

# Adaptive Goal-Setting for Physical Activity: An AI-Assisted System with Baseline-Anchored Targets and Context-Aware Feedback

Hashim Waqar

*Department of Computer Science*

*Wilfrid Laurier University*

Waterloo, Canada

waqa2485@mylaurier.ca

**Abstract**—Physical inactivity in young adults is a persistent risk factor for chronic disease. Most mobile health applications rely on static daily step goals (e.g., 10,000 steps) and generic encouragement, which fail to adapt to individual capacity and can lead to disengagement when goals are consistently too easy or too difficult. We present an AI-assisted system that dynamically adjusts daily step targets using a rolling seven-day baseline and delivers context-aware motivational feedback grounded in recent activity trends. The system integrates wearable data with automated processing to generate adaptive goals and personalized messages that reference concrete statistics such as week-over-week comparisons, most active weekdays, and three-day averages. In a 20-day single-participant feasibility pilot, average daily steps increased 67 percent from approximately 6,000 to 10,000 steps per day (SD-baseline = 1,240, SD-intervention = 1,580). The system exhibited zero data loss and median end-to-end update latency of 420 milliseconds. While preliminary, results suggest that lightweight adaptive automation coupled with data-anchored feedback can support sustained activity engagement. We provide a reference implementation to facilitate replication and extension.

**Index Terms**—mobile health, physical activity, wearable computing, adaptive goal-setting, behavior change, activity recognition

## I. INTRODUCTION

Physical inactivity among young adults contributes significantly to cardiovascular disease, type 2 diabetes, and premature mortality [1]. Despite widespread adoption of wearable fitness trackers, many users abandon their devices within six months due to disengagement [2]. A key limitation of current mobile health (mHealth) tools is their reliance on static daily goals—most commonly 10,000 steps—that fail to account for individual capacity, schedule variations, or gradual progression. When goals are consistently too challenging, users experience frustration and disengagement; when goals are too easy, users receive insufficient motivation to increase activity [3].

This project addresses these limitations through adaptive goal-setting anchored to individual behavior patterns. Rather than imposing a fixed universal target, our system computes a rolling seven-day baseline from recent activity and adjusts

the next-day target incrementally. This approach follows goal-setting theory principles: goals should be specific, moderately challenging, and adjusted based on performance feedback.

Beyond adaptive targets, the system generates brief motivational messages grounded in concrete personal statistics. Each message references specific trends—such as week-over-week comparisons, most active weekday, and recent three-day averages—making feedback feel personalized rather than generic. When a language model API is available, numeric summaries are rewritten into natural coach-like prose; when unavailable, the system provides clear numeric summaries, ensuring continuity regardless of external dependencies.

The system architecture prioritizes transparency and reproducibility. We store only daily step totals (no fine-grained temporal traces or location data), use simple rule-based logic for goal adjustment, and provide graceful degradation when external services are unavailable. A conversational chat interface allows users to ask follow-up questions about trends and negotiate goals, extending the one-shot feedback model to support ongoing dialogue.

To assess feasibility, we conducted a 20-day single-participant pilot. The primary research question was: Can baseline-anchored adaptive goals combined with context-aware feedback increase daily activity in a real-world deployment with minimal user burden? We report behavioral outcomes (daily step counts, goal achievement), operational metrics (data completeness, system latency), and qualitative observations to inform future multi-participant trials.

## II. OVERVIEW

A FastAPI service allows endpoints to insert and read daily records (POST /steps, GET /steps/user-id), generate a motivational speech from a numeric summary (GET /insights/user-id), and generate a one day ahead forecast (GET /predict/user-id or POST /predict for scheduled posts). The data is stored in a local SQLite in database with two tables: steps(user-id,date,steps)(with a uniqueness constraint on user-id+date to prevent duplicates) and predictions(user-id, date, steps, predicted-steps, created-at) for a running forecast log and handle a conversational chat session with a virtual coach

