Proceedings of the 2025 Winter Simulation Conference
E. Azar, A. Djanatliev, A. Harper, C. Kogler, V. Ramamohan, A. Anagnostou, and S. J. E. Taylor, eds.

A TWO-STAGE SIMULATION FRAMEWORK FOR EVALUATING AI POLICY RECOMMENDATIONS: A CASE STUDY OF COVID-19

Amir Abdollahi[1], Reyhaneh Mohammadi[1], Jacqueline Griffin[1], and Casper Harteveld[1,2]

[1]Dept. of Mechanical and Industrial Eng., Northeastern University, Boston, MA, USA
[2]College of Arts, Media and Design (CAMD), Northeastern University, Boston, MA, USA

ABSTRACT

As AI integration in critical domains grows, evaluating its effectiveness in complex policy environments remains challenging. We introduce a two-stage simulation framework for assessing AI policy recommendations in the COVID-19 pandemic. First, we train a deep reinforcement learning (DRL) agent using data from 186 countries to model optimal intervention timing and intensity. Results suggest the DRL agent outperforms average government outcomes within our simplified model under specific assumptions, reducing infections and fatalities, improving recovery rates. Second, we employ SEIRD (Susceptible-Exposed-Infected-Recovered-Dead) modeling to create a dynamic simulation environment, testing the agent across diverse scenarios beyond historical data. Unlike prior work lacking systematic evaluation, our framework provides a controlled testbed for high-stakes policy decisions before implementation. This presents a responsible approach to AI evaluation where real-world experimentation raises ethical concerns. It highlights the role of simulations in bridging development-deployment gaps while identifying financial constraints and human-AI interaction as future research priorities.

1 INTRODUCTION

As artificial intelligence becomes increasingly integrated into critical decision-making contexts, evaluating AI's effectiveness in complex policy environments presents significant challenges. The high-stakes nature of pandemic response—where policy decisions directly impact public health outcomes—makes it an ideal domain for studying AI assistance in complex decision-making (Asan et al. 2020; Maadi et al. 2021). While prior research has explored human-AI collaboration in various contexts (Amershi et al. 2019), fundamental questions remain about the potential of AI systems in complex policy formulation and how to systematically evaluate their performance before real-world deployment.

Pandemic management represents a particularly complex policy challenge. Without effective vaccines or treatments, highly contagious outbreaks require preventive interventions such as social distancing and travel restrictions to flatten infection curves and prevent healthcare system collapse (Wu et al. 2020; Chinazzi et al. 2020). Failure to implement timely and appropriate measures can lead to surging case numbers, increased mortality, and reduced recovery rates (Khadilkar et al. 2020). A research by Rahmandad et al. (2021) demonstrated that even minor improvements in policy responsiveness can reduce deaths by 14%—exceeding the impact of vaccinating half the population—underscoring the critical importance of effective policy-making during pandemic crises.

This paper explores the potential of AI-assisted policy formulation through a two-stage simulation framework designed to illuminate the underlying policy-making process. Specifically, it is designed for evaluating pandemic management scenarios. Our research questions include: (1) Can deep reinforcement learning (DRL) agents learn policy-making strategies that achieve favorable outcomes compared to historical government decisions in our simplified model? (2) Can simulation environments serve as preliminary testing grounds for exploring AI performance in simplified policy contexts?

Our main contributions include:

979-8-3315-8726-0/25/$31.00

- A two-stage simulation framework that first provides preliminary exploration of AI performance against real-world historical data, then explores performance in controlled simulation environments;
- Empirical evidence that a DRL agent trained on data from 186 countries can outperform average government responses in managing pandemic scenarios within our model constraints; and
- A stylized SEIRD-based simulation environment for preliminary evaluation of AI policy recommendations in controlled pandemic scenarios including diverse unseen pandemic scenarios.

In our first study, we conduct a preliminary exploration of a DRL agent against actual policy decisions by governments during the COVID-19 pandemic within our model constraints, categorizing interventions into no-action, moderate, and strict measures (Kwak et al. 2021). Our second study addresses the limitations of historical data by developing a controlled simulation environment based on the SEIRD epidemiological model, allowing for sequential policy-making evaluation rather than point-wise comparisons. This approach provides initial insights into AI's potential role in policy formulation within simplified models and offers a preliminary framework for future research on human-AI collaboration in high-stakes policy domains.

Our methodology and findings have implications for simulation research, reinforcement learning applications, and the design of human-AI collaborative systems across domains where complex decision-making occurs under uncertainty. By demonstrating how simulated environments can bridge the gap between AI development and responsible deployment, this work contributes to both the technical advancement of AI evaluation methodologies and the practical implementation of AI support systems in critical policy contexts.

2 BACKGROUND AND RELATED WORK

Integrating AI into complex systems poses multifaceted challenges beyond technical implementation (Gabsi 2024). Organizations face hurdles such as data integrity issues, scalability, and interoperability between AI and existing infrastructure (Ekundayo 2024). In parallel, organizational dynamics—like stakeholder acceptance, ethical risks, and mitigation of discriminatory outcomes—play a crucial role in AI adoption (Kubilay and Celiktas 2025). Adoption is often driven by perceived usefulness and usability, while trust depends on users' perceptions of competence, integrity, and benevolence (Lukyanenko et al. 2022).

Human-centered AI design emphasizes clear system capabilities, transparency in decision-making, and appropriate user control (Amershi et al. 2019; Chen and Zacharias 2024). While progress has been made in AI testing and deployment across domains, a persistent gap remains in evaluating AI effectiveness within complex policy environments (Valle-Cruz et al. 2020). This is especially critical in high-stakes scenarios like pandemics, where traditional experimentation is often infeasible or unethical (Tomašev et al. 2020). Our research addresses this gap through a two-stage evaluation framework, combining real-world data validation with simulation-based testing in controlled environments for policy-focused AI systems.

