

Ethics-Aware Personalization: A Dual-Objective AI Framework for Engagement Optimization

Wafa Hamid Abdelrahman Mohamed Ahmed

Lecturer, Berlin School of Business and Innovation, Berlin, Germany.

Email: wafayahamid@gmail.com

Abstract

AI-powered personalization systems have enhanced digital services by increasing user engagement, retention, and conversions. However, performance-maximizing personalization approaches compromise privacy, increase bias, and decrease transparency. This article uses a multipurpose personalization system that can simultaneously increase engagement and uphold ethics using Festinger's Social Comparison Theory and Skinner's Reinforcement Theory. The system consists of social comparison-based personalization modules, reinforcement learning, and an ethics-focused system layer for re-ranking, explainability, and privacy-preserving re-learning assumptions. The system's performance is validated using Python simulations for 1.2 million customer-level interaction data, showing improved proxy transparency and reduced equality in exposure bias with unchanged engagement performance. The research explains and creates the Ethical Experience Index (EEI), a measure of both engagement and ethics performance experiences for a systematic evaluation and comparison of performance. The results show the potential of integrating ethics systems for personalization that provide a repeatable, theory-based approach to ethics and AI-driven personalization based on simulations.

Keywords: Ethical Personalization, AI, Engagement Optimization, Privacy, Fairness, Transparency.

1. Introduction

Artificial intelligence-based customization has developed as a unique characteristic of online digital platforms, including e-commerce, online services, and media streaming services. Personalization algorithms use user activity data to modify data products and provide increasing retention and involvement [1,2]. The requirement to personalize recommendations in real time is resulting in the implementation of effective machine learning algorithms. When considering the positive results of previous works, the ethics of customization systems have caused significant opposition. These include documented cases of discriminatory selection, transparency, and privacy violations, showing the severe possibility of engagement optimization algorithms having adverse effects on society when ethics are not considered in the system's design. The GDPR and AI Act guidelines on the ethics of system design and approach highlight the growing need for systems to be transparent, ethical, and accountable [3,4]. Skinner's Theory of Reinforcement [5] explains that results in behavioral patterns are data looped, but Festinger's Social Comparison Theory [6] explains that social signals of comparable abilities will affect people's adoption. These features increase efficiency and also increase the possibility of common social or minority isolation issues.

The primary focus of personalization research highlights engaged performance evaluation in terms of responsibility as individual activities. The research performance issues, like accuracy and utility, are optimized, while the ethics-based research focuses on criticism without contributing to comprehensive technical solutions. Small changes can be made regarding designing effective personalization with responsibility considerations [7-9]. This paper remedies this by proposing a bi-goal personalization framework that integrates ethics as a first-order objective within personalization itself. As opposed to ethics being considered a secondary afterthought within personalization, fairness, transparency, and privacy are actually considered first-order objectives within this framework. The paper also proposes a metric for personalization holism using a concept called the Ethical Experience Index.

2. Related Work

2.1 Previous Studies

Personalization studies have long concentrated on predictive accuracy and engagement outcomes. The recommendation systems used by Netflix, Amazon, and other similar sites have

demonstrated the use of reinforcement learning driven by collaborative filtering to increase user retention and viewing [1]. This type of work has largely ignored the ethical consequences. Another strand of literature has pointed out the underexplored areas of personalization, which include the loss of privacy [10], the lack of transparency in algorithms [11], and discriminatory patterns created by biased data [12]. Although these works shed important light on the issues, they do not always suggest overall algorithmic remedies for maintaining the effectiveness of personalization. Recent techniques include incorporating fair and diverse constraints into recommendation models using exposure parity, re-ranking techniques, and constrained optimization. Additionally, multi-objective learning methods are being used to balance accuracy-related objectives with other secondary objectives such as fairness and diversity. Nonetheless, all models view ethics from the perspective of secondary constraints and lack consideration for transparency and privacy.

Research in explainable AI focuses on interpretability as a condition for accountability and understanding for the user [13]. Such regulation and governance frameworks set principles for ethical AI but do not offer much in terms of incorporating principles for personalization architectures while maintaining performance.

