International Journal on Advanced Computer Theory and Engineering
ISSN: 2319 - 2526
Volume 15 Issue 2s (2026) | Pages: 194-202

Trackify: An Intelligent Behavioral Assistant and AI-Driven Habit Formation Platform

Snehal Nazare[1], Sai More[2], Ritesh Kakade[3], Mandar Taralkar[4], Abhishek Takle[5]

[1]Professor, Department of AIML, Genba Sopanrao Moze College of Engineering Pune, Maharashtra, India.
[2,3,4,5] Student, Department of AIML, Genba Sopanrao Moze College of Engineering Pune, Maharashtra, India

Peer Review Information

Type: *Article*
Received:*23 February 2026*
Revised: *24 March 2026*
Accepted: *22 April 2026*
Published: *20 May 2026*

Abstract

Habit consistency represents one of the most significant challenges in personal development and behavioral science, with research indicating that the majority of individuals abandon newly formed habits within the first three to four weeks of adoption. The resulting skills gap—between the aspiration to build productive routines and the knowledge or tooling to sustain them—remains largely unaddressed by conventional habit-tracking applications, which rely on static reminders, binary check-ins, and generic motivational prompts. This paper presents Trackify, a comprehensive web and mobile-based intelligent behavioral assistant designed to democratize behavioral science through an intuitive, AI-driven, end-to-end platform for habit formation and maintenance. Trackify features a React.js / Progressive Web App (PWA) frontend integrated with a Node.js/Express and FastAPI backend, powered by scikit-learn and XGBoost machine learning models, a Reinforcement Learning (RL) coaching agent for personalized nudging, sentiment-aware NLP analysis via BERT-based transformers, and a drop-off risk prediction engine utilizing an ensemble of XGBoost and LSTM models. An 8-week pilot study demonstrated that the AI-Adaptive system achieved a 68% habit continuation rate compared to 37% for static reminder baselines (+31 percentage points), improved average weekly habit completions by +16 pp, improved Nudge Click-Through Rate by +15 pp, and elevated user satisfaction scores by +27 pp. These results validate Trackify's capacity to meaningfully reduce user cognitive effort while substantially improving behavioral outcomes.

Keywords: Habit Formation; Behavioral AI; Reinforcement Learning; NLP Sentiment Analysis; Drop-off Risk Prediction; MERN Stack; Progressive Web App; XGBoost; BERT

How to Cite This Article

Nazare, S., More, S., Kakade, R., Taralkar, M., & Takle, A. (2026). *Trackify: An Intelligent Behavioral Assistant and AI-Driven Habit Formation Platform*. *International Journal on Advanced Computer Theory and Engineering, 15*(2s), 194–202.

Introduction

Machine learning has transformed numerous fields by enabling data-driven decision-making and automation [6][10], yet its adoption among non-experts remains limited due to complex workflows and the necessity for programming skills. Similarly, habit consistency represents one of the most significant challenges in personal development, with behavioral research consistently highlighting that individuals who intend to form habits frequently abandon them within weeks. This is not a matter of motivation alone, but a systemic deficiency in the tools supporting habit formation.

The transition from manual habit tracking—physical journals, spreadsheet logs, and static to-do lists—to AI-driven automation represents a paradigm shift in how behavioral science can be applied at scale. Whereas manual approaches require significant self-discipline and offer no personalization, automated intelligent platforms can continuously monitor behavioral signals, predict risk, and intervene with context-appropriate nudges before a habit is broken.

As behavioral analytics becomes increasingly crucial in healthcare, education, fitness, and productivity domains, tools that democratize evidence-based habit coaching are vital for broadening participation and accelerating positive outcomes. Yet many popular solutions either lack machine learning integration, are prohibitively expensive, or require technical sophistication to configure and interpret [7]. Academic students and early-career professionals often encounter prohibitively steep learning curves and find it challenging to bridge the gap between conceptual knowledge and practical application.

