AI-Based Digital Twin Modeling of Human Preferences and Self-Concept for Personalized Recommendation Systems

Kevin J. Riley
Department of Computer Science, University of Central Florida, Orlando, FL, USA.
rileykevin@ucf.edu

Josea Heyes
Department of Computer Science, Binghamton University, Binghamton, NY, USA.
joseh@binghamton.edu

Joeal Jehnsten
Department of Computer Science, University of Houston, Houston, TX, USA.
joelmail@uh.edu

Abstract

Personalized recommendation systems have evolved from collaborative filtering engines into complex socio-technical infrastructures that mediate digital experience. A frontier emerging in this evolution is the use of AI-based digital twin modeling to capture not only explicit preferences but also the implicit and evolving self-concept of the individual. This paper develops a system-level analysis of such digital twins, focusing on architectural paradigms, data governance, computational trade-offs, and long-term sustainability. We argue that modeling human self-concept as a dynamic, multi-layered digital twin introduces distinct challenges for latency-sensitive recommendation delivery, privacy preservation, fairness, and infrastructural resilience. Drawing together perspectives from user modeling, cognitive psychology, distributed systems, and AI ethics, the paper examines how self-concept representations can be instantiated, synchronized with behavioral data streams, and protected against adversarial manipulation. We highlight the governance structures required to prevent exploitative personalization and the policy frameworks that must accompany large-scale deployment. Throughout, we emphasize the tension between rich inner-state modeling and the need for transparent, auditable, and ethically bounded recommender ecosystems. The analysis concludes by outlining forward-looking research directions that integrate digital twin sustainability, cross-domain interoperability, and participatory design into the next generation of human-centered recommendation infrastructure.

Keywords

Digital Twin, User Modeling, Self-Concept, Recommendation Systems, Personalization, Artificial Intelligence, Ethical AI, System Architecture.

1. Introduction

Recommendation systems are among the most pervasive computational artefacts of the present era, shaping information access, commercial transactions, and social interaction. Their underlying models have progressed from simple matrix factorization to deep neural architectures that ingest high-dimensional behavioral signals. Yet the majority of these

systems still operate on a narrow construct of the user: a vector of observed preferences that is treated as relatively stable, externally inferable, and separable from deeper cognitive structures. The resulting personalization often remains superficial, failing to account for the fluid, multifaceted nature of human identity. In response, research communities have begun to explore richer user representations, and the concept of a digital twin, originally conceived for industrial asset lifecycle management [1], has been proposed as a vehicle for capturing a holistic, evolving model of the individual.

The digital twin paradigm, when applied to human modeling, aims to create a persistent, data-driven virtual counterpart that mirrors not only what a person does but also the psychological constructs that organize those actions. Among these constructs, self-concept, the aggregation of beliefs, values, possible selves, and identity narratives, plays a central role in shaping preferences and decision-making [2, 3]. A recommendation system that understands the user's self-concept can anticipate needs that are not merely transactional but developmental, aligning suggestions with the person's aspirational identity rather than their immediate clickstream. Building such a system, however, requires a fundamental rethinking of data infrastructure, model architecture, and governance.

This paper provides a systems-oriented academic examination of AI-based digital twin modeling of human preferences and self-concept in the context of personalized recommendations. We adopt an interdisciplinary lens, integrating insights from user modeling [4], deep learning for recommendations [5], privacy research [6], fairness and accountability in machine learning [7], and digital twin frameworks from industrial and healthcare domains [8, 9]. The analysis emphasizes structural trade-offs rather than algorithmic specifics, focusing on architecture, deployment, robustness, and policy. After establishing conceptual foundations, we dissect the architectural choices that enable scalable self-concept modeling, evaluate system-level tensions among latency, accuracy, and privacy, and then pivot to the governance and ethical challenges that these models amplify. Finally, we discuss sustainability, long-term adaptation, and the policy landscape that will shape the responsible deployment of such systems.

