

Deep Reinforcement Learning

PIERROT Nicolas

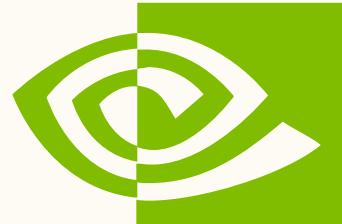
RENARD Adrien

TOUZART Antoine

Intervenant: VIDAL Nicolas

Technologies Utilisées

 PyTorch

 nVIDIA
CUDA

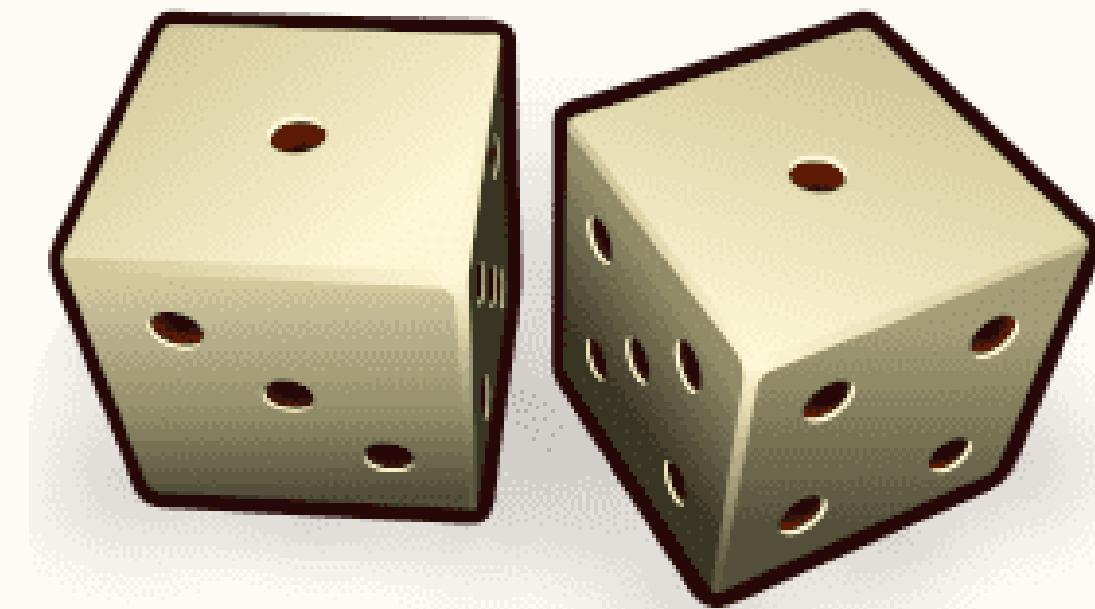
Architecture du Projet

Project Architecture



Environnements Implémentés

- Line World
- Grid World
- Tic Tac Toe
- Farkle



Modèles Implémentés

Random	On a
TabularQLearning (quand possible)	
DeepQLearning	
DoubleDeepQLearning	
DoubleDeepQLearningWithExperienceReplay	
DoubleDeepQLearningWithPrioritizedExperienceReplay	
REINFORCE	
REINFORCE with mean baseline	
REINFORCE with Baseline Learned by a Critic	
PPO A2C style	
RandomRollout	
Monte Carlo Tree Search (UCT)	
Expert Apprentice	
Alpha Zero	
MuZero	
MuZero stochastique	

Modèles Implémentés avec Enregistrement

Model	linew	gridworld	tictactoe	farkle
DQN	Monkey	Monkey	Monkey	Monkey
DDQN	Monkey	Monkey	Monkey	Monkey
DDQN avec Experience Replay	Monkey	Monkey	Monkey	Monkey
DDQN avec Prioritized Experience Replay	Monkey	Monkey	Monkey	Monkey
MCTS (Monte Carlo Tree Search)	Monkey	Monkey	Monkey	Monkey
PPO (Proximal Policy Optimization)	Monkey	Monkey	Monkey	Monkey
REINFORCE	Monkey	Monkey	Monkey	Monkey
REINFORCE avec Critic (Baseline apprise)	Monkey	Monkey	Monkey	Monkey
REINFORCE avec Mean Baseline	Monkey	Monkey	Monkey	Monkey
Tabular Q-Learning	Bear	Bear	Bear	Bear

Interface CLI Pretty Print

Choose an environment to play	
Key	Environment
1	LineWorld
2	GridWorld
3	TicTacToe
4	Farkle

Enter your choice (1-4): 1

Instructions
Enter 'Q' for left, 'D' for right to move.

Playing LineWorld



State vector: [1 0 0 0 0 0 0 6]

Actions
Available actions: [1]

Enter direction (Q: left, D: right):

Your move:

Interface CLI Pretty Print

Enter your choice (1-4): 3

Instructions —
Enter a number from 0 to 8 to place your mark.
You will be randomly assigned as X (Player 1) or O (Player 2).
The board is numbered as follows:

0	1	2
3	4	5
6	7	8

Playing TicTacToe

You are Player 2

0	1	2
3	4	5
6	7	8

State vector: [0 0 0 0 0 0 0 0]

AI Player 1 chose action: 1

Action vector: [0 1 0 0 0 0 0 0]

0	X	2
3	4	5
6	7	8

State vector: [0 1 0 0 0 0 0 0]

Actions —
Available actions: [0 2 3 4 5 6 7 8]

Enter your choice (1-4): 4

Instructions —
Choose dice to keep by entering their positions (1-6).
For example, '136' keeps the 1st, 3rd, and 6th dice.
Enter '0' to bank your points.
You will be randomly assigned as Player 1 or Player 2.
Aim to score points without 'farkling' (rolling no scoring dice).

Playing Farkle

You are Player 1

Current Dice —
Dice not visible - choose to roll or bank

Turn Score — Current turn score: 0

Player 1 Score — Player 1 score: 0

Player 2 Score — Player 2 score: 0

Current Player — Current player: Player 1

Do you want to:
-1: Bank your points (current turn score: 0)
0: Roll the dice
Enter your choice (-1 or 0): 0

Current Dice —

Turn Score — Current turn score: 0

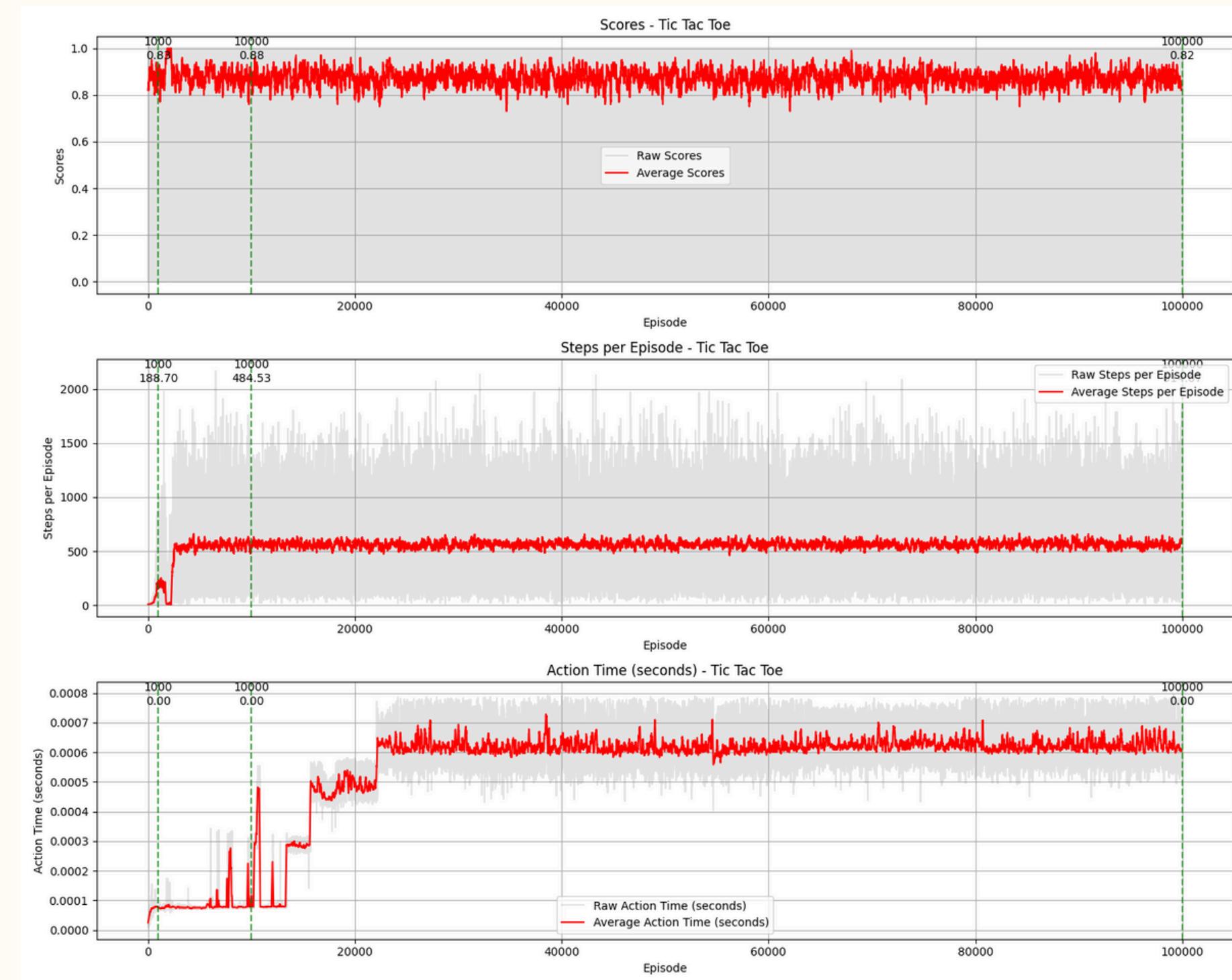
Player 1 Score — Player 1 score: 0

Player 2 Score — Player 2 score: 0

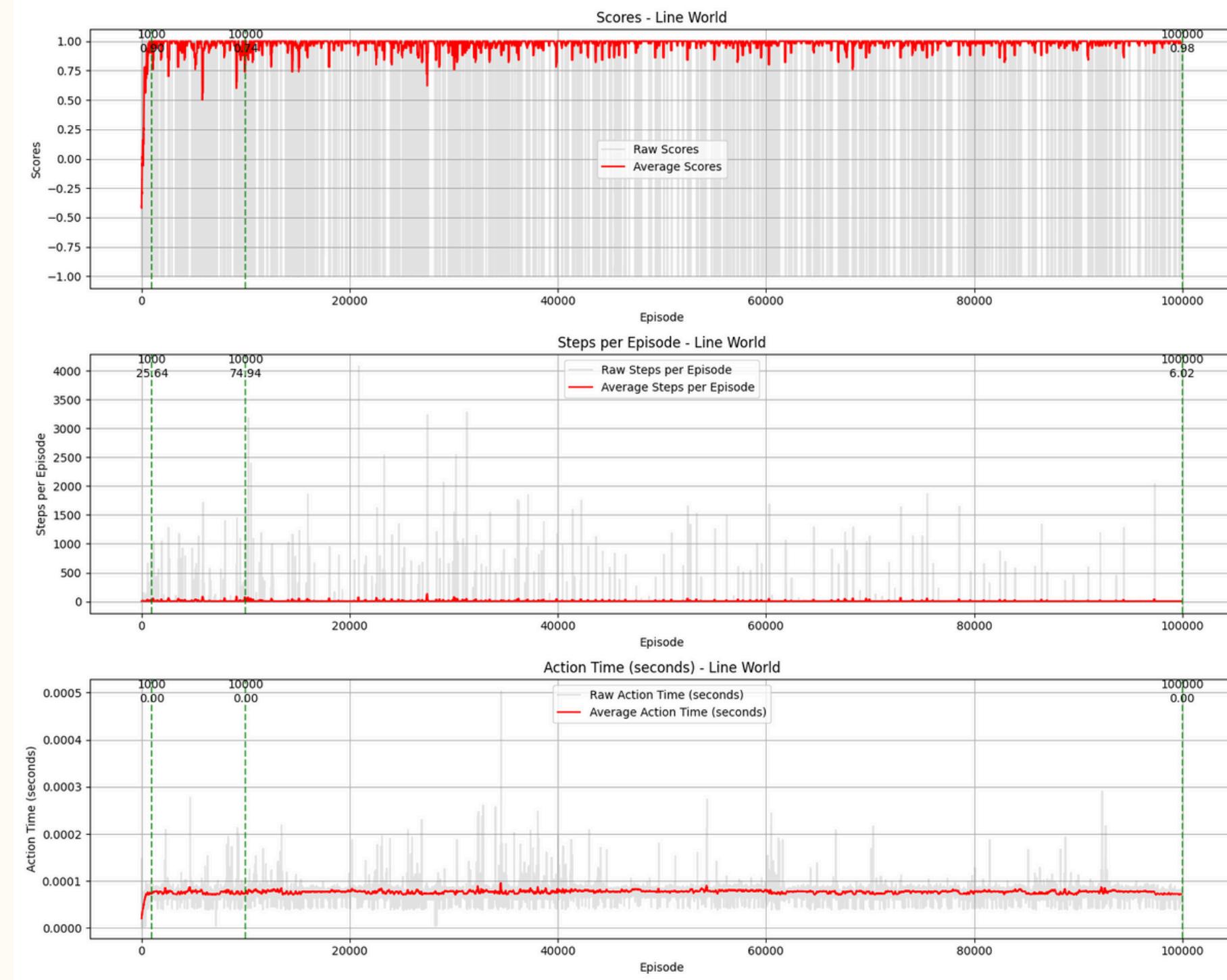
Current Player — Current player: Player 1

Choose dice to keep (1-6) or bank points (-1)
For example: '136' keeps the 1st, 3rd, and 6th dice
Enter your choice: ■

