Mitigating Problematic Social Media Use through Paired Recommender Systems with Contrasting Objectives

Extended Abstract

Stefano Livella
University of Milan-Bicocca
Milan, Italy
s.livella@campus.unimib.it

Luca Bolis
University of Milan-Bicocca
Milan, Italy
l.bolis3@campus.unimib.it

Sabrina Patania
University of Milan-Bicocca
Milan, Italy
sabrina.patania@unimib.it

Matteo Papini
University of Milan
Milan, Italy
matteo.papini@unimi.it

Dimitri Ognibene
University of Milan-Bicocca
Milan, Italy
dimitri.ognibene@unimib.it

ABSTRACT

Social media platforms enable connection and entertainment, but engagement-optimizing algorithms may drive compulsive overuse and harm well-being. We propose a paired-trained recommender with two shared modules: one maximizes engagement, while the other discourages excessive sessions. To investigate user-recommender interactions, we model users with a dual-system reinforcement learning framework from computational neuroscience, capturing individual differences such as variations in impulsivity and preference structure. Compared to a standard engagement-maximizing baseline evaluated over 200 trajectories per user type, our approach reduces addiction-like behaviors without sacrificing engagement, suggesting careful design can mitigate harmful overuse.

KEYWORDS

Social Media; Recommender Systems; Algorithm Auditing; User Behavior Modeling; Well-Being; Behavioural Addiction; SIM

ACM Reference Format:
Stefano Livella, Luca Bolis, Sabrina Patania, Matteo Papini, and Dimitri Ognibene. 2026. Mitigating Problematic Social Media Use through Paired Recommender Systems with Contrasting Objectives: Extended Abstract. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 3 pages. https://doi.org/10.65109/ROHQ5247

1 INTRODUCTION

Social media platforms have transformed communication and information consumption, yet they raise serious concerns about compulsive use. Unlike substance addiction, social media overuse stems from cognitive vulnerabilities, including difficulties in balancing immediate rewards against long-term well-being, the influence of engagement-maximizing recommendation systems and the complexity of exploration process. [9, 15, 16]. Although overall usage continues to grow, longitudinal evidence shows declining life satisfaction, revealing an engagement–utility gap acknowledged both by researchers and platform officials [13, 21, 22]. This issue is now recognized as a public health concern, particularly among adolescents, with links to depression and anxiety [8, 23].

Algorithmic feeds and infinite scroll amplify engagement by continuously tailoring content to users' past behavior, creating self-reinforcing consumption loops that hinder disengagement [1, 3]. Dual-system reinforcement learning (RL) models formalize the tension between habitual (model-free) and goal-directed (model-based) control, capturing how repeated engagement can become automatic even when it conflicts with users' long-term goals [4, 6, 18, 20]. By extending RL addiction models to incorporate explicit representations of recommender systems and the induced non-Markovian dynamics, it is possible to study how algorithmic interventions influence user behavior. [11, 16].

2 METHODS

We model the interaction between a user and a recommender system as a multi-agent system operating within a shared environment of psycho-physical states. This interaction is formalized as a two-player general-sum Markov game [12] $(\mathcal{S}, \mathcal{A}_{1,2}, P, R_{1,2})$, where $\mathcal{S}$ represents the user's state, $\mathcal{A}_1$ the user's actions, $\mathcal{A}_2$ the recommender's actions, P the transition probability function, and R_1, R_2 the respective reward functions. Notably, the recommender acts only in a subset of states, making its reward sparse and dependent on the user's policy.

2.1 User Modeling

The user is represented by a dual-system reinforcement learning (RL) model [20], combining a Model-Free (MF) Q-learning component [25] that captures habitual behavior and a Model-Based (MB) component using Prioritized Sweeping [14] for goal-directed planning. The balance between habitual and reflective decision-making is controlled by a parameter β, with the combined Q-value computed as:

$$Q_{MX}(s,a) = \beta Q_{MB}(s,a) + (1-\beta) Q_{MF}(s,a).$$

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). https://doi.org/10.65109/ROHQ5247

Additionally, the MBUS (Model-Based Updates per Step) parameter determines the number of internal updates per iteration, modulating planning depth.

