ESP-JETA
ESP Journal of Engineering & Technology Advancements
ISSN: 2583-2646 / Volume 6 Issue 1 March 2026
Paper Id: JETA-V6I1P121 / Page No: 175-185

Review Article

AI-Driven Behavioural Interventions Integrating Cognitive Frameworks with Intelligent Systems

Rahamath Mohamed Razikh Ulla
Capitol Technology University, Maryland.

Received Date: 28 February 2026 Revised Date: 11 March 2026 Accepted Date: 31 March 2026

Abstract: *Artificial intelligence (AI) is becoming critical in improving behavioral interventions in health, sustainability, and digital well-being. This review proposes a theory-based framework combining the Theory of Planned Behavior (TPB) with recommendation-system architecture and principles of EAST behavioral design (Easy, Attractive, Social, Timely). TPB offers interpretable cognitive predictors—attitudes, subjective norms, and perceived behavioral control whereas collaborative filtering enables personalization at scale, and EAST limits intervention delivery to action-oriented behavioral optimization. Empirical studies by just-in-time adaptive interventions (JITAIs), conversational agents and energy-feedback systems indicate that they have strong short-term effects but continue to face persistent challenges related to durability, interpretability, and fairness. The intelligent recommendation architecture suggested is TPB-compliant and overcomes prediction-oriented models in favor of cognitively anchored adaptive controllable behavioral systems. The proposed future research directions are longitudinal construct modelling, fairness-conscious personalization and cross-domain scalability. It represents a promising step toward towards responsible and scalable behavior change technologies by combining cognitive theory with intelligent recommendation systems.*

Keywords: *Artificial Intelligence, Collaborative Filtering, Digital Health, EAST Framework, Ethical AI, Explainable AI, Just-in-Time Adaptive Interventions, Recommendation Systems, Sustainable Behavior, Theory of Planned Behavior.*

I. INTRODUCTION

Cognitive science-based behavioral interventions have long played a central role in addressing complex societal challenges such as adherence to public health, environmental sustainability, energy conservation and mental well-being. Conceptual theories like Health Belief Model, Theory of Planned Behavior, Social Cognitive Theory, and Dual Process Models have been used to give systematic insights into how and why human decisions are made [1-3]. These models explain behavioral processes related to beliefs, intentions, perceived control, habits, and cognitive biases. Nevertheless, conventional interventions that are founded on these frameworks tend to use static messaging, generalized nudges, and a low level of personalization, which restricts long-term viability and scalability.

The current model is based mostly on the Theory of Planned Behavior (TPB) according to which behavioral intention is defined by three main constructs, namely attitudes to the behavior, subjective norms, and perceived behavioral control. These determinants are all predictors of the strength of intentions, which predicts behavioral performance when there is adequate control. TPB provides a causal framework linking cognition and behavior and has measurable mediators which can be implemented in smart systems. Meanwhile, the EAST framework (Easy, Attractive, Social, Timely) provides a practical behavioral design framework that develops psychological principles into practical intervention characteristics. Whereas TPB explains why behavior takes place, EAST explains how behavioral environments may be manipulated so as to enhance the likelihood of action. Combined, both TPB and EAST offer a complementary theoretical framework of AI-driven behavioral systems: one with cognitive determinants, the other with principles of environmental optimization. TPB constructs can be operationalized using AI systems through approximating latent attitudes, normative exposure, and perceived control based on behavioral traces, and EAST principles can be used to guide the intervention delivery scheme (e.g., lessening friction, capitalizing on social evidence, maximizing timing). Such integration enables intelligent systems not only to predict behavior but also to intervene in theory-consistent ways through construct-level adaptation and delivery optimization.

Recent developments in artificial intelligence (AI), specifically machine learning, natural language processing, affective computing and reinforcement learning have revolutionized the ability to provide adaptive, data-driven behavioral interventions [4-6]. It is possible to predict behaviors on large-scale behavioral streams, identify trends, and dynamically adapt interventions to specific cognition, contextual factors and longitudinal behavior patterns. It has

enabled the development of AI-based behavioral interventions that combine TPB-based cognitive construct modelling with intelligent recommendation systems, leading to the personalized behavior change at scale.

The relevance of this integration can be seen especially in areas like digital health, adoption of renewable energy and sustainable consumption. AI-powered coaching programs and chatbots offer medication and lifestyle change assistance and mental health care to individuals in the healthcare setting [7,8]. Smart feedback and adaptive nudging platforms in the context of renewable energy and sustainability encourage energy-saving actions, demand-level control, and low-carbon change of lifestyle [9,10]. The congruence between AI and behavioral science is thus placed at the center of the modern initiatives on solving global problems in terms of climate change, chronic illness, and resource optimization.

Even with the rapid development, there are a number of gaps in conceptualization and methods still exist. To start with, the majority of AI-based behavioral systems are engineered towards prediction accuracy, as opposed to theoretically interpretable models, making the connection between data-driven models and known cognitive constructs disjointed [11]. Second, AI systems barely support transparency and explainability, which makes it harder to trust them, ethically deploy them, and align it with psychological concepts like autonomy and perceived control [12]. Third, algorithmic architecture lacks sufficient integration of longitudinal adaptation mechanisms of cognition such as formation of habits and cognitive load control [13]. Moreover, the problems of bias, unfairness, privacy, and unwanted manipulation of behavior uphold an ethical perspective, which needs to be studied systematically [12,14].

