A Study of Genetic Algorithm in Evolving Agents for Autonomous Decision-Making in Dynamic Environments

Amir Bašović
Department of Information Technology,
International Burch University
Sarajevo, Bosnia and Herzegovina

Fatima Mašić
Department of Information Technology,
International Burch University
Sarajevo, Bosnia and Herzegovina

ABSTRACT
This thesis investigates the application of Genetic Algorithms (GAs) for evolving autonomous decision-making strategies in dynamic grid-world environments. The study focuses on scenarios in which obstacles appear, disappear, or move during agent navigation, creating conditions where traditional pathfinding and reinforcement learning (RL) approaches often struggle or require extensive retraining. Two GA-based agent models: rule-based agents and finite state machine (FSM) agents were evolved using population-based search to develop adaptive and generalizable behaviors. Their performance was evaluated in the MiniGrid DynamicObstacles environment and compared against a Q-learning agent across multiple metrics, including path efficiency, adaptability under environmental volatility, convergence time, and generalization to unseen map configurations. Experimental results show that GA-evolved agents achieve strong adaptability and high generalization performance, outperforming the RL baseline in dynamic and previously unseen environments. The findings demonstrate the viability of evolutionary methods for generating robust autonomous behaviors in uncertain, real-time settings, with implications for robotics, simulation platforms, and adaptive navigation systems.

General Terms
Evolutionary Computation, Machine Learning, Autonomous Systems, Pathfinding

Keywords
Genetic Algorithms, Autonomous Agents, Dynamic Environments, Pathfinding, Evolutionary Computation, Finite State Machines, Reinforcement Learning, Q-Learning, MiniGrid, Adaptive Decision-Making, Obstacle Avoidance, Grid-World Navigation

1. INTRODUCTION
Autonomous agents operating in dynamic and uncertain environments must continuously adapt their decision-making strategies to achieve mission objectives. Tasks such as navigation in the presence of moving or reconfigurable obstacles remain challenging in domains including robotics, logistics, and search and rescue. While classical pathfinding algorithms such as A* perform well in static settings, they often fail or require repeated replanning when environmental conditions change. RL provides a more adaptive alternative but typically relies on extensive exploration and dense reward feedback, which may be inefficient in highly dynamic scenarios.

Recent developments in evolutionary computation, particularly genetic programming and GAs, have shown promise in developing robust, interpretable, and flexible decision-making rules for autonomous systems. By evolving entire strategies over generations, GAs can optimize agent behavior with minimal feedback through population-based search. This makes them well-suited for environments with partial observability or unexpected dynamics.

Prior work has shown that evolutionary and hybrid evolutionary–RL approaches can achieve robust navigation performance in dynamic environments, particularly when combined with structured policy representations. However, limited attention has been given to directly comparing different evolutionary policy representations, such as rule-based and FSM-based agents, against RL baselines in highly volatile grid-world settings.

This study addresses this gap by investigating the effectiveness of GA-evolved rule-based and FSM-based agents in dynamic grid-world environments and benchmarking their performance against a Q-learning agent. The focus is placed on adaptability, convergence behavior, and generalization to unseen environments.

Specifically, this research examines how effectively genetic algorithms can evolve autonomous decision-making strategies for agents operating in dynamic grid-world environments, where obstacle configurations change over time. In addition, the study analyzes how varying levels of environmental dynamics influence the convergence behavior, robustness, and stability of evolved agent policies. Finally, the generalization capability of GA-evolved strategies is evaluated by testing their performance in previously unseen environments and obstacle configurations.

2. RELATED WORK
The study of autonomous decision-making in dynamic environments has expanded significantly, incorporating approaches ranging from RL to evolutionary computation.

Table 1 summarizes the key literature relevant to this domain, highlighting the methods employed and the primary contributions of each work. This section reviews key contributions that form the foundation of this thesis, with emphasis on GAs, evolutionary reinforcement learning (ERL), FSMs, and adaptive multi-agent navigation.

