Beyond Green Nudging: A Behavioral Ethics Framework for AI-Driven Sustainability in Smart Cities

Isslam Alhasan[1][0000−0003−0546−224X]

Zayed University, Abu Dhabi, United Arab Emirates
isslam.alhasan@zu.ac.ae

Abstract. The rise of smart cities has made artificial intelligence (AI) an important part of urban life. Yet, deploying AI in these systems creates a tension between enhancing efficiency and protecting human autonomy and independence. Traditional behavioral economics introduce the concept of nudges, which are known to be clear, straight forward, direct interventions that guide choices while preserving human freedom. In contrast, AI-driven nudges are highly personalized, adaptive, and are often unnoticed by users, raising complex ethical concerns about their influence on daily decision-making. To address this, we propose the BEHAVE framework (Behavioral Ethics for Human-Aware AI in Virtual Environments), which integrates sustainable AI, responsible machine learning, and human-centered design for smart cities. By analyzing applications in traffic control, energy consumption, and waste management, we identify three behavioral paradoxes: the cognitive load paradox, transparency trade-off, and sustainability rebound effect. The framework introduces six guiding principles: Behavioral Transparency, Ethical Timing, Human Agency Preservation, Adaptive Consent, Vulnerability Protection, and Evaluation Continuity to navigate these challenges. A case study in urban traffic management illustrates how ethical AI nudging can effectively promote sustainability while ensuring users retain meaningful control over their choices.

1 Introduction

The integration of artificial intelligence into smart city infrastructure marks a major technological advance of the 21st century. Smart cities promise unparalleled efficiency in resource management, traffic flow, and energy consumption through AI systems that predict, adapt, and influence human behavior at scale [2,3]. However, this evolution presents an ethical dilemma: as AI becomes more personalized and sophisticated in shaping human decisions, how can we ensure these benefits do not come at the expense of human autonomy? Imagine a typical morning in a modern smart city: your smartphone suggests an alternative route based on predicted traffic patterns, your smart thermostat adjusts heating to reduce peak demand, and your waste management app gamifies recycling through

personalized rewards. Each represents a form of *algorithmic nudging*: subtle influences designed to encourage beneficial behaviors. Unlike traditional nudges, these are dynamic, tailored for individuals, and often operate without the user's conscious awareness. This invisibility raises significant ethical concerns as AI-driven nudges can subtly influence decision-making without the user's conscious recognition, potentially undermining autonomy and informed consent [22,23,24]. This paper addresses a gap in research by developing an integrated framework for ethical AI nudging in smart cities. While prior work explores AI ethics, behavioral economics, and smart city technologies individually, few studies examine their intersection, especially regarding sustainability outcomes and long-term behavioral effects. The BEHAVE framework offers both theoretical foundations and practical guidelines for designing AI systems that achieve sustainability goals while protecting individual autonomy.

2 Background: The Evolution from Simple Nudges to Smart Persuasion

The original concept of a "nudge" comes from behavioral economics [1]. A nudge is defined as "any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any option or significantly changing their economic incentive" [1]. Nudging leverages behavioral science principles to encourage beneficial behaviors while preserving deliberative choice and freedom. A classic example involves a cafeteria placing fruit and vegetables at eye level and desserts on lower shelves, making the healthier options more noticeable without eliminating other choices.

Artificial intelligence transforms this model. In smart cities, AI-powered nudges are often integrated with sustainability goals under the banner of "Green AI" [4,5]. Unlike static nudges, AI can tailor interventions and recommendations to individual habits, contexts, and real-time conditions, often without explicit user awareness. For instance, a city might display real-time energy usage comparisons between households through a mobile app, which subtly encouraging conservation by leveraging social comparison mechanisms.

2.1 The Dual-System Framework in Digital Environments

Human decision-making operates through two cognitive systems [6]: System 1 (fast, automatic, emotional) handles approximately 95% of daily decisions, while System 2 (slow, deliberate, rational) is reserved more for complex reasoning. AI nudging systems frequently target System 1 processes, taking advantage of cognitive shortcuts and biases to subtly influence behavior efficiently.

