SmartPause: Designing a Reinforcement Learning Model for Reducing Screen Time

Gustavo Kruger[1,2,3], Nikhil Sachdeva[1,2,3], Yixue Zhao[4], Tianyi Li[5] and Michael Sobolev[6,*]

[1]Universitat de Barcelona, Barcelona, Spain
[2]Université Paris Cité, Paris, France
[3]Universidade Lusófona, Lisboa, Portugal
[4]Yixue Research Institute, Washington, D.C., United States of America
[5]Purdue University West Lafayette, Indiana, United States of America
[6]Cornell Tech, New York, United States of America

Abstract

Many users want to reduce their screen time and time spent on social media. Yet, most existing interventions rely on static rules that may not fully account for variations in user context and receptivity. In this paper, we propose the design of *SmartPause*, a Just-in-Time Adaptive Intervention (JITAI) based on Reinforcement Learning (RL) that adapts intervention timing by learning directly from users' natural compliance behavior. The system utilizes a Q-learning algorithm to learn the optimal moments for delivering vibrations as subtle digital nudges. Informed by an extension of the Predictability, Computability, and Stability (PCS) framework, we assess the algorithm's stability and potential intervention burden prior to deployment. We conducted offline simulations using historical usage logs from 26 participants (totaling 1,118 user-days) and five simulated personas, varying their responses to the intervention regarding usage time and in-session compliance. We evaluated four versions of the RL algorithm, differing in the frequency at which the model was queried and the criteria for separating different sessions. Our analysis suggests that all tested algorithms present a risk of highly burdening the users with too frequent interventions, especially for personas that are more compliant with the intervention. While personalizing querying frequencies might increase intervention opportunities, they also increase the risk of a high user burden. Further simulations should explore alternative approaches for balancing learning rates and user burden risk. This work contributes to the evolving field of using machine learning and RL to inform JITAI, and will be further evaluated in field studies using the *SmartPause* app.

Keywords

Smartphone addiction, Digital nudges, Reinforcement Learning, JITAI, Mobile Health

1. Introduction

One of the most frequent desires of smartphone users is reducing screen time, and in particular, time spent on social media. Moreover, some aspects of smartphone and social media use have been suggested to be linked with negative effects on cognitive, mental, and physical health [1, 2, 3, 4]. To address excessive screen time usage, a range of intervention strategies have been tested with small to moderate effectiveness and a vast design space [5, 6, 7, 8, 9]. Among these, digital nudging is a promising autonomy-preserving strategy, guiding users without strict limitations [6, 7, 10, 11, 12]. Moreover, the low cost and adaptivity of typical digital nudges (e.g., notifications, haptic feedback) provide opportunities for such nudges to be provided in a personalized, contextual, and adaptive fashion [7, 10].

Most intervention studies, using nudges or not, are static and one-size-fits-all interventions [5], which may signify lost opportunities to proactively offer support for usage regulation [13] as well as less

Joint Proceedings of the ACM Intelligent User Interfaces (IUI) Workshops 2026, March 23-26, 2026, Paphos, Cyprus

*Corresponding author.
gmagrokr7@alumnes.ub.edu (G. Kruger); a22401269@alunos.ulht.pt (N. Sachdeva); yzhao@isi.edu (Y. Zhao); li4251@purdue.edu (T. Li); michael.sobolev@cornell.edu (M. Sobolev)
https://github.com/nikhil-sachdeva (N. Sachdeva); https://yixuezhao.com/ (Y. Zhao); https://tl-tianyi-li.github.io/ (T. Li); https://www.michaelsobolev.com/ (M. Sobolev)
0000-0002-2771-8595 (G. Kruger); 0000-0001-7544-4552 (N. Sachdeva); 0000-0003-3046-6621 (Y. Zhao); 0000-0002-1145-2526 (T. Li); 0000-0002-8931-7682 (M. Sobolev)

than optimal intervention modality or timing [14, 15, 16]. Another approach is self-nudging; however, users might set suboptimal strategies for themselves [12], and effects decay over time [8]. This calls for personalized, adaptive, and context-aware approaches [17, 18, 19]. Drawing on the field of mobile health (mHealth), the Just-in-Time Adaptive Intervention (JITAI) paradigm, which seeks to deliver the right type and amount of support at the right moment (i.e., when receptivity and efficacy are highest), can help guide the design of such an intervention for reducing screen time [9, 20, 21, 22].

Recent systems explored more intelligent mechanisms, with personalized and adaptive interventions to reduce screen time. The *GoodVibe* app [7] employed haptic feedback in the form of vibration cues as nudges toward less phone usage. In that study, one of the experimental conditions was the nudge delivery when a personalized usage time was met, but the effect was not different from a static goal, indicating that such a simple personalization method is not sufficiently "personalized" and therefore does not effectively differentiate from one-size-fits-all solutions. The *Socialize* app uses clustering and association rule mining to discover specific habitual behaviors and identify the contexts (e.g., specific times, locations, physical activities, battery levels) that trigger certain patterns of app usage [23]. The app then prompts the user to employ guided self-nudges in the form of implementation intentions, a psychological strategy in the form of "if-then" plans (e.g., "If using Facebook while in bed in the morning, then a predefined alternative activity is triggered, such as drinking a glass of water") [24]. Another approach, deployed by the apps *Time2Stop* and *MindShift*, identifies opportune moments to suggest breaks within a JITAI framework [9, 22]. *Time2Stop* employs a classifier to predict overuse sessions [9]. Their model adapts intervention timing according to user feedback in the loop, considering app usage variables (current app, session duration, scroll speed, touch frequency) and contextual variables (time of day, day of week, location, physical activity, ringer mode, battery, and screen brightness). *MindShift* adapts intervention content using a Large Language Model (GPT-4) to generate persuasive content, tailoring interventions to mental states (boredom, stress, inertia) and user values in addition to usage and contextual data [22].

However, the previously described approaches rely on sustained and explicit user involvement, which can be burdensome and may hinder long-term adoption. The *Socialize* app requires users to manually design their own interventions, while *Time2Stop* depends on prompted, non-ecological user behavior to train its model, and *MindShift* relies on asking the user their intentions every time a target app is opened. An alternative route for designing adaptive screen-time interventions is to infer intervention adequacy directly from users' natural reactions, without repeatedly querying their intentions. In this framework, compliance (i.e., ceasing usage) indicates that an intervention was delivered in an appropriate context, whereas non-compliance suggests that the intervention instance was not. Across repeated interventions, idiosyncratic or exogenous factors influencing compliance are expected to average out, allowing intervention acceptance to function as a behavioral proxy for contextual adequacy. This formulation casts adaptive intervention design as a series of decision-making problems under uncertainty. A Reinforcement Learning (RL) approach is well-suited to this problem because it allows the system to gradually learn in which contexts interventions work best by observing users' responses over time.

In this paper, we introduce *SmartPause*, an RL model that learns when to deliver minimal, context-sensitive nudges to interrupt unfolding smartphone sessions (brief vibration cues, as employed by Okeke and colleagues [7]). Our approach combines a compact state representation (app type, time-of-day, weekday/weekend, session length) with a reward that balances session breaks against intervention burden. We evaluate candidate versions of *SmartPause*'s algorithm offline using simulations that replay real session logs from 26 participants over 1,118 user-days. That enables intervention burden and algorithm performance analyses before field deployment, informing the design of the final application. Accordingly, our contributions in this study are:

1. We propose an RL approach to adaptive screen-time interventions that learns directly from users' compliance behavior;
2. We introduce the *SmartPause* app model, a minimal, context-aware vibration-based intervention that adapts timing to reduce smartphone usage while limiting user burden;

3. We evaluate how the reward function design and user personas affect learning dynamics and intervention burden through offline simulations on real-world smartphone usage data, and derive design implications for deploying RL interventions in practice.

2. Screen Time Dataset

The dataset used to simulate the intervention comes from data collected as part of a previous unpublished study employing vibrations as a means of reducing screen time. The dataset has complete screen time data for 26 users during 6 weeks, amounting to 167,539 application logs, and is comprised of 5 variables: (1) a unique identifier for each of the users; (2) the date when the log was registered; (3) the application the log refers to; (4) the duration of the log in seconds; and (5) the duration of the log in milliseconds.

