General-Purpose User Modeling with Behavioral Logs

A Snapchat Case Study

Qixiang Fang*
q.fang@uu.nl
Utrecht University
Utrecht, Netherlands

Zhihan Zhou*
zhihanzhou2020@u.northwestern.edu
Northwestern University
Evanston, Illinois, USA

Francesco Barbieri
Yozen Liu
Leonardo Neves
Snap Inc.
Santa Monica, California, USA

Dong Nguyen
Daniel Oberski
Utrecht University
Utrecht, Netherlands

Maarten Bos
Ron Dotsch
Snap Inc.
Santa Monica, California, USA

ABSTRACT

Learning general-purpose user representations based on user behavioral logs is an increasingly popular user modeling approach. It benefits from easily available, privacy-friendly yet expressive data, and does not require extensive re-tuning of the upstream user model for different downstream tasks. While this approach has shown promise in search engines and e-commerce applications, its fit for instant messaging platforms, a cornerstone of modern digital communication, remains largely uncharted. We explore this research gap using Snapchat data as a case study. Specifically, we implement a Transformer-based user model with customized training objectives and show that the model can produce high-quality user representations across a broad range of evaluation tasks, among which we introduce three new downstream tasks that concern pivotal topics in user research: user safety, engagement and churn. We also tackle the challenge of efficient extrapolation of long sequences at inference time, by applying a novel positional encoding method.

CCS CONCEPTS

• **Human-centered computing** → **User models**.

KEYWORDS

Transformer, Representation Learning, User Safety, User Churn

ACM Reference Format:
Qixiang Fang, Zhihan Zhou, Francesco Barbieri, Yozen Liu, Leonardo Neves, Dong Nguyen, Daniel Oberski, Maarten Bos, and Ron Dotsch. 2024. General-Purpose User Modeling with Behavioral Logs: A Snapchat Case Study. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '24), July 14–18, 2024, Washington, DC, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3626772.3657908

*Both authors contributed equally to this research.

SIGIR '24, July 14–18, 2024, Washington, DC, USA

ACM ISBN 979-8-4007-0431-4/24/07
https://doi.org/10.1145/3626772.3657908

1 INTRODUCTION

Instant messaging (IM) apps, such as WhatsApp, WeChat and Snapchat, have become an indispensable part of modern lives [4]. For businesses running these platforms, understanding their users' needs and delivering a superior experience is paramount. For example, with growing concerns about harassment, scams, and bots [18], robust safety measures like timely identification and mitigation of malicious accounts are crucial. Furthermore, some IM apps (e.g., Snapchat and WeChat) suggest content (e.g., news, photos, videos) to users, requiring personalized recommendations. To this end, many IM platforms leverage *user modeling* techniques.

User modeling concerns the process of creating representations of users that can capture and predict their actions, preferences, or intentions [13]. In this process, various data sources may be used, including user attributes [39], social media posts [2, 8], and social networks [8, 26, 35]. An increasingly recognized alternative is *user behavioral logs*, which are records of high-resolution, low-level events triggered by user actions in an information system [1]. Examples are swiping left, viewing a video, and sending a text.

Using user behavioral logs for user modeling has many benefits. For instance, these logs contain rich and complex patterns, providing the basis for nuanced user representations. Also, because users tend to display consistent behaviors over time [10], behavioral logs remain more similar for the same user at different time points than between between two different users. This helps a model to learn personalized representations. Furthermore, the need for active user input like surveys is eliminated, thus reducing user burden. Lastly, behavioral logs are more privacy-friendly, unlike demographics, user-generated content (e.g., texts, images, videos), and locations, which users often consider private and sensitive [14, 16, 24, 31].

For brevity, we refer to user modeling approaches that leverage user behavioral logs as **BLUM** (**B**ehavioral **L**og-based **U**ser **M**odeling). BLUM can be either *task-specific* or *General-purpose*. In task-specific BLUM, separate user models are trained for different downstream tasks (e.g., [5, 6, 9, 22, 23, 36, 38, 40]). In contrast, in *General-purpose* BLUM (**G-BLUM**), a primary, upstream user model learns general-purpose user representations not specific to

a downstream task through some general training objectives [39]. Then, user representations are extracted from this upstream model and used as direct input in downstream tasks, which often use lightweight learners (e.g., linear regression [7, 11], logistic regression [7] and multilayer perceptron [39]). In this process, only the downstream models need updating, thus eliminating the need for extensive re-tuning of the upstream model for different tasks.

