SOCIAL MEDIA USE IS PREDICTABLE FROM APP SEQUENCES: USING LSTM AND TRANSFORMER NEURAL NETWORKS TO MODEL HABITUAL BEHAVIOR

Heinrich Peters
Columbia University
hp2500@columbia.edu

Joseph B. Bayer
Ohio State University
bayer.66@osu.edu

Sandra C. Matz
Columbia University
sm4409@columbia.edu

Yikun Chi
Stanford University
yikunchi@stanford.edu

Sumer S. Vaid
Harvard University
svaid@hbs.edu

Gabriella M. Harari
Stanford University
gharari@stanford.edu

ABSTRACT

The present paper introduces a novel approach to studying social media habits through predictive modeling of sequential smartphone user behaviors. While much of the literature on media and technology habits has relied on self-report questionnaires and simple behavioral frequency measures, we examine an important yet understudied aspect of media and technology habits: their embeddedness in repetitive behavioral sequences. Leveraging Long Short-Term Memory (LSTM) and transformer neural networks, we show that (i) social media use is predictable at the within and between-person level and that (ii) there are robust individual differences in the predictability of social media use. We examine the performance of several modeling approaches, including (i) global models trained on the pooled data from all participants, (ii) idiographic person-specific models, and (iii) global models fine-tuned on person-specific data. Neither person-specific modeling nor fine-tuning on person-specific data substantially outperformed the global models, indicating that the global models were able to represent a variety of idiosyncratic behavioral patterns. Additionally, our analyses reveal that the person-level predictability of social media use is not substantially related to the frequency of smartphone use in general or the frequency of social media use, indicating that our approach captures an aspect of habits that is distinct from behavioral frequency. Implications for habit modeling and theoretical development are discussed.

Keywords Social Media, Habits, User Modeling, LSTM, Transformer, Neural Networks

1 Introduction

Mobile app use is ubiquitous in our daily lives. On average, US smartphone users spend 4.5 hours per day on their phones [1], maintaining relationships, conducting business, and seeking information or entertainment. Among the long list of technology-mediated behaviors, social media use deserves special attention, not only because social media platforms are used by billions of people every day but also because they have been shown to impact almost every aspect of our lives. For example, social media use has revolutionized social interaction [2–5], changed the consumption of information and news [6–9], and been linked to a wide range of indicators of well-being [10–17].

Here, we investigate the predictability of social media use through the theoretical lens of media and technology habits. Habits are commonly defined as repeated behaviors that are automatically initiated by contextual cues, such as locations, situational elements, and preceding actions or activities [18–24]. Thus, habitual behaviors are performed in a semi-conscious manner in which individuals follow a rehearsed script once triggered by the contextual cue [25].

Past research indicates that social media use is, in large part, habitual behavior [11, 18]. For instance, one might habitually check one's phone while waiting for a bus on the way to work, respond to messages on Facebook before going to bed, or check Instagram by default when unlocking one's phone. Consequently, sequences of people's interactions with their devices tend to exhibit regularities over time. For example, social interaction and content consumption [26], instant messaging [27], and in-app action sequences [28], follow predictable patterns that have been characterized as habitual in nature. Moreover, there is a growing body of predictive work in human-computer interaction and user modeling, suggesting that smartphone user behaviors, including app engagement, are predictable to varying degrees from factors like device states, temporal context, location, and preceding actions [29–34].

While these findings demonstrate regularities in smartphone user behaviors that can be interpreted as habitual, the behavioral science literature has taken a different approach, often relying on relatively simple behavioral measures (e.g., behavioral frequency, reward devaluation [35, 36]) or self-report measures to assess habits (e.g., Self-Report Habit Index [37], Self-Report Behavioral Automaticity Index [38], Unified Theory of Acceptance and Use of Technology questionnaire [39], Response Frequency Measure of Media Habit [40]). These methods have proven useful in past research, but they are also limited in at least two important ways. First, people tend to be inaccurate in reporting past behavior, especially social media use [41]. Notably, such recall biases might be especially pronounced in the case of mobile and social media habits, as they often occur automatically and without conscious engagement in the midst of daily life [25, 42, 43]. Second, questionnaire-based methods are ill-equipped to detect more granular behavioral patterns, such as sequences of device interactions. Significantly, the same limitation applies to simple behavioral measures, which ignore the embeddedness of behaviors in preceding and subsequent activities along with surrounding contextual factors. Integrating more objective, granular, and context-sensitive methods grounded in digital trace data could provide a more holistic and accurate representation of media habits.

Here, we shift focus from self-report measures to tracking behavioral sequences via smartphone logs. Specifically, we build on recent research positing that repetitive behavioral sequences constitute an important, yet understudied, aspect of media and technology habits [44]. Under this scope, we can capture media habits as the extent to which a focal behavior (e.g. the use of a particular social media app) can be predicted from the history of preceding behaviors (e.g., the use of other applications). Hence, recurring sequences can indicate the extent to which an app (or combination of apps) acts as a contextual trigger for other app selections, aligning with past habit research on the role of preceding actions as contextual cues [see][][45]. This operationalization of media habits offers several important advantages compared to traditional approaches to measuring habit processes. First, we are able to directly extract habits from objectively trackable smartphone user behaviors in the real world, thereby circumventing the previously outlined response biases and artificial scenarios. Second, the reliance on sequences of behavior allows us to paint a more nuanced picture of habitual behaviors than merely focusing on behavioral frequencies.

The modeling of person-specific behavioral processes has a long tradition in the behavioral sciences [e.g., 46–51], but much of this work has relied on autoregressive models [47] or classic machine learning techniques [52]. While these approaches can offer insights into the structure of behavioral sequences, the increasing adoption of deep learning approaches that are specifically designed to represent complex time series, such as Long-Short-Term-Memory (LSTM) [53, 54] or transformer neural networks [55], opens the door for more sophisticated modeling approaches [51].

Given the fact that habits are - by definition - repetitive, the present paper examines the predictability of social media habits using LSTM and transformer neural network models to represent recurrent patterns in time series. First, we examine the extent to which social media use is predictable from intensive longitudinal app log data at the between and within-person levels and the extent to which the predictability of social media use varies across individuals. Second, we explore how the predictability of social media use relates to behavioral frequency measures and other properties of the underlying training data. Third, we examine the size of the relevant behavioral context window by analyzing the relationship between predictive performance and the length of the behavioral sequences available to the models. Our analyses are based on the predictive performance of (i) global models trained on pooled data from all users, (ii) person-specific models trained from scratch on person-level data, and (iii) global models fine-tuned on person-level data. The present study thus employs state-of-the-art machine learning approaches to explore a new operationalization of habits that is rooted in the predictability of objectively trackable behavior.

2 Method

2.1 Participants and Procedure

We collected data from a sample of N=182 adults (final N=99 after data cleaning; see below) between mid-January and early March 2021. Participants were recruited through Prolific, and all were located in the US. All participants were Android phone users. The current study focuses on Android app logs (time-stamped records of interactions with various smartphone apps), collected through the Screenomics app [56]. The app logs were recorded through event-based sampling each time an app was used. Each app was identified by its associated Android package name, a unique identifier used by the Android operating system and Google Play Store. All participants gave informed consent. The study received approval under the Stanford University IRB (Protocol #48234).

2.2 Data Preprocessing

In order to standardize the observation period per user, we discarded users with less than 14 full days of participation, and we truncated the participation period to 14 days for those who had participated for more than 14 days. From the app log sequences, we discarded instances of the Screenomics app, which was used for data collection (0.56% of observations), as well as empty events for which no app was logged (1.65% of observations). We also discarded logs associated with navigational (e.g., Launcher) and system UI processes (37.06% of observations), as these events represent engagement with the Android operating system rather than direct usage of apps. Finally, in order to guarantee sufficient training data at the person level, we removed users with less than 1,000 logged app sessions over the course of the observation period and users with minimal social media use (<5% of observations). Through this process, we reduced the participant sample size from N=182 to N=99, and the number of observations from 585,576 to 282,639.

The target variable, social media use, was derived directly from the app logs using a list of social media apps (e.g., Facebook, Instagram, TikTok, Snapchat; the full list can be found in SI A), based on the conceptualization of social media proposed by Bayer et al. [57]. Stand-alone messaging apps like WhatsApp or Telegram were not considered social media apps for the purpose of the study.

2.3 Modeling

We utilized two neural network architectures that are specifically designed to capture patterns in sequential data: LSTM [53, 54] and transformer neural networks [55]. LSTM neural networks are a type of recurrent neural network designed to represent long-term dependencies in sequential data, making them particularly effective for tasks like language processing or time series analysis [58–60]. Transformer neural networks are a type of architecture that relies on self-attention mechanisms [55], enabling the processing of entire sequences of data simultaneously, making them highly efficient and effective for tasks like natural language understanding and generation [61, 62]. In addition, they also excel at learning from other types of sequential data, such as health records [63] or life events [64].

In order to standardize the prediction task across individuals, we framed it as a binary classification task, predicting whether the next app opened by the user would be a social media app or not. For the main analyses, we constructed a sequence of the 20 preceding actions for each logged app use event, irrespective of the

temporal distance between events or whether they were separated by a period of inactivity. This was done separately within each of the training, validation, and testing sets in order to prevent information leakage across the splits. For the small subset of actions that did not have a sufficient number of preceding data points, we pre-padded the sequences with zeros. The app logs were fed into an embedding layer, which – in turn – fed into one or more LSTM or transformer layers and a single fully connected layer before the output layer. The output layer was a single neuron with a sigmoid activation function.