This targeting raises significant ethical concerns. While System 1 interventions can promote beneficial behaviors with minimal cognitive effort, they also create vulnerability to manipulation, particularly when users are unaware of influence attempts [22]. During times of cognitive vulnerability, mental fatigue,

Table 1. Classical vs. AI-powered nudging comparison

Dimension	Classical Nudging	AI-powered Nudging
Transparency	Obvious and visible	Opaque, embedded in algorithms
Scope	Static and general	Dynamic and personalized
Adaptability	Same for everyone	Tailored in real-time
Data Requirements	Minimal	Extensive behavioral tracking
Ethical Oversight	Clear attribution	Complex algorithmic decisions

distraction, or stress, emotional thinking takes over, making individuals more susceptible to automated influences.

Table 1 illustrates the differences between classical and AI-powered nudging approaches. While classical nudges are easily identifiable and universally applied, AI-powered systems operate with lack of transparency and personalization that complicate traditional ethical frameworks.

3 Behavioral Paradoxes in Smart City AI

Through systematic analysis of smart city implementations, we identify three paradoxes that challenge conventional approaches to ethical AI design:

3.1 The Cognitive Load Paradox

AI nudges often strategically target individuals during moments of cognitive vulnerability, when mental resources are used up, attention is divided, or stress levels are elevated [6]. During these periods, System 1 thinking predominates, making quick decisions with minimal effort and reducing rational evaluation capacity.

This creates an asymmetric power relationship where AI systems can detect vulnerable cognitive states through behavioral signals (quick interactions, location-based stress indicators, delayed responses) and time interventions for maximum influence. For example, an app automatically selecting a "green" energy plan as the default leverages users' tendency to accept defaults without any explanations or options, particularly when cognitive resources are limited.

The ethical dilemma emerges because while such targeting can efficiently promote beneficial behaviors, it potentially undermines informed consent and human autonomy decision-making by exploiting natural human limitations.

3.2 The Transparency Trade-off

As AI systems become more personalized and effective, their operations often become less visible and comprehensible to users [8,9]. This inverse relationship

between personalization sophistication and user awareness creates tension in ethical AI design.

Effective personalization requires complex algorithmic processing of user data, behavioral patterns, and contextual factors. The more tailored and context-aware a nudge becomes, the less likely users are to notice they've been influenced. This opacity raises concerns about algorithmic transparency and informed consent, which are foundational principles in ethical AI [10].

The transparency trade-off represents a core challenge: increasing behavioral precision often comes at the cost of user awareness and understanding. Users may be guided by systems they cannot comprehend, potentially compromising autonomy even when interventions promote beneficial outcomes.

3.3 The Sustainability Rebound Effect

A growing concern in AI-driven sustainability efforts involves behavioral rebound effects, patterns where positive actions in one domain psychologically justify less responsible behaviors in another domain [11]. This phenomenon, rooted in moral research, suggests that individuals who perform virtuous actions may subsequently feel entitled to less virtuous choices.

For instance, a gamified recycling app that successfully increases waste sorting behavior might lead users to increase overall consumption, reasoning that their recycling efforts justify additional purchases. Similarly, energy-saving behaviors promoted by smart home systems might be offset by increased energy use in other contexts.

The rebound effect shows a conflict between optimizing a specific part of an AI system and achieving a broader, long-term goal. Instead of focusing on just one metric, it suggests a need to look at the bigger picture and understand how changes in one area can have unexpected consequences in other areas. This raises ethical questions about whether AI nudges truly change behavior or just get people to comply for a short time [13].

4 The BEHAVE Framework

To address these paradoxes, we propose the BEHAVE framework, a comprehensive approach to ethical AI nudging in smart cities. The framework consists of six interconnected principles, each targeting specific aspects of the identified behavioral challenges.

See Figure 1 for a visual summary of the BEHAVE framework.

4.1 Behavioral Transparency

This principle requires making AI intentions, mechanisms, and rationale clear to users [12,20]. Behavioral transparency serves as a foundational requirement for ethical AI-driven nudging, directly addressing the transparency trade-off by ensuring that personalization doesn't compromise user understanding.