In this paper, we focus on G-BLUM systems for its efficiency, and make *three* contributions. *First*, while G-BLUM has shown success in platforms like search engines and e-commerce services, their applicability to IM apps remains largely uncharted and requires dedicated investigation due to distinct platform differences. For one, IM apps involve sustained and dynamic user interactions spanning minutes to hours [17], in contrast to the short and focused interactions of search engines [20], or occasional visits to e-commerce platforms [21]. For another, user intentions vary widely on IM apps, from casual chats to urgent communications, unlike the clearer intentions on e-commerce platforms and search engines [19]. We thus bridge this literature gap through rigorous experiments and evaluations with Snapchat data.

Second, we implement a customized *Transformer*-based [34] user model for Snapchat users, with two training objectives: *Masked Behavior Prediction* and *User Contrastive Learning* (§ 4.2). Furthermore, as users remain engaged, their behavior sequences naturally extend and can therefore exceed the maximum length encountered during training. Making inferences for such sequences remains a challenge in G-BLUM systems. To overcome this, we utilize a novel position encoding method: Attention with Linear Biases (ALiBi), which performs efficient extrapolation of varying-length sequences [28]. This marks an improvement over previous studies, where fixed input sequence lengths were used (e.g., [7, 33, 39]),

Lastly, we introduce three downstream tasks unseen in previous G-BLUM studies: Reported Account Prediction, Ad View Time Prediction, and Account Self-deletion Prediction. These concern user safety, engagement and churn, three pivotal topics in user research.

2 RELATED WORK

Yang et al. (2017) [37] applied the word2vec algorithm [25] to behavior sequences of Adobe Photoshop users. This research was extended by Tao et al. (2019) [33] with an encoder-decoder model. Chen et al. (2018) [7] used a *recurrent neural network* (RNN) to model user behavior sequences from commercial websites. More recent G-BLUM studies adopt the *Transformer* architecture [29, 34], which excels due to its self-attention mechanisms, which can capture nuanced interplay between behavior and context and address computational issues like vanishing gradients [27], making them better suited to encode long, complex sequences. For instance, Zhang et al. (2020) [39] modeled mobile phone users from their mobile app usage sequences (e.g., app installation, uninstallation) with two training objectives: Sequence Reconstruction and Masked Behavior Prediction and three evaluation tasks: Next Week's App Installation Prediction, Look-alike Audience Extension and Feed Recommendation. Chu et al. (2022) [11] modeled users of professional design software from their command sequences with contrastive learning and evaluated on predicting user responses to customer surveys.

In our paper, we also adopt the Transformer architecture. However, we base the design and evaluation of our model on our own proposed conceptualization of G-BLUM systems (§ 3), resulting in different training objectives and evaluation choices. We do not aim to compare ours with previous G-BLUM systems.

3 CONCEPTUALIZING G-BLUM SYSTEMS

Previous work [e.g., 11, 39] considers G-BLUM systems as BLUM systems that learn user representations generalizable to distinct downstream tasks. To guide our model design and evaluation, we break down this definition into five criteria for G-BLUM systems:

(1) The user model is solely trained on user behavioral data.
(2) The user model's training objective is not tied directly to specific downstream tasks.
(3) The user representations encode behavioral information.
(4) The user representations capture user-specific information that can distinguish one user from another.
(5) The user model can perform distinct downstream tasks.

Criterion 1 concerns that fact that user logs typically contain noisy events unrelated to user actions (e.g., app notifications, error reports), which should be left out. Criterion 2, 3 and 4 concern the choice of model training objectives. *Masked Behavior Prediction* and *User Contrastive Learning* are two suitable ones. The former involves randomly masking parts of users' behavior sequences, compelling the model to predict the masked behaviors based on their context and thus learns user behavioral information (fulfilling Criterion 3). The latter uses a contrastive loss function [15] to maximize the distance between representations of different users and minimize the distance between representations of the same user based on behavioral sequences from different time points. Hence, the model learns user-specific information that distinguishes one user from another (fulfilling Criterion 4). Since these two objectives are not tied to any downstream goal, Criterion 2 is also fulfilled. Criterion 5 concerns model evaluation, emphasizing that the learned user representations should generalize to different downstream tasks. To this end, we introduce three novel, distinct downstream tasks: Reported Account Prediction, Ad View Time Prediction, and Account Self-deletion Prediction. They concern predicting accounts/users who get reported by other users (e.g., for displaying malicious behaviors), who engage with an ad above a certain duration threshold, and who voluntarily delete their own accounts.