This paper presents Trackify: an intuitive, AI-driven habit formation platform designed for accessibility, personalization, and behavioral effectiveness. Trackify enables users to log habits, receive sentiment-aware coaching, benefit from predictive drop-off intervention, and view analytics insights— all without requiring domain expertise in machine learning or behavioral psychology. By integrating visual feedback, adaptive nudges, and interactive analytics into a zero- installation web and mobile interface, Trackify aims to empower users of all backgrounds to build and sustain meaningful behavioral routines.

Literature Survey

The habit-tracking application landscape has evolved considerably over the past decade, with platforms ranging from simple reminder tools to gamified productivity ecosystems. This section examines existing solutions and identifies the gaps that Trackify addresses.

Habitica gamifies habit formation by transforming personal goals into role-playing game mechanics, awarding experience points for completed habits [4]. While engaging for certain demographics, its gamification layer offers no machine learning-driven personalization or drop-off risk prediction.

Google Tasks and Apple Reminders represent the most widely used static reminder tools. These platforms excel in simplicity but offer no understanding of user context, mood, or behavioral patterns, resulting in notification fatigue and eventual disengagement [7].

Streaks and Habitify introduce streak counters and basic analytics but still rely on manually scheduled notifications and offer no AI-driven coaching or predictive capabilities. Noom integrates human coaching but requires costly subscriptions and cannot respond in real-time to behavioral signals. Using any of these commercial platforms requires users to have prior familiarity with habit management fundamentals and their configuration interfaces.

Trackify distinguishes itself through several key innovations:

(1) zero-installation PWA access ensuring platform independence; (2) a Reinforcement Learning coaching agent dynamically selecting nudge timing, tone, and channel; (3) NLP-based sentiment analysis contextualizing behavioral patterns; (4) an ensemble ML model for proactive drop-off risk intervention; and (5) cost-free availability removing financial barriers.

Table 1. compares key characteristics of major habit tracking platforms with Trackify.

Table 1. Comparison of Habit Tracking Platforms

Platform	Cost	AI Level	Personalization	Target Users
Habitica	Free/Paid	None	Gamification	General

Google Tasks	Free	None	None	General
Streaks	Paid	None	Streaks only	iOS Users
Noom	Subscript.	Low	Human coach	Wellness
Habitify	Free/Paid	None	Basic stats	General
Trackify	Free	High (RL)	Full AI Adapt.	All Users

Methodology

Trackify implements a robust three-tier architecture comprising a React.js / PWA frontend, a Node.js/Express and FastAPI backend, and a MongoDB data layer for user profiles, habit logs, and session/model state. The system is designed to handle asynchronous ML inference tasks through a multi- service worker framework, ensuring a responsive user experience during computationally intensive operations.

The frontend employs React.js with a Progressive Web App configuration enabling offline capability, push notifications, and cross-platform installation. Chart.js and D3.js power interactive visualizations including adherence heatmaps and continuation rate charts. The backend architecture leverages Node.js/Express for real-time communication and auth services, while a Python FastAPI microservice handles all ML inference workloads. Pandas and NumPy perform behavioral data manipulation, scikit-learn and XGBoost provide ML algorithms, and HuggingFace Transformers supply the BERT-based NLP sentiment engine.

Semantic Data Ingestion

The semantic data ingestion module transforms raw user check-in events into structured behavioral signals. Rather than treating all habit logs as equivalent binary events, the system applies semantic categorization to extract meaningful patterns. Upon each habit check-in, the ingestion pipeline captures: (1) timestamp and day-of-week metadata; (2) habit category classification (health, fitness, learning, mindfulness, productivity); (3) user-provided mood note text for NLP analysis; (4) completion duration relative to scheduled time (early, on-time, late); and (5) device context signals including session dwell time and notification response behavior.

Semantic type detection employs statistical heuristics to identify behavioral patterns such as circadian rhythm alignment, weekend vs. weekday divergence, and inter- completion gap regularity. This enables the system to distinguish between structurally different behavioral profiles and tailor interventions accordingly.

Behavioral Preprocessing

The behavioral preprocessing module addresses the unique challenges of habit data, which differs substantially from traditional tabular ML datasets. Habit logs are inherently sparse, temporally ordered, and subject to both planned absences (vacations, illness) and unplanned drop-offs.