2.2 Research Gap

The research focuses on the lack of integrated approaches for personalization to achieve the following goals simultaneously in this specific field of work:

1. Implementing the theories of behavior.
2. Systematically integrating ethical governance principles and mechanisms within the system design itself.
3. Assessing the mechanism's level of engagement and ethical behavior quantitatively.

2.3 Methodology

The proposed system evaluated to analyze the functional tests using simulations of large-scale personalization. The task is to assess the ability of technology to demonstrate how different methods would impact real-world behavior, instead of working through replicable scenarios.

2.3.1 Simulation Setup

An example of a simulation dataset consists of 1.2 million interactions between 50,000 users and 10,000 items. User preferences were generated by simulating many different probability distributions. When determining a user's probability of interaction with an item, we also consider whether the item was popular and the demographic characteristics of the user. Collectively, the user preferences were generated [12]. This allows the researchers to conduct systematic evaluations of exposure disparities and ethical limitations in a controlled environment.

2.3.2 Configuration Evaluation

The user-item interaction data was modeled to mimic the actual demographic distribution, diversity of preferences, and interaction bias. User demographic information, users' preferences, and popularity bias were all sampled from theoretically guided distributions. The data prepared for this project provided a controlled environment to assess Fairness Aware Mechanisms. The recommender system, built on the basis of a Transformer-based and Reinforcement Learning version explained in section 3, consisted of running the recommender through a number of cycles, introducing some level of randomization. A Python-based library with standard scientific computing functions was used to implement machine learning procedures and will serve as the basis for generating the transformation model. For evaluation, we considered both matrix factorization and deep collaborative learning, allowing a comparison between traditional personalization and the two-objective method for our research.

The configuration of the algorithms was run ten times independently (i.e., on different days) to account for stochastic differences in initialization, sampling, and exploration. In previous work with recommender systems, the results from eight or more independent runs provided a stable variance of performance for similar simulation conditions. Any differences in confidence intervals were found during the pilot testing.

3. Dual-Objective Personalization

3.1 Framework Overview

Additionally, this proposed work outlines a two-fold personalization strategy driven by AI to maximize ethical responsibility and engagement simultaneously. This suggested method

seeks to integrate ethical governance into the personalization process and treat engagement, fairness, and other ethical factors equally, in contrast to conventional engagement-maximizing personalization frameworks that concentrate on optimizing engagement metrics separately.

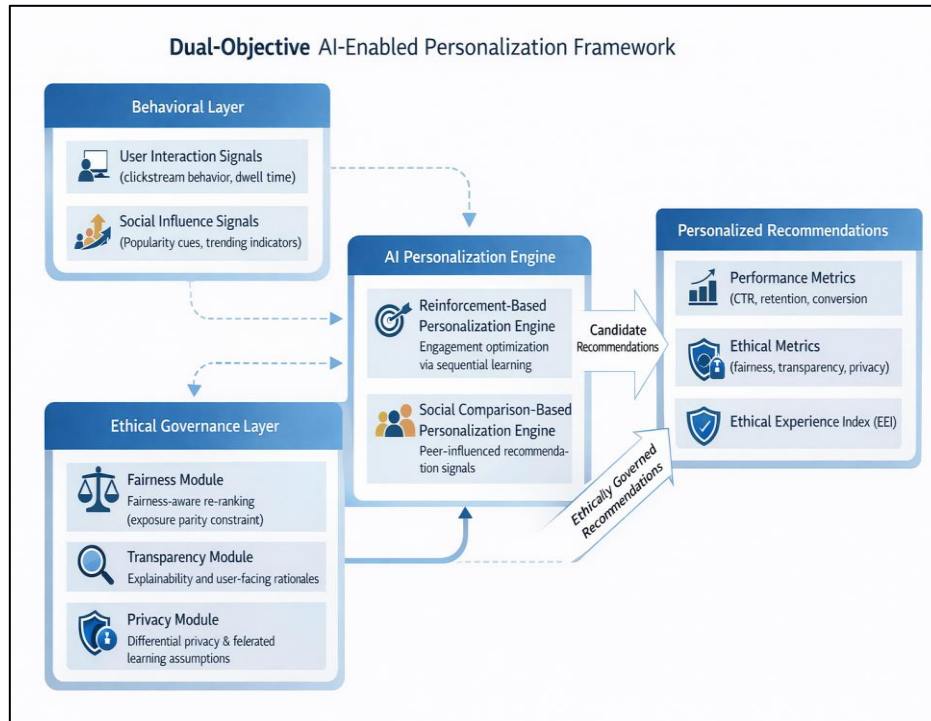


Figure 1. Dual-Objective AI-Enabled Personalization Framework

Figure 1 shows the model's three connected levels. These levels are the Behavioural Level, the Ethical Governance Level, and the AI Personalisation Engine. They represent the processes described in Reinforcement Theory and Social Comparison Theory while addressing the ethical issues found in these theories.