2.1 AI for Pandemic Modeling and Response

The COVID-19 crisis underscored the need for responsive, data-driven policymaking. AI approaches have been explored for modeling disease dynamics, including SEIRD models and agent-based simulations of population-level behaviors (Gawande et al. 2025). Machine learning techniques support outbreak prediction, resource optimization, and policy evaluation (Ondula 2024). Reinforcement learning (RL), in particular, has shown promise in formulating adaptive control strategies by learning optimal intervention sequences from data and simulations (Kwak et al. 2021; Li et al. 2024).

Simulated environments have become essential for training policymakers and testing AI assistants in risk-free settings (Elendu et al. 2024; Laamarti et al. 2014; Maheu-Cadotte et al. 2018; Kwak et al. 2024). These environments allow users to explore policy impacts and incorporate AI-generated insights. However, challenges persist. Data limitations hamper model training (Liu 2025), and many AI models lack mechanistic transparency and generalizability beyond observed contexts. Ethical issues—such as privacy, bias, and fairness—further complicate public health applications (Ekundayo 2024; Barocas et al. 2019).

Our research builds on these foundations while addressing key limitations. By combining DRL with a two-stage evaluation framework—first validating against real-world data, then testing in controlled simulations—we provide a structured approach to assess AI policy recommendations before deployment. This methodology specifically tackles the transparency and generalizability concerns by enabling systematic evaluation across diverse pandemic scenarios beyond historical observations.

2.2 AI Testing and Evaluation Methodologies

Traditional testing approaches fall short for AI systems due to properties like non-determinism, continuous learning, and opaque decision processes (Menzies and Pecheur 2005; Xiang et al. 2018). These limitations demand new evaluation frameworks. Key areas include robustness testing, fairness audits, and explainability assessments (Goodfellow et al. 2014). The quality of training data remains a central factor in reliability. Several standards offer guidance—ISO/IEC TR 29119-11 targets AI-specific testing challenges (Oviedo et al. 2024), while IEEE 1012 outlines lifecycle verification protocols (IEEE 2017). Emerging methods such as neurosymbolic AI seek to combine logical reasoning with learning-based approaches to enhance verification (Renkhoff et al. 2024). Yet, these typically focus on model-level validation rather than full-system testing in dynamic environments.

Evaluating AI in simulation environments introduces additional complexities. Unlike static benchmarks, interactive scenarios require assessing behavior over time using metrics like adaptability, learning speed, and decision quality (Ribeiro et al. 2020). While recent methods improve component verification (Renkhoff et al. 2024), few offer comprehensive evaluation across both historical data and future-oriented simulations. Our framework addresses this gap by integrating retrospective validation with simulation-based testing, enabling end-to-end assessment of AI policy systems. This dual-stage method aims to improve both reliability in known scenarios and adaptability in prospective ones—critical for real-world deployment in public policy domains.

2.3 Human-AI Collaboration in Complex Decision Making

Human-AI collaboration research highlights designs that improve decision-making in complex contexts (Elendu et al. 2024; Gomes et al. 2021). While such systems often outperform humans alone, they may lag behind AI-only performance (Gomes et al. 2021). However, autonomous AI is often unsuitable in ethically charged or legally sensitive settings, reinforcing the importance of effective human-AI partnerships.

In high-stakes areas like healthcare and policy, combining human judgment with AI capabilities enhances decision quality (Gomez et al. 2025). Effective collaboration depends on role clarity, trust, and transparent interaction (Bansal et al. 2019). The hybrid intelligence model emphasizes synergies between human insight and machine efficiency (Dellermann et al. 2021).

Key enablers include system explainability and intuitive interfaces (Akinnagbe 2024; Zhang et al. 2024; Gomez et al. 2025). Rather than replacing human judgment, the goal is to augment it through collaborative systems. Most prior research focuses on low-stakes or retrospective analyses. Our approach contributes by enabling prospective, simulation-based testing of AI recommendations in high-stakes contexts. This controlled environment allows for systematic evaluation of collaboration frameworks before real-world deployment, providing an ethically sound foundation for studying AI-supported policy decisions (Miller et al. 2024).

3 METHODS

AI has proven vital in pandemic management, notably demonstrated potential during the COVID-19 pandemic, aiding policy formulation and healthcare resource optimization. Reinforcement learning, coupled with deep learning, can model aspects of human learning processes. Mnih et al. (2015) showed that RL algorithms achieve strong performance in complex video games, suggesting potential applications to real-world socio-technical challenges (Sert et al. 2020; Kwak et al. 2021). This study seeks to design an agent that surpasses traditional methods like DQN (Mnih et al. 2015), addressing their limitations through advanced techniques such as double Q-learning (Van Hasselt et al. 2016) and dueling architectures (Wang et al. 2016), thereby enhancing AI's ability to support human decision-making in pandemic policy formulation.

3.1 Problem Formulation

We formulate pandemic policy-making as a Markov Decision Process (MDP) where an agent must select intervention strategies based on the current pandemic state. The state space consists of 7 features representing pandemic conditions (detailed in Section 3.3), while the action space comprises combinations of lockdown and travel restrictions (detailed in Section 3.4). Our objective is to learn a policy π that maximizes the expected cumulative reward, balancing infection reduction, mortality decrease, and recovery increase over a 100-step pandemic trajectory with decisions made every 7 days.