The synthesis of studies on the convergence of cognitive behavioral theory and intelligent systems across the board is thus opportune and important. This review discusses how AI-based systems can theoretically be based on established cognitive frameworks, how intelligent algorithms can operationalize constructs of behaviors, and how the systems can be designed to deliver scalable, ethical, and effective behavior change. The following sections will give: (i) a TPB-grounded theoretical foundation for behavioral intervention modelling that are useful in behavioral interventions; (ii) discussion on AI architectures that are implemented to adaptive behavior change; (iii) analysis of integration strategies that can be used between cognitive models and intelligent systems; (iv) critical analysis of technical, ethical, and methodological issues; and (v) future research directions on how interpretable, responsible, and domain-translatable AI-driven behavioral interventions can be developed.

II. LITERATURE REVIEW

The Theory of Planned Behavior (TPB) offers a theoretical framework that offers a causally organized approach to the association between cognition and action using three main determinants, which include attitudes toward the behavior, subjective norms, and perceived behavioral control. Behavioral intention is used as a proximal mediator between these determinants and behavioral performance. TPB provides operationally definable constructs and unlike broad behavioral taxonomies, these can be approximated computationally in intelligent systems. Engagement metrics, sentiment patterns and uniform behavior can be used to infer attitudes. The subjective norms can be estimated using peer behavior exposure, social comparison cue, and network-based patterns of activity. Based on the history of behavioral success, friction indicators and the contextual constraints, perceived behavioral control can be estimated. Such estimations enable the AI systems to go beyond making pure outcome prediction to cognitive interpretable state modelling. TPB constructs when coupled with recommendation systems are structured intermediate variables that can be used to make ranking decisions. The integration of behavioral theory and scalable personalization is achieved through alignment between machine learning predictions and the known psychological processes.

Table 1: Summary of key studies

Ref.	Focus	Findings (Key results and conclusions)
[15]	Core components and design principles for just-in-time adaptive interventions (JITAIs) in mobile/digital contexts	Defines core elements (distal/proximal outcomes, tailoring variables, decision points/rules, intervention options) and emphasizes time-varying personalization grounded in behavioural constructs; clarifies how “right time/type/amount” support is operationalized for adaptive intervention delivery [15].
[16]	Pragmatic framework for translating behavioural theory/evidence into dynamic models that inform adaptive intervention logic	Proposes a practical approach for organizing empirical and theoretical evidence into a time-sensitive model suitable for JITAI construction; highlights the need to explicitly model how contextual states and mechanisms evolve over time to guide decision rules [16].
[17]	Micro-randomized trials (MRTs) as an experimental design to optimize time-	Introduces MRTs as a method to estimate causal effects of intervention options at many decision points, supporting optimization of adaptive

	varying intervention components	policies and identification of time-varying moderators relevant to tailoring [17].
[18]	MRT design framework for evaluating JITAIs in mobile health (with emphasis on chronic disease contexts)	Provides a conceptual overview of MRTs for JITAI optimization, describing sequential randomization at decision points and how causal inference can be preserved while refining decision rules and components [18].
[19]	Reinforcement learning (RL) for learning and updating personalized treatment policies in a JITAI	Describes an RL approach embedded in a JITAI (HeartSteps) that updates intervention delivery probabilities using ongoing data; demonstrates how an algorithm can continuously improve policy decisions under repeated, context-dependent decision times [19].
[20]	Pre-implementation guidelines for designing RL algorithms for digital interventions	Identifies common failure modes for RL in real-world interventions (stability, learnability, operational constraints) and provides guidance for evaluation and simulation prior to deployment to support reliable personalization [20].
[21]	Data-driven personalization using smartphone sensing + recommendation-style algorithms for health feedback	Presents an approach that learns user behaviour patterns and preferences to generate personalized feedback; illustrates feasibility of automated, individualized recommendations derived from passively/actively collected behaviour data [21].
[22]	Standardized specification of intervention "active ingredients" for theory-based and AI-delivered behaviour change	Establishes a structured taxonomy for behaviour change techniques (BCTs), enabling clearer mapping between cognitive/behavioural mechanisms and digital intervention components that can be selected, combined, and operationalized in intelligent systems [22].
[23]	Conversational-agent delivery of automated behaviour change interventions across multiple behaviours	Reports evidence from a randomized trial that an automated counselling agent can influence targeted behaviours (e.g., physical activity and diet-related outcomes), supporting scalable, dialogue-based intervention delivery with potential for personalization [23].
[24]	State-of-the-field evaluation of JITAIs in mental health, including gaps in adaptivity/receptivity and analytics	Finds early-stage but generally positive feasibility/usability signals; reports that critical JITAI elements (e.g., receptivity/adaptivity, empirically grounded decision points/rules, and use of passive sensing/advanced analytics) are often incomplete or underutilized [24].

