Optimizing Household Waste Segregation Policy in the Municipality of Bacolod: An Agent-Based Modeling and Heuristic-Guided Deep Reinforcement Learning Approach

Hussam M. Bansao*, Jemar John J. Lumingkit*, and Dante D. Dinawanao*
*College of Computer Studies, MSU-Iligan Institute of Technology, Iligan City, Philippines
Email: {hussam.bansao, jemarjohn.lumingkit, dante.dinawanao}@g.msuiit.edu.ph

***Abstract*—The implementation of the Ecological Solid Waste Management Act (RA 9003) remains a critical challenge for Local Government Units (LGUs) in the Philippines. Compliance is often hindered by budget constraints, lack of behavioral data, and the complexity of enforcing segregation at the household level. This study presents a novel policy optimization framework combining Agent-Based Modeling (ABM) with Deep Reinforcement Learning (DRL). We model the Municipality of Bacolod, Lanao del Norte, as a dynamic environment where household agents react to policy interventions based on the Theory of Planned Behavior (TPB). A custom Heuristic-guided Deep Reinforcement Learning (HuDRL) agent acts as the LGU policymaker, learning to maximize segregation compliance under strict budget constraints. Our simulations reveal that the HuDRL agent outperforms traditional "Status Quo" policies and standard PPO agents by discovering a "Sequential Saturation" strategy—focusing resources on specific zones to build social norms before expanding. Furthermore, Global Sensitivity Analysis using Sobol Indices identifies "Cost of Effort" as the primary driver of non-compliance, suggesting that LGUs should prioritize logistical support over purely punitive measures.**

***Index Terms*—Agent-Based Modeling, Deep Reinforcement Learning, Waste Management Policy, Smart Governance, Sobol Sensitivity Analysis, Theory of Planned Behavior**

I. Introduction

The implementation of the Ecological Solid Waste Management Act (R.A. 9003) in the Philippines has faced significant hurdles at the local government level, particularly in achieving consistent household segregation compliance. While the law mandates a decentralized approach through barangay-level management, many municipalities continue to struggle with high waste generation and low participation rates. Traditional governance strategies often rely on static, linear policy models that fail to account for the complex, adaptive nature of human behavior and the socio-economic heterogeneity of local communities. This research proposes a computational shift toward smart governance by leveraging Agent-Based Modeling (ABM) and Deep Reinforcement Learning (DRL) to simulate and optimize policy interventions specifically tailored for the Municipality of Bacolod, Lanao del Norte.

A. Statement of the Problem

The study aims to find the most cost-effective way for the Municipality of Bacolod to allocate its limited budget across three solid waste management strategies—rewards, punishments, and educational campaigns—to achieve the highest long-term compliance in household waste segregation.
This study answered the following specific questions:

1) How do variations in synthesized behavioral parameters influence policy efficacy and stability within the simulated environment?
2) What is the optimal mathematical allocation ratio among incentives, enforcement, and education to maximize cost-effectiveness?
3) Which dynamic policy combination yields the highest aggregate compliance and optimal cost-benefit ratio while strictly adhering to the municipality's annual fiscal constraints?

B. Research Objectives

The primary objective of this study is to develop a coupled Agent-Based Model (ABM) and Heuristic-Guided Deep Reinforcement Learning (HuDRL) framework to optimize budget-constrained resource allocation, maximizing household solid waste segregation compliance in the Municipality of Bacolod. Specifically, the study aims:

1) To parameterize the ABM using synthesized academic literature, municipal financial data, and key-informant interviews.
2) To construct a Multi-Level ABM where household behavior is governed by the Theory of Planned Behavior and dynamically responds to local policy interventions.
3) To integrate a DRL algorithm that enables the municipal agent to autonomously learn optimal fund allocations across incentives, enforcement, and education under strict budget limits.
4) To evaluate the cost-effectiveness of isolated (Pure Incentive, Pure Penalty, Pure Education) and Hybrid policy regimes, providing data-driven recommendations for enforcing RA 9003.

C. Significance of the Study

This research contributes to the interdisciplinary fields of environmental science and computational social science by advancing the integration of the Theory of Planned Behavior (TPB) with Deep Reinforcement Learning (DRL) [1]. Academically, it demonstrates the utility of Deep Neural Networks (DNNs) in processing high-dimensional state spaces—specifically varying compliance rates across seven heterogeneous barangays—to discover adaptive policy strategies that traditional linear programming fails to capture [2], [3].

Practically, the study provides Local Government Units (LGUs) with a low-risk, data-driven decision-support tool to test policy mixes (incentives vs. enforcement) without the costs of real-world trials. On a national level, the successful implementation of these recommendations supports environmental sustainability and climate mitigation goals by improving waste segregation at the source, thereby promoting a circular economy and reducing landfill methane emissions.

D. Scope and Limitations

The study is geographically bounded to the seven coastal and urban barangays of the Municipality of Bacolod, Lanao del Norte, currently served by the municipal collection system. The remaining nine barangays are excluded due to logistical inaccessibility.

The scope is strictly focused on household-level segregation and the executive implementation of existing legislation (Municipal Ordinance No. 2018-05), rather than the drafting of new laws. While the model utilizes high-fidelity proxy data for behavioral weights [4] and primary qualitative data from Key Informant Interviews (KIIs) with MENRO and local officials, it remains an abstraction of reality. Specific limitations include:

- The model excludes downstream operations such as landfill management and collection routing.
- Micro-level agent values are derived from a meta-analysis of regional empirical data rather than primary household surveys.
- The simulation assumes honest interactions and excludes informal bypass mechanisms, such as tipping collectors to accept unsegregated waste, to focus on official policy optimization.

II. Related Work

A. Solid Waste Management in the Philippines

The national framework for solid waste management is anchored by the Ecological Solid Waste Management Act of 2000 (R.A. 9003), which mandates source segregation and the establishment of Materials Recovery Facilities (MRFs) [5], [6]. However, Local Government Units (LGUs) frequently struggle with sub-optimal implementation due to severe operational and budgetary constraints, chronic underfunding, and a lack of compliant infrastructure [7]–[9]. Compounding these systemic failures are behavioral challenges at the household level; resident non-compliance is often driven by structural friction—such as irregular collection services—rather than a simple lack of awareness [4], [10]. This indicates that educational campaigns alone are insufficient without robust structural support and community trust [11], [12]. Furthermore, policy design must account for socio-economic heterogeneity to ensure justice, as purely punitive measures like fines disproportionately burden low-income groups [13], [14]. Consequently, bridging the gap between national policy and local practice requires computational approaches that optimize for both cost-efficiency and policy equity by modeling the varying incentive and penalty sensitivities across diverse populations [15], [16].