2. Conceptual Foundations of Digital Twin Modeling for Personalization

In manufacturing, a digital twin is a high-fidelity virtual representation of a physical asset, continuously synchronized through sensor data to enable simulation, prediction, and optimization. Translating this paradigm to human users involves abstracting the "asset" as a cognitive-behavioral entity whose state is inferred from heterogeneous data streams. Instead of vibration and temperature, the system processes clickstreams, purchase histories, social media interactions, biometric signals, and self-reported inventories. The resulting twin must represent preferences, which are context-dependent and transient, alongside the more enduring structures of self-concept, which include ideal selves, feared selves, and autobiographical narratives [3]. This dual layer distinguishes a rich human digital twin from a conventional user profile.

Representing self-concept computationally demands a modeling substrate that can handle ambiguity, contradiction, and temporal drift. A user may simultaneously hold identities that pull in opposing directions, such as a health-conscious identity and an indulgent identity, or a professional persona and a leisure persona. The digital twin must encode these multiple selves and the situational triggers that activate them, moving beyond static attribute vectors toward dynamic relational graphs or memory-augmented neural architectures. The required conceptual depth aligns with foundational user modeling frameworks that advocate for

separating domain-dependent preferences from domain-independent psychological traits [4], while also incorporating the possible-selves framework that has long been influential in motivational psychology [3].

Integrating self-concept into a digital twin for recommendation also blurs the line between prediction and intervention. When a system models not just what a user likes but who they aspire to become, recommendations take on a prescriptive quality that can be empowering or manipulative. This duality foregrounds the ethical gravity of the infrastructure, as the system's internal representations become instruments of influence. Any robust design must therefore embed constraints within the modeling layer, ensuring that self-concept modeling serves the user's autonomy rather than pure engagement optimization [7]. The subsequent sections explore the architectures and system-level decisions that realize these conceptual aims at scale.

3. Architectural Considerations for Scalable AI-Based Digital Twin Personalization

The instantiation of a self-concept-aware digital twin in a recommendation pipeline introduces significant architectural complexity. A canonical architecture separates the system into a long-term twin store, a fast inference engine, and a synchronization layer that continuously updates the twin from raw event streams. The long-term store persists a structured representation of the user's identity graph, incorporating explicit values, semantified behavioral patterns, and narrative summaries of past interactions. This store is not a simple database; it requires graph-native or hybrid relational-graph storage capable of expressing the multi-faceted self and its temporal evolution [8].

The inference engine consumes the twin's current state, along with situational context, to generate recommendations. Here, latency requirements collide with model complexity. Industrial recommendation systems must serve predictions in tens of milliseconds, a constraint that historically precluded deep introspection. To reconcile rich self-concept models with low-latency serving, architectures often employ a two-path design: a compressed, pre-computed embedding of the user's salient identity facets is refreshed periodically and served from an in-memory cache, while deeper, exploratory personalization that draws on aspirational selves is handled asynchronously, influencing future model updates without blocking the critical path. This decoupling mirrors patterns established in large-scale deep learning recommendation pipelines [5] but introduces the added challenge of representing identity stability across update cycles.

Synchronization presents its own difficulties. The digital twin must absorb streaming data, reconcile it with existing self-concept structures, and detect meaningful shifts without overreacting to noise. Event-driven microservice architectures, commonly built on platforms like Apache Kafka or cloud-native streaming services, are well-suited to this task when paired with stateful stream processors that maintain a temporally consistent representation of the twin. The adoption of federated learning further alters the architecture by keeping raw behavioral data on user devices while only model updates or encrypted gradients travel to the central twin store, thereby strengthening privacy guarantees at the cost of increased synchronization complexity and reduced visibility for debugging [10]. The selection among centralized, federated, or hybrid architectures is therefore a foundational system-level decision that shapes data governance, auditability, and the ultimate fidelity of the self-concept model.

4. System-Level Trade-offs: Latency, Accuracy, and Privacy

Every design choice in a digital twin recommendation system is governed by a multi-dimensional trade-off space. The first axis, latency versus model depth, is well-known from production recommender systems. However, enriching the model with self-concept variables that require introspective inference, such as estimating which possible self is active in a given session, amplifies this tension. Pre-computing identity-aware embeddings and serving them from edge caches can mitigate latency, but stale embeddings degrade the system's ability to react to rapid context shifts, such as mood changes or immediate social influence. The system must therefore implement staleness budgets and adaptive refresh strategies informed by the volatility of different self-concept dimensions.

