CM3020 Artificial Intelligence Midterm Coursework

Student ID: 220516606

Introduction

In this project, I will explore how virtual creatures can evolve to climb a mountain using genetic algorithms. Through a series of experiments, I tweak parameters like mutation, structure, and terrain to see what helps or hinders their ability to move effectively.

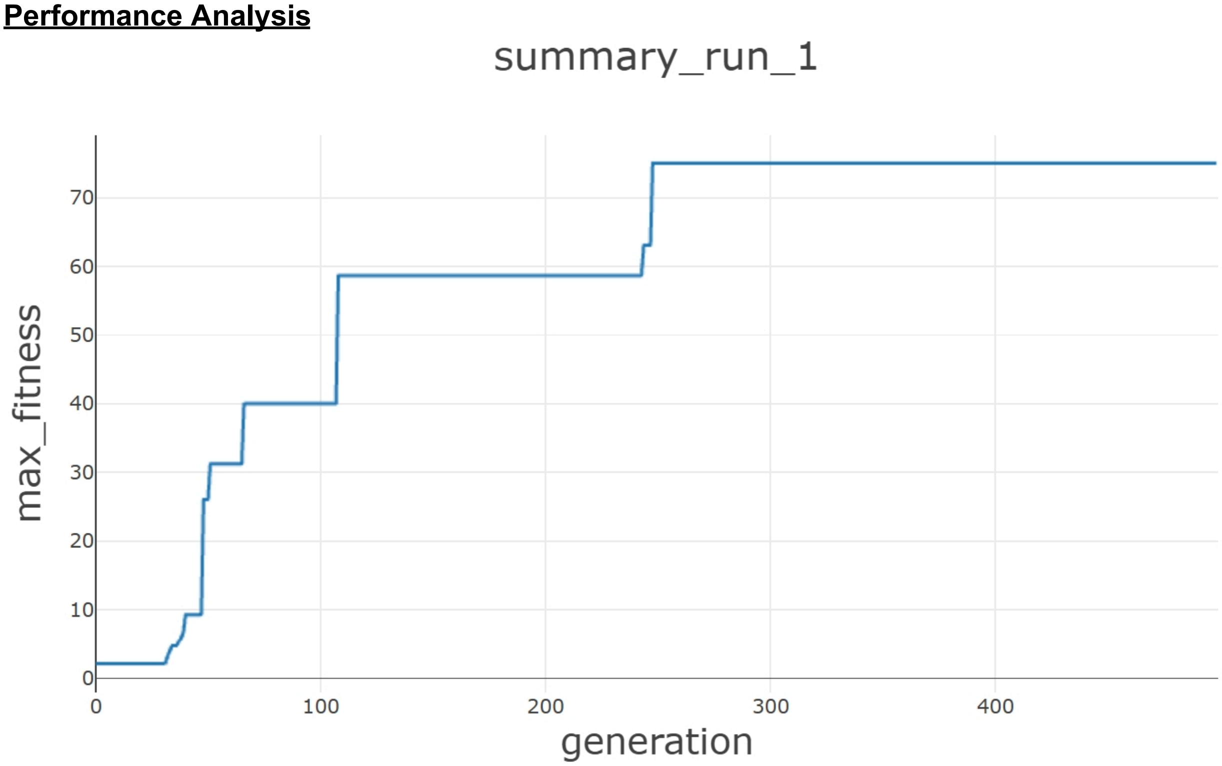
Basic Experiments

1. get_max_height()

This fitness function computes the highest z-coordinate (vertical axis) reached during simulation and assigns that as the fitness score. The rationale was to test if creatures would evolve to climb the mountain vertically, without explicitly guiding or penalising other behaviors such as lateral movement or falling off.

```
# experiment 1
    def get_max_height(self):
        return self.max_height
```

<u>Performance Analysis</u>



During playback, creatures began exploiting the system by evolving extremely tall and unstable morphologies. By generation 248, one creature reached a fitness of 75.05, largely due to its sheer size rather than meaningful movement. The final creature was so large it exceeded the mountain in height. It extended upwards instead of climbing. This was worsened by the low spawn point (z = 1), which allowed creatures to “launch” upon physics settling, creating artificially high vertical progress. While fitness scores improved, this behaviour was clearly exploitative and unrealistic.

Ultimately, this fitness function failed to encourage actual climbing. It rewarded vertical inflation and led to creatures that were structurally impressive but behaviorally useless. Hence, the approach was abandoned for the next experiment. To address these issues, the fitness function was redesigned in Experiment 2 and the spawn height was also adjusted to z = 2 to prevent launching exploits. Generations will also be lowered to 300 subsequently to reduce runtime, as fitness gains beyond that point often plateau.