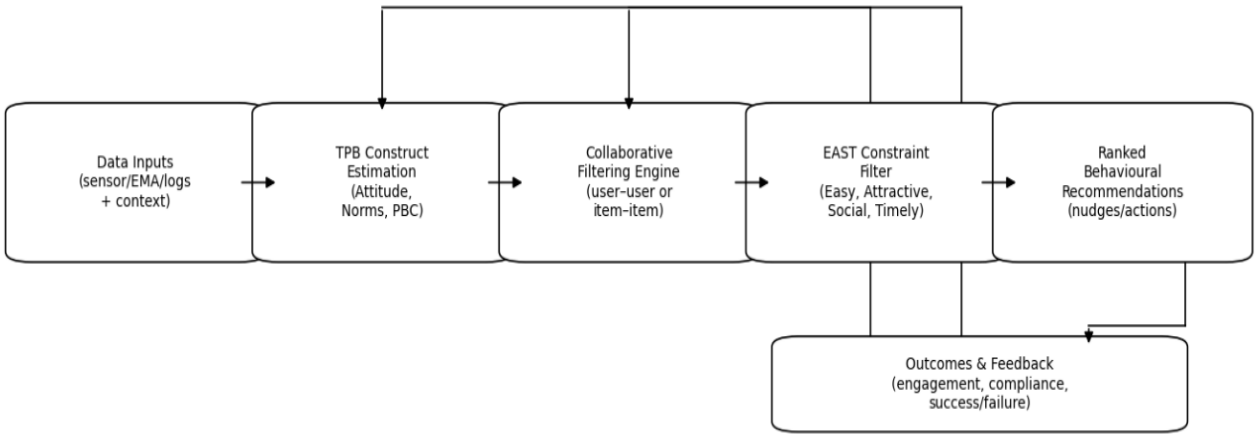

Figure 1: End-to-End TPB-Based Intelligent Recommendation Architecture

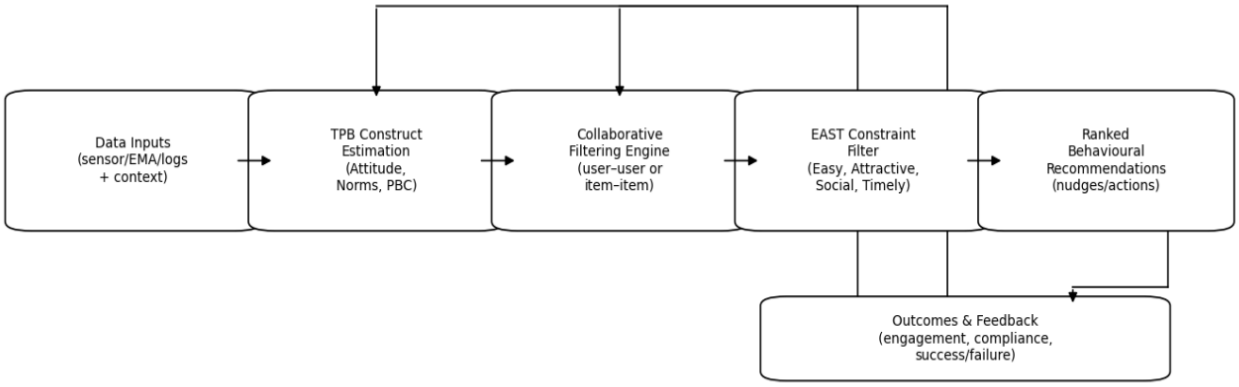

Figure 2: Closed-Loop TPB State Updating and Recommendation Optimization

Behavioral intervention scaling is a natural infrastructure that can be supported by recommendation systems that are widely used in digital platforms. In contrast to the process of static tailoring, recommendation algorithms provide intervention suggestions that are ranked dynamically by projected relevance. In collaborative filtering methods, behavioral similarity patterns are used to suggest useful prompts without necessarily specifying the behavioral psychology of the process involved.

The combination of recommendation systems and TPB constructs allows hybrid modelling where cognitive mediators are used to derive feature construction, and machine learning is used to optimize ranking and delivery. This is built in such a way that it fuses understandable behavioral theory and individual AI customization.

III. METHODOLOGY

The proposed model reorients AI-driven behavioural intervention around three interconnected layers, namely (1) TPB construct modelling, (2) collaborative filtering personalization, and (3) delivery of the intervention constrained by EAST.

1. TPB Construct Estimation Layer

The AI systems approximate the following latent constructs:

- Attitude cues: based on patterns of involvement, content sentiment reactions as well as history of behaviour.
- Subjective norm signals: derived based on social comparison feedback, peer exposure, and data of behaviour at network-level.
- Perceived behavioural control cues: based on prior success rates, time taken to complete a task, and behavioural friction.

The construct-level predictions are intermediary states as opposed to being dependent on outcome prediction.