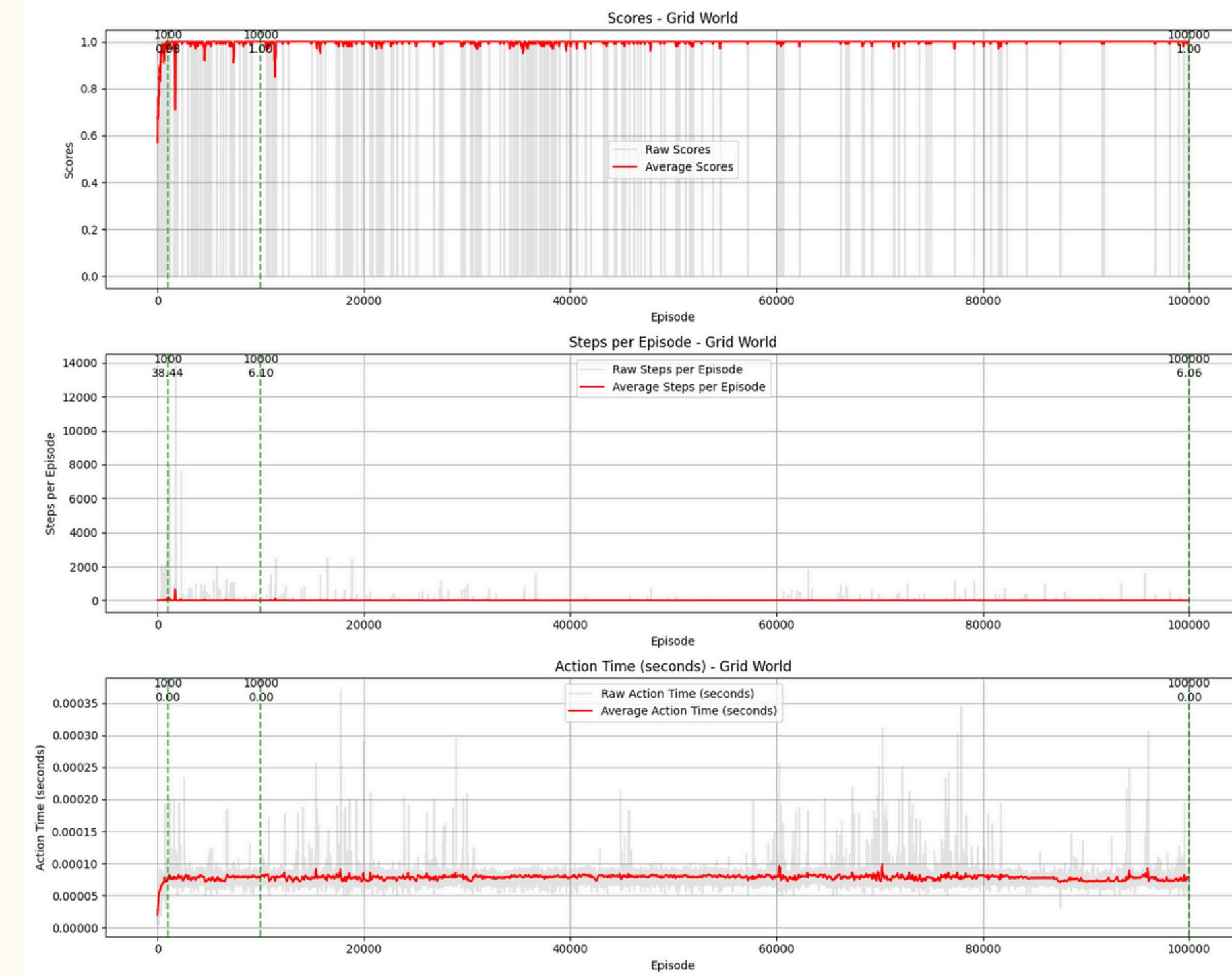
Résultats DDQN



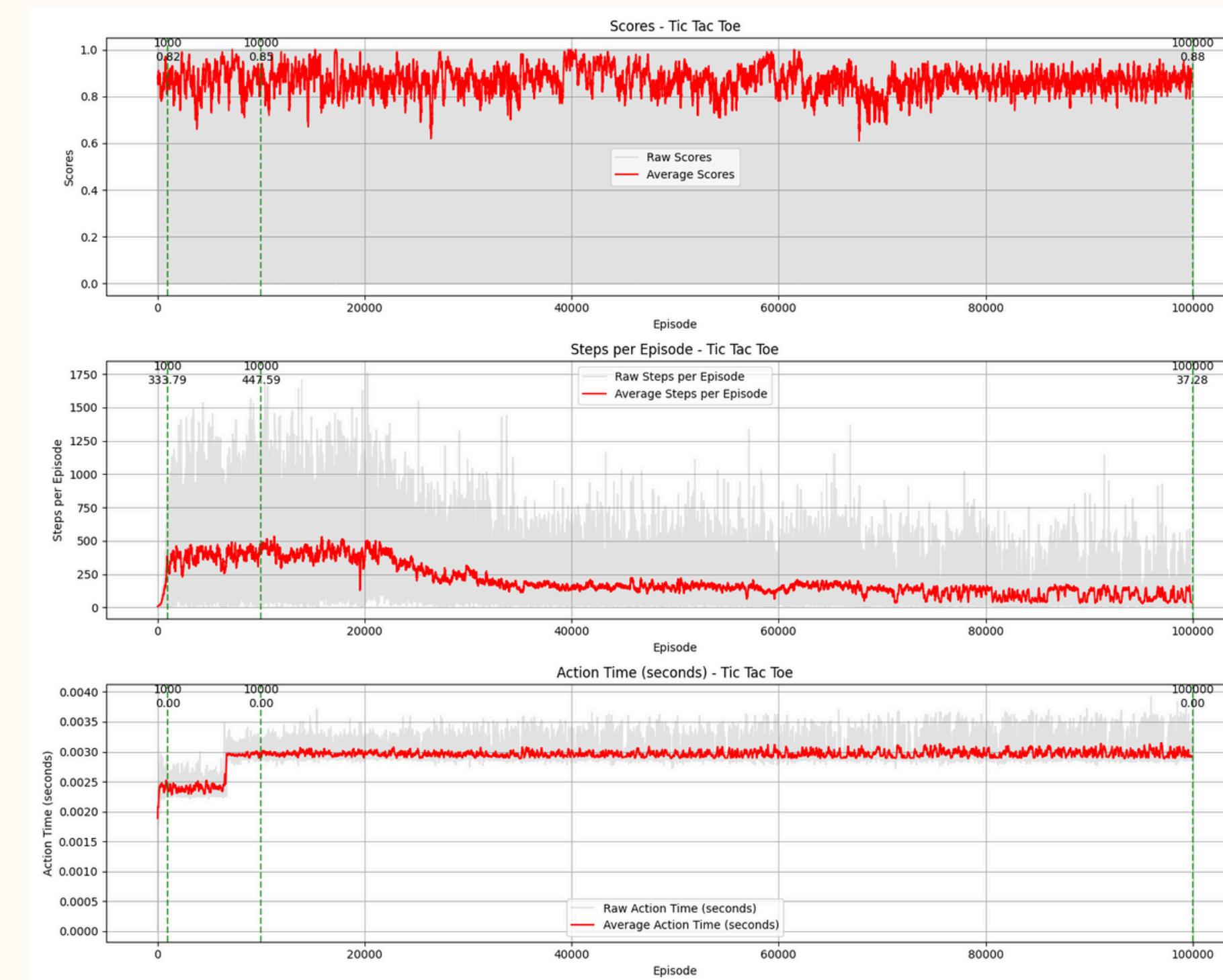
Résultats DDQN



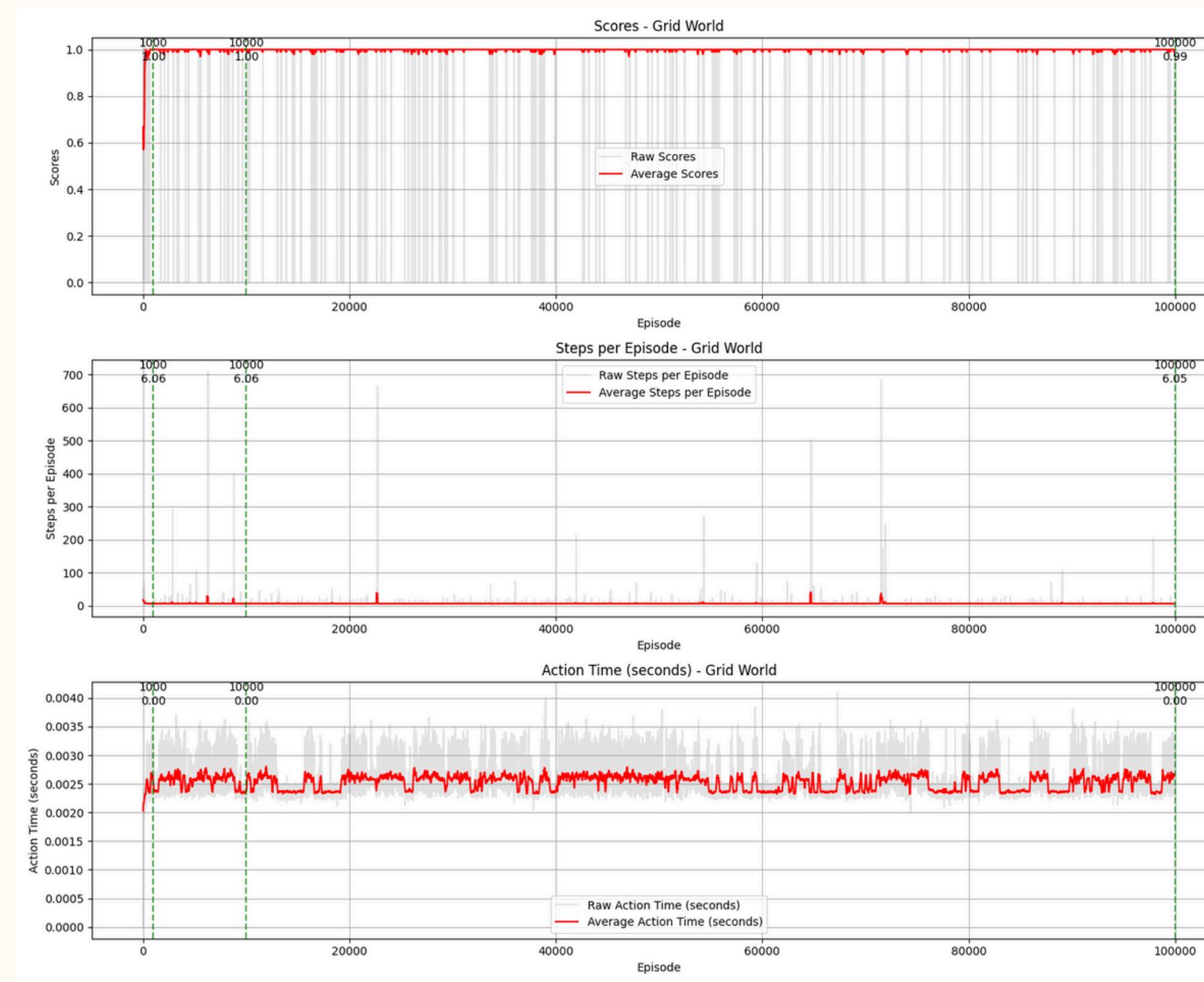
Résultats DDQN



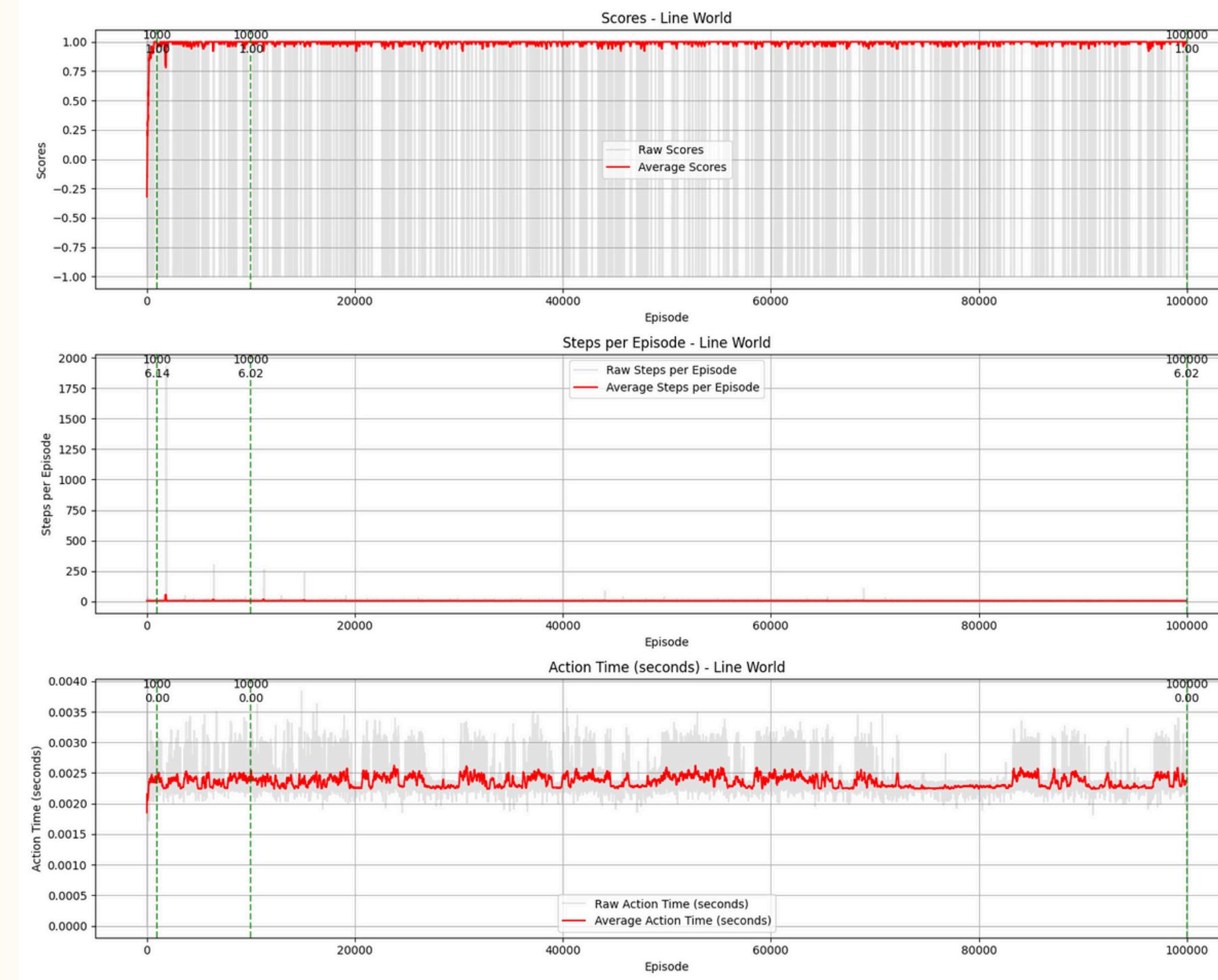
Résultats DDQN exp replay