B. Policy Behavioral Interventions in Waste Management

To effectively enforce waste segregation, Local Government Units (LGUs) must implement a strategic, budget-constrained policy mix of economic incentives, regulatory penalties, and educational campaigns [17], [18]. Economic and regulatory levers, specifically hybrid reward-penalty schemes, are powerful drivers of compliance because they directly alter a household's financial cost-benefit analysis [19], [20]. However, the impact of these financial policies varies significantly across different socio-economic profiles—such as low-income households being more sensitive to fines [21]. Therefore, optimization must account for household heterogeneity to balance policy effectiveness with cost-efficiency and avoid overspending on diminishing returns [22], [23]. To ensure long-term program sustainability, these financial tools must be complemented by educational and behavioral interventions [24]. Strategies like educational campaigns and Nudge Theory directly target the core constructs of the Theory of Planned Behavior by improving resident Attitude, Perceived Behavioral Control, and Subjective Norms [25], [26]. Because these "soft" behavioral factors are critical for sustained participation, an integrated Agent-Based Modeling (ABM) framework is ideally suited to simulate and predict the complex, community-level responses to this interconnected hybrid policy mix [27], [28].

C. From Behavioral Initiation to Habitual Persistence

Designing effective waste segregation policies requires a computational model grounded in established psychological theory. In this study, the Agent-Based Model (ABM) simulates household decision-making using the Theory of Planned Behavior (TPB), determining an individual's intention to segregate based on Attitude (A), Subjective Norms (SN), and Perceived Behavioral Control (PBC) [29], [30]. This cognitive process is mathematically operationalized through a linear utility function:

$$U_{\text{segregate}} = (w_A A + w_{SN} SN + w_{PBC} PBC) + \epsilon \quad (1)$$

where the ϵ term accounts for the inherent stochasticity, or "noise," in human decision-making [31]. Moving beyond individual initiation, the model incorporates non-linear "critical mass" dynamics. Research indicates that a committed minority reaching a tipping point of approximately 25% can

trigger a rapid cascade of adoption across the majority, implying that resource-constrained local governments can focus interventions to reach this threshold rather than funding the entire population indefinitely [32], [33]. Finally, to simulate resistance to behavioral decay once financial incentives are removed, the model integrates "Cultural Inertia". When waste segregation becomes a deeply established social habit, the social cost of deviating (such as peer pressure) exceeds the physical effort required, ensuring the behavior remains self-sustaining [34]–[36].

D. Agent-Based Modeling as a Markov Decision Process

Agent-Based Modeling (ABM) functions as a computational "virtual laboratory" uniquely equipped to analyze complex socio-environmental systems, such as municipal solid waste management, where system-wide compliance emerges directly from micro-level household decisions [1], [37]. Unlike aggregate methodologies such as System Dynamics that focus on macro-level stocks and flows [17], ABM is essential for this research because it accurately captures population heterogeneity and adaptive individual behavior [38], [39]. By representing households as autonomous agents with distinct socio-demographic profiles governed by the Theory of Planned Behavior (TPB), the model can simulate highly realistic, non-linear responses to dynamic policy changes [40]. Furthermore, ABM's flexible architecture allows for the integration of advanced computational techniques, including machine learning classifiers, which enhances the behavioral realism of the agents and significantly improves the accuracy of predicted policy outcomes [41], [42].

E. Heuristic-Guided Deep Reinforcement Learning

While Agent-Based Modeling (ABM) provides the simulation environment, Deep Reinforcement Learning (DRL) is essential for autonomously discovering optimal, budget-constrained policies by mapping complex, high-dimensional household states to precise municipal decisions [43], [44]. To achieve this, the ABM is mathematically framed as a Markov Decision Process (MDP) that acts as a high-fidelity data generator, providing the necessary state representations and reward signals to train the DRL agent [45]. However, applying DRL to public governance introduces the "Sparse Reward Problem," as delayed causal effects in human behavior make it difficult for standard agents to learn effective strategies without failing or randomly wandering [46], [47]. To overcome this inefficiency, the framework integrates Heuristic-Guided Reinforcement Learning (HuRL) to prune the action space and direct the agent's attention to relevant variables [48], alongside Potential-Based Reward Shaping to provide immediate, threshold-based feedback that prevents the agent from spreading resources too thinly [49]. Ultimately, these advanced techniques address a significant gap in current literature: while DRL is extensively utilized for physical, industrial waste sorting and logistics [3], [50], [51], its application in optimizing governance resources to actively influence human behavior remains critically under-explored [52], [53].

F. Research Gap

A comprehensive review of the literature reveals a critical implementation deficit in the Philippine Solid Waste Management system under R.A. 9003, driven by weak enforcement, chronic municipal budget constraints, and a persistent gap between public awareness and actual compliance [54], [55]. Addressing these heterogeneous behavioral challenges requires a paradigm shift toward smart, predictive management that blends hybrid reward-penalty schemes with non-monetary educational levers [56], [57]. However, a critical research gap exists at the intersection of Agent-Based Modeling (ABM) and Deep Reinforcement Learning (DRL) for adaptive public policy [14]. While existing ABM research relies on static scenario testing and DRL studies focus primarily on technical or logistical optimization, no study has developed an integrated framework where a resource-constrained Local Government Unit (LGU) autonomously optimizes the dynamic allocation of funds across enforcement, incentives, and education. To bridge this gap, this study proposes a dynamic structural modification to the Theory of Planned Behavior (TPB) by explicitly integrating an external policy term (C_{Net}) and time-variant weights:

$$\begin{aligned} U_{\text{segregate}} = (w_A(t)A) + (w_{SN}(t)SN_{\text{local}}) \\ + (w_{PBC}(t)PBC_{\text{infra}}) - C_{\text{Net}} + \epsilon \end{aligned} \tag{2}$$

This novel architecture simulates non-linear behavioral evolution—such as "public forgetting" and "psychological reactance"—while strategically decoupling the agent's internal psychological state from external policy levers. By transforming the TPB into a computational interface for a DRL agent, this coupled ABM-DRL framework elevates the LGU from a static administrator to a "strategic learner," providing a mathematically optimized, cost-effective decision-support tool to maximize long-term household segregation compliance [41], [58].

III. Methodology

A. Research Design

This study employed a computational simulation research design that integrated Agent-Based Modeling (ABM) with Deep Reinforcement Learning (DRL) optimization. This design created a virtual laboratory for testing Solid Waste Management (SWM) policies, allowing for the autonomous discovery of the optimal resource allocation strategy without the cost and risk of real-world trials. The research followed three main phases: (1) Model parameterization using literature synthesis, (2) DRL integration and training, and (3) Policy scenario simulation and analysis.

B. Data Sources and Model Parameterization

In lieu of collecting large-scale primary survey data, this study constructed a high-fidelity Agent-Based Model (ABM) by synthesizing data from academic literature, public government statistics, and operational records obtained through key-informant interviews. The behavioral core of the household

agents was grounded in the Theory of Planned Behavior (TPB), with parameters for Attitude (w_A), Subjective Norms (w_{SN}), and Perceived Behavioral Control (w_{PBC}) derived from a systematic review of environmental psychology literature [29], [59].

To ensure ecological validity, agent initialization utilized empirical Knowledge, Attitude, and Practices (KAP) data [4], explicitly calibrating agents to reflect a realistic "Intention-Action Gap" (High Attitude $A_0 \approx 0.66$ vs. Low Compliance $B_0 \approx 0.58$). The simulation environment was further contextualized to seven specific barangays in the Municipality of Bacolod (e.g., Liangan East, Poblacion) using validated socio-demographic profiles and LGU operational data.

