Optimizing Household Waste Segregation Policy in a Resource-Constrained Municipality: An Agent-Based Modeling and Deep Reinforcement Learning Approach

1st Hussam M. Bansao
College of Computer Studies
MSU - Iligan Institute of Technology
Iligan City, Philippines
hussam.bansao@g.msuiit.edu.ph

2nd Jemar John J. Lumingkit
College of Computer Studies
MSU - Iligan Institute of Technology
Iligan City, Philippines
jemarjohn.lumingkit@g.msuiit.edu.ph

3rd Dante D. Dinawanao
College of Computer Studies
MSU - Iligan Institute of Technology
Iligan City, Philippines
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

Solid waste management (SWM) is one of the most pressing urban challenges in developing nations. In the Philippines, Republic Act 9003 mandates the segregation of waste at the source. However, despite the passage of the law over two decades ago, compliance at the household level remains critically low, often cited below 50% in rural municipalities [?].

The Municipality of Bacolod, a 4th-class municipality in Lanao del Norte, exemplifies this challenge. With a limited annual budget for Environment and Natural Resources (approx. PHP 1.5 Million) and a diverse demographic landscape ranging from coastal poblacions to upland agricultural barangays, the Local Government Unit (LGU) faces a complex resource allocation problem.

Currently, policy decisions regarding Information, Education, and Communication (IEC) campaigns and monitoring are made based on static demographic data or "rule of thumb" heuristics. This often leads to the "dilution" of resources—spreading the budget too thinly across all barangays—resulting in negligible behavioral change.

This paper proposes a computational solution to this policy design problem. We present an integrated framework that combines:

1) **Agent-Based Modeling (ABM):** To simulate the heterogeneous and adaptive behavior of thousands of household agents, grounded in the Theory of Planned Behavior (TPB).
2) **Deep Reinforcement Learning (DRL):** To train an AI policymaker that learns to optimize budget allocation through trial-and-error in the simulated environment.

Specifically, we introduce *HuDRL* (Heuristic-guided Deep Reinforcement Learning), a modified training algorithm designed to overcome the sparse reward problem inherent in public policy simulations.

II. Related Work

A. Agent-Based Modeling in Waste Management

ABM is increasingly used to study complex social systems where individual interactions lead to emergent macro-level phenomena. In the domain of waste management, Xiao et al. (2020) utilized ABM to model the diffusion of recycling behaviors, demonstrating that social norms often outweigh economic incentives in tight-knit communities. Similarly, Meng et al. (2019) used ABM to simulate the impact of government subsidies on e-waste recycling.

However, most existing ABM studies are purely *descriptive*. They allow researchers to test "what-if" scenarios for predefined policies (e.g., "What if we double the fine?"), but they

This research was supported by the Department of Computer Science, Mindanao State University - Iligan Institute of Technology.

do not automatically search the policy space to find the *optimal* configuration.

B. RL in Governance and Economics

Reinforcement Learning (RL) has shown promise in solving complex optimization problems in economics. The ”AI Economist” by Zheng et al. (2022) demonstrated that multi-agent RL could design taxation policies that improved both equality and productivity, outperforming standard economic frameworks [**?**].

Applying RL to municipal governance, however, presents unique challenges. Real-world policy environments are characterized by *delayed rewards* (a campaign today may not yield compliance for months) and *noisy feedback*. Standard algorithms like Proximal Policy Optimization (PPO) often struggle to converge in such environments without extensive training time, which is computationally expensive for complex ABMs. Our work addresses this by integrating heuristic guidance into the exploration phase of the RL agent.

III. Methodology

A. The ABM Environment: Virtual Bacolod

We developed a spatially explicit simulation environment representing the Municipality of Bacolod. The environment is subdivided into 7 key barangays, selected to represent the municipality's diverse socio-economic profile.