Evolutionary and reinforcement learning methods have been widely explored for autonomous decision-making in dynamic environments. The MAPPER framework integrates evolutionary reinforcement learning with decentralized coordination to achieve robust multi-agent navigation under changing environmental conditions [1]. A comprehensive review highlights that evolutionary reinforcement learning methods enable real-time adaptation in non-stationary environments while reducing reliance on handcrafted rules [2].

Structured policy representations have also been explored to improve adaptability. FSM-based control combined with reinforcement learning has been shown to significantly

outperform deterministic controllers in dynamic discrete-event systems [3]. In addition, novelty-based evolutionary search methods have demonstrated that population diversity can lead to scalable and generalizable behaviors in uncertain environments [4].

The application of genetic algorithms to robot path planning in dynamic environments has received considerable attention in recent years. An improved GA for global dynamic path planning combining a static global optimal solution with dynamic obstacle avoidance has demonstrated that GA-based approaches can achieve near-optimal paths while adapting to moving obstacles in grid-world settings [5]. This work is closely related to the experimental setup used in this thesis and provides a direct basis for comparison in terms of path efficiency and obstacle handling.

Quality-Diversity (QD) evolutionary algorithms represent an important development in population-based search that is directly relevant to this study. QD methods have been shown to be a competitive alternative to information-theory-augmented RL for skill discovery, producing more diverse and transferable behaviors while being less sensitive to hyperparameter tuning [6], [7]. The emphasis on behavioral diversity in QD algorithms parallels the diversity-preserving role of the novelty search strategy discussed in this thesis and supports the argument that population-based methods generalize better than single-policy RL approaches.

Hybrid evolutionary approaches that combine GAs with deep reinforcement learning have also shown strong results in dynamic navigation tasks. A hierarchical framework combining evolutionary training environments with GA3C reinforcement learning has been proposed to handle both structured static environments and dynamic obstacle scenarios, achieving improved collision avoidance over standard RL baselines [8]. This work highlights the complementary strengths of evolutionary and gradient-based methods and motivates the comparative evaluation conducted in this thesis.

Path planning in multi-agent dynamic environments has been addressed using risk-aware evolutionary optimization. A genetic algorithm combined with a probabilistic roadmap for multi-agent path planning has demonstrated conflict-free navigation in environments with overlapping agent trajectories [9]. While this work focuses on multi-agent settings rather than single-agent grid-worlds, it confirms the scalability of GA-based planning approaches and supports their applicability to more complex real-world scenarios.

Despite these advances, few studies directly evaluate pure GA-based agents with different policy representations in highly dynamic grid-world scenarios or compare their adaptability and generalization against RL baselines. This work contributes to addressing this gap through a systematic experimental evaluation in dynamic grid-world benchmarks, including MiniGrid [10].

Table 1. Literature Review Summary

Reference	Method	Contribution
Liu et al. [1]	ERL, Multi-Agent GA	Robust navigation in mixed dynamic environments using evolutionary RL.
Lin et al. [2]	Evolutionary RL	Comprehensive overview of ERL methods for dynamic environments.
Zielinski et al. [3]	FSM + RL	Demonstrated improved adaptability using structured FSM-based control.
Lehman & Stanley [4]	GA, Novelty Search	Showed generalizable behaviors through diversity-driven evolution.
Li et al. [5]	GA, Dynamic Path Planning	Improved GA for mobile robot path planning with dynamic obstacle avoidance.
Flageat et al. [6]	QD, Neuroevolution	QD methods as competitive alternative to RL for diverse policy discovery.
Cully & Demiris [7]	QD Framework	Unified QD framework for behavioral diversity and policy transfer.
Faust et al. [8]	GA + Deep RL	Hierarchical evolutionary framework for dynamic obstacle avoidance.
Rekabi-Bana et al. [9]	Genetic Opt., Risk-aware PRM	Evolutionary path planning for heterogeneous multi-agent systems.
Chevalier-Boisvert et al. [10]	MiniGrid Benchmark	Standardized dynamic grid-world environment for evaluation.

3. METHODOLOGY

This study employs a quantitative experimental design to assess the effectiveness of GAs in evolving decision-making strategies for autonomous agents operating in dynamic grid-world environments. The methodology involves designing a controlled simulation environment, implementing GA for evolving agent behavior, and conducting comparative evaluations against FSM agents and Q-learning agents.