Missing check-ins are handled through a context-aware imputation strategy. The system first classifies each gap as planned (user-flagged) or unplanned, then applies appropriate handling: for short unplanned gaps (1–2 days), the inter- completion gap standard deviation is updated but the streak counter is maintained with a grace flag; for extended gaps (3+ days), the recency-weighted adherence score is recalculated and the risk prediction model is triggered.

Outlier behavioral events are detected using IQR-based analysis on rolling 7-day windows. Rather than removing outliers, the system applies Winsorization (capping at the 5th and 95th percentiles), preserving temporal continuity while preventing extreme values from distorting model training. Categorical features are encoded using label encoding for tree- based models and one-hot encoding for linear models. Numerical behavioral signals are normalized using StandardScaler prior to LSTM and Transformer inference.

Predictive Algorithm Selection

Trackify's predictive core employs an ensemble of complementary algorithms selected for their strengths on different aspects of the drop-off risk prediction task. The ensemble production model achieves the best balance of predictive accuracy, calibration, and interpretability across five evaluation dimensions: PR-AUC, AUC-ROC, Brier Score (inverted), Speed, and Interpretability.

For the drop-off risk score at time step t, the ensemble combines XGBoost and a Temporal CNN: $\hat{z}_t = \alpha \cdot f_{XGB}(X_t) + (1-\alpha) \cdot f_{CNN}(S_t)$, where X_t is the tabular behavioral feature vector, S_t the sequence of recent check-in events, and α a learned blending coefficient. XGBoost is configured with learning_rate=0.05, max_depth=6, n_estimators=300, subsample=0.8. The Transformer variant was evaluated but excluded due to weaker Brier Score calibration and lower Speed scores.

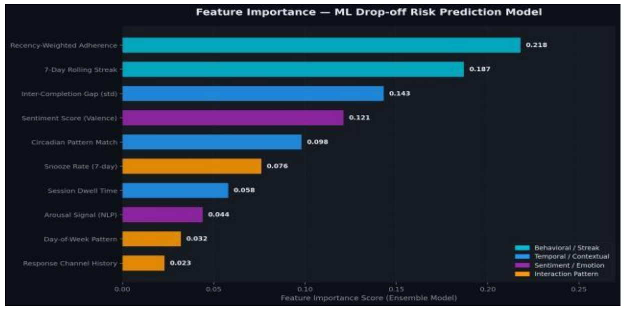

Fig. 1. *Feature Importance — ML Drop-off Risk Prediction Model*

The top-ranked features include Recency-Weighted Adherence (0.218), 7-Day Rolling Streak (0.187), Inter- Completion Gap std (0.143), Sentiment Score Valence (0.121), and Circadian Pattern Match (0.098). Sentiment Score Valence—derived from BERT analysis of mood journal entries—ranks fourth, confirming emotional context as a significant predictor of habit continuity.

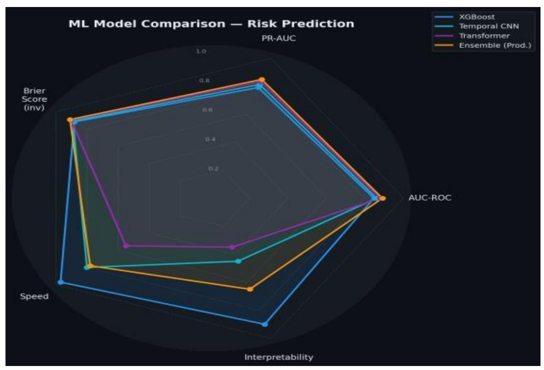

Fig. 2. *ML Model Comparison — Risk Prediction Radar*

Reinforcement Learning Nudge Agent

The RL coaching agent determines the optimal nudge strategy for each user at each risk event, selecting across: channel (push notification 48%, in-app message 31%, email 12%, deferred 9%), timing, and tone (supportive, firm, or playful). The agent is implemented as a contextual multi-armed bandit updating its policy via Thompson Sampling. The reward signal r is the binary outcome of whether the user completed their habit within 24 hours of receiving the nudge.