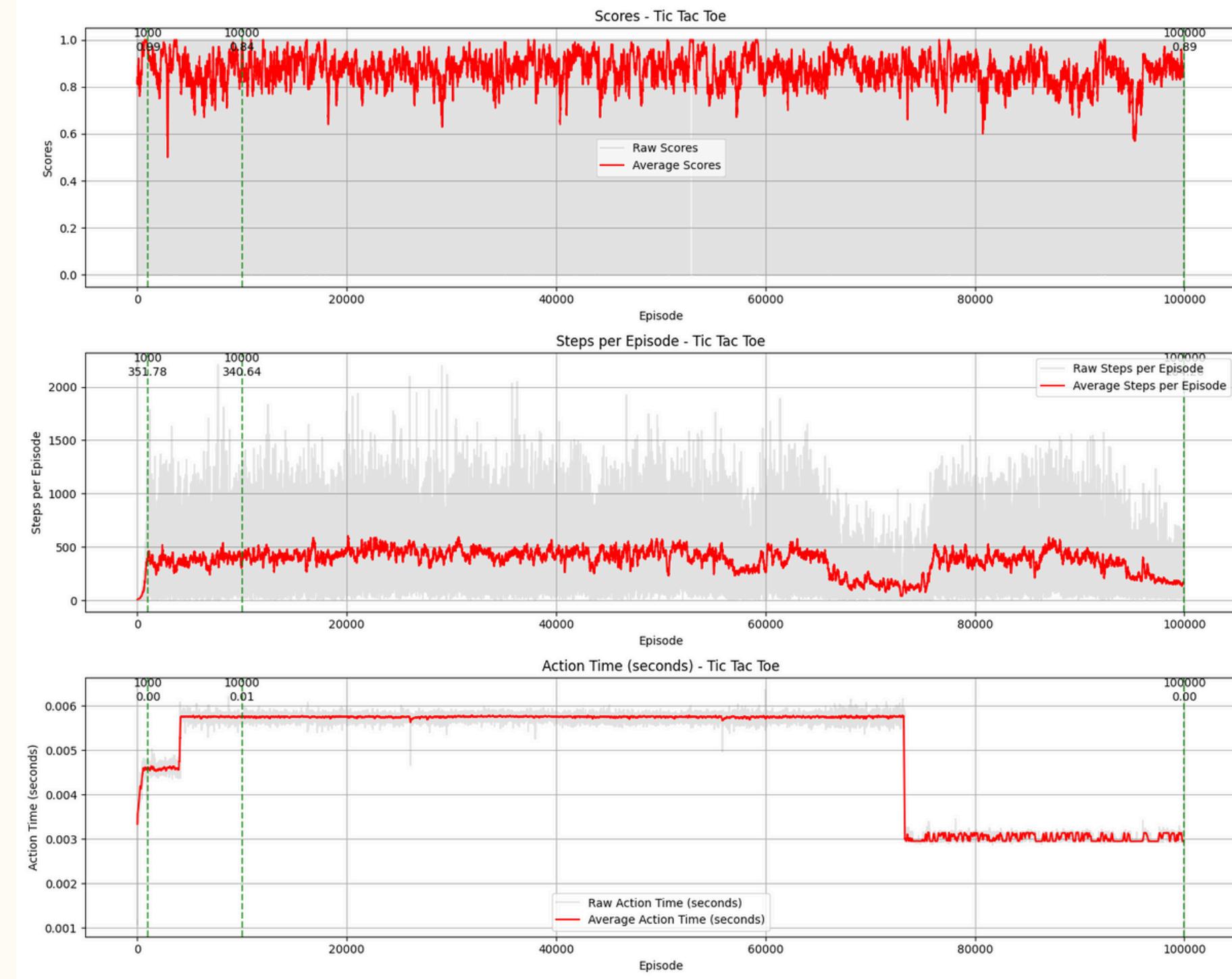
Résultats DDQN exp replay



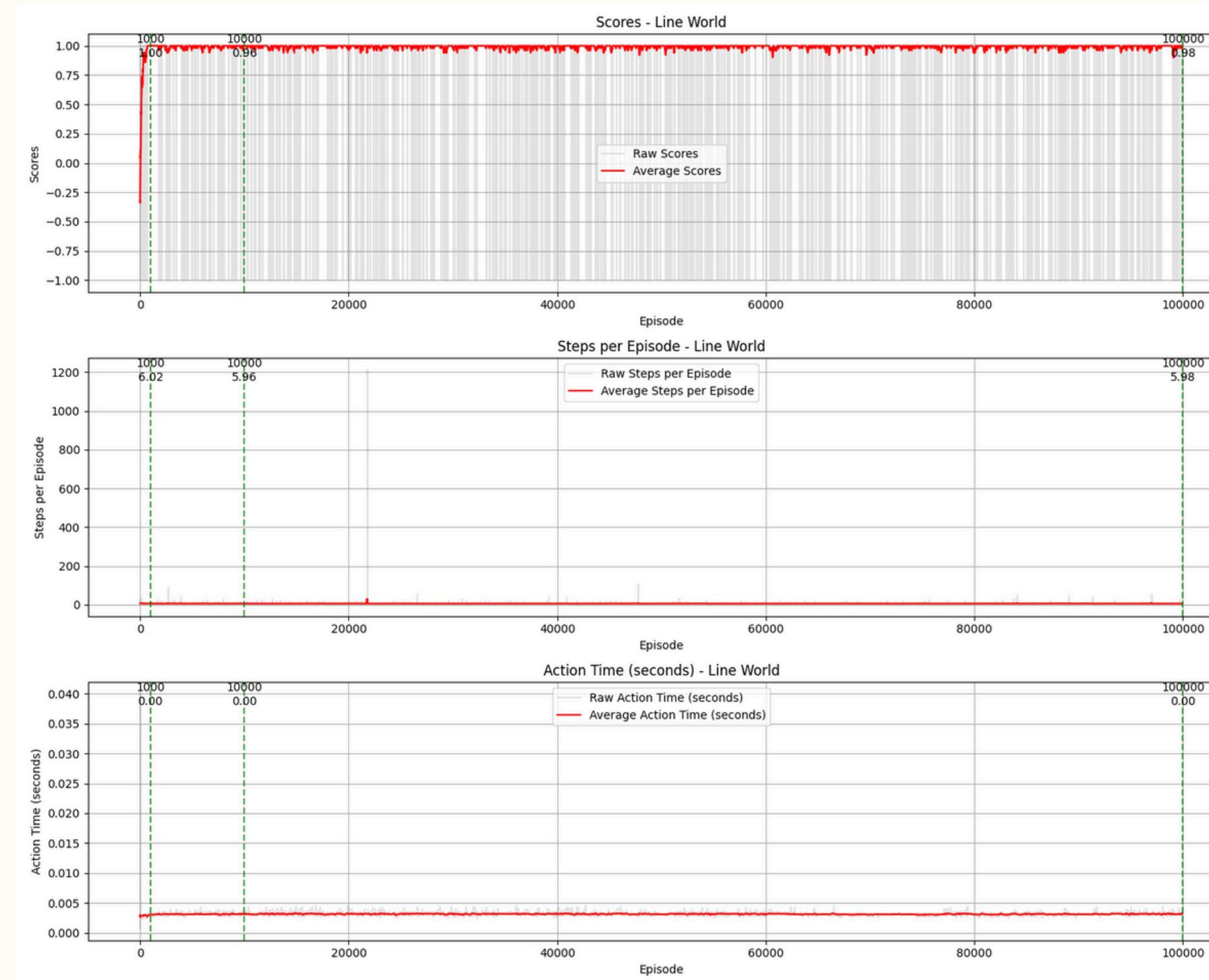
Résultats DDQN exp replay



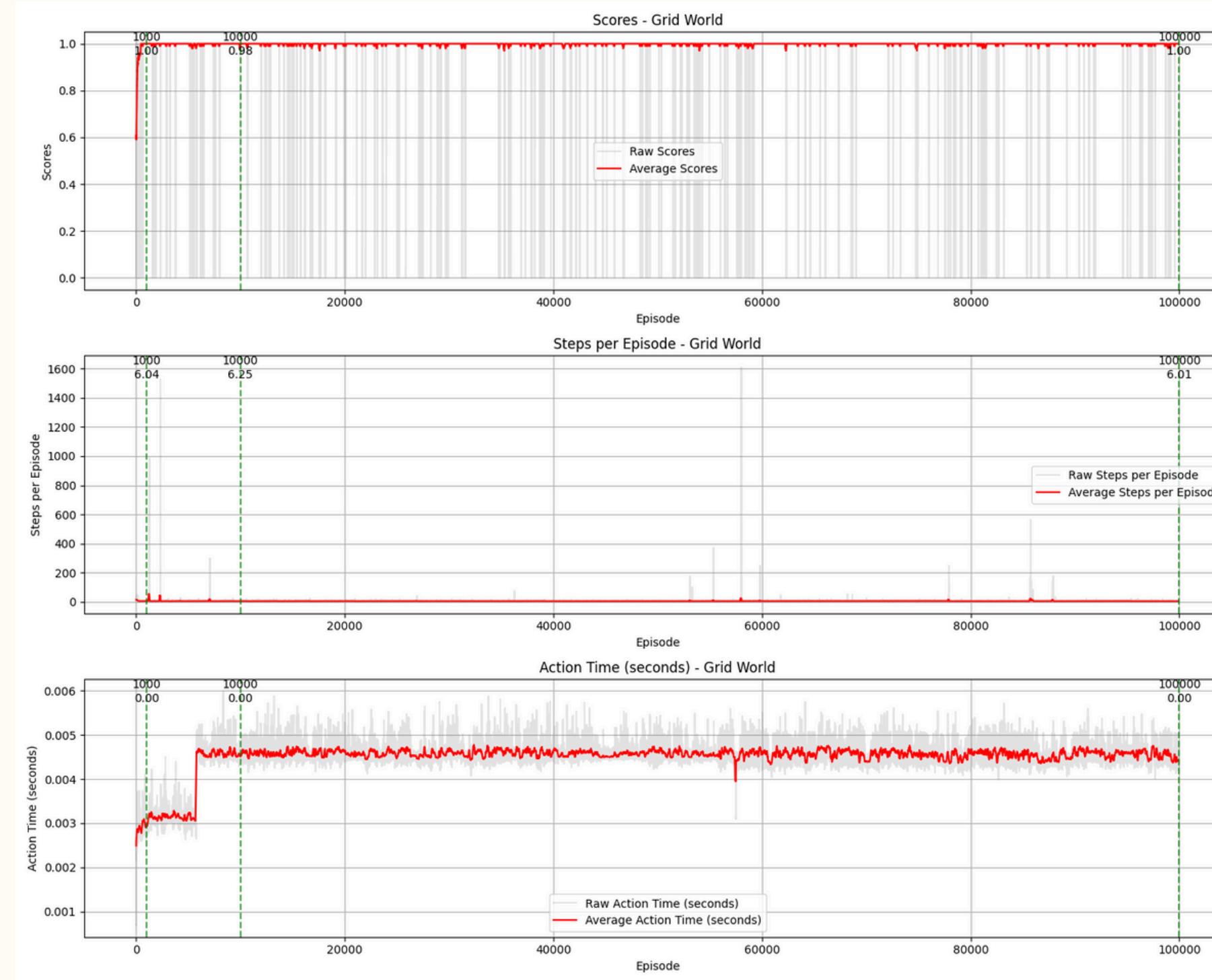
Résultats DDQN pri exp replay



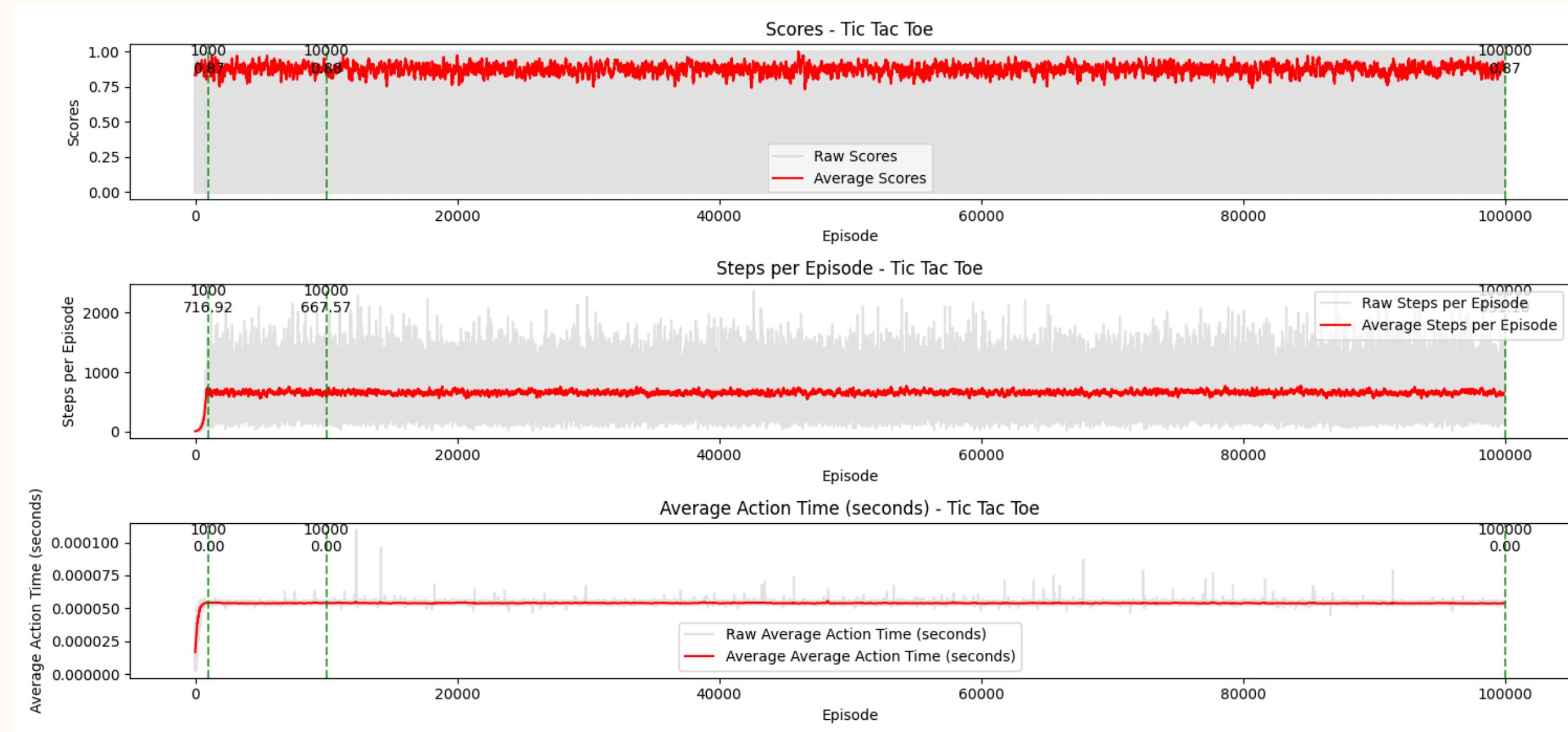
Résultats DDQN pri exp replay