4 EXPERIMENTS

4.1 Snapchat and Data

Snapchat is a multimedia IM platform with millions of daily active users and shares similar features (e.g., chats, camera, stories, videos) with other IM platforms [32], thus making it a suitable case for exploring the applicability of G-BLUM systems to IM apps. We identify 577 unique user behaviors from raw event logs (e.g., sending a chat, viewing a video, swiping up, and using a filter).

Training Dataset. Our training data is a random sample of 766,170 U.S. users *active* between April 1, 2023 and April 14, 2023. "Active" is defined as using Snapchat for cumulatively more than 1 minute during this period. The median length of user sequences is 466.

Evaluation Datasets. We construct datasets for 6 evaluation tasks. See § 4.4 for task and dataset descriptions. Note that users and dates do not overlap between training and evaluation datasets.

4.2 Transformer-based User Modeling

Our user model consists of 12 Transformer Encoder layers with a hidden size of 768 and 12 attention heads per layer. It is trained with a batch size of 512 and a learning rate of 4e−4 for 20 epochs. 95% user sequences are used for training while the remaining 5% is for validation. Training a model takes about 8 hours on 8 Nvidia V100 GPUs. For training, we truncate each behavioral sequence to maximum 128 tokens, as our preliminary analyses show that longer sequences increase computational overhead while providing minimal additional benefit for evaluation performance.

Training Objectives. During training, the model extracts a random pair of non-overlapping sequences from each user, performs Masked Behavior Prediction on each sequence independently (with 15% random masking and cross-entropy loss $\mathcal{L}_{\text{MBP}}$), and performs User Contrastive Learning within each mini-batch of behavior sequences. Specifically, let $\mathcal{B} = \{(s_i, s_{i^+})\}_{i=0}^{n}$ define a mini-batch of n behavior sequences, where (s_i, s_{i^+}) are the sequence pairs extracted from the same user's behavioral logs. The contrastive loss on each sequence pair (s_i, s_{i^+}) is $\ell^{i,i^+} = -\log \frac{\exp(\text{sim}(e_i, e_{i^+})/\tau)}{\sum_{j \neq i} \exp(\alpha_{ij} \cdot \text{sim}(e_i, e_j)/\tau)}$, where e_i and e_{i^+} are the vectors of sequences s_i and s_{i^+} that belong to user i, e_j is the vector of sequence s_j from user j, and τ is the temperature parameter. We obtain each user vector e by aggregating all the behavior/token vectors in the last layer of the user model. The User Contrastive Learning loss over the entire batch is calculated as $\mathcal{L}_{\text{UCL}} = \frac{1}{n} \sum_i \ell^{i,i^+}$. The total combined loss is $\mathcal{L} = \mathcal{L}_{\text{MBP}} + \mathcal{L}_{\text{UCL}}$.

ALiBi. ALiBi adds a linear bias (a function of the distance between query and key tokens) to the attention scores in a Transformer model [28]. Specifically, let $\mathbf{q}_i$ be the *i-th* query in the input sequence of length L, and $\mathbf{K}$ be the key matrix. The attention score of query i is then calculated as: $\texttt{softmax}(\mathbf{q}_i\mathbf{K} + m * [-(i-1), ..., -2, -1, 0, -1, -2, ..., -(L-1-i)])$, where m is a fixed head-specific constant. Following [28, 41], we set the values of different m as a geometric sequence (*i.e.,* $\frac{1}{2^1}, \frac{1}{2^2}, ..., \frac{1}{2^n}$). Thus, attention between distant tokens gets penalized, preventing the model from attending to tokens outside of its context window and this helping to generalize to sequences longer than seen previously.

