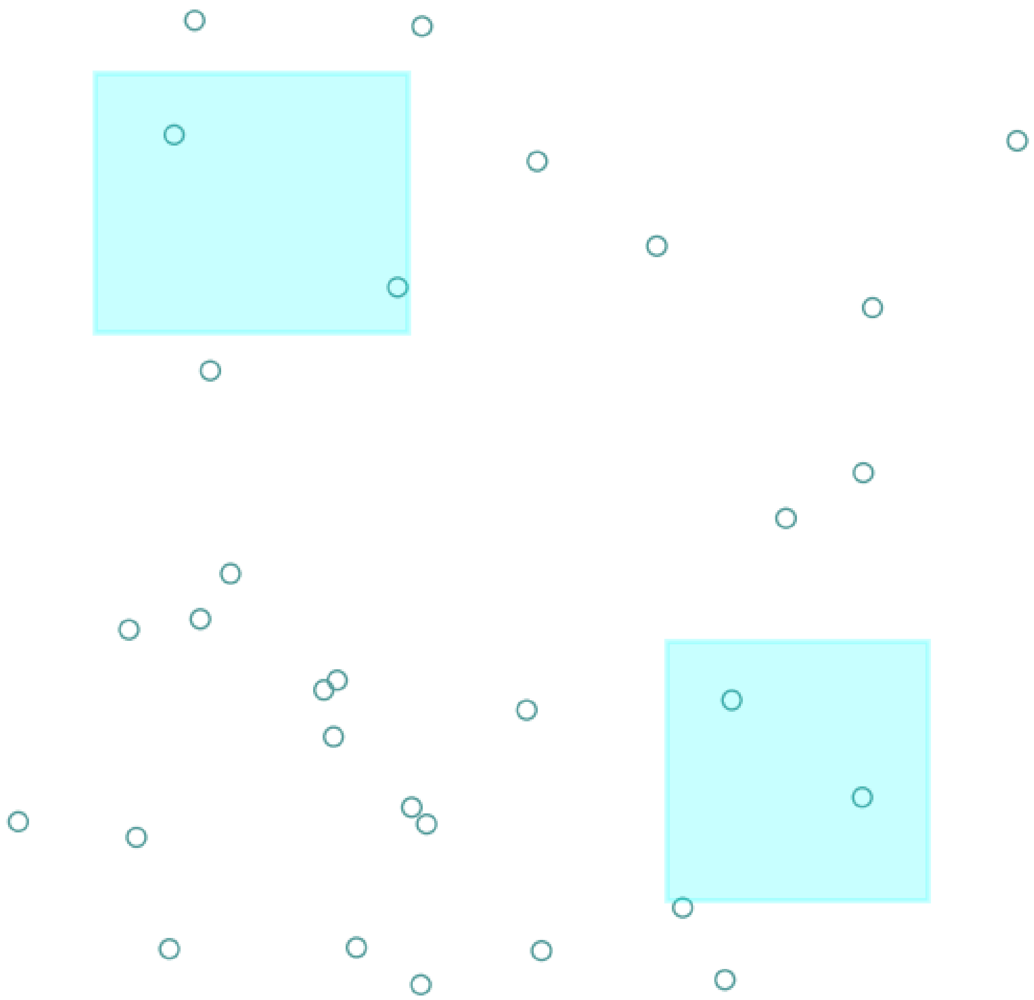
Darwin Meets Deep Learning: Teaching Neural Agents to Survive and Thrive through Neuro-Evolution

11 min read · 3 days ago

Krishna Barai

Listen Share More

We explored a neuro-evolution approach where swarms of simple “prey” agents learn to collect food and avoid predators in a cluttered 2D world. Each prey is controlled by a tiny neural network, whose parameters (weights and biases) are encoded as a genome. Over many generations, genetic operators (selection, crossover, mutation) evolve these genomes toward better survival strategies. The motivation is to blend machine learning with evolutionary biology: instead of training by gradient descent, we let selection on behavior drive the learning of control policies.



Generation 1000: Neural agents now instinctively navigate toward food zones and avoid poison, showcasing emergent behavior shaped purely through evolutionary pressure.

Environment: Food, Predators, and Obstacles

The virtual world is a square arena with food patches on one side and impassable walls on the other. Obstacles (vertical walls) block the middle of the field — specifically two vertical zones at x≈–4.5 and x≈+4.5 (see code defining obstacle_zones) — forcing agents to navigate around them. There are also hazard zones (poisonous areas) in the center that penalize agents stepping inside. Multiple predator agents roam this world hunting prey. A predator simply moves toward the nearest alive prey at each timestep (see the Predator step in code). If a prey gets too close to any predator, it is marked “caught” and effectively removed. Thus prey must learn both to find food and to avoid these predators. The prey receive a small fitness penalty if caught (parameter ε) and a bonus δ for staying near kin (agents of similar color) as a survival aid . Food is scattered in random patches on one side of the arena. Prey slowly lose energy each step and can regain it by eating food when they come within range. Their overall fitness is a weighted sum: α × (food eaten) + β × (survival time) + γ × (energy left), plus a bonus δ for nearby kin, minus a penalty ε if caught by a predator . In our runs we used α=1.0, β=0.25, γ=0.1, δ=0.5, ε=1.0 (see code comments at).

Neural Controller Architecture

Each prey has a feedforward neural network that maps its sensory inputs to actions. The network has no hidden layers — it is essentially a linear map with a tanh nonlinearity on the outputs. Specifically, we use a fixed-size genome vector that is reshaped into a weight matrix and bias vector, with an additional output scaling factor. In code, the genome layout is defined as w (weights), b (biases), and out_scale (output gains) . During agent initialization, a random genome is generated and then split into w , b , and out_scale as in the snippet

```
idx = 0
        self.w = genome[idx:idx+self.input_size*self.output_size].reshape((self
        idx += self.input_size*self.output_size
        self.b = genome[idx:idx+self.output_size]
        idx += self.output_size
        self.out_scale = np.abs(genome[idx:idx+self.output_size]) + 0.2

out = np.tanh(np.dot(x, self.w) + self.b) * self.out_scale
```

On each timestep, a prey computes its input vector and feeds it through forward() to get an output. The inputs include the prey's own 2D position (normalized), the direction vector to the nearest food source, the count of nearby allies, and the average "signal" being broadcast by those allies . (The last two inputs implement a simple group-awareness: each prey senses how many kin are nearby and what signal value they are sending.) The network's three outputs are (dx, dy) — a movement vector — and a scalar "signal" value that the prey can broadcast to others (clipped between –1 and 1) . In effect, each prey can communicate a single number to nearby peers each turn.

Genome Encoding and Evolution

We maintain a fixed population of prey, each with its own genome and color (used only for kin-group bonuses). At each generation, every prey is placed at a random start location, and the simulation runs for a fixed lifetime of steps (until energy runs out or caught). After the run, each agent's fitness is computed (food, survival, energy, kin bonus, predator penalty) . We then select parents for the next generation via roulette-wheel selection (higher fitness yields higher chance), with a small probability of selecting a random outsider ("drift") to maintain diversity. Offspring are created by standard sexual reproduction: for each new individual we pick two parents and perform uniform crossover on their genomes . In practice, each gene in the child's genome is randomly taken from one parent or the other. After crossover we apply Gaussian mutation to each gene with a small chance (flipping some weight by a small random amount). The two chosen parents each pay half of an "energy cost" for producing the child. The child's color is set to the average of its parents' colors (simulating lineage) , and its neural net is initialized from the combined genome. This process yields a new population of the same size. Over many generations, fitter behaviors tend to accumulate.