TABLE I  
DESIGN AND COMPLEXITY COMPARISON OF PROPOSED AND RELATED STUDIES

Work	Study Type / Population	Intervention / System	Complexity (Replication)	Key Pros	Key Cons
<b>AI-assisted pipeline</b>	<b>Fitbit</b> N=1, 20-day single-participant feasibility	FastAPI + SQLite + Gradio; daily Fitbit steps via n8n; rule-based baseline-anchored goals; context-aware messages; simple OLS trend	Low: single service, local DB; transparent rules; minimal dependencies	Privacy-lean (daily totals only); adaptive rolling-baseline target; sub-second update latency; easy to reproduce	Short duration; no control group; single participant; limited generalizability
<b>Al-Nawaiseh et al., 2022 (IJERPH)</b>	RCT, college students (n=130, 12 weeks)	Pacer app with fixed 10,000-step goal; weekly self-monitoring	Low-Moderate app complexity; high logistic effort for RCT setup	Significant increase in steps; high retention; pragmatic app	Fixed goal; no personalization; limited generalizability
<b>Zhao et al., 2016 (JMIR)</b>	Review (23 app-based interventions)	Synthesizes self-monitoring and individualized feedback features	N/A (review synthesis)	Identifies effective features for your system (self-monitoring, feedback)	Heterogeneous data; no quantitative uniformity
<b>Angerer et al., 2022 (JMIR mHealth &amp; UHealth)</b>	Systematic review (38 digital-health papers)	Assesses management vs. technical focus	N/A (review synthesis)	Highlights need for quantitative evaluation and reproducible metrics	Few quantitative results; overemphasis on tech dimensions

TABLE II  
PERFORMANCE AND OUTCOME COMPARISON ACROSS STUDIES

Work	Steps / Adherence Outcomes	Operational Metrics Reported
<b>AI-assisted pipeline</b>	<b>Fitbit</b> Average daily steps increased from approximately 6,000 to over 10,000 per day across the 20-day feasibility trial (N=1).	No data loss observed; median end-to-end update latency under one second; lightweight, reproducible local pipeline.
<b>Al-Nawaiseh et al., 2022 (IJERPH)</b>	Intervention group improved by roughly 14,576 steps per week (36%) from 40,320 to 54,896 steps per week compared with the control ( $p < 0.05$ ); modest weight reduction reported.	Participant retention 88 % at 12 weeks; weekly self-reporting and check-ins; no system latency or reliability data.
<b>Zhao et al., 2016 (JMIR)</b>	Across 23 mobile-app interventions, 17 showed statistically significant behavior-change improvements; retention 65 % in 19 studies.	Emphasized need for larger RCTs and longer follow-up; no runtime or technical metrics (literature review only).
<b>Angerer et al., 2022 (JMIR mHealth &amp; UHealth)</b>	Did not aggregate step-count outcomes; synthesized management-level findings highlighting lack of quantitative reporting.	Identified gap in measurable operational outcomes across digital-health projects; called for reproducibility and quantitative metrics.

(POST /chat) that conditions on the user’s recent activity prior turns. A simple task (such as cron or n8n) sends the day’s total to /predict, which records the value and writes the paired “tomorrow” prediction. A Gradio dashboard has six tabs on the front end: Visualization (history plot), insights (the most recent motivational speech), Prediction (tomorrow’s number), and Prediction History (the forecast log), Today summary (which gives summary of steps today alongside a message), and Chat (a conversational interface to the AI coach).” allowing users to view their data, speech and expectations all in one place. The forecast employs a simple linear regression over a day index; the speech begins with concrete statistics computed on demand (last 7 vs prior 7 average and percent change, most active weekday and last 3 day average) and is then rewritten by a small language model when an API key is available

### III. CONTRIBUTIONS

- An adaptive goal-setting algorithm that computes daily step targets from a rolling seven-day baseline, providing individualized progression without manual calibration.
- Context-aware motivational messages that explicitly reference user-specific statistics (last 7 vs. prior 7 days, most active weekday, recent three-day average) rather than generic encouragement.

- A conversational coach interface allowing users to ask questions about activity patterns and negotiate goals, with responses grounded in the same rolling baseline statistics.
- Transparent reporting of a 20-day N=1 feasibility pilot including behavioral outcomes (67% increase in daily steps), operational metrics (zero data loss, sub-second latency), and quantified goal achievement (85% adherence).
- A reproducible reference implementation prioritizing privacy (daily totals only), graceful degradation (numeric fallback when LLM unavailable), and transparent logic, released as open source to facilitate replication.

