Rapport projet DRL

- **BENDAVID Natane**
- **TARDY Louis**
- **WADE Cheikh Abdourahmane**

Dans ce rapport, vous trouverez l'intégralité des algorithmes que nous avons implementé ainsi que les résultats obtenus pour chaqun des environments testé sur les algos

Dynamic programming

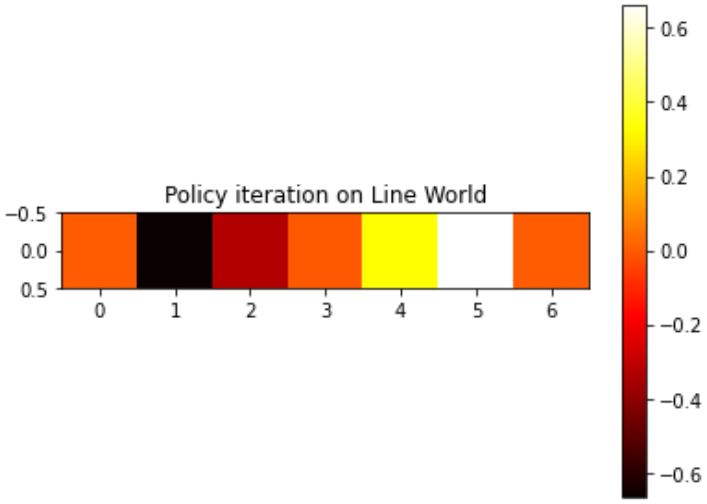
LINE WORD

Policy Evaluation & Policy Iteration

Nous avons testé la policy evaluation sur l'environnement Line World avec un gamma de 0.99 et un theta de 0.00001

Policy associée : Gauche, Droite, Droite, Droite, Droite

Graphique correspondant

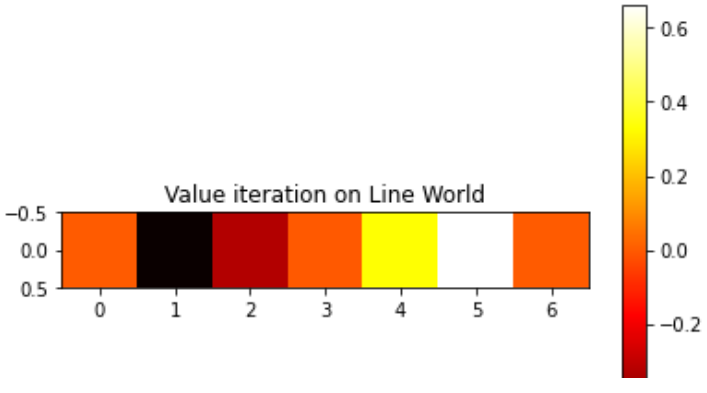


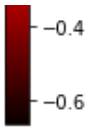
Value Iteration

Nous avons testé la value iteration sur l'environnement Line World avec un gamma de 0.99 et un theta de 0.00001