For the case study, this research uses the MiniGrid DynamicObstacles environment, which provides a reproducible dynamic grid-world benchmark with moving obstacles. This setup allows for controlled experiments on adaptability, path efficiency, convergence, and generalization across different levels of environmental volatility. The results of these experiments are reported in Section V.

The overall workflow of the proposed framework is illustrated in Figure 1.

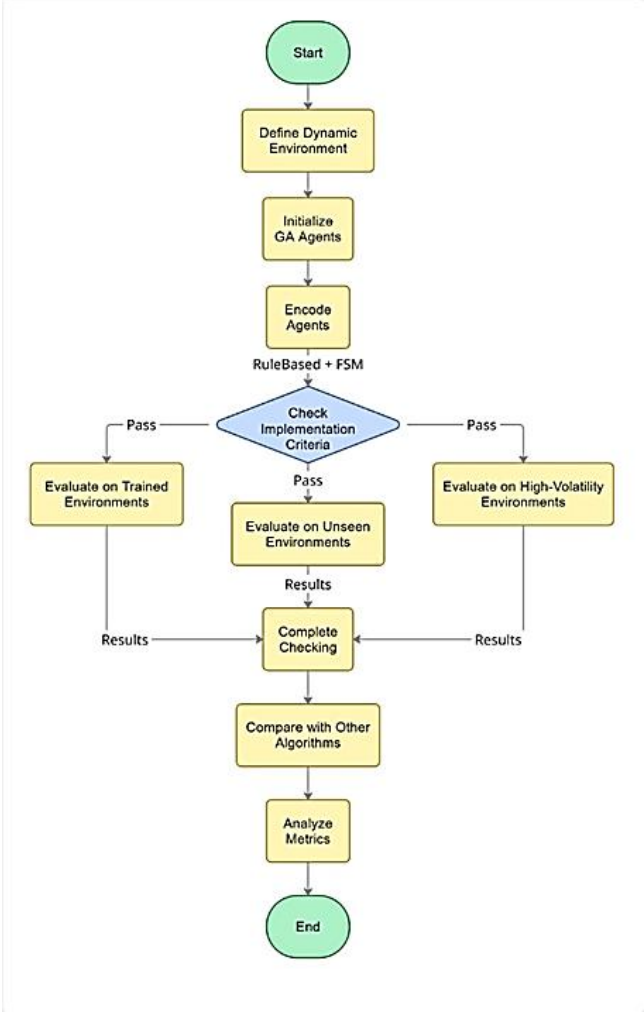

Fig. 1: Workflow Diagram

3.1 Experimental Setup

The environment consists of 10×10 grid worlds in which an agent must navigate from a fixed start position to a goal while avoiding dynamic obstacles. Obstacles may appear, disappear, or move at each simulation step, creating non-stationary conditions. Three categories of environments are considered: trained environments used during evolution, unseen environments for evaluating generalization, and high-volatility environments with increased obstacle movement.

Each evaluation category is tested across 30 independent runs using different random seeds to ensure statistical reliability. Results are reported as mean values with standard deviation where applicable, allowing for meaningful comparison across agent types and environment categories.

3.2 Agent Representation

Two GA-based agent models are implemented:

- Rule-based agents, which use encoded if–then logic as chromosomes.
- FSM agents, which use state-transition models encoded as chromosomes and evolved using genetic programming techniques.

Both agent types are evolved using standard GA operators, including selection, crossover, mutation, and fitness evaluation.

3.3 GA Configuration

The GA configuration is summarized in Table II.

Table 2. GA Configuration

Component	Configuration
Population size	50-100 agents
Selection method	Tournament selection
Crossover rate	0.8
Mutation rate	0.1
Elitism	Top 5% retained

3.4 Baseline Models for Comparison

GA-evolved agents are compared with the following baseline methods:

- Q-learning agent: a model-free reinforcement learning agent.
- FSM learning agent: a finite-state machine–based agent.

