommender with explanations, captioning and suggestions. *arXiv preprint arXiv:2406.15259*, 2024.

Sergei Popov, Stanislav Morozov, and Artem Babenko. Neural oblivious decision ensembles for deep learning on tabular data. In *International Conference on Learning Representations*, 2020.

Tim Postuvan, Claas Grohnfeldt, Michele Russo, and Giulio Lovisotto. Learning-based link anomaly detection in continuous-time dynamic graphs. *Transactions on Machine Learning Research*, 2024.

R. Prabhu and V. Seethalakshmi. A comprehensive framework for multi-modal hate speech detection in social media using deep learning. *Scientific Reports*, 15(1):1–21, 2025.

Ofir Press, Noah A. Smith, and Mike Lewis. Train short, test long: Attention with linear biases enables input length extrapolation. In *Proceedings of the Tenth International Conference on Learning Representations*, pp. 1–14, 2022.

Liudmila Prokhorenkova, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, and Andrey Gulin. CatBoost: unbiased boosting with categorical features. In *Advances in Neural Information Processing Systems*, volume 31, pp. 6639–6649, 2018.

Erasmo Purificato, Ludovico Boratto, and Ernesto William De Luca. User modeling and user profiling: A comprehensive survey. *arXiv preprint arXiv:2402.09660*, 2024.

Chuan Qin, Le Zhang, Yihang Cheng, Rui Zha, Dazhong Shen, Qi Zhang, Xi Chen, Ying Sun, Chen Zhu, Hengshu Zhu, and Hui Xiong. A comprehensive survey of artificial intelligence techniques for talent analytics. *Proceedings of the IEEE*, 113(2):125–171, 2025.

Xiao Qin, Nasrullah Sheikh, Chuan Lei, Berthold Reinwald, and Giacomo Domeniconi. Seign: A simple and efficient graph neural network for large dynamic graphs. In *Proceedings of the 2023 IEEE 39th International Conference on Data Engineering*, pp. 2850–2863, 2023.

Ruihong Qiu, Zi Huang, Hongzhi Yin, and Zijian Wang. Contrastive learning for representation degeneration problem in sequential recommendation. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, pp. 813–823, 2022.

Zhaopeng Qiu, Xian Wu, Jingyue Gao, and Wei Fan. U-BERT: Pre-training user representations for improved recommendation. In *Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence*, volume 35, pp. 4320–4327, 2021.

Zhenyu Qiu, Wenbin Hu, Jia Wu, Weiwei Liu, Bo Du, and Xiaohua Jia. Temporal network embedding with high-order nonlinear information. In *Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence*, volume 34, pp. 5436–5443, 2020.

Liang Qu, Huaisheng Zhu, Qiqi Duan, and Yuhui Shi. Continuous-time link prediction via temporal dependent graph neural network. In *Proceedings of The Web Conference 2020*, pp. 3026–3032, 2020.

Massimo Quadrana, Alexandros Karatzoglou, Balázs Hidasi, and Paolo Cremonesi. Personalizing session-based recommendations with hierarchical recurrent neural networks. In *Proceedings of the Eleventh ACM Conference on Recommender Systems*, pp. 130–137, 2017.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. *OpenAI Blog*, 2019. Accessed January 4, 2026.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140):1–67, 2020.

Jérémie Rappaz, Julian McAuley, and Karl Aberer. Recommendation on live-streaming platforms: Dynamic availability and repeat consumption. In *Proceedings of the 15th ACM Conference on Recommender Systems*, pp. 390–399, 2021.

André Gomes Regino, Victor Hochgreb, and Julio Cesar dos Reis. Humanizing answers for compatibility questions in e-commerce using large language models. In *Proceedings of the 39th Brazilian Symposium on Databases*, pp. 300–312, 2024.

Houxing Ren, Jingyuan Wang, Wayne Xin Zhao, and Ning Wu. RAPT: Pre-training of time-aware transformer for learning robust healthcare representation. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pp. 3503–3511, 2021.

Lin Ren, Yongbin Liu, Chunping Ouyang, Ying Yu, Shuda Zhou, Yidong He, and Yaping Wan. DyLas: A dynamic label alignment strategy for large-scale multi-label text classification. *Information Fusion*, 120: 103081, 2025.

Steffen Rendle. Factorization machines. In *Proceedings of the 2010 IEEE International Conference on Data Mining*, pp. 995–1000, 2010.

Steffen Rendle, Christoph Freudenthaler, and Lars Schmidt-Thieme. Factorizing personalized Markov chains for next-basket recommendation. In *Proceedings of the 19th ACM World Wide Web Conference*, pp. 811–820, 2010.

Emanuele Rossi, Benjamin Paul Chamberlain, Fabrizio Frasca, Davide Eynard, Federico Monti, and Michael M. Bronstein. Temporal graph networks for deep learning on dynamic graphs. *arXiv preprint arXiv:2006.10637*, 2020.

Benedek Rozemberczki, Paul Scherer, Yixuan He, George Panagopoulos, Alexander Riedel, Maria Astefanoaei, Oliver Kiss, Ferenc Beres, Guzmán López, Nicolas Collignon, and Rik Sarkar. PyTorch geometric temporal: Spatiotemporal signal processing with neural machine learning models. In *Proceedings of the 30th ACM International Conference on Information and Knowledge Management*, pp. 4564–4573, 2021.

Ramon Ruiz-Dolz and John Lawrence. An explainable framework for misinformation identification via critical question answering. *arXiv preprint arXiv:2503.14626*, 2025.

Milad Sabouri, Masoud Mansoury, Kun Lin, and Bamshad Mobasher. Towards explainable temporal user profiling with LLMs. In *Adjunct Proceedings of the 33rd ACM Conference on User Modeling, Adaptation and Personalization*, pp. 219–227, 2025.

Aravind Sankar, Yanhong Wu, Liang Gou, Wei Zhang, and Hao Yang. DySAT: Deep neural representation learning on dynamic graphs via self-attention networks. In *Proceedings of the Thirteenth ACM International Conference on Web Search and Data Mining*, pp. 519–527, 2020.

Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. The graph neural network model. *IEEE Transactions on Neural Networks*, 20(1):61–80, 2009.

Youngjoo Seo, Michaël Defferrard, Pierre Vandergheynst, and Xavier Bresson. Structured sequence modeling with graph convolutional recurrent networks. In *Neural Information Processing - 25th International Conference, Siem Reap, Cambodia, December 13–16, 2018, Proceedings, Part I*, volume 11301, pp. 362–373, 2018.

Gautam Kishore Shahi and Tim A. Majchrzak. Defining, understanding, and detecting online toxicity: Challenges and machine learning approaches. *arXiv preprint arXiv:2509.14264*, 2025.