We first trained general, global models on the combined data of all participants. Second, following an idiographic modeling approach, we trained person-specific models for each individual in the dataset. Third, we implemented a hybrid strategy by fine-tuning the global model with person-specific data for each participant. With the exception of the fine-tuning stage, we tuned the neural network architecture and regularization parameters for each model in order to avoid spurious differences in model performance as a result of a mismatch between model hyperparameter settings and idiosyncratic properties of each person's data. The tuned hyperparameters include architectural hyperparameters such as the dimensionality of the embedding layer ([5, 50]), the number ([1, 3]) and dimensionalities ([4, 64]) of the LSTM, transformer layers, and dense layers. In order to combat overfitting, we also tuned various regularization parameters such as dropout rate ([0.2, 0.5]) and recurrent dropout rate ([0.2, 0.5]), as well as L1 ([1e-5, 1e-3]) and L2 ([1e-4, 1e-2]) norms tuned separately for recurrent, transformer, and dense layers. Finally, we tuned the learning rate ([1e-4, 1e-2]) that was applied to the Adam optimizer [65]. Hyperparameter search was performed using Bayesian Optimization [66] with 20 iterations and five random starting points. The optimization target was the area under the receiver operating characteristic curve (AUC). The AUC was used as the optimization target because it is insensitive to changes in the class distribution. This characteristic makes it especially useful in imbalanced datasets where one class outnumbers the other. A detailed description of the hyperparameter spaces can be found in SI B.

A temporal train-validation-test split was performed for each individual, such that the first section of the data was allocated to the training set (50% of observations), the second section was allocated to the validation set (25% of observations), and the final section was allocated to the testing set (25% of observations). The training set was used to fit the model, the validation set was used for hyperparameter tuning, and the testing set was used to evaluate model performance. For the global model, the splits were comprised of the union of person-level splits across participants. In other words, we maintained the same splits that were used at the person level and combined all individual training/validation/test sets to create overall training/validation/test sets. Each model was trained for up to 1,000 epochs with a batch size of 1,024 and early stopping with a patience parameter value of 5 epochs. We chose the relatively low patience value to avoid overfitting. The stopping criterion was the binary cross-entropy loss computed on the validation set. In order to account for class imbalances in the testing set, we employed the area under the receiver-operating characteristic curve (AUC) as the main evaluation metric. Other commonly used binary classification metrics can be found in SI C.

All analyses were conducted using the Tensorflow Python library [67] and the Kerastuner Python library [68]. The code used to generate the results is available on this paper's OSF page (https://osf.io/rkswe/).

3 Results

3.1 Descriptive Statistics

On average, before excluding individuals from the analysis given cleaning criteria, each participant had 1980 logged app sessions (Median=1456.5, SD=1804.2). The average number of distinct apps (Android packages) used per person was 54.4 (Median=50.5, SD=24.5). The average proportion of app sessions that were social media sessions was 17.68% (Median=15.31, SD=15.29). After cleaning, the average number of app sessions was 2856 (Median=2374, SD=1875.5), the average number of distinct apps (Android packages) used per person was 64.8 (Median=59, SD=23.1), and the average proportion of social media sessions was 25.33% (Median=22.09, SD=14.43). A visual representation of the distributions can be found in Figure 1.

The most commonly used apps were Google Chrome (10.07% of observations, 9.34% after cleaning), Facebook (5.04% of observations, 6.17% after cleaning), Google Mail (4.71% of observations, 4.58% after cleaning), Saumsung's messaging app (4.64% of observations, 4.63% after cleaning), Google Quick Search Box (3.76% of observations, 3.65% after cleaning), Facebook Messenger (3.51% of observations, 3.97% after cleaning), Google's messaging app (3.47%, 2.91 after cleaning), Snapchat (3.13% of observations, 3.88% after cleaning), Instagram (3.07% of observations, 3.77% after cleaning), and Reddit (2.82% of observations, 3.33% after cleaning). A detailed list can be found in Table 1.

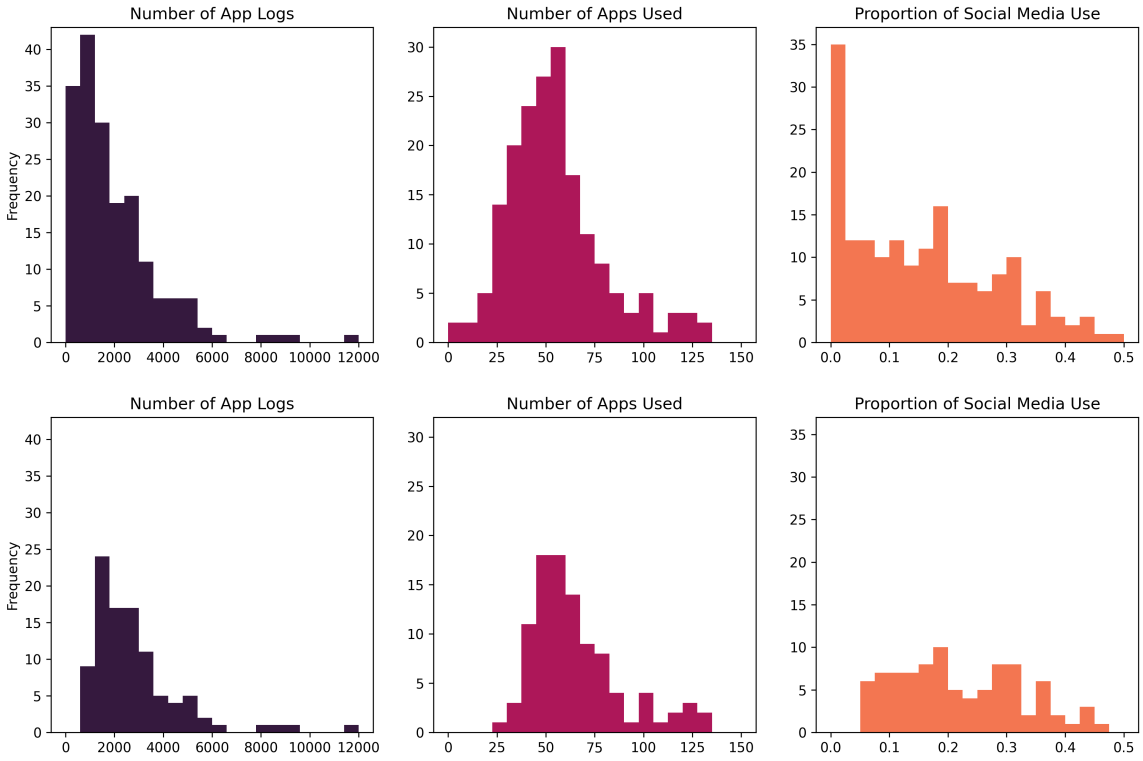

Figure 1: Frequency distributions across participants for the number of app logs, the number of distinct apps used during the observation period, and the proportion of social media sessions relative to other apps prior to excluding participants from the dataset pre (top) and post-cleaning (bottom).

3.2 Predictability of Social Media Use

Global Pre-Trained Models. We first trained a global model on the combined training sets of all individuals and performed a hyperparameter search using the combined validation sets of all individuals. Both the LSTM and the transformer model performed well on the combined test sets (AUC_{lstm}=0.782, AUC_{trans}=0.773), suggesting that social media app use is moderately to highly predictable at the between-person level (for additional evaluation metrics, please refer to SI C).

We then evaluated the global model on the individual test set of each participant to analyze the distribution of model performance scores across individuals. For this purpose, we computed separate AUC scores for each individual test set based on the predicted probability scores of the global models. We found considerable variance in the predictability of social media use across individuals for the global model. On average, the LSTM models yielded an AUC of 0.699 (SD=0.088, Min=0.515, Max=0.908), and the transformer models yielded an AUC of 0.679 (SD=0.072, Min=0.507, Max=0.853). The difference between global and person-level AUC scores indicates that the global model picks up on between-person variation in addition to within-person variation. A graphical representation of the distributions can be found in Figure 2.

To determine the sensitivity of the results to data cleaning decisions, we also evaluated an additional set of models trained on a less conservatively cleaned data set (i.e., including launcher and other system UI processes). While the predictive performance in these alternative models was generally higher than in the ones presented here (navigational events can be highly predictive of subsequent app use), the overall pattern of results was almost identical (see SI D). Additionally, we evaluated a set of models accounting for class imbalances in training by weighting the loss for each class with a value inverse to its frequency. The class weights were computed for each individual model based on the relative frequencies observed in its training set. While this approach led to slightly lower predictive performance in the case of transformer models, the overall pattern of results was almost identical to those of the main analysis as well (see SI F).

App	Proportion of Sessions Pre-Cleaning	Proportion of Users Pre-Cleaning	Proportion of Sessions Post-Cleaning	Proportion of Users Post-Cleaning
Google Chrome	10.07	92.86	9.34	95.96
Facebook	5.04	52.20	6.17	68.69
Google Mail	4.71	84.62	4.58	88.89
Samsung Messaging	4.64	39.56	4.63	44.44
Google Search	3.76	87.36	3.65	91.92
Facebook Messenger	3.51	53.30	3.97	66.67
Google Messaging	3.47	39.56	2.91	36.36
Snapchat	3.13	34.62	3.88	49.49
Instagram	3.07	47.80	3.77	70.71
Reddit	2.82	40.11	3.33	54.55
Discord	2.34	28.02	2.81	35.35
Twitter	1.89	31.87	2.32	44.44
Samsung Browser	1.82	18.13	1.80	25.25
Youtube	1.49	73.08	1.58	85.86
TikTok	1.36	25.27	1.62	34.34
Settings	1.17	95.05	0.92	96.97
Samsung Call UI	0.98	47.80	1.00	56.57
Gallery	0.96	47.25	1.03	57.58
Google Play Store	0.96	92.31	0.81	96.97
Spotify	0.69	31.87	0.83	42.42

Table 1: List of most used apps by proportion of sessions and proportion of users who used the app at least once during the observation period in percent.