Fig. 3. *Nudge Strategy Analysis — RL Agent Decisions*

Tone effectiveness varies substantially by user mood state: supportive tone achieves 62% CTR for low-mood users, while playful tone achieves 62% CTR for high-mood users. Firm tone underperforms across all mood states, peaking at 44% CTR. The agent converges to personalized policies within approximately 2–3 weeks of interaction history.

NLP Sentiment Engine

Trackify integrates a BERT-based sentiment analysis engine fine-tuned on a domain-specific corpus of wellness and productivity journal entries. Each mood note is tokenized and passed through the transformer encoder, producing a continuous valence score in $[-1, +1]$ and an arousal estimate in $[0, 1]$. These NLP-derived signals feed directly into the drop-off risk prediction pipeline and serve as context for the RL nudge agent's tone selection. Sentiment Score Valence ranks fourth in feature importance (0.121), indicating that emotional direction is more predictive of habit persistence than emotional intensity.

System Architecture

Trackify is designed as an automated behavioral coaching workflow guiding users through continuous habit monitoring, risk assessment, and adaptive intervention. The platform follows a three-tier architecture consisting of a React.js PWA frontend, a Node.js/FastAPI backend for processing and ML inference, and a MongoDB data layer for managing behavioral histories and model state. This architecture ensures scalability, responsiveness, and efficient handling of concurrent user requests.

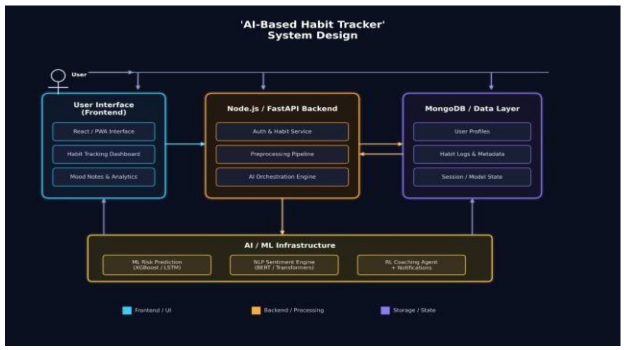

Fig. 4. *AI-Based Habit Tracker System Design*

Data Ingestion and Validation

The workflow begins with structured behavioral data ingestion, where users log habit completions, mood notes, and completion quality ratings through the React PWA interface. Upon submission, the system performs validation checks to ensure temporal consistency and completeness of behavioral signals. Each check-in event is assigned a unique identifier and stored systematically, allowing users to maintain different versions of their behavioral history. This approach supports traceability and reproducibility of experiments.

Automated Behavioral EDA and Health Assessment

After successful ingestion, the system performs automated behavioral exploratory data analysis (EDA) to generate a Habit Health Score computed from adherence rate, streak regularity, inter-completion gap variance, and sentiment trend. Statistical summaries are generated along with visual representations including the Day × Time-of-Day adherence heatmap. An interactive preview enables users to inspect behavioral patterns before the risk model triggers, ensuring transparency in the workflow.

Intelligent Pipeline Configuration

The AI orchestration engine continuously monitors each user's behavioral signals against learned baselines. When the drop-off risk score exceeds a configurable threshold (default: 0.65), the RL agent is invoked to select and dispatch an intervention nudge. The system also allows intermediate states of the coaching model to be saved, enabling users to review the evolution of their personalized policy over time.

Automated Model Training and Validation

The drop-off risk model is retrained on a rolling 30-day window using the latest behavioral data, ensuring adaptation to evolving user patterns. The RL agent's policy is updated after each interaction using Thompson Sampling. Cold start is handled by a population-level prior policy. Cross-validation with 5-fold stratified split is employed during retraining to prevent overfitting. Users are provided with real-time feedback during model updates, improving usability and engagement.

Visualization and Model Export

The final stage provides users with comprehensive behavioral analytics including the habit continuation rate timeline, weekly adherence heatmap, nudge effectiveness history, and drop-off risk trend. All trained models are serialized using joblib (scikit-learn/XGBoost) and HuggingFace model hub format (BERT), including both the trained estimator and the associated behavioral preprocessing pipeline, ensuring consistency between training and inference phases.