3.5 Evaluation Metrics

Agent performance is evaluated using the following metrics: path efficiency (steps to reach the goal), adaptability under environmental changes, convergence time, and generalization of success rate in unseen environments. Each metric is computed as an average over all evaluation runs within each environment category, and results across the three agent types are compared directly using the values reported in Section V.

4. EXPERIMENTAL SETUP AND DATASET

This section describes the experimental configuration used to generate the dataset in the MiniGrid DynamicObstacles environment and outlines the structure of the recorded data.

4.1 Environment Configuration

Table 3. MiniGrid Dataset Configuration

Attribute	Description
grid_size	Dimensions of the environment (e.g., 10 × 10 cells).
obstacle_density	Proportion of grid cells occupied by obstacles.
obstacle_dynamics	Frequency and pattern of obstacle movement during each episode.
start_position	Initial coordinates of the agent.
goal_position	Target cell to be reached.
steps_taken	Total steps required for the agent to reach the goal.

success_rate	Binary outcome indicating whether the goal was reached (Yes/No).
cumulative_reward	Aggregate reward assigned based on path efficiency and obstacle avoidance.
generation/episode	GA/FSM or Q-Learning

All experiments in this study are conducted using the MiniGrid DynamicObstacles environment, which provides a reproducible dynamic grid-world benchmark with moving obstacles. The environment is configured as follows:

- Grid size: a 10×10 grid world.
- Obstacle density: approximately 20% of the grid cells are occupied by obstacles.
- Obstacle dynamics: obstacles can move randomly at each simulation step, with validity checks to avoid blocking the agent's start or overwriting the goal cell.
- Start and goal positions: the agent starts from a fixed initial position and must reach a designated goal cell.

All runs use consistent random seeds and map-generation procedures to ensure reproducibility.

To assess generalization, a set of 10 previously unseen map configurations is generated with distinct obstacle placements and movement patterns. Agents trained on the primary environment are evaluated on these maps without any further evolution or retraining, providing a direct measure of policy transferability. High-volatility environments are created by increasing the obstacle movement frequency by a factor of two relative to the default configuration, introducing additional non-stationarity during evaluation. The key attributes of the dataset are summarized in Table III.

The dataset offers a consistent representation of each episode, capturing both environment-level and agent-level variables such as steps_taken, success_rate, cumulative_reward, and generation/episode. These features are later used for quantitative evaluation and analysis of learning dynamics in Section 5.

Although the dataset structure supports comparisons with additional baselines such as FSM-based agents and Q-learning, the MiniGrid-based analysis in this work focuses on GA-evolved agents. Comparative baselines are considered in separate simulation setups discussed in the results and discussion.

In this section, we focus exclusively on GA-evolved agents in the MiniGrid DynamicObstacles environment. Comparative experiments with FSM and Q-learning agents are conducted in the custom grid-world environment and are reported separately in the results and discussion.

5. RESULTS AND DISCUSSION

This section presents the experimental results obtained from two complementary setups:

- a custom dynamic grid-world environment used to compare rule-based GA agents, FSM-based GA agents, and a Q-learning agent,
- the MiniGrid DynamicObstacles benchmark used to analyze the behavior of GA-evolved agents in a standardized environment.

The subsection 5.1 focuses on the comparative results in the custom environment, while Section 5.2 reports the MiniGrid-based analysis.

5.1 Comparison of GA, FSM, and Q-Learning in the Custom Dynamic Grid-World Environment

To evaluate the effectiveness of the evolved decision-making strategies, a dynamic grid-world simulation environment was developed in Python. Two variations of this environment were created:

- GA/FSM environment: designed to support evolutionary simulations using rule-based and FSM agents. This version features a fixed start position, a goal at the opposite corner, and dynamically moving obstacles that simulate environmental volatility.
- Q-learning environment: includes random initial agent placement, dynamic obstacles, a fixed goal location, and step-limited episodes with negative shaping rewards to guide policy learning.

Both environments support dynamic obstacles, state visualization, and path rendering. The grid is 10×10 with roughly 20% obstacle coverage, where obstacles move randomly at each step while avoiding the agent and goal. Agents are evaluated across trained, unseen, and high-volatility environments, using consistent seeds and map configurations to ensure fair comparison.

