1. Optimise any genetic algorithm problem (any function estimation) with RL.

**K-armed Bandit Problem**:

* In the K-armed bandit problem, each arm can represent a different individual in the population, and the reward as the fitness of that individual.
* **Selection Strategy**: Use an RL-based strategy, such as the epsilon-greedy method or upper confidence bound (UCB) to decide which individuals to select based on their past fitness.
* With probability epsilon, select a random individual (exploration).
* With probability 1−ϵ, select the individual with the highest average fitness (exploitation).
* Select individuals based on the K-armed bandit strategy for reproduction and perform the other operations involved in genetic algorithm such as crossover or mutation
* Replace the least fit individuals in the population with the fittest offspring.
* The bandit approach can focus more on exploring diverse solutions when necessary or use the best solutions more aggressively when close to convergence.