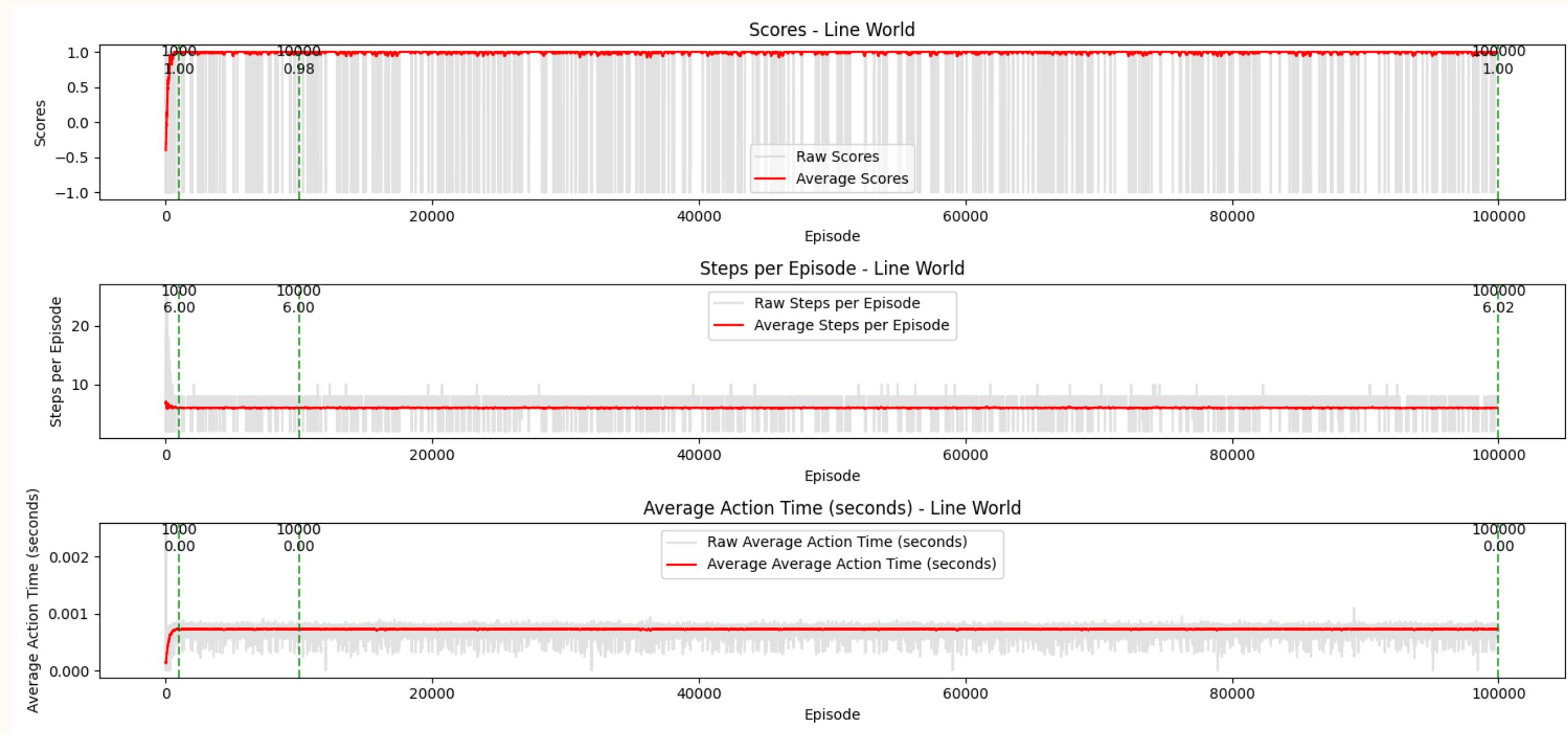
Résultats DDQN pri exp replay



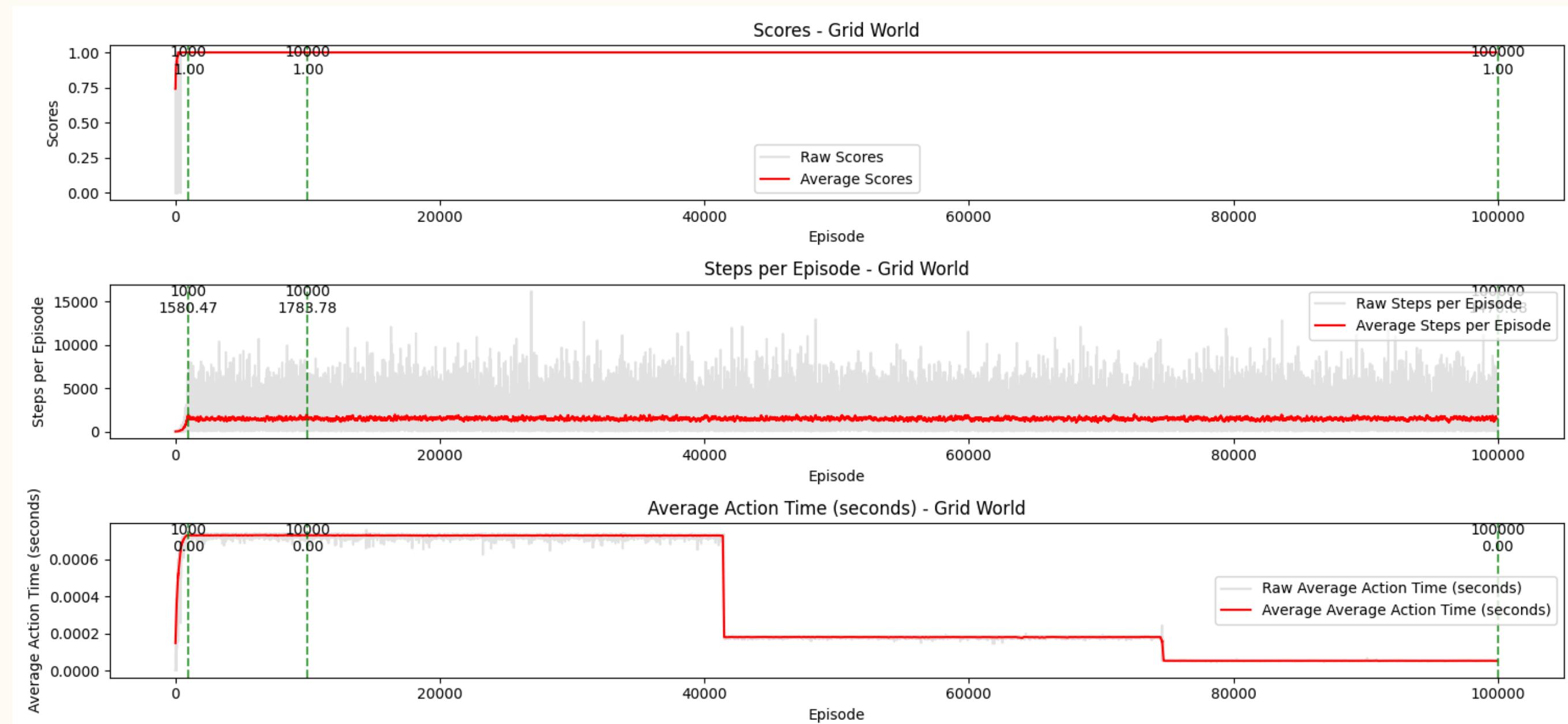
Résultats DQN



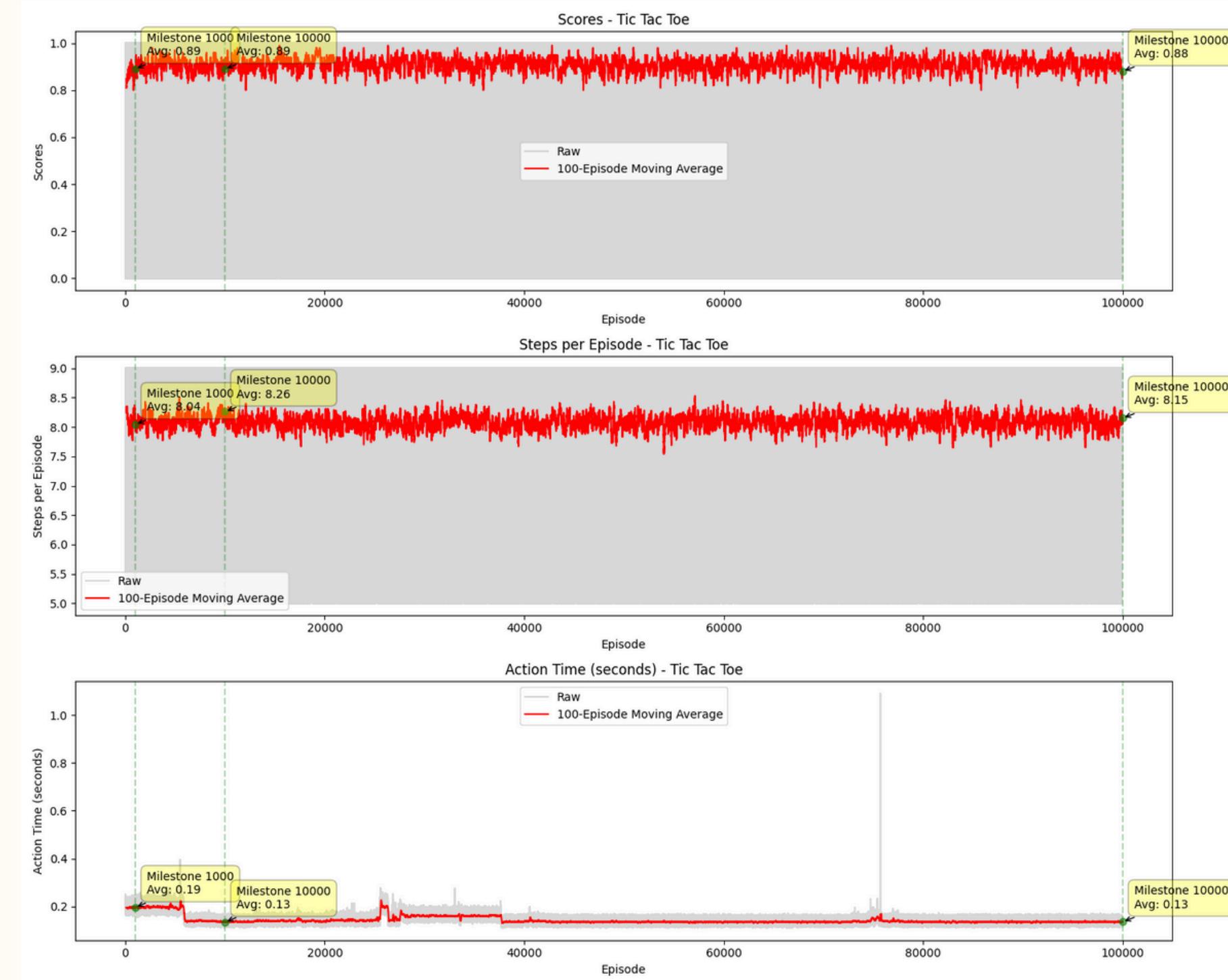
Résultats DQN



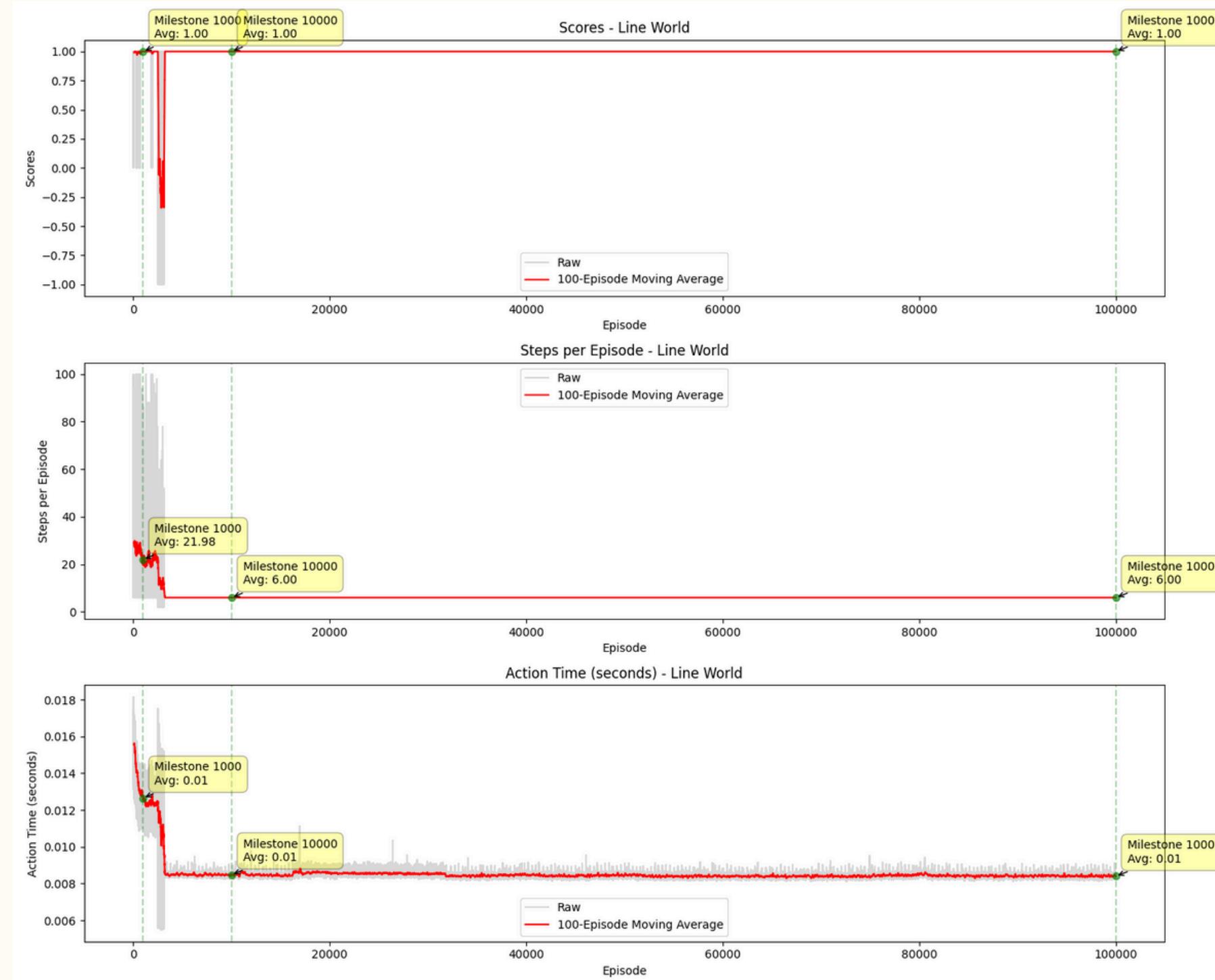
Résultats DQN