5.1.1 Fitness evolution of rule-based GA agents

The fitness evolution of the rule-based GA agents over 20 generations is shown in Figure 2.

Overall, the population fitness improves over time, with both maximum and average fitness increasing across generations. This behavior indicates successful convergence of the GA toward higher-quality decision-making strategies in the dynamic environment, while maintaining diversity in the population as reflected by the gap between minimum and maximum fitness.

The maximum fitness reaches a plateau early, stabilizing around 135–140 from approximately generation 2 onward, while the average fitness shows a more gradual increase, rising steeply through the first 10 generations before beginning to level off. The sustained gap between minimum and maximum fitness across all 20 generations suggests that the rule-based representation preserves sufficient diversity to avoid premature convergence. This is a direct consequence of the tournament selection strategy combined with the 0.1 mutation rate, which continuously introduces variation into lower-performing individuals. The continued increase in average fitness beyond generation 15 suggests that crossover operations are effectively combining high-performing rule sets from different individuals rather than disrupting them. This behavior is consistent with observations by Li et al [5], who reported similar convergence patterns in GA-based path planning under dynamic obstacle conditions.

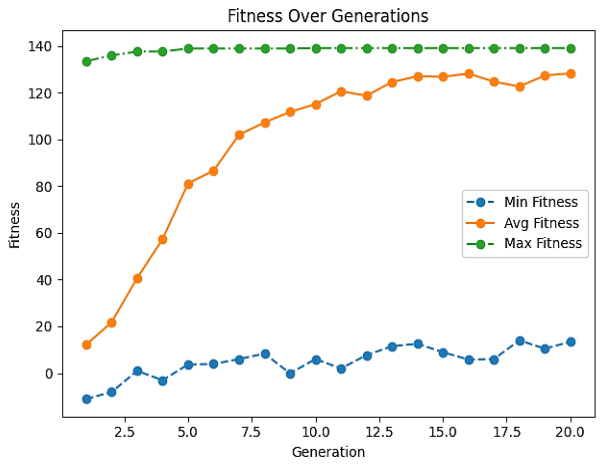

Fig. 2: Rule-based GA Fitness Plot

5.1.2 Fitness Evolution of FSM-Based GA Agents

Figure 3 illustrates the fitness evolution of FSM-based GA agents over 9 generations using minimum, average, and maximum fitness. The results show rapid early improvement, with maximum fitness stabilizing around generation 2, indicating that near-optimal FSM policies are found early. The average fitness increases steadily from approximately -75 in generation 1 to around 70 by generation 9, reflecting continuous population-level improvement.

However, the minimum fitness remains consistently low across all generations, dropping to values below -200 throughout the run, suggesting the persistence of poorly performing agents. This indicates reduced population diversity and a higher susceptibility to premature convergence compared to the rule-based GA.

The rapid stabilization of maximum fitness in FSM agents can be attributed to the more constrained structure of state-transition chromosomes compared to rule-based representations. Once a high-performing FSM structure is discovered, crossover operations tend to disrupt the transition logic, making it difficult for other individuals to reach the same fitness level. The persistent low minimum fitness further supports this interpretation, as poorly structured FSMs are difficult to repair through standard genetic operators. This structural rigidity results in a faster but narrower search, which explains the lower generalization score observed for FSM agents in Table IV. A similar dynamic was noted by Flageat et al [6], who observed that constrained representations in evolutionary algorithms limit behavioral diversity and reduce transferability to new task configurations.

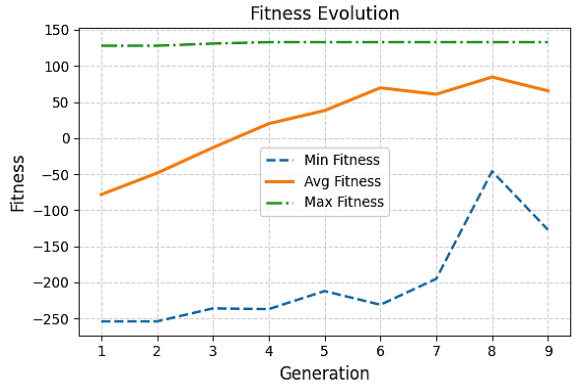

Fig. 3: FSM Fitness Plot

5.1.3 Training Dynamics of the Q-Learning Agent

The training performance of the Q-learning agent is depicted in Figure 4, which plots the total reward per episode over 1000 episodes. The x-axis represents the episode number, and the y-axis shows the total reward obtained in each episode.