3.2 Deep Reinforcement Learning Approach

Q-learning is a reinforcement learning approach where an agent learns to make optimal decisions at each state. Mnih et al. (2015) showed that Deep Q-Networks (DQN) have successfully addressed challenges like high dimensional input data using deep convolutional neural networks, and serve as a foundational algorithm underpinning RL approaches but are prone to overestimations. Van Hasselt et al. (2016) developed Double Q-Networks that resolve the issue of Q-value overestimation but lack in terms of stability, while Wang et al. (2016) showed that dueling Q-Networks segregate value and advantage estimation, enhancing stability and efficiency. We employ the Dueling Double DQN (D3QN) algorithm (Kwak et al. 2021), which extends standard Q-learning by learning two separate value functions: the *state value* $V(s)$, indicating the overall desirability of state s, and the *advantage function* $A(s,a)$, capturing the relative benefit of choosing action a in that state. The Q-value is then computed as shown in (1), where the mean advantage across all possible actions is subtracted to normalize $A(s,a)$ around zero:

$$Q(s,a) = V(s) + \left(A(s,a) - \frac{1}{|\mathscr{A}|} \sum_{a'} A(s,a') \right). \tag{1}$$

This architecture, shown in Figure 1, separates the evaluation of state importance from action selection—an essential feature for pandemic scenarios, where some states require immediate intervention while others allow more flexibility. By learning a state value function $V(s)$ and an advantage function $A(s,a)$, the agent can better assess when the choice of action matters most. As defined in (1), the Q-value combines these two streams, with the advantage normalized to prevent uniform inflation across actions. The dueling structure, paired with double Q-learning, reduces overestimation bias by decoupling action selection and evaluation (Wang et al. 2016; Mnih et al. 2015). Training is based on replayed experiences (s,a,r,s'), which improves sample efficiency and prevents the forgetting of rare but important transitions in dynamic policy environments (Kwak et al. 2021).

3.3 Features

In our study, we represented the pandemic state using a 7-feature vector based on the available dataset (Kwak et al. 2021). This vector included the number of confirmed cases, recovered cases, and deaths, along with the growth rates of confirmed cases, recovered cases, and deaths, as well as the population size. To ensure compatibility with the SEIRD model, integral for generating the simulated environment in our second study, we omitted six supplementary features from the original dataset—population density, average population age, GDP, latitude, longitude, and life expectancy. This choice aimed for consistency and comparability between our two studies, making the simulated SEIRD model align with real-world data. This simplification facilitates the generalization of results and makes the simulated environment more accessible for educational purposes too, focusing on key pandemic dynamics and streamlining the model's implementation.

3.4 Actions and Reward

In our study, the action space was composed of a combination of lockdown and travel restriction policies, with discrete values ranging from 0 to 2, resulting in a total of 9 possible actions following the approach used by Kwak et al. (2021). The different levels were defined as follows:

- Lockdown policy: No action (L0), restricted public social gathering (L1), and nationwide lockdown (L2)
- Travel policy: No action (T0), flight suspension (T1), and full closure of all borders (T2)

Given the multi-objective nature of pandemic control, we designed our reward function based on three main variables: the growth in confirmed cases, recovered cases, and deaths. Drawing insights from the literature, we assigned weights of 2:1:1 to these variables in the reward function, aligning with the approach used by Kwak et al. (2021). Consequently, our reward function can be formulated as follows:

$$r_t = r_t^i + 0.5 \times r_t^d + 0.5 \times r_t^r. \tag{2}$$

Here, i, d, and r represent the decrease in infection, decrease in death, and increase in recovery cases, respectively. The reward function is evaluated at each 7-day decision point, comparing the change in infection, death, and recovery rates between the current and previous time steps. This temporal granularity allows the agent to learn the delayed effects of interventions while maintaining computational efficiency.

3.5 Model Architecture and Training Protocol

We utilized the D3QN method introduced by Kwak et al. (2021) to construct our agent. Figure 1 illustrates the architecture, which implements the value and advantage streams described in Section 3.2. The environment varies by study: real-world data is used in Study 1, and a simulated SEIRD model in Study 2. The model includes two hidden layers of size 52 and uses a batch size of 4.

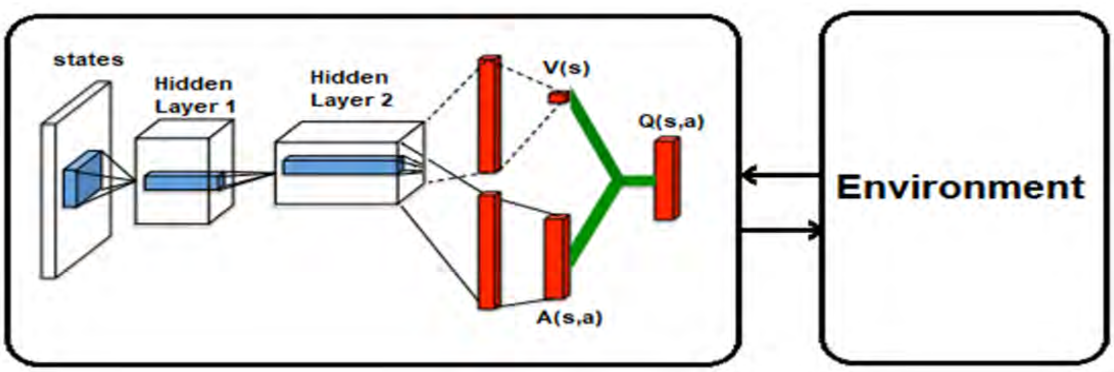

Figure 1: Architecture of AI agent (based on D3QN) integrated with an environment (real-world data in Study 1; simulated SEIRD model in Study 2). The neural network splits into value stream $V(s)$ and advantage stream $A(s,a)$, which combine to compute Q-values.