Umang Sharan and Jennifer Neville. Temporal-relational classifiers for prediction in evolving domains. In *Proceedings of the Eighth IEEE International Conference on Data Mining*, pp. 540–549, 2008.

Ira Shavitt and Eran Segal. Regularization learning networks: Deep learning for tabular datasets. In *Advances in Neural Information Processing Systems*, volume 31, pp. 1386–1396, 2018.

Peter Shaw, Jakob Uszkoreit, and Ashish Vaswani. Self-attention with relative position representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pp. 464–468, 2018.

Oleksandr Shchur, Ali Caner Türkmen, Tim Januschowski, and Stephan Günnemann. Neural temporal point processes: A review. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, pp. 4585–4593, 2021.

Guangming Sheng, Junwei Su, Chao Huang, and Chuan Wu. MSPipe: Efficient temporal GNN training via staleness-aware pipeline. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 2651–2662, 2024.

Alex Sherstinsky. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D: Nonlinear Phenomena*, 404:132306, 2020.

Wenqi Shi, Ran Xu, Yuchen Zhuang, Yue Yu, Jieyu Zhang, Hang Wu, Yuanda Zhu, Joyce C. Ho, Carl Yang, and May Dongmei Wang. EHRAgent: Code empowers large language models for few-shot complex tabular reasoning on electronic health records. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pp. 22315–22339, 2024.

Razieh Shirzadkhani, Shenyang Huang, Elahe Kooshafar, Reihaneh Rabbany, and Farimah Poursafaei. Temporal graph analysis with TGX. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*, pp. 1086–1089, 2024.

Ravid Shwartz-Ziv and Amitai Armon. Tabular data: Deep learning is not all you need. *Information Fusion*, 81:84–90, 2022.

Denis Sidorenko, Stefan Pushkov, Akhmed Sakip, Geoffrey Ho Duen Leung, Sarah Wing Yan Lok, Anatoly Urban, Diana Zagirova, Alexander Veviorskiy, Nina Tihonova, Aleksandr Kalashnikov, Ekaterina Kozlova, Vladimir Naumov, Frank W. Pun, Alex Aliper, Feng Ren, and Alex Zhavoronkov. Precious2GPT: the combination of multiomics pretrained transformer and conditional diffusion for artificial multi-omics multi-species multi-tissue sample generation. *npj Aging*, 10(1):37, 2024.

Uriel Singer, Ido Guy, and Kira Radinsky. Node embedding over temporal graphs. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, pp. 4605–4612, 2019.

Gurjot Singh, Prabhjot Singh, and Maninder Singh. Advanced real-time fraud detection using RAG-based LLMs. *arXiv preprint arXiv:2501.15290*, 2025.

Priya Singh, Manoj Jha, Mohamed Sharaf, Mohammed A. El-Meligy, and Thippa Reddy Gadekallu. Harnessing a hybrid CNN-LSTM model for portfolio performance: A case study on stock selection and optimization. *IEEE Access*, 11:104000–104015, 2023.

Ananya Singha, José Cambronero, Sumit Gulwani, Vu Le, and Chris Parnin. Tabular representation, noisy operators, and impacts on table structure understanding tasks in LLMs. *arXiv preprint arXiv:2310.10358*, 2023.

Karan Singhal, Tao Tu, Juraj Gottweis, Rory Sayres, Ellery Wulczyn, Mohamed Amin, Le Hou, Kevin Clark, Stephen R. Pfohl, Heather Cole-Lewis, Darlene Neal, Qazi Mamunur Rashid, Mike Schaekermann, Amy Wang, Dev Dash, Jonathan Chen, Nigam H. Shah, Sami Lachgar, Philip Mansfield, Sushant Prakash, Bradley Green, Ewa Dominowska, Blaise Agüera y Arcas, Nenad Tomasev, Yun Liu, Renee Wong, Christopher Semturs, S. Sara Mahdavi, Joelle K. Barral, Dale Webster, Greg S. Corrado, Yossi Matias, Shekoofeh Azizi, Alan Karthikesalingam, and Vivek Natarajan. Toward expert-level medical question answering with large language models. *Nature Medicine*, 31(3):943–950, 2025.

Koustuv Sinha, Amirhossein Kazemnejad, Siva Reddy, Joelle Pineau, Dieuwke Hupkes, and Adina Williams. The curious case of absolute position embeddings. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pp. 4449–4472, 2022.

Piotr Skalski, David Sutton, Stuart Burrell, Iker Perez, and Jason Wong. Towards a foundation purchasing model: Pretrained generative autoregression on transaction sequences. In *Proceedings of the 4th ACM International Conference on AI in Finance*, pp. 141–149, 2023.

Joakim Skarding, Bogdan Gabrys, and Katarzyna Musial. Foundations and modeling of dynamic networks using dynamic graph neural networks: A survey. *IEEE Access*, 9:79143–79168, 2021.

Gowthami Somepalli, Micah Goldblum, Avi Schwarzschild, C. Bayan Bruss, and Tom Goldstein. SAINT: Improved neural networks for tabular data via row attention and contrastive pre-training. *arXiv preprint arXiv:2106.01342*, 2021.

Shriyank Somvanshi, Subasish Das, Syed Aaqib Javed, Gian Antariksa, and Ahmed Hossain. A survey on deep tabular learning. *arXiv preprint arXiv:2410.12034*, 2024.

Weiping Song, Chence Shi, Zhiping Xiao, Zhijian Duan, Yewen Xu, Ming Zhang, and Jian Tang. AutoInt: Automatic feature interaction learning via self-attentive neural networks. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pp. 1161–1170, 2019.

Amauri Souza, Diego Mesquita, Samuel Kaski, and Vikas Garg. Provably expressive temporal graph networks. In *Advances in Neural Information Processing Systems*, volume 35, pp. 32257–32269, 2022.

Pragya Srivastava, Manuj Malik, Vivek Gupta, Tanuja Ganu, and Dan Roth. Evaluating LLMs' mathematical reasoning in financial document question answering. In *Findings of the Association for Computational Linguistics: ACL 2024*, pp. 3853–3878, 2024.

Jianlin Su, Murtadha Ahmed, Yu Lu, Shengfeng Pan, Bo Wen, and Yunfeng Liu. Roformer: Enhanced transformer with rotary position embedding. *Neurocomputing*, 568:127063, 2024a.

Junwei Su, Difan Zou, and Chuan Wu. PRES: Toward scalable memory-based dynamic graph neural networks. In *The Twelfth International Conference on Learning Representations*, pp. 1–14, 2024b.

Yuan Sui, Mengyu Zhou, Mingjie Zhou, Shi Han, and Dongmei Zhang. Table meets LLM: Can large language models understand structured table data? a benchmark and empirical study. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*, pp. 645–654, 2024a.