Usage statistics are presented in Table 1. The participants exhibited high variability in their mobile usage habits. Screen time averaged 3.73 hours per day (SD = 2.26 hours), ranging from 1.21 to 7.91 hours. Frequency of launches also varied widely, ranging from 1,316 to 13,466 (Mean = 6,443.81). Average Session Duration was 112.71 seconds (SD = 50.92).

Table 1
Descriptive statistics of baseline smartphone usage - average users' metrics

Measure	Mean	Median	SD	Min	Max
Log Duration (minutes)	1.88	1.70	0.85	0.87	5.17
Total App Launches	6,443.81	5,475.00	3,829.21	1,316.00	13,466.00
Screen Time per day (hours)	3.73	2.92	2.26	1.21	7.91

Figure 1 shows on the left side the baseline average daily usage over time, relative to the first day of data collection. The group average remained relatively stable across the observation period, between 200 and 250 minutes per day, but individual usage varied drastically (gray area), which signals the usefulness of this dataset for studying a personalized and adaptive intervention. The distribution of app session durations has a highly left-skewed, heavy-tailed distribution. The highest count of logs corresponds to sessions lasting 0 to 10 seconds, representing mostly system logs and Android launches, as shown on the right side of Figure 1, which displays the total usage time by app category. The dataset is dominated by social media, system utilities, and Productivity/Utilities. Social Media apps accumulate around 1300 hours of total usage time, 50% more than the next most used category (Productivity/Utilities). We later excluded all the logs from Android's system interface in order to only consider logs that represent time actually interacting with the screen in our simulations.

2.1. Timestamp Imputation

The original dataset included detailed daily data but omitted timestamps for the logs. The time of day has been shown to be a relevant variable in previous studies, impacting receptivity to intervention [9, 22, 25]. For that reason, we decided to input synthetic timestamps to each log so that it would be possible to include time of day as a variable during simulations. The process for timestamp imputation assigned specific POSIXct start timestamps to each usage log. We implemented a stochastic algorithm to reconstruct realistic circadian patterns. The algorithm processed data at the user-day level, preserving the original sequence, date, and duration of logs while enforcing the following constraints:

- **Non-overlapping sessions:** The start time of a subsequent log was strictly constrained to occur after the completion of the previous log.
- **Circadian anchoring:** The first log of the day was assigned a randomized start time within a morning window (06:00–11:00) to simulate variable wake-up times. Additionally, a sleep window (typically 02:00–06:00) was protected, without any activity placed in this period, to mimic human sleeping patterns.

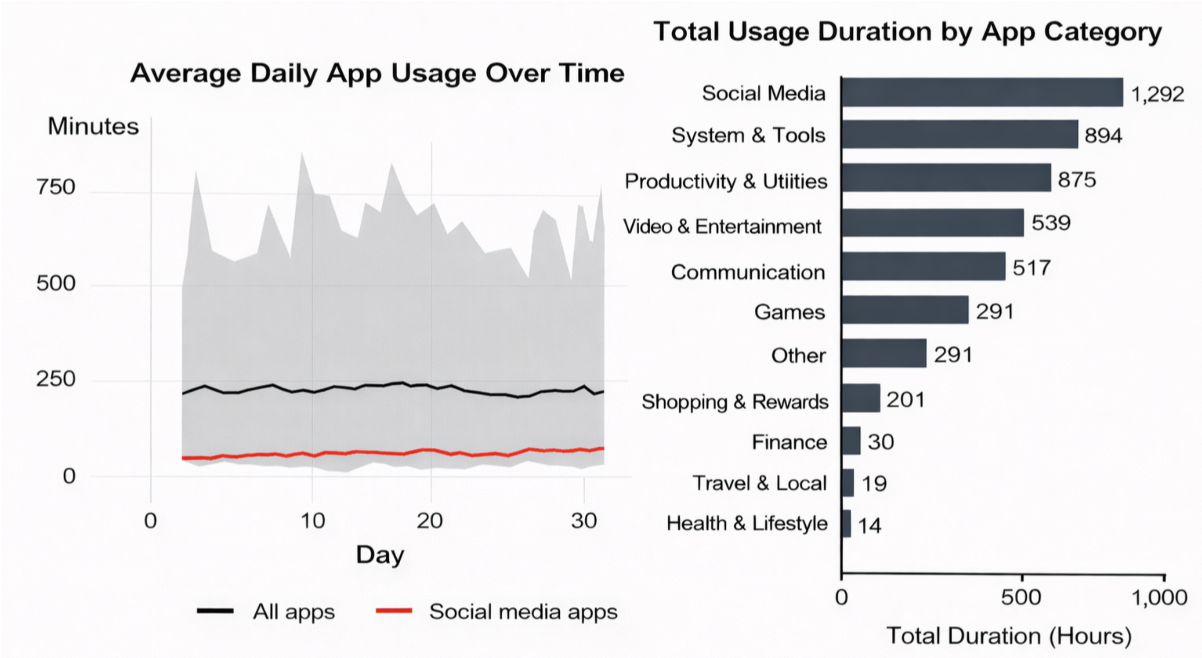

Figure 1: Original dataset average daily usage and total usage by app category.

- **Inter-session gaps:** Gaps between logs were generated using a time-dependent probabilistic model, simulating a realistic distribution at the session level [26]. Gap durations followed a log-normal distribution, with parameters shifting based on the time of day (e.g., shorter gaps were more probable during peak usage hours, while longer gaps were more likely in the evening).
- **Fatigue simulation:** To mimic human behavior, a "long break" (5–30 minutes) was enforced if cumulative continuous activity exceeded a threshold of 90 minutes.

3. Algorithm Design

To design an RL algorithm before deployment, our approach is informed by the extension of the PCS (Predictability, Computability, Stability) framework by Trella and colleagues [27]. This pre-implementation protocol emphasizes that RL agents designed for interventions with humans must be evaluated across diverse simulated environments, seeking computational feasibility, stability, and capability of predictable personalization.

The system implemented a tabular Q-learning algorithm to optimize the timing of interventions. The agent utilized a discount factor of $\gamma = 0.95$ and a learning rate of $\alpha = 0.1$. To balance exploration and exploitation, an ε-greedy strategy was employed with ε decaying from 1.0 to a minimum of 0.1 at a decay rate of 0.99 per update. The update rule is defined as:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[R_t + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t)] \quad (1)$$

3.1. Action Space and State Representation

The action space was binary, either producing a vibration or no vibration at each decision time, as represented by the equation:

$$a_t \in \{0 : \text{no vibration}, 1 : \text{vibrate}\} \quad (2)$$

Each decision state encoded contextual features expected to influence intervention receptivity and/or smartphone usage behavioral patterns. Therefore, the state space was constituted of the 32 possible combinations of the variables:

Type of the first app in a session (target / non-target): not all apps were considered targets for interventions. In the current simulation, we used social media apps as targets. The first app in a session was considered a proxy for the user's intention with the session, since previous research suggests intention is a strong predictor of later reports of regretting certain sessions [28] and intervening during sessions that users labeled as intentional ways of relaxation, for example, has been considered inappropriate [22]. It is assumed here that there is a higher probability that social media usage is intentional, in opposition to distracting, if the first app in a session was a social media one.

Time-of-day segment (morning, afternoon, evening, night): habitual phone use has been shown to be related to the different times of day. Previous studies found that receptivity to interventions also changes within a day [9, 22, 25].

Day of week (weekday/weekend): similarly to different times of day, different moments in the week are also expected to change smartphone usage due to changes in users' routines.

Session length (short/long): determining when to intervene during a usage session is one of the main expected outcomes of a JITAI approach. However, choosing the time granularity within sessions heavily impacts the computational power required for the algorithm to run, as well as how long it will take for the agent to be exposed to different state combinations several times and to update their corresponding Q-values. Therefore, we divided session length into two possible categories, short or long, in a personalized fashion, using each user's median session length in the baseline week as a cut-off point.