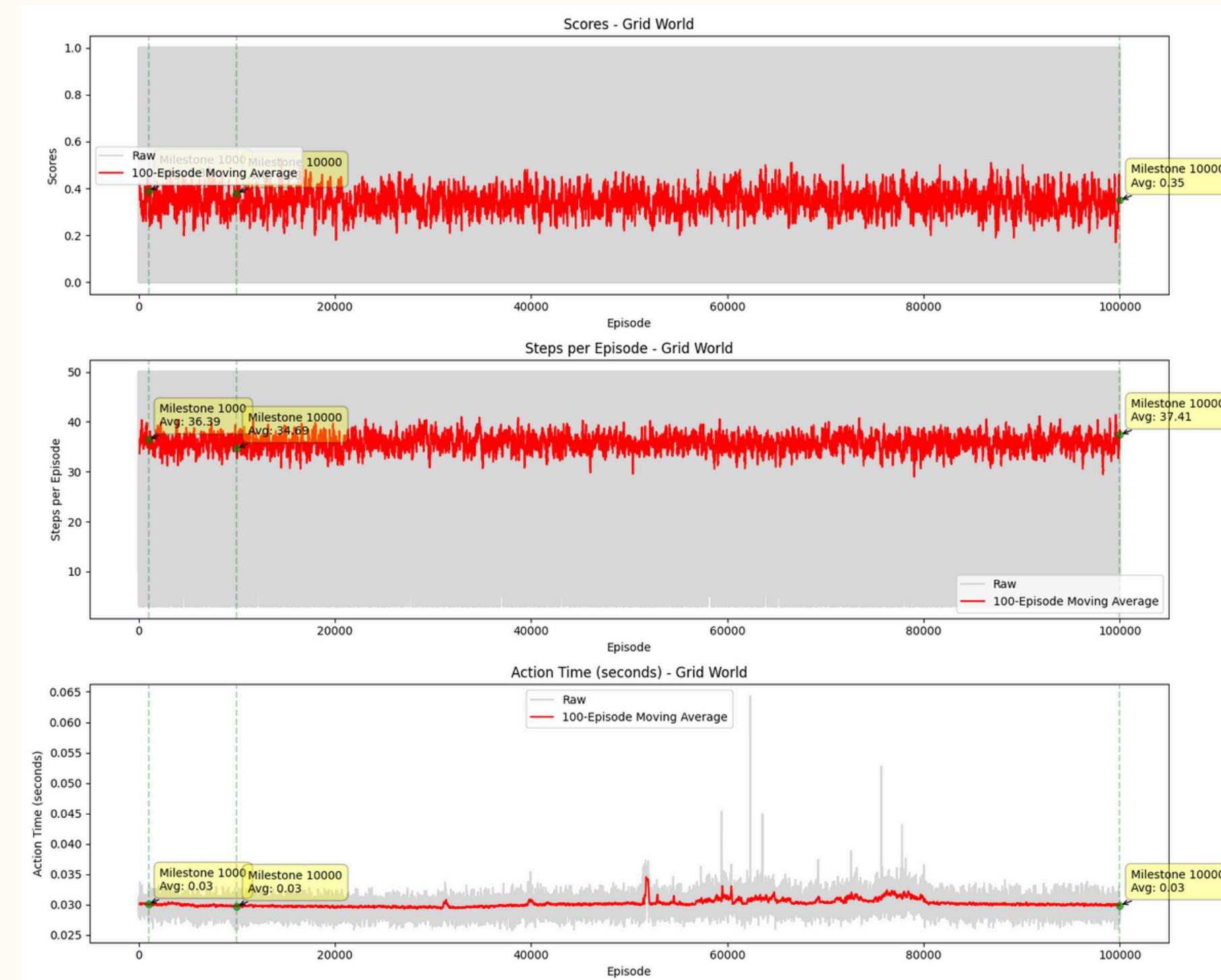
Résultats MCTS



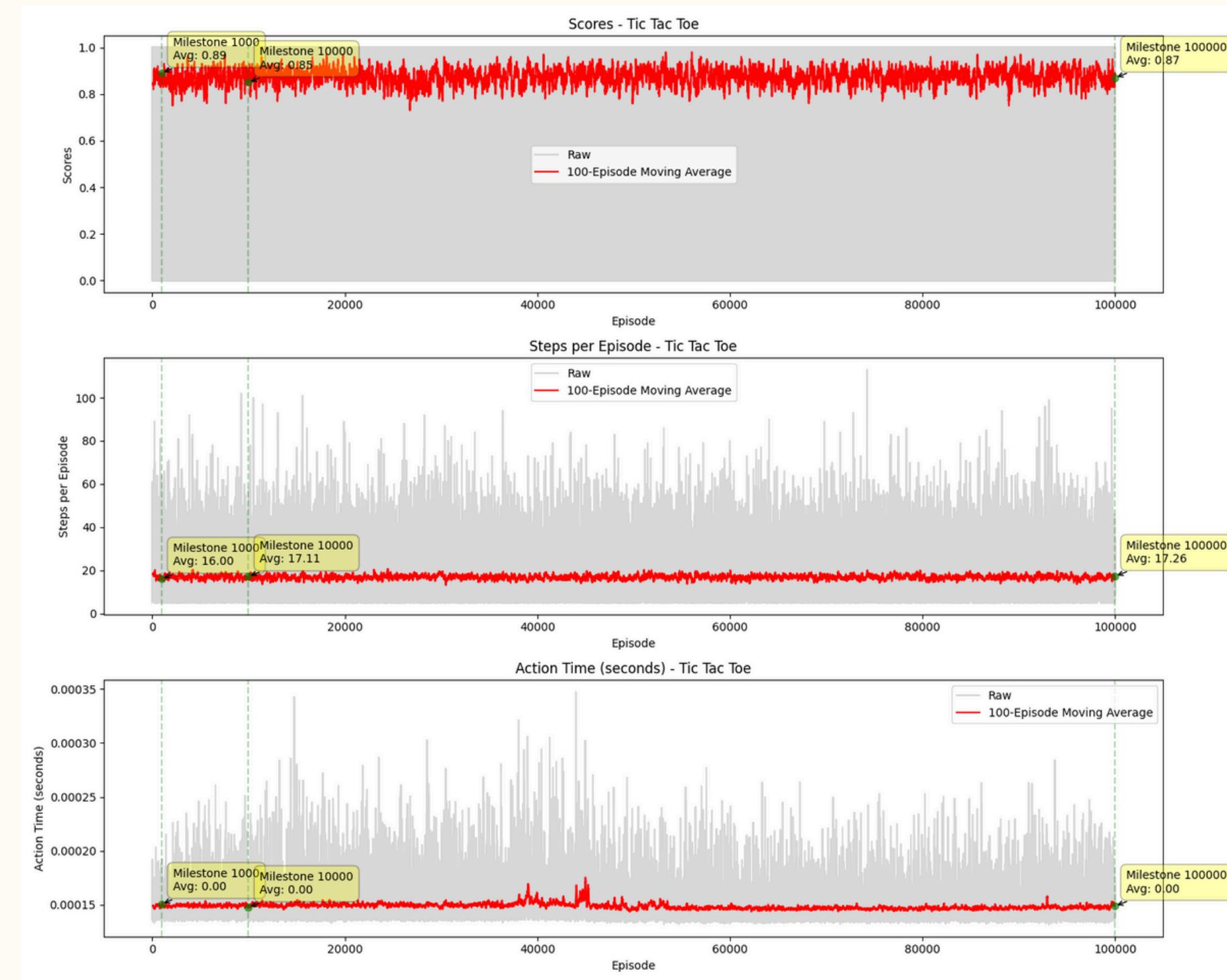
Résultats MCTS



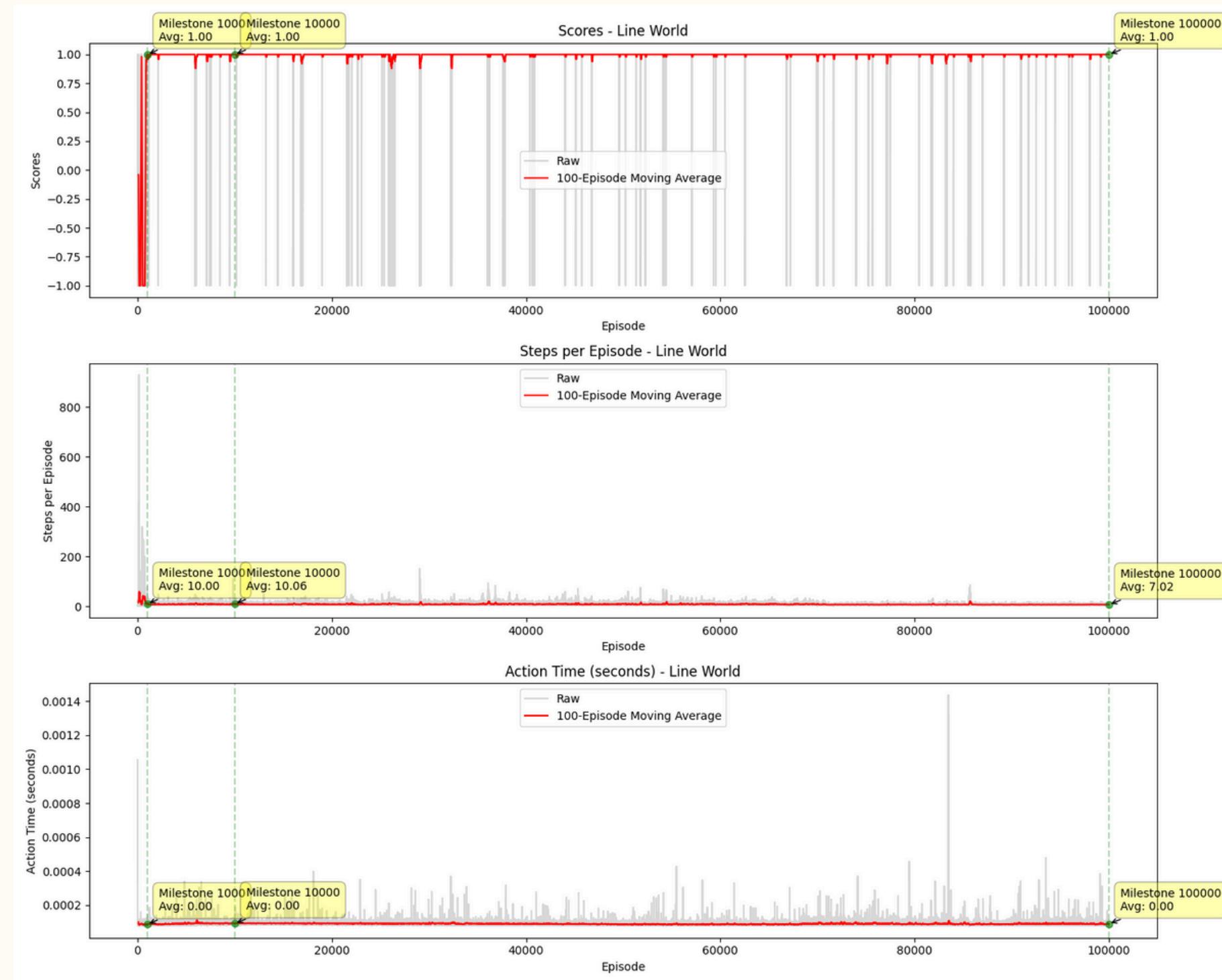
Résultats MCTS



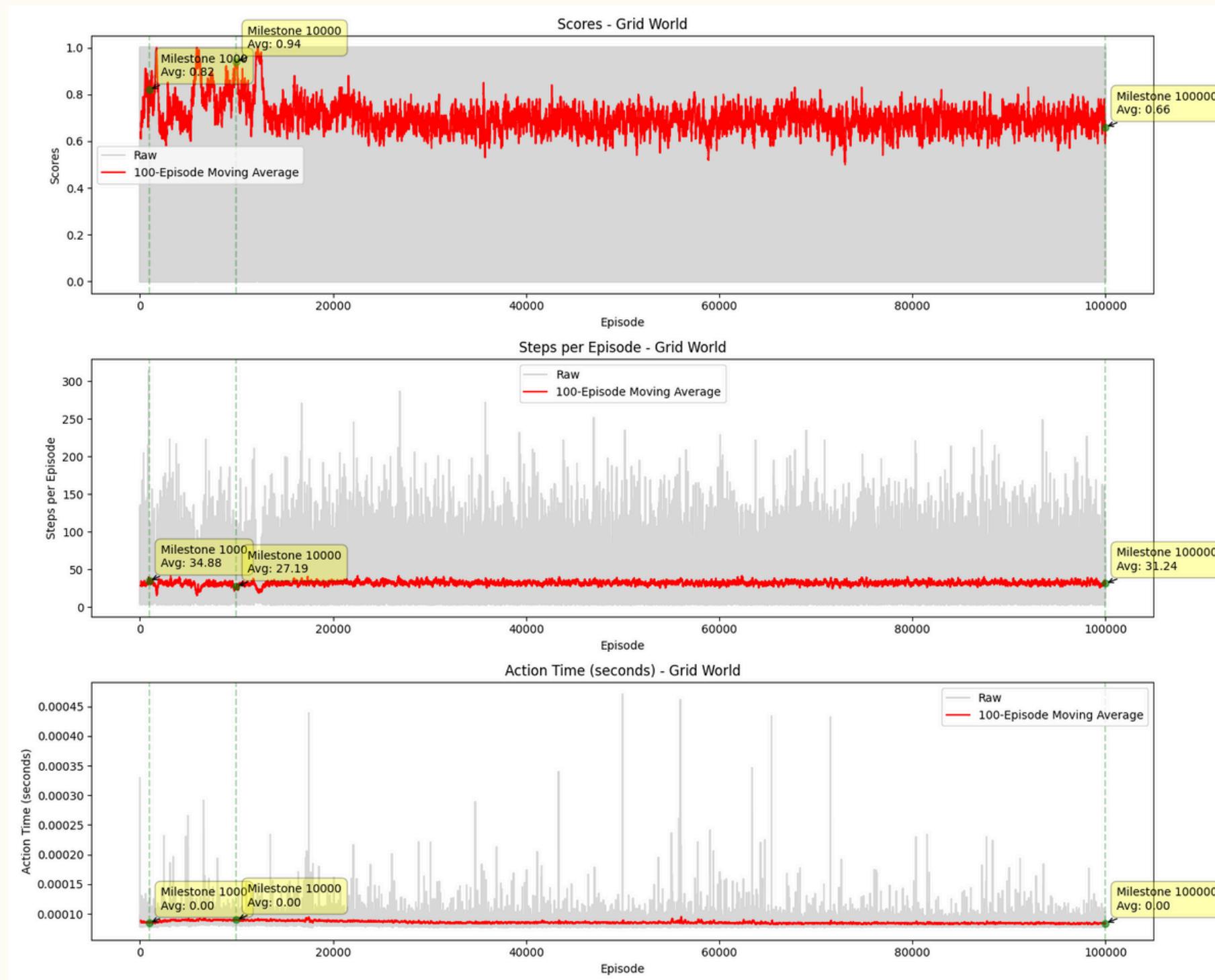
Résultats PPO



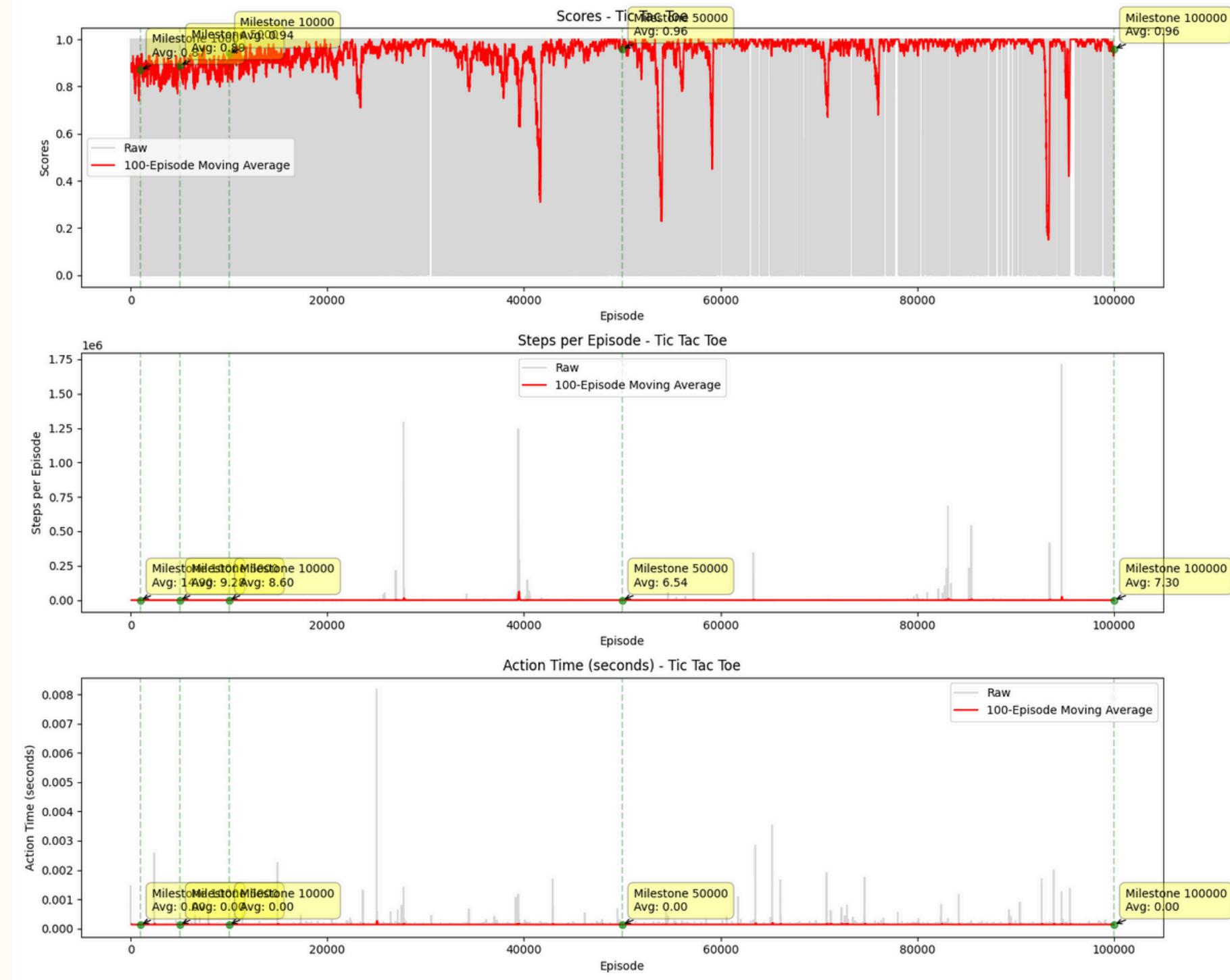