Table 2. Mapping Behavioral Paradoxes to BEHAVE Framework Principles

Behavioral Paradox	**BEHAVE Principles Addressing It**
Cognitive Load Paradox AI targets users during moments of cognitive vulnerability, risking exploitation of mental shortcuts.	*Ethical Timing*: Avoid interventions when users are vulnerable. *Human Agency Preservation*: Ensure users retain control despite automation.
Transparency Trade-off Greater personalization reduces user awareness and understanding of AI influence.	*Behavioral Transparency*: Make algorithmic intentions clear. *Adaptive Consent*: Enable ongoing, informed permission for personalization.
Sustainability Rebound Effect Positive action in one domain leads to compensatory negative behavior elsewhere.	*Evaluation Continuity*: Monitor long-term and cross-domain impacts. *Vulnerability Protection*: Prevent exploitation of psychological or demographic vulnerabilities.

Transparency-based nudges offer a workable design response by surfacing the rationale behind algorithmic suggestions. Examples include messages such as "Recommended because you purchased similar items last month" or "Suggested route based on your usual commute and current traffic conditions." These explanations allow users to reconstruct decision paths and assess relevance, enhancing trust while mitigating non-transparent concerns.

Nudges can be categorized based on both their cognitive targeting (automatic vs. reflective) and visibility (transparent vs. non-transparent) [15]. The most ethically problematic category involves non-transparent nudges targeting automatic systems, where intentions remain hidden while choices are subtly engineered. To explain this, let's look at an example of dynamic pricing algorithms. Dynamic pricing algorithms subtly changes the price of products based on your browsing and purchase history and psychological state. This is ethically problematic because it exploits cognitive biases without you even realizing it. In this example, the user's autonomy is undermined because they are not making rational decisions based on all available information. Instead, these choices are being hidden which exploits user's predictable psychological shortcuts for the benefit of the seller.

4.2 Ethical Timing

This principle mandates avoiding interventions during periods of cognitive vulnerability [14]. Ethical timing requires AI systems to recognize and respect users' cognitive states, preventing manipulation during moments of low cognitive effort or high susceptibility. An example of this is a food delivery application that pushes high-calorie, unhealthy options through notifications when a user has been logged in late at night. This is a violation of cognitive vulnerability and the system recognizes this period as such. Users are typically tired and stressed during these late hours and typically their ability to make rational choices are

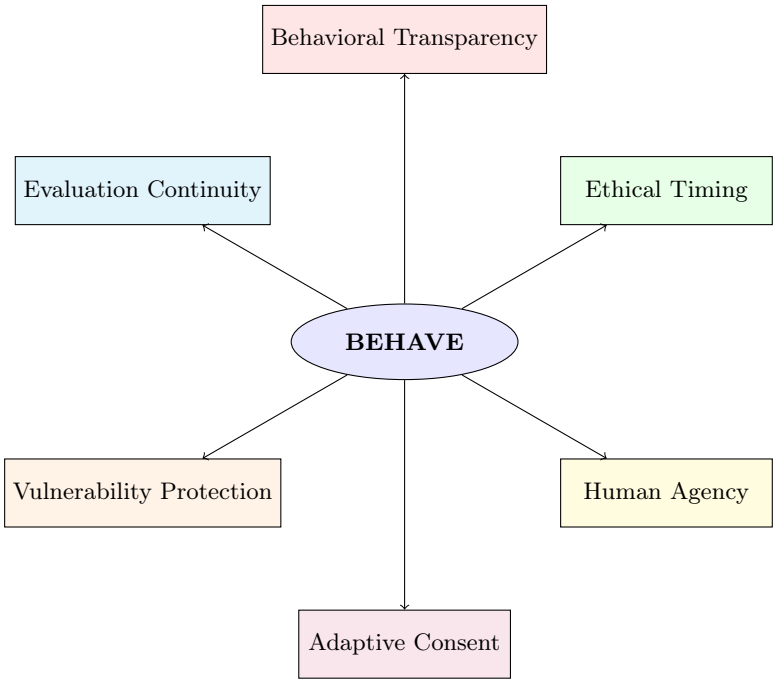

Fig. 1. Visual schematic of the BEHAVE framework principles

diminished. In this example, the system exploits user's weak cognitive state to drive a quick decision.