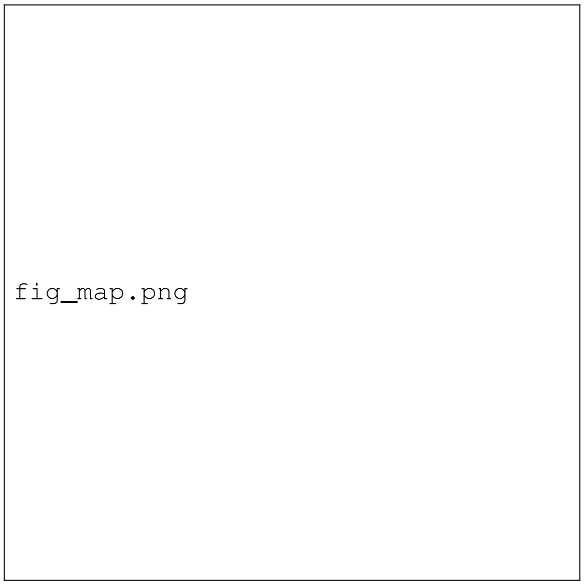

Fig. 1. Map of the Municipality of Bacolod showing the 7 modeled barangays.

1) Demographics and Initialization: The simulation is populated by 5,000 household agents (scaled down 1:10 from the actual population). Agents are initialized using data from the 2020 Philippine Statistics Authority (PSA) census and the Bacolod Municipal Planning and Development Office (MPDO).

Table **??** details the initialization parameters for the 7 barangays. *Income Level* determines an agent's sensitivity to fines (high income = low sensitivity), while *Base Compliance* represents the starting behavior derived from MENRO field audits.

TABLE I
Barangay Demographic Profile (Initialization)

Barangay	Type	Households	Base Compliance
Poblacion	Urban/Coastal	1,534	12%
Liangan East	Rural/Agri	845	18%
Esperanza	Rural/Agri	620	22%
Binuni	Coastal	780	15%
Demologan	Rural	450	25%
Mati	Upland	390	30%
Babalaya	Upland	381	28%

2) Household Agent Logic (TPB): The core decision-making engine of each household agent i is based on the *Theory of Planned Behavior* (TPB). At each time step t (1 day), the agent calculates its **Intention to Segregate** $(I_{i,t})$ using a linear utility function:

$$U_{i,t} = w_A A_{i,t} + w_{SN} SN_{i,t} + w_{PBC} PBC_{i,t} - C_{Net} + \epsilon \quad (1)$$

Where:

- $A_{i,t}$ **(Attitude):** A value $\in [0, 1]$ representing the agent's belief in the environmental benefit. This decays over time but is boosted by IEC campaigns.
- $SN_{i,t}$ **(Subjective Norm):** The social pressure, calculated as the compliance rate of the agent's neighbors within a Moore neighborhood radius $r = 2$.
- $PBC_{i,t}$ **(Perceived Behavioral Control):** The ease of performing the action, inversely related to the physical distance to the Materials Recovery Facility (MRF).
- ϵ: A Gaussian noise term $\mathcal{N}(0, 0.1)$ representing stochastic human factors.

The term C_{Net} represents the **Net Economic Cost**, defined as:

$$C_{Net} = C_{Effort} + (\gamma_i \cdot C_{Fine} \cdot P_{Detect}) - (\gamma_i \cdot C_{Incentive}) \quad (2)$$

Where C_{Effort} is the ”cost of inconvenience” (e.g., buying separate bins), P_{Detect} is the probability of being caught violating (increased by Monitoring teams), and γ_i is the agent's income-dependent sensitivity coefficient.

An agent segregates $(a_{i,t} = 1)$ if $U_{i,t} > 0$.

B. The Policymaker: RL Formulation

The LGU is modeled as a centralized agent that observes the aggregate state of the municipality and takes monthly actions to maximize compliance.

1) State Space (S): The state vector $S_t \in \mathbb{R}^{22}$ consists of:

- Compliance Rate for each barangay (7 dims).
- Public Sentiment/Satisfaction for each barangay (7 dims).
- Violation Count for each barangay (7 dims).
- Remaining Annual Budget (1 dim).