2. Collaborative Filtering Personalization Layer.

The personalization is also possible under collaborative filtering, which tries to identify the users with similar behavioural patterns and history of responding. There are two mechanisms that can be used:

- User-based collaborative filtering, whereby the system suggests interventions that have been successful on other users.
- Item-based collaborative filtering refers to behavioural prompts, which are co-occurring in performance across contexts.

This gives the ability to scale personalization beyond what can be done through a rule-based personalization and maintain interpretable similarity structures.

3. EAST-Constrained Delivery Layer

Interventions chosen with the help of collaborative filtering are filtered with EAST constraints:

- Easy - make it easy, cut back behavioural steps.
- Attractive - increase salience, incentives or personalization.
- Social - include peer comparisons/normative cues.
- Timely- achieve maximization of delivery at peaks of receptivity.

This ensures that AI-driven recommendations remain aligned with established behavioural design principles.

The behavioural outcomes update TPB construct estimates and collaborative similarity matrices, which create a continuously learning recommendation system based on the cognitive theory.

IV. DISCUSSION

Table 2: AI-driven behavioral interventions for physical activity

Ref.	Study / system type	Design	Sample / setting	Outcome(s)	Key experimental results
[45]	SNapp (JITAI walking app; tailoring moderated by perceived usefulness)	RCT	Adults (real-world walking promotion)	Daily steps + moderators	Average step effect was not significant, while perceived usefulness significantly moderated intervention impact (B = 344.38; 90% CI 40.4–648.3), indicating gains concentrated among users who judged the system as useful.
[46]	PIC vs UIC JITAI (personalized intervention criteria vs uniform criteria)	Pilot comparative study	University students (2 weeks)	Steps, calories, distance in first hour post-prompt	Both PIC and UIC increased activity in the first hour after a prompt (both P < .001); PIC outperformed UIC for calories (P = .02), steps (P = .007), and distance (P = .003), but weekly sustained change was not significant.

These studies reinforce a recurring pattern in AI-enabled behavior change: strong short-horizon effects (minutes to hours) with attenuation over weeks, unless engagement and perceived value stay high [45], [46].

Table 3: Mental health conversational/AI interventions (controlled trials)

Ref.	Study / system type	Design	Sample / setting	Outcome(s)	Key experimental results
[47]	Topic-based chatbot (rule-based sessions: stress management, emotion regulation, value clarification)	Assessor-blinded RCT (10 days + 1 month)	n=285 adults	Intentions, literacy, symptoms, well-being	Significant improvements vs control in behavioural intentions (F_2,379.74=15.02; P<.001) and mental health literacy (F_2,423.57=4.27; P=.02). Report emphasized short-term gains with limited sustained effects at follow-up.
[48]	PATH AI-enabled app (CBT-informed chat therapy + tools)	RCT (2, 8, 12 weeks)	n=316 UK adults	GAD-7, PHQ-9	At 2 weeks, intervention arm had significantly lower anxiety and depression scores vs control with medium effect sizes; continued-use subgroup showed sustained improvements at 8 and 12 weeks with moderate-to-large effect sizes.

Trials increasingly report a dose/continuation effect (continued use → stronger follow-up outcomes), consistent with cognitive-behavioral skill acquisition requiring repeated practice rather than one-time exposure [48].

Table 4: Experimental evidence for digital/feedback interventions that shift energy behavior

Ref.	Intervention type	Design	Sample / setting	Outcome(s)	Key experimental results
[49]	Real-time energy feedback via in-home display (gas +	RCT	>800 households, Netherlands	Gas %, electricity %, total energy	Real-time feedback reduced gas by 6.9%, electricity by 2.2%, and total energy by 5.8%; mechanism

	electricity)				evidence suggested higher cost salience [49].
[50]	Normative framing with smart-meter-enabled in-home display	Randomized control field experiment	Households	Electricity use vs control	Norm-framed feedback reduced usage by 9% in the first week and 7% over 3 months; simple feedback and cost-framed feedback were not significantly different from control [50].
[51]	Electronic home energy reports (eHER) (peer comparisons + tips; electronic delivery)	Randomized field trial (12 months)	~9,000 households	Electricity consumption	Program reduced electricity consumption by 2.9% (95% CI −5.0% to −0.76%), despite high non-compliance in delivery [51].
[52]	Peer comparison feedback (HER-style)	Large, randomized field experiments	Utility customers	Energy consumption	Reported reductions in energy consumption of 1.2%–2.1%, sustained for 7–12 months across sites [52].

Why this matters for renewable energy: Flexible demand and reduced consumption directly support renewable integration by lowering peaks and improving load-shifting feasibility. Experimental evidence shows that framing and real-time salience can produce materially larger effects than information-only designs [49], [50].

TPB constructs give an organization to the empirical findings that have been witnessed in the digital health and energy-feedback interventions. The short-term behavioural improvements found in JITAI research can be due to a short-term reinforcement of behavioural intention by salience and normative cues. Nevertheless, time attenuation implies a lack of reinforcement of the perceived ability to control behaviour or a lack of stable attitude. The subjective norm activation is directly related to norm-framed energy feedback, as predicted by the TPB. Equally, perceived usefulness as the mediator in intervention effect indicates the mediation of attitude. The sustained-effect of mental health interventions may reflect cumulative reinforcement of the perceived behavioural control by means of repeated successful enactment. These results indicate that longitudinal construct-level reinforcement is a may support more durable behaviour change than episodic prompting.