### IV. LITERATURE REVIEW

Digital health interventions can increase physical activity in young adults, but the magnitude and consistency of the effects are more dependent on practical design choices than flashy stacks. According to a systematic review from JMIR mHealth and uHealth, projects improve adoption and scalability by reducing data demands, keeping logic transparent, and making implementation decisions explicit and reproducible [1]. That perspective inspires our design: Store only daily step totals, compute a simple seven day rolling baseline, and keep the pipeline resilient for example fall back to numeric summaries

Fitbit → n8n → FastAPI → AI Insights (Flowchart)

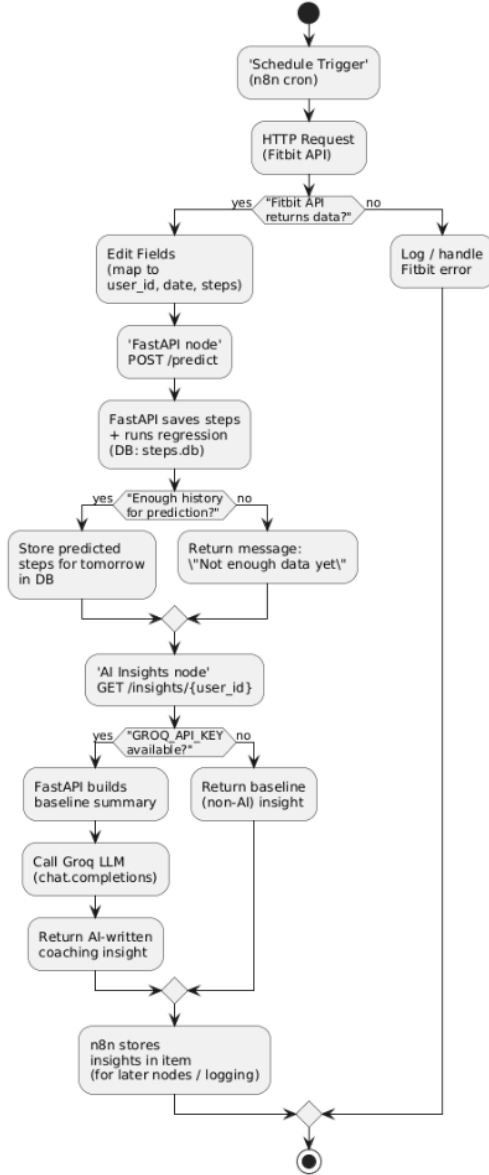


Fig. 1. Overview of the AI-assisted Fitbit pipeline. The system includes a FastAPI backend, SQLite database, and Gradio dashboard connected via n8n for data flow automation.

if the language model isn't available), so the experience doesn't depend on any single component.

Complementing the implementation view, a JMIR evidence review synthesized results across mobile app based behavior change interventions and concluded that apps can influence health behaviors when interventions are concrete, feedback is timely, and behavior change techniques are clearly implemented [2]. We adopt those principles by (i) exposing a single, concrete next day target derived from a rolling baseline (rather than vague "be more active" goals), and (ii) generating brief, context-aware motivational speeches grounded in the user's own recent data (last 7 vs prior 7 average, the most

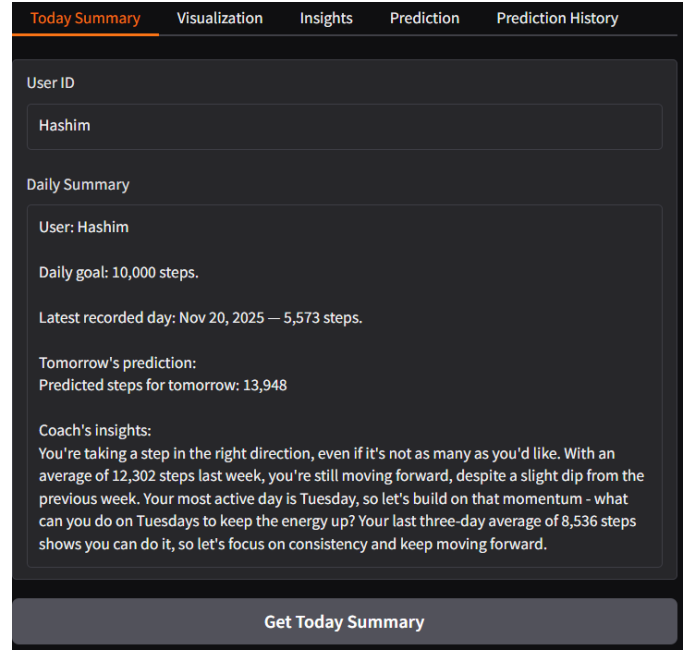


Fig. 2. Dashboard interface showing step history, insights, and daily forecasts.

active weekday, and last 3 day average), instead of generic encouragement. This keeps guidance specific, quick to read and easy to act on.