3.2. Reward Function Design

The reward function was developed to balance agentic learning and the potential for user burden. The immediate reward R_t is composed of four terms:

Intervention penalty: Every triggered vibration incurred a penalty of -40. This reflects the finding that while vibrations were effective, they were mildly irritating, functioning due to being aversive stimuli [7]. Consequently, this penalty discourages the agent from "spamming" the user.

Immediate compliance reward: Leaving the target apps during the vibration is considered a successful intervention and yields a $+60$ reward. This magnitude was selected to outweigh the intervention cost, resulting in a net positive reward of $+20$. This ensures that the agent is incentivized to intervene when it predicts a high probability of compliance, viewing the temporary irritation of the vibration as a necessary cost for a successful behavioral correction.

Pause bonus reward: To reinforce sustained disengagement, an additional bonus of $+40$ is awarded if the user remains away from all target apps for at least two minutes following an intervention.

Inactivity penalty: Sessions with significant target apps usage (higher than the target apps sessions median length during the baseline week) and no vibration are penalized to prevent the agent from maximizing rewards by simply remaining passive, avoiding the vibration penalty. A penalty of -40 is applied to those missed opportunities.

4. Evaluation Methods

4.1. Simulation Framework

SmartPause was evaluated through offline simulations that replayed historical usage sequences for each participant. At each decision time, the algorithm decided whether to deliver a subtle vibration cue (vibrate) or remain inactive (no vibration). Rewards were computed retrospectively at session termination. To simulate the response of the user, a compliance model governed by five different user personas defined the response to the intervention. The simulation pseudocode is presented in Figure 2. At a higher level, a compiled vibration successfully shortened the ongoing session, simulating a user stopping to use the phone. Otherwise, usage continued unaltered. Daily sessions were processed sequentially, and policy updates occurred daily (Figure 2).

```
for participant in participants:
  policy = initialize_policy(participant)

  for day in participant.days_in_history (chronological):

    for session in day.sessions (chronological):

      if not session.includes_any(Target_apps): # filters sessions with target apps
        continue                                # usage continues

      for t in session.decision_times (chronological):

      action = policy.decide(extract_state(session, t))    # {VIBRATE, NO_VIBRATION}

        if action == VIBRATE:
          persona = participant.persona         # one of 5 personas
          complied = persona.decide_compliance()  # True/False draw

          if complied:
            session.end(time=t)                 # terminate session early
            break                               # stop this session
          else:
            continue                            # usage continues
        else:  # NO_VIBRATION
          continue                              # usage continues

    policy = policy.update(using=day.simulated_experience)  # daily update
```

Figure 2: Offline simulation pseudocode for *SmartPause*. Historical usage sequences are replayed; the policy chooses whether to vibrate at each decision time; a persona-based compliance model determines whether vibration terminates the session; the policy is updated once per day.

4.2. User Personas

To be able to evaluate the RL algorithm offline, we must simulate how users would react to the interventions [29]. In order to test the algorithm stability across a diverse range of possible user behaviors, we simulated diverse responses to the interventions, varying compliance levels, and target application usage. We utilized user personas to represent these heterogeneous responses, focusing on having a diverse set of user behavior patterns to examine how the algorithm behaves in each of them, prioritizing differentiation over realism. The utilized personas are the following:

- **Ideal User:** Represents a highly receptive individual. This persona has a high compliance rate (80% probability of stopping a session when nudged) and maintains a steady reduction in overall usage (85% of the original volume) throughout the 42-day simulation. This magnitude of usage reduction is similar to the results found by field studies aiming at smartphone usage reduction [7, 9].
- **Non-Responder:** Represents a user who ignores the intervention. With a compliance rate of only 5% and a usage identical to the historical baseline, this persona tests the algorithm's ability to

minimize "spamming" (intervention costs) when nudges are ineffective. Since some users might ignore the vibrations, it is relevant to understand how the algorithm adapts.

- **Habituator:** Models the "novelty effect," where effectiveness fades over time. This persona starts with a high change in usage (80% usage reduction) but only in the first 7 days, while keeping compliance to vibration at 40%. This tests the agent's ability to detect changing reward probabilities. Follow-up measures on Okeke and colleagues' field study suggested that after an initial reduction, usage increases again [7]. Therefore, we simulate algorithmic performance on a similar longitudinal pattern.
- **Uses More:** Represents a user who reacts negatively to the intervention. This persona has low compliance (20%) and actually increases their usage to 115% of the baseline, following an increasing trajectory. Although this is not the behavior we would expect from users during the field deployment of this algorithm, it is relevant to understand how such patterns affect the algorithm's behavior.
- **Delayed Effect:** The effects of the intervention on usage are a delayed reduction of 15%, starting after a week. Compliance is moderate, set at 60%. This persona was designed to capture a pattern where habits take time to adjust despite consistent nudging. It evaluates the algorithm when the desired goal is achieved; however, it does not do so immediately.

Each of the original 26 unique user traces was simulated under the conditions of five distinct behavioral personas. This resulted in a total of 130 simulated user environments (26 users times 5 personas). This expansion allows us to assess whether the algorithm can effectively learn and adapt to more diverse usage patterns.

4.3. Evaluation Metrics

To evaluate the performance of the algorithm versions (combinations of varying parameters), we analyze metrics that were expected to capture the trade-off between personalization and learning (relative to agentic reward over time) and user burden (excessive vibrations). Following the recommendations for RL in mobile health proposed by Trella and colleagues [27], we do not rely solely on aggregate means, as they may mask poor performance for specific users. Instead, all metrics are reported for two distinct groups: all the simulated users and the subset of users with the lowest accumulated rewards (specifically, the 25th percentile of total rewards). The metrics are (a) Mean Reward per Day, (b) Mean Vibrations per Day, (c) Frequency of High-Burden Days (days when vibrations exceeded 20), (d) Mean Reward per Vibration, (e) Mean Vibration per Query, and (f) Missed Opportunities (number of long sessions without vibrations).

To identify a safe and efficacious configuration for the deployment version of *SmartPause*, we conducted simulations across a matrix of 4 experimental settings. These settings vary two core parameters: the frequency of decision times (model querying frequency), and how long an inactivity interval has to last for the following usage to be considered as a new session (session separation threshold). We compare a fixed querying frequency (decision times happened every 60 seconds) with a personalized frequency (considering the median session duration M for each user, decision times happened every M seconds). For session separation thresholds, we tested a short usage gap of 45 seconds [30] and a longer gap of 2 minutes.

5. Evaluation results

5.1. Intervention Burden

To assess the risk of intervention fatigue, we analyzed the frequency of vibrations generated by the agent across the different versions. Table 2 illustrates the range of daily vibrations (minimum and maximum), and the frequency of "high-burden days," defined here as days when users would be receiving more than 20 vibrations.

Table 2
Intervention Intensity (Daily Vibrations) and User Burden by Version and Persona.

Version	Persona	Max Vibs	% High Burden Days (> 20)
Fixed query/ Separation 120s	Non-responder	112	2.6%
	Uses More	102	4.6%
	Effective Habituator	126	11.2%
	Delayed Effects	96	17.7%
	Ideal User	100	18.4%
Fixed query/ Separation 45s	Non-responder	104	2.7%
	Uses More	93	5.8%
	Effective Habituator	114	11.0%
	Delayed Effects	107	15.6%
	Ideal User	98	15.7%
Personalized query/ Separation 120s	Non-responder	126	3.0%
	Uses More	139	4.5%
	Effective Habituator	117	11.3%
	Delayed Effects	134	19.9%
	Ideal User	121	21.6%
Personalized query/ Separation 45s	Non-responder	102	3.1%
	Uses More	110	6.7%
	Effective Habituator	97	12.8%
	Delayed Effects	130	19.1%
	Ideal User	118	20.0%

Ideal User, Effective Habituator, and Delayed Effects personas were subjected to the highest volume of interventions, with the most frequent high burden days. For the Ideal User, high-burden days occurred in 15.7% to 21.6% of the simulation period. In contrast, the Non-responder experienced high-burden volumes in less than 4% of days across all strategies.