2. get_combined_progress(w1=0.8, w2=0.2)

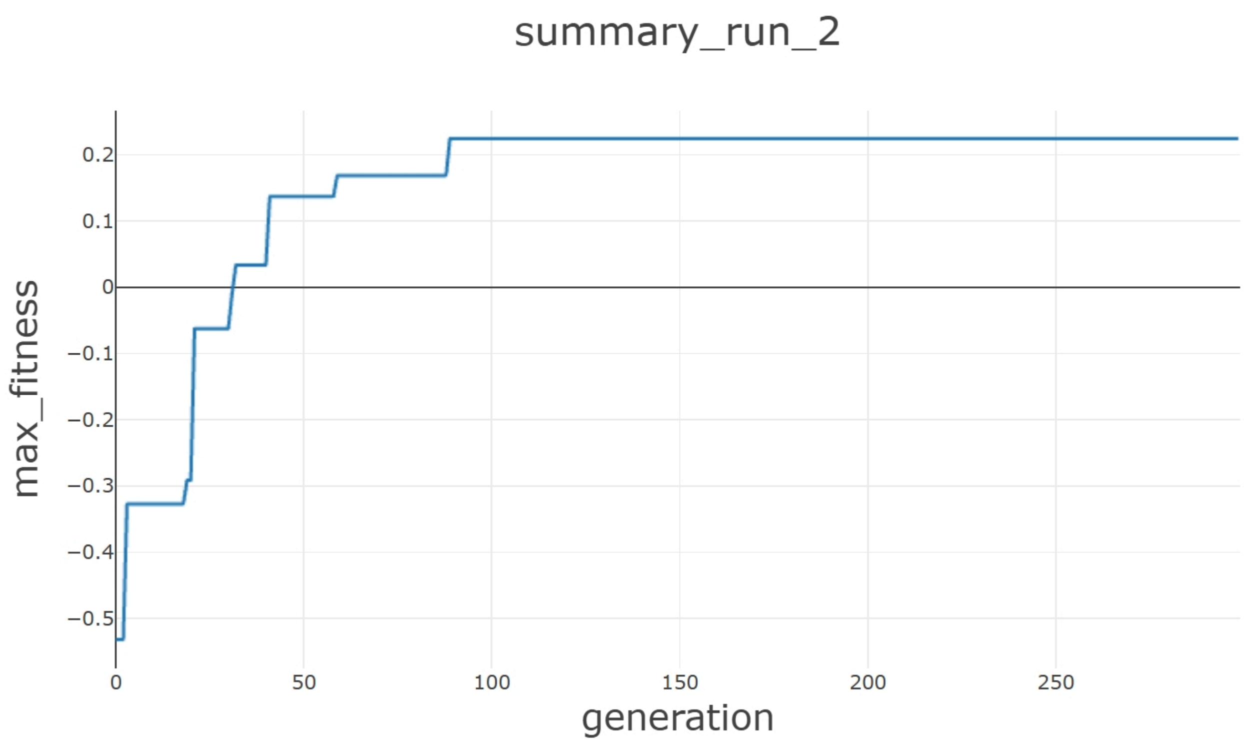
To improve upon the limitations of get_max_height, this function introduces a weighted combination of vertical and horizontal displacement.
The fitness = 0.8 * vertical_progress + 0.2 * horizontal_progress prioritises height (weight 0.8) while still rewarding lateral movement (weight 0.2) so that creatures are encouraged to evolve more efficient locomotion strategies rather than only maximizing vertical position.

```
# experiment 2
    def get_combined_progress(self, w1=0.8, w2=0.2):
        if self.start_position is None or self.last_position is None:
            return 0

        start = np.asarray(self.start_position)
        end = np.asarray(self.last_position)
        vertical_dist = end[2] - start[2]
        horizontal_dist = np.linalg.norm(end[:2] - start[:2])  # XY plane

        return w1 * vertical_dist + w2 * horizontal_dist
```

Performance Analysis



In early generations, creatures retained bulky, spherical bodies, but began evolving flatter bases and minimal limb structures. Over time, they adapted into simpler, more controlled shapes that attempted rolling or crawling strategies. A breakthrough occurred at generation 89, where a creature first reached the maximum fitness of 0.22, though it plateaued afterwards. While the number is low compared to Experiment 1, it reflects genuine terrain interaction rather than exploits.

Fitness remained low throughout, with the highest mean fitness at generation 32 (-0.76) and a final mean of -1.07, showing slow but gradual adaptation. The low values are likely due to most creatures falling or moving in suboptimal directions, which still counts negatively under the new metric.