The FASCAI (Fast and Slow Collaborative AI) framework directly addresses this issue by recommending System 1-based nudges only when absolutely necessary, prioritizing interventions that engage System 2 reflective thinking [16]. The timing of nudge delivery significantly affects whether the intervention qualifies as ethical guidance or manipulation.

Implementation involves developing algorithms that detect stress, fatigue, or distraction indicators and delay non-urgent nudges until users achieve more reflective states and are able to make informed decisions. Some systems implement "confrontational nudges" designed to interrupt impulsive behavior, such as brief delays before posting on social media to encourage reconsideration [17].

4.3 Human Agency Preservation

Human agency preservation reinforces human judgment. This means that AI systems should assist a person in making a decision, and not make the decision for them. As AI is becoming more integrated into our daily lives, it's important that it helps us strengthen our abilities to make decisions instead of taking it away [10]. This principle ensures that AI systems support rather than override user decision-making power. An example of this principle is a GPS application.

The ethical way to approach this is for the system to present you with several route options, each with different relevant information, such as travel time, traffic density, and potential tolls. The app provides users with information, but the user still should have to make the final decision based on the user's priorities.

A major issue exists between personalization efficiency and preserving human autonomy. Highly adaptive systems can anticipate behaviors and tailor interventions accordingly, but may subtly shift control away from individuals. Ethical nudging requires more than offering different choices, it demands that users perceive genuine ownership over decisions.

Research supports autonomy-supportive nudges that guide gently without pressure [18]. Messages like "You may find this option useful, feel free to explore other choices too" maintain system helpfulness while reinforcing user control. Such approaches increase acceptance and follow-through, particularly under high personalization conditions.

4.4 Adaptive Consent

The principle of adaptive consent explains that user permission for AI systems should be continuous and not follow the traditional one-time consent models. The traditional models prove insufficient for adaptive AI systems and calls for adaptive consent where dynamic permission mechanisms are evolved with user needs and system capabilities [19]. As technologies become more personalized and complex, consent must also adapt over time which ensures that users have ongoing control over the intensity of the AI's influence.

For example, a smart energy app beginning with simple thermostat adjustments might later introduce advanced recommendations based on lifestyle patterns. Adaptive consent enables users to accept, modify, or reject these deeper personalization levels, maintaining control over AI influence intensity. To proactively ask users for consent of new levels of personalization and optimization, the system could maybe provide a message saying "Our AI systems can now save you more energy by learning your daily routine. Do you want to enable this deeper personalization?" This approach allows users to accept, modify or even reject the new feature entirely. This method gives control back to the user.

AI systems often rely on continuous data collection and algorithmic feedback, potentially leading to outcomes users didn't initially anticipate or even sign up for. Adaptive consent restores control by making permissions flexible, direct, transparent, and responsive to changes in both system capabilities and user expectations and needs.

4.5 Vulnerability Protection

This principle mandates specific design constraints to prevent exploitation of psychological, social, or demographic vulnerabilities. The EU AI Act explicitly prohibits systems that use personal traits or vulnerabilities to manipulate behavior in harmful ways [20]. This creates legal and ethical obligations for developers to avoid designs disproportionately impacting at-risk individuals.

Examples of problematic targeting include financial apps promoting risky investments to low-income users or health applications collecting children's data without meaningful parental oversight. Research indicates that nudges targeting vulnerable groups can trigger public resistance, particularly when sensitive data misuse is perceived [21].

Instead of applying uniform design logic, AI systems should incorporate specific attention to vulnerable user needs and risks. This includes adapting interfaces, limiting personalization intensity, and ensuring transparency tailored to comprehension levels. Protection requires proactive design rather than passive exclusion from harmful targeting.

4.6 Evaluation Continuity

This principle emphasizes monitoring long-term behavioral impacts beyond initial deployment [6]. Many nudges demonstrate short-term effectiveness but may lose impact over time or generate unintended consequences. Some interventions may backfire, leading to compensating behaviors that cancel original goals.