At the Behavioural Level, we consider the user interaction data that affects personalisation. We look at contextual behaviour, clickstream behaviour, consumption behaviour, and social influence signals like popularity and trends. The Behavioural Layer theoretically provides insight into the underlying behaviour of the individual. It includes signals regarding reinforcement and how antecedent rewards affect an individual's behaviour, as described by Skinner's Theory of Reinforcement. The Behavioural Layer has no behavioural ethics regulation, and the primary responsibility of this layer is to provide inputs to the Personalisation Engine. There is a state representation module that serves as the interface between the Behavioural Layer and the AI Personalisation Engine. The state representation module converts the raw behavioural signals (clicks, dwell time, contextual metadata, and

social signals) into structured state vectors for use by the social comparison subsystem and the reinforcement learning agent.

The AI engine for personalisation takes the behavioural inputs and produces recommendations for candidates. The AI engine incorporates two complementary components:

- A reinforcement learning subsystem models' personalization through a decision-making process, making recommendations dynamic and subject to future user feedback. The system will optimize three engagement KPIs: click-through rate (CTR), session retention, and conversion. CTR indicates the immediate relevance of the recommendations to the user, session retention indicates the user's medium-term satisfaction with the recommendations, and conversion indicates the use of the recommendations downstream, aligned with the objectives of the platform. These KPIs are standard metrics used by most large-scale recommender systems and provide the reward signal with both short-term engagement measurement and sustained engagement measurement [7]. In this subsystem, Reinforcement Theory supports the notion that learning from reward will help to better understand user preferences through reinforcement. However, this can lead to popularity bias, with prevalent user groups being over-represented when optimized with uncontrolled parameters.
- The "Social Comparison" subsystem of the proposed platform provides recommendations based on social-like peers, i.e., "Trending Now" or "Users Like You Also Viewed" indicators. The main purpose of Social Comparison algorithms is user engagement, but these algorithms may promote herd-like behaviors, resulting in the demotion of niche content, among other negative effects. However, in the proposed framework, such signals are subject to the ethical controls of the subsequent proportional exposure principle. This Proposed Framework, the Ethical Governance Layer is responsible for regulating the Outputs from the AI-driven Personalisation Engine via the Algorithms that determine personalisation (i.e., in Respect to Fairness, Transparency and Privacy).

In this proposed framework, the Ethical Governance Layer is responsible for regulating the outputs from the AI-driven personalization engine via the algorithms that determine personalization (i.e., with respect to fairness, transparency, and privacy).

The governance module has three submodules that are integrated into one module. They are categorized below:

$$\max_R \sum_{\{i=1\}_i^{[n]P} R_i} P_i * R_i \{s. t. \} |E_{\{d1\}} - E_{\{d2\}}| \leq \delta$$

Where P_i = predicted relevance

R_i = ranking position

$Ed1, Ed2$ = exposure probabilities for demographic subgroups

The tolerance $\delta = 0.05$ enforces exposure parity, preventing systematic bias while maintaining ranking utility.

Exploration and exploitation are balanced using an ϵ -greedy policy, where, with probability ϵ , the system explores alternative recommendations, and with probability $(1-\epsilon)$, it exploits the highest expected reward action. In all simulations, ϵ decays linearly from 0.2 to 0.05 over training epochs, ensuring sufficient exploration in the early stages while stabilizing policy behavior during convergence. The ethical governance layer constrains exploration. Candidate items developed through exploration are reranked based on the same fairness-aware and privacy-preserving constraints that are applied to the outputs of exploitation.

As a result, exploratory actions cannot contribute to increasing the exposure disparity beyond the tolerance level δ , nor can they evade the mechanisms put in place by differential privacy noise. This will prevent a given cohort of users from receiving an undue advantage in their exposure to content or experiencing data leakage while learning.