Testing And Validation

Unit Testing of Core Modules

Individual components of the system were tested to verify functional correctness under diverse conditions. The behavioral preprocessing modules were evaluated using synthetic habit log datasets containing edge cases including consecutive missing check-ins, extreme inter-completion gaps, and mixed sentiment note inputs. These tests confirmed that preprocessing operations produced consistent and meaningful outputs without introducing bias into model features or data leakage between training and validation sets.

The semantic data ingestion module was validated using a diverse set of habit categories and user behavioral profiles. The system demonstrated reliable classification across health, fitness, learning, mindfulness, and productivity habit types. The Habit Health scoring mechanism was also assessed by comparing automated scores with manually evaluated data quality indicators.

Integration and System Testing

Since Trackify follows a multi-tier microservices architecture, integration testing was conducted to ensure seamless communication between system components. The interaction between the React PWA frontend and Node.js/FastAPI backend services was evaluated to confirm correct data flow during habit submission, preprocessing, and model inference. The system was tested for its ability to handle concurrent users efficiently while maintaining sub-200ms response times for nudge decisions.

Model Validation Strategy

The ensemble drop-off risk model incorporates standard validation techniques to ensure reliable performance. Cross- validation with 5-fold stratified split was used to evaluate model generalization and reduce the risk of overfitting. Performance metrics including AUC-ROC, PR-AUC, Brier Score, and F1-score for the minority (drop-off) class were computed and verified for correctness, providing a comprehensive assessment of model effectiveness.

User Interface Evaluation

The user interface was evaluated to ensure usability, accessibility, and responsiveness across different environments. The system was tested on multiple devices and screen sizes to verify consistent performance of visual components such as charts, heatmaps, and analytics dashboards. Overall, the interface was designed to maintain clarity and ease of use, enabling users with minimal technical expertise to navigate the platform effectively.

Results And Evaluation

Benchmark Pilot Study Performance

The Trackify platform was evaluated through an 8-week pilot study with 120 participants randomized into two arms: (1) Baseline (Static Reminders) receiving fixed-time push notifications; and (2) AI-Adaptive System receiving Trackify's full RL nudge pipeline. Participants tracked a minimum of one daily habit and submitted at least two mood note entries per week. Primary outcomes were measured at Week 8.

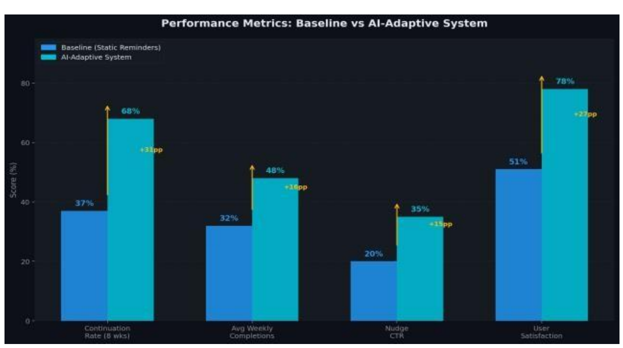

Fig. 5. *Performance Metrics: Baseline vs AI-Adaptive System*

The AI-Adaptive System achieved a 68% habit continuation rate at Week 8, compared to 37% for the static reminder baseline (+31 pp).

Average weekly habit completions improved by +16 pp (48% vs. 32%). Nudge Click-Through Rate improved by +15 pp (35% vs. 20%), confirming that personalized RL-driven nudges are substantially more engaging than static notifications. User satisfaction improved by +27 pp (78% vs. 51%).