Study 1 uses a dataset of daily pandemic observations from 186 countries, split into training (70%), validation (10%), and testing (20%) sets with balanced pandemic conditions. Training involves 1000 episodes, each sampling 100-step trajectories (700 days) from historical data in Study 1, or simulating 40-step trajectories (280 days) in the SEIRD environment for Study 2. Both studies follow identical learning protocols: agents observe states every 7 days and select actions using ε-greedy strategy (ε decays from 1.0 to 0.1 over 500 episodes). Actions are random with probability ε or Q-value-maximizing with probability $1-\varepsilon$. We employ experience replay and soft target network updates every 100 steps ($\tau = 0.001$).

4 RESULTS

To evaluate the performance differences between our DRL agent and government policies, we conducted a counterfactual analysis using the test dataset (2,926 samples). We ran parallel simulations with identical initial conditions based on real-world data—one following actual government decisions, and the other using the actions recommended by our trained agent. Figure 2 displays three heatmaps illustrating the respective policy distributions and their differences. The results reveal a clear behavioral contrast: the agent consistently favored balanced, moderate interventions, whereas government decisions tended to swing between inaction and strict enforcement.

Figure 2: Comparison of policy action distributions between the AI agent and the governments. The left and middle heatmaps show the normalized frequencies of actions taken by the D3QN-based agent and the governments across lockdown levels (y-axis) and travel restrictions (x-axis). The right heatmap shows their difference (Agent – Government), where blue cells (positive values) indicate actions more frequently chosen by the agent, and red cells (negative values) those favored by the governments.

The left heatmap in Figure 2 reveals the agent's highly concentrated strategy, with 57% of actions at moderate lockdown with moderate travel restriction [1,1] and 41% at no lockdown with moderate travel restriction [0,1]. The agent notably avoids strict lockdown measures entirely, demonstrating a learned preference for balanced, moderate interventions. The middle heatmap shows government actions dispersed across the action space, with the highest concentration (57%) at the no-intervention position [0,0], followed by scattered frequencies across various intervention levels. This pattern reflects a reactive approach, maintaining status quo until circumstances force action. The right heatmap highlights the policy differences, with deep blue at [1,1] (+0.50) indicating the agent's strong preference for proactive moderate measures, while deep red at [0,0] (-0.56) reveals governments' tendency toward initial inaction. This contrast demonstrates that our DRL agent has learned to implement consistent preventive measures in this stylized environment, avoiding both extremes of no action and strict interventions, while governments exhibit a pattern of delayed response potentially requiring more drastic actions later.

Furthermore, Figure 3 illustrates the cumulative reward trajectories over 100 simulation weeks for both the DRL agent and government policies. The reward metric, calculated as weighted reductions in infections and deaths plus recovery increases, is expressed as a percentage of the population. The agent's cumulative reward (blue line) exceeds government policies (red line) with the gap widening over time. By week 100, the agent achieves approximately 12.5% cumulative reward compared to 9.5% for government policies. While these results within our simplified model align with prior work by Kwak et al. (2021). It is crucial to acknowledge the comparison's limitations. Government policies involved a broader portfolio of interventions beyond our model's two categories (discrete lockdowns and travel restrictions), operating under complex economic, social, and political constraints. Thus, our findings reflect performance within a stylized framework rather than the full complexity of real-world pandemic policy decisions.

Figure 3: Agent vs Government performance: Cumulative reward (y-axis) over 100 simulation steps (x-axis). Reward represents weighted average reduction in infection rate ($-2 \times \Delta$infection), mortality rate ($-1 \times \Delta$deaths), and increase in recovery rate ($+1 \times \Delta$recoveries).

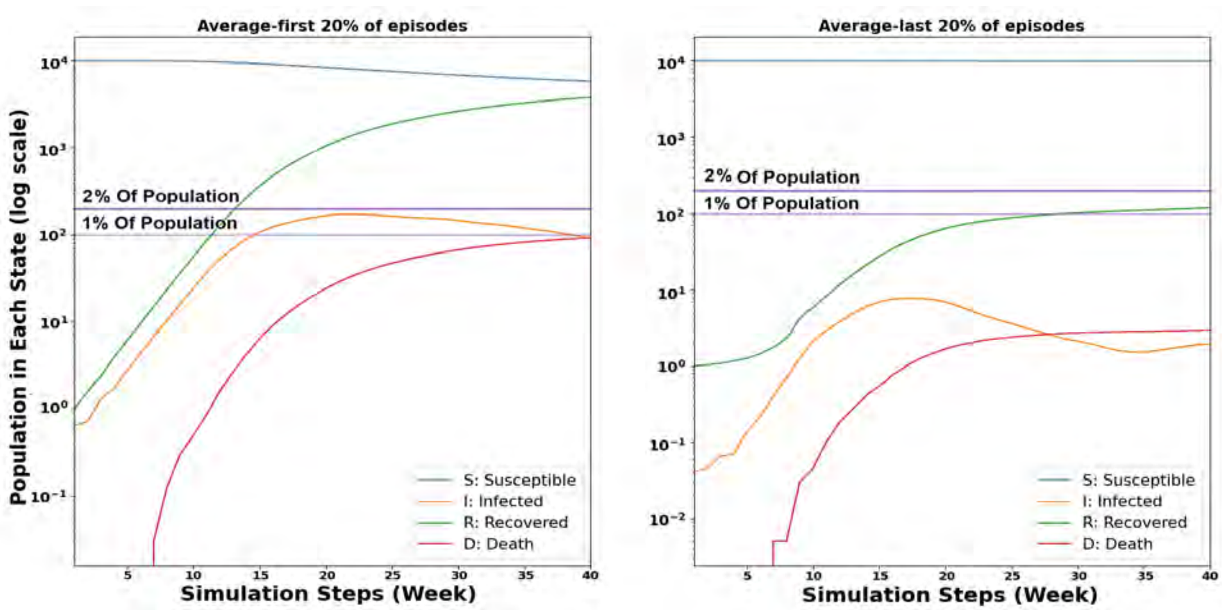

Figure 4: Mean population distribution across different states of the system (S: Susceptible, I: Infected, R: Recovered, or D: Dead) during the learning process (initial and final phases). The figure excludes the E: Exposed state for simplicity. Horizontal lines indicate baseline thresholds for healthcare system capacity: the thicker line represents 2% of the population, and the thinner line represents 1%. Infection rates exceeding these thresholds are associated with a significant increase in mortality due to system overload.