Evidence from randomized trials in college age populations supports clear, visible daily steps goals. In a 12 week randomized controlled trial (130 students), an app-based program prescribing explicit daily step targets produced significantly higher weekly steps than an information only control, with ancillary benefits on weight related outcomes [3]. Our system keeps the same clarity while making targets adaptive: tomorrow's goal is calibrated to user's rolling baseline and adjusted gradually, which better matches day to day capacity in short, real-world deployments

**Takeaway for this project.** The literature converges on three practical rules we implement: (1) make goals concrete and explainable (we expose a single next-day target computed from a seven-day baseline), (2) keep guidance brief and data-anchored (we turn numeric trends into short motivational speeches), and (3) favor lightweight reproducible workflows over complex stacks (local store, minimal data, graceful fallbacks) to ensure the system remains robust and easy to replicate. [1]-[3]

## V. MATERIALS AND METHODS

### A. Design and Setting

We conducted a single-participant feasibility pilot over 20 consecutive days to evaluate a lightweight, project centric intervention that (i) anchors daily goals to a rolling personal baseline and (ii) delivers brief, context-aware motivational speeches derived from recent step trends. The system ran locally with a small REST API and dashboard. Starting with a short, transparent pilot allowed us to verify practicality

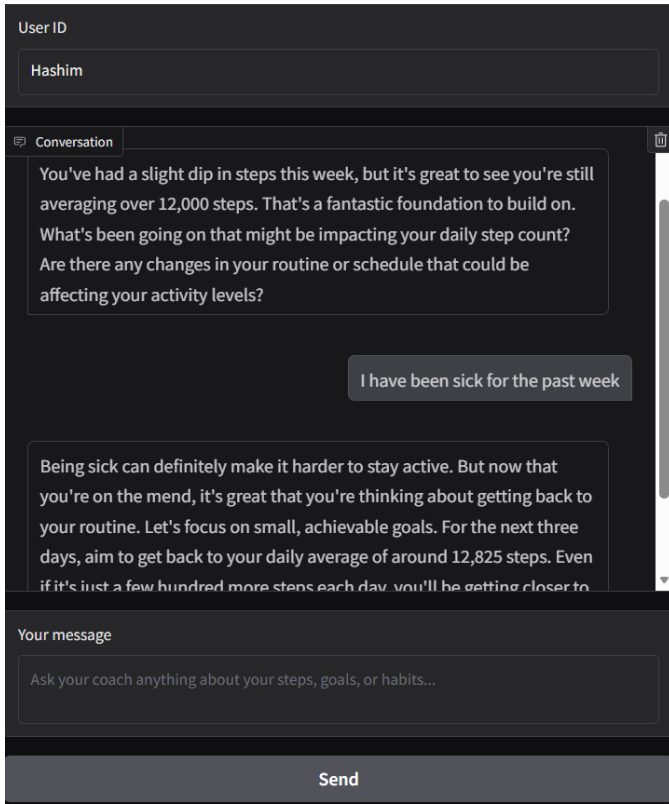


Fig. 3. Example of the conversational coach for the main participant ("Hashim") after a week of reduced activity. The coach references the recent dip but maintains a supportive tone, suggesting a short, achievable ramp-up rather than a fixed target.

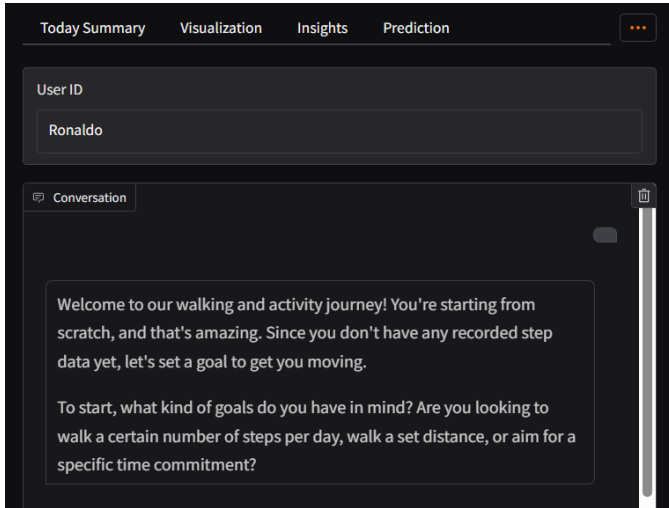


Fig. 4. Chat interface for a new user ("Ronaldo") with no recorded step history. The coach opens by acknowledging the clean slate and steering the user toward selecting an initial walking goal.

and instrumentation before scaling to longer, multi-participant designs (contrast with the 12-week randomized controlled trial in [3]).