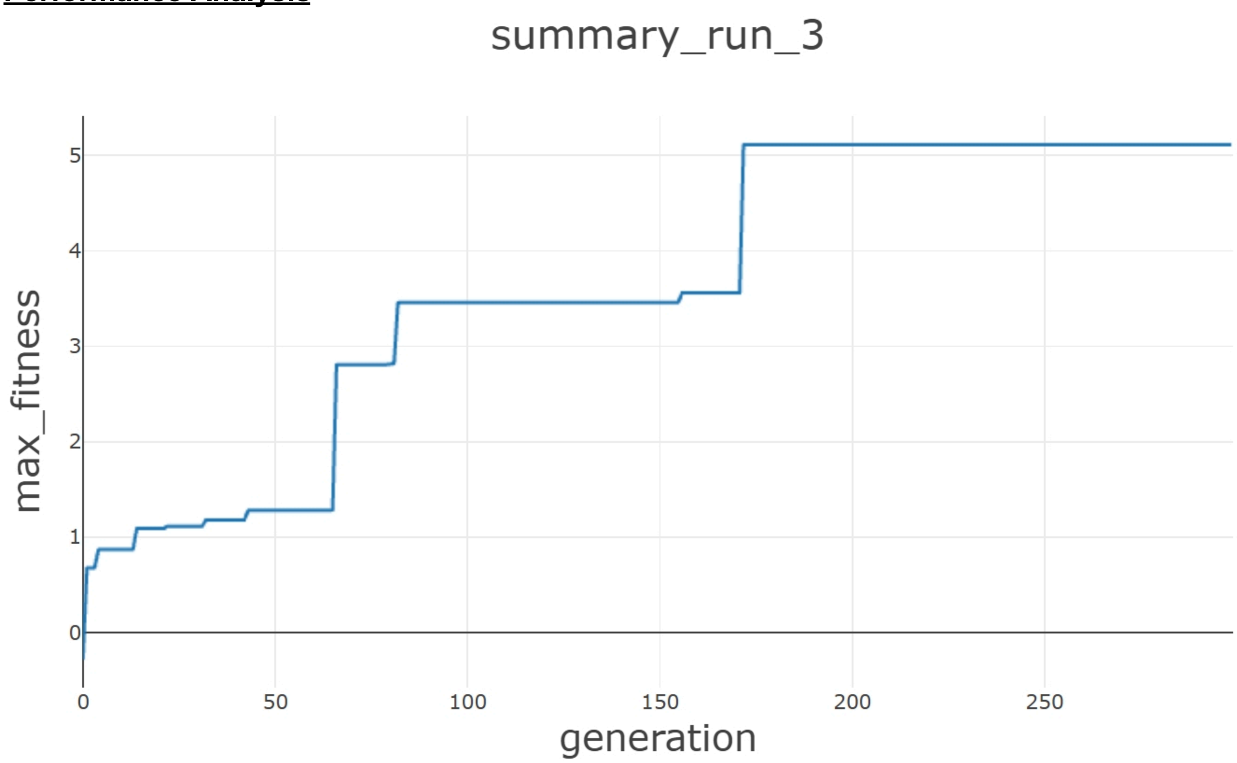
Overall, this revised fitness function successfully eliminated the vertical exploitation seen in Experiment 1. Though scores were lower, behaviour became more grounded and meaningful, laying a fairer foundation for future experiments. Therefore, I will continue using this fitness function in subsequent experiments.

3. Adjusting Population Size (pop_size = 25)

This experiment increased the population size from the default 10 to 25, to explore whether a larger gene pool could improve fitness and climbing behaviours under the same combined fitness function.

```
pop = population.Population(pop_size=25,
                            gene_count=3)
```

Performance Analysis



Earlier generations evolved unstable spherical forms, but by generation 80, creatures began showing structured shapes with more coordinated movement. The breakthrough emerged at generation 172, achieving a fitness of 5.11, which remained the peak throughout. The highest mean fitness occurred at generation 210 (0.57), suggesting a brief phase of population-wide coherence. However, by generation 299, the mean dropped to -0.13, showing that while elite creatures performed well, the average stability declined.

While the experiment produced stronger individuals, the population also grew more variable, and some creatures began to resemble larger, bulkier designs that risked relapsing into the exploitative behaviours seen in Experiment 1.

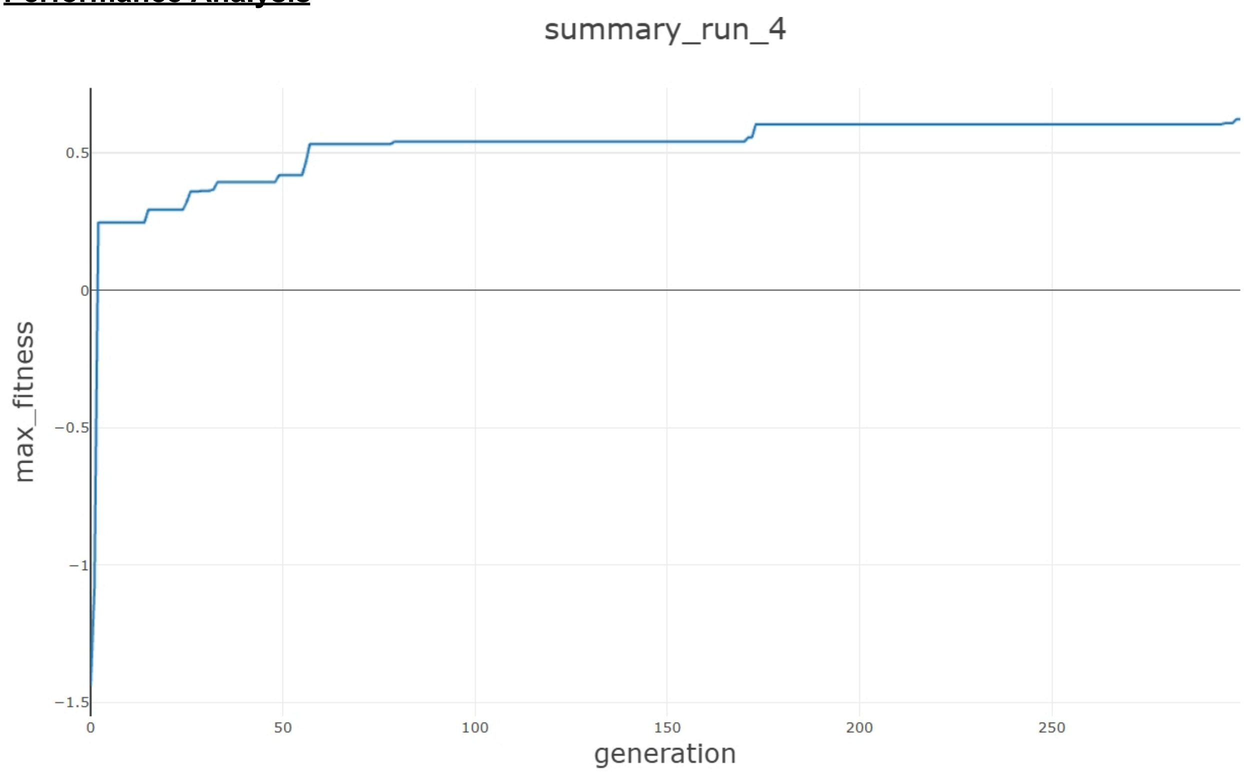
To prevent uncontrolled growth and isolate the effects of mutation changes, I decided to revert to a population size of 10 for subsequent experiments. This allows more controlled evolution and better evaluation of parameter adjustments.