The second axis, accuracy versus privacy, becomes particularly acute when the digital twin draws upon intimate psychological signals. The more granular and longitudinal the data, the more accurate the representation of possible selves, but the greater the risk of sensitive inference and re-identification. Techniques such as differential privacy, secure multi-party computation, and on-device federated learning have been proposed to bound these risks [10, 11]. Yet applying them to self-concept modeling is non-trivial because the utility of the model hinges on preserving the very subjective textures that privacy mechanisms are designed to obscure. Overly aggressive noise injection can flatten the nuanced distinction between an aspirational self and a feared self, rendering the twin no more useful than a conventional profile. System designers must therefore calibrate privacy budgets in collaboration with domain experts who understand the psychological structures being protected, a governance practice that is rarely realized in contemporary industrial settings.

A further trade-off involves fairness and representational equity. Self-concept models trained predominantly on data from digitally engaged, affluent populations may fail to capture the identity structures of marginalized groups whose self-narratives are shaped by systemic constraints. This can lead to recommendations that reinforce existing disparities, for example by suggesting aspirational products that are economically inaccessible or by failing to represent communal value systems that differ from individualistic preference frameworks [7, 12]. Addressing these issues requires not only algorithmic fairness interventions but also infrastructural commitments to representative data collection and participatory labeling of self-concept dimensions, which in turn introduces additional cost and latency in the data pipeline.

5. Governance, Fairness, and Ethical Deployment

The incorporation of self-concept into recommendation digital twins transforms the system from a passive reflector of taste into an active shaper of identity. The governance frameworks that accompany such systems must therefore address a broader set of ethical hazards than those covered by standard fairness, accountability, and transparency guidelines. A primary risk is that of identity manipulation, where the twin's model of the user's possible self is deliberately nudged toward profiles that maximize platform engagement or commercial conversion rather than user wellbeing. This risk is structural, not merely a matter of bad actor intent, because the objective functions of advertising-driven platforms inherently reward prolonged attention and repeated transactions [6].

Effective governance demands a multi-layered approach. At the infrastructure layer, audit trails must log every substantive alteration to the self-concept representation, capturing the provenance of data that triggered the change and the algorithmic logic that mediated the update. Explainability mechanisms must be extended beyond surface-level recommendation justifications to offer users insight into how their identities are being modeled and which

possible selves are being prioritized [13]. At the organizational layer, independent ethics boards with genuine authority to veto model changes should be established, and external auditing regimes should assess the alignment between modeled possible selves and user-reported values over longitudinal periods.

Cross-domain lessons from digital twins in healthcare and education illuminate governance challenges. In those sectors, strict regulatory frameworks, such as HIPAA in the United States or GDPR in Europe, already constrain the processing of sensitive personal data, and digital twin implementations must demonstrate clinical or pedagogical validity [9, 14]. In the consumer recommendation domain, no equivalent regulatory specificity exists. This regulatory gap places an outsized burden on voluntary corporate governance and academic standard-setting. A promising direction involves adapting the safety-case and impact-assessment methodologies developed for algorithmic systems to specifically evaluate the influence of self-concept modeling on vulnerable populations, including adolescents, individuals with mental health conditions, and economically precarious users [12, 15].

Fairness in self-concept-aware recommendation also intersects with representational harm. When a digital twin internalizes narrow cultural templates of success, attractiveness, or normalcy, it can systematically disadvantage users whose identity constructs diverge from those templates. Mitigating this requires participatory design practices that include diverse users in defining the vocabulary of self-concept dimensions and in validating the fidelity of the digital twin over time. Such practices are costly and slow, running counter to the rapid iteration cycles of industry, but they are essential if the technology is to avoid becoming a vehicle for homogenization masquerading as personalization.

6. Robustness, Adaptation, and Long-Term Sustainability

A digital twin that models human self-concept must be robust to several classes of failure that are distinct from conventional recommendation system vulnerabilities. Concept drift in user identity is inevitable: major life events, developmental transitions, or therapeutic interventions can rapidly reshape an individual's self-concept in ways that violate the stationarity assumptions of many machine learning models. The system must therefore incorporate mechanisms for detecting and adapting to regime changes in identity structure, potentially drawing on the concept drift literature that emphasizes adaptive windowing, ensemble diversity, and gradual forgetting [16]. However, unlike standard supervised learning contexts where drift is detected through prediction error, drift in self-concept may manifest subtly in shifts of the distribution over possible selves, requiring unsupervised or self-supervised drift detectors.