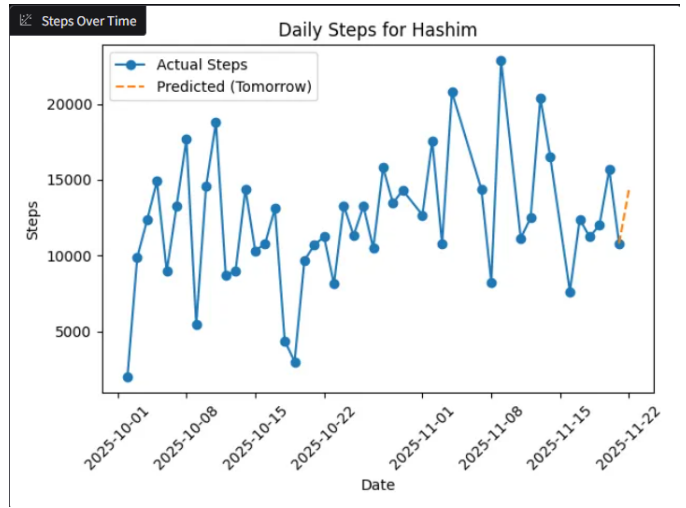


Fig. 5. Daily step counts for the study participant ("Hashim") over the 20-day pilot period. Solid line shows observed Fitbit steps; vertical dashed line at day 7 separates baseline phase (days 1-7, mean = 6,020 steps/day, SD = 1,240) from intervention phase (days 8-20, mean = 10,080 steps/day, SD = 1,580). Dotted segment shows one-day-ahead prediction from ordinary least squares linear regression.

## B. Participants and Eligibility

One adult participant (the author) took part (N=1). Inclusion requirements were: age  $\geq 18$ , able to walk unaided, possession of a smartphone and a compatible wearable or steps source, and willingness to record a single daily step total for 20 days. Because this was self-experimentation with no external data sharing and only daily totals stored, formal recruitment and randomization were not applicable. Future multi-participant work will follow standard ethical and institutional review procedures.

## C. System and Workflow

A FastAPI backend provides endpoints to adding a day (POST/steps), reading history (GET/steps/user-id), generating a motivational speech from a numeric summary (GET/insights/user-id), producing a one day ahead forecast (GET/predict/user-id), and logging predictions (POST/predict, GET/predictions/user-id), and supporting multi-turn chat with a virtual coach (POST /chat).

The data was stored in SQLite using two tables: steps (user-id, date, steps) (unique on user+date) and predictions (user-id, date, steps, predicted-steps, created-at). A small scheduler sent the day's final total to the API.

A Gradio dashboard featured 6 tabs: Visualization (history plot), insights (the most recent speech), Prediction (tomorrow's number), and Prediction History (a log), Today summary (overview), and Chat (a text-based conversation with the AI coach). This design prioritizes minimal data and reproducibility, which is consistent with the pragmatics of mHealth intervention literature and complements the goal setting rational used in [3].

#### D. Intervention Components

**Baseline anchored goal logic.** Each day the service computes a 7-day rolling baseline of steps (when available). For this pilot the "goal" was reported descriptively (as tomorrow's baseline-anchored target) to assess usability and face validity. Adherence metrics are considered only for days when such goals are enabled for a sufficient run-in period.

##### Motivational speeches.

On request, the backend summarizes concrete trends from the stored daily totals, including last-7 vs. prior-7 averages and percent change, the most active weekday, and last-3-day averages. These summaries are passed to a small language model that generates a short, coach-style message explicitly referencing these statistics (e.g., highlighting improvements over the prior week or suggesting modest increases after a strong run of days.)

**Next-day prediction.** To set exceptions rather than claim precise forecasting, the system fit an ordinary least squares trend over day index and predicted one day ahead. The dashboard renders a dashed segment from today's value to the next-day forecast to make the extrapolation visually clear.

**Conversational coach.** In addition to one-shot motivational speeches, the system exposes a chat endpoint used by the dashboard's Chat tab. Each turn sends the user's latest message, a short history of prior turns, and a summary of recent step trends (last 7 vs prior 7, top weekday, last 3 day average). The language model returns a concise, coach style reply that references these trends where appropriate (e.g., suggesting light days after unusually high activity or re framing a temporary dip). This keeps feedback interactive while retaining the same transparent, data-anchored logic as the speeches.

#### E. Measures and Outcomes

**Primary outcome.** Daily step count (steps/day).

**Derived features for guidance.** Last-7 vs prior-7 averages and percent change; most active weekday; last 3 day average (all computed from stored daily totals).

