

# **Analysis Report: Hangman Q-Learning Agent with HMM Integration**

## **1. Key Observations**

### **Challenges**

Developing a reinforcement learning agent to play Hangman presented several non-trivial challenges: Sparse and delayed rewards made it difficult for the agent to assign credit effectively. The large and variable state space also reduced sample efficiency. Balancing exploration and exploitation was tricky, as the agent could easily overfit or explore too randomly. Finally, integrating probabilistic priors from the HMM into the Q-learning framework required careful weighting.

### **Insights Gained**

Reward shaping proved essential to learning stability. A curriculum approach that gradually increased word length improved generalization. The hybrid model—using both HMM priors and Q-learning—consistently outperformed either component alone. A gradual epsilon decay maintained exploration over extended training.

## **2. Strategies**

### **HMM Design Choices**

The HMM modeled English letter frequencies and conditional dependencies. Each hidden state represented possible letter emissions, allowing the agent to leverage linguistic patterns. HMM probabilities were integrated into Q-values using a dynamic weighting factor that decreased as exploration decayed, allowing the agent to rely more on learned Q-values over time.

### **Reinforcement Learning State Design**

The state was represented as a compressed feature vector containing revealed letter ratio, guessed ratio, average HMM probability, and normalized word length. This simplified representation improved generalization while keeping the Q-table size manageable.

### **Reward Design**

Reward shaping included  $+3.0 + 2.0 \times \text{progress}$  for correct guesses,  $-1.5$  for wrong guesses,  $+15$  for a win, and  $-5$  for failure. This structure provided dense feedback, encouraging steady progress and rewarding complete word recovery.

## **3. Exploration vs. Exploitation**

An epsilon-greedy policy managed the balance between exploration and exploitation. The agent began with high exploration ( $\epsilon \approx 1.0$ ) and decayed gradually toward 0.1. Early training focused on exploring simple patterns, while later phases emphasized exploiting learned strategies. Curriculum learning was also applied—shorter words first, longer words later—to progressively build difficulty.

## 4. Future Improvements

If given more time, improvements could include: 1) Using a neural Q-network for function approximation, 2) Enhancing feature encoding with linguistic and positional cues, 3) Dynamic reward scaling to encourage rare correct guesses, 4) Prioritized experience replay for stability, 5) Integrating transformer-based priors instead of HMMs, and 6) Adding analytics for deeper insight into learning dynamics.

Overall, the hybrid Q-learning + HMM approach demonstrated measurable learning progress, achieving a success rate of around 4–8%. While modest, this shows promise for integrating probabilistic and reinforcement-based reasoning in natural language tasks.