Yuan Sui, Jiaru Zou, Mengyu Zhou, Xinyi He, Lun Du, Shi Han, and Dongmei Zhang. TAP4LLM: Table provider on sampling, augmenting, and packing semi-structured data for large language model reasoning. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pp. 10306–10323, 2024b.

Baohua Sun, Lin Yang, Wenhan Zhang, Michael Lin, Patrick Dong, Charles Young, and Jason Dong. Supertml: Two-dimensional word embedding for the precognition on structured tabular data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 2973–2981, 2019a.

Fei Sun, Jun Liu, Jian Wu, Changhua Pei, Xiao Lin, Wenwu Ou, and Peng Jiang. BERT4Rec: Sequential recommendation with bidirectional encoder representations from transformer. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pp. 1441–1450, 2019b.

Anirudh S. Sundar and Larry Heck. cTBLS: Augmenting large language models with conversational tables. In *Proceedings of the 5th Workshop on NLP for Conversational AI*, pp. 59–70, 2023.

Susheel Suresh, Mayank Shrivastava, Arko Mukherjee, Jennifer Neville, and Pan Li. Expressive and efficient representation learning for ranking links in temporal graphs. In *Proceedings of the ACM Web Conference 2023*, pp. 567–577, 2023.

Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, Massachusetts, 1998.

Van-Dai Ta, Chuan-Ming Liu, and Direselign Addis Tadesse. Portfolio optimization-based stock prediction using long-short term memory network in quantitative trading. *Applied Sciences*, 10(2):437, 2020.

Sabera J Talukder, Yisong Yue, and Georgia Gkioxari. Totem: Tokenized time series embeddings for general time series analysis. *Transactions on Machine Learning Research*, 2024.

Haoran Tang, Shiqing Wu, Guandong Xu, and Qing Li. Dynamic graph evolution learning for recommendation. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 1589–1598, 2023.

Jiaxi Tang and Ke Wang. Personalized top-N sequential recommendation via convolutional sequence embedding. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pp. 565–573, 2018.

Anton Frederik Thielmann, Arik Reuter, Thomas Kneib, David Rügamer, and Benjamin Säfken. Interpretable additive tabular transformer networks. *Transactions on Machine Learning Research*, 2024.

Yuxing Tian, Yiyan Qi, and Fan Guo. FreeDyG: Frequency enhanced continuous-time dynamic graph model for link prediction. In *The Twelfth International Conference on Learning Representations*, pp. 1–14, 2024.

Zhen Tian, Changwang Zhang, Wayne Xin Zhao, Xin Zhao, Ji-Rong Wen, and Zhao Cao. UFIN: Universal feature interaction network for multi-domain click-through rate prediction. In *Database Systems for Advanced Applications: 30th International Conference, Singapore, May 26–29, 2025, Proceedings, Part II*, volume 15987, pp. 3–18, 2025.

Ghazaleh Haratinezhad Torbati, Anna Tigunova, Andrew Yates, and Gerhard Weikum. Recommendations by concise user profiles from review text. *arXiv preprint arXiv:2311.01314*, 2023.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. LLaMA: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*, 2023.

Rakshit Trivedi, Hanjun Dai, Yichen Wang, and Le Song. Know-evolve: Deep temporal reasoning for dynamic knowledge graphs. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70, pp. 3462–3471, 2017.

Rakshit Trivedi, Mehrdad Farajtabar, Prasenjeet Biswal, and Hongyuan Zha. DyRep: Learning representations over dynamic graphs. In *International Conference on Learning Representations*, pp. 1–14, 2019.

Ciprian-Octavian Truică, Elena-Simona Apostol, Marius Marogel, and Adrian Paschke. GETAE: Graph information Enhanced deep neural NeTwork ensemble ArchitecturE for fake news detection. *arXiv preprint arXiv:2412.01825*, 2024.

Talip Ucar, Ehsan Hajiramezanali, and Lindsay Edwards. Subtab: Subsetting features of tabular data for self-supervised representation learning. In *Advances in Neural Information Processing Systems*, volume 34, pp. 18853–18865, 2021.

Dennis Ulmer, Lotta Meijerink, and Giovanni Cinà. Trust issues: Uncertainty estimation does not enable reliable OOD detection on medical tabular data. In *Proceedings of the Machine Learning for Health NeurIPS Workshop*, volume 136, pp. 341–354, 2020.

Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, pp. 5998–6008, 2017.

Ryan Volum, Sudha Rao, Michael Xu, Gabriel DesGarennes, Chris Brockett, Benjamin Van Durme, Olivia Deng, Akanksha Malhotra, and Bill Dolan. Craft an iron sword: Dynamically generating interactive game characters by prompting large language models tuned on code. In *Proceedings of the 3rd Wordplay: When Language Meets Games Workshop*, pp. 25–43, 2022.

Lukas von Ziegler, Oliver Sturman, and Johannes Bohacek. Big behavior: Challenges and opportunities in a new era of deep behavior profiling. *Neuropsychopharmacology*, 46(1):33–44, 2021.

Mengting Wan and Julian J. McAuley. Modeling ambiguity, subjectivity, and diverging viewpoints in opinion question answering systems. In *Proceedings of the 2016 IEEE 16th International Conference on Data Mining*, pp. 489–498, 2016.

Mengting Wan and Julian J. McAuley. Item recommendation on monotonic behavior chains. In *Proceedings of the 12th ACM Conference on Recommender Systems*, pp. 86–94, 2018.

Mengting Wan, Jianmo Ni, Rishabh Misra, and Julian McAuley. Addressing marketing bias in product recommendations. In *Proceedings of the Thirteenth ACM International Conference on Web Search and Data Mining*, pp. 618–626, 2020.

Benyou Wang, Donghao Zhao, Christina Lioma, Qiuchi Li, Peng Zhang, and Jakob Grue Simonsen. Encoding word order in complex embeddings. In *International Conference on Learning Representations*, pp. 1–14, 2020.

Chia-Hung Wang, Jinchen Yuan, Yingping Zeng, and Shengming Lin. A deep learning integrated framework for predicting stock index price and fluctuation via singular spectrum analysis and particle swarm optimization. *Applied Intelligence*, 54(2):1770–1797, 2024a.

Guanzhi Wang, Yuqi Xie, Yunfan Jiang, Ajay Mandlekar, Chaowei Xiao, Yuke Zhu, Linxi Fan, and Anima Anandkumar. Voyager: An open-ended embodied agent with large language models. *Transactions on Machine Learning Research*, 2024b.

Jun Wang, Lantao Yu, Weinan Zhang, Yu Gong, Yinghui Xu, Benyou Wang, Peng Zhang, and Dell Zhang. IRGAN: A minimax game for unifying generative and discriminative information retrieval models. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 515–524, 2017.