Robustness also encompasses resilience against adversarial manipulation. Attackers could poison the training data streams of a digital twin to induce harmful self-concept representations, for example by injecting content that reinforces negative self-perceptions or addictive cycles. Research on shilling attacks in recommender systems provides a partial blueprint, but the stakes are higher when the attacked artifact is a person's digital identity rather than a product ranking [15]. Defenses must operate at multiple levels: robust aggregation of behavioral signals with anomaly detection, cryptographic identity-binding to prevent unauthorized twin updates, and human-in-the-loop verification for major shifts in self-concept representations.

Long-term sustainability of such systems poses an underappreciated challenge. The computational and carbon costs of maintaining continuously synchronized, deep-learning-

driven digital twins for millions of users are substantial, raising questions about the environmental footprint of hyper-personalization. A single rich digital twin that ingests multimodal data and frequently recomputes identity representations can consume orders of magnitude more resources than a static profile-based recommender. Sustainable AI principles, including model compression, sparse update strategies, and energy-proportional infrastructure scaling, must be integrated into the design from the outset [11]. Moreover, data retention policies must balance the continuity necessary for longitudinal self-concept modeling against the privacy and security risks of indefinite storage, a tension that current data minimization regulations address only superficially.

7. Policy Implications and Infrastructure Design

The deployment of self-concept digital twins in recommendation systems sits at the intersection of multiple regulatory and infrastructural regimes. Existing data protection frameworks, such as the GDPR, establish rights to access, rectification, and erasure of personal data, but do not clearly articulate how these rights translate to inferred psychological constructs like possible selves. If a user exercises the right to erasure, should the entire identity graph be deleted, or only the behavioral traces from which it was derived? The answers have profound architectural implications, mandating data lineage tracking and the ability to retroactively excise portions of the twin without destabilizing the entire model [17].

Beyond privacy regulation, emerging AI governance instruments increasingly require impact assessments for high-risk AI systems. The European Union's proposed AI Act, for instance, classifies certain uses of AI that manipulate human behavior as prohibited or high-risk. A self-concept digital twin that makes recommendations designed to shape long-term identity could plausibly fall under such scrutiny, especially if deployed to vulnerable populations. Infrastructure designers must therefore build in compliance capabilities from the start, including model documentation formats such as model cards and datasheets that disclose the psychological constructs being modeled and the validation procedures employed [12, 18].

Infrastructure design for self-concept-aware recommendation also has geopolitical dimensions. The concentration of digital twin infrastructure within a small number of multinational technology firms creates data sovereignty risks, as the intimate identity data of entire populations becomes subject to foreign legal regimes. This has motivated proposals for national or regional digital twin commons, where identity models are managed as public-interest utilities with governance structures that include citizen representatives, academic researchers, and civil society organizations [19]. Such commons would require interoperable standards for self-concept representation, a line of work that is still in its infancy but could draw on existing semantic web ontologies and W3C provenance specifications.

Finally, the policy conversation must address the asymmetry of power inherent in the digital twin paradigm. When a platform possesses a high-fidelity model of a user's self-concept, including their insecurities and aspirations, it operates with an informational advantage that far exceeds that of the individual. Counterbalancing this requires not only transparency but also mechanisms for user agency, such as interactive dashboards that allow users to explore, edit, and simulate alternative versions of their digital twin. Research in human-computer interaction has begun to prototype such tools, demonstrating that they can enhance user understanding and trust [20], but scaling them to production-grade reliability while maintaining real-time synchronization remains an open systems challenge.