4. Adjusting point mutate

This experiment increased the point mutation rate to 0.2 and mutation amount to 0.5 to test whether stronger genetic variation could drive better climbing behaviour under the same combined fitness function.

```
dna = genome.Genome.point_mutate(dna, rate=0.2, amount=0.5)
```

Performance Analysis



Fitness improved gradually across generations, with a breakthrough at generation 298, where a creature reached a maximum fitness of 0.62. The best mean fitness occurred at generation 219 (-0.60), but the average dropped to -1.11 by the final generation, indicating instability and inconsistent progress across the population. Behaviourally, creatures developed rounder bodies with fewer limbs and began rolling as a primary movement style.

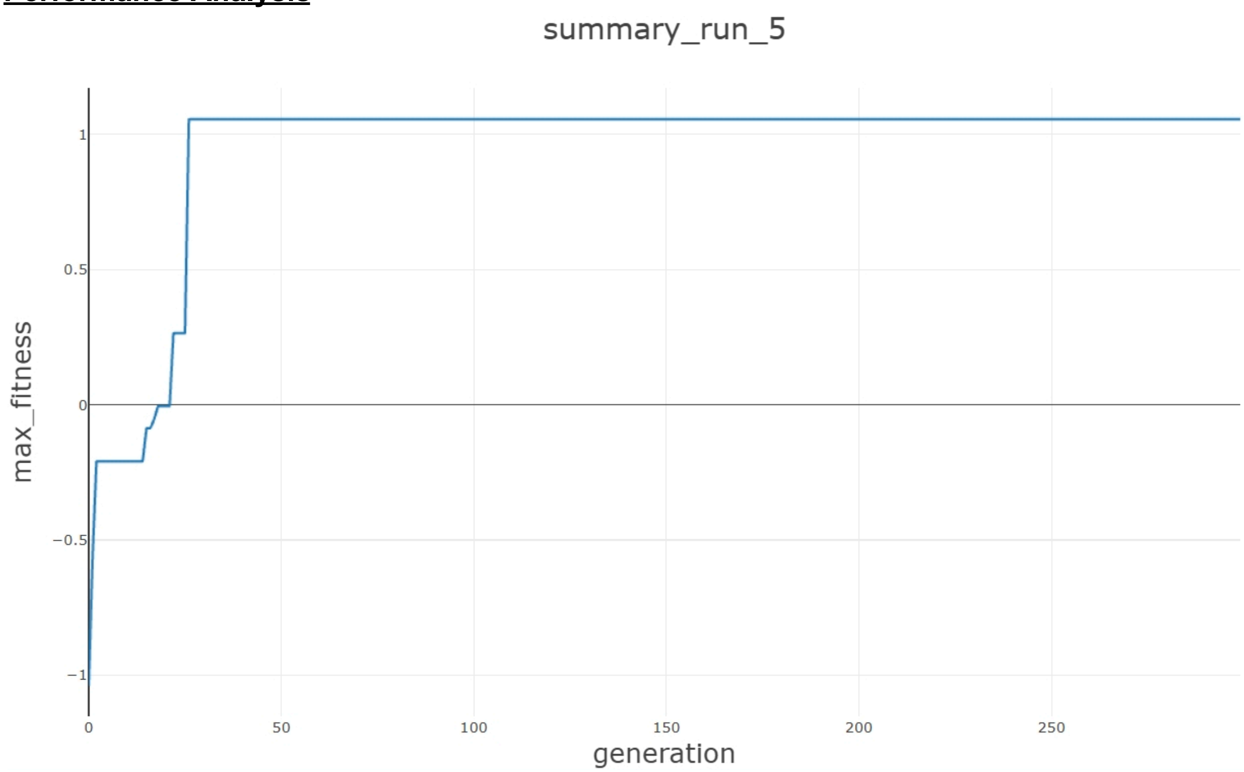
By observation, this experiment was less effective at climbing compared to previous runs. Hence, to reduce variability, improve evaluation consistency, and establish a clearer baseline for testing structural mutation changes, default mutation values will be restored in Experiment 5.