The LGU-DRL agent operates within a strict annual budget of ₱1,500,000, discretized quarterly, to optimize a "Policy Mix" across three cost-constrained levers: (1) *Enforcement* (C_{Enf}), calculated via Equation 3 based on personnel wages and coverage ratios; (2) *Monetary Incentives* (C_{Inc}), modeled in Equation 4 as a variable function of $N_{\text{Compliant}}$ to introduce a fiscal "victim of success" risk; and (3) *IEC Campaigns* (C_{IEC}), defined in Equation 5 as tiered fixed costs for media dissemination.

1) Cost of Enforcement (C_{Enf}): Modeled as a resource-constrained operational expense based on personnel wages and coverage capacity. Assuming a logistical limit of 30 households per officer per day, the cost relies on the regional minimum wage for Region X over 66 working days per quarter.

$$C_{\text{Enf}} = (N_{\text{Enforcers}} \times W_{\text{Daily}} \times 66) \tag{3}$$

2) Cost of Incentives (C_{Inc}): Structured as a dynamic, variable liability directly proportional to the rate of public compliance. Because the total cost scales with the number of compliant households, the agent faces a fiscal "Victim of Success" risk, requiring it to balance positive reinforcement against rapid treasury depletion.

$$C_{\text{Inc}} = (V_{\text{Reward}} \times N_{\text{Compliant}}) \tag{4}$$

3) Cost of Information & Educational Campaign (C_{IEC}): Calculated as tiered fixed costs associated with discrete public awareness efforts. This formulation captures the explicit expenses of procuring local radio broadcast spots and mobilizing community events.

$$C_{\text{IEC}} = (N_{\text{Spots}} \times R_{\text{Radio}}) + (N_{\text{Events}} \times C_{\text{Mobilization}}) \tag{5}$$

4) Multi-Level ABM Architecture: The simulation, developed using Python's *MESA* framework, employs a hierarchical structure to mirror the decentralized governance of the Municipality of Bacolod. Formulated as a Markov Decision Process (MDP), the model operates on quarterly time steps and features a Deep Reinforcement Learning (DRL) agent that optimizes budget allocations across enforcement, incentives, and educational campaigns to maximize compliance.

The architecture is driven by three interacting agent classes:

- *Barangay Agents (Local Government):* Representing seven distinct barangays, these agents manage fiscal allocation, policy formulation, and personnel deployment. They aggregate local compliance data to serve as the observational state and reward signal for the DRL mechanism.
- *Enforcement Agents (Operational Arm):* Tasked with stochastic monitoring, these agents conduct random inspections to enforce the "No Segregation, No Collection" policy. Their presence deters non-compliance but incurs high operational costs, forcing the system to balance strict environmental policing with fiscal sustainability.
- *Household Agents (Citizens):* Geographically assigned to barangays, these agents make binary segregation decisions based on the Theory of Planned Behavior. Their choices are dynamically influenced by policy strictness, perceived enforcement intensity, and the social norms of neighboring households.

A critical innovation in this environment is the Norm Internalization Mechanism, which dictates that sustained community compliance exceeding 70% transforms waste segregation into a resilient social habit, resisting decay even when the LGU reduces external interventions.

C. Household Agent Design

The decision-making architecture of the `HouseholdAgent` is governed by a dynamic utility function grounded in the Theory of Planned Behavior (TPB), which posits that an individual's intention to perform a behavior is a function of their attitude, subjective norms, and perceived behavioral control [27]. In contrast to static behavioral models, this framework incorporates time-variant weights ($w(t)$) for psychological constructs, allowing agent behavior to evolve non-linearly in response to LGU interventions and social feedback loops [28], [29]. The core decision logic is represented by the utility of segregation ($U_{\text{segregate}}$):

$$\begin{aligned} U_{\text{segregate}} = (w_A(t)A) + (w_{SN}(t)SN_{\text{local}}) \\ + (w_{PBC}(t)PBC_{\text{infra}}) - C_{\text{Net}} + \epsilon \end{aligned} \tag{6}$$

In this formulation, $w_A(t)$ represents the temporal evolution of the agent's internal valuation of segregation, functioning as a decay model that increases in response to Information, Education, and Communication (IEC) investment and decays stochastically in the absence of reinforcement to simulate "public forgetting" [26]. The social component, SN_{local}, is an endogenous variable derived from the observed compliance rate of the agent's immediate spatial neighborhood (radius r), capturing the effects of social pressure and observational learning [30], [40]. To account for the inherent uncertainty in human decision-making, the term ϵ introduces stochastic noise, ensuring the model reflects real-world behavioral variance [31].

A critical innovation in this model is the calculation of C_{Net}, the net behavioral cost, which characterizes the perceived friction of compliance. This variable is defined as:

$$C_{\text{Net}} = C_{\text{Effort}} + (\gamma C_{\text{Monetary}}) - (\gamma I) - (\gamma FP_{\text{Detection}}) \tag{7}$$

This cost-benefit sub-model integrates both physical and financial barriers, defined as follows:

$\boldsymbol{C}_{\mathbf{Effort}}$: The physical hassle associated with the washing, sorting, and storage of waste.

$\boldsymbol{C}_{\mathbf{Monetary}}$: Tangible financial expenses, such as the procurement of color-coded bins or sacks.

$\boldsymbol{\gamma}$ *(Income Sensitivity)*: A weighting factor derived from household income level; for lower-income households, $\gamma > 1$ as financial levers (I and F) carry greater psychological weight, whereas $\gamma < 1$ for higher-income households.

$\boldsymbol{I}$ *and* $\boldsymbol{F}$: The objective magnitudes of monetary incentives and punitive fines, respectively.

$\boldsymbol{P}_{\mathbf{Detection}}$: The likelihood of enforcement detection, which is dynamically linked to the LGU's resource allocation for enforcement personnel.

Furthermore, the model captures complex psychological feedbacks such as "psychological reactance." While LGU investments in IEC campaigns generally improve Attitude (A) and Subjective Norms (SN) [26], the model stipulates that if enforcement intensity crosses a specific threshold, Attitude (A) may paradoxically decrease. This inverse reaction reflects the tendency of individuals to resist coercive mandates when they perceive a loss of autonomy, a concept central to Nudge Theory applications in waste management [25]. Finally, Perceived Behavioral Control (PBC) is calibrated by the availability of barangay-level infrastructure, ensuring the simulation acknowledges that even highly motivated agents may fail to comply if functional bins or Materials Recovery Facilities (MRFs) are absent.

D. Behavioral Hysteresis and the Internalization of Social Norms

While the baseline `HouseholdAgent` architecture operates on the Theory of Planned Behavior (TPB), standard models often exhibit unrealistic behavioral decay where compliance collapses immediately upon the cessation of external stimuli. To more accurately simulate the real-world stability of established habits, this study incorporates a *Norm Internalization Mechanism*. This mechanism addresses the phenomenon of "Cultural Inertia," where a behavior transitions from a calculated, incentive-driven decision to an internalized social habit once a critical tipping point is reached [27], [34].