This behavior indicates that the Q-learning agent successfully learned the negative reward structure: for non-responsive users, the penalty for intervening was not offset by the compliance reward, causing the agent to reduce intervention frequency to minimize costs. However, for the more compliant users, the agent maximized cumulative reward by triggering vibrations more frequently, potentially leading to excessive interruption. However, even the less compliant user personas still had days with very high numbers of vibrations (over 100).

5.1.1. Query Frequency

Comparing the algorithmic strategies, the Fixed query configurations consistently demonstrated a lower risk of burden compared to the personalized frequencies, and this separation was proportional to the compliance levels. For example, for the Ideal User persona, the 45 seconds separation and Fixed Query frequency version resulted in a maximum of 98 daily vibrations and a 15.7% high-burden rate, whereas the 45 seconds separation and Personalized Query frequency reached 118 vibrations with a 20.0% high-burden rate. Personalized strategies appear to be more aggressive in exploiting compliant behavior. The session separation parameter showed a less pronounced impact on burden than the query method, though the 120-second separation versions had slightly more frequent high burden days for high compliance personas.

These results suggest that while the RL agent effectively identifies when to stop bothering non-responsive users, it lacks an intrinsic upper bound for responsive users, risking intervention fatigue for the most compliant participants. Consequently, the Fixed Query frequencies approach appears to be the safer candidate for deployment to mitigate the risk of extreme outlier days.

5.2. Algorithmic Learning

To verify the agent's ability to learn optimal policies and adapt to the specific constraints of the environment, we analyzed the evolution of rewards and decision probabilities over the simulated 42-day period.

5.2.1. Reward Trajectories and Policy Adaptation

The agent demonstrated a clear learning curve across all scenarios. As shown in the Mean Reward per Day plots (Figure 3), the average daily reward typically started low or negative and improved. The improvement was slow for the personas with lower compliance levels, and rapid for the Ideal User, happening within the first 7 to 10 days and then stabilizing. For the *Ideal User* persona, daily rewards rose from an initial range of 0-200 to stabilize between 200 and 400. Conversely, for the *Non-responder* and *Uses More* personas, it appears that the agent learned to minimize losses, stabilizing negative rewards around -200.

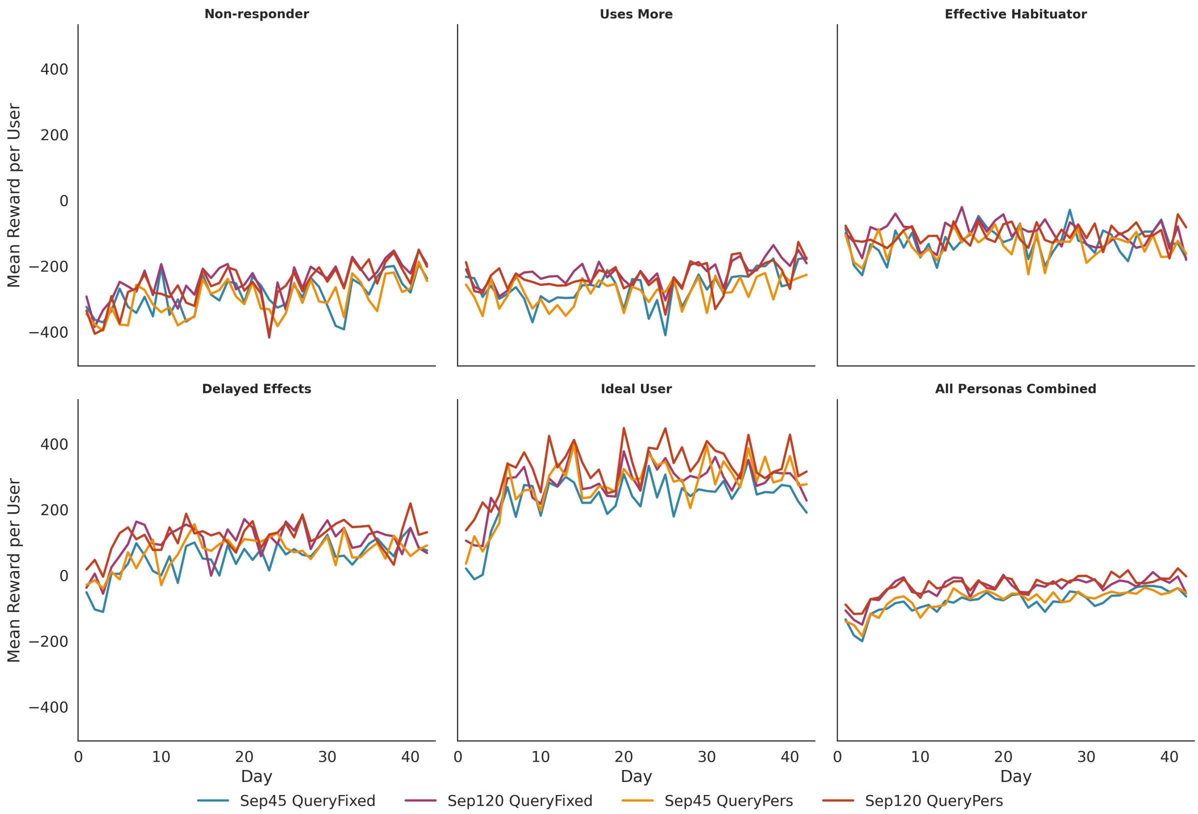

Figure 3: Mean reward per day for each persona and algorithm version.

This adaptation is mechanistically visible in the *Vibrations per Query* metric (Figure 4). This metric acts as a proxy for the learned policy, representing the probability of taking action (vibrate) given a query.

For more compliant personas (*Ideal User*, *Delayed Effects*), the vibration probability increased from an initial greedy baseline of 0.5 to nearly 0.8. The agent learned that the expected return of vibrating was positive, thus exploiting the action most of the time the state permitted. For the *Non-responder*, the metric dropped from 0.5 to below 0.2. The agent correctly identified that vibrations yielded net negative utility (cost vs. negligible compliance reward) and restricted interventions.

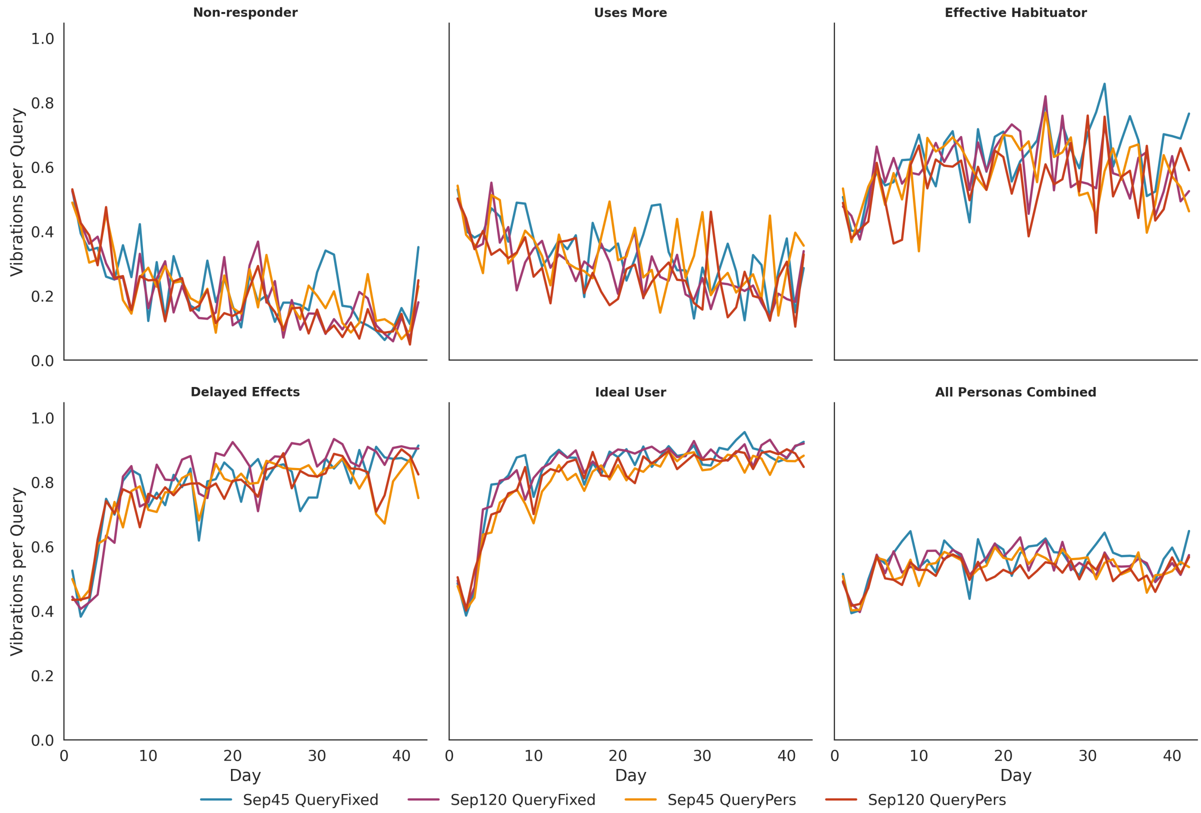

Figure 4: Mean vibration per decision time (query) for each persona and algorithm version.