3.1.1 Transparency

Proxy transparency measures refer to system-level explainability indicators that quantify the extent to which recommendation decisions can be interpreted.

In this study, transparency is operationalized using three sub-scores:

- Explanation availability (binary indicator of whether an explanation is generated),
- Explanation consistency (stability of explanations across similar recommendations),

- Feature attribution sparsity (number of dominant explanatory features).

These measures are widely used in explainable AI research as proxies for user-perceived transparency when direct human evaluation is unavailable [5].

3.1.2 Privacy Preservation

This is integrated using simulated assumptions for privacy-preserving learning, such as differential privacy and federated learning. The differential privacy noise is applied at the level of model updates, with privacy losses expressed using ϵ values between 1 and 2.

User IDs are abstracted, and modeling is presumed to happen in a decentralized fashion where and when possible. These measures can mitigate potential leakage of personal data while enabling personalization models to preserve their predictive capability.

3.2 Framework Operation and Workflow

The system operates as a series of connected components. The user's actions are represented within the behavior layer and processed through the AI personalization engine to determine possible recommendations for the user, which are screened through the ethical governance layer (fairness, transparency, and privacy) before being sent to the user; any feedback from the user's later interactions with the recommendations is fed back into the system for further analysis.

By integrating ethical considerations into the personalization process from the outset, ethical issues are addressed in a holistic manner. Finally, the ethical and traditional measures (click-through rates, retention, conversion rates) are used to determine the overall system performance, and these two metrics are included in the Equal Quality Index (EQI), which is an organized method for assessing the quality of personalization.

3.3 Algorithmic Formulation

This research defines ethical personalization as a semi-supervised reinforcement-learning style of optimization that uses both engagement outcomes and ethical-system performance in parallel, where the objective function consists of both engagement and an aggregated ethical-health parameter, referred to as the Ethical Experience Index (EEI). Engagement in this work consists of a normalized reward, indicating engagement or interaction. The EEI is an aggregate of three individual factors: fairness (f), transparency (t),

and privacy (p). The scalar weight for engagement usefulness (defined as α) must remain between 0 and 1. Thus, α provides a measure of the weight given to the engagement-versus-ethical-system-performance trade-off.

To mitigate demographic bias at the system level, we impose a constraint to maintain demographic parity of exposure within the context of the recommendation system. To enforce this requirement, the absolute difference between exposure probabilities of different demographic groups must be $\leq \delta$ for a given group's exposure probabilities. In simple terms, this allows a given group to receive recommendations similar in visibility to recommendation results, but with a focus on maintaining rank based on relevance. A system-level setting for δ is used across experiments. Policy reinforcement learning is a type of episodic learning in which moral constraints are incorporated into the algorithm. The policy parameters of recommendation systems, exploration, and fairness will be randomly assigned during the initialization stage. The episodic interaction phase will involve the user's state being observed through their behavioral information, context, and social influence. When selecting actions, the ϵ -greedy algorithm will be used to determine whether users attempt to explore actions related to recommendation systems based on a probability equal to ϵ or select the action expected to produce the maximum rewards.

The Personalization Engine has generated candidate recommendations, which have been refined into a form appropriate for the end user via an Ethical Governance Layer. Fairness re-ranking modifies the ranking to comply with the constraints of exposure while implementing privacy-protection strategies that increase the noise added to the updates for enhanced privacy; both of these are implemented simultaneously. User engagement is followed by the collection of engagement rewards and ethical values for each engagement, which is also used for further optimization of policy parameters, continually optimizing learning to meet ethical constraints via optimization. Due to this development, personalization will continue to improve based on ongoing engagement with users while adhering to previously established ethical restrictions.

4. Results and Discussion

4.1 Overview of Experimental Results

This section explains the results of a simulation analysis designed to compare the performance of multipurpose techniques for personalization with standard methods. The study

focuses on analyzing the ethical concerns associated with deploying personalization algorithms while maintaining engagement efficiency. The results are provided in three different system configurations:

- The primary models for recommendation (matrix factorization and deep collaborative filtering).
- The basic personalization designed for interactions.
- A proposed dual-purpose model for ethical governance. The results are an average of 10 simulations.