In this framework, the standard attitude update function, which typically suffers from a decay constant (δ) due to enforcement fatigue or apathy, is refined by a dynamic damping factor (D_{factor}). This modification aligns with recent findings on the persistence of pro-environmental behavior, suggesting that established social norms act as a buffer against the rapid erosion of compliance [35]:

$$A_{t+1} = A_t - (\delta \times D_{\text{factor}}) \tag{8}$$

The damping factor is governed by the strength of local social norms (SN), representing the protective effect of community-wide compliance. Specifically, the model identifies a *tipping point at 70% compliance*, beyond which the behavior is considered internalized. This threshold reflects the "critical mass" theory in social dynamics, where minority behaviors cascade into majority conventions once a specific saturation point is exceeded [32], [33]:

$$D_{\text{factor}} = \begin{cases} 0.1 & \text{if } SN > 0.70 \text{ (Strong Norms / Internalized)} \\ 0.5 & \text{if } SN > 0.50 \text{ (Moderate Norms)} \\ 1.0 & \text{otherwise (Weak Norms)} \end{cases} \tag{9}$$

Furthermore, to prevent the total erosion of community standards during periods of fiscal austerity or low LGU intervention, the model introduces a *Social Norm Floor set at 40%*. This floor ensures that once a barangay has achieved a baseline level of awareness, the perceived social pressure does not drop to zero, representing the residual collective memory and existing communal infrastructure of the municipality [60]. This refinement is critical for addressing the sustainability objectives of this research, as it allows the Deep Reinforcement Learning agent to discover strategies that foster self-sustaining compliance rather than perpetual reliance on high-cost monetary incentives .

E. Multi-Objective Heuristic Reward Function

The Local Government Unit (LGU) agent's learning behavior is guided by a Composite Reward Function designed to optimize environmental policy while strictly managing fiscal limitations and social acceptability. To accelerate the agent's training, the model employs Reward Shaping to provide immediate, dense feedback.

The total reward at each time step (t) is calculated as:

$$R_{\text{total}} = w_1 R_{\text{Comp}} + w_2 R_{\text{Sustain}} - w_3 P_{\text{Backlash}} + R_{\text{Shaping}} \tag{10}$$

This function balances four distinct operational components:

- *Environmental Compliance (R_{Comp}):* The primary objective, measured by the population-weighted waste segregation rate across all barangays.
- *Fiscal Sustainability (R_{Sustain}):* A regularization term that prevents premature budget exhaustion by penalizing deviations from a steady, ideal spending rate:

$$R_{\text{Sustain}} = -\left| \frac{S_{\text{Actual}}}{B_{\text{Total}}} - \frac{1}{12} \right| \tag{11}$$

- *Political Backlash Penalty (P_{Backlash}):* A penalty applied when strict enforcement is paired with low public compliance, discouraging draconian policies that could trigger social resistance.
- *Heuristic Shaping Rewards (R_{Shaping}):* Targeted training bonuses designed to guide the agent out of sub-optimal strategies and toward a "Sequential Saturation" approach. This includes an Allocation Focus Bonus for directing $> 40\%$ of the budget to a critical, underperforming barangay, and an Intensity Threshold Jackpot ($+20,000$) for pushing local enforcement intensity beyond a critical 0.80 threshold.

Finally, the function's weights (w_1, w_2, w_3) calibrate these competing objectives to accurately mirror the real-world constraints of a 4th Class Municipality, ensuring the AI's solutions are practically viable.

F. Heuristic-Guided Deep Reinforcement Learning (HuDRL)

To overcome the "Sparse Reward Problem" inherent in high-dimensional municipal resource allocation ($d = 21$), this study implements a Heuristic-Guided Deep Reinforcement Learning (HuDRL) framework utilizing Deep Proximal Policy Optimization (Deep PPO). The agent employs a custom Actor-Critic architecture where two deep neural networks (64-neuron dense layers with ReLU activation) process a multi-modal State Vector (S_t) comprising barangay compliance rates, fiscal liquidity, and political capital. To strictly enforce the municipality's fiscal constraints, the Actor network utilizes a Softmax Normalization Layer, mathematically guaranteeing that the continuous action vector (A_t)—representing allocations for IEC, Enforcement, and Incentives—never exceeds the quarterly budget cap ($B_{\text{Quarterly}}$).

Crucially, the framework integrates "Reward Shaping" to guide the agent away from sub-optimal, equitable distributions (the "Status Quo" trap) and toward a "Sequential Saturation" strategy. The Composite Reward Function (R_{total}) balances environmental compliance against fiscal sustainability and political backlash, while adding heuristic bonuses ($+20,000$) when the agent concentrates $> 40\%$ of resources on the weakest performing barangay. This logic is operationalized in the HuDRL Algorithm (Algorithm 1), which utilizes a "Targeted Amplification" mechanism to act as a saliency filter, multiplying intent signals to critical nodes by a factor of $\alpha = 100$ to accelerate the discovery of optimal tipping points.

Algorithm 1 Heuristic-Guided Action Selection & Reward Shaping (HuDRL)

Input: S_t (State Vector), $\mathcal{B}_{\text{list}}$ (Barangay Agents), B_{cap} (Quarterly Budget)
Output: A_{final} (Optimized Budget Allocation)
Step 1: Observation
$S_t \leftarrow$ GetCurrentState()
Step 2: Identify Critical Node
$B_{\text{crit}} \leftarrow \min_{b \in \mathcal{B}_{\text{list}}}(b.\text{compliance\_rate})$ {Identify weakest performing barangay}
Step 3: Initial Policy Prediction
$A_{\text{raw}} \leftarrow \pi_\theta(S_t)$ {Neural Network prediction}
Step 4: Heuristic Targeted Amplification
if $A_{\text{raw}}[B_{\text{crit}}] > \delta_{\text{intent}}$ **then**
    $A_{\text{raw}}[B_{\text{crit}}] \leftarrow A_{\text{raw}}[B_{\text{crit}}] \times \alpha$ {Amplify signal (e.g., alpha=100)}
    Decrease other allocations to balance
end if
Step 5: Budget Normalization (Softmax-style)
$A_{\text{final}} \leftarrow \text{Softmax}(A_{\text{raw}}) \times B_{\text{cap}}$
Step 6: Execution
Apply A_{final} to ABM Environment
Step 7: Reward Shaping
$R_{\text{total}} \leftarrow R_{\text{base}} + R_{\text{bonus}}$ {Includes Jackpot for Saturation Strategy}
Step 8: Network Update
PPO.Update(S_t, A_{final}, R_{total})

G. Simulation and Analysis

This study employed a four-stage experimental design to evaluate the ABM-RL framework. First, *Initialization and Calibration* grounded the model in operational data from seven barangays, matching the municipality's $\approx 10\%$ baseline compliance [10], [14]. Second, the HuDRL agent trained over 10,000 simulated periods to explore the policy space [2]. Third, we simulated three *Policy Scenarios—Pure Penalty*, *Pure Incentive*, and an AI-optimized *Hybrid Regime* [1]—evaluating them on maximum compliance, cost-effectiveness, policy equity, and resource allocation. Finally, a Sobol *Global Sensitivity Analysis* [61] validated model robustness by stress-testing core behavioral parameters (w_a, w_{sn}, c_{effort}).