Continuous evaluation helps answer critical questions: Does the nudge provide meaningful education? Do positive behaviors persist after nudge removal? Do new risks emerge over time? Without ongoing monitoring, determining whether AI nudges genuinely help or quietly cause harm becomes impossible.

Long-term monitoring addresses the sustainability rebound effect by tracking cross-domain behavioral changes and compensating behaviors. Ethical systems must incorporate these checks throughout real-world deployment, not merely during initial design phases.

5 Case Study: Ethical Traffic Management

To demonstrate practical BEHAVE framework application, we present detailed analysis of AI-driven traffic management systems, among the most common smart city applications with significant sustainability implications.

5.1 Current Implementation Challenges

Modern traffic systems utilize technology-mediated nudging to influence routing decisions, often employing mechanisms designed to facilitate quick choices [6]. However, current implementations frequently prioritize efficiency over ethical considerations, creating potential manipulation risks.

Typical systems collect location data, analyze traffic patterns, and provide route recommendations designed to minimize travel time. Some platforms prioritize overall traffic flow optimization, potentially routing individual users through suboptimal paths for system-wide benefits without explicit disclosure. The ethical problem here is the lack of transparency. The app does not disclose or prioritize the system's efficiency over the individual autonomy and optimal outcome.

5.2 BEHAVE Framework Application

Behavioral Transparency Implementation Ethical traffic systems must explicitly state rationale behind routing recommendations [6]. This requires adopting transparency-based nudges that disclose personalization logic to build algorithmic trust [10].

Design implementation involves clearly communicating route selection criteria: "Quickest route for your destination," "Eco-friendly option reducing emissions by 15% with 2 additional minutes," or "System-optimal route helping reduce city congestion in this location." Without transparency, routing nudges risk becoming dark patterns that leverage non-transparency to steer decisions away from user interests.

Ethical Timing Considerations Commuting and driving, particularly during stress, place users in cognitive vulnerability states dominated by System 1 processing [6]. In demanding situations, individuals become susceptible to cognitive biases, and AI can exploit these vulnerabilities by determining optimal intervention moments.

The FASCAI framework proposes using AI-driven System 1 nudges with caution, favoring techniques that encourage users to use reflective thinking (System 2) when making important decisions [16]. Implementation might involve subtle feedback signaling congestion without demanding immediate high-pressure decisions, or automatic delays for significant route changes during detected stress periods.

Human Agency Preservation Responsible design ensures AI increases rather than threatens user autonomy and satisfaction [10]. The system should explicitly employ autonomy-supportive nudges ("You may find this useful, but feel free to explore other options as well") that empirically yield higher perceived autonomy scores and offset control reduction in high-personalization conditions.

Risk mitigation addresses potential resistance (when a user feels their freedom of choice is being taken away, causing them to push back against AI's suggestions), and active rejection when users feel autonomy is threatened, and long-term critical thinking reduction risks [16]. Multiple meaningful route alternatives with clear trade-offs must always be provided to the user. This approach empowers the user rather than controlling them.

Adaptive Consent Mechanisms Traffic systems rely on continuous location and behavioral data collection to generate personalized recommendations [21]. Since dynamic personalization struggles to warrant autonomy and transparency, static consent proves insufficient.

Implementation requires continuous and specific permissions regarding location and behavioral data usage for personalized interventions. The AI4SG norm of privacy protection and data subject consent demands respecting consent thresholds [19]. Without adaptability, personal information usage might be perceived as intrusive and untrustworthy.

Vulnerability Protection Measures While routing scenarios focus on situational stress, this principle guards against using predicted behavioral traits for manipulation. The EU AI Act prohibits AI systems exploiting vulnerabilities (including predicted personality traits or socioeconomic factors) to materially distort behavior causing significant harm [20].

Systems should not target specific drivers based on observed stress patterns or risk aversion tendencies to promote choices benefiting system operators over drivers. This aligns with AI4SG situational fairness norms.