Lu Wang, Wei Zhang, Xiaofeng He, and Hongyuan Zha. Supervised reinforcement learning with recurrent neural network for dynamic treatment recommendation. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2447–2456, 2018.

Lu Wang, Xiaofu Chang, Shuang Li, Yunfei Chu, Hui Li, Wei Zhang, Xiaofeng He, Le Song, Jingren Zhou, and Hongxia Yang. TCL: Transformer-based dynamic graph modelling via contrastive learning. *arXiv preprint arXiv:2105.07944*, 2021a.

Shengkun Wang, Taoran Ji, Linhan Wang, Yanshen Sun, Shang-Ching Liu, Amit Kumar, and Chang-Tien Lu. StockTime: A time series specialized large language model architecture for stock price prediction. *arXiv preprint arXiv:2409.08281*, 2024c.

Shoujin Wang and Longbing Cao. Inferring implicit rules by learning explicit and hidden item dependency. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 50(3):935–946, 2020.

Shoujin Wang, Liang Hu, Yan Wang, Longbing Cao, Quan Z. Sheng, and Mehmet A. Orgun. Sequential recommender systems: Challenges, progress and prospects. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, pp. 6332–6338, 2019.

Tianze Wang, Maryam Honari-Jahromi, Styliani Katsarou, Olga Mikheeva, Theodoros Panagiotakopoulos, Sahar Asadi, and Oleg Smirnov. player2vec: A language modeling approach to understand player behavior in games. *arXiv preprint arXiv:2404.04234*, 2024d.

Tianze Wang, Maryam Honarijahromi, Styliani Katsarou, Olga Mikheeva, Theodoros Panagiotakopoulos, Oleg Smirnov, Lele Cao, and Sahar Asadi. Understanding players as if they are talking to the game in a customized language: A pilot study. In *Proceedings of the 1st Workshop on Customizable NLP: Progress and Challenges in Customizing NLP for a Domain, Application, Group, or Individual*, pp. 47–52, 2024e.

Wuhong Wang, Jianhui Ma, Yuren Zhang, Kai Zhang, Junzhe Jiang, Yihui Yang, Yacong Zhou, and Zheng Zhang. Intent oriented contrastive learning for sequential recommendation. In *Proceedings of the Thirty-Ninth AAAI Conference on Artificial Intelligence*, volume 39, pp. 12748–12756, 2025.

Xinzhi Wang, Jiayu Guo, Xiangfeng Luo, and Hang Yu. DyHDGE: Dynamic heterogeneous transaction graph embedding for safety-centric fraud detection in financial scenarios. *Journal of Safety Science and Resilience*, 5(4):486–497, 2024f.

Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc V. Le, Ed H. Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. Self-consistency improves chain of thought reasoning in language models. In *International Conference on Learning Representations*, 2023a.

Xuhong Wang, Ding Lyu, Mengjian Li, Yang Xia, Qi Yang, Xinwen Wang, Xinguang Wang, Ping Cui, Yupu Yang, Bowen Sun, and Zhenyu Guo. APAN: Asynchronous propagation attention network for real-time temporal graph embedding. In *Proceedings of the 2021 International Conference on Management of Data*, pp. 2628–2638, 2021b.

Yanbang Wang, Yen-Yu Chang, Yunyu Liu, Jure Leskovec, and Pan Li. Inductive representation learning in temporal networks via causal anonymous walks. In *International Conference on Learning Representations*, pp. 1–14, 2021c.

Yanbang Wang, Pan Li, Chongyang Bai, and Jure Leskovec. TEDIC: Neural modeling of behavioral patterns in dynamic social interaction networks. In *Proceedings of the Web Conference 2021*, pp. 693–705, 2021d.

Yancheng Wang, Ziyan Jiang, Zheng Chen, Fan Yang, Yingxue Zhou, Eunah Cho, Xing Fan, Yanbin Lu, Xiaojiang Huang, and Yingzhen Yang. Recmind: Large language model powered agent for recommendation. In *Findings of the Association for Computational Linguistics: NAACL 2024*, pp. 4351–4364, 2024g.

Ye Wang, Jiahao Xun, Minjie Hong, Jieming Zhu, Tao Jin, Wang Lin, Haoyuan Li, Linjun Li, Yan Xia, Zhou Zhao, and Zhenhua Dong. EAGER: Two-stream generative recommender with behavior-semantic collaboration. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 3245–3254, 2024h.

Ziming Wang, Qianru Wu, Baolin Zheng, Junjie Wang, Kaiyu Huang, and Yanjie Shi. Sequence as genes: An user behavior modeling framework for fraud transaction detection in e-commerce. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 5194–5203, 2023b.

Jason Wei, Yi Tay, Rishi Bommasani, Colin Raffel, Barret Zoph, Sebastian Borgeaud, Dani Yogatama, Maarten Bosma, Denny Zhou, Donald Metzler, Ed H. Chi, Tatsunori Hashimoto, Oriol Vinyals, Percy Liang, Jeff Dean, and William Fedus. Emergent abilities of large language models. *Transactions on Machine Learning Research*, 2022a.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc V Le, and Denny Zhou. Chain-of-thought prompting elicits reasoning in large language models. In *Advances in Neural Information Processing Systems*, volume 35, pp. 24824–24837, 2022b.

Zhihao Wen and Yuan Fang. TREND: TempoRal event and node dynamics for graph representation learning. In *Proceedings of the ACM Web Conference 2022*, pp. 1159–1169, 2022.

Ulme Wennberg and Gustav Eje Henter. The case for translation-invariant self-attention in transformer-based language models. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pp. 130–140, 2021.

Robert C. Wilson and Anne G. E. Collins. Ten simple rules for the computational modeling of behavioral data. *eLife*, 8:e49547, 2019.

Cheng Wu, Chaokun Wang, Jingcao Xu, Ziwei Fang, Tiankai Gu, Changping Wang, Yang Song, Kai Zheng, Xiaowei Wang, and Guorui Zhou. Instant representation learning for recommendation over large dynamic graphs. In *Proceedings of the 2023 IEEE 39th International Conference on Data Engineering*, pp. 82–95, 2023a.

Chuhan Wu, Fangzhao Wu, and Yongfeng Huang. DA-transformer: Distance-aware transformer. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 2059–2068, 2021.

Jiahao Wu, Qijiong Liu, Hengchang Hu, Wenqi Fan, Shengcai Liu, Qing Li, Xiao-Ming Wu, and Ke Tang. Leveraging ChatGPT to empower training-free dataset condensation for content-based recommendation. In *Companion Proceedings of the ACM on Web Conference 2025*, pp. 1402–1406, 2025a.