5. Adjusting shrink mutate and grow mutate

Assess whether adjusting the mutation pressure towards growth (by increasing grow_mutate) and reducing the frequency of shrinkage (shrink_mutate) can improve fitness and creature complexity over 300 generations.

```
dna = genome.Genome.shrink_mutate(dna, rate=0.2)
dna = genome.Genome.grow_mutate(dna, rate=0.15)
```

Performance Analysis



Fitness climbed quickly in early generations, with a sharp breakthrough at generation 26, where a creature reached the maximum fitness of 1.05. Population-level adaptation also improved, with the highest mean fitness of 0.14 at generation 161. However, this was not sustained as the final generation's mean fitness declined to -0.44, suggesting that while some individuals were effective, general performance was still volatile.

Structurally, creatures developed compact round forms that combined rolling and partial crawling. They remained smaller than those in earlier experiments,avoiding launching or vertical exploits.

Although mean fitness dropped toward the end, this experiment yielded the strongest individual without relying on inflated structures (such as in experiment 1 and 3). Therefore, this mutation setting will be retained as they strike a good balance between growth potential and control.

Results and Analysis

Experiment ID	1	2	3	4	5
Generations	500	300	300	300	300
Gene count	3	3	3	3	3
Population size	10	10	25	10	10
Point mutate	0.1	0.1	0.1	0.2	0.1
Shrink mutate	0.25	0.25	0.25	0.25	0.2
Grow mutate	0.1	0.1	0.1	0.1	0.15
control_freq	1	1	1	1	1
control_amp	1	1	1	1	1
Best fitness generation (elite_x.csv)	248	89	210	298	26
Best fitness score	75.05	0.22	5.11	0.62	1.05

While Experiment 1 recorded the highest raw fitness score, this was due to creatures exploiting the vertical-only fitness function by growing excessively tall. They inflate their z-axis giving them an unfair advantage, making this result misleading and not indicative of true climbing.

Experiments 2 to 5 used the combined progress function, which better rewarded actual movement. Although the maximum scores were lower, the creatures showed more genuine attempts to crawl, roll, and climb. Experiment 3 had a high fitness score (5.11) but was excluded from comparison due to large body sizes that may have inflated scores via vertical bias, similar to Experiment 1.

Mean fitness across all runs was mostly negative, likely because many creatures failed to move uphill and were penalized for downward or sideways movement.

Another key observation is that creatures did evolve some locomotion, but they often moved in random directions instead of heading toward the mountain. This lack of goal direction suggests a need for further refinement, such as sensor-based inputs or directional fitness shaping.

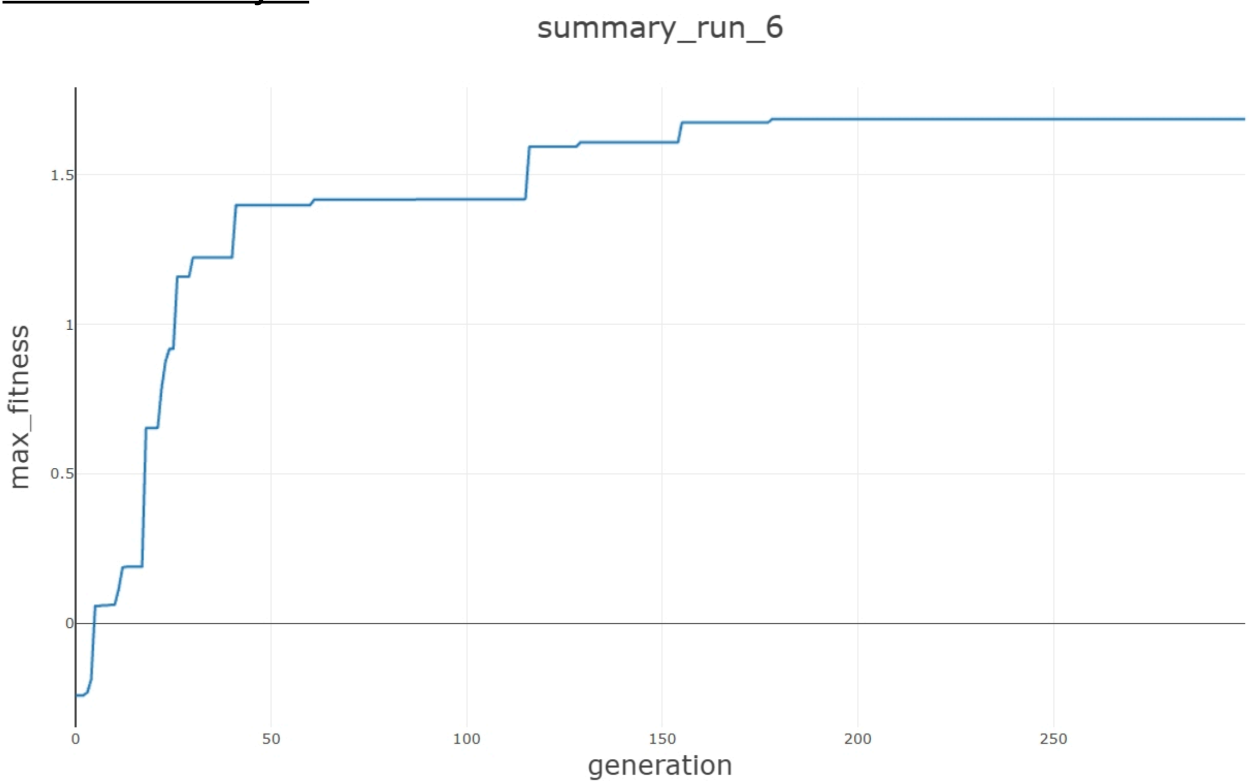
Advanced Experiments

6. Limit links to 10

This experiment introduced a cap of 10 links per creature to constrain structural complexity and observe whether simpler forms could still evolve effective climbing strategies. The goal was to reduce bloated morphologies and encourage more compact, functional designs.

```python
# exp 6: Limit number of links to 10
if len(exp_links) > 10:
    exp_links = exp_links[:10]
```

<u>Performance Analysis</u>



Fitness rose steadily across generations, with a breakthrough at generation 178, achieving a maximum fitness of 1.68. The mean fitness peaked at 0.14 (generation 113) before declining to -0.39 by the final generation, indicating reasonable adaptation despite some late-generation volatility.