Results: Learning to Survive

The evolution consistently improved the agents. ***Figure 1*** (below) shows a typical run's fitness curves over generations: both the best individual (solid) and the population average (dashed) climb rapidly, indicating that prey are learning to eat more food and survive longer in a dangerous environment .

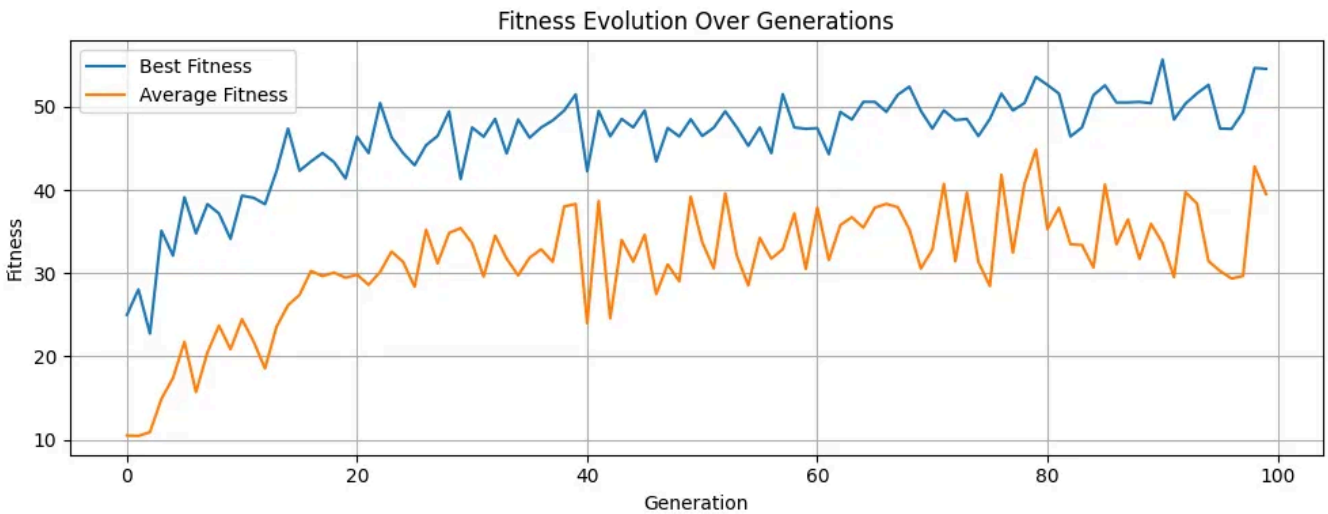


Figure: Evolution of the best (solid) and average (dashed) fitness across generations in one run. Fitness improves as agents evolve more effective foraging and evasion behaviors (fitness = α·food + β·survival + γ·energy, plus kin bonus, minus predator penalty).

Over the course of evolution, the diversity of genomes in the population steadily contracts as the most successful genetic lineages outcompete others. ***Figure 2*** (below) plots a standard diversity metric — here the average pairwise distance between genomes — over successive generations. At first, diversity is high, reflecting random initialization and exploration. As selection favors high-fitness controllers, diversity plunges: a handful of winning genomes proliferate, much like a genetic bottleneck in natural populations where a particularly advantageous trait sweeps through the gene pool. By generation Y, diversity stabilizes at a low baseline, indicating that most individuals share very similar genomes. This pattern mirrors nature's own "selective sweep," where the success of one genotype leads to reduced variation until new mutations or environmental pressures reintroduce novelty.

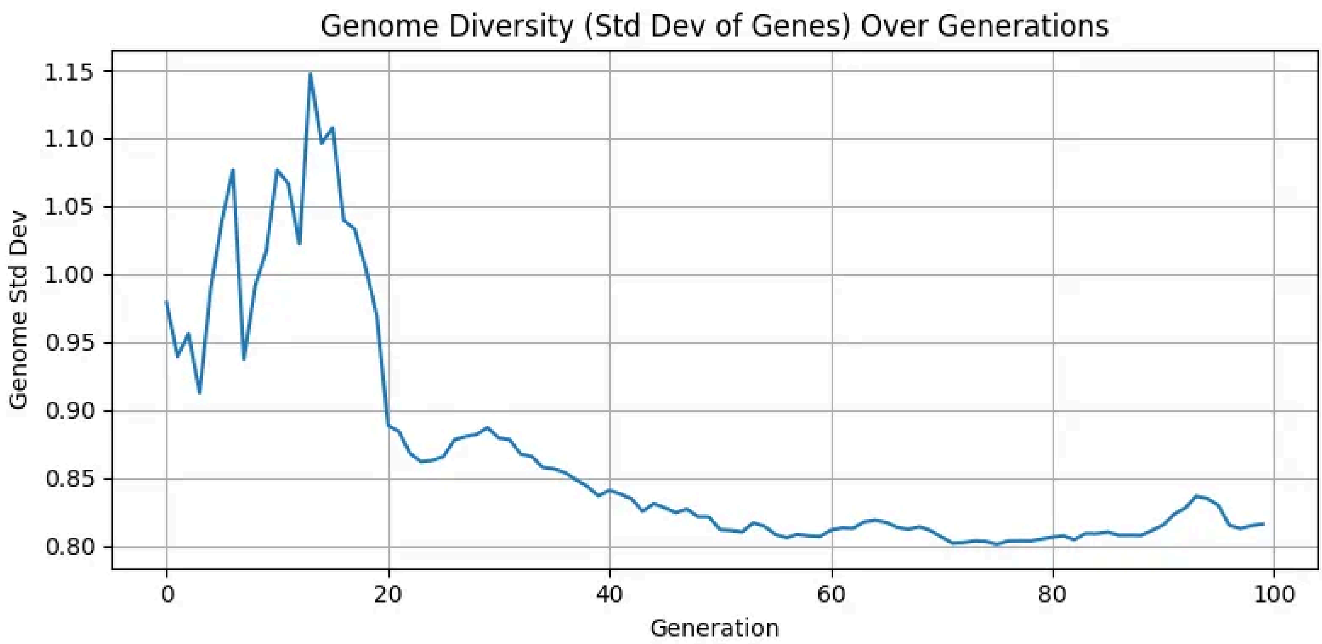


Figure: Decline in average pairwise genome distance over generations. Initial exploration gives way to a sharp drop as high-fitness genomes dominate, echoing natural selective sweeps that reduce genetic diversity.

We also compiled a heatmap of all positions visited by the population in the last generation. As shown in ***Figure 3***, the density of visits is highest in the safe, food-rich zones and along corridors around the obstacles. The learned policy focuses the swarm's activity where food is available but predators are relatively easier to avoid.