IV. Results

A. Data and Input Parameters Analysis

The validity of the Agent-Based Model (ABM) was established through empirical grounding of initialization parameters derived from legislative documents, interviews, and municipal financial data. The environment was bounded by an Annual Budget Cap of P1,500,000 (approx. P375,000 per quarter), a constraint creating a "Personnel Trap" where enforcement costs frequently crowded out education and incentives.

As detailed in Table I, profound heterogeneity prevented a "one-size-fits-all" policy. Brgy. Poblacion (the "Urban Resource Trap") possessed the largest budget (P200,000) but the lowest per-capita spending power, necessitating a 60% legislative lock on enforcement. Brgy. Binuni represented a "Wealthy Anomaly," while Babalaya and Demologan exemplified the "Poverty Trap," with 90% of funds consumed by mandatory personnel costs.

TABLE I
Barangay Demographic and Financial Parameters

		Annual	Initial	Income Profile (%)		
Barangay	Households	Budget (P)	Comp.	Low	Mid	High
Poblacion	1,534	200,000	2%	70	25	5
Liangan East	608	30,000	14%	40	40	20
Ezperanza	574	90,000	14%	20	50	30
Binuni	507	126,370	15%	50	30	20
Demologan	463	21,000	11%	80	15	5
Babalaya	171	15,000	14%	80	10	10
Mati	165	80,000	11%	90	5	5

The simulation incorporated distinct psychosocial profiles (Table II). Agrarian and commercial center disparities altered the marginal utility of money; low-income concentrations in Mati and Demologan resulted in high price sensitivity (γ), making uniform fines regressive.

TABLE II
CALIBRATED BEHAVIORAL AND ALLOCATION PROFILES

Barangay	Agent Parameters				Allocation (%)		
	w_a	w_{sn}	C_e	γ	Enf	Inc	IEC
Binuni	0.65	0.80	0.64	0.02	40	40	20
Ezperanza	0.40	0.70	0.55	0.03	50	30	20
Babalaya	0.60	0.90	0.62	0.05	90	5	5
Liangan East	0.65	0.60	0.58	0.04	25	65	10
Poblacion	0.55	0.20	0.48	0.03	60	20	20
Demologan	0.60	0.60	0.62	0.05	85	10	5
Mati	0.60	0.50	0.60	0.05	30	50	20

B. Model Calibration and Data Reconciliation

Calibration aligned the simulation with MENRO audit data, reconciling discrepancies between micro-level reports and macro-level reality. While interviews suggested 40–50% compliance, audits confirmed only 10–15%. This "Act vs. Reality" gap highlighted social desirability bias. The model utilized Brgy. Poblacion (2%) as an accurate "Anchor Point" to establish a valid Status Quo baseline (≈12.5%).

TABLE III
COMPLIANCE CALIBRATION (S_0)

Barangay	Acquired (Interview)	Calibrated S_0 (Audit)
Babalaya	100%	14%
Binuni	95%	15%
Mati	70%	11%
Liangan East	65%	14%
Demologan	60%	11%
Ezperanza	20%	14%
Poblacion	2%	2%

C. Comparative Results of Policy Strategies

Simulation experiments showed a stark divergence between equal-distribution strategies and DRL-driven dynamic allocation. Without "Sequential Saturation," static rebalancing was insufficient to resolve non-compliance.

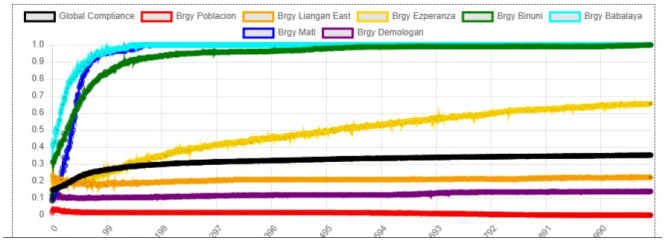

Fig. 1. Baseline Status Quo

The *Status Quo* strategy distributed funds equally (P53,571/barangay), acting as a "dilution mechanism." Small communities like Mati achieved compliance via social momentum, but critical zones like Poblacion stagnated at 0.33%.

The *HuDRL* agent outperformed all manual strategies, achieving *92.8%* terminal compliance. The AI discovered *Sequential Saturation*, concentrating 51%–69% of the total municipal budget on single barangays sequentially (e.g., Demologan in Q2, Poblacion in Q8-12). This ensured intervention intensity exceeded resistance thresholds, prioritizing sequential efficacy over simultaneous fairness.

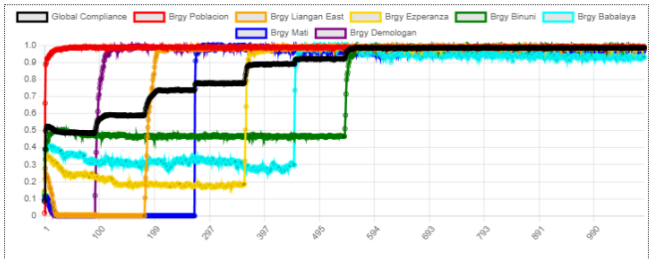

Fig. 2. Heuristic-Guided Deep Reinforcement Learning Approach

TABLE IV
GLOBAL COMPLIANCE ACROSS REGIMES (Q_1 - Q_{12})

Policy Regime	Q1	Q4	Q8	Q12
Status Quo (Baseline)	10.7%	52.4%	55.3%	57.1%
Pure Incentives	10.7%	34.6%	34.2%	34.3%
Pure Enforcement	10.7%	14.9%	14.0%	13.9%
Adaptive (HuDRL)	*10.7%*	*45.3%*	*72.2%*	*92.8%*

D. Global Sensitivity Analysis

A Sobol Global Sensitivity Analysis (GSA) revealed the *Cost of Effort* (c_{effort}) as the primary driver, accounting for *83%* of variance. This provides evidence for the "Convenience Hypothesis," where structural deficiencies override awareness. *Social Norms* (w_{sn}) acted as a force multiplier, while *Attitude* (w_a) showed near-zero sensitivity, quantifying the "Value-Action Gap" where awareness alone fails without structural support.

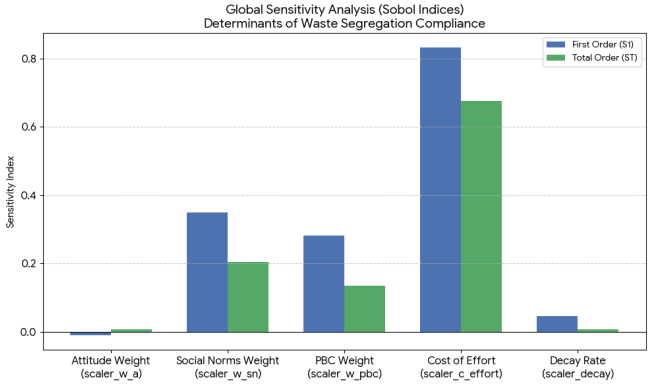

Fig. 3. Global Sensitivity Analysis Results illustrating First-Order Sobol Indices (S_1) and Total Order Indices (ST) for behavioral parameters.