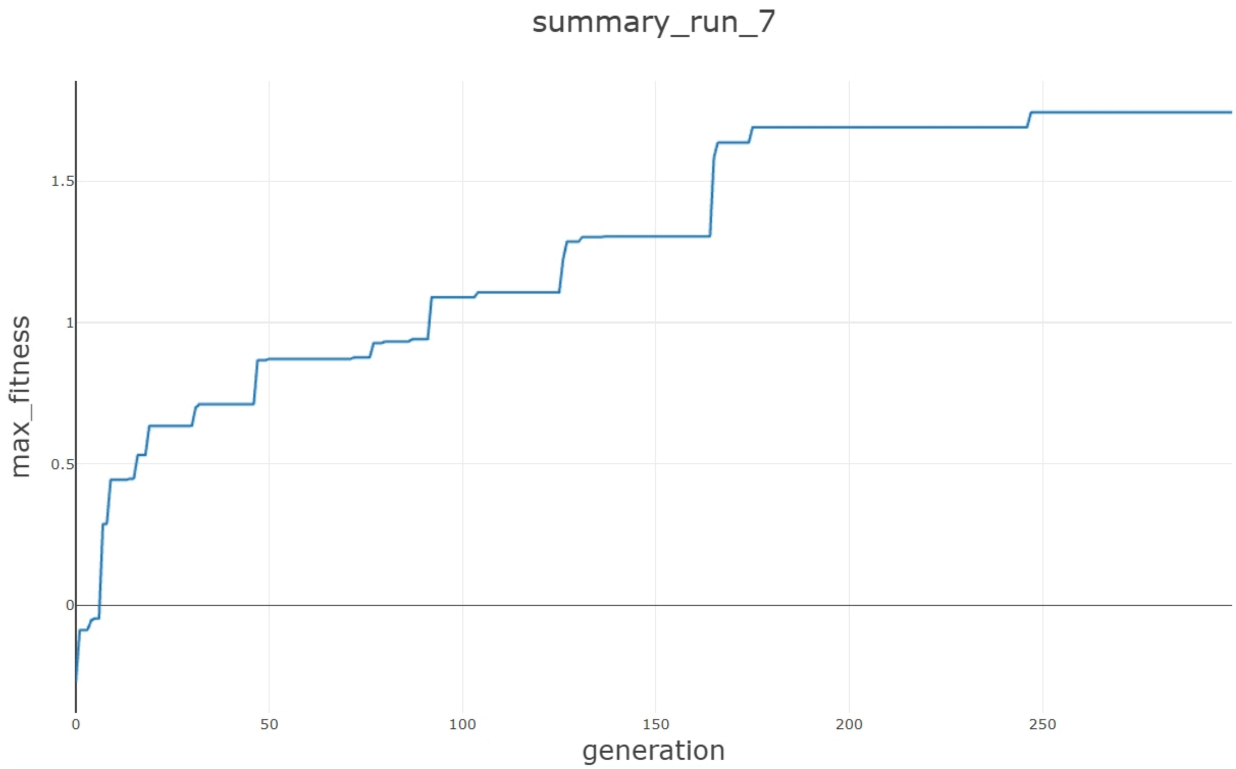
Visually, creatures evolved into tightly packed, efficient forms, stable builds and minimal excess limbs. Movement was achieved through controlled rolling and occasional crawling, aided by more deliberate limb positioning compared to earlier experiments. The link cap appeared to prevent chaotic overgrowth, improving evaluation consistency and reducing erratic behaviors. This result suggests that limiting complexity does not hinder fitness, and may in fact focus the evolutionary process on refining useful traits. However, the cap was removed in Experiment 7 to isolate whether joint limitations alone could still prevent excessive growth. This allows clearer analysis of whether joint count alone is the key factor driving fitness and stability.

7. Limit joints to 5

This experiment restricted the number of joints to 5 (6 links including the root), aiming to promote simpler morphologies while maintaining enough complexity for movement. The goal was to observe whether reducing joint count alone, without limiting overall link number, could still yield efficient climbers.

```
# exp 7: Limit number of joints (joints = links - 1)
if len(exp_links) - 1 > 5:
    exp_links = exp_links[:6]  # 1 root + 5 links = 5 joints
```

Performance analysis



Fitness improved steadily, with the best individual reaching 1.76 at generation 264. The population mean also improved, peaking at 0.43 in generation 259 and ending at 0.27, which is the highest population-level consistency so far. Early generations displayed compact, bulbous bodies, while later creatures evolved more appendage-like extensions, showing locomotion strategies that combined rolling with limb flailing or crawling.