2.2 Recommender System Modeling

We introduce two architectures: the PutIn-PutOut Recommender, composed of two modules, and a Paired-Training Recommender, which updates both modules at each interaction. The PutIn module maximizes engagement through positive rewards for accepted recommendations (+1) and penalizes rejections (-1), while the PutOut module promotes healthy disengagement, assigning positive rewards (+1) when users avoid content and negative rewards (-1) for prolonged use. Both modules are implemented as non-stationary multi-armed bandits [24], using an exponentially weighted average to adapt to shifting user preferences [27]. The Paired-Training system extends this approach by allowing feedback from one module to inform updates in both modules.

2.3 Environment Design

The environment maps state-action pairs to subsequent states and rewards, with states grouped into: Healthy (reflecting the user's physical and psychological well-being), Neutral (reflecting no immediate positive or negative effects on users), Recommender System (representing interaction with social media, divided into RecShort (brief use) and RecLong (prolonged use)), Aftereffects (modeling the negative consequences of excessive social media use) and Balanced (states that allow short social media interactions with limited penalties).

Users can perform three actions: aG (action Goal, well-being), aW (action Wait, no effect) and aD (action Drug, possibly incurring in penalties). Probabilistic transitions (e.g. 50% chance of going from RecShort to RecLong) capture the chance that, with repeated use, a user drifts into excessive time spent, leading to negative consequences. The escape point from RecLong provide a mechanism for balanced engagement, reflecting realistic attempts to avoid excessive use.

This environment extends previous models of behavioral addiction [15], while capturing the complex interplay between dual-system decision-making and recommender interventions. By simulating these dynamics, our framework allows the study of how algorithmic design influence both compulsive and balanced engagement [17].

3 EXPERIMENTS

To evaluate model performance, we conducted 200 simulations of 100,000 steps for each user parameter configuration, corresponding to different endophenotypes. To enhance robustness, we applied a bootstrapping procedure: for each configuration, 50 trajectories were randomly sampled from the 200 simulations over 100 resampling iterations.

We explored three main parameters. First, β, which captures the balance between model-based (goal-directed) and model-free (habitual) control. Second, recommender-assigned rewards to simulate three user populations: ADD (users for whom most content is highly addictive), NOSM (users for whom most content is non-addictive) and NTRL (users for whom some content is addictive and other content is non-addictive). Third, we varied the PutIn learning rate to assess how the speed of adaptation influences user behavior. The MBUS parameter was fixed at 2, as higher values produced negligible differences in this simplified environment.

To assess the benefits of our proposed recommender, we evaluated its performance compared to a baseline architecture: an aggressive dual-PutIn system designed to maximize user engagement.

4 RESULTS

User behavior was evaluated in segments of 200 steps and classified into four categories: Healthy, Addicted, Balanced and Uncertain. These categories reflect reinforcement-learning–based distinctions rather than clinical diagnoses. Healthy agents follow the long-term optimal policy, Addicted agents over-select short-term rewarding states, Balanced agents alternate between social media consumption and healthy actions and, finally, Uncertain agents show no dominant strategy. Consistently with prior work [2, 5, 15], addiction in our framework emerges from model-free dominance and exploration complexity.

Under the engagement-maximizing recommender, the population for whom most content is addictive (popADD) predominantly converges to addictive behavior, whereas non-addictive population (popNOSM) gradually adopts healthier strategies by disengaging from the social media platform. Balanced behavior rarely emerges in this setting.

The PutIn–PutOut architecture mitigates addiction by fostering balanced behavior, but over time many users disengage entirely, raising sustainability concerns. Adjusting the learning rate of PutIn improves stability, particularly when PutIn adapts more slowly. In this configuration, PutOut can adapt more effectively to discourage transitions into prolonged sessions, thereby reducing the likelihood that users develop harmful usage patterns.

The paired-trained recommender further stabilizes balanced behavior by updating both modules at each interaction, substantially reducing addiction across populations.

Finally, the scalability test with a larger content set (16 arms instead of 4) confirms that the proposed recommender architecture maintains its effectiveness.

Additional results, along with the code and environment details are available at https://github.com/DimNeuroLab/SocialMediaAddiction.