Person-Specific Modeling. Next, we trained a series of models for each participant to test whether idiosyncratic patterns of social media use would be better represented in person-specific models. Hyperparameter search was performed on person-level validation data, and model performance was evaluated on person-level test data. The splits were held constant with respect to those used for the global model. Similar to the global model, the person-specific models showed varying predictive performance across individuals. On average, the LSTM models yielded an AUC of 0.609 (SD=0.092, Min=0.414, Max=0.817), and the transformer models yielded an AUC of 0.627 (SD=0.087, Min=0.447, Max=0.861). The models did not outperform the global model on average, but for a small subset of participants (LSTM: 13.13% of participants; transformer: 26.26% of participants), the individual models showed improved model performance. A graphical representation of the distributions can be found in Figure 2.

To test whether the relative strength of the global models compared to person-specific models was based on its ability to represent diverse behavioral patterns or simply because individuals exhibited very similar behavioral patterns in the first place, we examined the generalizability of person-specific models across participants. If the person-specific models were to generalize well across participants, this would indicate that their patterns of social media use were indeed similar. If the person-specific models did not generalize across individuals, we would conclude the inverse. This would suggest that the global models, too, were able to represent a multitude of idiosyncratic behavioral patterns.

We evaluated each person-specific model on the testing sets of all other individuals across the whole sample and then compared the average performance to the original results. We found that person-specific models did not generalize well across individuals with an average performance of AUC=0.526 for LSTM and AUC=0.538 for transformer models, both of which are considerably lower than their performance on within-person testing data and marginally beat the chance baseline of AUC=0.5. These results indicate that the person-specific models do not pick up on general behavioral patterns that are common across individuals. Instead, the strength of the models, including the pre-trained global models, seems to be driven by their ability to capture a multitude of individual behavioral patterns.

Fine-Tuning Global Models on Person-Specific Data. We fine-tuned the pre-trained global model to test whether personalized models could benefit from general representations of patterns of social media use and vice versa. For this purpose, we continued to train the pre-trained global models for up to 1000 additional epochs on the training data of each individual and performed early stopping using the binary cross-entropy loss on the individual's validation set as the stopping criterion. In order to avoid overriding previously learned relationships, a low learning rate of 0.0001 was applied. This procedure resulted in a fine-tuned LSTM and a fine-tuned transformer model for each individual. On average, the fine-tuned LSTM models yielded an AUC of 0.702 (SD=0.089, Min=0.522, Max=0.908), and the fine-tuned transformer models yielded an AUC of 0.683 (SD=0.077, Min=0.497, Max=0.858). While the fine-tuned models outperformed the global model in most cases (LSTM: 68.68% of participants; transformer: 63.63% of participants), the fine-tuning only resulted in

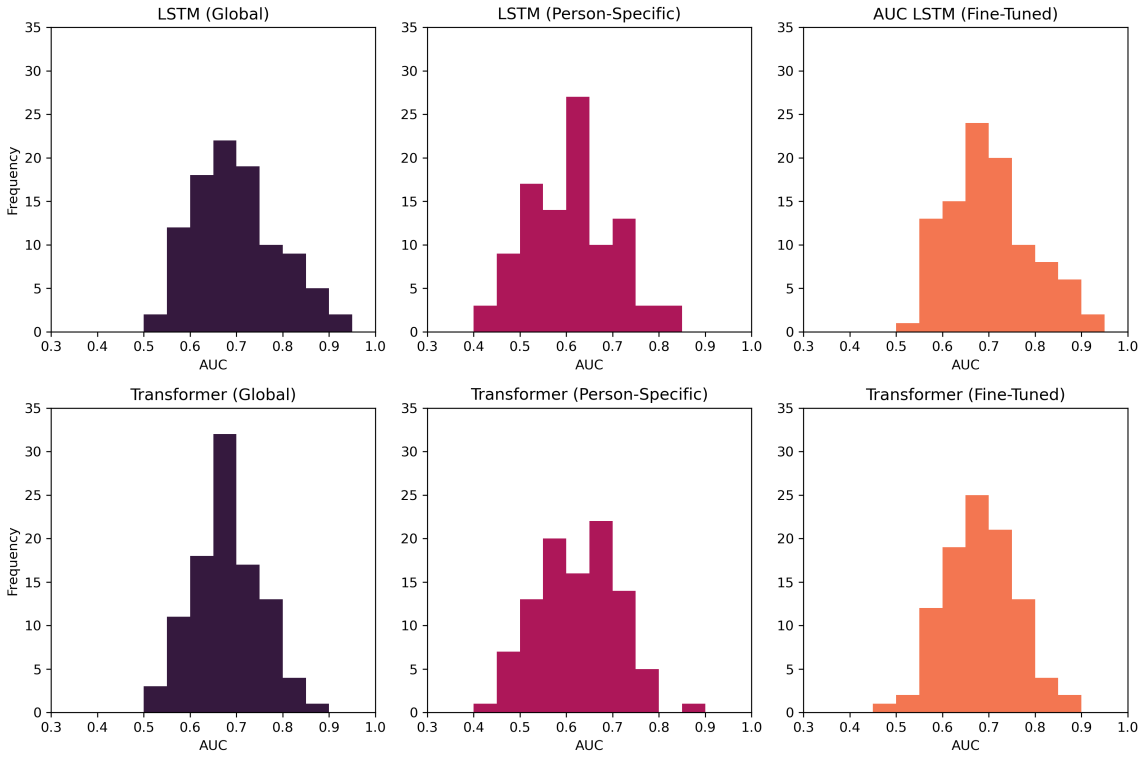

Figure 2: Distributions of model performance scores (AUC) across participants for global (left), person-specific (middle), and fine-tuned (right) models.

a very minor improvement over and above the global pre-trained model. A graphical representation of the distributions can be found in Figure 2.

In order to incorporate hyperparameter tuning into the fine-tuning approach, an alternative set of fine-tuned models was trained by freezing the model weights of the embedding layers, LSTM layers, and transformer layers while retraining the dense layers and output layers from scratch with another round of hyperparameter search (i.e., adjusting layer dimensions, regularization, and learning rate). This approach places a higher emphasis on finding appropriate hyperparameters at the cost of discarding the model weights of layers for which hyperparameters are tuned. The procedure did not result in improved model performance (see SI E). Taken together, the results show that the pre-trained global models already captured the vast majority of explainable within-person variation in social media use and that fine-tuning provided only limited value.

3.3 Relationships Between Predictability and Behavioral Frequency

To gain a better understanding of the factors driving individual differences in the predictability of social media use, we analyzed its associations with properties of the underlying training data. Specifically, we analyzed the association between predictability scores and frequency measures as indicators of habitual behavior. We found that predictability scores were weakly correlated with the overall volume of data (number of app logs) per person in the global (r_{lstm}=-0.014; r_{trans}=0.058) and fine-tuned models (r_{lstm}=-0.017; r_{trans}=0.029). As would be expected, the correlation was higher for person-specific models trained on person-level data (r_{lstm}=0.273, r_{trans}=0.322). Similarly, predictive performance scores were not related to the absolute number of social media sessions for global models (r_{lstm}=-0.049, r_{trans}=0.004) and fine-tuned models (r_{lstm}=0.035, r_{trans}=0.003) but were somewhat related for person-specific models (r_{lstm}=0.247; r_{trans}=0.258). The proportion of social media sessions relative to the total number of app sessions was weakly negatively related to model performance in global models (r_{lstm}=-0.082; r_{trans}=-0.118) and fine-tuned models (r_{lstm}=-0.052; r_{trans}=-0.092) but weakly positively related for person-specific models (r_{lstm}=0.178; r_{trans}=0.124). These findings indicate that (i) differences in the predictability of social media use are not primarily driven by the availability of training

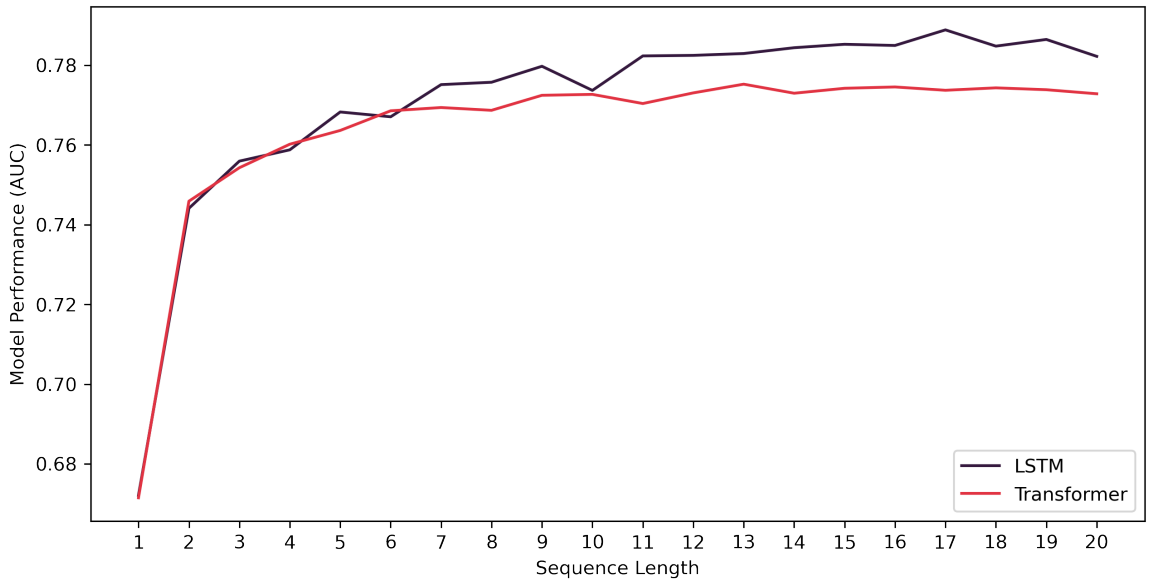

Figure 3: Global model performance as a function of sequence length. While longer sequences were generally associated with higher predictive performance, the incremental performance gains decreased with sequence length.

data, even in the case of person-specific models, and (ii) they support the proposition that the predictability of social media use is indeed distinct from behavioral frequency.