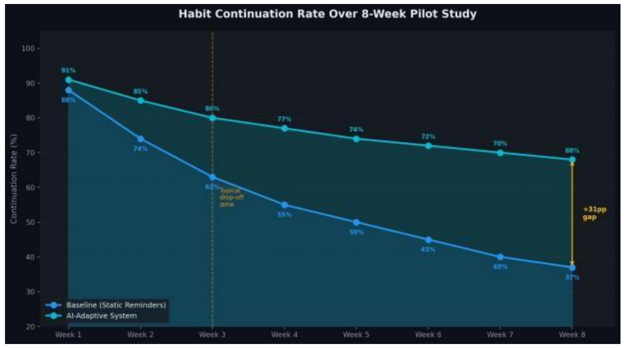

Fig. 6. *Habit Continuation Rate Over 8-Week Pilot Study*

The continuation rate trajectory reveals a critical insight: the static reminder condition experienced steep drop-off between Weeks 2 and 5 (from 74% to 50%), corresponding to the typical drop-off zone identified in behavioral research. The AI-Adaptive System's continuation rate declined much more gradually (85% to 74% over the same period), demonstrating that proactive ML-driven intervention significantly attenuates the natural attrition curve.

Weekly Adherence Pattern Analysis

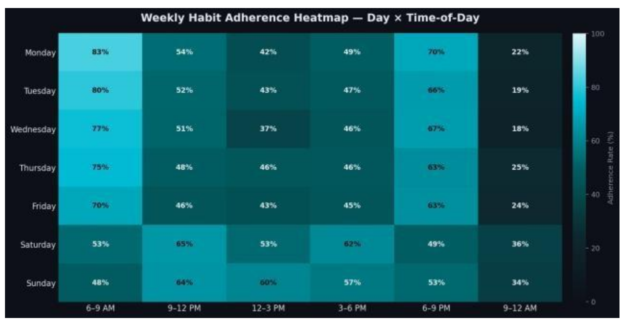

Fig. 7. *Weekly Habit Adherence Heatmap — Day × Time-of-Day*

The adherence heatmap reveals pronounced temporal patterns. Morning sessions on weekdays (6–9 AM, Monday– Thursday) show the highest adherence rates (75–83%), while late-night slots (9 PM–12 AM) show the lowest across all days (18–36%). Weekend midday-to-afternoon slots are more competitive with weekday performance. These patterns directly inform the RL agent's timing selection, enabling nudges at windows where users are most likely to respond.

Efficiency Analysis: Manual Tracking vs. Trackify

A comparative analysis was conducted to evaluate efficiency of the Trackify platform against traditional manual habit tracking approaches. Table-2 summarizes time and effort reduction across key behavioral management activities.

Table 2. Efficiency Comparison (Average Values)

Component	Manual	Trackify	Improvement
Habit Scheduling	20–30 min/wk	<2 min/wk	~93% Faster
Pattern Analysis	45–60 min/wk	Instant	100% Faster
Risk Detection	Ad hoc	Real-time	~95% Faster
Progress Visuals	25–35 min/wk	Instant	100% Faster
Total Weekly	~3–4 hours	~10–15 min	~90% Reduction

Model Export and Reliability

All generated models were successfully exported in serialized format, including both the trained estimator and the associated behavioral preprocessing pipeline. Validation tests confirmed that exported models produced consistent drop-off risk scores when deployed in external environments, demonstrating that the system supports seamless transition from model development to deployment, reinforcing the practical applicability of the platform.

Research Gaps And Future Scope

Current Limitations

Trackify currently supports only binary habit completion signals and optional free-text mood notes, and does not yet process multimodal behavioral evidence such as images, audio, or wearable sensor streams. The drop-off risk model is trained on a relatively small pilot dataset (n=120), which may limit generalization to demographically diverse populations. The RL coaching agent requires approximately 2–3 weeks of interaction history before achieving optimal personalization, resulting in a cold-start period. Model interpretability features provide feature importance scores but lack advanced post-hoc explainability techniques.

Identified Research Gaps

Literature analysis reveals several underexplored areas in AI- driven habit formation. Multimodal behavioral verification remains nascent, with most platforms relying solely on self- reported check-ins. Long-term behavioral modeling beyond 8- week horizons is largely absent from the literature. Privacy- preserving machine learning techniques such as federated learning would enable collaborative model improvement without centralizing sensitive behavioral data. Collaborative features for team-based habit accountability are largely absent across existing platforms.