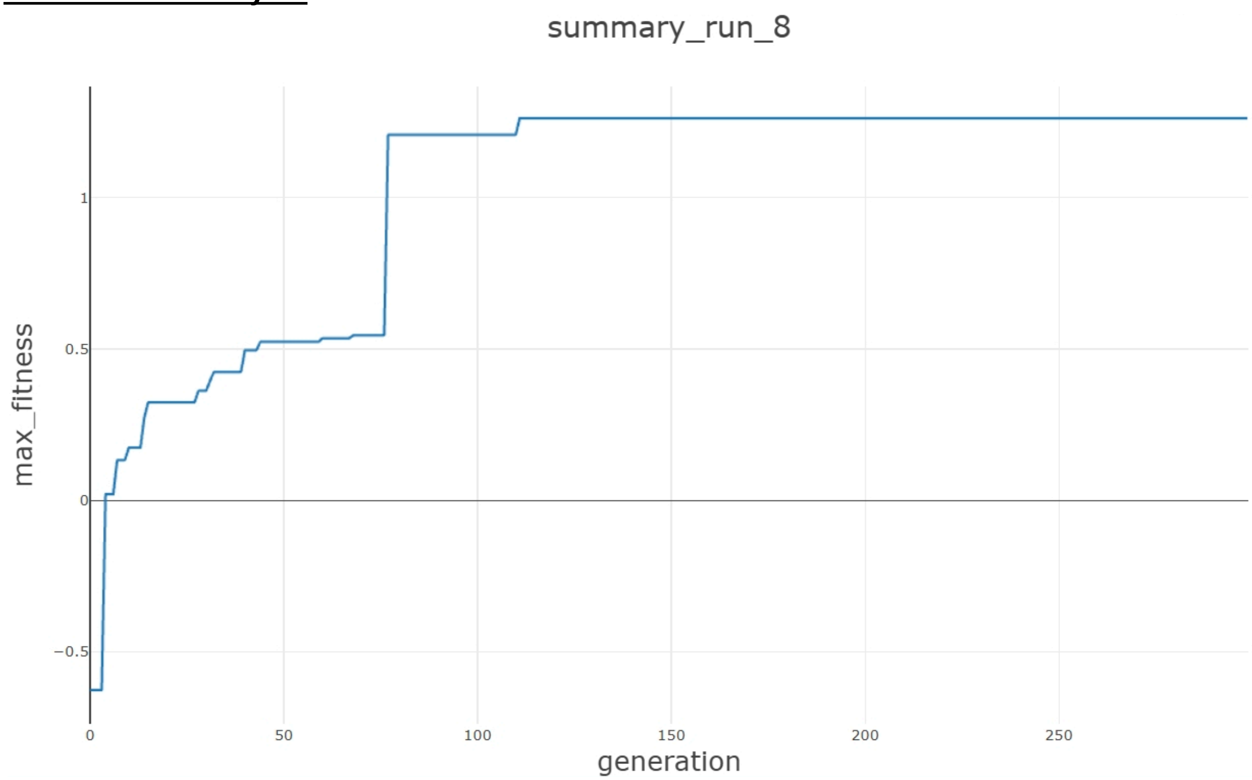
While the 5-joint cap constrained excessive structural sprawl, it still allowed sufficient variety for functional adaptation without triggering physics failures or overcomplexity. Based on this, limiting joints offers better control over structural growth while preserving behavioral potential. This balance makes joint capping a preferable strategy for subsequent runs.

8. Fixed frequency = 0.5

This experiment fixed all joint motor frequencies to 0.5, removing frequency as a mutable parameter to evaluate whether simplified control would still allow effective locomotion to evolve.

```
FIXED_FREQ = 0.5
m = Motor(l.control_waveform, l.control_amp,  FIXED_FREQ)
```

Performance analysis



Despite this constraint, creatures quickly achieved competitive performance. A breakthrough occurred at generation 111, where maximum fitness jumped to 1.26, which remained stable until the final generation. While the score was slightly lower than Experiments 6 and 7, this run showed stable convergence and minimal regression in later generations. The best mean fitness was recorded at generation 189 (0.15), with a final mean of -0.08.

Structurally, creatures evolved small, symmetric forms that crawled. While motion types varied, limiting frequency did not severely hinder behaviour. Instead, it reduced the search space, possibly helping evolution focus on refining movement patterns rather than oscillating wildly between frequency extremes.

This suggests that a fixed mid-range frequency is sufficient for enabling effective control, especially when paired with adaptive morphologies. It may be useful in future work to compare fixed vs. evolvable frequencies across multiple ranges.

Results and Analysis

Experiment ID	6	7	8
Generations	300	300	300
Gene count	3	3	3
Population size	10	10	10
Point mutate	0.1	0.1	0.1
Shrink mutate	0.2	0.2	0.2
Grow mutate	0.15	0.15	0.15
Max links	10	Default	Default
Max joints	Default	5	5
control_freq	1	1	0.5
control_amp	1	1	1
Best fitness generation (elite_x.csv)	178	264	111
Best fitness score	1.68	1.74	1.26

Experiments 6-8 tested structural and motor encoding constraints. Limiting links (Experiment 6) reduced bloat and improved stability, while limiting joints (Experiment 7) offered better control over functional complexity, producing the highest fitness overall. In contrast, fixing frequency (Experiment 8) removed motor adaptability and led to weaker performance.