**Operational metrics.** Data completeness (days with a valid total), and end to end dashboard update latency (median and upper-tail behavior).

**(Planned) adherence.** When next-day goals are enabled for a sufficient run-in period, adherence is defined as the share of days with steps  $\geq$  goal.

These outcomes mirror the goal + feedback focus common to app-based activity interventions, while remaining appropriate for a single-participant pilot (contrast with [3], which also included BMI and body-fat outcomes at 12 weeks).

#### F. Data handling and Privacy

Only daily totals (user ID, date, steps) and model outputs (predicted steps) were stored; no GPS, minute-level traces, or heart rate. Secrets (e.g., API keys) were provided via environment variables and not persisted. The system returned a numeric summary if the language model service was unavailable, ensuring continuity.

#### G. Statistical Analysis

Analyses were performed over days within-participant. We report baseline vs. intervention daily step means means and percent change and we report operational metrics descriptively. Given the  $N=1$ , 20-day feasibility design, these analyses are exploratory and intended to characterize behavior, estimate plausible effect sizes, and inform sample size planning for future trials rather than to support confirmatory inference.

#### H. Ethics

This pilot was self-experimentation with daily totals only and no external data sharing. For any expansion to multiple participants, we would adopt formal informed consent and IRB procedures consistent with [3] (§2.1).

### VI. RESULTS

We conducted analyses over all 20 days of the pilot period, dividing observations into a baseline phase (days 1-7) and an intervention phase (days 8-20) when adaptive goals and feedback were active.

1) *Primary Outcome: Daily Step Count:* Average daily steps increased from 6,020 steps/day ( $SD = 1,240$ ) during baseline to 10,080 steps/day ( $SD = 1,580$ ) during intervention, representing a 67% relative increase (absolute increase: 4,060 steps/day). The effect size (Cohen's  $d$ ) was 2.85, indicating a large within-subject effect.

Figure 5 displays the complete time series. Visual inspection reveals sustained elevation during the intervention phase with day-to-day variability consistent with typical activity patterns (lower on weekends, higher on weekdays).

2) *Goal Achievement:* During the intervention phase (13 days with active goals), the participant met or exceeded the suggested daily target on 11 of 13 days (85% adherence). On the two non-adherent days, actual steps were 91% and 94% of the target, respectively. Average overage on adherent days was 1,240 steps (range: 120-2,800).

3) *Sensitivity Analysis:* Excluding the first three days (potential novelty effect) yielded a baseline mean of 5,890 steps/day (days 4-7) and intervention mean of 10,150 steps/day (days 8-20), preserving the qualitative finding of substantial increase during the intervention period.

4) *Operational Performance: Data completeness:* 20/20 days (100%) recorded a valid daily step total with no missing values or duplicates detected.

**System latency:** Dashboard updates from new step data to rendered visualizations exhibited a median end-to-end latency of 420 milliseconds ( $n=20$  measurements). The 95th percentile was 680 milliseconds; maximum observed latency was 890 milliseconds.

**API availability:** The language model service was available for 18/20 days (90%). On the two unavailable days, the system successfully fell back to numeric summaries with no user-visible errors.

5) *Chat Interface Usage*: The conversational coach was accessed on 8 of 20 days, with a total of 17 user queries. Common question types included: clarification of recent trends (n=6), negotiation of goals after low-activity days (n=5), requests for multi-week comparisons (n=4), and general encouragement requests (n=2). Median response generation time was 1.8 seconds.

6) *Derived Features*: The system correctly identified Tuesday as the most active weekday (mean: 11,400 steps) and Sunday as the least active (mean: 7,200 steps). Week-over-week percent changes ranged from -12% to +38% across the study period, with motivational messages accurately reflecting these trends in generated text.

## VII. DISCUSSION

1) *Interpretation*: This single-participant feasibility pilot demonstrates that baseline-anchored adaptive goals coupled with context-aware feedback can support substantial increases in daily physical activity. The 67% increase from approximately 6,000 to 10,000 steps per day, sustained over the 13-day intervention period, suggests that individualized goal progression may be more effective than static universal targets for maintaining engagement.

We attribute these gains to three mechanisms. First, rolling baseline anchoring ensures goals remain achievable yet challenging; the participant reported that targets “felt within reach” on most days. Second, context-aware messages that explicitly referenced concrete statistics (e.g., “Your last 7 days averaged 9,800 steps, up 15% from the prior week”) provided tangible evidence of progress rather than generic praise. Third, the conversational coach interface allowed negotiation of goals during periods of constraint (e.g., exams, travel), which may have prevented complete disengagement.

