Rapport TP3

1. Agent SARSA (On-Policy)

- Apprentissage "on-policy" : mise à jour basée sur l'action réellement choisie
- Paramètres : taux d'apprentissage, facteur de discount (gamma)
- Plus conservateur dans ses choix d'actions

2. Agent Q-Learning Classique (Off-Policy)

- Apprentissage "off-policy" : mise à jour basée sur la meilleure action possible
- Paramètres :
    - Learning rate (α) : contrôle la vitesse d'ajustement des Q-values
    - Epsilon (ε) : balance exploration/exploitation
    - Gamma (γ) : pondération des récompenses futures
- Stratégie ε-greedy pour l'exploration

3. Agent Q-Learning avec Epsilon Scheduling

- Extension du Q-Learning avec décroissance progressive de l'exploration
- Paramètres additionnels :
    - epsilon_start : valeur initiale élevée pour favoriser l'exploration
    - epsilon_end : valeur minimale pour garantir une exploitation stable
    - epsilon_decay_steps : durée de la transition
- Adaptation dynamique de l'exploration au cours de l'apprentissage

Résultats Expérimentaux

```
Total training time for QLearningArgent: 1.7633159160614014 seconds, Reward mean: 1.37
Total training time for QLearningAgentEpsScheduling: 1.1672828197479248 seconds, Reward mean : 4.75
Total training time for SarsaAgent: 1.27296781539917 seconds, Reward mean : 7.8
```

Analyse Comparative:

1. **SARSA** : Meilleure performance globale (récompense moyenne : 7.80), temps d'entraînement modéré, convergence stable vers une politique efficace
2. **Q-Learning avec ε Scheduling** : Performance intermédiaire (récompense moyenne : 4.75), temps d'entraînement le plus court, balance efficace entre exploration et exploitation
3. **Q-Learning Classique** : Performance la plus faible (récompense moyenne : 1.37), temps d'entraînement le plus long, difficulté à converger vers une politique optimale

L'agent SARSA s'est révélé le plus efficace pour cette tâche spécifique, démontrant une meilleure capacité à apprendre et à maximiser les récompenses. L'ajout du scheduling d'epsilon au Q-Learning améliore significativement ses performances par rapport à la version classique, soulignant l'importance d'une stratégie d'exploration adaptative. Les résultats suggèrent que l'approche on-policy de SARSA est particulièrement adaptée à l'environnement Taxi-v3, probablement en raison de sa nature plus conservative dans la sélection des actions