V. DISCUSSION

A. The Determinants of Compliance

The findings from the evaluation phase provided a critical computational justification for the strategic trajectory of the

Deep Reinforcement Learning (DRL) agent, specifically its preference for the "Sequential Saturation" strategy over the existing education-centric "Status Quo." By examining the behavioral sensitivity of the model, this discussion bridges the gap between the algorithmic outputs of the AI and the socioeconomic realities of waste management in the Municipality of Bacolod [14].

1) The "Convenience Barrier" and the Dominance of Cost: The overwhelming sensitivity of the model to the Cost of Effort (c_{effort}), which accounted for approximately 83% of the variance in global compliance rates ($S_1 \approx 0.83$), provided mathematical validation for what is known in environmental sociology as the "Convenience Hypothesis." Within the framework of this study, "Cost" was not defined solely by financial penalties or fines; it represented the total "friction" of the activity, including temporal, physical, and cognitive effort [34].

In the context of Bacolod, Lanao del Norte, this sensitivity implied that if the segregation process was characterized by high friction—such as the requirement to purchase specific color-coded liners or the need to traverse significant distances to collection points—household agents would default to non-compliance regardless of their pro-environmental beliefs. This finding was strongly aligned with the empirical work of [54] and [10], whose research suggested that structural deficiencies often overrode individual intent.

Consequently, the DRL agent's aggressive funding of Enforcement (to increase the cost of non-compliance) and Incentives (to offset the cost of compliance) represented a rational response to this "Convenience Barrier." The AI learned that it could not simply "educate" away the physical friction of the waste system; it had to fundamentally alter the individual utility calculus [19].

2) The "Value-Action Gap": Why Education Exhibited Diminishing Returns: A significant revelation of the sensitivity analysis was the near-zero sensitivity of the Attitude (w_a) parameter. While traditional municipal policies often focused heavily on IEC campaigns, the model demonstrated that awareness alone had a low ceiling of effectiveness in the absence of structural support. This phenomenon was a computational realization of the "Value-Action Gap" [40].

As noted by [21] in their research on waste sorting in developing nations, extrinsic motivators such as incentives and regulatory pressure were significantly more effective than intrinsic motivators (education) in promoting consistent household participation. For the Municipality of Bacolod, these results suggested a state of diminishing returns on education. Residents likely understood the ecological importance of segregation, but the current policy framework failed to address the gap where that knowledge should turn into action. The DRL agent's decision to pivot away from IEC funding was therefore a strategic recognition that the most effective lever for immediate compliance lay in structural and extrinsic motivators [62].

3) Cultural Inertia and the "Tipping Point": The significant sensitivity to Social Norms ($w_{sn}, S_1 \approx 0.35$) validated the "King of the Hill" or "Sequential Saturation" strategy favored by the AI. Social Norms functioned as a non-linear behavioral multiplier within the social fabric of the barangays. The mechanism operated on the principle of communal visibility: when compliance was low (e.g., $< 10\%$), the social norm reinforced non-compliance as the acceptable communal standard, creating a state of "Cultural Inertia" that resisted change [35].

However, the DRL agent identified a critical threshold—a "Tipping Point"—where social influence flipped from a barrier to a facilitator. Once enforcement and incentives pushed a barangay's compliance past a critical mass (approximately 70%), the Social Norm parameter began to act as a reinforcement mechanism [32]. In this state, the pressure to conform to the now-visible majority behavior sustained high compliance even as the agent reallocated resources to other areas. This explained the "Graduation Effect" observed in the longitudinal data.

VI. Conclusion

This study demonstrates that a Heuristic-Guided DRL agent can successfully optimize municipal waste segregation policies under strict budget constraints, discovering a "Sequential Saturation" strategy that outperforms status quo equitable distribution [43]. Global Sensitivity Analysis revealed that logistical friction, or the Cost of Effort (c_{effort}), is the primary barrier to compliance (accounting for 83% of variance), indicating policies must prioritize convenience over pure awareness [54]. Furthermore, the near-zero sensitivity of the Attitude (w_a) parameter highlights a profound "Value-Action Gap," mathematically justifying the agent's pivot toward extrinsic motivators [21]. Ultimately, leveraging Social Norms ($w_{sn} \approx 0.35$) creates a self-sustaining "Graduation Effect," allowing Local Government Units (LGUs) to build resilient compliance ecosystems without perpetual funding reliance [14], [35].

Future research should focus on scaling this framework into a Multi-Agent Reinforcement Learning (MARL) environment, where individual barangays operate as autonomous, self-optimizing sub-agents negotiating for resources with the central LGU. Additionally, transitioning the ABM's observational state from delayed quarterly audits to real-time empirical data streams—such as IoT-enabled smart bin sensors or digitized waste collection routing—would allow the DRL agent to execute dynamic, micro-targeted policy interventions. Finally, incorporating the informal waste sector (e.g., independent scavengers and private recyclers) into the agent architecture would further enhance the ecological validity of the municipality's simulation.

VII. Acknowledgement

The authors thank the Bacolod LGU and its component barangays for their support and data.

Use of Generative AI

The authors used Gemini 3.1 Pro for grammatical corrections and maintain full responsibility for the final manuscript.

REFERENCES

[1] X. Tian, F. Peng, G. Wei, C. Xiao, Q. Ma, Z. Hu, and Y. Liu, "Agent-based modeling in solid waste management: Advantages, progress, challenges and prospects," *Environmental Impact Assessment Review*, vol. 110, p. 107723, 2024.

[2] G. Dey, "AI-driven community-centric waste management: Leveraging reinforcement learning for dynamic waste sorting and public engagement," 2025, researchGate. [Online]. Available: https://www.researchgate.net/publication/392726834

[3] V. T. Ha and N. Q. Minh, "Intelligent route planning for waste collection in smart cities via reinforcement learning," *Journal of Information Systems Engineering and Management*, vol. 10, no. 51s, pp. 434–445, 2025.

[4] S. J. L. Paigalan, D. M. Paigalan, and J. M. R. Zoleta, "Knowledge, attitude, and practices on solid waste management among residents of a riverside barangay: Basis for sustainable policies and programs," *Scientia. Technology, Science and Society*, vol. 2, no. 5, pp. 3–14, 2025.

[5] Republic of the Philippines, "Ecological Solid Waste Management Act of 2000, rep. act no. 9003," January 2001, philippines. [Online]. Available: https://www.officialgazette.gov.ph/2001/01/26/republic-act-no-9003-s-2001/

[6] R. Santos, "A critical analysis of Republic Act No. 9003 Ecological Solid Waste Management Act of 2000," *SSRN Electronic Journal*, 2025.

[7] M. E. C. Camarillo and L. M. Bellotindos, "A study of policy implementation and community participation in the municipal solid waste management in the Philippines," *Applied Environmental Research*, vol. 43, no. 2, pp. 1–26, 2021.