Evaluation Continuity Requirements Long-term monitoring proves necessary to detect AI system dependency or behavioral backfires. If drivers become overly reliant on AI guidance, they risk losing independent route assessment capacity. Furthermore, nudges may backfire through compensating behaviors, such as more aggressive driving because navigation AI ensures timely arrival.

Implementation mandates continuous post-deployment monitoring to assess whether nudge effects sustain after intervention removal and to detect educational impact absence [6]. Current research demonstrates limited understanding of long-term nudging effects, making continuous evaluation essential.

6 Discussion and Broader Applications

The BEHAVE framework addresses a critical gap in smart city development by providing integrated guidelines for ethical AI nudging. While our detailed case study focuses on traffic management, the framework principles apply broadly across urban AI applications.

Energy management systems influencing home consumption through dynamic pricing and automated controls can implement all six principles to ensure ethical influence while achieving sustainability goals. Gamified waste reduction systems can use the framework to avoid manipulation while promoting lasting behavioral change. Urban planning AI systems influencing residential, work, and recreational choices through recommendation algorithms require careful ethical consideration.

However, several challenges must be addressed for widespread adoption. Developing interpretable AI models without sacrificing performance requires advances in explainable artificial intelligence. Implementing real-time ethical decision-making needs efficient algorithms for ethical reasoning and constraint satisfaction. Detecting cross-domain rebound effects requires comprehensive data integration and analysis capabilities.

The framework also carries significant governance implications. It aligns with emerging AI regulations such as the EU AI Act while providing practical implementation guidance. Transparent, ethical AI systems are more likely to gain public acceptance, crucial for smart city success. Most importantly, the framework helps make sure that technological efficiency doesn't compromise personal autonomy.

7 Limitations and Future Work

Our work has several important limitations requiring future research. The framework needs extensive empirical validation across different populations and regions, contexts, and smart city applications. Principles may require adaptation for varying cultural contexts with different expectations around privacy, autonomy, and government influence.

Technical feasibility represents another challenge, as some framework requirements may conflict with current capabilities or economic constraints. Implementation difficulties may increase significantly when scaling from pilot projects to city-wide deployments.

Priority areas for future research include studies examining how transparent AI nudging affects user behavior, autonomy, and well-being over extended periods. Cross-cultural validation testing framework applicability across different cultural contexts remains essential. Technical development advancing capabilities needed for ethical AI implementation, particularly in interpretability and real-time ethical reasoning, requires continued investment.

8 Conclusion

As artificial intelligence becomes increasingly integrated in urban infrastructure, the tension between technological efficiency and human autonomy represents a defining ethical challenge. The BEHAVE framework provides principles for developing AI systems that achieve sustainability goals while preserving human autonomy and dignity.

Through analysis of behavioral paradoxes in smart cities and detailed examination of traffic management applications, we demonstrate that ethical AI design can enhance rather than undermine system effectiveness. Users who understand and trust AI systems engage more productively over the long term.

The future of smart cities relies on technological progress and the smart use of that technology to enhance human well-being. The BEHAVE framework represents progress toward cities that are not just smart, but smart environments where efficiency serves human values rather than replacing them.

Our contribution extends beyond technical solutions to questions about human-AI relationships in smart city contexts. As we approach truly smart cities, we must make sure efficiency pursuits never compromise the values making urban life meaningful: autonomy, transparency, dignity, and genuine choice. The framework provides one pathway toward ensuring our technological inheritance reflects humanity's highest aspirations for both capability and dignity.