Ablation Analysis. We also perform an ablation study on the impact of the training objectives and ALiBi on our user model.

4.3 Baselines and User Representation Extraction Methods

We compare our model with several baseline models, including:

- Term Frequency (**TF**) and TF-Inverse Document Frequency (**TF-IDF**).
- Skip-Gram with Negative Sampling (**SGNS**) [25], similar to the word2vec approach in Yang et al. (2017) [37].
- **Untrained** user representations, where a fixed vector is randomly generated for each unique user behavior. This "high-dimensional" TF approach has shown competitive performances in various downstream applications [3].
- Transformer Encoder (**Enc**) and Decoder (**Dec**) are our implementation of `BERT-base` [12] and `GPT2-117M` [30] for user modeling. We train them respectively with Masked Behavior Prediction and Next Behavior Prediction objective.

User Representation Extraction. TF and TF-IDF are fixed-length vectors that summarize user behavior on the sequence level and are thus directly used as user representations. In contrast, our user model and the other baseline models learn vector representations on the token/behavior level. Therefore, pooling the behavior-level vectors to a sequence-level vector is necessary. We experiment with four pooling strategies, including mean pooling, max pooling, weighted mean pooling, and weighted max pooling. We then select the best performing pooling strategy for each of the user models.

4.4 Model Evaluation

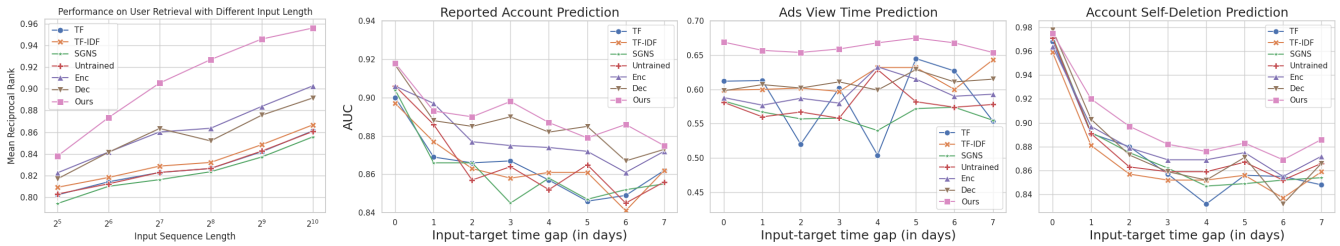
Following Criterion 3 and 4, we expect user representations from our model to encode both behavioral and user-specific information. For the former, we check whether the model can accurately predict a masked behavior given its context (i.e., *Masked Behavior Prediction*). For the latter, we check whether the model can 1) generate similar representations for the same user based on the user's behavioral sequences from non-overlapping periods and dissimilar representations for different users (i.e., *User Representation Similarity Analysis*), and 2) among behavioral sequences of different users, retrieve the two that belong to the same user (i.e., *User Retrieval*).

Masked Behavior Prediction. We use a random sample of 23,596 users, randomly mask behaviors in the user sequences and check how accurately the model can predict those masked behaviors. To account for varying frequencies of behaviors, we also used a stratified masking procedure, masking each behavior equally frequently.

User Representation Similarity Analysis. We use a random sample of 2,001 users and check whether randomly chosen sequences from the same user would be more similar (based on cosine similarity) than those from different users. Larger differences are preferred.

User Retrieval. Using the same previous dataset, we assess, given a random user's sequence and a sample of sequences from both that user and other users, whether the same user's sequence can be retrieved. Specifically, each sample consists of 101 user sequences: 1 *query* and 100 *candidates*. Within the 100 candidates, there is 1 *positive candidate* that belongs to the query user and 99 *negative candidates* that belong to other users. For each sample, we obtain a vector representation for each behavior sequence, and rank the 100 candidates based on their cosine similarity with the query. We measure the models' performance with Mean Reciprocal Rank (MRR). Additionally, we vary the input sequence length at inference time. For baseline models with an input length limitation, we split each behavior sequence into non-overlapping 128-length segments, obtain a vector representation for each segment, and take the average of these representations as the final user representation.