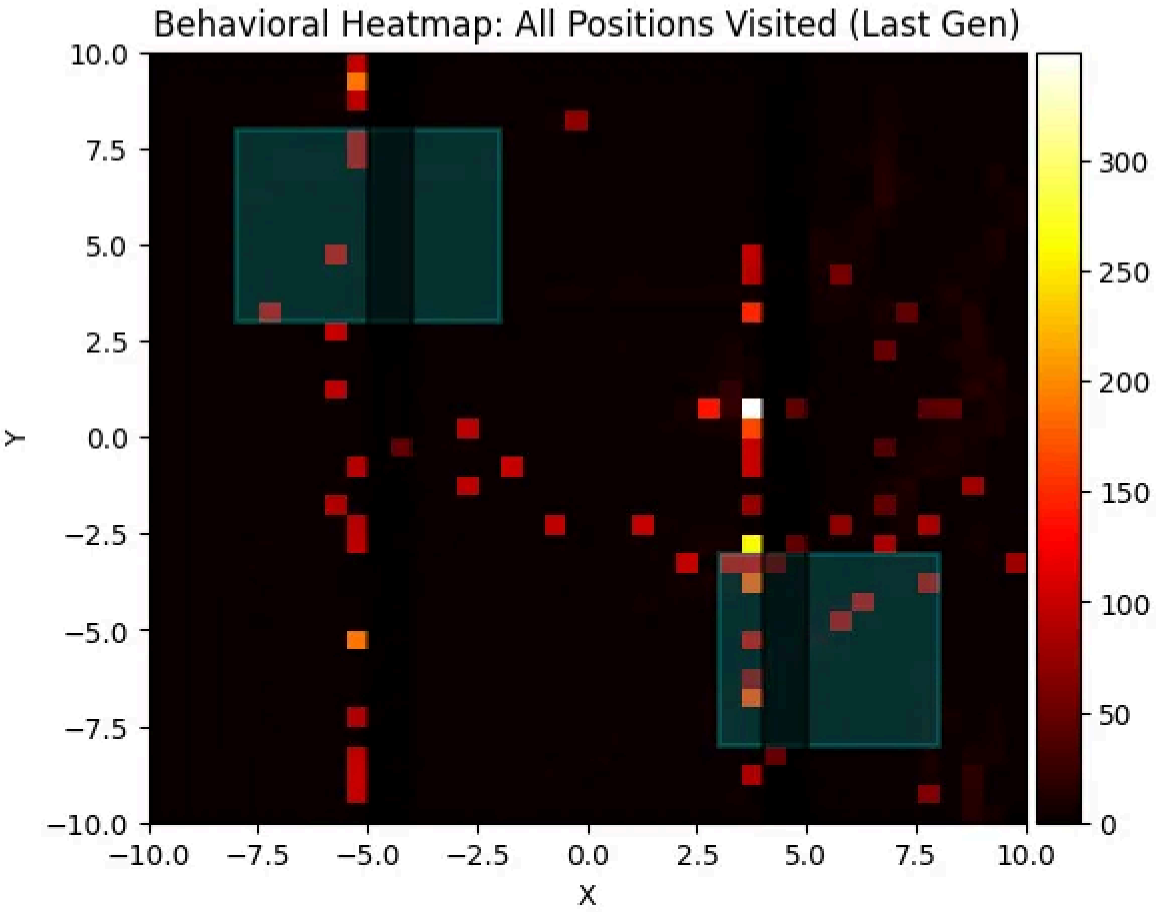
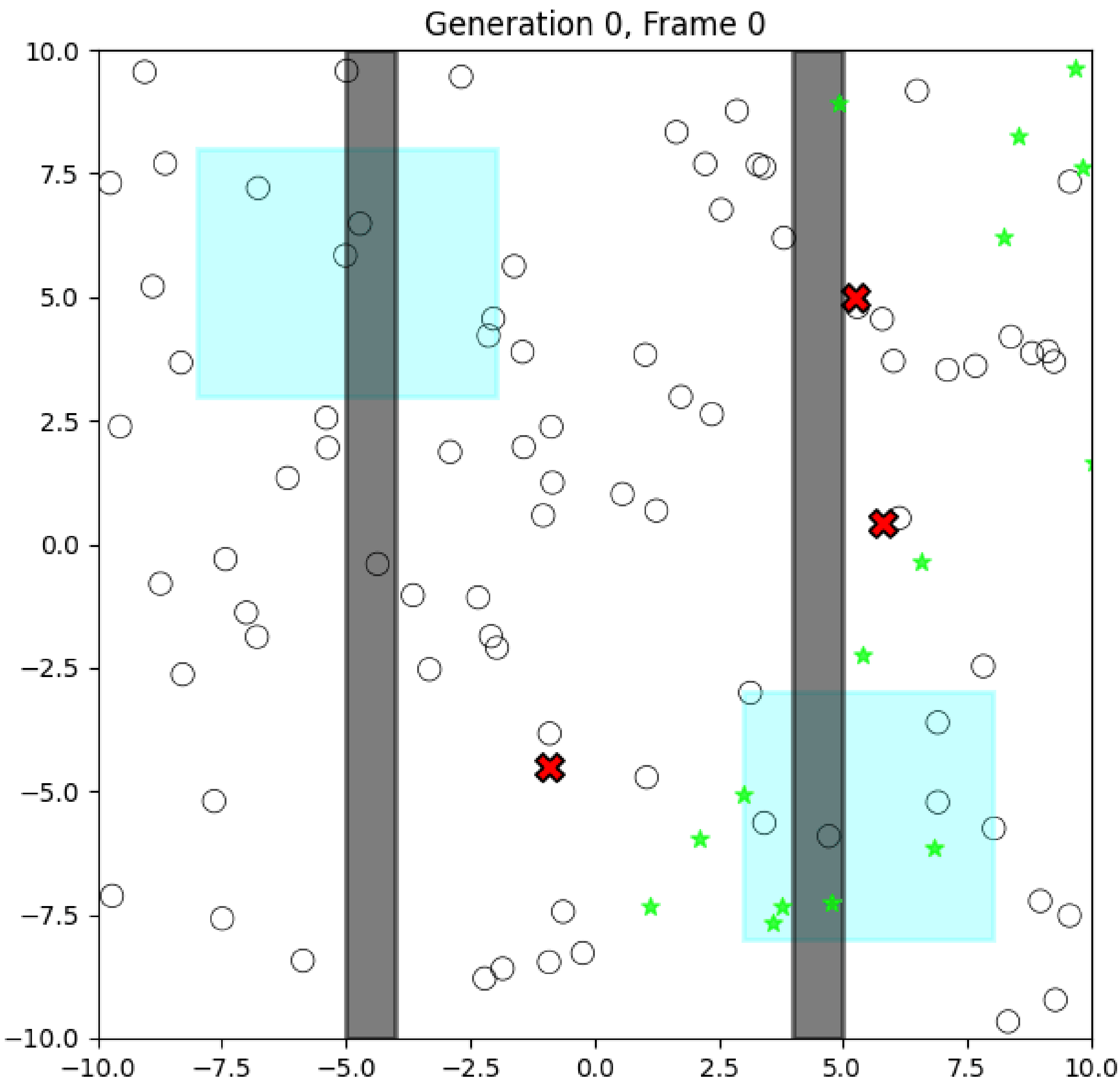


Figure: Heatmap of all positions visited by the final-generation population (darker = more time spent). Agents concentrate in the food area (right) and near safe borders of the obstacles, reflecting the evolved navigation strategy

Across runs we observed **emergent behaviors** like grouping and implicit signaling. Because we gave a bonus for being near kin (and each agent could broadcast a signal), many populations evolved to stay together when possible. In effect they share information about predator threats and food locations. Even though this communication channel was very simplistic, it encouraged a “schooling” strategy that increases survival: agents cluster when predators lurk, then fan out to gather food.