5.2.2. Efficiency and Coverage

The *Mean Reward per Vibration* (Figure 5) remained relatively stable for all personas except the *Non-responder*, confirming that the agent's state representation was sufficient to distinguish productive contexts, but kept exploring for the full duration of the simulation when compliance was set to 5%. For the *Ideal User* and the *Delayed Effects*, each vibration consistently yielded a positive net utility, while for *Non-responders*, it remained negative. The impact of this learning on user coverage is evidenced by the *Missed Opportunities* metric (Figure 6), defined as long sessions where no vibration was triggered. For the *Ideal User*, missed opportunities dropped drastically from an average of 4-7 per day in the first week to 1-2 per day by the end of the simulation. This confirms the agent effectively maximized coverage for receptive users. In contrast, for *Non-responders*, the number of missed opportunities remained higher (5 per day) and increased during the simulation, probably as a result of consistent punishment of vibrations, leading to fewer session interruptions.

5.2.3. Performance in Low-Reward Scenarios (P25)

To assess robustness, we examined the 25th percentile of users (P25) in regard to accumulated rewards over the intervention duration, representing the most difficult cases for the algorithm. The data for P25 users exhibits higher volatility (Figure 7), and the vibration frequency was probably more sensitive to daily variations in usage, suggesting the current algorithm was not able to optimize its learning for part of the users. However, some of the fundamental trends persist, for example, the average rewards and the average rewards per query both trended upwards over the simulation duration.

5.2.4. Session Separation Threshold

The separation threshold parameter showed a subtler impact on overall performance. Intuitively, a shorter separation allows for higher intervention density, theoretically reducing missed opportunities

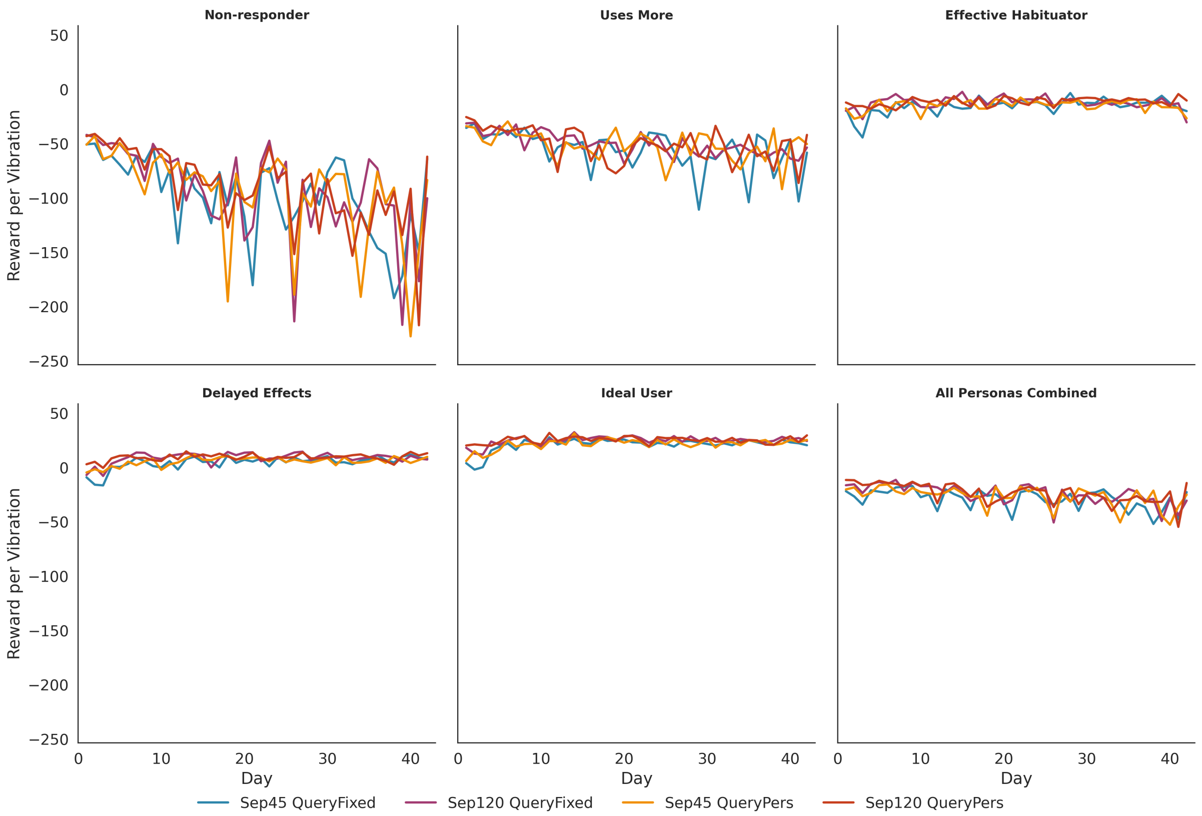

Figure 5: Mean reward per vibration for each persona and algorithm version.

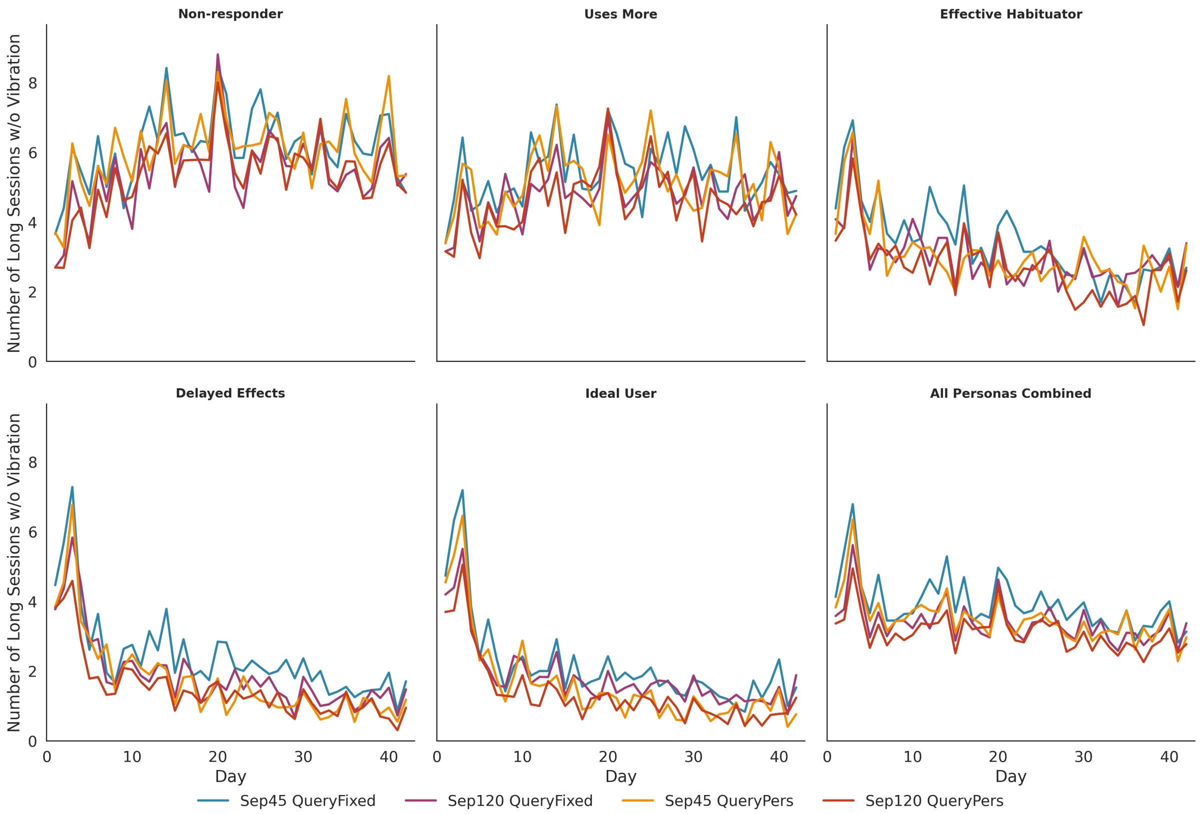

Figure 6: Number of long sessions without vibrations (missed opportunities) for each persona and algorithm version.