3.4 Relationships Between Predictability and Context Size

In order to examine the boundaries of the relevant behavioral context, we analyzed model performance as a function of input sequence length (i.e., the number of preceding app sessions available to the model). For this purpose, we re-trained the global model for sequence lengths between one and twenty time steps and another model with 50 preceding app sessions, accounting for the possibility that far-removed behaviors would still contain useful information. The results show that longer sequences are generally associated with higher model performance, with diminishing incremental contributions of each additional time step. The LSTM model reached its best test performance at a sequence length of 17 with an AUC of 0.789, while the transformer model reached its best test performance at a sequence length of 13 with an AUC of 0.775. Models trained on sequences of 50 time steps did not perform better (AUC_{lstm}=0.787; AUC_{trans}=0.775). The results also reveal that even extremely short sequences are predictive of subsequent social media use. For example, both models performed considerably above chance based on information about only the last preceding app session (AUC_{lstm}=0.672; AUC_{trans}=0.671). Generally, the incremental performance improvement for each additional unit in sequence length leveled off dramatically for models trained on sequences of more than three to ten app sessions. For a visual presentation of the results, please refer to Figure 3.

4 Discussion

The present paper is among the first to examine the predictability of social media use through the lens of media and technology habits. The results demonstrate that social media use is moderately to highly predictable from preceding smartphone user behaviors. This finding highlights the repetitive, predictable nature of social media habits and is aligned with past research on the predictability of more fine-grained online social behaviors such as active and passive social media use [26], instant messaging [27] or in-app action sequences [28], as well as app usage in general [29–34].

Contributing specifically to the habits literature, our study highlights the importance and predictive validity of preceding action cues [45], which have not been examined at this level of granularity in previous work. The results indicate that objectively captured preceding action cues contain considerable information about

momentary social media use and that the sequential structure of app use matters above and beyond more general behavioral tendencies that have been studied in most previous work [see 11]). Relatedly, our analysis of the relationship between sequence lengths and predictive performance indicates that the relevant context window spans sequences of between roughly three and ten apps, as predictive performance did not increase substantially for longer sequences. While this finding is highly specific to the research setting and the predictive task at hand, it provides an intuitive way to distinguish between longer-term behavioral tendencies and preceding action cues. Behavioral tendencies are captured by the consistency of behavioral frequencies over time, whereas preceding action cues can be defined by their incremental predictive contribution with respect to a focal behavior.

Our results also address the question of the generalizability of social media habits across individuals. As the person-specific models did not generalize well across individuals, our results indicate that the models pick up on idiosyncratic habits [48, 69, 70]. At the same time, the fact that person-specific models did not systematically outperform global models indicates that the global models not only represent general behavioral trends but simultaneously learn a broad range of idiosyncratic patterns, given sufficient training data. These findings raise the question as to whether individual differences in the predictability of social media use can be interpreted as an objective measure of habit strength, focusing on the predictably recurrent nature of habit. The fact that predictive performance scores were not highly correlated with frequency measures of social media use indicates that they capture a fundamentally different aspect of social media habits. While the interpretation of predictability as habit strength still needs to be substantiated by dedicated psychometric analyses, it could open up new opportunities for studying habits based on the wealth of objective behavioral data that is available through consumer electronics and online social platforms [71, 72].

More broadly, the current research represents a novel approach to the study of media and technology habits and affirms the potential of modeling sequential behavior for the future of habit research. By utilizing objective records of smartphone use, our study circumvents these limitations of self-report methods [41] while retaining the ecological validity of studying individuals' own real-world habits. Moreover, the focus on predictive models designed to represent temporal dependencies presents a new step in the identification of media and technology habits. This approach aligns with the broader shift in behavioral science research towards leveraging objectively trackable data and machine learning to uncover patterns that are not readily apparent through traditional methods [72–74].

Notably, our analytical framework can be extended to generate a wide range of features related to the predictability of behavior that are not examined in the current research. Aside from analyzing predictive performance purely based on preceding actions, it would be possible to analyze predictive performance with regard to other theoretically relevant constructs. For example, the per-person uptick in predictive performance after adding location data to the model could be interpreted as a "context-contingency score", indicating the varying degrees to which people's behaviors are determined by their environments. Similarly, the degree of generalizability of person-specific models could be interpreted as an "idiosyncrasy score", indicating the extent to which individuals exhibit unique behavioral associations. Relatedly, measures of model complexity (in the absence of regularization) in person-specific models could serve as a measure of habit complexity. Future research should investigate the viability of these suggestions and perform dedicated psychometric analyses.

Finally, our approach may not be limited to the study of social media habits but could also be applied to other behavioral phenomena and theoretical perspectives, such as personality and person-situation transactions. For example, the Cognitive-Affective Personality System (CAPS) [75, 76] defines personality not as a set of static traits but rather as a collection of individual differences in how people respond to varying situations. According to this theory, personality is understood through the distinctive patterns of thoughts, feelings, and behaviors that emerge as individuals interact with specific situational contexts. Our modeling approach would lend itself to examining this notion of personality based on extensive behavioral records. Finding a large predictive contribution of context features and relatively low generalizability across person-specific models would indicate the validity of CAPS with regard to online behaviors and would extend prior work on the behavioral consistency of smartphone use [77].

4.1 Limitations and Directions for Future Research

Our work has several limitations that need to be addressed in future research. Firstly, while the overall volume of data was sufficient to train complex machine learning models, we faced data limitations for the person-specific models. The lack of training data per person is one potential explanation for the finding that person-specific models were outperformed by global models. The relatively low correlation between person-level AUC scores and the number of data points per person, however, suggests that differences in predictability are not merely an artifact introduced by varying degrees of data availability across individuals. Nonetheless, additional research is

needed in order to fully rule out concerns related to data coverage. Mirroring the concerns about data coverage for training person-specific models, insufficient fine-tuning data could have decreased the effectiveness of the fine-tuning stage. Importantly, the limited effectiveness of person-specific modeling and fine-tuning could also be related to the decision to frame the predictive task as a binary classification task (i.e., collapsing various social media apps into the single category of "social media use"). This approach likely attenuated some of the idiosyncratic patterns that person-specific or fine-tuned models might have otherwise picked up on. Future research should therefore consider examining a more granular range of target behaviors.

Second, by condensing all social media use into a single category, our approach does not distinguish between different platforms, their unique affordances [78], or the nature of the engagement (e.g., active versus passive use) [79–81]. While this decision was necessary to standardize the target variable across models, it leaves room for further exploration, as different social media platforms have distinct features and may engender different patterns of usage. For instance, the usage patterns on a visually oriented platform like Instagram may differ from text-based platforms like X or Reddit. Additionally, active use (such as posting or commenting) likely follows different habitual patterns compared to passive use (such as browsing without interacting) – and these differences may have implications for well-being [16]. Future research should take these distinctions into account and examine how different types of social media engagement impact user behavior and experience. This could lead to a more nuanced understanding of social media habits and their effects on individuals, potentially offering more targeted insights for the development of interventions for healthier technology use.

Third, we did not analyze the relationships between person-level predictive performance and self-reported habit measures. Future research should alleviate this concern and also analyze the relationships between predictability scores and other variables known to be related to habitual social media use. However, as we maintain that the predictability of behavior captures an aspect of habits not sufficiently represented in self-report measures, we would not necessarily expect a high correlation. A better empirical test of criterion validity would be whether individual differences in the predictability of social media use show incremental explanatory power above and beyond existing self-report measures. Additionally, future work can compare how preceding sequences relate to other indirect indicators of habitual behavior (e.g., devaluation insensitivity).

Finally, we exclusively focused on preceding user behaviors as input features without incorporating other contextual features that could provide a more complete picture of social media usage patterns. In future research, including other contextual data such as time of day or location could allow for a more explicit examination of the context-contingent nature of social media habits. Similarly, examining the effects of technical cues (e.g., notifications) and data from other smartphone sensors could add nuance to our research. This approach could potentially improve the predictive performance of our models by identifying contextual conditions under which certain behaviors are more or less likely to occur.

4.2 Practical Implications

Aside from its theoretical implications, our work lays the foundations for more habit-centered approaches to user modeling. This shift could, in turn, lead to the development of new technologies and user experiences that are more aligned with users' needs and preferences. For example, the ability to predict when and what type of app a user is likely to engage with can be utilized for prefetching content [32]. This means that apps could preload content before a user is likely to engage. This not only enhances the user experience by reducing loading times but can also help to manage network traffic more effectively by spreading content download over time.