Jiapeng Wu, Meng Cao, Jackie Chi Kit Cheung, and William L Hamilton. TeMP: Temporal message passing for temporal knowledge graph completion. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, pp. 5730–5746, 2020.

Jing Wu, Suiyao Chen, Qi Zhao, Renat Sergazinov, Chen Li, Shengjie Liu, Chongchao Zhao, Tianpei Xie, Hanqing Guo, Cheng Ji, Daniel Cociorva, and Hakan Brunzell. SwitchTab: Switched autoencoders are effective tabular learners. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 15924–15933, 2024a.

Likang Wu, Zhilan Zheng, Zhaopeng Qiu, Hao Wang, Hongchao Gu, Tingjia Shen, Chuan Qin, Chen Zhu, Hengshu Zhu, Qi Liu, Hui Xiong, and Enhong Chen. A survey on large language models for recommendation. *arXiv preprint arXiv:2305.19860*, 2023b.

Shijie Wu, Ozan Irsoy, Steven Lu, Vadim Dabravolski, Mark Dredze, Sebastian Gehrmann, Prabhanjan Kambadur, David Rosenberg, and Gideon Mann. BloombergGPT: A large language model for finance. *arXiv preprint arXiv:2303.17564*, 2023c.

Wenhao Wu, Xiaojie Li, Lin Wang, Jialiang Zhou, Di Wu, Qinye Xie, Qingheng Zhang, Yin Zhang, Shuguang Han, Fei Huang, and Jufeng Chen. IU4Rec: Interest unit-based product organization and recommendation for e-commerce platform. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, volume 2, pp. 5039–5048, 2025b.

Yuxia Wu, Yuan Fang, and Lizi Liao. On the feasibility of simple transformer for dynamic graph modeling. In *Proceedings of the ACM Web Conference 2024*, pp. 1–11, 2024b.

Zihao Wu, Xin Wang, Heng Chang, Hong Chen, Lifeng Sun, and Wenwu Zhu. Aligning large multimodal model with sequential recommendation via content-behavior guidance. In *Proceedings of the 2025 International Conference on Multimedia Retrieval*, pp. 1507–1516, 2025c.

Maciej Wysocki and Paweł Sakowski. Investment portfolio optimization based on modern portfolio theory and deep learning models. Technical Report Working Papers No. 12/2022 (388), Faculty of Economic Sciences, University of Warsaw, 2022.

Isadora Xavier, Edvaldo Veiga, Andrea Maria NC Ribeiro, Rodrigo de Paula Monteiro, Marcos Oliveira, Rodrigo Amaro, and Diego Pinheiro. Towards the identification of money laundering in bank transactions using recurrent neural networks. In *Proceedings of the 2024 IEEE Latin American Conference on Computational Intelligence*, pp. 1–5, 2024.

Yunjia Xi, Weiwen Liu, Jianghao Lin, Xiaoling Cai, Hong Zhu, Jieming Zhu, Bo Chen, Ruiming Tang, Weinan Zhang, and Yong Yu. Towards open-world recommendation with knowledge augmentation from large language models. In *Proceedings of the 18th ACM Conference on Recommender Systems*, pp. 12–22, 2024.

Shuai Xiao, Junchi Yan, Mehrdad Farajtabar, Le Song, Xiaokang Yang, and Hongyuan Zha. Learning time series associated event sequences with recurrent point process networks. *IEEE transactions on neural networks and learning systems*, 30(10):3124–3136, 2019.

Yijia Xiao, Edward Sun, Di Luo, and Wei Wang. Tradingagents: Multi-agents LLM financial trading framework. *arXiv preprint arXiv:2412.20138*, 2024.

Xu Xie, Fei Sun, Zhaoyang Liu, Shiwen Wu, Jinyang Gao, Jiandong Zhang, Bolin Ding, and Bin Cui. Contrastive learning for sequential recommendation. In *Proceedings of the 2022 IEEE 38th International Conference on Data Engineering*, pp. 1259–1273, 2022.

Luolin Xiong, Haofen Wang, Xi Chen, Lu Sheng, Yun Xiong, Jingping Liu, Yanghua Xiao, Huajun Chen, Qing-Long Han, and Yang Tang. Deepseek: Paradigm shifts and technical evolution in large AI models. *IEEE/CAA Journal of Automatica Sinica*, 12(5):841–858, 2025.

Chenwei Xu, Yu-Chao Huang, Jerry Yao-Chieh Hu, Weijian Li, Ammar Gilani, Hsi-Sheng Goan, and Han Liu. Bishop: Bi-directional cellular learning for tabular data with generalized sparse modern hopfield model. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235, pp. 55048–55075, 2024.

Da Xu, Chuanwei Ruan, Evren Korpeoglu, Sushant Kumar, and Kannan Achan. Self-attention with functional time representation learning. In *Advances in Neural Information Processing Systems*, volume 32, pp. 15915–15925, 2019.

Da Xu, Chuanwei Ruan, Evren Korpeoglu, Sushant Kumar, and Kannan Achan. Inductive representation learning on temporal graphs. In *Proceedings of the 8th International Conference on Learning Representations*, pp. 1–14, 2020.

Junhao Xu. A multimodal adaptive graph-based intelligent classification model for fake news. *arXiv preprint arXiv:2411.06097*, 2024.

Zhiyu Xu, Yi Liu, Yuchi Wang, Ruihan Bao, Keiko Harimoto, and Xu Sun. Modeling interactions between stocks using LLM-enhanced graphs for volume prediction. In *Proceedings of the Joint Workshop of the 9th Financial Technology and Natural Language Processing, the 6th Financial Narrative Processing, and the 1st Workshop on Large Language Models for Finance and Legal*, pp. 153–163, 2025.

Hansheng Xue, Luwei Yang, Wen Jiang, Yi Wei, Yi Hu, and Yu Lin. Modeling dynamic heterogeneous network for link prediction using hierarchical attention with temporal RNN. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2020, Ghent, Belgium, September 14–18, 2020, Proceedings, Part I*, volume 12457, pp. 282–298, 2021.

Siqiao Xue, Yan Wang, Zhixuan Chu, Xiaoming Shi, Caigao Jiang, Hongyan Hao, Gangwei Jiang, Xiaoyun Feng, James Y. Zhang, and Jun Zhou. Prompt-augmented temporal point process for streaming event sequence. In *Advances in Neural Information Processing Systems*, volume 36, pp. 18885–18905, 2023.

An Yan, Zhankui He, Jiacheng Li, Tianyang Zhang, and Julian J. McAuley. Personalized showcases: Generating multi-modal explanations for recommendations. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 2251–2255, 2023.