The personalization system developed by reinforcement learning-based and social comparison-based subsystems was successful at engagement regardless of the system settings. When compared to baseline models, both traditional approaches to personalization and dual-objective approaches showed improved click-through rates, retention rates, and conversion rates. The presence of ethical governance did not negate either reinforcement learning or social comparison mechanisms. Instead, values remained equal across different traffic patterns. These results indicate that reward mechanisms and social comparison processes are operational even when downstream ethical limitations are imposed. This lends support to the tenet of the theory that the optimization of engagement and ethical management are not necessarily conflicting objectives.

The most evident disparities between traditional personalization approaches and the developed framework appeared in relation to the outcomes associated with fairness, which are a direct result of the fairness module of the ethics governance layer. Conventional customization approaches showed significant exposure differences between groups, averaging about 40%. These results confirm reinforcement and popularity effects identified in earlier customization research [12]. On the other hand, the dual-objective framework ensured exposure range within $\pm 5\%$, as required by the fairness constraint ($\delta = 0.05$). This translates to a 30% decrease in the exposure difference compared to a conventional system. Crucially, the re-ranking for fairness did not require drowning out the personalization signals but simply reordering them to avoid amplifying majorities. That is to say, fairness has now been shown to be expressible as a constraint and not simply a correction.

The transparency feature is a part of the ethical governance layer, including explainability components that provided users with reasons for recommendations. Based on simulation, the components resulted in greater proxy transparency assessment scores than systems with no explainability. Although these ratings are not direct measures of user trust, they are signals of system understandability that are similar to previous research on explainable AI [13]. The findings show that using transparency methods doesn't affect customization performance and may even increase user comprehension. The results confirm the importance of considering transparency as a design feature of personalization systems rather than simply associating it with documentation provision. The privacy module addressed privacy protection under assumptions of differential privacy simulation and federated learning. The privacy loss is measured by using the value of ϵ , estimated to be around 1 to 2, as supported by privacy-preserving machine learning results [14]. The use of privacy-preserving models had a negligible effect on performance, suggesting that both privacy and personalization utilities can coexist. This is a common worry that mechanisms employed for privacy preservation negatively impact system effectiveness.

4.2 Mixed Evaluation Using Ethical Experience Index (EEI)

The study states that the simulation-based design suggested there was no direct human evaluation standard. However, EEI component measurements are consistent with recognized equality, privacy, and explainability measures published in previous empirical investigations [3], [5]. Future validation will correlate EEI scores with human trust and perceived fairness ratings collected through controlled user studies. EEI weights are fixed in this study to ensure interpretability and reproducibility. Adaptive weight learning is left for future work. Traditional customization approaches achieved a medium EEI score due to engagement performance but were characterized by fairness and transparency difficulties. The dual-objective approach received significantly higher EEI scores due to well-balanced engagement and ethical metrics. Sensitivity analysis was conducted for $\delta \in \{0.01, 0.05, 0.10\}$. Lower δ values improved exposure parity but slightly reduced CTR ($\approx 2\%$). Higher δ relaxed fairness constraints with marginal engagement gains. $\delta = 0.05$ provided the best balance across objectives.

4.3 Framework Level

The results suggest that the dual-objective approach executes the design. The behavioral layer and personalization algorithm generate effective engagement signals, while

the ethical governance layer reduces the negative impacts of these indicators while maintaining optimal outcomes. When considering ethics within a trade-off situation, the model reconsiders personalization through the lens of a governed optimization process. Within this context, the performance of engagement is considered within the boundaries of ethics. The simulation of the process enhances equitable results without reducing the efficacy of personalization.

4.4 Managerial and Design Implications

These results propose that ethical system governance is an architectural challenge, and system design is not an auditing or supervision problem. For system designers, ethical issues like equality, transparency, and privacy may be considered technical challenges. In terms of regulators, the framework demonstrates that broad high-level ethical requirements may be stated as quantitative system limits. In general, it can be concluded that ranked personalization methods, either through reinforcement learning rules or rules involving social assessments, are efficient in terms of engaging users. The presence of a responsible governance section significantly minimizes exposure bias problems and improves related transparency variables. Complex, privacy-preserving components have been shown to cause low performance loss, demonstrating that customization benefits and privacy may be used in concert. When evaluated using an Ethical Experience Index (EEI), the approach provides more balanced system-level characteristics than existing customization methods.