To establish a benchmark for future studies and support consistent comparison, we conducted a second investigation. We acknowledge that government decisions are informed by a wide array of information sources, data of which we do not possess in its entirety. Our trained agent may have overlooked significant factors influencing government decisions, to which we lack access. To enhance the applicability of our findings, we utilized the SEIRD model to simulate pandemic scenarios, addressing data limitations and constraints imposed by available data. In this vein, our second study was based on this simulated environment, allowing us to test our designed DRL agent and validate the potential of simulation-based testing in evaluating AI performance.

Figure 4 illustrates the outcomes of our analysis, specifically comparing the average population in each state (Susceptible, Infected, Recovered, Dead) during different stages of the learning process (initial and final 20%). These results demonstrate the promise of our agent to effectively learn and make policies in the simulated environment. Notably, the agents display proficient management of the pandemic by significantly increasing overall recovery rates and reducing infection and mortality rates over time.

5 DISCUSSION

The integration of AI into complex policy domains offers both promise and challenges. Our research examines AI's potential contribution to pandemic management through a framework combining real-world data analysis and simulation-based testing. We investigate whether DRL shows potential for application to support high-stakes policy decisions under uncertainty, providing insights into AI's capabilities, limitations, and pathways for responsible integration.

5.1 Potential of AI (specifically DRL) for Pandemic Policy-Making

Our research shows that DRL has potential for application to complex policy decision-making in pandemic scenarios within controlled environments. Our two-stage framework shows that the DRL agent (based on D3QN) takes a more balanced approach to pandemic interventions compared to government responses. While governmental policies during COVID-19 often oscillated between extremes—no restrictions or severe lockdowns—our agent consistently recommended moderate, calibrated interventions that achieved better outcomes in the stylized environment. These findings align with findings by Rahmandad et al. (2021) and Kwak et al. (2021), supporting the utility of our trained agent in simulated policy settings.

Our approach extends previous work by exploring DRL performance relative to historical government policies within our simplified two-action framework. Training on data from 186 diverse countries—spanning different demographics, healthcare systems, and governance structures—suggests the model's capacity to identify patterns despite substantial heterogeneity. Within our stylized environment, the DRL agent achieves higher cumulative rewards than government policies. However, these results require cautious interpretation. Government policies were developed without training data during a novel pandemic, and our static comparison cannot capture real-time adaptive decision-making. Our comparison is limited to two intervention types while real-world implementation involves factors beyond our model's scope: economic trade-offs, social acceptance, political constraints, and operational complexities (Ramezani et al. 2023). The agent's apparent success partly reflects this asymmetry—it learns from complete historical data while policymakers faced unprecedented uncertainty. Whether these findings generalize to different pathogens or social contexts remains uncertain (Ali 2024). Thus, while our results suggest potential value in AI-assisted policy analysis, they should not imply that AI can replace human judgment in pandemic management.

5.2 Simulation-Based Testing of AI Integration into Complex Systems

Our approach provides a framework for evaluating AI in policy domains before real-world deployment—essential for responsible high-stakes implementation. It addresses data limitations (Ali 2024) by generating synthetic scenarios beyond historical observations, enables controlled testing across diverse conditions, and allows rapid iteration without real-world risks. However, our SEIRD model simplifies population heterogeneity, regional healthcare variations, social compliance behaviors, and economic factors. These simplifications limit policy recommendation comprehensiveness (Li et al. 2024), and model assumptions may not hold for novel pandemics. Furthermore, training on simulated data risks circularity—agents may learn simulation artifacts rather than genuine dynamics, producing policies that fail to generalize (Bengio et al. 2011). While our two-stage validation partially mitigates this risk, distinguishing meaningful patterns from artifacts remains challenging. Despite these limitations, our approach addresses the need for benchmarks assessing real-world capabilities (Renkhoff et al. 2024), providing controlled evaluation environments supporting responsible AI development (Akinnagbe 2024).

5.3 Going from AI Simulation to Human-AI Interaction

While our research explores AI's potential in simulated pandemic policy evaluation, real-world implementation would require meaningful human-AI integration. Complex crisis decisions would likely benefit from collaboration rather than autonomous AI. Our preliminary findings suggest AI might support decision-making by identifying balanced strategies within our stylized environment, particularly under time pressure. However, these insights remain limited to our simplified model and require extensive validation before real-world application. The transition to effective human-AI interaction presents several challenges. AI recommendations must be interpretable and explainable to gain stakeholder trust (Dellermann et al. 2021). Interfaces must facilitate appropriate task allocation, leveraging complementary strengths of human judgment and AI analysis (Bansal et al. 2019). The collaborative system must adapt to different stakeholder needs while maintaining consistent recommendations. Simulated environments offer promising platforms for developing and testing human-AI interaction approaches. These environments create safe spaces where policymakers can interact with AI advisors and incorporate recommendations into decision processes (Maheu-Cadotte et al. 2018; Laamarti et al. 2014). They provide structured settings where interactions can be measured, support experiential learning about AI's strengths and limitations, and allow observation of decision consequences without real-world risks.