8. Conclusion

The convergence of digital twin technology, artificial intelligence, and recommendation systems opens a pathway toward personalization that respects the full depth of human identity. Modeling preferences alongside self-concept promises to transform recommender engines from shallow predictors of next-click behavior into partners in identity navigation. Realizing this vision, however, demands a comprehensive systems perspective that treats architecture, governance, fairness, robustness, and sustainability as interconnected design dimensions rather than afterthoughts to algorithm development. This paper has argued that rich self-concept modeling introduces structural tensions: latency must be managed without sacrificing fidelity, privacy protections must be calibrated to preserve psychological nuance, and fairness interventions must account for the cultural embeddedness of identity. Governance frameworks must evolve from generic AI ethics principles to domain-specific accountability structures that constrain how platforms deploy models of user identity. Infrastructure choices, from stream processing topologies to federated learning architectures, are not merely technical but fundamentally shape the distribution of power, risk, and agency between users and platform operators. Future research should pursue cross-disciplinary validation of self-concept representations, develop standardized benchmarks for identity-aware recommendation, and prototype governance mechanisms that are both technically enforced and democratically legitimate. The next generation of recommendation infrastructure will not only require faster models or larger datasets but a deeper engagement with the question of what it means to represent a person inside a machine, and how that representation can be marshaled in service of human flourishing rather than extraction.

References

1. Grieves, M., & Vickers, J. (2017). Digital twin: mitigating unpredictable, undesirable emergent behavior in complex systems. In F.-J. Kahlen, S. Flumerfelt, & A. Alves (Eds.), Transdisciplinary perspectives on complex systems (pp. 85–113). Springer.
2. Ricci, F., Rokach, L., & Shapira, B. (2011). Introduction to recommender systems handbook. In F. Ricci, L. Rokach, B. Shapira, & P. B. Kantor (Eds.), Recommender systems handbook (pp. 1–35). Springer.
3. Markus, H., & Nurius, P. (1986). Possible selves. American Psychologist, 41(9), 954–969.
4. Kobsa, A. (2001). Generic user modeling systems. User Modeling and User-Adapted Interaction, 11(1-2), 49–63.
5. Zhang, S., Yao, L., Sun, A., & Tay, Y. (2019). Deep learning based recommender system: A survey and new perspectives. ACM Computing Surveys, 52(1), 1–38.
6. Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. Science, 347(6221), 509–514.
7. Barocas, S., Hardt, M., & Narayanan, A. (2019). Fairness and machine learning. fairmlbook.org.
8. Semeraro, C., Lezoche, M., Panetto, H., & Dassisti, M. (2021). Digital twin paradigm: A systematic literature review. Computers in Industry, 130, 103469.
9. Tao, F., Zhang, H., Liu, A., & Nee, A. Y. C. (2019). Digital twin-driven product design, manufacturing and service with big data. The International Journal of Advanced Manufacturing Technology, 94(9-12), 3563–3576.

10. Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning: Concept and applications. ACM Transactions on Intelligent Systems and Technology, 10(2), 1–19.

11. van Wynsberghe, A. (2021). Sustainable AI: AI for sustainability and the sustainability of AI. AI and Ethics, 1(3), 213–218.

12. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Schafer, B. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines, 28(4), 689–707.

13. Zhang, Y., & Chen, X. (2020). Explainable recommendation: A survey and new perspectives. Foundations and Trends in Information Retrieval, 14(1), 1–101.

14. Liu, Y., Zhang, L., Yang, Y., Zhou, L., Ren, L., Wang, F., ... & Deen, M. J. (2019). A novel cloud-based framework for the elderly healthcare services using digital twin. IEEE Access, 7, 49088–49101.

15. Adomavicius, G., & Zhang, J. (2012). Impact of data characteristics on recommender systems performance. ACM Transactions on Management Information Systems, 3(1), 1–17.

16. Gama, J., Zliobaite, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. ACM Computing Surveys, 46(4), 1–37.

17. Solanki, D., Hsu, H. M., Zhao, O., Zhang, R., Bi, W., & Kannan, R. (2020, July). The way we think about ourselves. In International Conference on Human-Computer Interaction (pp. 276-285). Cham: Springer International Publishing.

18. Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2), 2053951716679679.

19. Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389–399.

20. Mobasher, B., Burke, R., Bhaumik, R., & Williams, C. (2007). Toward trustworthy recommender systems: An analysis of attack models and algorithm robustness. ACM Transactions on Internet Technology, 7(4), 23-es.

21. Tao, F., Zhang, M., Liu, Y., & Nee, A. Y. C. (2018). Digital twin driven prognostics and health management for complex equipment. CIRP Annals, 67(1), 169–172.