Policy associée: Gauche, Droite, Droite, Droite, Droite

Graphique correspondant



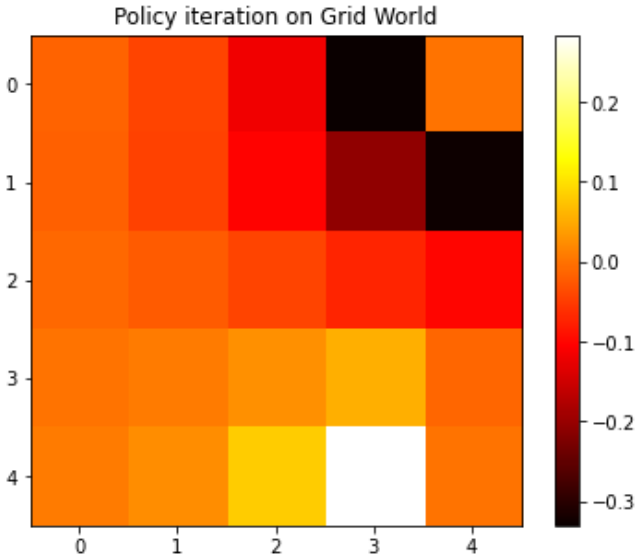


GRID WORD

Policy Evaluation & Policy Iteration

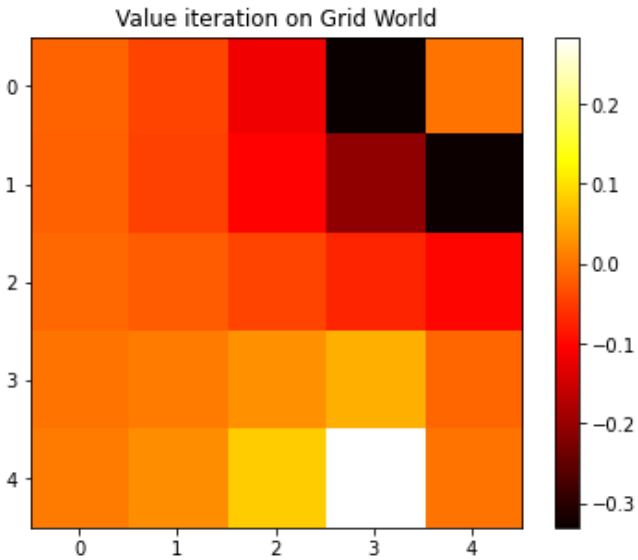
Nous avons testé la policy evaluation sur l'environnement Grid World avec un gamma de 0.99 et un theta de 0.00001

Graphique correspondant:



Value Iteration

Nous avons testé la value iteration sur l'environnement Grid World avec un gamma de 0.99 et un theta de 0.00001



SECRET ENV 1

Policy Evaluation & Policy Iteration

Nous avons testé la policy evaluation sur l'environnement secret 1 avec un gamma de 0.99 et un theta de 0.00001

Impossible de visualiser cet environnement mais on obtient cette value function:

{1: -0.5533333498239517, 2: -0.3333333333333333, 3: -0.3333333333333333, 4: -0.3333333333333333, 5: 0.0}

Et cette Policy associée :

{0: {1: 1.0, 2: 0.0, 3: 0.0}, 1: {1: 1.0, 2: 0.0, 3: 0.0}, 2: {1: 0.0, 2: 1.0, 3: 0.0}, 3: {1: 0.0, 2: 0.0, 3: 1.0}, 4: {1: 1.0, 2: 0.0, 3: 0.0}}

Value Iteration

Nous avons testé la value iteration sur l'environnement Grid World avec un gamma de 0.99 et un theta de 0.00001

Nous obtenons la value function suivante:

{1: -0.5533333498239517, 2: -0.3333333333333333, 3: -0.3333333333333333, 4: -0.3333333333333333, 5: 0.0}

Et cette Policy associée:

{0: {1: 1.0, 2: 0.0, 3: 0.0}, 1: {1: 1.0, 2: 0.0, 3: 0.0}, 2: {1: 0.0, 2: 1.0, 3: 0.0}, 3: {1: 0.0, 2: 0.0, 3: 1.0}, 4: {1: 1.0, 2: 0.0, 3: 0.0}}