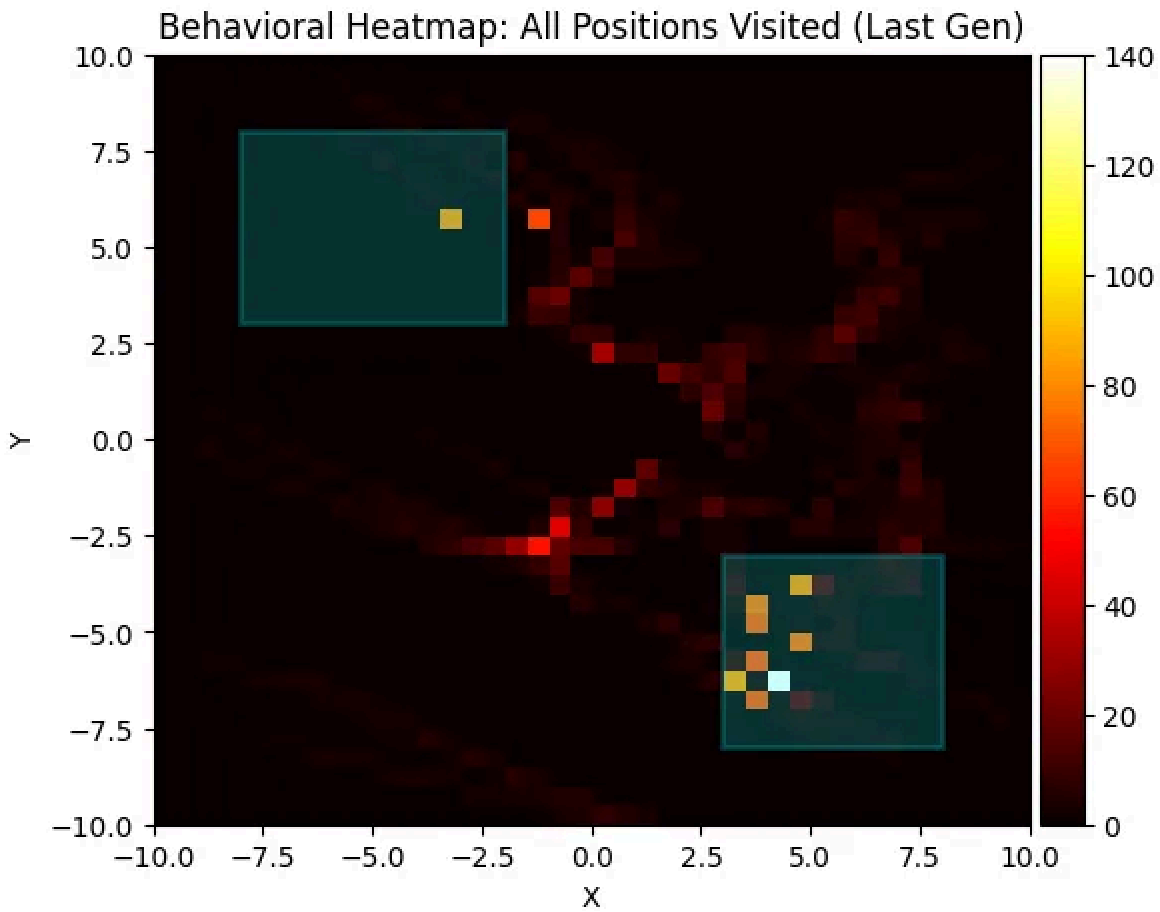


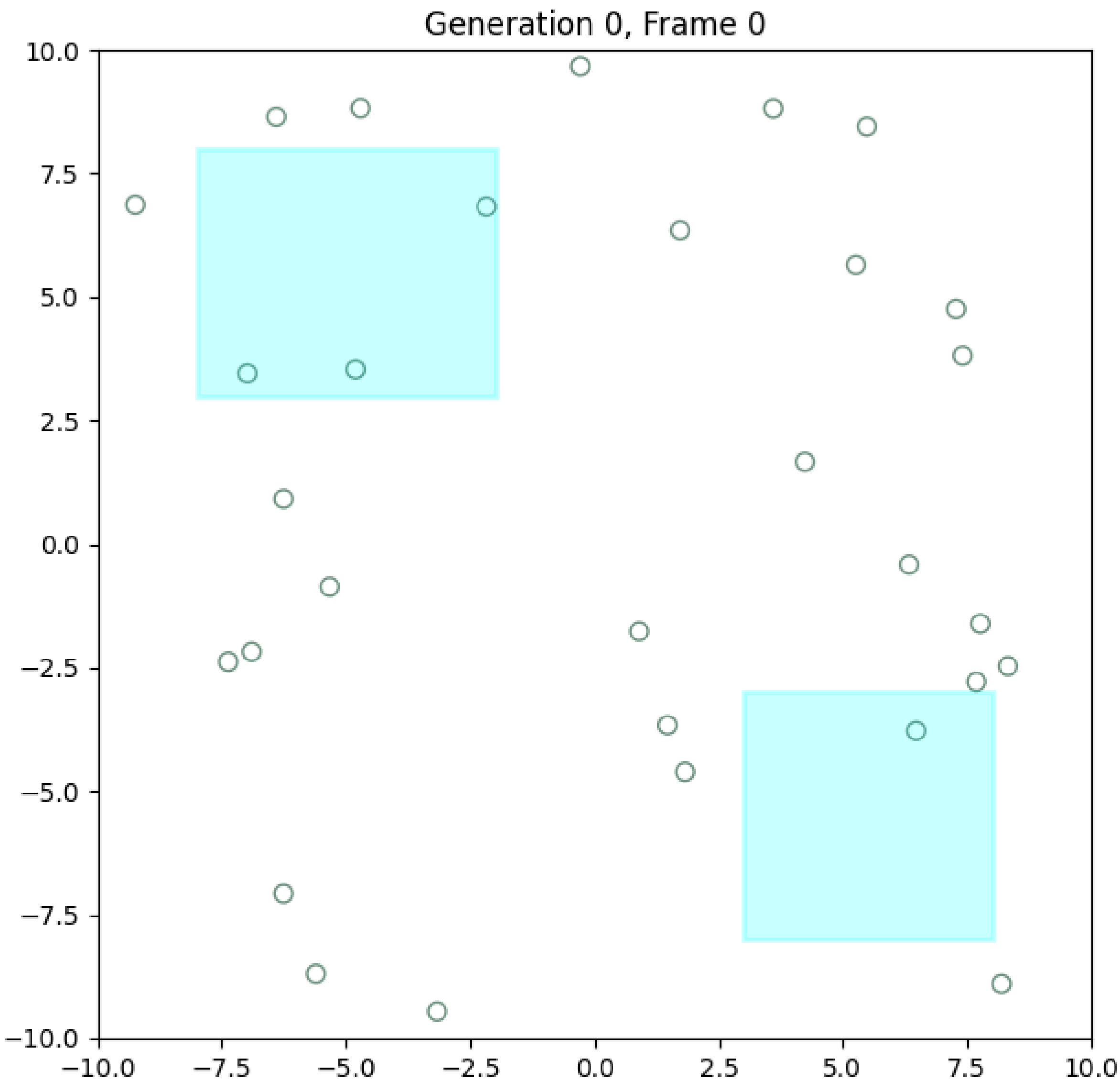
500th generation Simulation

Emergent Migration and Instinct-Like Movement Patterns:

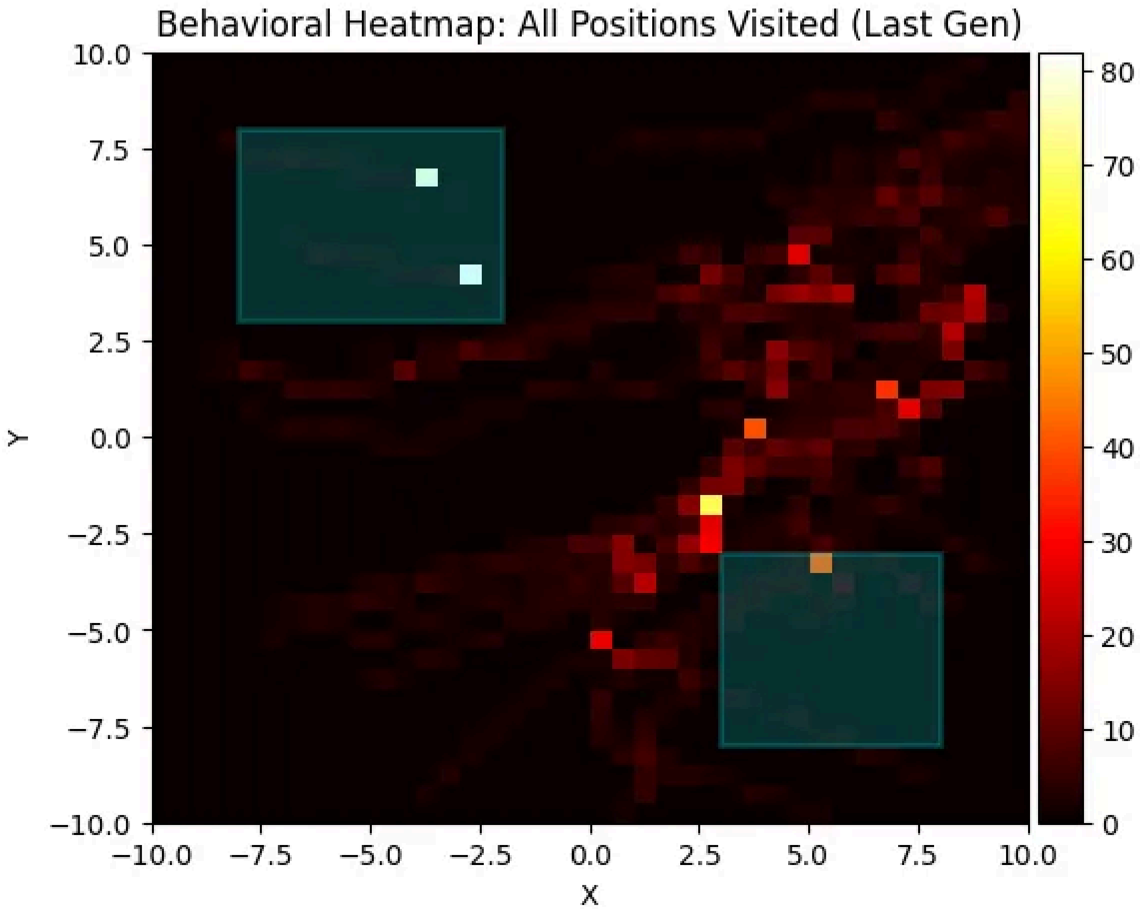
In the obstacle-free environment — where the only danger zones lie at the extreme left and right edges of the arena — the simulated agents exhibited striking, evolution-driven spatial behaviors. While initial generations showed random and dispersed movement patterns, by the final generation a distinct migratory pattern had emerged: agents consistently navigated toward the food-rich zone on the right, while carefully avoiding the hazardous boundary zones. This shift was not programmed explicitly; rather, it arose through evolutionary selection pressures alone.

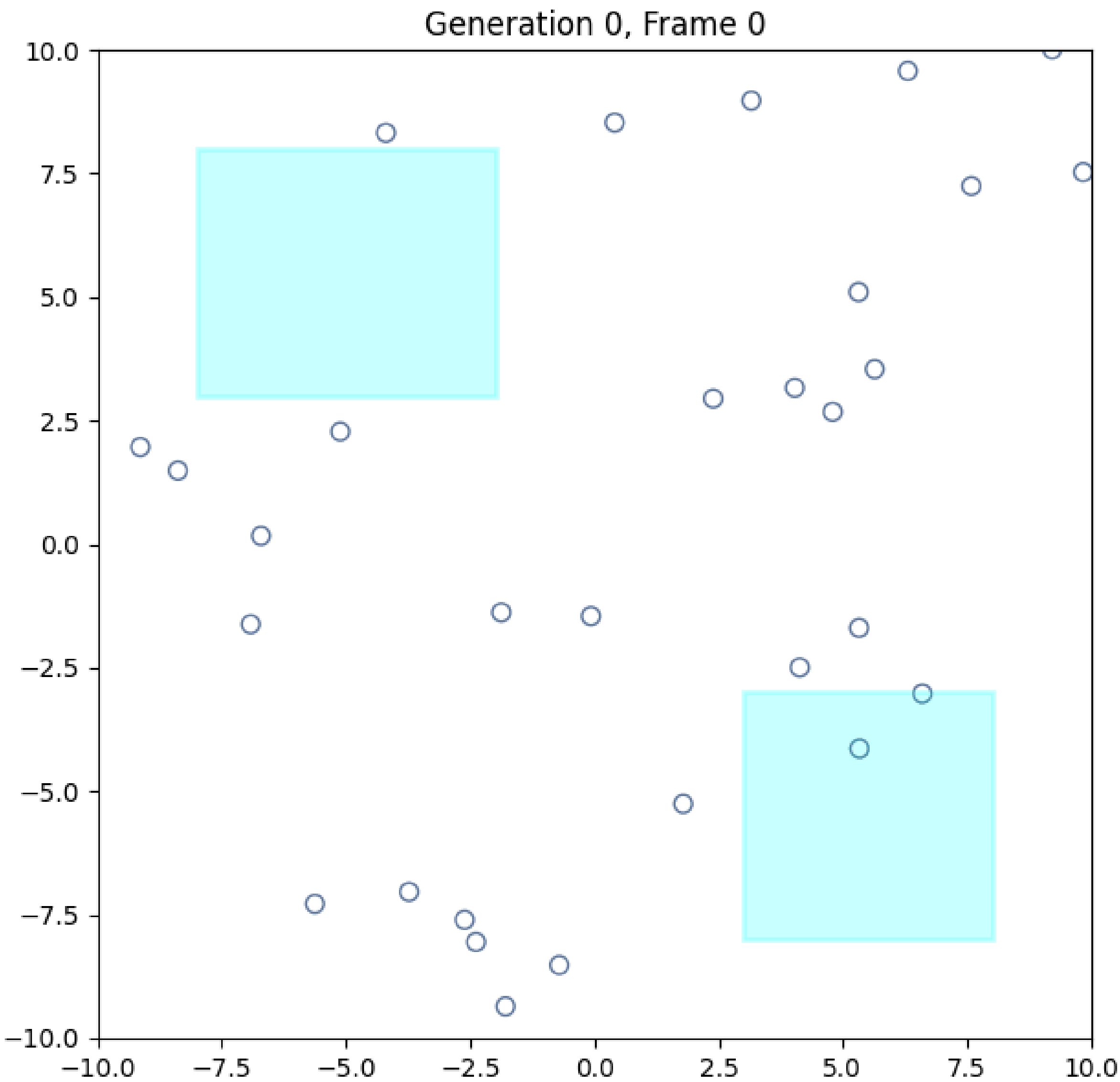
As evolution progressed, neural controllers that successfully guided agents toward food while minimizing exposure to poisonous zones were preferentially selected. These successful genomes were replicated and gradually came to dominate the population. The result, visible in the final-generation heatmap (**Figure 4**), is a concentrated corridor of movement along safe paths, particularly around the periphery of the hazard zones. This corridor is not hard-coded but instead reflects evolved behavioral instincts, much like how migratory animals in nature — such as birds or wildebeests — navigate toward seasonal resources without ever being taught the route.