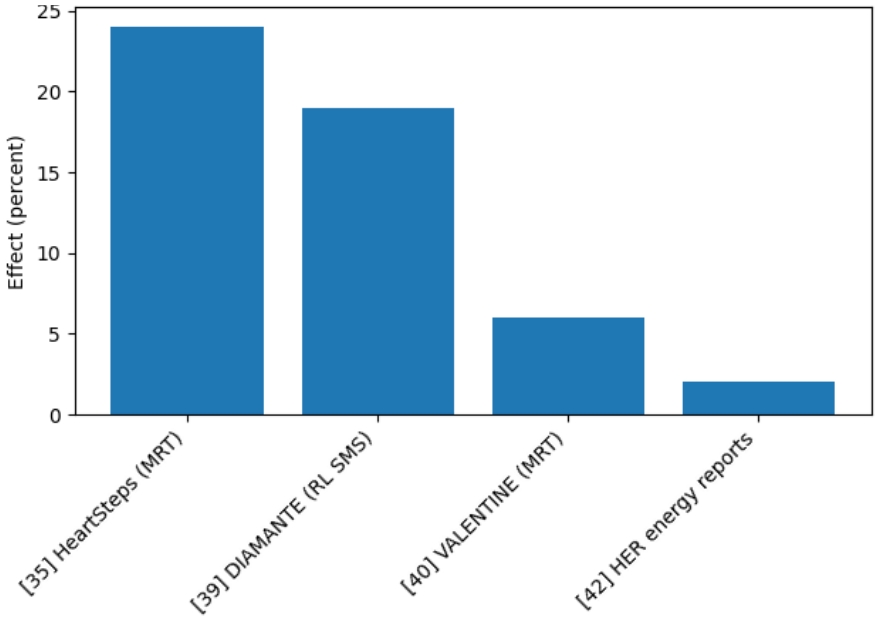

Figure 3: Effect magnitude comparison across outcome domains

Figure 3 compares representative effect magnitudes reported across selected intervention domains. The comparison is illustrative and is not intended as a formal meta-analytic synthesis because the studies differ in design, duration, and outcome metrics. This graph typically shows that high-frequency feedback + norm framing achieves larger short-horizon reductions than monthly report formats, while longer programs often stabilize around 1–3% unless salience remains high [49]–[52].

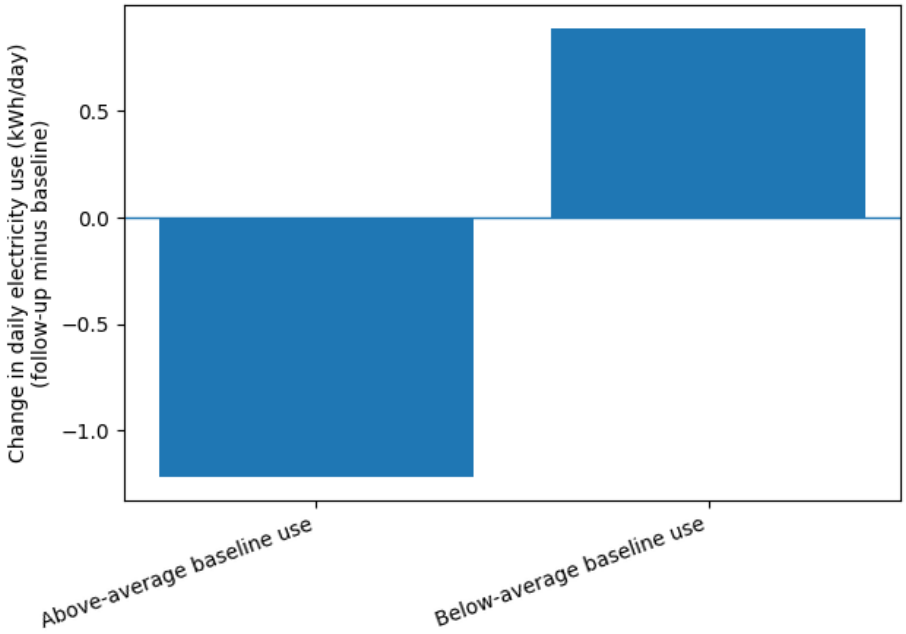

Figure 4: Short-horizon bump vs long-horizon fade (behavioural intervention durability)

Figure 4 illustrates the common durability pattern observed across selected studies, in which short-term gains attenuate over time unless engagement is sustained. The general trend of durability can be represented by a line plot based on qualitative trend anchors based on controlled results:

- PIC/UIC JITAI: no significant increase on a weekly basis but an important increase of hours within the first hour [46].
- Topic chatbot: meaningful short-term effects of intentions/literacy and minimal long-term effects [47].
- PATH continued-use subgroup continued to gain; discontinuation weakened durability [48].