5.4 Ethical Considerations and Future Research

The integration of AI into policy decision-making raises important ethical considerations. Our methodology represents an ethical approach by advocating for simulation-based testing before high-stakes deployment. However, several concerns require attention. Bias in AI policy recommendations remains significant—while our model trains on data from 186 countries, these data reflect existing healthcare disparities that AI systems might inadvertently amplify (Barocas et al. 2019). Our research suggests maintaining human oversight is essential, with AI serving in an advisory rather than autonomous capacity (Ekundayo 2024).

Several promising research directions emerge. First, enhancing simulation realism by incorporating economic impacts, heterogeneous population characteristics, and adaptive social behaviors would better capture real-world complexity (Renkhoff et al. 2024). Second, developing frameworks for human-AI collaboration in policy contexts requires careful attention to system design, performance metrics, and presentation formats that may influence trust and decision quality (Akinnagbe 2024). Building on prior work that demonstrates the value of keeping experts in the loop for system refinement (Troiano et al. 2025), advancing AI for policy similarly demands continuous stakeholder involvement to shape, evaluate, and build trust in AI outputs. Third, more robust evaluation methodologies are needed, including standardized benchmarks that capture epidemiological outcomes alongside economic impacts and equity considerations (Ribeiro et al. 2020). Finally, investigating the transferability of AI policy approaches across different crisis types represents an important extension beyond pandemic management.

6 CONCLUSION

The integration of AI into critical decision-making contexts requires robust evaluation frameworks, particularly where real-world experimentation raises ethical concerns. Our two-stage simulation framework addresses this challenge by first training a DRL agent on data from 186 countries to model optimal pandemic interventions, then testing it in a SEIRD-based simulation environment beyond historical data. Results indicate the DRL agent outperforms governments' average outcomes in our test environment, implementing more balanced intervention strategies that reduce infections and fatalities while improving recovery rates. The first stage demonstrates the agent's potential using real-world data within model constraints, while the second enables evaluation across unobserved scenarios—a methodological contribution to simulation-based testing. This research presents a responsible AI evaluation methodology for domains where direct experimentation is problematic. By demonstrating how simulations bridge the AI development-deployment gap, we provide a pathway prioritizing safety and effectiveness. Future research should enhance simulation fidelity, develop comprehensive evaluation metrics, investigate human-AI collaboration frameworks, and address implementation constraints. Building effective, context-sensitive human-AI collaboration designs remains central to realizing the potential of AI-supported decision-making in critical domains like pandemic management and other complex policy environments.

A DATASET

The pre-processed dataset (Kwak et al. 2021) aggregates data from two primary sources:

- Johns Hopkins CSSE COVID-19 data (Johns Hopkins University 2020)
- Oxford OxCGRT policy tracker (University of Oxford 2020)

B SEIRD MODEL

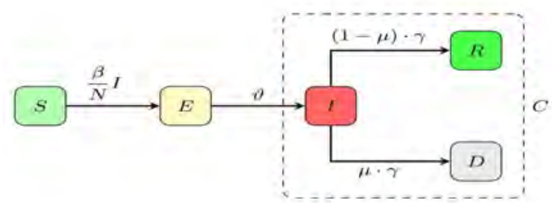

Figure 5: SEIRD epidemic model showing transitions between states with their respective transition rates.

Mathematical expressions illustrating the interplay between distinct disease stages within the SEIRD model, constructed using ordinary differential equations (ODE):

$$\begin{aligned}
\frac{dS}{dt} &= -\frac{\beta}{N}SI, && S(t_0) = S_0 := N - E_0 - I_0,\\
\frac{dE}{dt} &= \frac{\beta}{N}SI - \vartheta E, && E(t_0) = E_0,\\
\frac{dI}{dt} &= \vartheta E - \gamma I, && I(t_0) = I_0,\\
\frac{dR}{dt} &= (1-\mu)\cdot\gamma I, && R(t_0) = 0,\\
\frac{dD}{dt} &= \mu\cdot\gamma I, && D(t_0) = 0.
\end{aligned} \tag{3}$$

REFERENCES

Akinnagbe, O. B. 2024. “Human-AI Collaboration: Enhancing Productivity and Decision-Making”. *International Journal of Education, Management, and Technology* 2(3):387–417.

Ali, H. 2024. “AI for Pandemic Preparedness and Infectious Disease Surveillance: Predicting Outbreaks, Modeling Transmission, and Optimizing Public Health Interventions”. *International Journal of Research Publication and Reviews* 5(8):4605–19.

Amershi, S., D. Weld, M. Vorvoreanu, A. Fourney, B. Nushi, P. Collisson, *et al*. 2019. “Guidelines for Human-AI Interaction”. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–13. New York, NY, USA: ACM.

Asan, O., A. E. Bayrak, and A. Choudhury. 2020. “Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians”. *Journal of Medical Internet Research* 22(6):e15154.

Bansal, G., B. Nushi, E. Kamar, W. S. Lasecki, D. S. Weld, and E. Horvitz. 2019. “Beyond Accuracy: The Role of Mental Models in Human-AI Team Performance”. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, edited by E. Law and J. W. Vaughan, Volume 7, 2–11.

Barocas, S., M. Hardt, and A. Narayanan. 2019. “Fairness and Abstraction in Sociotechnical Systems”. In *Proceedings of the FAT* ’19: Conference on Fairness, Accountability, and Transparency*, 59–68. New York, NY, USA: ACM.