Résultats PPO



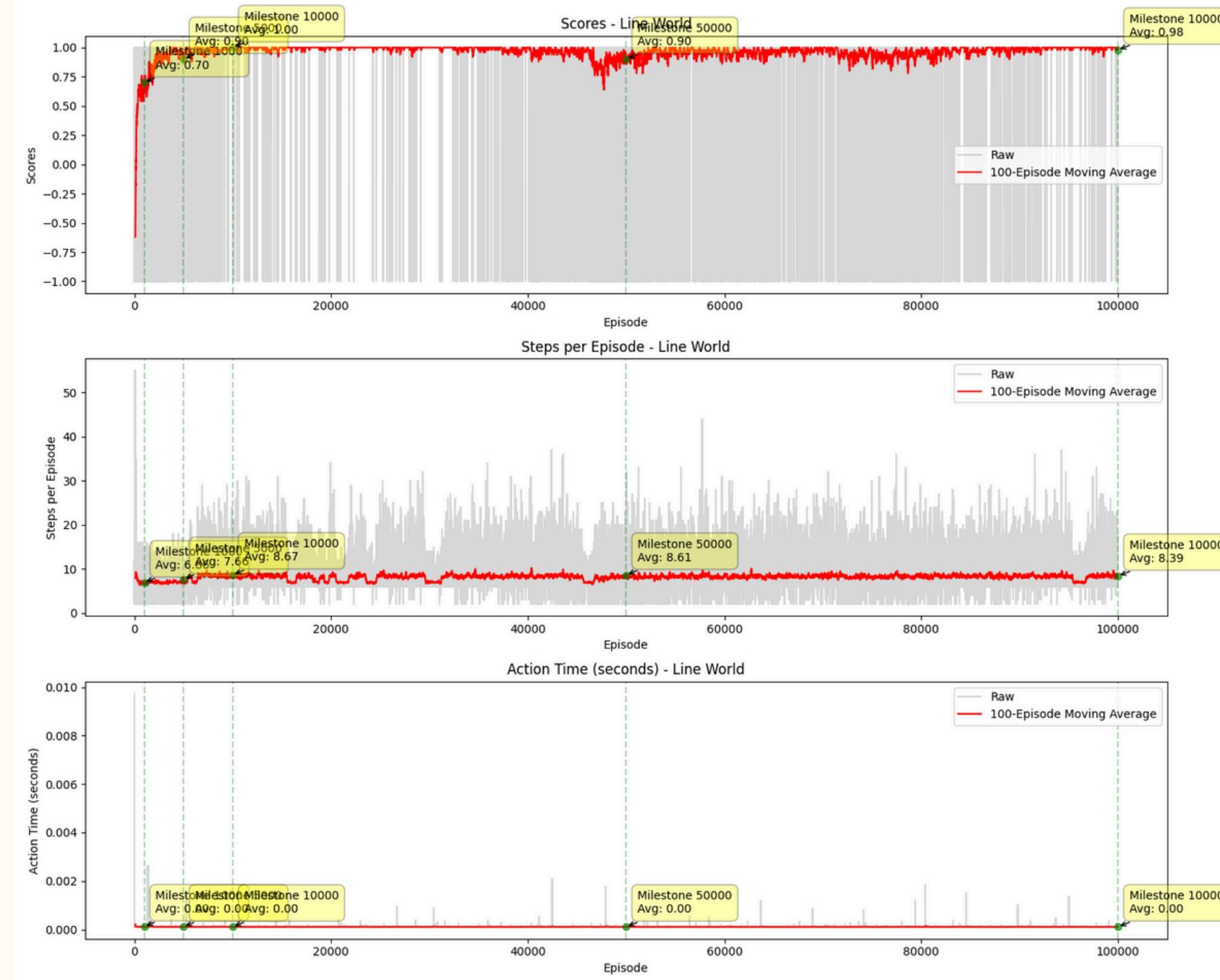
Résultats PPO



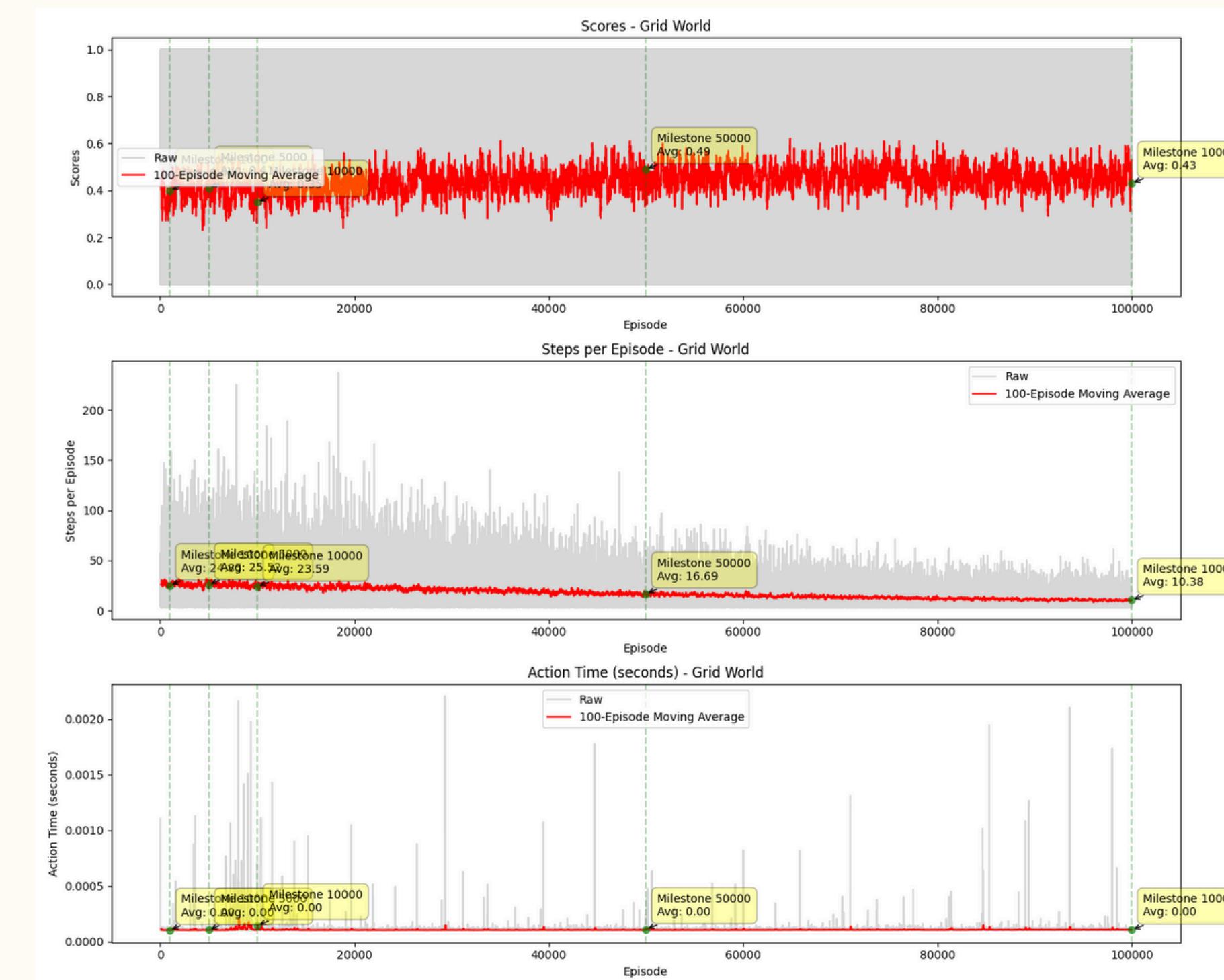
Résultats REINFORCE



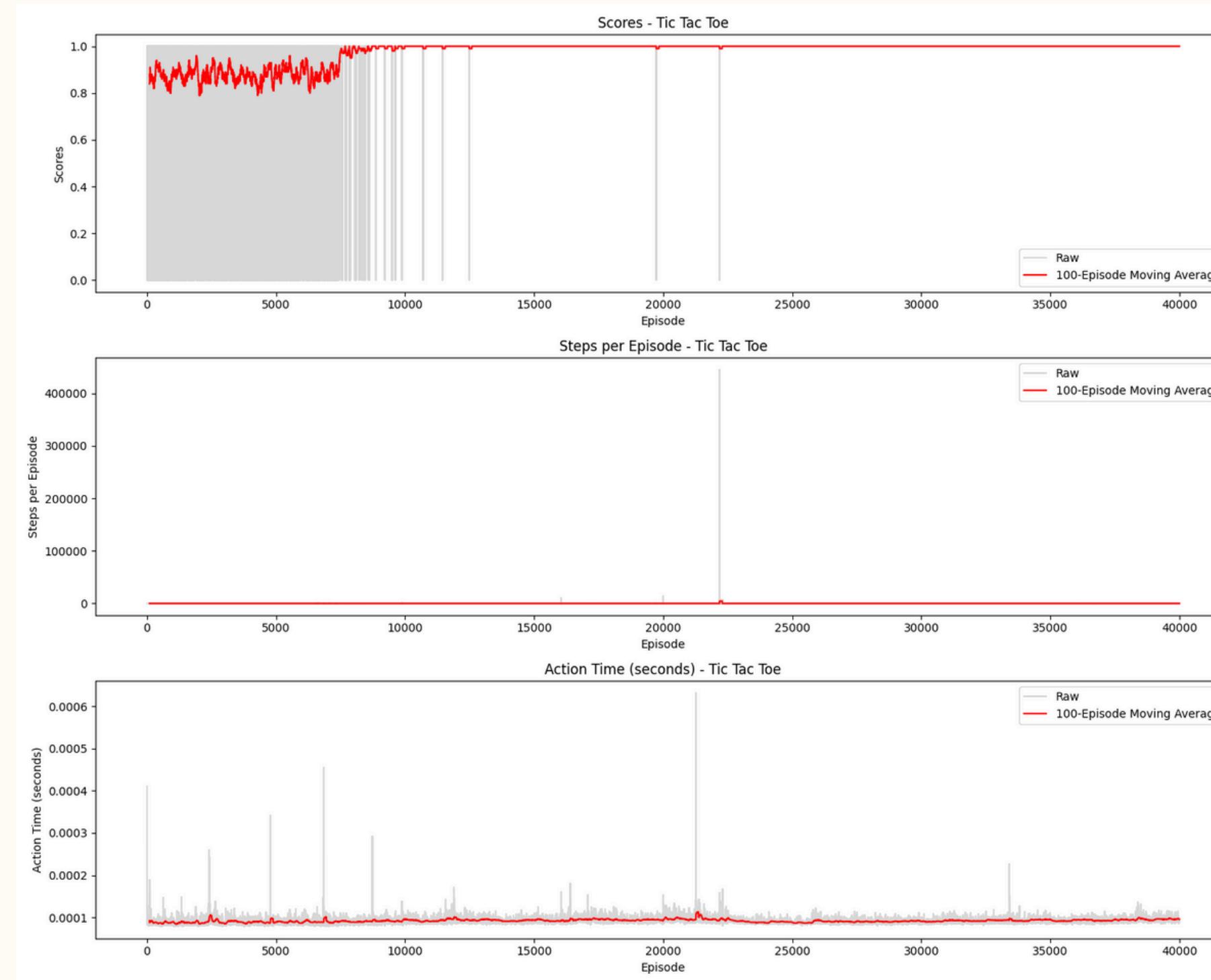
Résultats REINFORCE



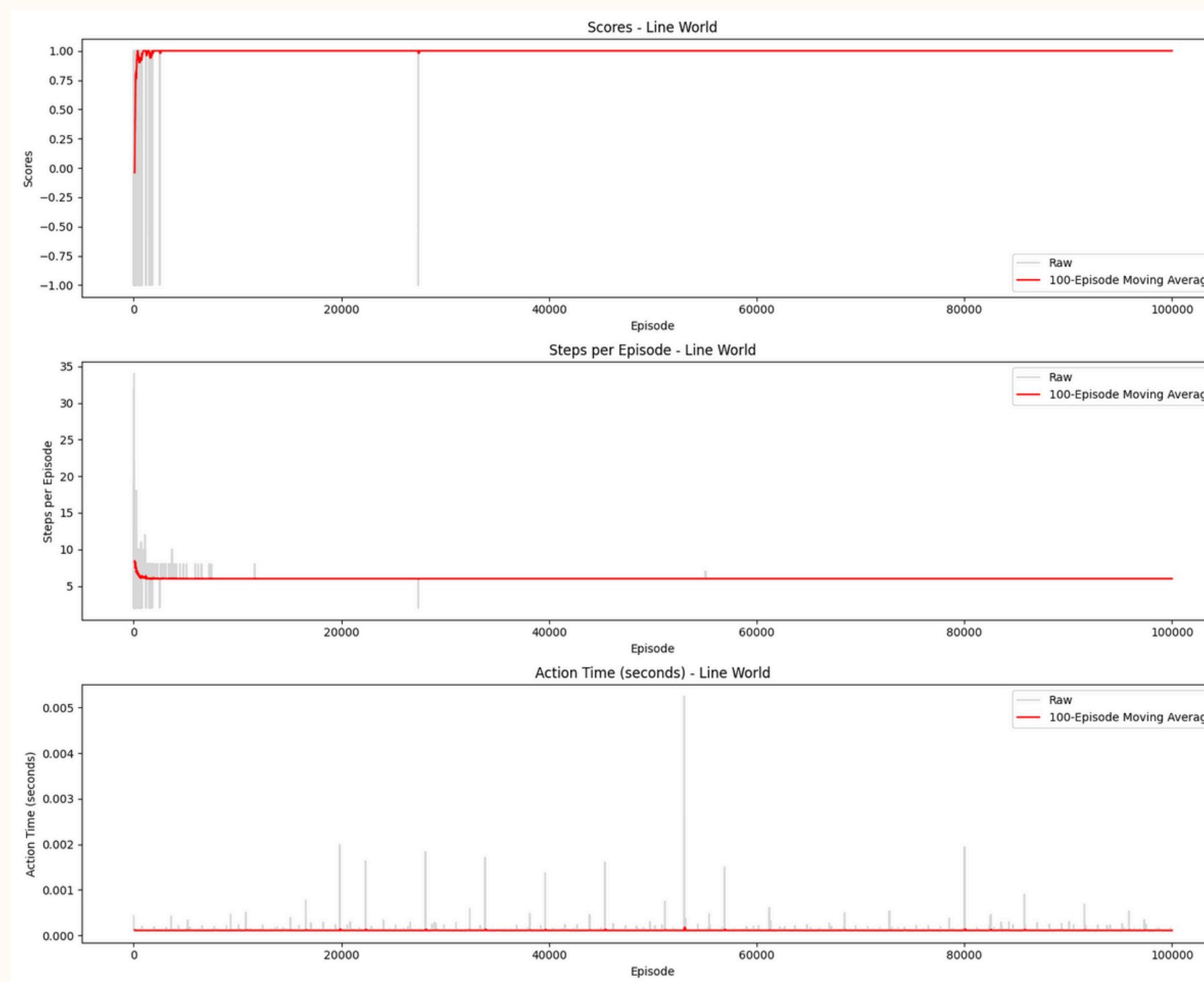
Résultats REINFORCE