This pattern highlights one of the more fundamental gaps in the field: instant responsiveness and personalizing and aligning theory enhances that, but it needs to be maintained by long-term engagement, lighter load, and pacing adaptation [45]-[48].

Key experimental gaps highlighted by these results

1. Fidelity to mechanism is irregularly measured. Most of the trials provide results but fail to measure mediators (capability, opportunity, motivation proxies; autonomy support; habit strength), preventing theory-testing even in the case of theory-inspired content [45]-[48].
2. Durability and habituation are the significant weaknesses. The short-term payoffs are normal; it is common to find a multi-week or multi-month persistence that relies on continued use and finding it useful [45]-[48].
3. The preponderance of intervention framing is over the information. Test studies indicate that feedback format and framing (normative vs cost vs raw consumption) can be the difference between an emerging effect [50].
4. Generalization across contexts is uncertain. eHER delivery may be comparable or even greater than paper reports in certain situations, however, delivery compliance and market context determine the extent of impact [51], [52].

V. FUTURE DIRECTIONS

A. Machine Learning with theory constraints

One key research direction is the direct incorporation of behavioral theory constructs into model architectures (as opposed to considering theory as post hoc interpretation). Hybrid human-AI systems Hybrid human-AI systems are

meant to enable enhanced interpretability and fidelity to mechanisms by integrating computational learning with structured psychological models [53]. Reinforcement learning models of JITAIs show promise but need to be converted into practical use of theoretically relevant state variables and reward functions [54]. The operationalization of constructs like autonomy support, perceived competence, habit strength, and readiness to change are measurable states that should be operationalized in future systems as updateable states.

B. Longitudinal Adaptivity and Habit Formation

Short-horizon responsiveness is common, but modelling dynamics of the habit formation and cognitive load is necessary to achieve lasting behavior changes. Habit and automatic theories help stress the importance of repetitions in consistent situations to convert behavioral processing to automatic processing [57]. Adaptive fading, timing optimization, and burden minimization should therefore be included in the AI system to avoid the fatigue of intervention. Long-term longitudinal studies (several months, several years) are scarcely utilized, and long-term measures should not be confined to instant compliance.

C. Explainability and Calibration of Trust

Black-box personalization can reduce user trust, particularly in the field of health and sustainability. Explainable AI algorithms, e.g. local interpretable model explanations, give insight into reasoning behind recommendations [58]. According to research on trust calibration, the process of clarifying and creating cognitive simplicity in explanation should not result in overconfidence or misconception [59]. Autonomy support and ethical alignment can be improved by integrating the mechanism of explanation in the design of behavior intervention.

D. Fairness, Bias, and Equity on Adaptive Systems

Behaviorally trained AI systems may encode structural bias, which will not only result in disparate intervention efficacies among demographic groups but also mean that the AI system is more effective in certain groups than in others. The studies on algorithmic auditing have shown that predictive systems may display disproportionate impact despite no obvious demographic inputs [60]. Future generation behavioral AI systems must be continuously monitored on performance of subgroups, optimization with fairness and inclusive data collection policies to prevent increasing inequities.

E. Privacy-Preserving Personalization

Behavioral AI is a major product that builds on a constant stream of data provided by smartphones, wearables, and sensors of the environment. Federated learning and other privacy-saving methods allow the model training of decentralized data without aggregating raw data in a centralized location [61]. Multi-party computation and differential privacy system provides further protection. The trade-off between the quality of personalization and privacy assurances will be one of the focal points of design.

F. Cross-Domain Translation: Well to Sustainability

Digital health research offers methodological blueprints [e.g., JITAI optimization and reinforcement learning personalization] which can be applied to other systems of renewable energy adoption and sustainable consumption [54], [55]. Experimental studies of behavioral energy prove that effects of framing and integration of social norms play an important role in shaping consumption patterns [56]. Smart systems that change sustainability prompts depending on contextual receptivity can enhance the peak-load regulation, demand response involvement, and carbon-reduction compliance.

G. The models of human-AI Collaboration

Instead of entirely automated interventions, the hybrid solution might be more accountable and relational by including AI advice with a human coach. Digital mental health evidence indicates that minimal human supervision will enhance adherence and clinical outcomes in contrast to fully automated systems [55]. The development of formalized collaboration frameworks between the human specialists and AI systems is a research problem.

VI. CONCLUSION

Combining the Theory of Planned Behavior with recommendation-system architecture and EAST principles of behavioral design can increase the theoretical consistency and scalability of the AI-based behavioral interventions. TPB offers interpretable cognitive determinants, collaborative filtering offers scalable customization and EAST limits the delivery of interventions to actionable behavioral maximization. The impact of empirical evidence in short- term behavioral responsiveness has been high but there have been enduring durability and equity problems. The countermeasures to these challenges need construct-level longitudinal modelling, fairness-rewarding personalization and explicit recommendation policies. The integration of behavioral theory and intelligent recommendation systems may

provide a meaningful advance of the prediction-based models to cognitively committed, adaptive, and behavior change technologies guided by ethical considerations.

Interest Conflicts

The author declares that there is no conflict of interest regarding the publication of this paper.