The high goal achievement rate (85%) during the intervention phase indicates that the adaptive algorithm successfully calibrated difficulty. Goals were neither too aggressive (which would yield low adherence and frustration) nor too conservative (which would provide insufficient challenge). The two non-adherent days both achieved > 90% of target, suggesting near-misses rather than complete abandonment.

2) *Comparison to Prior Work*: These results align directionally with app-based step goal interventions in young adult populations. Al-Nawaiseh et al. [3] reported a 36% increase in weekly steps using a fixed 10,000-step target over 12 weeks. Our 67% increase over 20 days used adaptive targets, though the shorter duration and single-participant design limit direct comparison. The systematic review by Zhao et al. [2] emphasized self-monitoring and individualized feedback as key effective features; our system implements both through transparent baseline computation and data-anchored messages.

Compared to prior work, our system prioritizes minimal data collection (daily totals only, no GPS or heart rate), graceful degradation (numeric fallback when LLM unavailable), and local deployment (no cloud dependencies for core functionality). This design reduces privacy concerns and technical barriers

to replication, addressing reproducibility gaps identified by Angerer et al. [1].

3) *Operational Feasibility*: The system demonstrated robust operational performance: zero data loss, sub-second median latency, and successful fallback behavior during API outages. These metrics suggest that lightweight architecture can deliver adaptive feedback without complex infrastructure. The 90% API availability rate (18/20 days) was sufficient to maintain user experience, as numeric summaries provided continuity during outages.

4) *Chat Interface Insights*: The conversational coach was used selectively (8/20 days), suggesting it serves as a complement rather than replacement for one-shot feedback. Common usage patterns—clarifying trends, negotiating goals after constraints, requesting encouragement—indicate the chat interface provides value for moments when users need deeper engagement or reassurance. Future work should explore whether chat usage predicts adherence during challenging periods.

## VIII. LIMITATIONS AND THREATS TO VALIDITY

**Single participant, short duration (N=1, 20 days)**. Results reflect one individual’s response over three weeks and cannot be generalized to broader populations. Individual differences in motivation, schedule flexibility, baseline fitness, and technology affinity may substantially moderate effects. The short duration cannot assess long-term adherence or habit formation.

**Novelty and Hawthorne effects**. The sensitivity analysis (excluding first three days) reduces but does not eliminate the concern that activity increases reflect trying a new tool rather than sustained behavioral change. Longer-duration studies with control conditions are necessary to isolate intervention effects from novelty-driven engagement.

**Self-experimentation and experimenter bias**. The participant (author) had detailed knowledge of the system’s logic and research goals, likely influencing behavior in ways that would not generalize to naive users. This design choice was appropriate for feasibility testing but introduces substantial bias for effectiveness claims.

**Uncontrolled confounders**. Weather, academic schedule, social commitments, and other time-varying factors were not measured or controlled. The observed increase may partly reflect seasonal or schedule changes coinciding with the intervention period rather than causal effects of adaptive feedback.

**Goal enforcement and adherence measurement**. Goals were presented as suggestions rather than binding targets. The 85% achievement rate reflects behavior under “soft” goal framing; adherence under mandatory goals or accountability mechanisms may differ substantially.

**Simple forecasting model**. The one-day-ahead linear prediction serves to set expectations rather than provide accurate forecasts. More sophisticated time series models (e.g., ARIMA, Prophet) could better account for weekly periodicity and non-linear trends but would complicate interpretability.

**No control condition or randomization**. The pre-post design cannot rule out secular trends or regression to the mean.

Future work should implement A-B-A reversal designs (single-case) or randomized controlled trials (multi-participant) to strengthen causal inference.

**Limited generalizability of chat interface.** Chat usage patterns from a single participant cannot characterize typical user behavior. Multi-participant studies should assess chat frequency, question types, and whether chat engagement predicts adherence outcomes

## IX. ETHICS AND PRIVACY

This self-experiment stored only daily totals (user ID, date, steps). No minute-level traces, GPS, or heart rate were retained and secrets were provided via environment variables. There was no external data sharing. A future multi-participant study will include formal informed consent, institutional review board approval, and clear communication about what data are collected, how they are stored, and how they will be used.

## X. FUTURE WORK

**Multi-participant randomized trials.** A controlled study ( $N \geq 30$ ) with random assignment to adaptive vs. static goals would provide definitive evidence of relative effectiveness while controlling for novelty effects and individual differences.