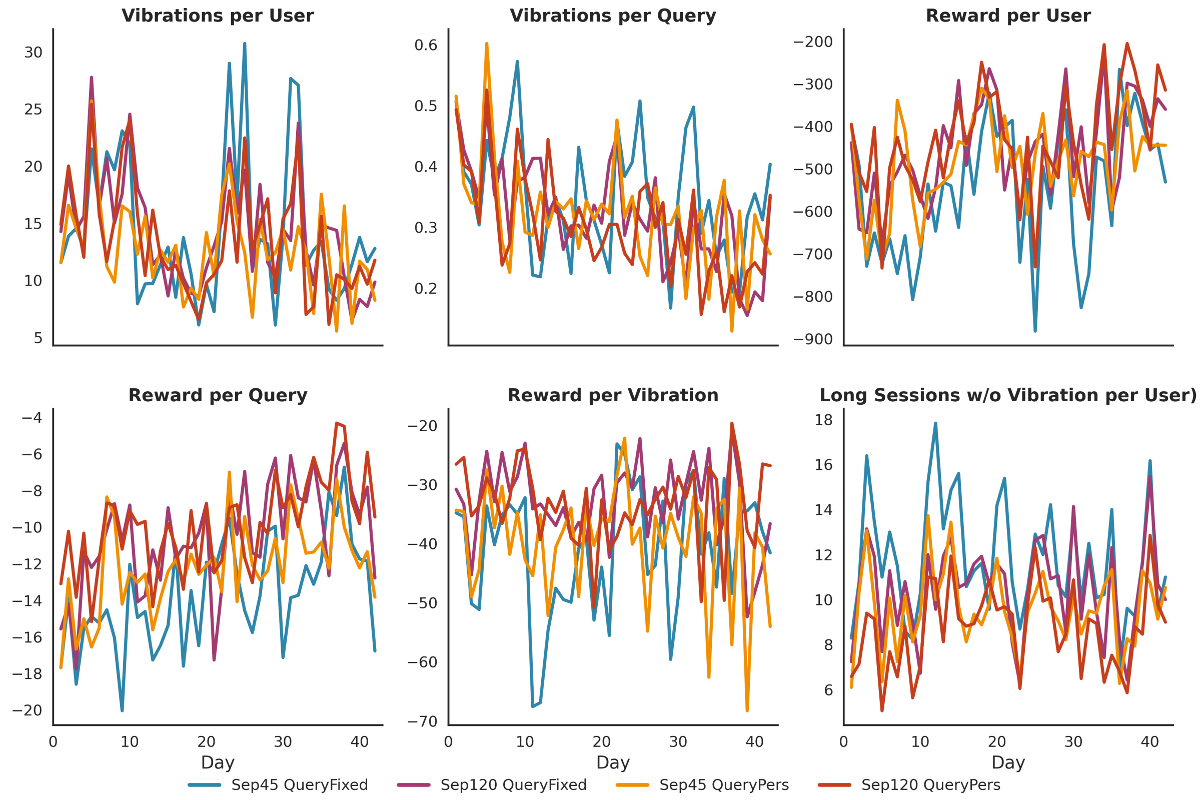

Figure 7: Metrics for the 25th percentile of users with the lowest average rewards, combining the personas.

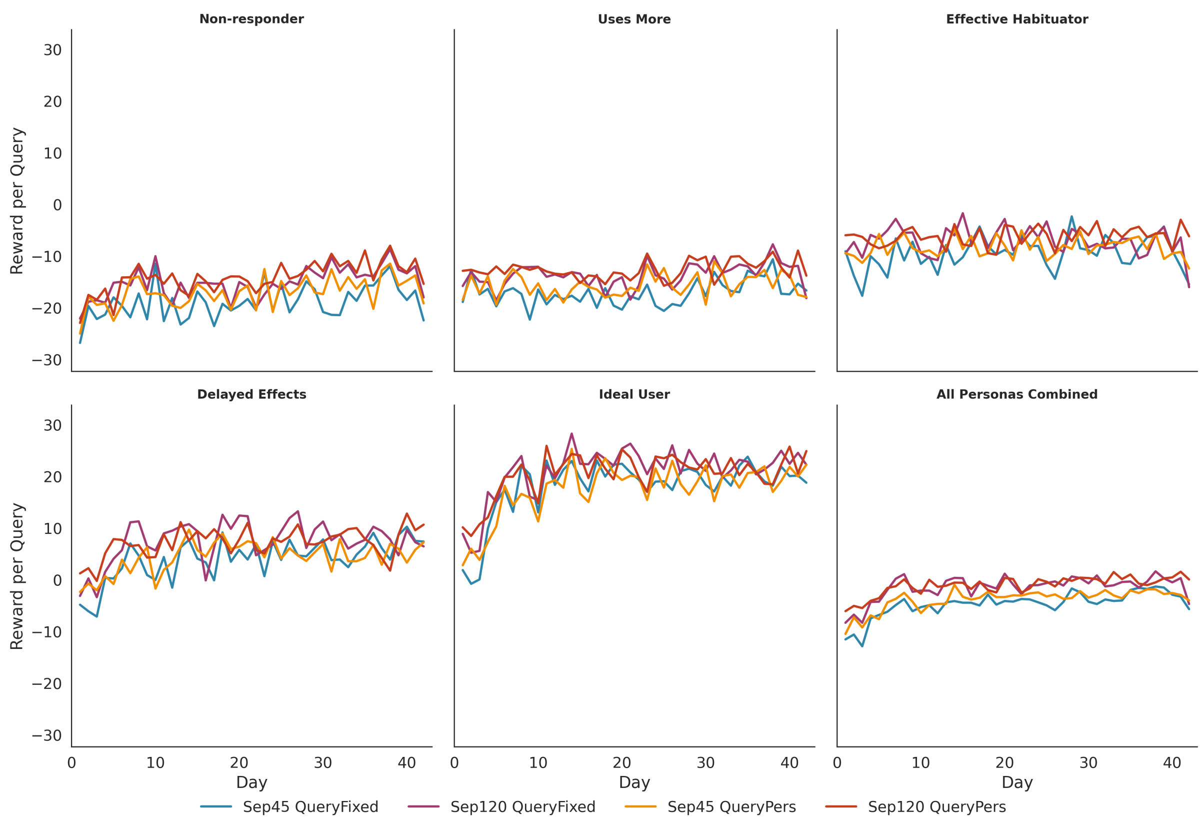

Figure 8: Mean reward per decision time (query) for each persona and algorithm version.

in long sessions. However, the data in Figure 6 indicates that the reduction in "Missed Opportunities" (long sessions without vibration) was marginal when moving from 120 seconds to 45 seconds.

5.2.5. Convergence and Policy Stability

All four algorithmic versions converged, as shown in the Vibrations per Query and Reward per Query metrics (Figures 4 and 8). The learning curves for all versions followed the same trajectory: rapidly increasing vibration probability for compliant personas (converging to 0.8) and decreasing for non-responders (converging to less than 0.2). This suggests that the state representation is enough to support learning regardless of the specific timing parameters simulated so far. The choice of version, considering the results of the simulations presented, does not compromise the core learning capability, but acts as a hyperparameter for user burden.