VII. REFERENCES

[1] Ajzen, I., The theory of planned behaviour, Organ. Behav. Hum. Decis. Process. 50(2) (1991) 179–211.

[2] Bandura, A., Social foundations of thought and action: A social cognitive theory, Englewood Cliffs, NJ: Prentice-Hall (1986).

[3] Kahneman, D., Thinking, fast and slow, New York: Farrar, Straus and Giroux (2011).

[4] Jordan, M. I. and Mitchell, T. M., Machine learning: Trends, perspectives, and prospects, Science 349(6245) (2015) 255–260.

[5] Russell, S. and Norvig, P., Artificial intelligence: A modern approach, 4th ed., Pearson (2021).

[6] Sutton, R. S. and Barto, A. G., Reinforcement learning: An introduction, 2nd ed., Cambridge, MA: MIT Press (2018).

[7] Laranjo, L., Dunn, A. G., Tong, H. L., et al., Conversational agents in healthcare: A systematic review, J. Am. Med. Inform. Assoc. 25(9) (2018) 1248–1258.

[8] Fitzpatrick, K. K., Darcy, A. and Vierhile, M., Delivering cognitive behaviour therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent, JMIR Ment. Health 4(2) (2017) e19.

[9] Delmas, M. A., Fischlein, M. and Asensio, O. I., Information strategies and energy conservation behaviour: A meta-analysis, Energy Policy 61 (2013) 729–739.

[10] Allcott, H., Social norms and energy conservation, J. Public Econ. 95(9–10) (2011) 1082–1095.

[11] Rudin, C., Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, Nat. Mach. Intell. 1(5) (2019) 206–215.

[12] Floridi, L. and Cowls, J., A unified framework of five principles for AI in society, Harv. Data Sci. Rev. 1(1) (2019) 1–15.

[13] Wood, W. and Rünger, D., Psychology of habit, Annu. Rev. Psychol. 67 (2016) 289–314.

[14] Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S. and Floridi, L., The ethics of algorithms: Mapping the debate, Big Data Soc. 3(2) (2016) 1–21.

[15] Nahum-Shani, I., Smith, S. N., Spring, B. J., Collins, L. M., Witkiewitz, K., Tewari, A. and Murphy, S. A., Just-in-time adaptive interventions in mobile health, Am. J. Prev. Med. 52(4) (2018) 446–462.

[16] Nahum-Shani, I., Hekler, E. B., Spruijt-Metz, D. and Murphy, S. A., Building health behaviour models for JITAIs, Health Psychol. 34(S) (2015) 1209–1219.

[17] Klasnja, P., Hekler, E. B., Shiffman, S., et al., Microrandomized trials for developing just-in-time adaptive interventions, Health Psychol. 34(S) (2015) 1220–1228.

[18] Golbus, J. R., Dempsey, W., Jackson, E. A., Nallamothu, B. K. and Klasnja, P., Microrandomized trial design for evaluating JITAIs, Circ. Cardiovasc. Qual. Outcomes 14(2) (2021) e006760.

[19] Liao, P., Greenewald, K., Klasnja, P. and Murphy, S. A., Personalized HeartSteps reinforcement learning algorithm, Proc. ACM IMWUT 4(1) (2020) Article 18.

[20] Trella, A. L., Zhang, K. W., Nahum-Shani, I., Shetty, V., Doshi-Velez, F. and Murphy, S. A., Designing reinforcement learning algorithms for digital interventions, Algorithms 15(8) (2022) 255.

[21] Rabbi, M., Aung, M. H., Zhang, M. and Choudhury, T., MyBehavior: Automatic personalized health feedback using smartphones, Proc. ACM UbiComp (2015) 707–718.

[22] Michie, S., Richardson, M., Johnston, M., et al., The behaviour change technique taxonomy (v1), Ann. Behav. Med. 46(1) (2013) 81–95.

[23] Bickmore, T. W., Schulman, D. and Sidner, C., Automated interventions for multiple health behaviours using conversational agents, Patient Educ. Couns. 92(2) (2013) 142–148.

[24] van Genugten, C. R., et al., Just-in-time adaptive interventions in mental health: A systematic review, Front. Digit. Health 7 (2025) 1460167.

[25] Michie, S., van Stralen, M. M. and West, R., The behaviour change wheel, Implement. Sci. 6 (2011) 42.