Reported Account Prediction, Ad View Time Prediction, Account Self-deletion Prediction. We formulate these downstream tasks as binary classification problems. We extract user representations using the latest 128 behaviors, and fit a multi-layer perception (MLP) classifier using random search with 5-fold cross validation



(a) (b)

Figure 1: Model Performance on (a) User Retrieval across Input Sequences Lengths and (b) Three Downstream Tasks across time gaps.

for hyperparameter selection. We use AUC as the evaluation metric. In addition, we vary the gap between the user behavior sequences and the labels from 0 to 7 days. For these three tasks, we collected balanced datasets of sample sizes between 10,000 and 13,000.

4.5 Results

4.5.1 Masked Behavior Prediction. Table 1 summarizes the performance of our user model (*Ours*) and the baseline model (*Enc*) on Masked Behavior Prediction across the two masking strategies (i.e., *Random* and *Stratified*). We exclude the other baselines, as they do not have a Masked Behavior Prediction training objective. Our model achieves substantially higher prediction accuracy than *Majority Vote* for both masking strategies. Remarkably, even with stratified masking, our model's performance goes down by only about 5 percentage points. While Enc has the best performance for both masking strategies, this is to be expected, as Enc is optimized for this task. That our model performs on par with Enc demonstrates its ability to encode user behaviors.

Table 1: Accuracy on Masked Behavior Prediction.

	Random	Stratified
Majority Vote	0.2450	0.063
Enc	0.9379	0.8908
Ours	0.9221	0.8693

Table 2: Within- vs Between-user Cosine Similarity.

Model	Within	Between	Difference ↑
TF	0.7607	0.5610	0.1997
TF-IDF	0.6572	0.3811	0.2762
Untrained	0.7563	0.5540	0.2023
SGNS	0.8880	0.7908	0.0972
Enc	0.8046	0.6350	0.1696
Dec	0.6761	0.4589	0.2172
Ours	0.5960	0.1910	**0.4050**

4.5.2 User Representation Similarity Analysis. Table 2 shows the results from User Representation Similarity Analysis. The *Within* column is the cosine similarity between representations of the same user, while *Between* column is the cosine similarity between representations of different users. Because both within-user and between-user cosine similarity scores can be high due to overlapping user behaviors, we focus on the difference between *within-user* and *between-user* cosine similarity (i.e., the *Difference* column), which better reflects the ability of user representations to encode user-specific information and make one user's representation distinguishable from another's. Remarkably, our user model outperforms the best baseline approach (TF-IDF) by about 13 percentage points. BERT (Enc) and GPT2 (Dec), however, have led to cosine differences even lower than simple approaches like TF-IDF and Untrained; thus, naively applying BERT/GPT2 to user modeling should be avoided.

4.5.3 User Retrieval. Figure 1a summarizes results on the User Retrieval task, varying the length of input user sequences from 32 to 1024 by a progression ratio of 2. Our model consistently and substantially outperforms all baselines, with performance steadily improving as the input sequence length increases. This shows that our model benefits from longer user sequences and, thanks to ALiBi, can effectively handle longer sequences than seen during training.

4.5.4 Downstream Task Performance. Figure 1b shows the results on the three downstream tasks across different *time gaps*. We highlight two observations. *First*, our model consistently outperforms all baselines (except for time gap 1 and 5 in Reported Account Prediction). *Second*, our model can detect malicious accounts (reported by other users) and predict user account self-deletion with high AUC scores up to one week in advance. However, our model predicts ad view time less well (still better than the baselines and chance). This is expected, as ad view time also depends on other important factors like ad length, content and users' cognitive states.

4.5.5 Ablation Analysis. Table 3 shows the impact of User Contrastive Learning (UCL) and ALiBi on our model's performance. Without UCL, our model improves slightly on Masked Behavior Prediction, but significantly underperforms other tasks. Also, without ALiBi, performance suffered across all evaluation tasks.

Table 3: Ablation Study on Masked Behavior Prediction (MBP) objective, User Contrastive Learning (UCL) objective, and ALiBi, evaluated on 6 tasks: MBP, User Representation Similarity Analysis (URSA), User Retrieval (UR), Reported Account Prediction (RAP), Ad View Time Prediction (AVTP), and Account Self-deletion Prediction (ASP). ACC (accuracy), COS (cosine difference), MRR (Mean Reciprocal Rank), and AUC are converted to the same scale.