Another potential opportunity lies in adaptive user interfaces and notifications. Interfaces could adjust their layout and content to align with the user's predicted preferences and habits at different times. For example, apps that are likely to be used could be presented on the home screen for easy access. Similarly, knowing when users are most likely to engage with social media could help time notifications and alerts more effectively. As smartphone use can have negative effects on productivity [82–84], predictions of social media engagement could also be used to optimize the delivery timing of messages and avoid unwanted interruptions. A similar idea has been discussed as "bounded deferral" [85, 86], where messages are held back while a user is busy, but only up to a maximum amount of time. For users who wish to balance their social media exposure, predictive models can make interventions more personalized and effective. For example, it has been shown that introducing friction (e.g., a short wait period before opening a social media app) can be highly effective in reducing social media use [87]. Such interventions could be dynamically adjusted based on ongoing or preceding user behavior, making them more relevant and effective. For instance, users might want to intervene in habit-driven use but not goal-directed use.

Finally, a better understanding of social media habits can help policymakers develop user-centric policies to protect young users from the potential harms of excessive media consumption and promote digital well-being [88].

4.3 Conclusion

Taken together, our results indicate that social media use is predictable within and across individuals and that global models are capable of making accurate predictions, even at the within-person level. By utilizing objective behavioral time series data, employing advanced predictive models, and examining the predictability of social media use at both the within and between-person level, our work highlights a new direction in habit research. Our findings shed light on the predictive validity of preceding action cues and the value of sequential information in predicting social media habits. In addition, our findings reveal individual differences in the predictability of social media use and raise the question as to whether they can be interpreted as a novel indicator of habit strength. Finally, our study not only enhances our understanding of media consumption behaviors but also paves the way for more data-driven strategies to manage everyday media and technology habits.

Ethics approval

The study was approved by the Stanford University IRB (Protocol #48234). All methods were carried out in accordance with relevant guidelines and regulations.

Availability of data and materials

The code used to generate the results is available on this paper's OSF page (https://osf.io/rkswe/). The data needed to reproduce the analyses will be made available after peer review of the manuscript is complete.

Competing interests

The authors declare no potential conflicts of interest.

Author contributions

HP: Conceptualization, Methodology, Software, Formal analysis, Visualization, Writing - Original Draft; JBB: Conceptualization, Writing - Review & Editing; SCM: Conceptualization, Writing - Review & Editing; YC: Data Curation, Software; SSV: Writing - Review & Editing; GMH: Conceptualization, Resources, Writing - Review & Editing, Supervision, Project administration, Funding acquisition

Acknowledgments

This research was supported in part by the Stanford Institute for Human-Centered Artificial Intelligence (HAI) with a Google Cloud Credit Award and a Stanford HAI Seed Grant. The authors would like to thank Katherine Roehrick and Serena Soh for their research assistance with this study and their role in data collection.

References

[1] Statista, *U.S.: Mobile phone daily usage time 2024*, 2024. [Online]. Available: https://www.statista.com/statistics/1045353/mobile-device-daily-usage-time-in-the-us/.

[2] C. L. Coyle and H. Vaughn, "Social networking: Communication revolution or evolution?" *Bell Labs Technical Journal*, vol. 13, no. 2, pp. 13–17, 2008.

[3] T. Diehl, B. E. Weeks, and H. Gil de Zúñiga, "Political persuasion on social media: Tracing direct and indirect effects of news use and social interaction," *New Media & Society*, vol. 18, no. 9, pp. 1875–1895, 2016.

[4] C. Licoppe and Z. Smoreda, "Are social networks technologically embedded?: How networks are changing today with changes in communication technology," *Social Networks*, vol. 27, no. 4, pp. 317–335, 2005.

[5] T. Zeitzoff, "How Social Media Is Changing Conflict," *Journal of Conflict Resolution*, vol. 61, no. 9, pp. 1970–1991, 2017.

[6] H. Gil de Zúñiga, B. Weeks, and A. Ardèvol-Abreu, "Effects of the News-Finds-Me Perception in Communication: Social Media Use Implications for News Seeking and Learning About Politics," *Journal of Computer-Mediated Communication*, vol. 22, no. 3, pp. 105–123, 2017.

[7] S. K. Lee, N. J. Lindsey, and K. S. Kim, "The effects of news consumption via social media and news information overload on perceptions of journalistic norms and practices," *Computers in Human Behavior*, vol. 75, pp. 254–263, 2017.

[8] S. Vermeer, D. Trilling, S. Kruikemeier, and C. de Vreese, "Online News User Journeys: The Role of Social Media, News Websites, and Topics," *Digital Journalism*, vol. 8, no. 9, pp. 1114–1141, 2020.

[9] M. Walker and K. E. Matsa, *News Consumption Across Social Media in 2021*, 2021. [Online]. Available: https://www.pewresearch.org/journalism/2021/09/20/news-consumption-across-social-media-in-2021/.

[10] H. Allcott, L. Braghieri, S. Eichmeyer, and M. Gentzkow, "The Welfare Effects of Social Media," *American Economic Review*, vol. 110, no. 3, pp. 629–676, 2020.

[11] J. B. Bayer, I. A. Anderson, and R. S. Tokunaga, "Building and breaking social media habits," *Current Opinion in Psychology*, vol. 45, pp. 279–288, 2022.

[12] J. Brailovskaia, F. Ströse, H. Schillack, and J. Margraf, "Less Facebook use – More well-being and a healthier lifestyle? An experimental intervention study," *Computers in Human Behavior*, vol. 108, p. 106 332, 2020.

[13] J. Fox and J. J. Moreland, "The dark side of social networking sites: An exploration of the relational and psychological stressors associated with Facebook use and affordances," *Computers in Human Behavior*, vol. 45, pp. 168–176, 2015.

[14] M. G. Hunt, R. Marx, C. Lipson, and J. Young, "No More FOMO: Limiting Social Media Decreases Loneliness and Depression," *Journal of Social and Clinical Psychology*, vol. 37, no. 10, pp. 751–768, 2018.

[15] B. A. Primack *et al.*, "Social Media Use and Perceived Social Isolation Among Young Adults in the U.S.," *American Journal of Preventive Medicine*, vol. 53, no. 1, pp. 1–8, 2017.

[16] P. Verduyn *et al.*, "Passive Facebook usage undermines affective well-being: Experimental and longitudinal evidence," *Journal of Experimental Psychology: General*, vol. 144, no. 2, pp. 480–488, 2015.

[17] S. S. Vaid *et al.*, "Variation in social media sensitivity across people and contexts," *Scientific Reports*, vol. 14, no. 1, p. 6571, 2024.

[18] I. A. Anderson and W. Wood, "Habits and the electronic herd: The psychology behind social media's successes and failures," *Consumer Psychology Review*, vol. 4, no. 1, pp. 83–99, 2021.

[19] J. B. Bayer and R. LaRose, "Technology Habits: Progress, Problems, and Prospects," in *The Psychology of Habit*, B. Verplanken, Ed., Cham: Springer International Publishing, 2018, pp. 111–130.

[20] B. Gardner and A. L. Rebar, "Habit Formation and Behavior Change," in *Oxford Research Encyclopedia of Psychology*, Oxford University Press, 2019.

[21] A. Schnauber-Stockmann and T. K. Naab, "The process of forming a mobile media habit: Results of a longitudinal study in a real-world setting," *Media Psychology*, vol. 22, no. 5, pp. 714–742, 2019.

[22] R. S. Tokunaga, "Media Use as Habit," in *The International Encyclopedia of Media Psychology*, Hoboken, Nj: John Wiley & Sons, Ltd, 2020, pp. 1–5.

[23] B. Verplanken, "Beyond frequency: Habit as mental construct," *British Journal of Social Psychology*, vol. 45, no. 3, pp. 639–656, 2006.

[24] W. Wood and D. Rünger, "Psychology of Habit," *Annual Review of Psychology*, vol. 67, no. 1, pp. 289–314, 2016.

[25] J. B. Bayer, S. Dal Cin, S. W. Campbell, and E. Panek, "Consciousness and Self-Regulation in Mobile Communication," *Human Communication Research*, vol. 42, no. 1, pp. 71–97, 2016.

[26] H. Peters, Y. Liu, F. Barbieri, R. A. Baten, S. C. Matz, and M. W. Bos, "Context-Aware Prediction of User Engagement on Online Social Platforms," *arXiv:2310.14533 [cs]*, 2023.

[27] H. Peters, R. Dotsch, M. Triguero Roura, Y. Liu, S. C. Matz, and M. W. Bos, "Predicting Attentiveness and Responsiveness in Instant Messaging - Evidence from a Large Online Social Platform," in *6th-annual Psychology of Technology Conference*, Philadelphia, PA, 2022.

[28] Y. Liu, X. Shi, L. Pierce, and X. Ren, "Characterizing and Forecasting User Engagement with In-App Action Graph: A Case Study of Snapchat," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, New York, USA: ACM, 2019, pp. 2023–2031.

[29] R. Baeza-Yates, D. Jiang, F. Silvestri, and B. Harrison, "Predicting The Next App That You Are Going To Use," in *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, New York, NY, USA: ACM, 2015, pp. 285–294.

[30] K. Huang, C. Zhang, X. Ma, and G. Chen, "Predicting mobile application usage using contextual information," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, New York, NY, USA: ACM, 2012, pp. 1059–1065.

[31] N. Natarajan, D. Shin, and I. S. Dhillon, "Which app will you use next? collaborative filtering with interactional context," in *Proceedings of the 7th ACM conference on Recommender systems*, New York, NY, USA: ACM, 2013, pp. 201–208.

[32] A. Parate, M. Böhmer, D. Chu, D. Ganesan, and B. M. Marlin, "Practical prediction and prefetch for faster access to applications on mobile phones," in *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, New York, NY, USA: ACM, 2013, pp. 275–284.

[33] T. Xia *et al.*, "DeepApp: Predicting Personalized Smartphone App Usage via Context-Aware Multi-Task Learning," *ACM Transactions on Intelligent Systems and Technology*, vol. 11, no. 6, pp. 1–12, 2020.