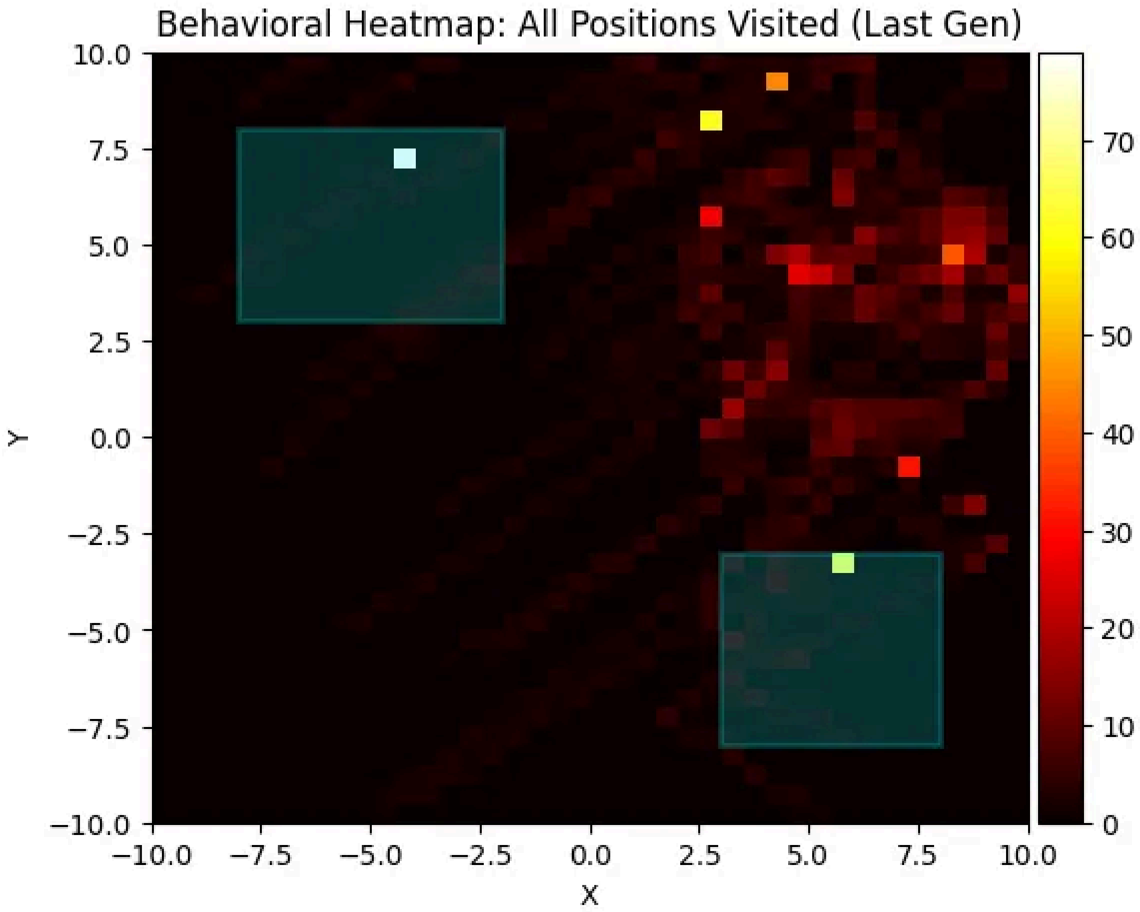


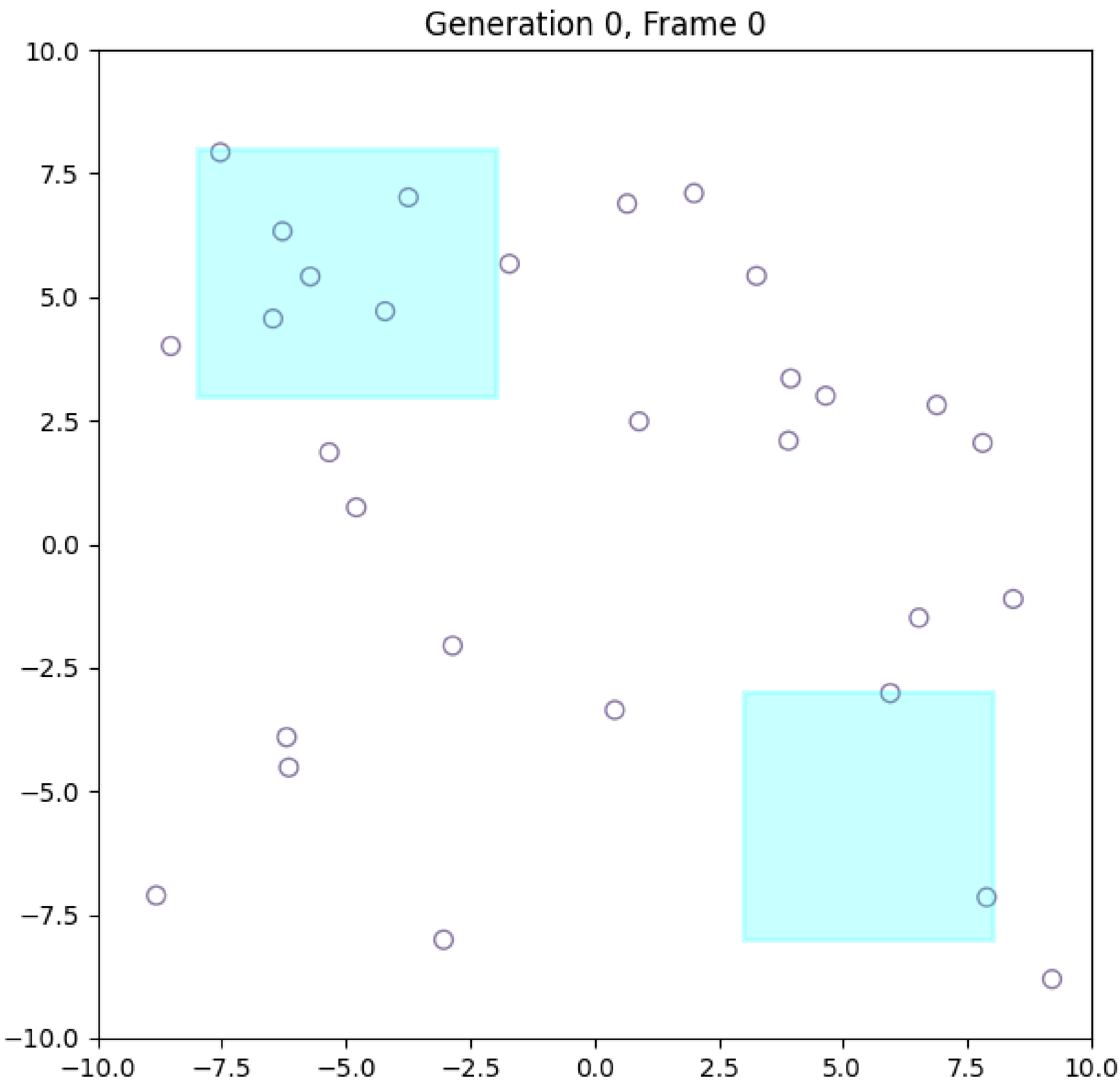
By Generation 100, basic selection pressures have begun to nudge agents away from random movement. While there's no clear strategy yet, we start to see the earliest hints of directional preference. The population still roams widely, with many agents venturing into hazardous zones — natural selection is just beginning to carve the first faint grooves of adaptive behavior.


Behavioral Heatmap: All Positions Visited (Last Gen)
Y
X
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
10.0
80
70
60
50
40
30
20
10
0



800 generations in, a dramatic shift is underway. Agents increasingly cluster near the resource-rich zones, with fewer individuals straying into danger. The once-chaotic landscape now shows early signs of an evolved "safe corridor" — an implicit, collectively learned migratory strategy driven solely by neural evolution and survival outcomes.