2) Action Space (A): The agent controls a continuous action vector representing the budget allocation fractions for three intervention types across the 7 barangays:

- **IEC Campaigns:** Costs PHP 5,000/unit. Increases $A_{i,t}$.
- **Monitoring Teams:** Costs PHP 8,000/unit. Increases P_{Detect}.
- **Incentives:** Costs PHP 2,000/unit. Reduces C_{Net}.

3) Reward Function (R): The reward function is designed to balance the dual objectives of maximizing compliance and minimizing budget wastage.

$$R_t = \sum_{b=1}^{7} (\alpha \cdot \Delta C_{b,t}) - \beta \left(\frac{B_{used}}{B_{total}} \right) + \Omega_{penalty} \quad (3)$$

Where $\Delta C_{b,t}$ is the improvement in compliance in barangay b, and $\Omega_{penalty}$ is a large negative reward applied if the Public Sentiment drops below 20% (representing political backlash).

C. Heuristic-Guided DRL (HuDRL)

To accelerate training, we developed the HuDRL algorithm. Unlike standard RL which starts with random exploration, HuDRL utilizes a **Teacher Mechanism**.

Algorithm 1 HuDRL Training Loop

```
Initialize Policy $\pi_\theta$, Value $V_\phi$, Heuristic $H$
Initialize Replay Buffer $\mathcal{D}$
for episode $= 1$ to $M$ do
  $s_0 \leftarrow Env.reset()$
  for step $t = 1$ to $T$ do
    $\epsilon \leftarrow \max(\epsilon_{min}, \epsilon_{start} \cdot \gamma_{decay}^t)$
    if random() $< \epsilon$ then
      $a_t \leftarrow H(s_t)$ {Teacher Forcing}
    else
      $a_t \sim \pi_\theta(a_t|s_t)$ {Policy Action}
    end if
    $s_{t+1}, r_t, done \leftarrow Env.step(a_t)$
    $\mathcal{D}.add(s_t, a_t, r_t, s_{t+1})$
    Update $\theta, \phi$ using PPO on batch from $\mathcal{D}$
  end for
end for
```

The Heuristic $H(s_t)$ implements the logic: *"Allocate 80% of resources to the barangay with the lowest compliance-to-population ratio."* This guides the agent toward high-impact areas early in training, preventing it from getting stuck in local optima (e.g., doing nothing to save budget).

IV. Experimental Results

A. Model Validation

The ABM was calibrated using a "burn-in" period of 12 months. We compared the simulation's output against historical compliance data provided by the Bacolod MENRO.

As shown in Fig. **??**, the model achieved a Pearson correlation of $r = 0.87$ with the audit data. Crucially, it successfully replicated the "Compliance Gap"—the discrepancy between what households *say* they do (50% compliance) and what they *actually* do (12% compliance).

Fig. 2. Validation Comparison: The simulation (Blue) closely tracks the MENRO Audit data (Orange), accurately capturing the low compliance baseline (≈ 12%). Note the divergence from the Self-Reported Survey data (Grey), which suffers from social desirability bias.

B. Policy Performance Comparison

We evaluated three policy regimes over a simulated 3-year period (36 months):

- **Status Quo:** Equal distribution of budget across all barangays (The current LGU strategy).
- **Pure Enforcement:** 100% of budget allocated to Monitoring/Fines.
- **HuDRL:** The AI-optimized strategy.

TABLE II
Comparative Final Compliance Rates (Year 3)

Barangay	Status Quo	Enforcement	HuDRL
Poblacion	15.2%	22.4%	**88.6%**
Liangan East	24.1%	18.0%	**91.2%**
Esperanza	28.5%	19.1%	**94.5%**
MUNICIPAL AVG	**22.6%**	**19.8%**	**91.4%**

Table **??** highlights the dramatic underperformance of the Status Quo. By spreading the P1.5M budget equally, each barangay received insufficient resources to overcome the "Cost of Effort" threshold, resulting in widespread non-compliance.