The Q-learning agent initially achieves low and highly variable rewards, with episodic rewards dropping as low as -75 in the early stages, before converging after approximately 200 episodes, as indicated by the stabilized moving average. Unlike evolutionary methods, convergence occurs through repeated interaction rather than generation-based fitness.

The high variance in episodic rewards during the first 200 episodes reflects the exploratory behavior driven by the epsilon-greedy policy, where the agent frequently selects suboptimal actions in order to build the Q-table. Once sufficient state-action pairs have been visited, the policy stabilizes and rewards improve, with the moving average consistently approaching 100 beyond episode 400. However, when the same agent is evaluated in unseen or high-volatility environments, the learned Q-values may no longer correspond to optimal actions due to changes in obstacle positions, which directly explains the lower generalization score compared to GA-based agents. This sensitivity to environmental shift is a known limitation of model-free RL methods and has been previously documented by Faust et al [8], who showed that standard RL agents require additional evolutionary training mechanisms to maintain performance under dynamic obstacle conditions.

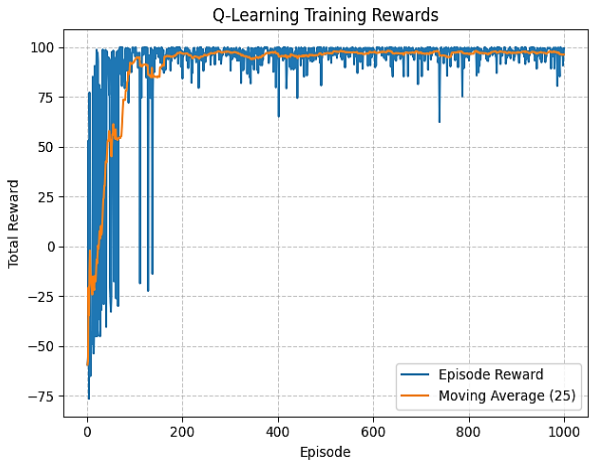

Fig. 4: Q-Learning Training Rewards Plot

5.1.4 Quantitative Comparison Across Evaluation Metrics

A quantitative comparison of the three agent models is summarized in Table IV. The agents are evaluated using four key performance metrics: path efficiency, adaptability score, convergence time, and generalization score.

Table IV indicates that the rule-based GA provides the best balance between path efficiency, adaptability, and generalization, with moderate convergence time. The FSM-based GA converges rapidly but shows reduced adaptability, while the Q-learning agent requires substantially more training and generalizes less reliably in dynamic environments.

Across all three environment categories, rule-based GA agents consistently achieve the highest generalization score at 85%, compared to 80% for FSM agents and 70–75% for Q-learning. In trained environments, the performance gap between agent

types is smaller, which suggests that all three approaches are capable of learning effective strategies when the evaluation conditions match training. The gap widens significantly in unseen and high-volatility environments, where the Q-learning agent's reliance on environment-specific Q-values becomes a disadvantage. These results are consistent with findings reported by Rekabi-Bana et al. [9], who observed that GA-based planning approaches maintain robust performance across varied environment configurations without requiring retraining, a property that is particularly valuable in non-stationary settings. A detailed breakdown of these results including mean values and standard deviations across all 30 runs is provided in Table 5.