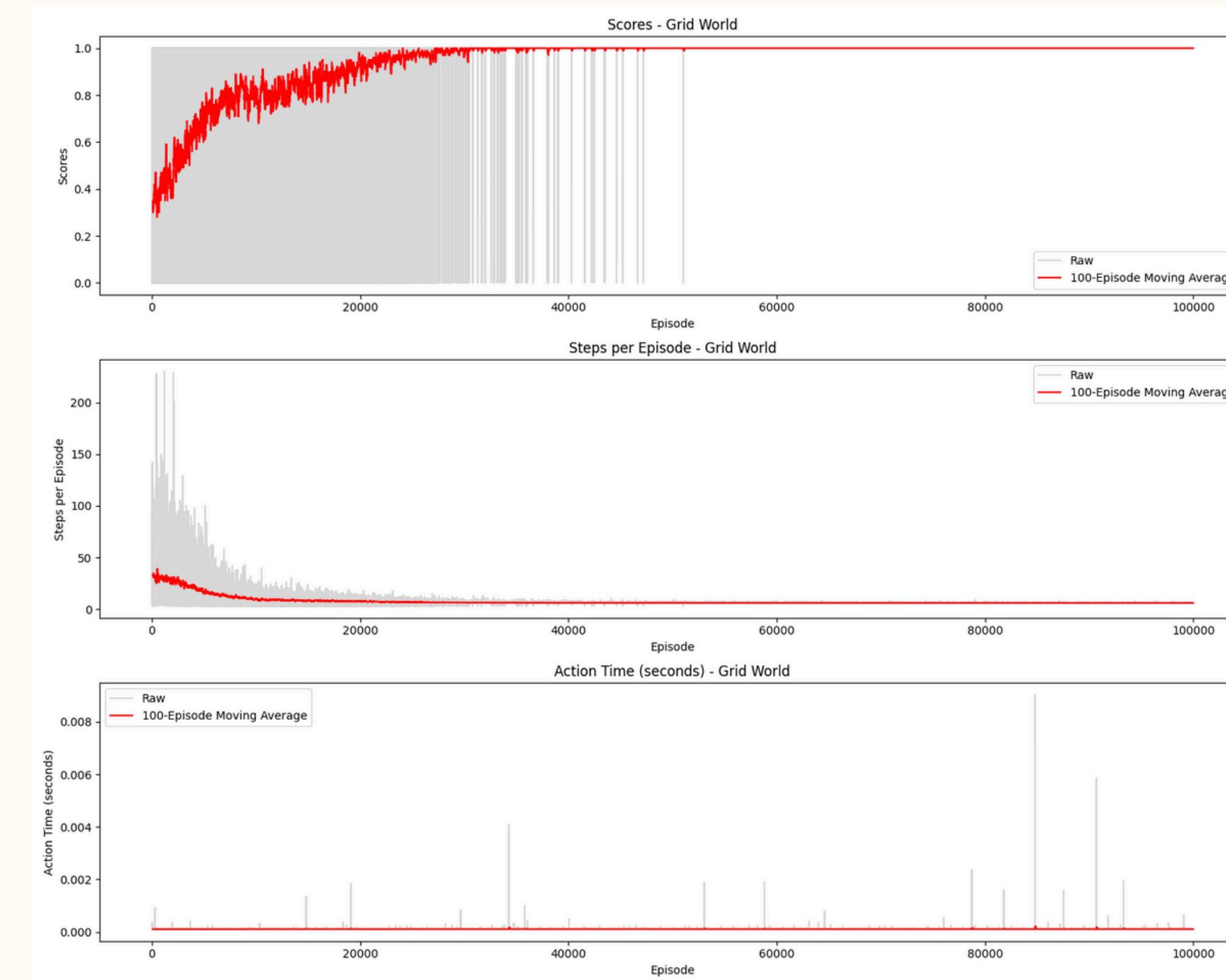
Résultats REINFORCE baseline



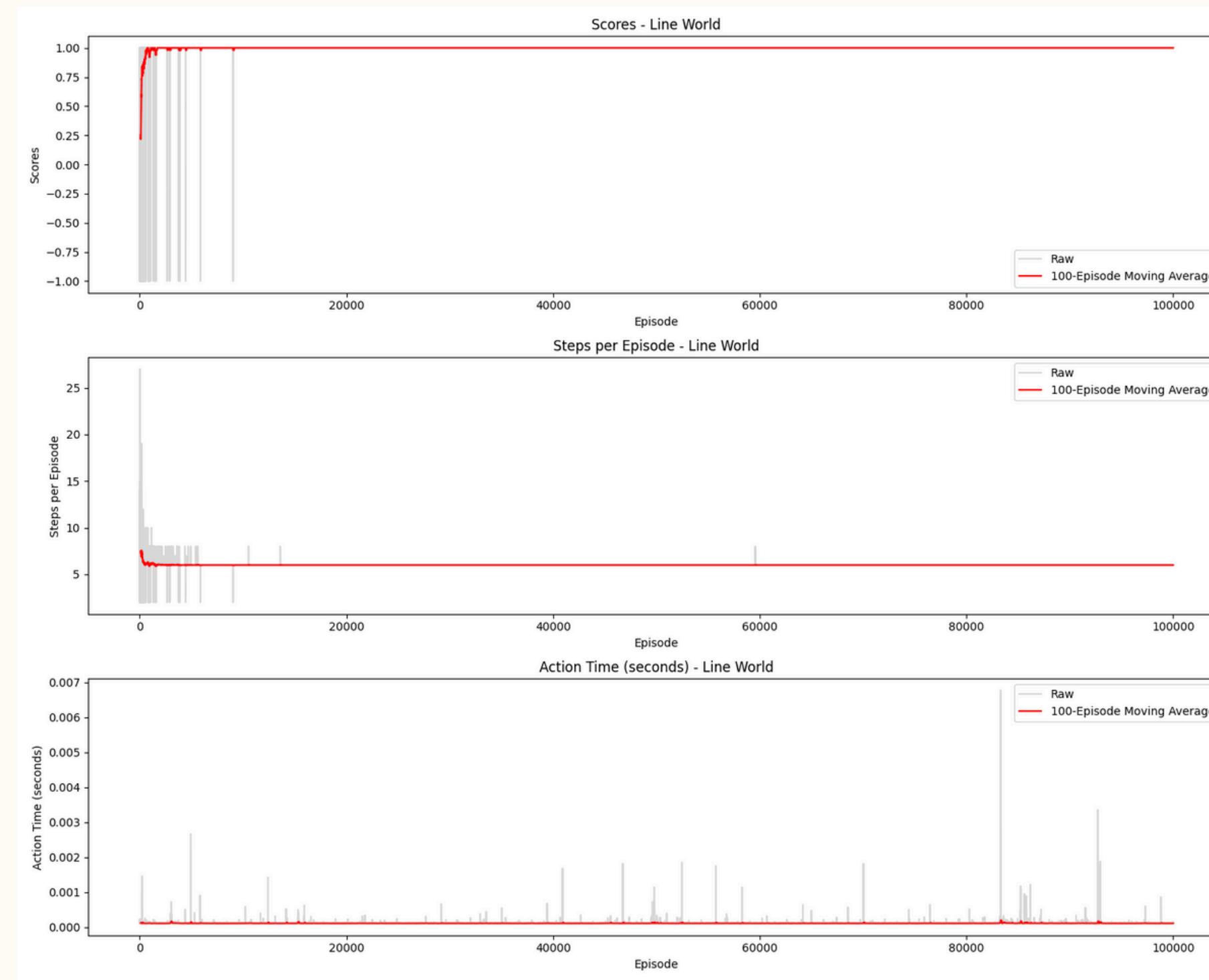
Résultats REINFORCE baseline



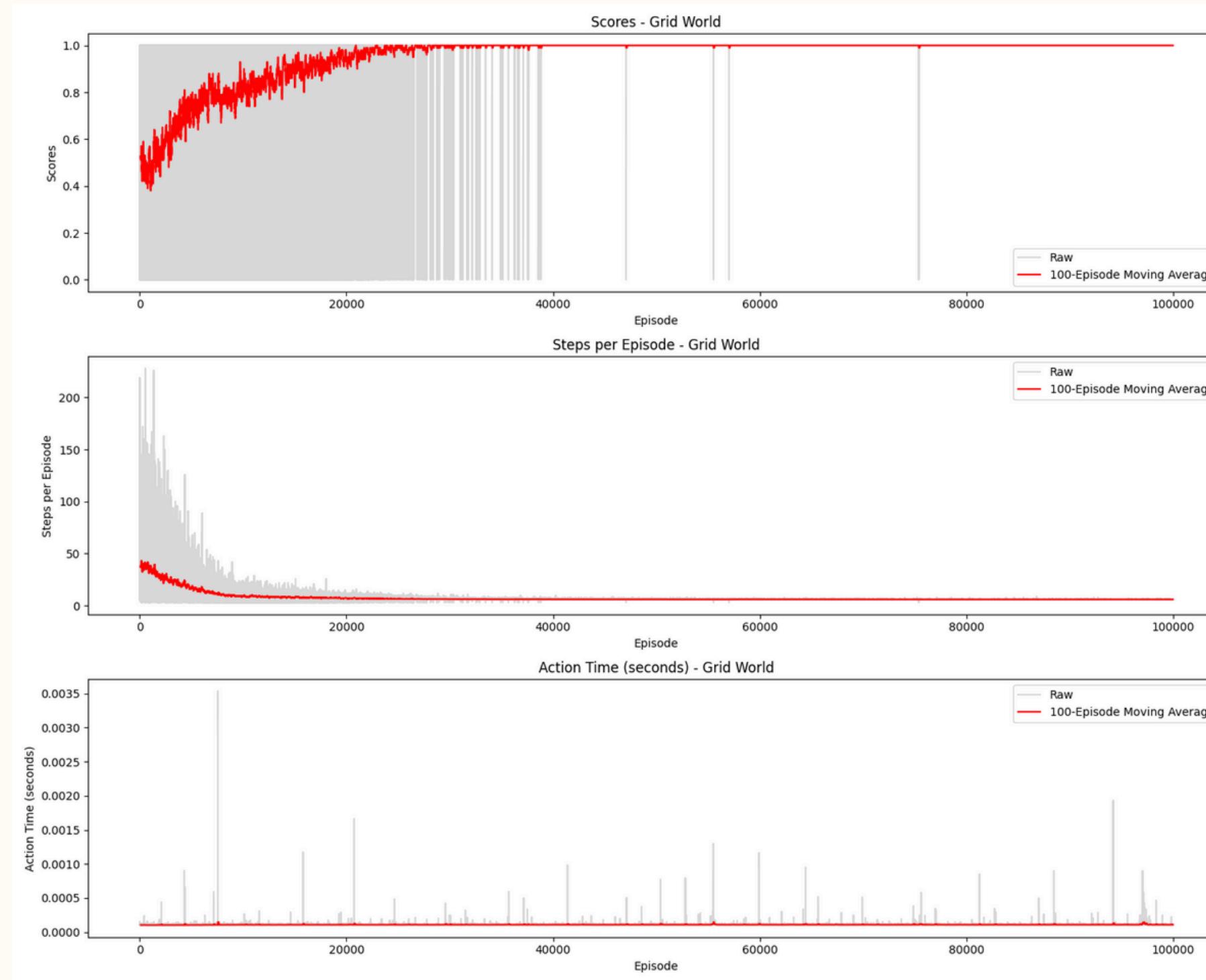
Résultats REINFORCE baseline



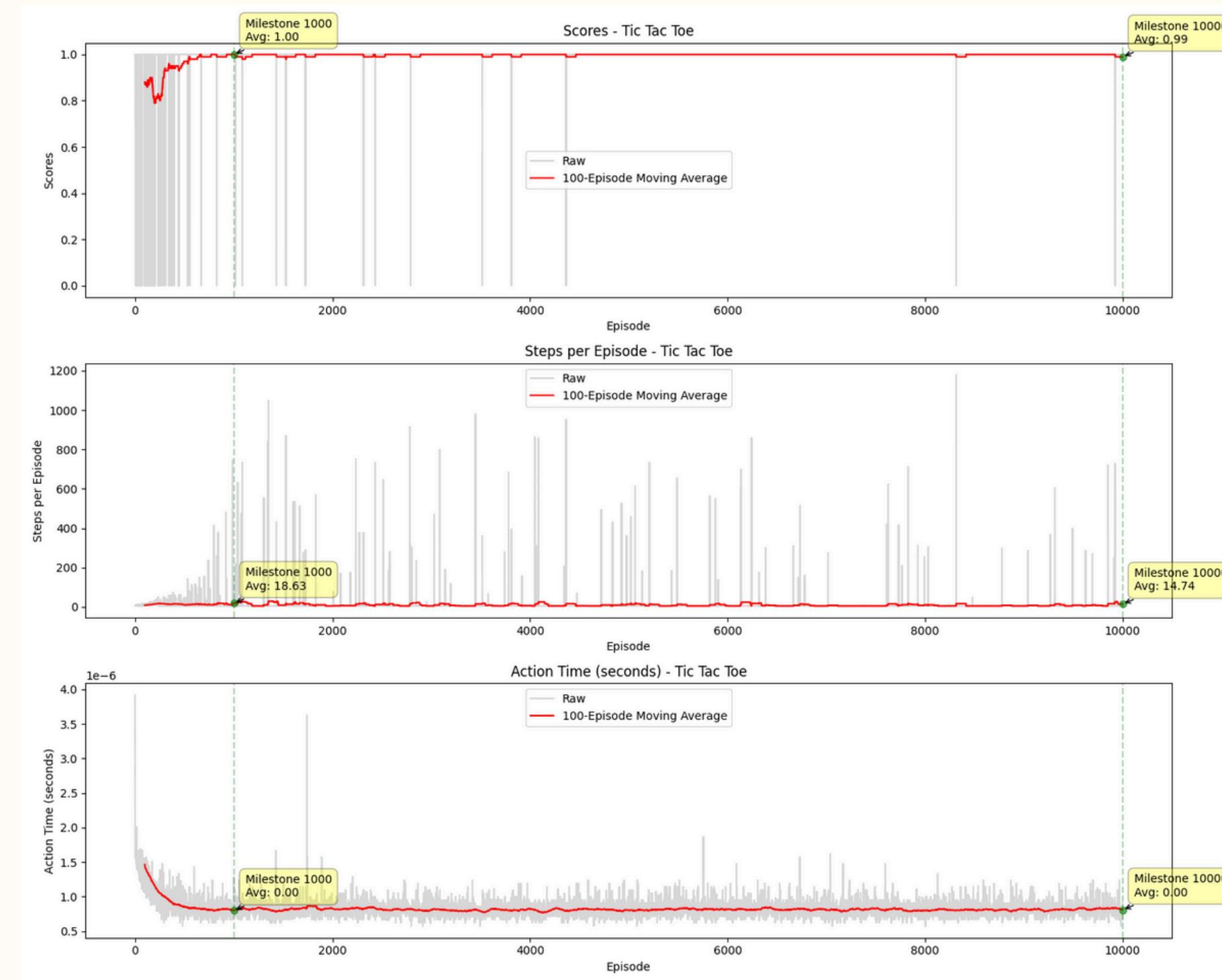