[34] S. Xu, W. Li, X. Zhang, S. Gao, T. Zhan, and S. Lu, "Predicting and Recommending the next Smartphone Apps based on Recurrent Neural Network," *CCF Transactions on Pervasive Computing and Interaction*, vol. 2, no. 4, pp. 314–328, 2020.

[35] S. Nebe, A. Kretzschmar, M. C. Brandt, and P. N. Tobler, "Characterizing Human Habits in the Lab," *Collabra: Psychology*, vol. 10, no. 1, p. 92 949, 2024.

[36] A. I. Mendelsohn, "Creatures of Habit: The Neuroscience of Habit and Purposeful Behavior," *Biological psychiatry*, vol. 85, no. 11, e49–e51, 2019.

[37] B. Verplanken and S. Orbell, "Reflections on Past Behavior: A Self-Report Index of Habit Strength," *Journal of Applied Social Psychology*, vol. 33, no. 6, pp. 1313–1330, 2003.

[38] B. Gardner, C. Abraham, P. Lally, and G.-J. de Bruijn, "Towards parsimony in habit measurement: Testing the convergent and predictive validity of an automaticity subscale of the Self-Report Habit Index," *International Journal of Behavioral Nutrition and Physical Activity*, vol. 9, no. 1, p. 102, 2012.

[39] V. Venkatesh, J. Y. L. Thong, and X. Xu, "Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology," *MIS Quarterly*, vol. 36, no. 1, pp. 157–178, 2012.

[40] T. K. Naab and A. Schnauber, "Habitual Initiation of Media Use and a Response-Frequency Measure for Its Examination," *Media Psychology*, vol. 19, no. 1, pp. 126–155, 2016.

[41] S. C. Boyle, S. Baez, B. M. Trager, and J. W. LaBrie, "Systematic Bias in Self-Reported Social Media Use in the Age of Platform Swinging: Implications for Studying Social Media Use in Relation to Adolescent Health Behavior," *International Journal of Environmental Research and Public Health*, vol. 19, no. 16, p. 9847, 2022.

[42] J. B. Bayer, S. W. Campbell, and R. Ling, "Connection Cues: Activating the Norms and Habits of Social Connectedness," *Communication Theory*, vol. 26, no. 2, pp. 128–149, 2016.

[43] J. B. Bayer and S. W. Campbell, "Texting while driving on automatic: Considering the frequency-independent side of habit," *Computers in Human Behavior*, vol. 28, no. 6, pp. 2083–2090, 2012.

[44] A. Roffarello and L. Russis, "Understanding and Streamlining App Switching Experiences in Mobile Interaction," *International Journal of Human-Computer Studies*, vol. 158, no. 2, p. 102 735, 2021.

[45] W. Wood, "Habit in Personality and Social Psychology," *Personality and Social Psychology Review*, vol. 21, no. 4, pp. 389–403, 2017.

[46] E. L. Hamaker and C. V. Dolan, "Idiographic data analysis: Quantitative methods—from simple to advanced," in *Dynamic process methodology in the social and developmental sciences*, J. Valsiner, P. C. M. Moleenar, M. C. D. P. Lyra, and N. Chaudhary, Eds., New York, NY, US: Springer Science + Business Media, 2009, pp. 191–216.

[47] J. M. B. Haslbeck and O. Ryan, "Recovering Within-Person Dynamics from Psychological Time Series," *Multivariate Behavioral Research*, vol. 57, no. 5, pp. 735–766, 2022.

[48] P. Lally, C. H. M. Van Jaarsveld, H. W. W. Potts, and J. Wardle, "How are habits formed: Modelling habit formation in the real world," *European Journal of Social Psychology*, vol. 40, no. 6, pp. 998–1009, 2010.

[49] P. Molenaar, "A Manifesto on Psychology as Idiographic Science: Bringing the Person Back Into Scientific Psychology, This Time Forever," *Measurement: Interdisciplinary Research & Perspective*, vol. 2, pp. 201–218, 2004.

[50] P. C. Molenaar and C. G. Campbell, "The New Person-Specific Paradigm in Psychology," *Current Directions in Psychological Science*, vol. 18, no. 2, pp. 112–117, 2009.

[51] N. Ram, N. Haber, T. N. Robinson, and B. Reeves, "Binding the Person-Specific Approach to Modern AI in the Human Screenome Project: Moving past Generalizability to Transferability," *Multivariate Behavioral Research*, pp. 1–9, 2023.

[52] E. D. Beck and J. J. Jackson, "Personalized Prediction of Behaviors and Experiences: An Idiographic Person–Situation Test," *Psychological Science*, vol. 33, no. 10, pp. 1767–1782, 2022.

[53] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[54] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.

[55] A. Vaswani *et al.*, "Attention is all you need," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Red Hook, NY: Curran Associates, 2017, pp. 6000–6010.

[56] B. Reeves *et al.*, "Screenomics: A Framework to Capture and Analyze Personal Life Experiences and the Ways that Technology Shapes Them," *Human–Computer Interaction*, vol. 36, no. 2, pp. 150–201, 2021.

[57] J. B. Bayer, P. Trieu, and N. B. Ellison, "Social Media Elements, Ecologies, and Effects," *Annual Review of Psychology*, vol. 71, no. 1, pp. 471–497, 2020.

[58] A. Graves, N. Jaitly, and A.-r. Mohamed, "Hybrid speech recognition with Deep Bidirectional LSTM," in *2013 IEEE Workshop on Automatic Speech Recognition and Understanding*, Olomouc, Czech Republic, 2013, pp. 273–278.

[59] S. Han *et al.*, "ESE: Efficient Speech Recognition Engine with Sparse LSTM on FPGA," in *Proceedings of the 2017 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays*, New York, NY, USA: ACM, 2017, pp. 75–84.

[60] S. Siami-Namini, N. Tavakoli, and A. S. Namin, "The Performance of LSTM and BiLSTM in Forecasting Time Series," in *2019 IEEE International Conference on Big Data (Big Data)*, Los Angeles, CA, 2019, pp. 3285–3292.

[61] J. Devlin, M.-W. Chang, K. Lee, and K. 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)*, Minneapolis, Minnesota: Association for Computational Linguistics, 2019, pp. 4171–4186.

[62] OpenAI, *GPT-4 Technical Report*, 2023. [Online]. Available: `https://cdn.openai.com/papers/gpt-4.pdf`.

[63] Y. Li *et al.*, "BEHRT: Transformer for Electronic Health Records," *Scientific Reports*, vol. 10, no. 1, p. 7155, 2020.

[64] G. Savcisens *et al.*, "Using sequences of life-events to predict human lives," *Nature Computational Science*, vol. 4, no. 1, pp. 43–56, 2024.

[65] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv:1412.6980 [cs]*, 2017.

[66] S. Falkner, A. Klein, and F. Hutter, "BOHB: Robust and Efficient Hyperparameter Optimization at Scale," in *Proceedings of the 35th International Conference on Machine Learning*, Stockholm, Sweden: PMLR, 2018, pp. 1437–1446.

[67] M. Abadi *et al.*, "TensorFlow: A system for large-scale machine learning," in *Proceedings of the 12th USENIX conference on Operating Systems Design and Implementation*, Savannah, GA: USENIX Association, 2016, pp. 265–283.

[68] T. O'Malley, E. Bursztein, J. Long, C. Francois, H. Jin, and L. Invernizzi, *Keras Tuner*, 2019. [Online]. Available: `https://github.com/keras-team/keras-tuner`.

[69] B. Gardner, "A review and analysis of the use of 'habit' in understanding, predicting and influencing health-related behaviour," *Health Psychology Review*, vol. 9, no. 3, pp. 277–295, 2015.

[70] B. Gardner and P. Lally, "Modelling Habit Formation and Its Determinants," in *The Psychology of Habit: Theory, Mechanisms, Change, and Contexts*, B. Verplanken, Ed., Cham: Springer International Publishing, 2018, pp. 207–229.

[71] G. M. Harari, N. D. Lane, R. Wang, B. S. Crosier, A. T. Campbell, and S. D. Gosling, "Using Smartphones to Collect Behavioral Data in Psychological Science: Opportunities, Practical Considerations, and Challenges," *Perspectives on Psychological Science*, vol. 11, no. 6, pp. 838–854, 2016.

[72] H. Peters, Z. Marrero, and S. D. Gosling, "The Big Data toolkit for psychologists: Data sources and methodologies," in *The psychology of technology: Social science research in the age of Big Data*, Washington, DC, US: American Psychological Association, 2022, pp. 87–124.

[73] D. M. J. Lazer *et al.*, "Computational social science: Obstacles and opportunities," *Science*, vol. 369, no. 6507, pp. 1060–1062, 2020.

[74] C. Stachl *et al.*, "Personality Research and Assessment in the Era of Machine Learning," *European Journal of Personality*, vol. 34, no. 5, pp. 613–631, 2020.

[75] W. Mischel, "Toward a cognitive social learning reconceptualization of personality.," *Psychological Review*, vol. 80, no. 4, p. 252, 1974.

[76] W. Mischel and Y. Shoda, "A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure," *Psychological Review*, vol. 102, no. 2, pp. 246–268, 1995.

[77] H. Shaw, P. J. Taylor, D. A. Ellis, and S. M. Conchie, "Behavioral Consistency in the Digital Age," *Psychological Science*, vol. 33, no. 3, pp. 364–370, 2022.

[78] T. Bucher and A. Helmond, "The Affordances of Social Media Platforms," in *The SAGE Handbook of Social Media*, J. Bruges, A. Marwick, and T. Poell, Eds., Thousand Oaks, CA: Sage Publications, 2018, pp. 233–253.