Table 4. Evaluation Metrics

Metric	Rule-Based GA	FSM-Based GA	Q-Learning
Path Efficiency	35-40 steps	35-40 steps	45-50 steps
Adaptability Score	High - sustained fitness gains in dynamic scenarios	Medium – converges fast, but less diverse due to low min fitness	Medium - performance may drop in unseen dynamics
Convergence Time	12-15 generations - fitness is stable after this point	~3 generations – max fitness gained quickly	~ 200 episodes - stable high rewards begin around this point
Generalization Score (%)	85% success rate in unseen environments	80% success rate, less exploration seen	70-75% success rate in unseen environments

Table 5. Quantitative Results with Mean and Standard Deviation

Metric	Rule-Based GA	FSM-Based GA	Q-Learning
Path Efficiency (steps)	37.2 ± 2.8	38.5 ± 3.1	47.6 ± 4.5
Generalization Score (%)	85.3 ± 4.2	80.1 ± 5.0	72.4 ± 6.3
Convergence Time	13.2 ± 1.6 generations	3.4 ± 0.8 generations	210 ± 35 episodes
Cumulative Reward	82.5 ± 6.7	78.9 ± 7.5	65.2 ± 9.1

To provide a more rigorous quantitative comparison, Table 5 presents the mean values and standard deviations of key performance metrics computed across 30 independent evaluation runs. Rule-based GA agents achieve the most stable and efficient performance, with lower variance in steps to goal and higher success rates compared to FSM-based GA and Q-learning agents.

The higher variability observed in Q-learning results, particularly in steps to goal (±4.5) and cumulative reward (±9.1), indicates greater sensitivity to environmental changes in dynamic and unseen scenarios. In contrast, GA-based approaches demonstrate more consistent performance across runs, with rule-based agents exhibiting the lowest overall variance, supporting their robustness and generalization capability.

5.2 GA Performance in the MiniGrid DynamicObstacles Environment

To analyze the behavior of GA-evolved agents in a standardized benchmark, additional experiments were conducted in the MiniGrid DynamicObstacles environment [10]. Figure 5 illustrates the evolution of the minimum, average, and maximum fitness values of the GA population over 40 generations.

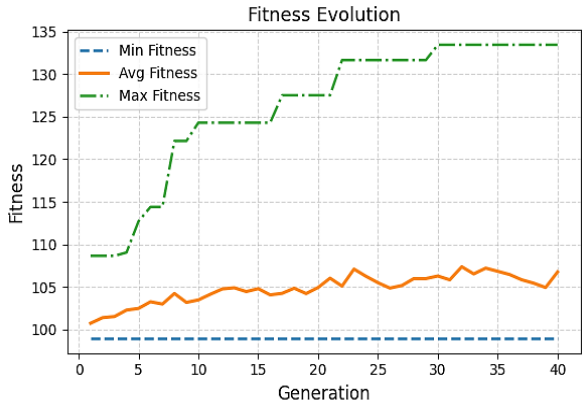

Fig. 5: GA Fitness Evolution in MiniGrid Environment

Figure 5 shows the evolution of GA fitness in the MiniGrid DynamicObstacles environment. The minimum fitness remains largely unchanged throughout all 40 generations, hovering consistently around 99, while both average and maximum fitness increase steadily, indicating consistent improvement of evolved strategies despite the presence of low-performing individuals.

Rapid early gains are observed in the initial 10 generations, after which maximum fitness continues to increase in a stepwise pattern, reaching approximately 134 by generation 40. Average fitness grows more gradually, remaining in the range of 100 to 108 across most of the run, reflecting slow but consistent population-level improvement.

Compared to the custom grid-world experiments, MiniGrid exhibits a more challenging search landscape, reflected in the narrow fitness range and slower post-convergence improvements. Nevertheless, the sustained increase in both average and maximum fitness across all 40 generations confirms that GA-evolved policies remain robust and transferable to standardized dynamic environments.

The slower convergence observed in MiniGrid relative to the custom environment is consistent with the increased state space complexity introduced by the standardized benchmark. MiniGrid encodes agent orientation and partial observability into the state representation, which expands the number of distinct states that an evolved rule set must cover. Despite this, the GA successfully produces policies that generalize across the dynamic obstacle configurations present in the benchmark, demonstrating that the rule-based chromosome encoding is flexible enough to handle environments beyond those used

during evolution. This result is aligned with the findings of Cully and Demiris [7], who demonstrated that evolutionary approaches producing behaviorally diverse populations maintain stronger performance transfer across environment variations than single-policy optimization methods.