Behavioral Heatmap: All Positions Visited (Last Gen)
Y
X
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
10.0
70
60
50
40
30
20
10
0



By Generation 1000, a robust behavioral instinct has clearly emerged. The heatmap shows a high-density corridor through which prey consistently travel to reach the food-rich areas while skirting hazards. This isn't hardcoded or learned through backpropagation — it's the result of pure evolutionary dynamics, echoing natural migration patterns seen across biological systems.

Interestingly, many agents appear to move as a loose collective, despite no communication being encoded between them. This emergent group cohesion can be understood as a byproduct of shared ancestry: as evolution favors certain movement strategies, genetically similar agents exhibit nearly identical behaviors. This convergence creates the illusion of coordinated migration, though each agent operates entirely independently. The behavioral consistency across individuals reflects evolutionary canalization — a phenomenon observed in nature where development and behavior become stabilized through selective pressures.

In summary, what emerges is a compelling analog to biological migration: agents evolve not just to survive, but to move purposefully through space in ways that resemble instinctual, adaptive behavior in real animals. The heatmap serves as a final snapshot of this evolutionary trajectory, capturing how simple selection forces can produce complex, spatially organized group dynamics in artificial neural systems.

Adaptive Minds in Conflict : Co-Evolution

How shared neural anatomy becomes a battleground for emergent evolutionary behavior.

At the heart of this project lies a co-evolutionary duel between two neural populations — prey and predators — each powered by independently evolving feed-forward neural networks. Importantly, both groups share an identical genome structure: the same layout of weights, biases, and output scaling. This symmetry ensures that any differences in behavior are not due to design bias, but emerge purely through evolutionary dynamics. The prey, shaped by selective pressure, evolve to maximize their food intake, survival duration, and energy efficiency. In contrast, the predators evolve under opposing incentives: their fitness depends on how effectively they capture prey, maintain energy, and prolong their own survival. Together, this creates a tightly coupled feedback loop — each side adapting in response to the other, generation after generation — mirroring the evolutionary arms races seen in natural ecosystems.

Reproduction is handled via a biologically inspired process: roulette-wheel selection ensures that fitter individuals are more likely to reproduce, while uniform crossover and mutation introduce variation across generations. The agents must navigate a non-trivial environment — a food gradient rewards movement toward one side, while hazard zones penalize the other. This spatial pressure forces both populations to weigh foraging, aggression, and risk in increasingly nuanced ways. What's striking is that, despite no explicit communication or coordination, certain prey appear to migrate in loosely cohesive groups as generations progress, potentially exploiting the safest and most rewarding pathways — an emergent phenomenon reminiscent of flocking or collective risk-avoidance in nature.

Over 200 generations, I tracked multiple indicators — peak vs. average fitness, genome-wide variability, and spatial dispersion of individuals. These metrics uncover the hidden dynamics of the system: bursts of co-adaptation, temporary stability, and even evolutionary bottlenecks. And finally, watching the final-generation simulation unfold — predators darting and converging on evasive prey, prey weaving through danger zones in search of sustenance — feels like witnessing an accelerated documentary on evolution itself. What emerges is not merely code or animation, but a living, evolving system — one that surprises even its creator.

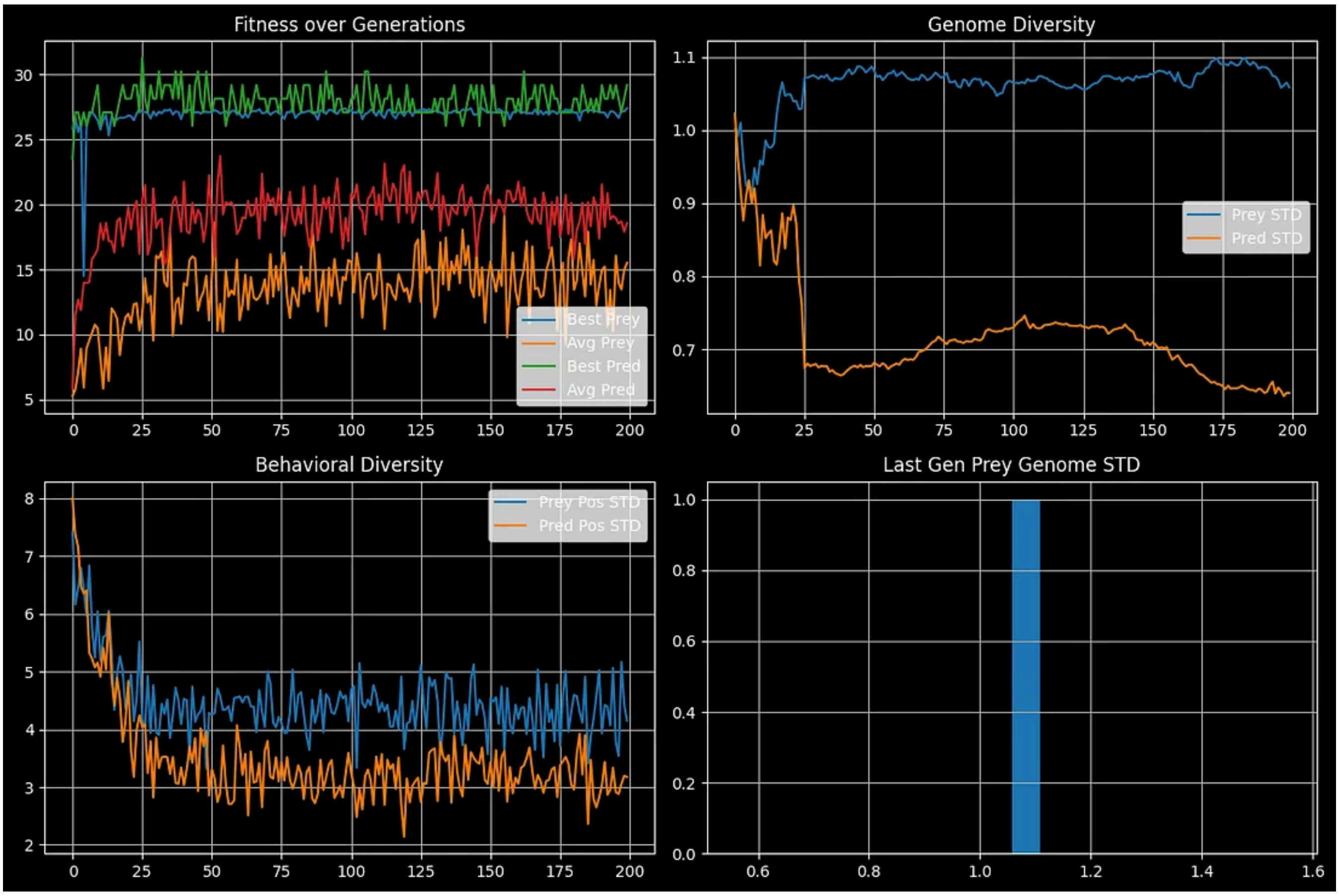


Figure: Co-Evolutionary Dynamics Over 200 Generations.
Top left: Fitness trajectories show sustained improvement and co-adaptive stability, with prey consistently outpacing predators in both average and best-performing individuals.
Top right: Genome diversity (standard deviation of weights) diverges sharply — prey maintain high genetic variation, while predator diversity collapses early, hinting at premature convergence.
Bottom left: Behavioral diversity (spatial spread) mirrors genomic trends, with prey exhibiting greater positional variability, suggesting more exploratory or evasive strategies.
Bottom right: Final-generation genome diversity for prey clusters tightly around ~1.05, reinforcing the persistence of genetic exploration despite selection pressure.
Together, these panels reveal a dynamic evolutionary arms race, punctuated by diversity bottlenecks, asymmetric adaptation rates, and emergent specialization.