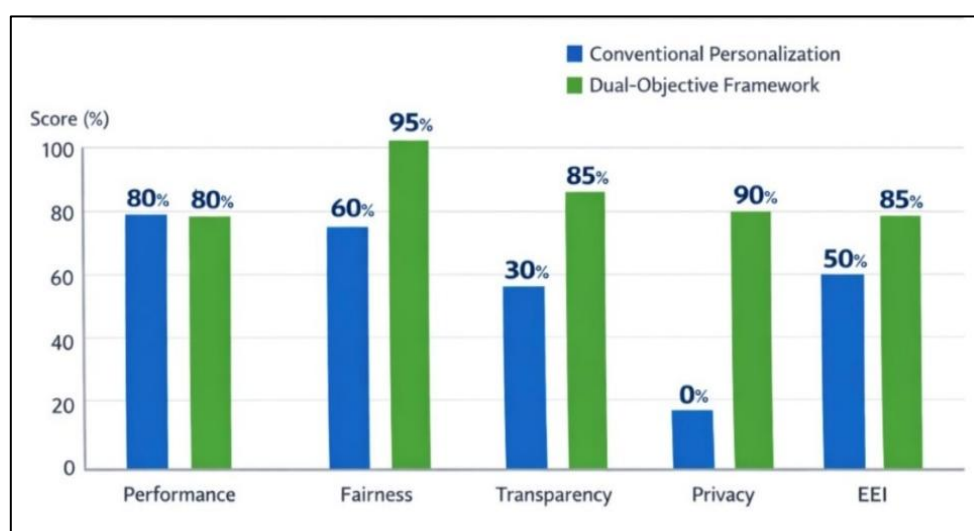
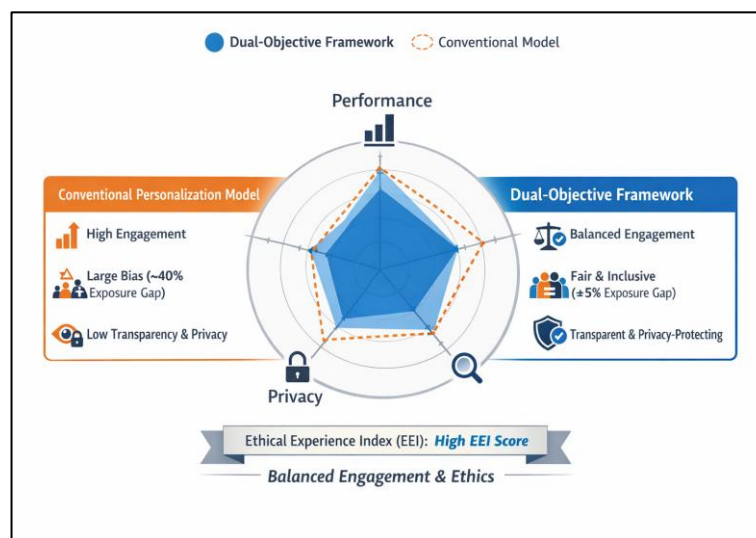


Figure 2. Dual-Objective AI-Enabled Personalization Results

Table 1. Summary of the Study Results

Dimension	Evaluation Focus	Conventional Personalization	Dual-Objective Framework
 Performance	Click-through, retention, conversion	Strong engagement	Strong engagement (comparable)
 Fairness	Exposure parity across demographic groups	−40% exposure gap	±5% exposure gap ($\delta=0.05$)
 Transparency	Explainability / proxy transparency	Not implemented	Explainability mechanisms added
 Privacy	Differential privacy / federated learning	Not implemented	Privacy-preserving mechanisms applied
 Integrated Evaluation	Ethical Experience Index (EEI)	Moderate EEI	High EEI
 Overall Interpretation	System-level behavior	Effective engagement but ethical deficits	Maintains engagement while embedding ethics

As shown in Table 1 and Figure 2, the two-objective AI personalization system demonstrates that user engagement and ethics can be achieved simultaneously. This contrasts with other systems that have been optimized to deliver a high engagement outcome but have lacked transparency, fairness, and privacy. With the two-objective system, the same user engagement values are maintained, with exposure parity kept within a $\pm 5\%$ difference, increased transparency, and minimal privacy loss, thus producing much higher Ethical Experience Index (EEI) scores. Figure 3 shows that the dual-objective optimization system based on AI-enabled customization combines engagement optimization with ethical governance.