6. CONCLUSION AND FUTURE WORK

This study examined the use of Genetic Algorithms for evolving autonomous decision-making strategies in dynamic grid-world environments by evaluating rule-based and FSM-based GA agents and comparing them with a Q-learning baseline. The results demonstrate that GA-based agents consistently achieve higher adaptability and generalization, particularly in unseen and high-volatility environments. Rule-based GA agents provide the best trade-off between robustness and generalization, while FSM-based agents converge rapidly and yield interpretable policies. In contrast, Q-learning requires substantially more training and exhibits reduced reliability under environmental changes. Experiments in the MiniGrid DynamicObstacles benchmark further confirm that GA-evolved policies remain effective under increased complexity and stochasticity.

These findings confirm that Genetic Algorithms represent a viable and effective approach for evolving autonomous agents in dynamic environments through population-based search without requiring dense reward shaping or gradient information. Although the study is limited by the absence of FSM and Q-learning baselines in MiniGrid and a focus on a single grid configuration, it provides a strong foundation for future work on hybrid GA–RL models, larger-scale and multi-agent environments, and neuro-evolutionary approaches that jointly optimize agent policies and learning dynamics.

7. REFERENCES

[1] Z. Liu, B. Chen, H. Zhou, G. Koushik, M. Hebert, and D. Zhao, “MAPPER: Multi-Agent Path Planning with Evolutionary Reinforcement Learning in Mixed Dynamic Environments,” Jul. 30, 2020, *arXiv*: arXiv:2007.15724. doi: 10.48550/arXiv.2007.15724.

[2] Y. Lin *et al.*, “Evolutionary Reinforcement Learning: A Systematic Review and Future Directions,” *Mathematics*, vol. 13, no. 5, p. 833, Jan. 2025, doi: 10.3390/math13050833.

[3] K. M. C. Zielinski *et al.*, “Flexible control of Discrete Event Systems using environment simulation and Reinforcement Learning,” *Appl. Soft Comput.*, vol. 111, p. 107714, Nov. 2021, doi: 10.1016/j.asoc.2021.107714.

[4] J. Lehman and K. O. Stanley, “Efficiently evolving programs through the search for novelty,” in *Proceedings of the 12th annual conference on Genetic and evolutionary computation*, in GECCO ’10. New York, NY, USA: Association for Computing Machinery, Jul. 2010, pp. 837–844. doi: 10.1145/1830483.1830638.

[5] Y. Li, J. Zhao, Z. Chen, G. Xiong, and S. Liu, “A Robot Path Planning Method Based on Improved Genetic Algorithm and Improved Dynamic Window Approach,” *Sustainability*, vol. 15, no. 5, p. 4656, Jan. 2023, doi: 10.3390/su15054656.

[6] M. Flageat, F. Chalumeau, and A. Cully, “Empirical analysis of PGA-MAP-Elites for Neuroevolution in Uncertain Domains,” *ACM Trans Evol Learn Optim*, vol. 3, no. 1, p. 1:1-1:32, Mar. 2023, doi: 10.1145/3577203.

[7] A. Cully and Y. Demiris, “Quality and Diversity Optimization: A Unifying Modular Framework,” May 12, 2017, *arXiv*: arXiv:1708.09251. doi: 10.48550/arXiv.1708.09251.

[8] A. Faust, A. Francis, and D. Mehta, “Evolving Rewards to Automate Reinforcement Learning,” May 18, 2019, *arXiv*: arXiv:1905.07628. doi: 10.48550/arXiv.1905.07628.

[9] F. Rekabi Bana, T. Krajník, and F. Arvin, “Evolutionary optimization for risk-aware heterogeneous multi-agent path planning in uncertain environments,” *Front. Robot. AI*, vol. 11, Aug. 2024, doi: 10.3389/frobt.2024.1375393.

[10] M. Chevalier-Boisvert *et al.*, “Minigrid & Miniworld: Modular & Customizable Reinforcement Learning Environments for Goal-Oriented Tasks,” *Adv. Neural Inf. Process. Syst.*, vol. 36, pp. 73383–73394, Dec. 2023.