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

Hussam M. Bansao, Jemar John J. Lumingkit, Dante D. Dinawanao
College of Computer Studies, MSU-Iligan Institute of Technology

***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

This study systematically quantified and determined the optimal settings for policy instruments, specifically incentive, punitive, and information and educational campaign measures, required to maximize sustained household solid waste segregation compliance within the Municipality of Bacolod, Lanao del Norte, by taking into account local socioeconomic determinants and budget-constraints.

This study answered the following specific questions:

1) How did variations in the synthesized household behavioral parameters (e.g., the relative weight of Subjective Norms vs. Perceived Behavioral Control, derived from literature and LGU records) affect the stability and efficacy of policy outcomes within the Agent-Based Model?
2) What was the optimal long-term resource allocation ratio among the three policy levers (monetary incentives, punitive enforcement, and educational campaigns) that maximized compliance per peso spent, as determined by the Reinforcement Learning agent?
3) Which dynamic policy strategy yielded the highest overall compliance and cost-benefit ratio for the LGU while strictly adhering to the defined annual budget constraint?

B. Research Objectives

The primary objective of this study was to develop and apply a coupled Agent-Based Model (ABM) and Deep Reinforcement Learning (DRL) framework to determine the optimal, budget-constrained allocation of resources across policy levers for maximizing household solid waste segregation compliance in the Municipality of Bacolod.

Specific objectives were:

1) To conduct a comprehensive synthesis of academic literature and utilize contextual financial and operational data from the Philippine Statistics Authority and LGU records, including interviews with key implementing officers, to rigorously parameterize the ABM.
2) To develop a Multi-Level Agent-Based Model where household agent behavior was governed by a utility function incorporating Theory of Planned Behavior constructs and socioeconomic variables, and where policy levers dynamically updated behavioral constructs.
3) To integrate a Reinforcement Learning algorithm that enabled the LGU agent to autonomously learn the opti-

mal policy (allocating funds among incentives, enforcement staff, and education campaign) that maximized a composite reward function balancing compliance and financial cost, while adhering to a defined budget constraint.
4) To simulate and validate the efficacy and cost-effectiveness of budget allocation strategies (Pure Incentive, Pure Penalty, Pure Information Education Campaign, and Hybrid regimes) and provide actionable, data-driven recommendations on the optimal resource mix for the LGU 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 waste management is governed by the Ecological Solid Waste Management Act of 2000 (R.A. 9003), which mandates source segregation, recycling, and the establishment of Materials Recovery Facilities (MRFs). However, implementation remains sub-optimal across various Local Government Units (LGUs), necessitating computational approaches to bridge the gap between national policy and local practice.

1) Systemic and Budgetary Constraints on LGUs: The primary challenge in implementing R.A. 9003 is the significant operational and financial burden placed on LGUs as the chief implementers. Solid waste management constitutes a high financial drain on municipal budgets, a situation compounded by a scarcity of compliant sanitary landfills and chronic underfunding for local initiatives. These deficiencies often lead to systemic institutional failures that weaken the effectiveness of the law. Consequently, overcoming these constraints requires significant political will and initiative from local officials to substantially improve municipal performance.

2) Behavioral Drivers and Resident Non-Compliance: Understanding the behavioral drivers of residents is essential for effective policy design. Non-compliance is frequently driven by structural friction—specifically, frustration with irregular collection services—rather than a simple lack of awareness. In some communities, minor environmental offenses have become normalized, requiring targeted interventions. While educational campaigns such as the SURWEM project raise awareness, they do not guarantee sustained behavioral change without accompanying structural support. Therefore, building trust through participatory governance and community involvement is vital for increasing compliance rates.

3) Heterogeneity and Policy Equity: Policy design must account for the socio-economic diversity of the population to ensure justice. Purely financial penalties, such as fines, are known to be regressive as they disproportionately affect low-income groups. To address this, reinforcement learning agents must optimize for both cost-efficiency and policy equity. This requires simulating agents with varying sensitivities to incentives and penalties based on their distinct socio-economic profiles.

B. Household Behavior and Theory of Planned Behavior

C. Agent-Based Modeling as a Markov Decision Process

D. Heuristic-Guided Deep Reinforcement Learning

E. Research Gap

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 [5], [6].

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 **??** based on personnel wages and coverage ratios; (2) **Monetary Incentives** (C_{Inc}), modeled in Equation **??** 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 **??** as tiered fixed costs for media dissemination.