5 CONCLUSION AND FUTURE DIRECTIONS

This study introduces a framework to mitigate compulsive social media use, demonstrating its effectiveness through simulation-based evidence, by modeling users with a dual-system reinforcement learning approach [9] interacting with a dynamic multi-armed recommender system. The PutIn–PutOut architecture reduced addictive patterns and tuning the PutIn learning rate further stabilized balanced engagement. The paired-trained recommender system resolved update imbalances, fostering more consistent balanced social media usage across user types, with these effects remaining robust even when larger content sets are considered.

Limitations include reliance on synthetic data, simplified user modeling, and lack of social or content-aware interactions [7, 10, 26], while persistent addiction under challenging conditions [9, 19] highlights directions for future works.

ACKNOWLEDGMENTS

This research was supported by the Italian Ministry of University and Research under Grant No. 2023-NAZ-0206, PsyFuture – Dipartimento di Eccellenza 2023-2027.

REFERENCES

[1] Keyi Chen. 2024. If it is bad, why don't I quit? Algorithmic recommendation use strategy from folk theories. *Global Media and China* 9, 3 (2024), 344–361. https://doi.org/10.1177/20594364231209354 arXiv:https://doi.org/10.1177/20594364231209354

[2] Nathaniel D Daw, Samuel J Gershman, Ben Seymour, Peter Dayan, and Raymond J Dolan. 2011. Model-based influences on humans' choices and striatal prediction errors. *Neuron* 69, 6 (2011), 1204–1215.

[3] Cynthia A. Dekker, Susanne E. Baumgartner, and Sindy R. Sumter. 2025. For you vs. for everyone: The effectiveness of algorithmic personalization in driving social media engagement. *Telematics and Informatics* 101 (2025), 102300. https://doi.org/10.1016/j.tele.2025.102300

[4] Lorenz Deserno, Quentin JM Huys, Rebecca Boehme, Ralph Buchert, Hans-Jochen Heinze, Anthony A Grace, Raymond J Dolan, Andreas Heinz, and Florian Schlagenhauf. 2015. Ventral striatal dopamine reflects behavioral and neural signatures of model-based control during sequential decision making. *Proceedings of the National Academy of Sciences* 112, 5 (2015), 1595–1600.

[5] Ray J Dolan and Peter Dayan. 2013. Goals and habits in the brain. *Neuron* 80, 2 (2013), 312–325.

[6] Vincenzo G Fiore, Dimitri Ognibene, Bryon Adinoff, and Xiaosi Gu. 2018. A multilevel computational characterization of endophenotypes in addiction. *eneuro* 5, 4 (2018).

[7] Sabrina Guidotti, Gregor Donabauer, Simone Somazzi, Udo Kruschwitz, Davide Taibi, and Dimitri Ognibene. 2024. Modeling Social Media Recommendation Impacts Using Academic Networks: A Graph Neural Network Approach. In *International Workshop on Recommender Systems for Sustainability and Social Good*. Springer, 63–72.

[8] Zaheer Hussain and Mark D Griffiths. 2018. Problematic social networking site use and comorbid psychiatric disorders: A systematic review of recent large-scale studies. *Frontiers in psychiatry* 9 (2018), 686.

[9] Ayaka Kato, Kanji Shimomura, Dimitri Ognibene, Muhammad A Parvaz, Laura A Berner, Kenji Morita, and Vincenzo G Fiore. 2023. Computational models of behavioral addictions: State of the art and future directions. *Addictive behaviors* 140 (2023), 107595.

[10] Bosen Lian, Wenqian Xue, Frank L. Lewis, and Tianyou Chai. 2022. Inverse reinforcement learning for multi-player noncooperative apprentice games. *Automatica* 145 (2022), 110524. https://doi.org/10.1016/j.automatica.2022.110524

[11] Yuanguo Lin, Yong Liu, Fan Lin, Lixin Zou, Pengcheng Wu, Wenhua Zeng, Huanhuan Chen, and Chunyan Miao. 2023. A survey on reinforcement learning for recommender systems. *IEEE Transactions on Neural Networks and Learning Systems* 35, 10 (2023), 13164–13184.