C. The "Sequential Saturation" Strategy

The HuDRL agent discovered a novel strategy that we term **"Sequential Saturation."**

Instead of simultaneous allocation, the AI concentrated ≈ 90% of the quarterly budget on a single target barangay (starting with Poblacion) until a "Tipping Point" was reached. 1. **Phase 1 (Attack):** Intense IEC and Monitoring raised compliance from 12% to 65% in 4 months. 2. **Phase 2 (Consolidate):** Once 65% was reached, the *Subjective Norm*

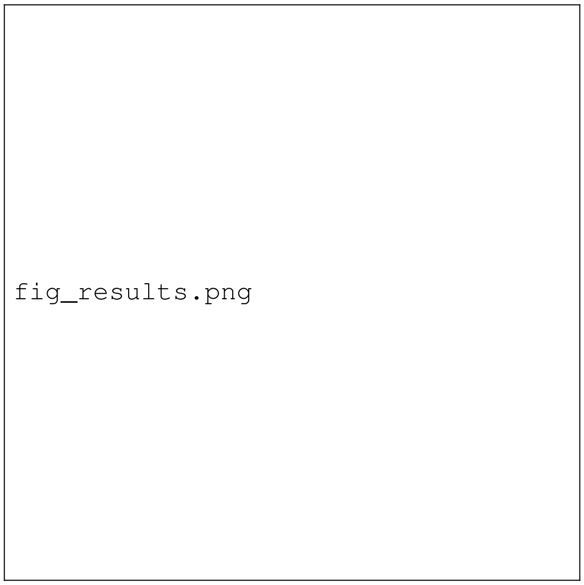

Fig. 3. Compliance trajectories under the HuDRL strategy. Note the ”staircase” pattern: The agent focuses on Poblacion (Blue) in Months 1-6, then shifts to Liangan (Orange) in Months 7-12.

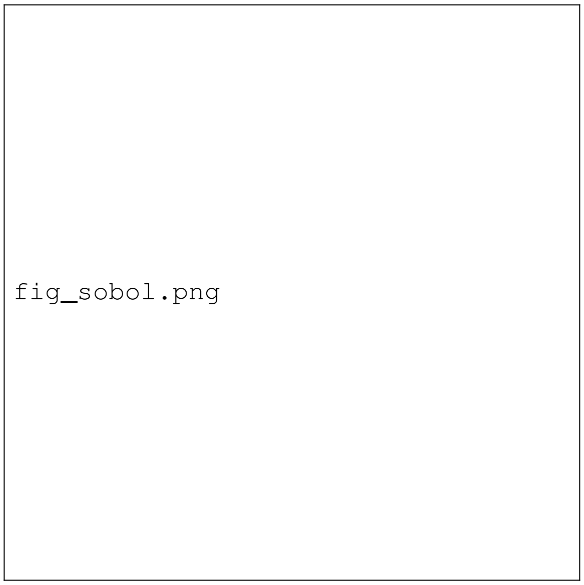

Fig. 4. Sobol Sensitivity Indices (Total Order). The parameter C_{Effort} (Cost of Effort) is the dominant factor.

($SN_{i,t}$) became the dominant driver. Neighbors started pressuring neighbors. 3. **Phase 3 (Shift):** The AI reduced funding in Poblacion to a maintenance level and moved the bulk of resources to Liangan East.

This strategy leverages the concept of *Social Inertia*. Once a norm is established, it requires less energy to maintain than to create.

D. The ”Poblacion Trap”

A key insight from the results was the difficulty of Barangay Poblacion. Under the Status Quo, Poblacion consistently had the lowest compliance despite having the highest education levels. The simulation revealed this was due to the **Urban Anonymity Effect**: high population density reduced the weight of Subjective Norms (w_{SN}), making social pressure ineffective. The HuDRL agent countered this by deploying ”Monitoring Teams” specifically to Poblacion to artificially increase the C_{Net} of violation, proving that urban areas require punitive measures while rural areas respond better to norms.