[79] C. G. Escobar-Viera *et al.*, "Passive and Active Social Media Use and Depressive Symptoms Among United States Adults," *Cyberpsychology, Behavior, and Social Networking*, vol. 21, no. 7, pp. 437–443, 2018.

[80] K. Hemmings-Jarrett, J. Jarrett, and M. B. Blake, "Evaluation of User Engagement on Social Media to Leverage Active and Passive Communication," in *2017 IEEE International Conference on Cognitive Computing (ICCC)*, Honolulu, HI, 2017, pp. 132–135.

[81] M. L. Khan, "Social media engagement: What motivates user participation and consumption on YouTube?" *Computers in Human Behavior*, vol. 66, pp. 236–247, 2017.

[82] D. Derks, A. B. Bakker, and M. Gorgievski, "Private smartphone use during worktime: A diary study on the unexplored costs of integrating the work and family domains," *Computers in Human Behavior*, vol. 114, p. 106 530, 2021.

[83] É. Duke and C. Montag, "Smartphone addiction, daily interruptions and self-reported productivity," *Addictive Behaviors Reports*, vol. 6, pp. 90–95, 2017.

[84] A. F. Ward, K. Duke, A. Gneezy, and M. W. Bos, "Brain Drain: The Mere Presence of One's Own Smartphone Reduces Available Cognitive Capacity," *Journal of the Association for Consumer Research*, vol. 2, no. 2, pp. 140–154, 2017.

[85] T. Dingler and M. Pielot, "I'll be there for you: Quantifying Attentiveness towards Mobile Messaging," in *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, New York, NY: ACM, 2015, pp. 1–5.

[86] M. Bazire and P. Brézillon, "Understanding Context Before Using It," in *Modeling and Using Context*, D. Hutchison *et al.*, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 29–40.

[87] D. J. Grüning, F. Riedel, and P. Lorenz-Spreen, "Directing smartphone use through the self-nudge app one sec," *Proceedings of the National Academy of Sciences*, vol. 120, no. 8, e2213114120, 2023.

[88] M. M. P. Vanden Abeele, "Digital Wellbeing as a Dynamic Construct," *Communication Theory*, vol. 31, no. 4, pp. 932–955, 2021.

Supplementary Information

A List of Social Media Apps

- Discord
- Facebook
- Facebook Lite
- Facebook Local
- Facebook Messenger
- Instagram
- Kik
- LinkedIn
- Pinterest
- Reddit
- Snapchat
- TikTok
- Tumblr
- Twitter
- Youtube

B Hyperparameter Spaces

B.1 LSTM Models

Hyperparameter Name	Min	Max	Step Size
Embedding Dimensions	5	50	5
Number of LSTM Layers	1	3	1
LSTM Units	4	64	4
Recurrent Dropout	0.2	0.5	0.1
LSTM L1 Regularization	$1e-5$	$1e-3$	Continuous (log)
LSTM L2 Regularization	$1e-4$	$1e-2$	Continuous (log)
Dense Layer Units	4	64	4
Dense Layer L1 Regularization	$1e-5$	$1e-3$	Continuous (log)
Dense Layer L2 Regularization	$1e-4$	$1e-2$	Continuous (log)
Dropout Rate Top Layer	0.2	0.5	0.1
Learning Rate	$1e-5$	$1e-2$	Continuous (log)

B.2 Transformer Models

Hyperparameter Name	Min	Max	Step Size
Embedding Dimensions	5	50	5
Number of Transformer Layers*	1	3	1
Transformer Units	4	64	4
Recurrent Dropout*	0.2	0.5	0.1
Transformer L1 Regularization	$1e-5$	$1e-3$	Continuous (log)
Transformer L2 Regularization	$1e-4$	$1e-2$	Continuous (log)
Dense Layer Units	4	64	4
Dense Layer L1 Regularization	$1e-5$	$1e-3$	Continuous (log)
Dense Layer L2 Regularization	$1e-4$	$1e-2$	Continuous (log)
Dropout Rate Top Layer	0.2	0.5	0.1
Learning Rate	$1e-4$	$1e-2$	Continuous (log)

B.3 Fine-Tuned Models

Hyperparameter Name	Min	Max	Step Size
Dense Layer Units	4	64	4
Dense Layer L1 Regularization	$1e-5$	$1e-3$	Continuous (log)
Dense Layer L2 Regularization	$1e-4$	$1e-2$	Continuous (log)
Dropout Rate Top Layer	0.2	0.5	0.1
Learning Rate	$1e-4$	$1e-2$	Continuous (log)

C Additional Model Evaluation Metrics

C.1 Performance of Global LSTM Model Across 20 Rounds of Hyperparameter Search

rank	acc	pre	rec	f1	auc
1	0.779	0.713	0.651	0.668	0.782
2	0.781	0.728	0.631	0.648	0.780
3	0.780	0.716	0.652	0.669	0.778
4	0.769	0.700	0.615	0.628	0.764
5	0.771	0.705	0.618	0.632	0.763
6	0.765	0.728	0.571	0.566	0.761
7	0.770	0.707	0.607	0.618	0.761
8	0.768	0.705	0.605	0.615	0.760
9	0.769	0.706	0.605	0.616	0.759
10	0.767	0.695	0.617	0.629	0.758
11	0.768	0.695	0.625	0.639	0.759
12	0.769	0.705	0.604	0.615	0.758
13	0.769	0.701	0.613	0.626	0.757
14	0.768	0.703	0.604	0.614	0.755
15	0.768	0.702	0.608	0.620	0.754
16	0.769	0.704	0.608	0.620	0.753
17	0.767	0.702	0.598	0.607	0.749
18	0.741	0.371	0.500	0.426	0.507
19	0.741	0.371	0.500	0.426	0.500
20	0.741	0.371	0.500	0.426	0.568

C.2 Performance of Global Transformer Model Across 20 Rounds of Hyperparameter Search

rank	acc	pre	rec	f1	auc
1	0.775	0.706	0.643	0.659	0.773
2	0.777	0.716	0.628	0.644	0.775
3	0.771	0.697	0.659	0.672	0.774
4	0.776	0.712	0.636	0.652	0.774
5	0.774	0.709	0.630	0.645	0.773
6	0.776	0.716	0.626	0.641	0.773
7	0.775	0.707	0.645	0.661	0.774
8	0.775	0.717	0.621	0.635	0.773
9	0.775	0.709	0.636	0.651	0.771
10	0.776	0.715	0.626	0.641	0.774
11	0.773	0.701	0.646	0.661	0.772
12	0.775	0.712	0.626	0.641	0.772
13	0.773	0.703	0.642	0.658	0.772
14	0.775	0.708	0.636	0.651	0.773
15	0.775	0.710	0.634	0.649	0.773
16	0.775	0.710	0.633	0.649	0.773
17	0.772	0.707	0.620	0.634	0.767
18	0.774	0.711	0.621	0.635	0.769
19	0.774	0.716	0.614	0.627	0.771
20	0.771	0.709	0.613	0.625	0.766

C.3 Distributions of Model Performance Scores (LSTM, Global Model)

	auc	acc	pre	rec	f1
mean	0.699	0.785	0.627	0.580	0.578
std	0.088	0.100	0.106	0.075	0.092
min	0.515	0.552	0.386	0.486	0.432
25%	0.635	0.714	0.570	0.522	0.508
50%	0.690	0.795	0.621	0.556	0.554
75%	0.752	0.875	0.691	0.631	0.645
max	0.908	0.978	0.886	0.798	0.806

C.4 Distributions of Model Performance Scores (Transformer, Global Model)

	auc	acc	pre	rec	f1
mean	0.679	0.779	0.620	0.560	0.554
std	0.072	0.103	0.089	0.054	0.068
min	0.507	0.566	0.408	0.494	0.447
25%	0.630	0.701	0.579	0.515	0.493
50%	0.677	0.768	0.629	0.546	0.543
75%	0.724	0.855	0.669	0.592	0.597
max	0.853	0.981	0.823	0.703	0.728

C.5 Distributions of Model Performance Scores (LSTM, Person-Specific Models)

	auc	acc	pre	rec	f1
mean	0.609	0.772	0.451	0.520	0.469
std	0.092	0.114	0.114	0.051	0.080
min	0.414	0.422	0.287	0.483	0.320
25%	0.544	0.689	0.374	0.500	0.425
50%	0.611	0.776	0.415	0.500	0.451
75%	0.681	0.857	0.468	0.500	0.479
max	0.817	0.981	0.829	0.724	0.740

C.6 Distributions of Model Performance Scores (Transformer, Person-Specific Models)

	auc	acc	pre	rec	f1
std	0.087	0.105	0.136	0.059	0.083
min	0.447	0.523	0.295	0.455	0.371
25%	0.560	0.691	0.432	0.500	0.452
50%	0.635	0.769	0.501	0.500	0.479
75%	0.685	0.841	0.643	0.569	0.570
max	0.861	0.981	0.915	0.752	0.757

C.7 Distributions of Model Performance Scores (LSTM, Fine-Tuned Models)

	auc	acc	pre	rec	f1
mean	0.702	0.791	0.626	0.577	0.569
std	0.089	0.098	0.136	0.080	0.102
min	0.522	0.605	0.372	0.492	0.426
25%	0.640	0.720	0.541	0.503	0.480
50%	0.689	0.796	0.625	0.556	0.553
75%	0.754	0.877	0.700	0.627	0.627
max	0.908	0.978	0.972	0.800	0.807

C.8 Distributions of Model Performance Scores (Transformer, Fine-Tuned Models)

	auc	acc	pre	rec	f1
mean	0.683	0.783	0.625	0.568	0.560
std	0.077	0.102	0.100	0.068	0.085
min	0.497	0.552	0.388	0.489	0.437
25%	0.635	0.707	0.572	0.513	0.486
50%	0.681	0.788	0.630	0.548	0.552
75%	0.731	0.859	0.688	0.609	0.616
max	0.858	0.981	0.823	0.753	0.766

The full set of results is available for download in CSV format on this project's OSF page (https://osf.io/rkswe/).