[12] Michael L. Littman. 1994. Markov Games as a Framework for Multi-Agent Reinforcement Learning. In *ICML*. Morgan Kaufmann, 157–163.

[13] Natasha Lomas. 2017. Google to ramp up AI efforts to ID extremism on YouTube. *TechCrunch* 24 (2017), 2019. https://techcrunch.com/2017/06/19/google-to-ramp-up-ai-efforts-to-id-extremism-on-youtube/

[14] Andrew W Moore and Christopher G Atkeson. 1993. Prioritized sweeping: Reinforcement learning with less data and less time. *Machine learning* 13, 1 (1993), 103–130.

[15] Dimitri Ognibene, Vincenzo G Fiore, and Xiaosi Gu. 2019. Addiction beyond pharmacological effects: The role of environment complexity and bounded rationality. *Neural Networks* 116 (2019), 269–278.

[16] Dimitri Ognibene, Rodrigo Wilkens, Davide Taibi, Davinia Hernández-Leo, Udo Kruschwitz, Gregor Donabauer, Emily Theophilou, Francesco Lomonaco, Sathya Bursic, Rene Alejandro Lobo, et al. 2023. Challenging social media threats using collective well-being-aware recommendation algorithms and an educational virtual companion. *Frontiers in Artificial Intelligence* 5 (2023), 654930.

[17] Alfonso Pellegrino, Alessandro Stasi, and Veera Bhatiasevi. 2022. Research trends in social media addiction and problematic social media use: A bibliometric analysis. *Frontiers in psychiatry* 13 (2022), 1017506.

[18] Srinivasan A Ramakrishnan, Riaz B Shaik, Tamizharasan Kanagamani, Gopi Neppala, Jeffrey Chen, Vincenzo G Fiore, Christopher J Hammond, Shankar Srinivasan, Iliyan Ivanov, V Srinivasa Chakravarthy, et al. 2025. Impaired arbitration between reward-related decision-making strategies in Alcohol Users compared to Alcohol Non-Users: a computational modeling study. *NPP—Digital Psychiatry and Neuroscience* 3, 1 (2025), 1.

[19] A David Redish. 2004. Addiction as a computational process gone awry. *Science* 306, 5703 (2004), 1944–1947.

[20] A David Redish, Steve Jensen, and Adam Johnson. 2008. Addiction as vulnerabilities in the decision process. *Behavioral and brain sciences* 31, 4 (2008), 461–487.

[21] Manoel Horta Ribeiro, Raphael Ottoni, Robert West, Virgílio AF Almeida, and Wagner Meira Jr. 2020. Auditing radicalization pathways on YouTube. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*. 131–141.

[22] Kevin Roose et al. 2020. Rabbit hole. *The New York Times* (2020). https://www.nytimes.com/column/rabbit-hole

[23] Arianna Sala, Lorenzo Porcaro, and Emilia Gómez. 2024. Social Media Use and adolescents' mental health and well-being: An umbrella review. *Computers in Human Behavior Reports* 14 (2024), 100404. https://doi.org/10.1016/j.chbr.2024.100404

[24] Richard S Sutton, Andrew G Barto, et al. 1998. *Reinforcement learning: An introduction*. Vol. 1. MIT press Cambridge.

[25] Christopher JCH Watkins and Peter Dayan. 1992. Q-learning. *Machine learning* 8, 3 (1992), 279–292.

[26] Siliang Zeng, Chenliang Li, Alfredo Garcia, and Mingyi Hong. 2023. When Demonstrations meet Generative World Models: A Maximum Likelihood Framework for Offline Inverse Reinforcement Learning. In *Advances in Neural Information Processing Systems*, A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (Eds.), Vol. 36. Curran Associates, Inc., 65531–65565. https://proceedings.neurips.cc/paper_files/paper/2023/file/ce9d3c592712d23f2ec3671941d67fa1-Paper-Conference.pdf

[27] Lixin Zou, Long Xia, Zhuoye Ding, Jiaxing Song, Weidong Liu, and Dawei Yin. 2019. Reinforcement learning to optimize long-term user engagement in recommender systems. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. 2810–2818.