Jiahuan Yan, Bo Zheng, Hongxia Xu, Yiheng Zhu, Danny Z. Chen, Jimeng Sun, Jian Wu, and Jintai Chen. Making pre-trained language models great on tabular prediction. In *International Conference on Learning Representations*, pp. 1–14, 2024.

Bufang Yang, Yunqi Guo, Lilin Xu, Zhenyu Yan, Hongkai Chen, Guoliang Xing, and Xiaofan Fred Jiang. SocialMind: LLM-based proactive AR social assistive system with human-like perception for in-situ live interactions. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 9(1): 1–30, 2025a.

Chaoqi Yang, Ruijie Wang, Shuochao Yao, Shengzhong Liu, and Tarek F. Abdelzaher. Revisiting over-smoothing in deep GCNs. *arXiv preprint arXiv:2003.13663*, 2020a.

Menglin Yang, Min Zhou, Marcus Kalander, Zengfeng Huang, and Irwin King. Discrete-time temporal network embedding via implicit hierarchical learning in hyperbolic space. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 1975–1985, 2021.

Shenghao Yang, Weizhi Ma, Peijie Sun, Qingyao Ai, Yiqun Liu, Mingchen Cai, and Min Zhang. Sequential recommendation with latent relations based on large language model. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 335–344, 2024a.

Susen Yang, Yong Liu, Yonghui Xu, Chunyan Miao, Min Wu, and Juyong Zhang. Contextualized graph attention network for recommendation with item knowledge graph. *arXiv preprint arXiv:2004.11529*, 2020b.

Yang Yang, Lijun Zhang, and Jiaquan Liu. Temporal user interest modeling for online advertising using Bi-LSTM network improved by an updated version of parrot optimizer. *Scientific Reports*, 15(1):18858, 2025b.

Yuxin Yang, Hongkuan Zhou, Rajgopal Kannan, and Viktor K. Prasanna. Towards ideal temporal graph neural networks: Evaluations and conclusions after 10,000 GPU hours. *Proceedings of the VLDB Endowment*, 18(4):956–969, 2024b.

Liang Yao, Chengsheng Mao, and Yuan Luo. Graph convolutional networks for text classification. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence*, volume 33, pp. 7370–7377, 2019.

Lin Yao, Luning Wang, Lv Pan, and Kai Yao. Link prediction based on common-neighbors for dynamic social network. *Procedia Computer Science*, 83:82–89, 2016.

Ghim-Eng Yap, Xiao-Li Li, and Philip S. Yu. Effective next-items recommendation via personalized sequential pattern mining. In *Database Systems for Advanced Applications: 17th International Conference, Busan, South Korea, April 15-18, 2012, Proceedings, Part II*, volume 7239, pp. 48–64, 2012.

Chao Ye, Guoshan Lu, Haobo Wang, Liyao Li, Sai Wu, Gang Chen, and Junbo Zhao. Towards cross-table masked pretraining for web data mining. In *Proceedings of the ACM Web Conference 2024*, pp. 4449–4459, 2024a.

Han-Jia Ye, Si-Yang Liu, Hao-Run Cai, Qi-Le Zhou, and De-Chuan Zhan. A closer look at deep learning methods on tabular datasets. *arXiv preprint arXiv:2407.00956*, 2024b.

Wen Ye, Wei Yang, Defu Cao, Yizhou Zhang, Lumingyuan Tang, Jie Cai, and Yan Liu. Domain-oriented time series inference agents for reasoning and automated analysis. *arXiv preprint arXiv:2410.04047*, 2024c.

Yunan Ye, Hengzhi Pei, Boxin Wang, Pin-Yu Chen, Yada Zhu, Ju Xiao, and Bo Li. Reinforcement-learning based portfolio management with augmented asset movement prediction states. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence*, volume 34, pp. 1112–1119, 2020.

Jinsung Yoon, Yao Zhang, James Jordon, and Mihaela van der Schaar. VIME: Extending the success of self- and semi-supervised learning to tabular domain. In *Advances in Neural Information Processing Systems*, volume 33, pp. 11033–11043, 2020.

Jiaxuan You, Yichen Wang, Aditya Pal, Pong Eksombatchai, Chuck Rosenberg, and Jure Leskovec. Hierarchical temporal convolutional networks for dynamic recommender systems. In *Proceedings of the 28th ACM World Wide Web Conference*, pp. 2236–2246, 2019.

Jiaxuan You, Tianyu Du, and Jure Leskovec. ROLAND: Graph learning framework for dynamic graphs. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 2358–2366, 2022.

Bowen Yu, Cheng Fu, Haiyang Yu, Fei Huang, and Yongbin Li. Unified language representation for question answering over text, tables, and images. In *Findings of the Association for Computational Linguistics: ACL 2023*, pp. 4756–4765, 2023a.

Le Yu, Leilei Sun, Bowen Du, and Weifeng Lv. Towards better dynamic graph learning: New architecture and unified library. In *Advances in Neural Information Processing Systems*, volume 36, pp. 67686–67700, 2023b.

Yangyang Yu, Zhiyuan Yao, Haohang Li, Zhiyang Deng, Yuechen Jiang, Yupeng Cao, Zhi Chen, Jordan W. Suchow, Zhenyu Cui, Rong Liu, Zhaozhuo Xu, Denghui Zhang, Koduvayur Subbalakshmi, Guojun Xiong, Yueru He, Jimin Huang, Dong Li, and Qianqian Xie. FinCon: A synthesized LLM multi-agent system with conceptual verbal reinforcement for enhanced financial decision making. In *Advances in Neural Information Processing Systems*, volume 37, pp. 137010–137045, 2024.

Fajie Yuan, Alexandros Karatzoglou, Ioannis Arapakis, Joemon M. Jose, and Xiangnan He. A simple convolutional generative network for next item recommendation. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, pp. 582–590, 2019.

Haoqi Yuan, Chi Zhang, Hongcheng Wang, Feiyang Xie, Penglin Cai, Hao Dong, and Zongqing Lu. Skill reinforcement learning and planning for open-world long-horizon tasks. *arXiv preprint arXiv:2303.16563*, 2023.

Ruoxin Yuan and Guanhua Fang. Residual TPP: A unified lightweight approach for event stream data analysis. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267, pp. 73455–73477, 2025.

Yiwen Yuan, Zecheng Zhang, Xinwei He, Akihiro Nitta, Weihua Hu, Manan Shah, Blaž Stojanovič, Shenyang Huang, Jan Eric Lenssen, Jure Leskovec, and Matthias Fey. ContextGNN: Beyond two-tower recommendation systems. In *Proceedings of the International Conference on Learning Representations*, pp. 1–14, 2025.