D Models Trained on Data Including Launcher and SystemUI Events

D.1 Visualization of Model Performance Scores

D.2 Performance of Global LSTM Model Across 20 Rounds of Hyperparameter Search

rank	acc	pre	rec	f1	auc
1	0.847	0.737	0.622	0.649	0.826
2	0.847	0.737	0.626	0.653	0.827
3	0.848	0.747	0.608	0.634	0.827
4	0.847	0.737	0.620	0.646	0.826
5	0.847	0.758	0.590	0.611	0.826
6	0.847	0.735	0.624	0.651	0.827
7	0.845	0.734	0.607	0.632	0.818
8	0.842	0.716	0.652	0.674	0.823
9	0.845	0.729	0.616	0.641	0.822
10	0.846	0.734	0.623	0.649	0.823
11	0.846	0.737	0.612	0.637	0.821
12	0.846	0.733	0.614	0.639	0.819
13	0.838	0.733	0.551	0.553	0.793
14	0.839	0.728	0.562	0.570	0.792
15	0.830	0.415	0.500	0.454	0.704
16	0.830	0.415	0.500	0.454	0.701
17	0.830	0.415	0.500	0.454	0.692
18	0.830	0.415	0.500	0.454	0.495
19	0.830	0.415	0.500	0.454	0.500
20	0.830	0.415	0.500	0.454	0.501

D.3 Performance of Global Transformer Model Across 20 Rounds of Hyperparameter Search

rank	acc	pre	rec	f1	auc
1	0.846	0.730	0.636	0.662	0.825
2	0.848	0.752	0.602	0.627	0.824
3	0.847	0.741	0.613	0.639	0.824
4	0.847	0.740	0.613	0.639	0.823
5	0.847	0.736	0.621	0.647	0.823
6	0.849	0.757	0.606	0.632	0.824
7	0.848	0.743	0.618	0.644	0.822
8	0.848	0.740	0.618	0.644	0.822
9	0.849	0.746	0.622	0.650	0.823
10	0.847	0.744	0.610	0.636	0.821
11	0.847	0.745	0.608	0.634	0.821
12	0.846	0.748	0.593	0.615	0.820
13	0.847	0.740	0.610	0.636	0.817
14	0.844	0.728	0.616	0.641	0.818
15	0.846	0.735	0.612	0.638	0.816
16	0.844	0.725	0.616	0.641	0.813
17	0.839	0.733	0.557	0.562	0.802
18	0.841	0.726	0.585	0.603	0.805
19	0.840	0.719	0.589	0.609	0.801
20	0.837	0.710	0.564	0.575	0.787

E Fine-Tuned Models (Top Layers Only)

E.1 Distributions of Model Performance Scores (LSTM)

	auc	acc	pre	rec	f1
mean	0.678	0.771	0.532	0.554	0.521
std	0.090	0.118	0.147	0.081	0.109
min	0.394	0.251	0.285	0.483	0.251
25%	0.626	0.689	0.410	0.500	0.447
50%	0.678	0.782	0.496	0.500	0.481
75%	0.723	0.856	0.638	0.601	0.599
max	0.902	0.978	0.906	0.801	0.805

E.2 Distributions of Model Performance Scores (Transformer)

	auc	acc	pre	rec	f1
mean	0.654	0.766	0.585	0.554	0.538
std	0.083	0.111	0.116	0.065	0.085
min	0.384	0.517	0.369	0.406	0.405
25%	0.599	0.681	0.496	0.500	0.475
50%	0.655	0.766	0.589	0.532	0.524
75%	0.703	0.846	0.650	0.592	0.590
max	0.819	0.981	0.915	0.767	0.776

F Models Trained with Weighted Loss to Adjust for Class-Imbalances

F.1 Performance of Global LSTM Model Across 20 Rounds of Hyperparameter Search

rank	acc	pre	rec	f1	auc
1	0.728	0.481	0.658	0.556	0.781
2	0.719	0.471	0.697	0.562	0.785
3	0.715	0.466	0.698	0.559	0.783
4	0.714	0.465	0.699	0.559	0.783
5	0.720	0.471	0.690	0.560	0.784
6	0.703	0.451	0.682	0.543	0.767
7	0.681	0.429	0.711	0.535	0.761
8	0.677	0.428	0.733	0.540	0.766
9	0.685	0.434	0.719	0.541	0.765
10	0.645	0.404	0.786	0.534	0.763
11	0.682	0.430	0.711	0.536	0.760
12	0.683	0.432	0.716	0.539	0.762
13	0.697	0.444	0.680	0.537	0.760
14	0.711	0.457	0.625	0.528	0.757
15	0.679	0.428	0.718	0.536	0.761
16	0.695	0.441	0.668	0.531	0.755
17	0.680	0.428	0.711	0.534	0.758
18	0.687	0.434	0.691	0.533	0.757
19	0.674	0.413	0.616	0.494	0.714
20	0.741	0.000	0.000	0.000	0.500

F.2 Performance of Global Transformer Model Across 20 Rounds of Hyperparameter Search

rank	acc	pre	rec	f1	auc
1	0.702	0.451	0.707	0.550	0.775
2	0.726	0.477	0.644	0.548	0.774
3	0.700	0.449	0.702	0.548	0.773
4	0.710	0.459	0.685	0.550	0.772
5	0.675	0.427	0.756	0.546	0.773
6	0.684	0.435	0.741	0.548	0.774
7	0.697	0.445	0.703	0.545	0.771
8	0.687	0.437	0.731	0.547	0.774
9	0.691	0.441	0.727	0.549	0.773
10	0.708	0.457	0.681	0.547	0.773
11	0.674	0.426	0.754	0.545	0.770
12	0.696	0.445	0.701	0.544	0.769
13	0.667	0.420	0.756	0.540	0.769
14	0.670	0.423	0.751	0.541	0.768
15	0.688	0.437	0.716	0.543	0.765
16	0.692	0.439	0.691	0.537	0.762
17	0.680	0.425	0.673	0.521	0.745
18	0.698	0.441	0.632	0.520	0.745
19	0.683	0.431	0.704	0.535	0.738
20	0.741	0.000	0.000	0.000	0.545

F.3 Distributions of Model Performance Scores (LSTM, Global Model)

	auc	acc	pre	rec	f1
mean	0.694	0.732	0.621	0.623	0.605
std	0.088	0.113	0.060	0.075	0.066
min	0.502	0.522	0.514	0.505	0.450
25%	0.634	0.648	0.578	0.573	0.565
50%	0.678	0.722	0.610	0.612	0.598
75%	0.739	0.815	0.668	0.656	0.645
max	0.909	0.969	0.776	0.828	0.774

F.4 Distributions of Model Performance Scores (Transformer, Global Model)

	auc	acc	pre	rec	f1
mean	0.681	0.705	0.611	0.612	0.583
std	0.076	0.124	0.058	0.064	0.064
min	0.519	0.479	0.490	0.492	0.457
25%	0.626	0.605	0.567	0.561	0.534
50%	0.679	0.682	0.605	0.616	0.585
75%	0.730	0.803	0.649	0.649	0.618
max	0.848	0.964	0.810	0.773	0.724

F.5 Distributions of Model Performance Scores (LSTM, Person-Specific Models)

	auc	acc	pre	rec	f1
mean	0.630	0.595	0.568	0.586	0.507
std	0.087	0.166	0.082	0.066	0.120
min	0.375	0.157	0.113	0.447	0.152
25%	0.580	0.504	0.535	0.541	0.459
50%	0.633	0.609	0.566	0.579	0.528
75%	0.688	0.688	0.612	0.632	0.582
max	0.834	0.945	0.787	0.774	0.765

F.6 Distributions of Model Performance Scores (Transformer, Person-Specific Models)

	auc	acc	pre	rec	f1
mean	0.554	0.580	0.419	0.518	0.398
std	0.075	0.251	0.173	0.044	0.145
min	0.411	0.083	0.041	0.442	0.076
25%	0.500	0.357	0.330	0.500	0.297
50%	0.546	0.646	0.459	0.500	0.447
75%	0.592	0.792	0.536	0.514	0.487
max	0.770	0.953	0.866	0.731	0.681

F.7 Distributions of Model Performance Scores (LSTM, Fine-Tuned Models)

	auc	acc	pre	rec	f1
mean	0.696	0.739	0.622	0.629	0.616
std	0.088	0.106	0.059	0.075	0.063
min	0.501	0.544	0.517	0.516	0.503
25%	0.636	0.658	0.581	0.574	0.571
50%	0.686	0.728	0.612	0.616	0.606
75%	0.743	0.813	0.666	0.661	0.658
max	0.909	0.970	0.776	0.826	0.774

F.8 Distributions of Model Performance Scores (Transformer, Fine-Tuned Models)

	auc	acc	pre	rec	f1
mean	0.682	0.720	0.608	0.623	0.605
std	0.076	0.104	0.056	0.064	0.057
min	0.520	0.529	0.499	0.499	0.495
25%	0.635	0.653	0.574	0.576	0.566
50%	0.679	0.693	0.601	0.625	0.604
75%	0.731	0.785	0.643	0.656	0.641
max	0.849	0.966	0.748	0.779	0.730

The full set of results is available for download in CSV format on this project's OSF page (https://osf.io/rkswe/).