	MBP	**URSA**	**UR**	**RAP**	**AVTP**	**ASP**
	ACC	COS	MRR	AUC	AUC	AUC
Our Model	92.21	40.50	90.77	89.07	66.30	89.85
- MBP	N/A	-0.27	-1.28	+0.23	-0.80	-1.53
- UCL	+1.61	-28.38	-4.66	-0.34	-3.98	-0.49
- ALiBi	-0.47	-1.05	-0.43	-1.01	-3.52	-1.53
- UCL & AliBi	+1.58	-23.54	-4.53	-1.15	-6.76	-1.34

5 CONCLUSION

In this paper, we demonstrate the use of G-BLUM systems for IM apps using Snapchat data. We highlight three findings. *First*, our user model achieves high prediction performance on two important downstream domains: user safety and user churn. *Second*, sequences as short as consisting of only 128 events can already result in effective general-purpose user representations for IM platforms. G-BLUM thus has the potential to replace traditional user modeling approaches that rely on months worth of (sensitive) user data. *Third*, ALiBi can effectively overcome the challenge of making inferences for sequences longer than seen at training time.

REFERENCES

[1] Luka Abb and Jana-Rebecca Rehse. 2022. A Reference Data Model for Process-Related User Interaction Logs. In *Business Process Management (Lecture Notes in Computer Science)*, Claudio Di Ciccio, Remco Dijkman, Adela del Río Ortega, and Stefanie Rinderle-Ma (Eds.). Springer International Publishing, Cham, 57–74. https://doi.org/10.1007/978-3-031-16103-2_7

[2] Nicholas Andrews and Marcus Bishop. 2019. Learning Invariant Representations of Social Media Users. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Association for Computational Linguistics, Hong Kong, China, 1684–1695. https://doi.org/10.18653/v1/D19-1178

[3] Simran Arora, Avner May, Jian Zhang, and Christopher Ré. 2020. Contextual Embeddings: When Are They Worth It?. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Online, 2650–2663. https://doi.org/10.18653/v1/2020.acl-main.236

[4] Shannon K. T. Bailey, Bradford L. Schroeder, Daphne E. Whitmer, and Valerie K. Sims. 2016. Perceptions of Mobile Instant Messaging Apps Are Comparable to Texting for Young Adults in the United States. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 60 (2016), 1235 – 1239. https://api.semanticscholar.org/CorpusID:113826805

[5] Raiyan Abdul Baten, Yozen Liu, Heinrich Peters, Francesco Barbieri, Neil Shah, Leonardo Neves, and Maarten W. Bos. 2023. Predicting Future Location Categories of Users in a Large Social Platform. *Proceedings of the International AAAI Conference on Web and Social Media* (2023). https://api.semanticscholar.org/CorpusID:259373138

[6] Yue Cao, Xiaojiang Zhou, Jiaqi Feng, Peihao Huang, Yao Xiao, Dayao Chen, and Sheng Chen. 2022. Sampling is all you need on modeling long-term user behaviors for CTR prediction. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. 2974–2983.

[7] Charles Chen, Sungchul Kim, Hung Bui, Ryan Rossi, Eunyee Koh, Branislav Kveton, and Razvan Bunescu. 2018. Predictive Analysis by Leveraging Temporal User Behavior and User Embeddings. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. ACM, Torino Italy, 2175–2182. https://doi.org/10.1145/3269206.3272032

[8] Di Chen, Qinglin Zhang, Gangbao Chen, Chuang Fan, and Qinghong Gao. 2018. Forum User Profiling by Incorporating User Behavior and Social Network Connections. In *Cognitive Computing – ICCC 2018 (Lecture Notes in Computer Science)*, Jing Xiao, Zhi-Hong Mao, Toyotaro Suzumura, and Liang-Jie Zhang (Eds.). Springer International Publishing, Cham, 30–42. https://doi.org/10.1007/978-3-319-94307-7_3

[9] Haonan Chen, Zhicheng Dou, Yutao Zhu, Zhao Cao, Xiaohua Cheng, and Ji-Rong Wen. 2022. Enhancing user behavior sequence modeling by generative tasks for session search. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. 180–190.