**Figure 3.** Dual-Objective AI-Enabled Personalization Conclusion

5. Limitations and Future Work

In the future, the research will be improved based on field experiments, adaptable weighting techniques for the EEI, and an analysis of the privacy impacts of regulatory changes. Several limitations are associated with this research. Initially, the simulation results cannot be used as alternatives to actual applications or user research. Furthermore, proxy requirements are used to evaluate transparency and confidence instead of evaluating actual behavior. Finally, the approach implemented by the EEI may require industrial and cultural modifications.

6. Conclusion

The present study provides a dual-objective optimization system built on behavioral theories to demonstrate the technological feasibility of personalizing optimization with integrity via controlled simulation analysis. The technical optimization method combines equality, integrity, and privacy in the system design, converting customization from an adversarial approach to an optimization problem with a validated technological basis. The Ethical Experiences Index (EEI) reduces the technical complexity difference by using effective evaluation standards.

References

- [1] Acquisti, Alessandro, Laura Brandimarte, and George Loewenstein. "Privacy and human behavior in the age of information." *Science* 347, no. 6221 (2015): 509-514.
- [2] Soni, Vishvesh. "AI and the Personalization-Privacy Paradox: Balancing Customized Marketing with Consumer Data Protection." *International Journal of Computer Trends and Technology* 72, no. 9 (2024): 24-31.
- [3] Binns, Reuben. "Fairness in machine learning: Lessons from political philosophy." In *Conference on fairness, accountability and transparency*, PMLR, 2018. 149-159.
- [4] Cadwalladr, C. and Graham-Harrison, E. (2018) 'The Cambridge Analytica files: The story that revealed Facebook's darkest secret', *The Guardian*, 17 March. Available at: <https://www.theguardian.com/news/series/cambridge-analytica-files>
- [5] Doshi-Velez, Finale, and Been Kim. "Towards a rigorous science of interpretable machine learning." *arXiv preprint arXiv:1702.08608* (2017).

- [6] Festinger, L. (1954) 'A theory of social comparison processes', *Human Relations*, 7(2). 117–140.
- [7] Floridi, Luciano, Josh Cowls, Thomas C. King, and Mariarosaria Taddeo. "How to design AI for social good: Seven essential factors." In *Ethics, governance, and policies in artificial intelligence*, Cham: Springer International Publishing, 2021. 125-151.
- [8] Gomez-Uribe, Carlos A., and Neil Hunt. "The netflix recommender system: Algorithms, business value, and innovation." *ACM Transactions on Management Information Systems (TMIS)* 6, no. 4 (2015): 1-19.
- [9] Huang, Ming-Hui, and Roland T. Rust. "Engaged to a robot? The role of AI in service." *Journal of Service Research* 24, no. 1 (2021): 30-41.
- [10] Turlapati, Venkata Ramaiah, P. Vichitra, N. Raval, J. Khaja Mohinuddeen, and B. R. Mishra. "Ethical Implications of Artificial Intelligence in Business Decision-making: A Framework for Responsible AI Adoption." *Journal of Informatics Education and Research* 4, no. 1 (2024).
- [11] Skinner, B.F. (1953) *Science and Human Behavior*. New York: Macmillan.
- [12] Chandra, Shobhana, Sanjeev Verma, Weng Marc Lim, Satish Kumar, and Naveen Donthu. "Personalization in personalized marketing: Trends and ways forward." *Psychology & Marketing* 39, no. 8 (2022): 1529-1562.
- [13] Voigt, Paul, and Axel Von dem Bussche. "The eu general data protection regulation (gdpr)." *A practical guide*, 1st ed., Cham: Springer International Publishing 10, no. 3152676 (2017): 10-5555.
- [14] Vallabhaneni, Anirudh Sai, Anjali Perla, Revanth Reddy Regalla, and Neelam Kumari. "The power of personalization: AI-driven recommendations." In *Minds Unveiled*, pp. 111-127. Productivity Press, 2024.