C. Multi-Level Agent-Based Model Architecture

The simulation was developed using the Python library *MESA* [7], a choice supported by recent reviews of computational tools in solid waste management [1], [8]. The architecture employed a strict hierarchical design within a single Agent-Based Model (ABM) environment, containing seven distinct `BarangayAgent` objects (representing Liangan East, Esperanza, Poblacion, Binuni, Demologan, Mati, and Babalaya), which in turn encapsulated their respective populations of `HouseholdAgent` objects. This multi-level structure accurately captured the decentralized governance flow and heterogeneity of the Municipality of Bacolod [9]. The model operated on quarterly time steps, where `BarangayAgents` managed fiscal allocation, policy implementation (e.g., "No Segregation, No Collection"), and enforcement deployment, bridging municipal mandates with local execution [10]. `HouseholdAgents` made binary segregation decisions based on a Theory of Planned Behavior (TPB) utility function, weighing policy strictness, social norms, and infrastructure availability against the "cost of effort" [11].

To enable autonomous policy optimization, the system was formalized as a finite-horizon Markov Decision Process (MDP), where an LGU-DRL agent dynamically allocated a continuous budget across Enforcement, Incentives, and IEC levers to maximize long-term compliance while minimizing costs [12]. Crucially, the model incorporated a "Norm Internalization Mechanism" to simulate behavioral hysteresis, where sustained high compliance ($> 70\%$) created internalized social habits that resisted decay even when external interventions were reduced [13].

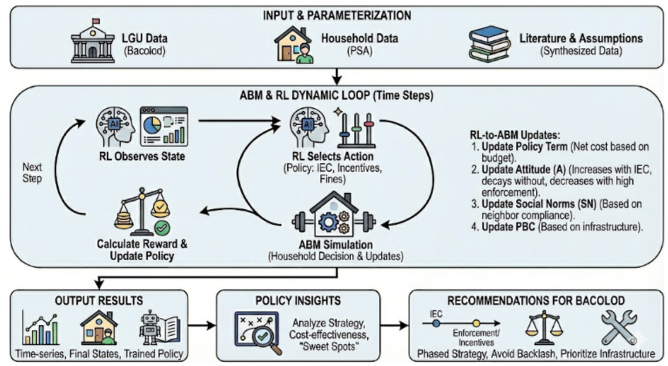

Fig. 1. Baseline Status Quo

D. Heuristic-Guided Deep Reinforcement Learning (HuDRL) Framework

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

E. Simulation and Analysis

This study employed a rigorous four-stage experimental design to operationalize the coupled ABM-RL framework and isolate the causal impact of specific policy interventions. The process began with **Initialization and Calibration**, where the model was grounded in demographic and operational data from seven specific barangays and iteratively tuned to replicate the municipality's baseline compliance rate (approx. 10%) [14], [15]. Following calibration, the Heuristic-Guided Deep Reinforcement Learning (HuDRL) agent underwent extensive training over 12 simulated lifetimes (approx. 10,000 periods) to explore the high-dimensional policy space [2]. To evaluate governance efficacy, the study simulated three distinct **Policy Scenarios**: a *Pure Penalty Regime* (representing the status quo/enforcement-heavy approach), a *Pure Incentive Regime* ("soft" governance), and a *Hybrid Regime* (an adaptive, AI-optimized mix of IEC, incentives, and enforcement) [1]. System performance was quantified using four key metrics: Maximum Sustainable Compliance, Cost-Effectiveness (compliance gained per peso), Policy Equity (variance across barangays), and Optimal Resource Allocation. Finally, the model's structural robustness was validated through a **Global Sensitivity Analysis** using the Sobol method [16], which stress-tested the influence of core behavioral parameters (w_a, w_{sn}, c_{effort}) on global compliance outcomes.

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

Barangay	Households	Annual Budget (P)	Initial Comp.	Income Profile (%) 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 w_a	w_{sn}	C_e	γ	Allocation (%) 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 ($\approx$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 AI-driven dynamic allocation. Without "Sequential Saturation," static rebalancing was insufficient to resolve non-compliance.