Résultats REINFORCE baseline critic



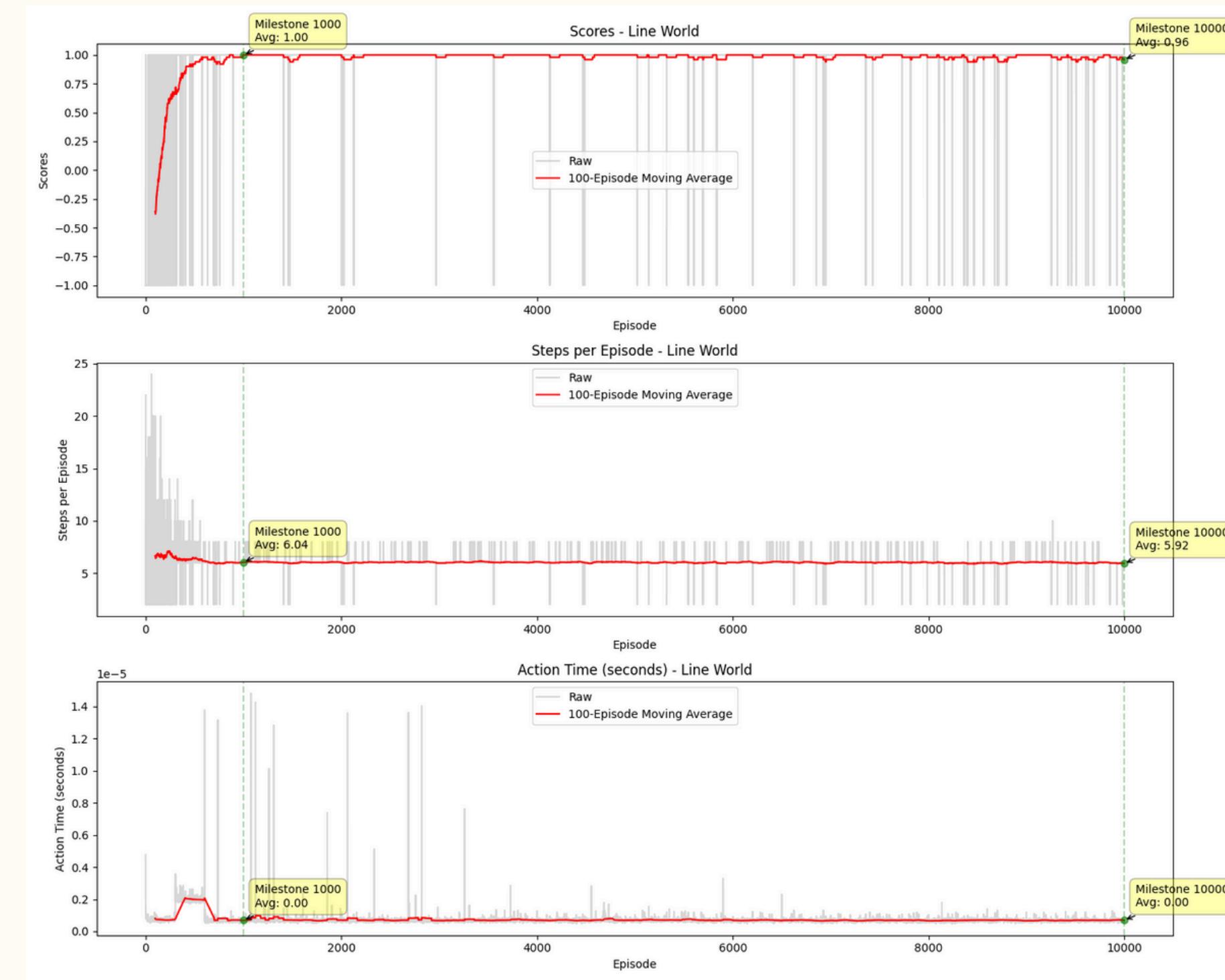
Résultats REINFORCE baseline critic



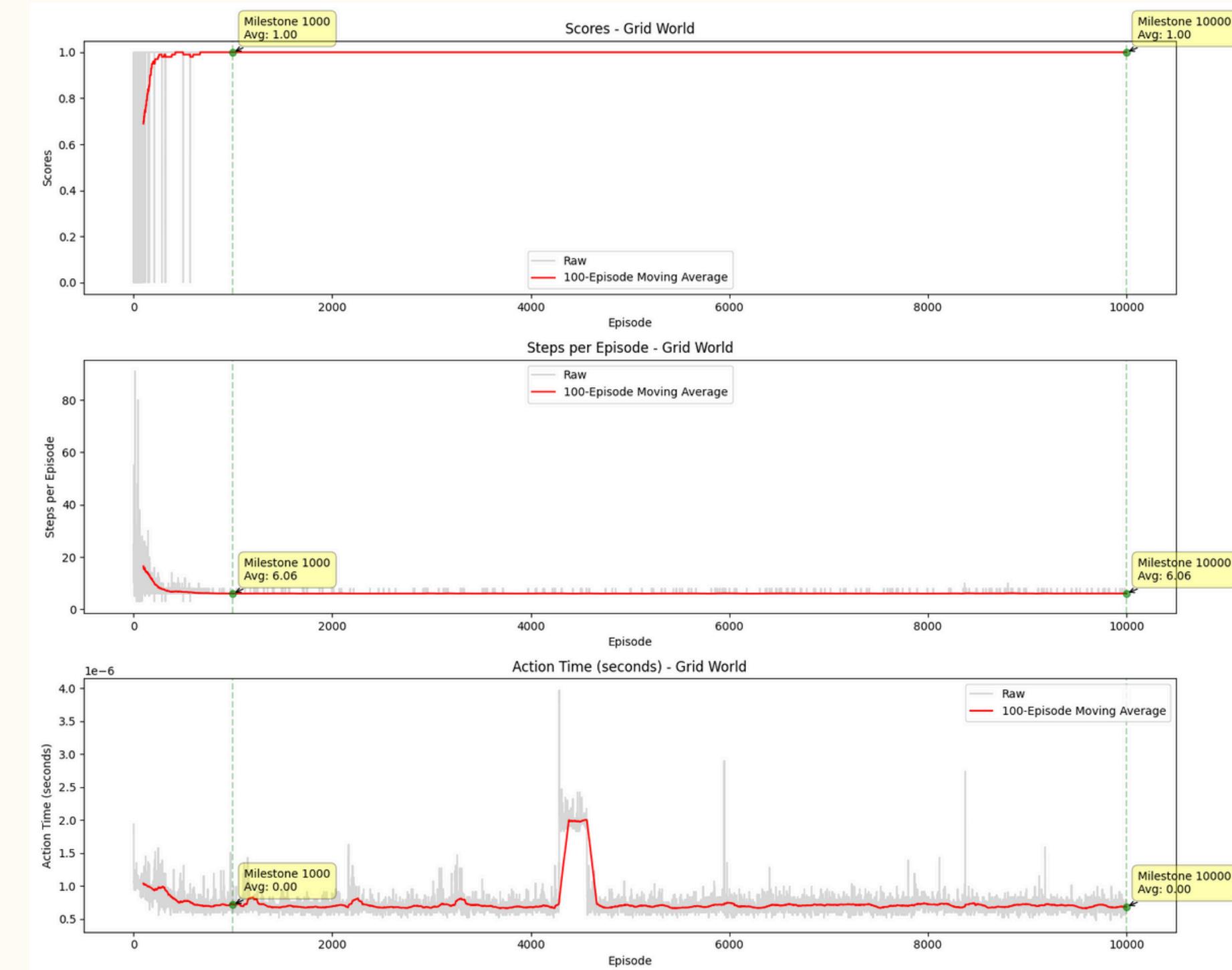
Résultats tab q learn



Résultats tab q learn



Résultats tab q learn



Résultats tab q learn