[10] Farhan Asif Chowdhury, Yozen Liu, Koustuv Saha, Nicholas Vincent, Leonardo Neves, Neil Shah, and Maarten W. Bos. 2021. CEAM: The Effectiveness of Cyclic and Ephemeral Attention Models of User Behavior on Social Platforms. *Proceedings of the International AAAI Conference on Web and Social Media* 15 (May 2021), 117–128. https://doi.org/10.1609/icwsm.v15i1.18046

[11] Hang Chu, Amir Hosein Khasahmadi, Karl D. D. Willis, Fraser Anderson, Yaoli Mao, Linh Tran, Justin Matejka, and Jo Vermeulen. 2022. SimCURL: Simple Contrastive User Representation Learning from Command Sequences. In *Proceedings of 21st IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, Bahamas. https://doi.org/10.1109/ICMLA55696.2022.00186

[12] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. 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)*. Association for Computational Linguistics, Minneapolis, Minnesota, 4171–4186. https://doi.org/10.18653/v1/N19-1423

[13] Gerhard Fischer. 2001. User Modeling in Human–Computer Interaction. *User Modeling and User-Adapted Interaction* 11 (2001), 65–86. https://api.semanticscholar.org/CorpusID:1526097

[14] European Union Agency for Fundamental Rights. 2020. *Your Rights Matter: Data Protection and Privacy: Fundamental Rights Survey.* Technical Report. Luxembourg.

[15] Tianyu Gao, Xingcheng Yao, and Danqi Chen. 2021. SimCSE: Simple Contrastive Learning of Sentence Embeddings. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, 6894–6910. https://doi.org/10.18653/v1/2021.emnlp-main.552

[16] Sarah Gilbert, Jessica Vitak, and Katie Shilton. 2021. Measuring Americans' Comfort With Research Uses of Their Social Media Data. *Social Media + Society* 7, 3 (July 2021), 20563051211033824. https://doi.org/10.1177/20563051211033824 Publisher: SAGE Publications Ltd.

[17] Rebecca Grinter and Margery Eldridge. 2003. Wan2tlk? Everyday text messaging. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 441–448.

[18] Ankit Kumar Jain, Somya Ranjan Sahoo, and Jyoti Kaubiyal. 2021. Online social networks security and privacy: comprehensive review and analysis. *Complex & Intelligent Systems* 7 (2021), 2157 – 2177. https://api.semanticscholar.org/CorpusID:236389184

[19] Bernard J Jansen, Danielle L Booth, and Amanda Spink. 2008. Determining the informational, navigational, and transactional intent of Web queries. *Information Processing & Management* 44, 3 (2008), 1251–1266.

[20] Bernard J Jansen, Soon-gyo Jung, and Joni Salminen. 2022. Measuring user interactions with websites: A comparison of two industry standard analytics approaches using data of 86 websites. *Plos one* 17, 5 (2022), e0268212.

[21] Richard J Lee, Ipek N Sener, Patricia L Mokhtarian, and Susan L Handy. 2017. Relationships between the online and in-store shopping frequency of Davis, California residents. *Transportation Research Part A: Policy and Practice* 100 (2017), 40–52.

[22] Chao Li, Zhiyuan Liu, Mengmeng Wu, Yuchi Xu, Huan Zhao, Pipei Huang, Guoliang Kang, Qiwei Chen, Wei Li, and Dik Lun Lee. 2019. Multi-Interest Network with Dynamic Routing for Recommendation at Tmall. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. ACM, Beijing China, 2615–2623. https://doi.org/10.1145/3357384.3357814

[23] Yozen Liu, Xiaolin Shi, Lucas Pierce, and Xiang Ren. 2019. Characterizing and Forecasting User Engagement with In-App Action Graph: A Case Study of Snapchat. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (2019). https://api.semanticscholar.org/CorpusID:173991009

[24] Kirsten Martin and Helen Nissenbaum. 2020. What Is It About Location? *Berkeley Technology Law Journal* 35, 1 (2020). https://doi.org/10.2139/ssrn.3360409

[25] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality. In *Advances in Neural Information Processing Systems*, Vol. 26. Curran Associates, Inc. https://papers.nips.cc/paper_files/paper/2013/hash/9aa42b31882ec039965f3c4923ce901b-Abstract.html

[26] Alexander Modell, Jonathan Larson, Melissa Turcotte, and Anna Bertiger. 2021. A Graph Embedding Approach to User Behavior Anomaly Detection. In *2021 IEEE International Conference on Big Data (Big Data)*. 2650–2655. https://doi.org/10.1109/BigData52589.2021.9671423

[27] Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio. 2013. On the difficulty of training recurrent neural networks. In *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28 (ICML'13)*. JMLR.org, Atlanta, GA, USA, III–1310–III–1318.