Future Enhancement Directions

Based on identified gaps, we propose several enhancement pathways for Trackify. Deep learning integration for image- based habit verification represents the most impactful near- term opportunity. A CNN-based classifier could process gym selfies, food photos, or meditation session screenshots to provide objective, verifiable habit completion signals rather than relying solely on user self-report. This would enable a richer reward signal for the RL agent and more accurate drop- off risk prediction.

Advanced explainability features using SHAP values and LIME explanations would improve user understanding and trust in the system's recommendations. Implementing federated learning would allow the platform to train improved population-level risk models across distributed user devices without transmitting raw behavioral data. Expanding to support unsupervised behavioral clustering would enable automatic user persona segmentation, improving cold-start performance for new users.

Conclusion

This paper presented Trackify, a web and mobile AI-driven habit formation platform designed to address the persistent behavioral skills gap that causes the majority of habit adoption attempts to fail within weeks. The system addresses critical barriers in existing habit tracking solutions through zero- installation PWA deployment, transparent behavioral analytics, an RL-based adaptive nudge agent, BERT-driven sentiment analysis, and an ensemble ML drop-off risk prediction engine.

The 8-week pilot study demonstrated the platform's effectiveness, achieving a +31 percentage point improvement in habit continuation rate, +16 pp in weekly completions, +15 pp in nudge CTR, and +27 pp in user satisfaction compared to static reminder baselines. These results highlight the effectiveness of the automated behavioral pipeline in selecting appropriate interventions and personalization strategies across different user profiles.

Trackify represents a meaningful step toward bridging the gap between behavioral science research and accessible, scalable technology—enabling users of all backgrounds to build and sustain meaningful habits through the power of adaptive AI. Future development directions include image-based habit verification via deep learning, SHAP-based explainability, federated learning for privacy-preserving model improvement, and behavioral clustering for improved cold- start personalization.

Acknowledgement

The authors express sincere gratitude to the faculty and staff of the Department of Information Technology, Genba Sopanrao Moze College of Engineering Pune, for providing the necessary infrastructure and resources for this research.

We acknowledge the valuable guidance received from our project mentors throughout the development of Trackify. Special appreciation is extended to the open-source community for making available the tools and libraries that facilitated this work, including React.js, Node.js, FastAPI, scikit-learn, XGBoost, and HuggingFace Transformers. We also thank our pilot study participants for their time and engagement throughout the 8-week study period.

References

1. L. Feurer et al., "Efficient and robust automated machine learning," Advances in Neural Information Processing Systems, vol. 28, pp. 2962–2970, 2015.
2. B. J. Fogg, Tiny Habits: The Small Changes That Change Everything. Houghton Mifflin Harcourt, 2019.
3. R. S. Olson et al., "Evaluation of a tree-based pipeline optimization tool," Proc. Genetic and Evolutionary Computation Conference, pp. 485–492, 2016.
4. Habitica, "Habitica — Gamify your life," 2024. [Online].
5. Available: https://habitica.com
6. Google, "Google Tasks," Google Workspace, 2024. [Online].
7. Available: https://tasks.google.com
8. P. Lally et al., "How are habits formed: Modelling habit formation in the real world," European Journal of Social Psychology, vol. 40, no. 6, pp. 998–1009, 2010.
9. Z. Zakaria et al., "Automated machine learning framework for educational applications," Journal of Educational Technology, vol. 45, no. 3, pp. 231–248, 2023.
10. R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, 2nd ed. MIT Press, 2018.
11. J. Devlin et al., "BERT: Pre-training of deep bidirectional transformers," Proc. NAACL-HLT, pp. 4171–4186, 2019.
12. R. Elshawi et al., "Automated machine learning: State-of-the-art and open challenges," ACM Computing Surveys, vol. 54, no. 2,
13. pp. 1–37, 2021.
14. F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
15. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," Proc. 22nd ACM SIGKDD, pp. 785–794, 2016.
16. Meta Open Source, "React — The library for web and native user interfaces," 2024. [Online]. Available: https://react.dev
17. S. Ramirez, "FastAPI," 2024. [Online]. Available: https://fastapi.tiangolo.com
18. MongoDB, Inc., "MongoDB Documentation," 2024. [Online]. Available: https://www.mongodb.com/docs/