Mateo Espinosa Zarlenga, Zohreh Shams, Michael Edward Nelson, Been Kim, and Mateja Jamnik. Tabcbm: Concept-based interpretable neural networks for tabular data. *Transactions on Machine Learning Research*, 2023.

Chiyu Zhang, Yifei Sun, Jun Chen, Jie Lei, Muhammad Abdul-Mageed, Sinong Wang, Rong Jin, Sem Park, Ning Yao, and Bo Long. SPAR: Personalized content-based recommendation via long engagement attention. *arXiv preprint arXiv:2402.10555*, 2024a.

Hengyuan Zhang, Peng Chang, and Zongcheng Ji. Bridging the gap: Deciphering tabular data using large language model. *arXiv preprint arXiv:2308.11891*, 2023a.

Jia-Dong Zhang, Chi-Yin Chow, and Yanhua Li. LORE: Exploiting sequential influence for location recommendations. In *Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 103–112, 2014.

Jiarui Zhang. Guided profile generation improves personalization with large language models. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pp. 4005–4016, 2024.

Jiasheng Zhang, Jialin Chen, Menglin Yang, Aosong Feng, Shuang Liang, Jie Shao, and Rex Ying. DTGB: A comprehensive benchmark for dynamic text-attributed graphs. In *Advances in Neural Information Processing Systems*, volume 37, 2024b.

Junjie Zhang, Wenqi Sun, Yupeng Hou, Wayne Xin Zhao, and Ji-Rong Wen. Review-enhanced universal sequence representation learning for recommender systems. *ACM Transactions on Information Systems*, 43:1–31, 2025a.

Mengqi Zhang, Shu Wu, Xueli Yu, Qiang Liu, and Liang Wang. Dynamic graph neural networks for sequential recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 35(5):4741–4753, 2023b.

Qiang Zhang, Aldo Lipani, Omer Kirnap, and Emine Yilmaz. Self-attentive Hawkes process. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119, pp. 11183–11193, 2020.

Shuai Zhang, Yi Tay, Lina Yao, and Aixin Sun. Next item recommendation with self-attention. *arXiv preprint arXiv:1808.06414*, 2018.

Si Zhang, Hanghang Tong, Jiejun Xu, and Ross Maciejewski. Graph convolutional networks: a comprehensive review. *Computational Social Networks*, 6(11):1–23, 2019a.

Siwei Zhang, Yun Xiong, Yao Zhang, Yiheng Sun, Xi Chen, Yizhu Jiao, and Yangyong Zhu. RDGSL: Dynamic graph representation learning with structure learning. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, pp. 3174–3183, 2023c.

Siwei Zhang, Yun Xiong, Yao Zhang, Xixi Wu, Yiheng Sun, and Jiawei Zhang. iLoRE: Dynamic graph representation with instant long-term modeling and re-occurrence preservation. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, pp. 3216–3225, 2023d.

Siwei Zhang, Xi Chen, Yun Xiong, Xixi Wu, Yizhu Jiao, Yao Zhang, Mingyang Zhang, Tengfei Liu, Weiqiang Wang, and Jiawei Zhang. Interactions exhibit clustering rhythm: A prevalent observation for advancing temporal link prediction. In *Openreview*, 2024c.

Siwei Zhang, Xi Chen, Yun Xiong, Xixi Wu, Yao Zhang, Yongrui Fu, Yinglong Zhao, and Jiawei Zhang. Towards adaptive neighborhood for advancing temporal interaction graph modeling. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 4290–4301, 2024d.

Siwei Zhang, Yun Xiong, Yateng Tang, Jiarong Xu, Xi Chen, Zehao Gu, Xuezheng Hao, Zian Jia, and Jiawei Zhang. Unifying text semantics and graph structures for temporal text-attributed graphs with large language models. In *Advances in Neural Information Processing Systems*, 2025b.

Tingting Zhang, Pengpeng Zhao, Yanchi Liu, Victor S. Sheng, Jiajie Xu, Deqing Wang, Guanfeng Liu, and Xiaofang Zhou. Feature-level deeper self-attention network for sequential recommendation. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, pp. 4320–4326, 2019b.

Tonghui Zhang, Ming Da Huo, Zhaozhao Ma, Jiajun Hu, Qian Liang, and Heng Chen. Prediction model of stock return on investment based on hybrid DNN and TabNet model. *PeerJ Computer Science*, 10:e2057, 2024e.

Weixu Zhang, Yifei Wang, Yuanfeng Song, Victor Junqiu Wei, Yuxing Tian, Yiyan Qi, Jonathan H. Chan, Raymond Chi-Wing Wong, and Haiqin Yang. Natural language interfaces for tabular data querying and visualization: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 36(11):6699–6718, 2024f.

Wentao Zhang, Lingxuan Zhao, Haochong Xia, Shuo Sun, Jiaze Sun, Molei Qin, Xinyi Li, Yuqing Zhao, Yilei Zhao, Xinyu Cai, Longtao Zheng, Xinrun Wang, and Bo An. A multimodal foundation agent for financial trading: Tool-augmented, diversified, and generalist. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 4314–4325, 2024g.

Yao Zhang, Yun Xiong, Yongxiang Liao, Yiheng Sun, Yucheng Jin, Xuehao Zheng, and Yangyong Zhu. TIGER: Temporal interaction graph embedding with restarts. In *Proceedings of the 32nd ACM World Wide Web Conference*, pp. 478–488, 2023e.

Zeyang Zhang, Ziwei Zhang, Xin Wang, Yijian Qin, Zhou Qin, and Wenwu Zhu. Dynamic heterogeneous graph attention neural architecture search. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 11307–11315, 2023f.

Bowen Zhao, Changkai Ji, Yuejie Zhang, Wen He, Yingwen Wang, Qing Wang, Rui Feng, and Xiaobo Zhang. Large language models are complex table parsers. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pp. 14786–14802, 2023a.

Kai Zhao, Yukun Zheng, Tao Zhuang, Xiang Li, and Xiaoyi Zeng. Joint learning of e-commerce search and recommendation with a unified graph neural network. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, pp. 1461–1469, 2022.

Ling Zhao, Yujiao Song, Chao Zhang, Yu Liu, Pu Wang, Tao Lin, Min Deng, and Haifeng Li. T-GCN: A temporal graph convolutional network for traffic prediction. *IEEE Transactions on Intelligent Transportation Systems*, 21(9):3848–3858, 2020.

Sha Zhao, Yizhi Xu, Zhiling Luo, Jianrong Tao, Shijian Li, Changjie Fan, and Gang Pan. Player behavior modeling for enhancing role-playing game engagement. *IEEE Transactions on Computational Social Systems*, 8(2):464–474, 2021.