[28] Ofir Press, Noah A Smith, and Mike Lewis. 2022. Train Short, Test Long: Attention with Linear Biases Enables Input Length Extrapolation. In *Proceedings of the Tenth International Conference on Learning Representations*. Virtual Event.

[29] Erasmo Purificato, Ludovico Boratto, and Ernesto William De Luca. 2024. User Modeling and User Profiling: A Comprehensive Survey. *arXiv preprint arXiv:2402.09660* (2024).

[30] Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language Models are Unsupervised Multitask Learners. (2019).

[31] Aaron Smith. 2018. *Public Attitudes towards Computer Algorithms*. Technical Report. Washington, D.C., USA.

[32] Statista. 2023. *Snapchat daily active users 2023.* Retrieved August 29, 2005 from https://www.statista.com/statistics/545967/snapchat-app-dau/

[33] Zhiqiang Tao, Sheng Li, Zhaowen Wang, Chen Fang, Longqi Yang, Handong Zhao, and Yun Fu. 2019. Log2Intent: Towards Interpretable User Modeling via Recurrent Semantics Memory Unit. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '19)*. Association for Computing Machinery, New York, NY, USA, 1055–1063. https://doi.org/10.1145/3292500.3330889

[34] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *Advances in Neural Information Processing Systems*, Vol. 30. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

[35] Isaac Waller and Ashton Anderson. 2019. Generalists and Specialists: Using Community Embeddings to Quantify Activity Diversity in Online Platforms. In *The World Wide Web Conference (WWW '19)*. Association for Computing Machinery, New York, NY, USA, 1954–1964. https://doi.org/10.1145/3308558.3313729

[36] Carl Yang, Xiaolin Shi, Luo Jie, and Jiawei Han. 2018. I Know You'll Be Back: Interpretable New User Clustering and Churn Prediction on a Mobile Social Application. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (2018). https://api.semanticscholar.org/CorpusID:47017747

[37] Longqi Yang, Chen Fang, Hailin Jin, Matthew D. Hoffman, and Deborah Estrin. 2017. Personalizing Software and Web Services by Integrating Unstructured Application Usage Traces. In *Proceedings of the 26th International Conference on World Wide Web Companion (WWW '17 Companion)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE,

485–493. https://doi.org/10.1145/3041021.3054183

[38] Zeping Yu, Jianxun Lian, Ahmad Mahmoody, Gongshen Liu, and Xing Xie. 2019. Adaptive User Modeling with Long and Short-Term Preferences for Personalized Recommendation. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*. International Joint Conferences on Artificial Intelligence Organization, Macao, China, 4213–4219. https://doi.org/10.24963/ijcai.2019/585

[39] Junqi Zhang, Bing Bai, Ye Lin, Jian Liang, Kun Bai, and Fei Wang. 2020. General-Purpose User Embeddings based on Mobile App Usage. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '20)*. Association for Computing Machinery, New York, NY, USA, 2831–2840. https://doi.org/10.1145/3394486.3403334

[40] Guorui Zhou, Xiaoqiang Zhu, Chenru Song, Ying Fan, Han Zhu, Xiao Ma, Yanghui Yan, Junqi Jin, Han Li, and Kun Gai. 2018. Deep Interest Network for Click-Through Rate Prediction. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '18)*. Association for Computing Machinery, New York, NY, USA, 1059–1068. https://doi.org/10.1145/3219819.3219823

[41] Zhihan Zhou, Yanrong Ji, Weijian Li, Pratik Dutta, Ramana Davuluri, and Han Liu. 2023. DNABERT-2: Efficient Foundation Model and Benchmark For Multi-Species Genome. arXiv:2306.15006 [q-bio.GN]