Bengio, Y., F. Bastien, A. Bergeron, N. Boulanger-Lewandowski, T. Breuel, Y. Chherawala, *et al*. 2011. "Deep Learners Benefit More from Out-of-Distribution Examples". In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, 164–172. New York, NY, USA: JMLR Workshop and Conference Proceedings.

Chen, J., and J. Zacharias. 2024. "Design Principles for Collaborative Generative AI Systems in Software Development". In *19th International Conference on Design Science Research in Information Systems and Technology*, 341–354. Cham, Switzerland: Springer.

Chinazzi, M., J. T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, *et al*. 2020. "The Effect of Travel Restrictions on the Spread of the 2019 Novel Coronavirus (COVID-19) Outbreak". *Science* 368(6489):395–400.

Dellermann, D., A. Calma, N. Lipusch, T. Weber, S. Weigel, and P. Ebel. 2021. "The Future of Human-AI Collaboration: A Taxonomy of Design Knowledge for Hybrid Intelligence Systems". *arXiv preprint arXiv:2105.03354*.

Ekundayo, F. 2024. "Leveraging AI-Driven Decision Intelligence for Complex Systems Engineering". *International Journal of Research Publication and Reviews* 5(11):1–10.

Elendu, C., D. C. Amaechi, A. U. Okatta, E. C. Amaechi, T. C. Elendu, C. P. Ezeh *et al*. 2024. "The Impact of Simulation-Based Training in Medical Education: A Review". *Medicine* 103(27):e38813.

Gabsi, A. E. H. 2024. "Integrating Artificial Intelligence in Industry 4.0: Insights, Challenges, and Future Prospects–A Literature Review". *Annals of Operations Research*:1–28.

Gawande, M. S., N. Zade, P. Kumar, S. Gundewar, I. N. Weerarathna, and P. Verma. 2025. "The Role of Artificial Intelligence in Pandemic Responses: From Epidemiological Modeling to Vaccine Development". *Molecular Biomedicine* 6(1):1.

Gomes, M. L. A., C. M. da Silva Carvalho, and A. C. B. Rocha. 2021. "Serious Games as Social Innovation Tools". *Product: Management and Development* 19(2):0–0.

Gomez, C., S. M. Cho, S. Ke, C.-M. Huang, and M. Unberath. 2025. "Human-AI Collaboration Is Not Very Collaborative Yet: A Taxonomy of Interaction Patterns in AI-Assisted Decision Making from a Systematic Review". *Frontiers in Computer Science* 6:1–15.

Goodfellow, I. J., J. Shlens, and C. Szegedy. 2014. "Explaining and Harnessing Adversarial Examples". *arXiv preprint arXiv:1412.6572*.

IEEE 2017. *IEEE Standard for System, Software, and Hardware Verification and Validation*. IEEE. IEEE Std 1012-2016 (Revision of IEEE Std 1012-2012).

Johns Hopkins University 2020. "COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE)". Available at: https://coronavirus.jhu.edu, accessed August 2025.

Khadilkar, H., T. Ganu, and D. P. Seetharam. 2020. "Optimising Lockdown Policies for Epidemic Control Using Reinforcement Learning: An AI-Driven Control Approach Compatible with Existing Disease and Network Models". *Transactions of the Indian National Academy of Engineering* 5(2):129–132.

Kubilay, B., and B. Celiktas. 2025. "Relationships among Organizational-Level Maturities in Artificial Intelligence, Cybersecurity, and Digital Transformation: A Survey-Based Analysis". *IEEE Access* 13:88399–88411.

Kwak, G. H., L. Ling, and P. Hui. 2021. "Deep Reinforcement Learning Approaches for Global Public Health Strategies for COVID-19 Pandemic". *PLOS ONE* 16(5):e0251550.

Kwak, M., B.-J. Kim, J.-B. Chung *et al*. 2024. "Serious Game Development for Public Health: Participatory Design Approach to COVID-19 Quarantine Policy Education". *JMIR Serious Games* 12(1):e54968.

Laamarti, F., M. Eid, and A. El Saddik. 2014. "An Overview of Serious Games". *International Journal of Computer Games Technology* 2014(1):358152.

Li, X., L. Yin, K. Liu, K. Zhu, and Y. Cui. 2024. "Deep-Reinforcement-Learning-Based Optimization for Intra-Urban Epidemic Control Considering Spatiotemporal Orderliness". *International Journal of Geographical Information Science*:1–26.

Liu, W. 2025, June. "Bracing the Artificial Intelligence Technology in Viral Infectious Disease Control". *Infectious Medicine* 4(2):100186.

Lukyanenko, R., W. Maass, and V. C. Storey. 2022. "Trust in Artificial Intelligence: From a Foundational Trust Framework to Emerging Research Opportunities". *Electronic Markets* 32(4):1993–2020.

Maadi, M., H. Akbarzadeh Khorshidi, and U. Aickelin. 2021. "A Review on Human-AI Interaction in Machine Learning and Insights for Medical Applications". *International Journal of Environmental Research and Public Health* 18(4):2121.

Maheu-Cadotte, M.-A., S. Cossette, V. Dubé, G. Fontaine, T. Mailhot, P. Lavoie, *et al*. 2018. "Effectiveness of Serious Games and Impact of Design Elements on Engagement and Educational Outcomes in Healthcare Professionals and Students: A Systematic Review and Meta-Analysis Protocol". *BMJ open* 8(3):e019871.

Menzies, T., and C. Pecheur. 2005. "Verification and Validation and Artificial Intelligence". *Advances in Computers* 65:153–201.

Miller, S., P. Johnson, and K. Garcia. 2024. "Digital Gamification Tools to Enhance Vaccine Uptake: Scoping Review". *JMIR Serious Games* 12(1):e47257.