[8] E. Coracero, "A long-standing problem: A review on the solid waste management in the Philippines," *Indonesian Journal of Social and Environmental Issues (IJSEI)*, vol. 2, no. 3, 2021. [Online]. Available: https://ojs.literacyinstitute.org/index.php/ijsei/article/view/144

[9] W. D. Dalugdog, "Level of compliance of the Local Government Units (LGUs) in the implementation and enforcement of R.A. 9003 (known as Ecological Solid Waste Management Act of 2000) in CALABARZON," *Asian Journal of Multidisciplinary Studies*, vol. 4, no. 1, pp. 25–38, 2021. [Online]. Available: https://asianjournals.org/online/index.php/ajms/article/view/339

[10] J. Villanueva, A. Magsino, F. Hernandez, and A. M. Hernandez, "Solid waste management (SWM) conditions, practices, and challenges of select barangays in Lipa City," 2021, philippine Association of Institutions for Research, Inc. [Online]. Available: https://www.researchgate.net/publication/360176307

[11] V. F. Collado, A. M. Badua, J. L. Aban, and R. F. Deleña, "Serve, understand, respect waste education management (SURWEM): An extension project to strengthen solid waste management in a service Barangay (Arosip) in northern Philippines," *International Journal of Biosciences (IJB)*, vol. 25, no. 2, pp. 218–229, 2024.

[12] L. Salsabila, E. P. Purnomo, and H. D. Jovita, "The importance of public participation in sustainable solid waste management," *Journal of Governance and Public Policy*, vol. 8, no. 2, 2021.

[13] N. O. Carpio, J. E. Angsinco, R. B. P. Pineda, and A. M. B. Donaire, "Environmental awareness and practices in waste management: A green criminological perspective," *European Journal of Social Sciences Studies*, vol. 11, no. 3, 2025.

[14] A. N. Jiménez, "Agent-based modeling and reinforcement learning for equitable waste transitions in Costa Rica: A national policy stress-test," *Journal of Cleaner Production*, vol. 535, p. 147103, 2025.

[15] R. Medina-Mijangos, A. De Andrés, H. Guerrero-Garcia-Rojas, and L. Seguí-Amórtegui, "A methodology for the technical-economic analysis of municipal solid waste systems based on social cost-benefit analysis with a valuation of externalities," *Environmental Science and Pollution Research*, vol. 28, no. 15, pp. 18 807–18 825, 2020.

[16] A. E. Torkayesh, H. R. Vandchali, and E. B. Tirkolaee, "Multi-objective optimization for healthcare waste management network design with sustainability perspective," *Sustainability*, vol. 13, no. 15, p. 8279, 2021.

[17] M. Dhanshyam and S. K. Srivastava, "Effective policy mix for plastic waste mitigation in India using system dynamics," *Resources, Conservation and Recycling*, vol. 168, p. 105455, 2021.

[18] L. Fontaine, R. Legros, and J.-M. Frayret, "Sustainability and environmental performance in selective collection of residual materials: Impact of modulating citizen participation through policy and incentive implementation," *Resources*, vol. 13, no. 11, p. 151, 2024.

[19] S.-W. Chen, S.-W. Huang, J. Chen, K.-Y. Huang, and Y.-X. He, "Effects of incentives and penalties on farmers' willingness and behavior to separate domestic waste—analysis of farm household heterogeneity based on chain multiple intermediary effects," *Sustainability*, vol. 15, no. 7, p. 5958, 2023.

[20] D. Mu and S. Zhang, "The impact of reward–penalty policy on different recycling modes of recyclable resources in residential waste," *Sustainability*, vol. 13, no. 14, p. 7883, 2021.

[21] A. Zhao, L. Zhang, X. Ma, F. Gao, and H. Zhu, "Effectiveness of extrinsic incentives for promoting rural waste sorting in developing countries: Evidence from China," *The Developing Economies*, vol. 60, no. 3, 2022.

[22] X. Wang, Z. Li, and P. Liu, "Recycling/production and incentive-penalty strategies in closed-loop supply chains under remanufacturing policies," *Sustainability*, vol. 15, no. 12, p. 9750, 2023.

[23] B. Cheng, J. Huang, J. Li, S. Chen, and H. Chen, "Improving contractors' participation of resource utilization in construction and demolition waste through government incentives and punishments," *Environmental Management*, vol. 70, no. 4, pp. 666–680, 2022.

[24] D. Vorobeva, I. J. Scott, T. Oliveira, and M. Neto, "Adoption of new household waste management technologies: The role of financial incentives and pro-environmental behavior," *Journal of Cleaner Production*, vol. 362, no. 1, p. 132328, 2022.

[25] L. T. T. Loan and R. M. Balanay, "Towards reinforcing the waste separation at source for Vietnam's waste management: Insights from the Nudge Theory," *Environmental Challenges*, vol. 10, p. 100660, 2023.

[26] T. Trushna, K. Krishnan, R. Soni, S. Singh, M. Kalyanasundaram, K. S. Annerstedt, A. Pathak, M. Purohit, C. S. Lundbog, Y. Sabde, S. Atkins, K. C. Sahoo, K. Rousta, and V. Diwan, "Interventions to promote household waste segregation: A systematic review," *Heliyon*, vol. 10, no. 2, p. e24332, 2024.

[27] A. Ceschi, R. Sartori, S. Dickert, A. Scalco, E. M. Tur, F. Tommasi, and K. Delfini, "Testing a norm-based policy for waste management: An agent-based modeling simulation on nudging recycling behavior," *Journal of Environmental Management*, vol. 294, p. 112938, 2021.

[28] H. Ma, M. Li, X. Tong, and P. Dong, "Community-level household waste disposal behavior simulation and visualization under multiple incentive policies—an agent-based modelling approach," *Sustainability*, vol. 15, no. 13, p. 10427, 2023.

[29] M. Taraghi and L. Yoder, "Integrating the theory of planned behavior in agent-based models: A systematic review of applications of pro-environmental behaviors," *Ecological Modelling*, vol. 508, p. 111231, 2025.

[30] X. Meng, Z. Wen, and Y. Qian, "Multi-agent based simulation for household solid waste recycling behavior," *Resources, Conservation and Recycling*, vol. 128, pp. 535–545, 2018.

[31] A. Subedi, S. Shrestha, A. Ghimire, and S. R. Paudel, "Leveraging machine learning for sustainable solid waste management: A global perspective," *Sustainable Futures*, vol. 10, p. 101098, 2025.

[32] D. Centola, J. Becker, D. Brackbill, and A. Baronchelli, "Experimental evidence for tipping points in social convention," *Science*, vol. 360, no. 6393, pp. 1116–1119, 2018.

[33] K. Nyborg, J. M. Anderies, A. Dannenberg, T. Lindahl, C. Schill, M. Schluter, W. N. Adger, K. J. Arrow, S. Barrett, S. Carpenter, F. S. Chapin, A. Crepin, G. Daily, P. Ehrlich, C. Folke, W. Jager, N. Kautsky, S. A. Levin, O. J. Madsen, S. Polasky, M. Scheffer, B. Walker, E. U. Weber, J. Wilen, A. Xepapadeas, and A. de Zeeuw, "Social tipping dynamics in environmental policy," *Review of Environmental Economics and Policy*, vol. 18, no. 1, pp. 3–23, 2024.

[34] J. Corcoran, T. Kelly, and K. Rousta, "Habit formation in waste management: The role of differing cognitive scripts," *Journal of Environmental Psychology*, vol. 71, p. 101445, 2020.