[26] Michie, S., Atkins, L. and West, R., The Behaviour Change Wheel: A guide to designing interventions, Silverback Publishing (2014).
[27] Ryan, R. M. and Deci, E. L., Self-determination theory and well-being, Am. Psychol. 55(1) (2000) 68–78.
[28] Oinas-Kukkonen, H. and Harjumaa, M., Persuasive systems design, Commun. Assoc. Inf. Syst. 24(1) (2009) 485–500.
[29] Gollwitzer, P. M., Implementation intentions, Am. Psychol. 54(7) (1999) 493–503.
[30] Prochaska, J. O. and Velicer, W. F., The transtheoretical model of health behaviour change, Am. J. Health Promot. 12(1) (1997) 38–48.
[31] Ribeiro, M. T., Singh, S. and Guestrin, C., Explaining the predictions of any classifier (LIME), Proc. ACM SIGKDD (2016) 1135–1144.
[32] Obermeyer, Z., Powers, B., Vogeli, C. and Mullainathan, S., Dissecting racial bias in an algorithm used to manage the health of populations, Science 366(6464) (2019) 447–453.
[33] Shiffman, S., Stone, A. A. and Hufford, M. R., Ecological momentary assessment, Annu. Rev. Clin. Psychol. 4 (2008) 1–32.
[34] Dwork, C., McSherry, F., Nissim, K. and Smith, A., Calibrating noise to sensitivity in private data analysis, Lect. Notes Comput. Sci. 3876 (2006).
[35] Schafer, J. B., Konstan, J. A. and Riedl, J., Recommender systems in e-commerce, IEEE Internet Comput. 5(3) (2001) 38–45.
[36] Ricci, F., Rokach, L. and Shapira, B., Recommender Systems Handbook, 2nd ed., Springer (2015).
[37] Sarwar, B., Karypis, G., Konstan, J. and Riedl, J., Item-based collaborative filtering recommendation algorithms, WWW Conf. (2001) 285–295.
[38] Burke, R., Hybrid recommender systems, User Model. User-Adapt. Interact. 12(4) (2002) 331–370.
[39] Guidotti, R., Monreale, A., Ruggieri, S., et al., A survey of methods for explaining black box models, ACM Comput. Surv. 51(5) (2018) 93.
[40] Miller, T., Explanation in artificial intelligence: Insights from the social sciences, Artif. Intell. 267 (2019) 1–38.
[41] Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K. and Galstyan, A., A survey on bias and fairness in machine learning, ACM Comput. Surv. 54(6) (2021) 115.
[42] Doshi-Velez, F. and Kim, B., Towards a rigorous science of interpretable machine learning, arXiv preprint arXiv:1702.08608 (2017).
[43] Breiman, L., Random forests, Mach. Learn. 45(1) (2001) 5–32.
[44] Dwork, C. and Roth, A., The algorithmic foundations of differential privacy, Found. Trends Theor. Comput. Sci. 9(3–4) (2014) 211–407.
[45] Vos, A. L., De Bruijn, G.-J., Klein, M. and Boerman, S. C., Effectiveness of a JITAI app to increase daily steps, Am. J. Prev. Med. (2025).
[46] Ikegaya, M., Foo, J. C., Murata, T., Oshima, K. and Kim, J., Mobile JITAI for physical activity in university students, JMIR Hum. Factors (2025).
[47] Tong, A. C. Y., Wong, K. T. Y., Chung, W. W. T. and Mak, W. W. S., Chatbots for mental health self-care, J. Med. Internet Res. (2025).
[48] Allen, A., Young, A. H., Jellesma, F. C., et al., AI-enabled mental health intervention for generalized anxiety, J. Affect. Disord. 401 (2026) 121275.
[49] Boomsma, M., Vringer, K. and van Soest, D., Real-time energy feedback and household energy usage, J. Environ. Econ. Manag. 132 (2025) 103163.
[50] Schultz, P. W., Estrada, M., Schmitt, J., Sokoloski, R. and Silva-Send, N., Smart meter feedback and social norms, Energy 90 (2015) 351–358.
[51] Henry, M. L., Ferraro, P. J. and Kontoleon, A., Electronic home energy reports and behavioural change, Energy Policy 132 (2019) 1256–1261.
[52] Ayres, I., Raseman, S. and Shih, A., Peer comparison feedback and residential energy usage, J. Law Econ. Organ. 29(5) (2013) 992–1022.

[53] Bickmore, T. W. and Picard, R. W., Establishing and maintaining long-term human–computer relationships, ACM Trans. Comput.-Hum. Interact. 12(2) (2005) 293–327.

[54] Murphy, S. A., Optimal dynamic treatment regimes, J. R. Stat. Soc. B 65(2) (2003) 331–355.

[55] Mohr, D. C., Weingardt, K. R., Reddy, M. and Schueller, S. M., Problems in digital mental health research, Psychiatr. Serv. 68(5) (2017) 427–429.

[56] Cialdini, R. B., Crafting normative messages to protect the environment, Curr. Dir. Psychol. Sci. 12(4) (2003) 105–109.

[57] Lally, P., van Jaarsveld, C. H. M., Potts, H. W. W. and Wardle, J., Habit formation in the real world, Eur. J. Soc. Psychol. 40(6) (2010) 998–1009.

[58] Lundberg, S. M. and Lee, S. I., A unified approach to interpreting model predictions (SHAP), Adv. Neural Inf. Process. Syst. 30 (2017).

[59] Lee, J. D. and See, K. A., Trust in automation: Designing for appropriate reliance, Hum. Factors 46(1) (2004) 50–80.

[60] Barocas, S. and Selbst, A. D., Big data's disparate impact, Calif. Law Rev. 104(3) (2016) 671–732.

[61] Koren, Y., Bell, R. and Volinsky, C., Matrix factorization techniques for recommender systems, IEEE Comput. 42(8) (2009) 30–37.