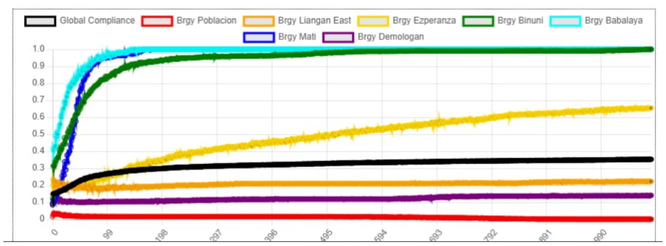

Fig. 2. Baseline Status Quo

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%.

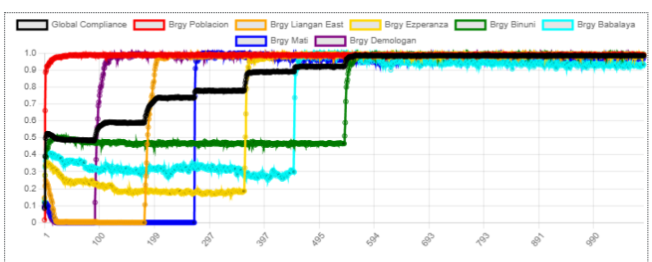

Fig. 3. Heuristic-Guided Deep Reinforcement Learning Approach

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.

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

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%**

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.

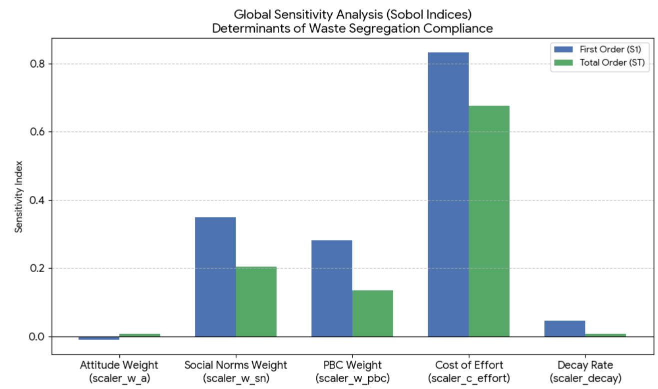

Fig. 4. Global Sensitivity Analysis Results illustrating First-Order Sobol Indices (S_1) and Total Order Indices (ST) for behavioral parameters.

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 [15].

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 [17].

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 [18] and [14], 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" [20].

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 [22].

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 [23].

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 [13]. 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

A. Summary of Findings

The evaluation of the Agent-Based Model through Global Sensitivity Analysis (GSA) and behavioral determinant mapping provided a robust computational foundation for the DRL agent's optimized policies. By transitioning from a "Status Quo" approach to a "Sequential Saturation" strategy, the model demonstrated how municipal resources could be leveraged more effectively by targeting the underlying mechanics of household decision-making [24].

The primary discovery of this evaluation was the dominance of convenience as a behavioral driver. With the Cost of Effort (c_{effort}) accounting for 83% of output variance, the study confirmed that logistical friction was the single greatest inhibitor of waste segregation in the Municipality of Bacolod [18]. Policies that did not actively reduce this "cost" were statistically likely to fail regardless of the level of public support.

Furthermore, the results highlighted a distinct "Value-Action Gap" characterized by the near-zero sensitivity of the Attitude (w_a) parameter. This mathematically illustrated that while IEC campaigns were foundational, they had reached a point of diminishing returns [21]. The DRL agent's strategic shift toward extrinsic motivators was thus justified; awareness alone could not bridge the gap between intention and action when systemic barriers remained high.

Finally, the model identified social multipliers as the key to long-term sustainability. The significant influence of Social Norms ($w_{sn} \approx 0.35$) provided the mechanism for the "Graduation Effect," allowing for the eventual reallocation of funds without a collapse in segregation rates [23]. In conclusion, the DRL agent's preference for localized enforcement and incentives was a sophisticated response to the socio-technical barriers in Philippine waste management, offering a data-driven roadmap toward a resilient compliance ecosystem [15].

VII. Acknowledgement

The authors would like to thank the Local Government Unit of Bacolod, Lanao del Norte, specifically the Municipal Environment and Natural Resources Office (MENRO), for providing the crucial data for this study.

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] 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.

[6] 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.

[7] e. a. ter Hoeven, E., "Mesa 3: Agent-based modeling with python in 2025," *Journal of Open Source Software*, vol. 10, no. 107, p. 7668, 2025.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[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] 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.

[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] 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.

[23] 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.

[24] 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.