[35] H. Andre, J. Bone, M. Silva, and G. Audley, "Social norms and the persistence of pro-environmental behavior," *Global Environmental Change*, vol. 63, pp. 102–115, 2021.

[36] K. Farrow, G. Grolleau, and L. Ibanez, "Social norms and pro-environmental behavior: A review of the evidence," *Annual Review of Resource Economics*, vol. 12, pp. 400–423, 2020.

[37] A. Brugière, N. Doanh, and A. Drogoul, "Handling multiple levels in agent-based models of complex socio-environmental systems: A comprehensive review," *Frontiers in Applied Mathematics and Statistics*, vol. 8, 2022.

[38] V. M. de Souza, J. Bloemhof, and M. Borsato, "Assessing the eco-effectiveness of a solid waste management plan using agent-based modelling," *Waste Management*, vol. 125, pp. 235–248, 2021.

[39] X. Tian, F. Peng, J. Xie, and Y. Liu, "Agent-based modeling for an end-of-life power battery cross-regional recycling system and subregional policy analysis: A case study in China," *Journal of Cleaner Production*, vol. 441, p. 141054, 2024.

[40] C.-H. Liao, "Exploring social media determinants in fostering pro-environmental behavior: Insights from social impact theory and the theory of planned behavior," *Frontiers in Psychology*, vol. 15, 2024.

[41] M. R. Mousavi and K. Niazmand, "A deep reinforcement learning approach for creating diverse and adaptive agents in agent-based simulations," *SIMULATION*, vol. 97, no. 6, pp. 375–389, 2021.

[42] L. Biré, Q. N. Phung, P. Taillandier, D. A. Phung, N. D. Nguyen, and A. Drogoul, "RÁC: A serious agent-based simulation game to drive discussion on waste management in Vietnamese irrigation systems," *Journal of Artificial Societies and Social Simulation*, vol. 28, no. 2, 2025.

[43] S. Zheng, A. Trott, S. Srinivasan, N. Naik, M. Gruesbeck, D. C. Parkes, and R. Socher, "The AI Economist: Taxation policy design via two-level deep multiagent reinforcement learning," *Science Advances*, vol. 8, no. 18, p. eabk2607, 2022.

[44] K. Rajesh and S. R. Kumar, "Deep reinforcement learning for urban air quality management: Multi-objective optimization of pollution mitigation booth placement," *IEEE Access*, vol. 13, p. 146504, 2025.

[45] V. Kompella, R. Capobianco, S. Jong, K. Browne, S. Fox, L. Meyers, P. Wurman, and P. Stone, "Agent-based Markov modeling for improved COVID-19 mitigation policies," *Journal of Artificial Intelligence Research*, vol. 69, pp. 1–13, 2020.

[46] M. Vecerik, T. Hester, J. Scholz, F. Wang, O. Pietquin, B. Piot, N. Heess, T. Rothörl, T. Lampe, M. Riedmiller, and R. Guryeva, "Leveraging demonstrations for deep reinforcement learning on robotics tasks with sparse rewards," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.

[47] D. Amodei, C. Olah, J. Steinhardt, P. Christiano, J. Schulman, and D. Mané, "Concrete problems in AI safety," in *Proceedings of the 30th Conference on Neural Information Processing Systems (NeurIPS 2016)*, 2016.

[48] Y. Cheng, L. Huang, X. Ma, and Y. Pan, "Heuristic-guided reinforcement learning for efficient exploration in large-scale scenarios," *IEEE Transactions on Cybernetics*, vol. 52, no. 10, pp. 10 452–10 465, 2021.

[49] A. Y. Ng, D. Harada, and S. Russell, "Policy invariance under reward transformations: Theory and application to reward shaping," in *Proceedings of the 16th International Conference on Machine Learning (ICML)*, vol. 99, 1999, pp. 278–287.

[50] M. A. Duhayyim, T. A. E. Eisa, F. N. Al-Wesabi, A. Abdelmaboud, M. A. Hamza, A. S. Zamani, M. Rizwanullah, and R. Marzouk, "Deep reinforcement learning enabled smart city recycling waste object classification," *Computers, Materials & Continua*, vol. 71, no. 3, pp. 5699–5715, 2022.

[51] N. Khan, K. Kulkarni, Y. Mahale, S. Kolhar, and S. Mahajan, "Waste objects segregation using deep reinforcement learning with deep Q networks," *Ingénierie des Systèmes d'Information*, vol. 29, no. 6, pp. 2219–2229, 2024.

[52] C. Hertweck and V. Dignum, "Graduated punishment in a deep reinforcement learning agent-based model of common-pool resources," in *Multi-Agent-Based Simulation XXIII*, vol. 13833. Springer, 2023, pp. 109–122.

[53] S. S. Mousavi, M. Schukat, and E. Howley, "Deep reinforcement learning: An overview," in *Proceedings of SAI Intelligent Systems Conference*. Springer, 2021, pp. 426–440.

[54] T. Yazawa, K. N. Tablada, K. M. Baring, K. L. Alojipan, and M. Watanabe, "Act and reality of the ecological solid waste management act on barangay-level waste management in Barbaza, the Philippines," *Discover Sustainability*, vol. 6, no. 1, 2025.

[55] S. M. Apostol-Jamoralin, "Assessing the implementation of Republic Act 9003: Ecological solid waste management in Sorsogon City," *Sorsogon Multidisciplinary Research Journal*, vol. 3, no. 1, pp. 15–29, 2024.

[56] D. B. Olawade, O. Fapohunda, O. Z. Wada, S. O. Usman, A. O. Ige, O. Ajisafe, and B. I. Oladapo, "Smart waste management: A paradigm shift enabled by artificial intelligence," *Waste Management Bulletin*, vol. 2, no. 2, 2024.

[57] R. Udayakumar, R. Elankavi, V. R. Vimal, and R. Sugumar, "Improved particle swarm optimization with deep learning-based municipal solid waste management in smart cities," *Revista de Gestão Social e Ambiental*, vol. 17, no. 4, p. e03561, 2023.

[58] C. Hertweck and V. Dignum, "Values in reinforcement learning: A definition and categorization," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 12, 2023, pp. 14 782–14 790.

[59] B. Moeini, E. Ayubi, M. Barati, S. Bashirian, L. Tapak, K. Ezzati-Rastgar, and M. Hashemian, "Effect of household interventions on promoting waste segregation behavior at source: A systematic review," *Sustainability*, vol. 15, no. 24, p. 16546, 2023.

[60] K. Nishimura, "How does decentralization affect the performance of municipalities in urban environmental management in the Philippines?" *Lex Localis - Journal of Local Self-Government*, vol. 20, no. 4, pp. 715–738, 2022.

[61] I. M. Sobol, "Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates," *Mathematics and Computers in Simulation*, vol. 55, no. 1–3, pp. 271–280, 2001.

[62] R. Badua, "Effectiveness of the implementation on "No Segregation, No Collection" measures (Bacnotan MO No. 481 s. 2014 CII S7): The case of DMMMSU adjacent barangays," *DMMMSU Research and Extension Journal*, vol. 6, no. 6, pp. 21–38, 2022.