Though the hypothesis in Experiment 8 was not supported, it revealed the importance of motor flexibility. Overall, joint limits proved most effective, offering a strong balance of control and expressiveness for climbing behaviour.

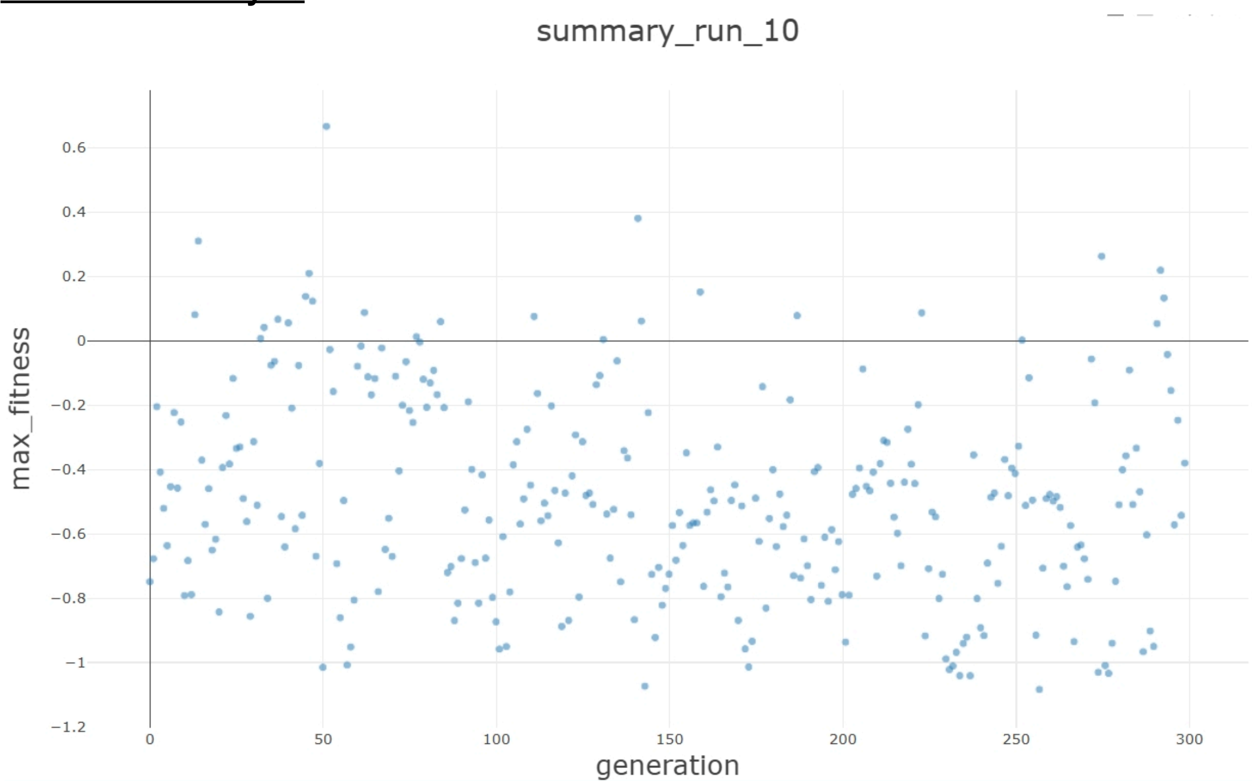
Extension

9. Noisy Terrain

As an extension, I evaluated the adaptability of creatures evolved in Experiment 7 by testing them on procedurally generated noisy terrain (Perlin noise) and rocks, with a higher spawn height (z=3) to avoid launching.

```
# experiment 9 extension
ga_mountain.make_rocks(arena_size=20)
ga_mountain.mountain = p.loadURDF("shapes/gaussian_pyramid_noisy.urdf", mountain_position, mountain_orientation, useFixedBase=1)
```

Performance analysis



Fitness fluctuated across generations with no consistent upward trend, and the best fitness peaked at generation 51 (0.67). Mean fitness remained negative, suggesting most creatures failed to make meaningful progress. Unlike smoother environments, the uneven terrain punished inefficient movement, showing limitations in prior evolved strategies.

Results and Analysis

Experiment ID	9
Generation	300
Gene count	3
Population size	10
Point mutate	0.1
Shrink mutate	0.2
Grow mutate	0.15
Max links	Default
Max joints	5
control_freq	1
control_amp	1
Best fitness generation (elite_x.csv)	51
Best fitness score	0.67

While some individuals managed brief motion, most failed to adapt to the irregular topology, often remaining stuck or rolling in unproductive directions. This suggests that creatures optimised for smooth terrain do not generalise well to uneven environments as their evolved strategies lack robustness.

In summary, the experiment highlights the need for terrain-aware design or sensory control mechanisms if general adaptability is a goal in future evolutionary setups.

Conclusion

In conclusion, this project explored how evolutionary algorithms shape creature behaviour across varying conditions. While I was not successful in evolving creatures that could consistently climb the mountain, some configurations led to brief progress. The fitness function's dependence on z-axis height introduced unfair advantages when spawn positions caused launching, and the lack of terrain awareness meant creatures frequently moved in the wrong direction. These limitations suggest that meaningful climbing may require better fitness shaping, sensory feedback, or explicit goal alignment in the future.

(1935 words)