6. Discussion and Future Directions

This study evaluated an RL algorithm designed to reduce screen time through JITAI vibrations, learning directly from users' compliance with interventions. By simulating 1,118 user-days, derived from a real dataset collected from 26 users, across distinct behavioral personas, we assessed the algorithm's stability, learning capability, and potential intervention burden before field deployment. The results indicate that a tabular Q-learning agent with a compact state representation can differentiate between receptive and non-receptive contexts, adapting its policy to maximize rewards or minimize costs depending on the user's compliance profile. However, the analysis also highlighted significant trade-offs between maximizing intervention opportunities and mitigating the risk of user annoyance among highly compliant users.

6.1. Versions Comparison

The comparison of algorithmic configurations reveals that the mechanism for triggering state queries (Fixed vs. Personalized) has a more profound impact on user burden than the time threshold to separate sessions (45s vs. 120s). All versions led to excessive user burden, which can hinder the viability of this approach in deployment, and this was particularly evident in personalized query strategies. This resulted in "high-burden days" exceeding 20% for higher compliance personas, a volume of interruption likely to cause attrition in real-world settings. The fixed-interval query frequencies somewhat limited the vibrations, although not sufficiently to be acceptable to a typical user. The personalized frequencies meant shorter separation thresholds, which translates into more sessions and therefore more queries. That theoretically offers higher density for intervention, as was confirmed by the simulation data and reflected in the observed higher burden. Among the ones simulated, the configuration with the lowest average burden was the combination of a lower threshold for session separation and fixed every-minute queries to the agent, therefore emerging as the most conservative choice for deployment. However, further simulations should explore different parameters or even vibration budgets to avoid excessive user burden.

6.2. Current Limitations

The current analysis is limited to the algorithmic versions that were explored and discussed here. More generally, our approach relies on simulated user responses, which are defined by certain assumptions about human behavior. Different simulation goals require and benefit from varying assumptions. To analyze stability, the variability level employed between personas allowed valuable results. However, if the goal becomes to determine the algorithmic conversion rates for different state variables, such as times of day, the lack of state-dependent compliance variations would be problematic.

Additionally, the use of synthetic timestamps for the historical logs introduces uncertainty. Although the imputation algorithm respected circadian constraints, sleep windows, and usage breaks, the specific alignment of usage sessions with time-of-day segments in the simulation may not perfectly reflect the original users' routines. Since time-of-day is a state variable, the agent's learned preference for specific temporal windows in the simulation might differ from the policies it would learn from the ground-truth timing. Finally, the Q-learning approach requires discrete state variables, necessitating the binning of continuous variables (e.g., session length), which inevitably results in a loss of information granularity compared to continuous state-space models.

6.3. Future Directions

Further simulations should address these three important areas, with necessary adaptation of the current settings. Specifically, we aim to expand the work in three directions:

- **Varying the tests of algorithm performance** Simulations should test more ways in which algorithmic performance could differentiate the tested versions, for example, designing personas that vary their compliance levels linearly, non-linearly, and depending on the state variables.
- **Expanding the state space** Future algorithmic iterations should explore the expansion of the state space to include more granular context, such as the daily vibration count and/or the compliance level for vibrations in the last hour. However, as the state space grows, transitioning from tabular Q-learning to Deep Q-Networks (DQN) may be necessary to handle the dimensionality without sacrificing convergence speed.
- **Optimizing the reward function** Future simulations should explore ways to reduce the number of vibrations users are exposed to in high burden days. This could be accomplished by setting a per-day or per-session vibration budget while evaluating the learning metrics. Another route would be to change current reward values, such as increasing the penalty for vibrating.

These steps are designed to inform the deployment of the *SmartPause* app in a longitudinal field study. This deployment will enable the assessment of the agent's "cold start" period and determine whether the pre-trained Q-tables from the simulation can serve as effective initializations to accelerate on-device learning.

6.4. App Design and Implementation

The comparative analysis reveals a trade-off between maximizing screen time interruptions and minimizing user annoyance. Although personalized query frequencies enabled more frequent interventions, they also increased the number of high-burden days, raising the risk of user disengagement or app abandonment. As a result, the configuration using a 45-second session separation with a fixed 60-second query frequency appears to be the most balanced option among those tested. Nonetheless, this setup still imposes an unacceptable level of user burden, indicating the need for further simulations before deployment.

Following offline evaluation through simulations, the *SmartPause* framework will be deployed as an Android application running primarily as a background service. The source code for the Android application and the reinforcement learning environment are available on GitHub, and will be updated as the project progresses [31]. The app monitors smartphone usage in real time to deliver JITAI vibrations and collect behavioral data. During installation, users select target applications for intervention from their installed apps; this list is fixed for the duration of the study to ensure consistency.

SmartPause consists of three core components: (1) a locally stored Q-table enabling real-time decision-making on the device, (2) a cloud-based relational database storing historical usage data and models, and (3) a backend API where the Q-Learning agent performs model training. A continuously running background service processes Android "UsageStatsManager" events, grouping consecutive foreground events within a certain time window into usage sessions. From each session, a four-dimensional state representation is extracted and used to query the local Q-table to decide whether to intervene.

When an intervention is triggered, the phone vibrates with the same patterns of the *GoodVibes* app [7], providing a distinct but non-disruptive tactile cue. Session-level data (timestamps, app identifiers, duration, vibration events, and user compliance) is stored locally, allowing the system to learn from real user behavior rather than simulated responses. To support continuous learning while minimizing battery, bandwidth, and user disruption, model updates occur once daily during low-usage hours. The app securely uploads the previous day's session data to the backend, which updates the Q-table. The updated model is then downloaded and loaded onto the device for immediate use. This architecture enables offline, real-time intervention decisions while benefiting from daily cloud-based learning.

7. Conclusion

Our work aims to design and evaluate low-burden, adaptive screen-time interventions that learn directly from users' behavior. We introduced *SmartPause*, a reinforcement learning–based framework that infers intervention adequacy from compliance to minimal vibration cues, an alternative for designing adaptive screen time interventions. Using offline simulations replaying 1,118 user-days from real usage logs, we demonstrated that a compact, tabular Q-learning agent can learn context-sensitive intervention policies and differentiate between receptive and non-receptive usage contexts. Our results reveal substantial trade-offs between intervention density and user burden. All tested configurations produced levels of interruption that would likely hinder real-world adoption. Future work shall investigate expanding the state space, refining reward structures, and exploring alternative learning architectures to reduce user burden while preserving adaptive capacity. These refinements will inform the longitudinal deployment of the *SmartPause* app, enabling the study of the validity of simulation-trained policies in ecological settings. This work contributes to the evolving field of using machine learning and RL to inform JITAI for health and behavior change.

Acknowledgments

We thank Fabian Okeke for designing the original app and conducting the study that collected data for this simulation.

Declaration on Generative AI

During the preparation of this work, the authors used Gemini-3-Pro and Grammarly in order to: grammar/spelling check, rephrasing, and double-check the reference list. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] R. M. S. Santos, S. D. A. Ventura, Y. J. D. A. Nogueira, C. G. Mendes, J. J. D. Paula, D. M. Miranda, M. A. Romano-Silva, The associations between screen time and mental health in adults: A systematic review, Journal of Technology in Behavioral Science (2024) 1–21. doi:10.1007/s41347-024-00384-y.

[2] L. A. S. Brautsch, L. Lund, M. M. Andersen, P. J. Jennum, A. P. Folker, S. Andersen, Digital media use and sleep in late adolescence and young adulthood: A systematic review, Sleep Medicine Reviews 68 (2023) 101742. doi:10.1016/j.smrv.2022.101742.

[3] P. Haghjoo, G. Siri, E. Soleimani, M. A. Farhangi, S. Alesaeidi, Screen time increases overweight and obesity risk among adolescents: A systematic review and dose-response meta-analysis, BMC Primary Care 23 (2022). URL: https://doi.org/10.1186/s12875-022-01722-2. doi:10.1186/s12875-022-01722-2.