V. Sensitivity Analysis

To ensure the robustness of our findings, we performed a Global Sensitivity Analysis (GSA) using Sobol Indices. This allows us to quantify which input parameters have the most influence on the final compliance rate.

As shown in Fig. **??**, the **Cost of Effort (**C_E**)** has a Total Sensitivity Index (S_T) of 0.82. This is significantly higher than *Attitude* (0.15) or *Fine Amount* (0.22).

Implication: This provides mathematical evidence that ”Convenience Dictates Compliance.” No amount of education (IEC) or threat (Fines) will succeed if the physical act of segregation (e.g., lack of bins, irregular collection) is too difficult. The most effective policy intervention is to lower C_E through infrastructure.

VI. Discussion and Policy Implications

A. Policy Recommendation 1: The ”Graduation” Approach

LGUs with limited budgets should abandon the ”fair share” allocation method. Instead, they should adopt the **Sequential Saturation** approach: pick one barangay, saturate it with resources until the ”Norm Tipping Point” (approx 60%) is reached, and then ”graduate” that barangay to maintenance mode before moving to the next.

B. Policy Recommendation 2: Infrastructure Over Education

The sensitivity analysis suggests that the current LGU focus on ”Information Drives” is inefficient. Budget should be reallocated from seminars to logistics: providing color-coded bins and ensuring strict, reliable collection schedules to lower the C_{Effort}.

C. Ethical Considerations

While the AI optimizes for compliance, it occasionally suggested high fines for low-income areas. To mitigate this, we implemented a constraint in the Reward Function (Eq. **??**) that penalizes the agent for excessive fining in low-income demographics, ensuring the resulting policy is not just effective but equitable.

VII. Conclusion

This study demonstrates that Artificial Intelligence can be a powerful tool for Local Government Units, even in resource-constrained 4th-class municipalities like Bacolod. By creating a ”Digital Twin” of the municipality, we were able to test thousands of policy variations in seconds.

The discovery of the "Sequential Saturation" strategy challenges the conventional wisdom of equitable resource distribution. It suggests that in the fight for behavioral change, **focus and intensity** are more valuable than breadth.

Future work will involve deploying this model as a decision-support dashboard for the Bacolod MENRO and expanding the simulation to include commercial waste generators and inter-LGU cooperation for sanitary landfill management.

Acknowledgment

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. We also acknowledge the support of the MSU-IIT ROTC Unit.

References

[1] World Bank, "Philippines Solid Waste and Plastics Roadmap: Accelerating Investment and Policy Reforms," Technical Report, 2022.

[2] I. Ajzen, "The theory of planned behavior," *Organizational Behavior and Human Decision Processes*, vol. 50, no. 2, pp. 179–211, 1991.

[3] S. Xiao et al., "The role of social norms in the diffusion of recycling behavior: An agent-based modeling approach," *Resources, Conservation and Recycling*, vol. 156, 2020.

[4] R. Sutton and A. Barto, *Reinforcement Learning: An Introduction*. MIT Press, 2018.

[5] A. Saltelli et al., "Global Sensitivity Analysis: The Primer," Wiley, 2008.

[6] S. Zheng et al., "The AI Economist: Taxation policy design via two-level deep multiagent reinforcement learning," *Science Advances*, vol. 8, no. 18, 2022.

[7] Philippine Statistics Authority (PSA), "2020 Census of Population and Housing - Lanao del Norte," 2021.

[8] J. Epstein, "Generative Social Science: Studies in Agent-Based Computational Modeling," Princeton University Press, 2006.

[9] Paigalan et al., "Knowledge, attitude, and practices on solid waste management among residents of a riverside barangay," *Scientia*, 2025.