References

1. Thaler, R., Sunstein, C.: Nudge: Improving Decisions About Health, Wealth, and Happiness. Penguin Books (2009)

2. Singh, R.K., Shah, M.S.: AI-Driven Smart Cities: Improving Urban Infrastructure and Services. International Journal of Advanced Research in Science and Technology 3(2), 156–170 (2025)
3. Rahbarianyazd, R.: Human-Centric Smart Cities for Inclusive and Ethical Urban Development. Smart Design Policies 8(4), 245–267 (2024)
4. Radchenko, K.: Applying Nudge Theory to Foster the SDGs in Smart Cities. In: CEEeGov Days 2023, pp. 78–92 (2023)
5. Bjørlo, L., Moen, Ø., Pasquine, M.: The Role of Consumer Autonomy in Developing Sustainable AI. Sustainability 13(8), 4332 (2021)
6. Caraban, A., Karapanos, E., Gonçalves, D., Campos, P.: 23 Ways to Nudge: A Review of Technology-Mediated Nudging in Human-Computer Interaction. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1–15. ACM (2019)
7. Nallur, V., Renaud, K., Gudkov, A.: Nudging Using Autonomous Agents: Risks and Ethical Considerations. arXiv preprint arXiv:2401.12345 (2024)
8. Mark, R., Anya, G.: Ethics of Using Smart City AI and Big Data: Challenges and Opportunities. ORBIT Journal 2(3), 145–162 (2019)
9. Popescul, D., Radu, L.D.: Socially Responsible Use of AI in Smart Cities: A Framework for Ethical Implementation. European Financial Resilience 4(1), 23–45 (2025)
10. Deepa, V., Halder, S.R., Kalhapure, B., Sonkamble, S., Patel, C.: AI and Consumer Psychology: Personalization, Ethical Nudging, and Digital Decision-Making. Advances in Consumer Research 2(4), 3216–3224 (2025)
11. Voisin, A.: A Green-by-Design Methodology to Increase Sustainability of Smart City Systems. In: RCIS 2019, pp. 267–284 (2019)
12. Bosco, G., Riccardi, V., et al.: AI-Driven Innovation in Smart City Governance: Ethical Considerations and Practical Applications. Transforming Government 15(2), 123–145 (2024)
13. D. Mhlanga, "AI beyond efficiency, navigating the rebound effect in AI-driven sustainable development," *Frontiers in Energy Research*, vol. 13, 2025.
14. Kulaklioglu, D.: Ethical AI in Autonomous Systems and Decision-making: Challenges and Solutions. Human Computer Interaction 8(3), 45–67 (2024)
15. Hansen, P.G., Jespersen, A.M.: Nudge and the Manipulation of Choice: A Framework for the Responsible Use of the Nudge Approach to Behaviour Change in Public Policy. European Journal of Risk Regulation 4(1), 3–28 (2013)
16. Ganapini, M.B., Fabiano, F., Horesh, L., Loreggia, A., Mattei, N., Murugesan, K., Pallagani, V., Rossi, F., Srivastava, B., Venable, B.: Value-based Fast and Slow AI Nudging. arXiv preprint arXiv:2307.07628 (2023)
17. Samuelson, W., Zeckhauser, R.: Status Quo Bias in Decision Making. Journal of Risk and Uncertainty 1(1), 7–59 (1988)
18. Ganapini, M.B., et al.: Fast and Slow Nudging: Using AI to Support Human Decision-Making. In: Proceedings of the IJCAI Workshop on Ethics and Trust in Human-AI Collaboration, pp. 45–52 (2023)
19. Capasso, M., Umbrello, S.: Responsible Nudging for Social Good: New Healthcare Skills for AI-driven Digital Personal Assistants. Medicine, Health Care and Philosophy 25(1), 11–22 (2022)
20. Zhong, H., O'Neill, E., Hoffmann, J.A.: Regulating AI: Applying Insights from Behavioural Economics and Psychology to the Application of Article 5 of the EU AI Act. The Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI-24), pp. 20001–20009 (2024)
21. Yamazaki, Y.: An Empirical Study for The Acceptance of Original Nudges and Hypernudges. In: Societal Challenges in the Smart Society, pp. 323–336 (2020)

22. Nallur, V., Renaud, K., Gudkov, A.: Nudging Using Autonomous Agents: Risks and Ethical Considerations. arXiv preprint arXiv:2201.12345 (2022)
23. Liu, Y., Smith, J., Zhang, H.: The nudging effect of AI-generated content labeling on users' perceptions and behaviors. Frontiers in Psychology (2023)
24. Suroto, S.: Ethical Challenges in Digital Nudging: Transparency and User Trust. Journal of Digital Ethics, 12(1), 45-57 (2025)