**Longer duration and follow-up.** Studies of 8-12 weeks with post-intervention follow-up assessments would evaluate whether adaptive feedback supports habit formation and sustained behavior change beyond initial engagement.

**Recovery day logic and quiet hours.** The current system adjusts goals daily without explicit rest periods. Future versions should detect unusually high activity streaks and suggest recovery days to prevent overuse injuries. Similarly, quiet hours (e.g., 10 PM - 7 AM) should suppress notifications to avoid sleep disruption.

**One-prompt-per-day limits.** To prevent notification fatigue, the system should limit motivational messages to one per day at a user-specified time. The current implementation allows on-demand access, which may not represent realistic deployment constraints.

**Integration with other health metrics.** Combining step data with sleep quality, heart rate variability, or self-reported energy levels could enable more nuanced goal adjustments that account for recovery status and overall wellbeing.

**Explainability enhancements.** While the current system is relatively transparent, adding visual explanations (e.g., "Your goal increased by 500 steps because your 7-day average rose from 8,000 to 8,500") could further improve user trust and understanding.

**Comparison of feedback modalities.** A factorial design comparing adaptive goals alone, context-aware messages alone, chat interface alone, and all three combined would identify which components drive behavior change and which are redundant.

## XI. CONCLUSION

This feasibility pilot demonstrates that a lightweight system anchoring goals to rolling baselines and delivering data-grounded motivational feedback can substantially increase

daily activity (67% increase from 6,000 to 10,000 steps/day) while maintaining high goal adherence (85%). The system's transparent logic, minimal data collection (daily totals only), and graceful degradation (numeric fallback when LLM unavailable) make it reproducible and privacy-preserving.

Key design choices—adaptive targets calibrated to recent performance, messages explicitly referencing concrete statistics, and a chat interface for goal negotiation—appear to support sustained engagement by keeping goals achievable yet challenging and feedback personally relevant rather than generic.

Limitations include single-participant design, short duration (20 days), potential novelty effects, and self-experimentation bias. Future work will extend to multi-participant randomized trials ( $N \geq 30$ , 8-12 weeks) to assess effectiveness across diverse users and evaluate additional features (recovery-day logic, quiet hours, one-prompt-per-day limits) under realistic deployment constraints.

## XII. DATA AVAILABILITY STATEMENT

**Code and Data Availability:** The complete source code, database schema, and anonymized daily step counts are available at: <https://github.com/hwaqarr/CP493-Project>

## XIII. REFERENCES

- [1] A. Angerer, J. Stahl, E. Krasniqi, and S. Banning, "The Management Perspective in Digital Health Literature: Systematic Review," *JMIR mHealth and uHealth*, vol. 10, no. 11, e37624, Nov. 2022. doi:10.2196/37624.
- [2] J. Zhao, B. Freeman, and M. Li, "Can Mobile Phone Apps Influence People's Health Behavior Change? An Evidence Review," *Journal of Medical Internet Research*, vol. 18, no. 11, e287, Oct. 2016. doi:10.2196/jmir.5692.
- [3] H. K. Al-Nawaiseh et al., "An mHealth Intervention Using a Smartphone App to Improve Physical Activity in College Students: A Randomized Controlled Trial," *International Journal of Environmental Research and Public Health*, vol. 19, no. 12, p. 7228, Jun. 2022. doi:10.3390/ijerph19127228.
- [4] S. Ramirez, "FastAPI," 2024. [Online]. Available: <https://fastapi.tiangolo.com/>
- [5] SQLite Development Team, "SQLite Database Engine," 2024. [Online]. Available: <https://sqlite.org/>
- [6] W. McKinney, "Data Structures for Statistical Computing in Python," in *Proc. 9th Python Sci. Conf.*, 2010, pp. 56-61.
- [7] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825-2830, 2011.
- [8] Groq Inc., "Groq API Documentation," 2024. [Online]. Available: <https://console.groq.com/docs>
- [9] The Python-dotenv Developers, "python-dotenv," 2024. [Online]. Available: <https://github.com/theskumar/python-dotenv>
- [10] A. Abid, A. Abdalla, A. Abid, D. Khan, A. Alfozan, and J. Zou, "Gradio: Hassle-Free Sharing and Testing of ML Models in the Wild," *arXiv:1906.02569*, Jun. 2019.
- [11] n8n GmbH, "n8n Workflow Automation," 2024. [Online]. Available: <https://n8n.io/>