Xiangyu Zhao, Maolin Wang, Xinjian Zhao, Jiansheng Li, Shucheng Zhou, Dawei Yin, Qing Li, Jiliang Tang, and Ruocheng Guo. Embedding in recommender systems: A survey. *arXiv preprint arXiv:2310.18608*, 2023b.

Yibo Zhao, Jiapeng Zhu, Can Xu, Yao Liu, and Xiang Li. Enhancing LLM-based hatred and toxicity detection with meta-toxic knowledge graph. In *Findings of the Association for Computational Linguistics: ACL 2025*, pp. 24747–24760, 2025.

Ziwei Zhao, Xi Zhu, Tong Xu, Aakas Lizhiyu, Yu Yu, Xueying Li, Zikai Yin, and Enhong Chen. Time-interval aware share recommendation via bi-directional continuous time dynamic graphs. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 822–831, 2023c.

Bowen Zheng, Yupeng Hou, Hongyu Lu, Yu Chen, Wayne Xin Zhao, Ming Chen, and Ji-Rong Wen. Adapting large language models by integrating collaborative semantics for recommendation. In *Proceedings of the 2024 IEEE 40th International Conference on Data Engineering*, pp. 1435–1448, 2024.

Shangfei Zheng, Hongzhi Yin, Tong Chen, Quoc Viet Hung Nguyen, Wei Chen, and Lei Zhao. DREAM: Adaptive reinforcement learning based on attention mechanism for temporal knowledge graph reasoning. *arXiv preprint arXiv:2304.03984*, 2023.

Yuchen Zhong, Guangming Sheng, Tianzuo Qin, Minjie Wang, Quan Gan, and Chuan Wu. GNNFlow: A distributed framework for continuous temporal GNN learning on dynamic graphs. *arXiv preprint arXiv:2311.17410*, 2023.

Doudou Zhou, Han Tong, Linshanshan Wang, Suqi Liu, Xin Xiong, Ziming Gan, Romain Griffier, Boris P. Hejblum, Yun-Chung Liu, Chuan Hong, Clara-Lea Bonzel, Tianrun Cai, Kevin Pan, Yuk-Lam Ho, Lauren Costa, Vidul Ayakulangara Panickan, J. Michael Gaziano, Kenneth D. Mandl, Vianney Jouhet, Rodolphe Thiebaut, Zongqi Xia, Kelly Cho, Katherine P. Liao, and Tianxi Cai. Representation learning to advance multi-institutional studies with electronic health record data. *arXiv preprint arXiv:2502.08547*, 2025a.

Fan Zhou, Xovee Xu, Ce Li, Goce Trajcevski, Ting Zhong, and Kunpeng Zhang. A heterogeneous dynamical graph neural networks approach to quantify scientific impact. *arXiv preprint arXiv:2003.12042*, 2020a.

Feng Zhou, Quyu Kong, Jie Qiao, Cheng Wan, Yixuan Zhang, and Ruichu Cai. Advances in temporal point processes: Bayesian, neural, and llm approaches. *arXiv preprint arXiv:2501.14291*, 2025b.

Hongkuan Zhou, Da Zheng, Israt Nisa, Vasileios Ioannidis, Xiang Song, and George Karypis. TGL: A general framework for temporal GNN training on billion-scale graphs. *Proceedings of the VLDB Endowment*, 15 (8):1572–1580, 2022.

Hongkuan Zhou, Da Zheng, Xiang Song, George Karypis, and Viktor Prasanna. DistTGL: Distributed memory-based temporal graph neural network training. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage, and Analysis*, pp. 1–11, 2023.

Joyce Zhou, Yijia Dai, and Thorsten Joachims. Language-based user profiles for recommendation. *arXiv preprint arXiv:2402.15623*, 2024.

Kun Zhou, Hui Wang, Wayne Xin Zhao, Yutao Zhu, Sirui Wang, Fuzheng Zhang, Zhongyuan Wang, and Ji-Rong Wen. S3-Rec: Self-supervised learning for sequential recommendation with mutual information maximization. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management*, pp. 1893–1902, 2020b.

Lekui Zhou, Yang Yang, Xiang Ren, Fei Wu, and Yueting Zhuang. Dynamic network embedding by modeling triadic closure process. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*, volume 32, 2018.

Yujia Zhou, Hexi Wang, Qingyao Ai, Zhen Wu, and Yiqun Liu. Investigating prosocial behavior theory in LLM agents under policy-induced inequities. *arXiv preprint arXiv:2505.15857*, 2025c.

Bingzhao Zhu, Xingjian Shi, Nick Erickson, Mu Li, George Karypis, and Mahsa Shoaran. Xtab: Cross-table pretraining for tabular transformers. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202, pp. 43181–43204, 2023a.

Feng Zhu, Chaochao Chen, Yan Wang, Guanfeng Liu, and Xiaolin Zheng. DTCDR: A framework for dual-target cross-domain recommendation. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pp. 1533–1542, 2019a.

Han Zhu, Xiang Li, Pengye Zhang, Guozheng Li, Jie He, Han Li, and Kun Gai. Learning tree-based deep model for recommender systems. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1079–1088, 2018.

Han Zhu, Daqing Chang, Ziru Xu, Pengye Zhang, Xiang Li, Jie He, Han Li, Jian Xu, and Kun Gai. Joint optimization of tree-based index and deep model for recommender systems. In *Advances in Neural Information Processing Systems*, volume 32, pp. 3973–3982, 2019b.

Yifan Zhu, Fangpeng Cong, Dan Zhang, Wenwen Gong, Qika Lin, Wenzheng Feng, Yuxiao Dong, and Jie Tang. WinGNN: Dynamic graph neural networks with random gradient aggregation window. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 3650–3662, 2023b.

Yitan Zhu, Thomas S. Brettin, Fangfang Xia, Alexander Partin, Maulik Shukla, Hyunseung Yoo, Yvonne A. Evrard, James H. Doroshow, and Rick L. Stevens. Converting tabular data into images for deep learning with convolutional neural networks. *Scientific Reports*, 11(1):11325, 2021.

Jingwei Zhuo, Ziru Xu, Wei Dai, Han Zhu, Han Li, Jian Xu, and Kun Gai. Learning optimal tree models under beam search. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119, pp. 11650–11659, 2020.

Tao Zou, Yuhao Mao, Junchen Ye, and Bowen Du. Repeat-aware neighbor sampling for dynamic graph learning. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 4722–4733, 2024.

Simiao Zuo, Haoming Jiang, Zichong Li, Tuo Zhao, and Hongyuan Zha. Transformer hawkes process. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119, pp. 11692–11702, 2020.

Yuan Zuo, Guannan Liu, Hao Lin, Jia Guo, Xiaoqian Hu, and Junjie Wu. Embedding temporal network via neighborhood formation. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2857–2866, 2018.