Mnih, V., K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, *et al*. 2015. "Human-Level Control through Deep Reinforcement Learning". *Nature* 518(7540):529–533.

Ondula, E. A. 2024. "Navigating Uncertainty in Epidemic Contexts with Reinforcement Learning". In *Proceedings of the 38th AAAI Conference on Artificial Intelligence (AAAI-24)*, Volume 38, 23407–23408.

Oviedo, J., M. Rodriguez, A. Trenta, D. Cannas, D. Natale, and M. Piattini. 2024. "ISO/IEC Quality Standards for AI Engineering". *Computer Science Review* 54:100681.

Rahmandad, H., T. Y. Lim, and J. Sterman. 2021. "Behavioral Dynamics of COVID-19: Estimating Underreporting, Multiple Waves, and Adherence Fatigue across 92 Nations". *System Dynamics Review* 37(1):5–31.

Ramezani, M., A. Takian, A. Bakhtiari, H. R. Rabiee, S. Ghazanfari, and H. Mostafavi. 2023. "The Application of Artificial Intelligence in Health Policy: A Scoping Review". *BMC Health Services Research* 23(1):1416.

Renkhoff, J., K. Feng, M. Meier-Doernberg, A. Velasquez, and H. H. Song. 2024. "A Survey on Verification and Validation, Testing and Evaluations of Neurosymbolic Artificial Intelligence". *IEEE Transactions on Artificial Intelligence* 5(8):3765–3779.

Ribeiro, M. T., T. Wu, C. Guestrin, and S. Singh. 2020. "Beyond Accuracy: Behavioral Testing of NLP Models with CheckList". *arXiv preprint arXiv:2005.04118*.

Sert, E., Y. Bar-Yam, and A. J. Morales. 2020. "Segregation Dynamics with Reinforcement Learning and Agent Based Modeling". *Scientific Reports* 10(1):11771.

Tomašev, N., J. Cornebise, F. Hutter, S. Mohamed, A. Picciariello, B. Connelly, *et al*. 2020. "AI for Social Good: Unlocking the Opportunity for Positive Impact". *Nature Communications* 11(1):2468.

Troiano, G. M., M. Cassidy, D. E. Morales, G. Pons, A. Abdollahi, G. Robles, *et al*. 2025. "CT4ALL: Towards Putting Teachers in the Loop to Advance Automated Computational Thinking Metric Assessments in Game-Based Learning". In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 1–23. New York, NY, USA: ACM.

University of Oxford 2020. "Oxford COVID-19 Government Response Tracker (OxCGRT)". Available at: https://github.com/OxCGRT/covid-policy-tracker, accessed August 2025.

Valle-Cruz, D., J. I. Criado, R. Sandoval-Almazán, and E. A. Ruvalcaba-Gomez. 2020. "Assessing the Public Policy-Cycle Framework in the Age of Artificial Intelligence: From Agenda-Setting to Policy Evaluation". *Government Information Quarterly* 37(4):101509.

Van Hasselt, H., A. Guez, and D. Silver. 2016. "Deep Reinforcement Learning with Double Q-Learning". In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI-16)*, Volume 30, 2094–2100. Washington, DC, USA: AAAI Press.

Wang, Z., T. Schaul, M. Hessel, H. Hasselt, M. Lanctot, and N. Freitas. 2016. "Dueling Network Architectures for Deep Reinforcement Learning". In *33rd International Conference on Machine Learning*, Volume 48, 1995–2003. New York, NY, USA: JMLR Workshop and Conference Proceedings.

Wu, J. T., K. Leung, and G. M. Leung. 2020. "Nowcasting and Forecasting the Potential Domestic and International Spread of the 2019-nCoV Outbreak Originating in Wuhan, China: A Modelling Study". *The Lancet* 395(10225):689–697.

Xiang, W., P. Musau, A. A. Wild, D. M. Lopez, N. Hamilton, X. Yang, *et al*. 2018. "Verification for Machine Learning, Autonomy, and Neural Networks Survey". *arXiv preprint arXiv:1810.01989*.

Zhang, S., J. Yu, X. Xu, C. Yin, Y. Lu, B. Yao, *et al*. 2024. "Rethinking Human-AI Collaboration in Complex Medical Decision Making: A Case Study in Sepsis Diagnosis". In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 1–18. New York, NY, USA: ACM.

AUTHOR BIOGRAPHIES

AMIR ABDOLLAHI is a PhD student of Industrial Engineering at Northeastern University, skilled in ML, probability and statistics, GenAI, and reinforcement learning. His research focuses on "leveraging gamified experiments to explore the inclusion of context-specific AI-based recommendations in societal systems." His email address is abdollahi.a@northeastern.edu.

REYHANEH MOHAMMADI earned her Ph.D. in Industrial Engineering from Northeastern University and brings more 9+ years of expertise in machine learning, optimization, and time-series analysis. She integrates simulation with data science to transform complex datasets into actionable insights and decision-support tools. Her email address is mohammadi.re@northeastern.edu.

JACQUELINE GRIFFIN is an Associate Professor of Industrial Engineering at Northeastern University. Her research focuses on developing novel simulation models and operations research methods that support complex, multi-objective decision making in real-world health and humanitarian systems. Her email address is ja.griffin@northeastern.edu.

CASPER HARTEVELD is a Professor of Game Design at Northeastern University with appointments spanning Computer Science, Engineering, Law, and Public Policy. He designs games to analyze decisions in real-world applications, gathering rich behavioral data under controlled conditions. He bridges research and teaching, bringing game-based insights into classrooms and informal venues to prepare future generations for complex societal challenges. His e-mail address is c.harteveld@northeastern.edu.