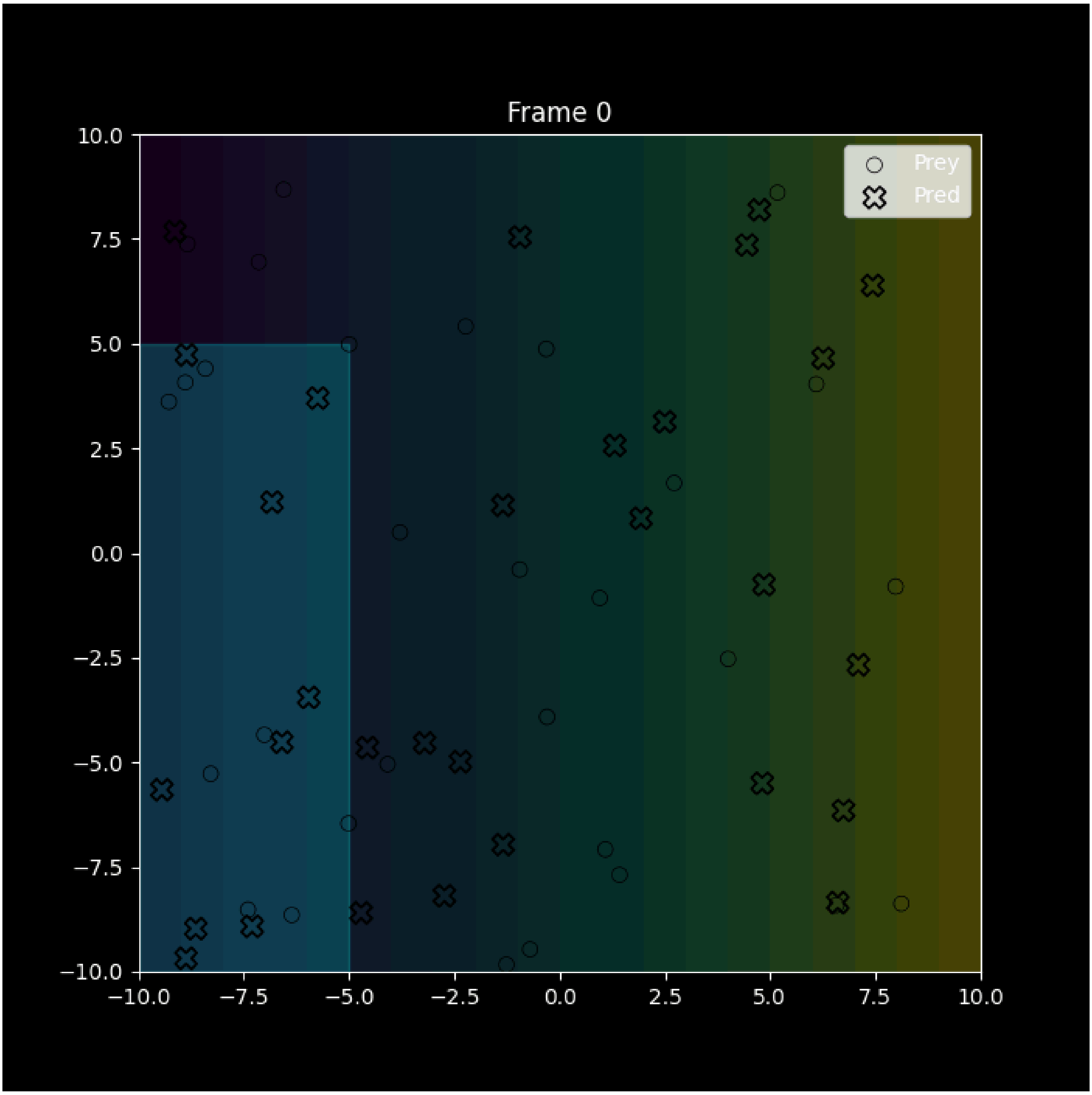


Figure: Frames of a Co-evolutionary Simulation.
Prey (circles) and predators (crosses) begin in a shared environment featuring a spatial gradient — food-rich zones to the right (warmer hues) and a hazardous dead-zone to the left (blue overlay). The stage is set for an evolutionary contest of speed, strategy, and survival.

In the very first frame of the simulation, both prey and predators are randomly distributed across a heterogeneous landscape. To the right, a nutrient-rich gradient lures prey toward energy gain. But on the left lies a hazardous region — effectively a survival penalty zone — that both species must avoid. This initial configuration poses an evolutionary dilemma: prey must balance the pull of food against the push of predators, while predators must learn to anticipate evasive prey behavior amid environmental risks. What begins as stochastic positioning soon gives rise to emergent tactics — herding, flanking, avoidance — all driven by evolving neural controllers. This is the opening gambit of an arms race written in code and competition.

Discussion and Future Directions :

This experiment demonstrates that a simple neuro-evolutionary system can discover sophisticated behaviors in a multi-agent environment. Without explicit programming, the prey learned to navigate obstacles, seek food, and avoid predators. The fitness components naturally balance trade-offs: agents that run straight for food but ignore predators get caught (high food but low survival), while overly timid agents survive but starve. Evolution finds a middle ground. There is much room to extend this setup. One could evolve the predators' neural controllers in tandem (coevolution), leading to an arms race of tactics. The input space could be enriched (e.g. richer communication signals, sensory noise, more obstacles). Scaling to larger populations or continuous evolution in an open-ended world would model more realistic ecology. Ultimately, this approach — combining evolutionary search with neural control — could inspire adaptive controllers for robotics or swarm systems facing uncertain, adversarial environments.

There is much room to extend this setup. The input space could be enriched (e.g. richer communication signals, sensory noise, more obstacles). Scaling to larger populations or continuous evolution in an openended world would model more realistic ecology. Ultimately, this approach — combining evolutionary search with neural control — could inspire adaptive controllers for robotics or swarm systems facing uncertain, adversarial environments.

📌 Key Takeaways :

- This project demonstrates how Darwinian selection and neural control can work together to evolve adaptive agents without gradient-based training.
- A population of simple prey agents controlled by shallow neural networks evolved over generations to forage for food, avoid predators, and signal kin.
- The system employs roulette-wheel selection, crossover, and mutation to evolve agent behaviors based on a custom multi-factor fitness function.
- Over time, agents naturally developed grouping, navigation, and risk-reward balancing behaviors—despite having no explicit programming or training feedback.

- The experiment highlights the potential of neuro-evolution to discover emergent strategies, making it promising for complex environments and swarm robotics.

————————————————————

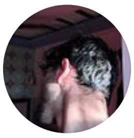
Sources: *The above description is based directly on the author's simulation code. Key implementation details (network genome layout, fitness formula, and evolutionary operators) come from the notebook (e.g. genome→weights mapping in the NeuralNetwork class , input/output definitions , and reproduction logic). Observed behaviors are inferred from the simulation outputs shown in the figures.*

🔬 Explore the Code

Curious to dive into the simulation yourself with the evolving neural agents?

The full codebase is available on GitHub:
github.com/HeroicKrishna160905/Neuro-Evolution

Clone it, run it, experiment with it — and maybe even evolve your own version of intelligence.



Edit profile

Written by Krishna Barai

0 followers · 2 following

1st year Undergrad at IIEST Shibpur, with interests in Finance, DeepLearning and Data Analytics. LinkedIn: https://www.linkedin.com/in/krishna-barai-9a3889227/

No responses yet

Krishna Barai he/him