[4] H. Wen, M. Sobolev, R. Vitale, J. Kizer, J. Pollak, F. Muench, D. Estrin, et al., mpulse mobile sensing model for passive detection of impulsive behavior: exploratory prediction study, JMIR Mental Health 8 (2021) e25019.

[5] A. M. Roffarello, M. E. Minchia, D. Fogli, H. Simoes, F. Paterno, E. Velloso, Achieving digital wellbeing through digital self-control tools: A systematic review and meta-analysis, ACM Transactions on Computer-Human Interaction 30 (2023) 1–38. doi:10.1145/3571810.

[6] L. Zimmermann, M. Sobolev, Digital strategies for screen time reduction: A randomized field experiment, Cyberpsychology, behavior and social networking 26 (2024) 42–49. doi:10.1089/cyber.2022.0027.

[7] F. Okeke, M. Sobolev, N. Dell, D. Estrin, Good vibrations: Can a digital nudge reduce digital overload?, in: Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI ’18), ACM, Barcelona, Spain, 2018, pp. 1–12. doi:10.1145/3229434.3229463.

[8] D. J. Grüning, F. Riedel, P. Lorenz-Spreen, Directing smartphone use through the self-nudge app one sec, Proceedings of the National Academy of Sciences 120 (2023) e2213114120. doi:10.1073/pnas.2213114120.

[9] A. Orzikulova, H. Xiao, Z. Li, Y. Yan, Y. Wang, Y. Shi, X. Xu, Time2stop: Adaptive and explainable human-AI loop for smartphone overuse intervention, in: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI ’24), ACM, Honolulu, HI, USA, 2024, pp. 1–20. doi:10.1145/3613904.364274.

[10] M. Sobolev, Digital nudging: Using technology to nudge for good, SSRN Electronic Journal (2021). URL: https://ssrn.com/abstract=3889831.

[11] T. Lu, H. Zheng, T. Zhang, X. Xu, A. Guo, Interactout: Leveraging interaction proxies as input manipulation strategies for reducing smartphone overuse, in: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI ’24), ACM, Honolulu, HI, USA, 2024, pp. 1–19. doi:10.48550/arXiv.2401.16668.

[12] G. Kruger, N. Sachdeva, P. Sleboda, L. Zimmermann, M. Sobolev, Self-nudging using the digital wellbeing app, in: Proceedings of the 9th International Conference on Computer-Human Interaction Research and Applications (CHIRA 2025), SCITEPRESS, Marbella, Spain, 2025, pp. 1–15. doi:10.2139/ssrn.5476626.

[13] A. M. Roffarello, L. D. Russis, Nudging users or redesigning interfaces? evaluating novel strategies for digital wellbeing through incontrol, in: Proceedings of the 2023 ACM International Conference on Information Technology for Social Good (GoodIT ’23), ACM, 2023, pp. 17–26. doi:10.1145/3582515.3609523.

[14] S. Y. Schoenebeck, Giving up Twitter for Lent: How and why we take breaks from social media, in: Proceedings of the 2014 CHI Conference on Human Factors in Computing Systems (CHI ’14), ACM, Toronto, Canada, 2014, pp. 773–782. doi:10.1145/2556288.2556983.

[15] E. Oduor, C. Neustaedter, W. Odom, A. Tang, N. Moallem, M. Tory, P. Irani, The frustrations and benefits of mobile device usage in the home when co-present with family members, in: Proceedings of the 2016 ACM Conference on Designing Interactive Systems (DIS ’16), ACM, Brisbane, Australia, 2016, pp. 1315–1327. doi:10.1145/2901790.2901809.

[16] A. M. Roffarello, L. D. Russis, Coping with digital wellbeing in a multi-device world, in: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI ’21), ACM, Yokohama, Japan, 2021, pp. 1–14. doi:10.1145/3411764.3445076.

[17] F. I. Rahmillah, A. Tariq, M. King, O. Oviedo-Trespalacios, Evaluating the effectiveness of apps designed to reduce mobile phone use and prevent maladaptive mobile phone use: Multimethod study, JMIR mHealth and uHealth 11 (2023) e46813. doi:10.2196/42541.

[18] A. M. Roffarello, L. D. Russis, Nudging users towards conscious social media use, in: Proceedings of the 25th International Conference on Mobile Human-Computer Interaction (MobileHCI ’23 Companion), ACM, 2023, pp. 1–6. doi:10.1145/3565066.3608703.

[19] Y. Zhao, T. Li, M. Sobolev, Digital wellbeing redefined: Toward a user-centric approach for positive social media engagement, in: Proceedings of the 11th International Conference on

Mobile Software Engineering and Systems (MOBILESoft '24), ACM, Lisbon, Portugal, 2024, pp. 1–6. doi:10.1145/3647632.3651392.

[20] I. Nahum-Shani, S. N. Smith, B. J. Spring, L. M. Collins, K. Witkiewitz, A. Tewari, S. A. Murphy, Just-in-time adaptive interventions (JITAIs) in mobile health: Key components and design principles for ongoing health behavior support, Annals of Behavioral Medicine 52 (2018) 446–462. doi:10.1007/s12160-016-9830-8.

[21] O. Perski, E. T. Hébert, F. Naughton, E. B. Hekler, J. Brown, M. S. Businelle, Technology-mediated just-in-time adaptive interventions (JITAIs) to reduce harmful substance use: A systematic review, Addiction 117 (2022) 1220–1241. doi:10.1111/add.15687.

[22] R. Wu, C. Yu, X. Pan, Y. Liu, N. Zhang, Y. Shi, Mindshift: Leveraging large language models for mental-states-based problematic smartphone use intervention, in: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24), ACM, Honolulu, HI, USA, 2024, pp. 1–24. doi:10.1145/3613904.3642790.

[23] A. M. Roffarello, L. D. Russis, Towards detecting and mitigating smartphone habits, in: Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers (UbiComp/ISWC '19 Adjunct), ACM, London, UK, 2019, pp. 149–152. doi:10.1145/3341162.3343770.

[24] A. Cuadra, O. Bankole, M. Sobolev, Planning habit: daily planning prompts with alexa, in: International Conference on Persuasive Technology, Springer, 2021, pp. 73–87.

[25] L.-M. Meinhardt, M. Elhaidary, M. Colley, M. Rietzler, J. O. Rixen, A. K. Purohit, E. Rukzio, Scrolling in the deep: Analysing contextual influences on intervention effectiveness during infinite scrolling on social media, in: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25), ACM, Yokohama, Japan, 2025, pp. 1–17. doi:10.1145/3706598.3713187.

[26] G. Kruger, N. Sachdeva, M. Sobolev, Synthetic data generation for screen time and app usage, arXiv preprint arXiv:2509.13892 (2025). doi:https://doi.org/10.48550/arXiv.2509.13892. arXiv:2509.13892.

[27] A. L. Trella, K. W. Zhang, I. Nahum-Shani, V. Shetty, F. Doshi-Velez, S. A. Murphy, Designing reinforcement learning algorithms for digital interventions: Pre-implementation guidelines, Algorithms 15 (2022) 255. doi:10.3390/a15080255.

[28] L. Guo, Y. Fu, X. Lin, X. Xu, Y.-J. Chang, A. Hiniker, What social media use do people regret? an analysis of 34k smartphone screenshots with multimodal LLM, in: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25), ACM, Yokohama, Japan, 2025, pp. 1–23. doi:10.1145/3706598.3713724.

[29] A. L. Trella, K. W. Zhang, I. Nahum-Shani, V. Shetty, F. Doshi-Velez, S. A. Murphy, Reward design for an online reinforcement learning algorithm supporting oral self-care, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 37, 2023, pp. 15724–15730. doi:10.1609/aaai.v37i13.26866.

[30] N. van Berkel, C. Luo, T. Anagnostopoulos, D. Ferreira, J. Goncalves, S. Hosio, V. Kostakos, A systematic assessment of smartphone usage gaps, in: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, ACM, New York, NY, USA, 2016, pp. 4711–4721. doi:10.1145/2858036.2858348.

[31] N. Sachdeva, G. Kruger, Smartpause app: Core logic and reinforcement learning environment for smartpause, 2026. URL: https://github.com/nikhil-sachdeva/SmartpauseApp.