Monte carlo methods

TIC TAC TOE

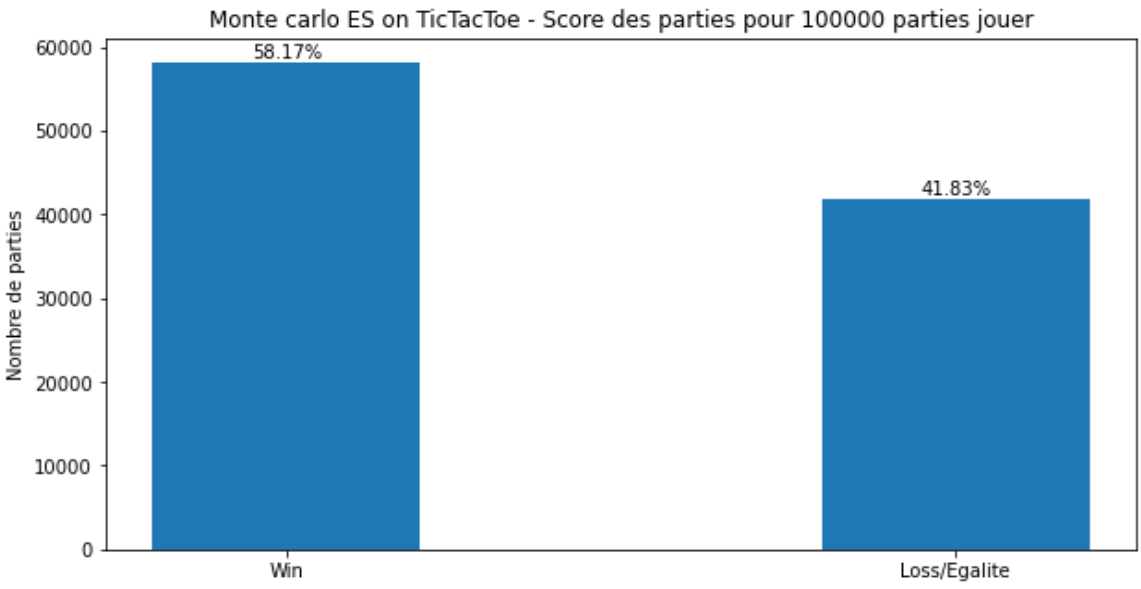
MONTE CARLO ES

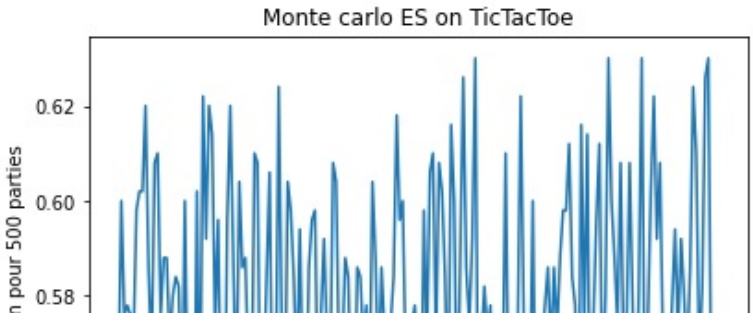
Nous avons testé l'algo Monte Carlo ES sur l'env TIC TAC TOE avec un gamma de 0.99

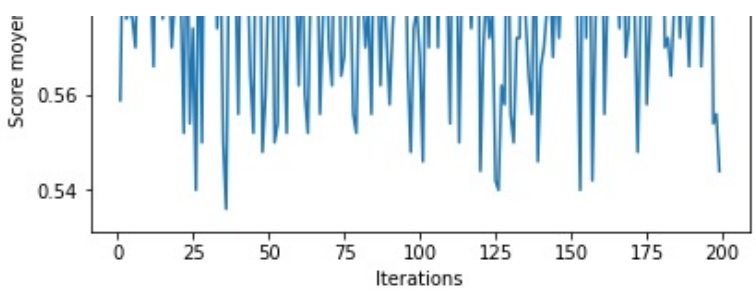
Nous avons obtenu les résultats suivant pour 100 000 parties jouées :

- **58.17% de parties gagnées**
- **41.83% de parties perdues ou égalités**

Voici les graphiques :







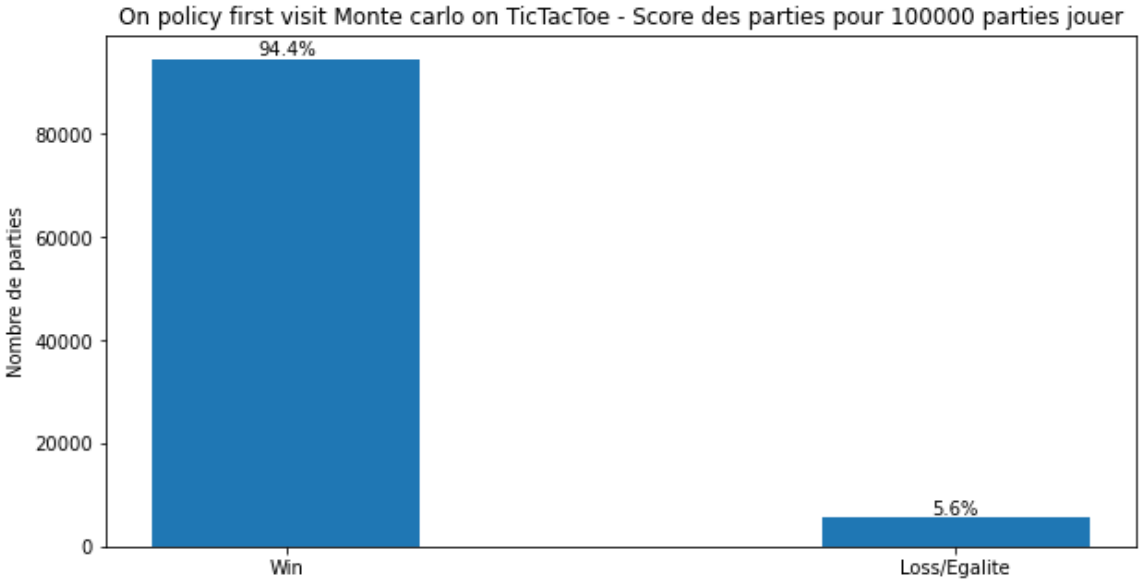
On policy first visit monte carlo control

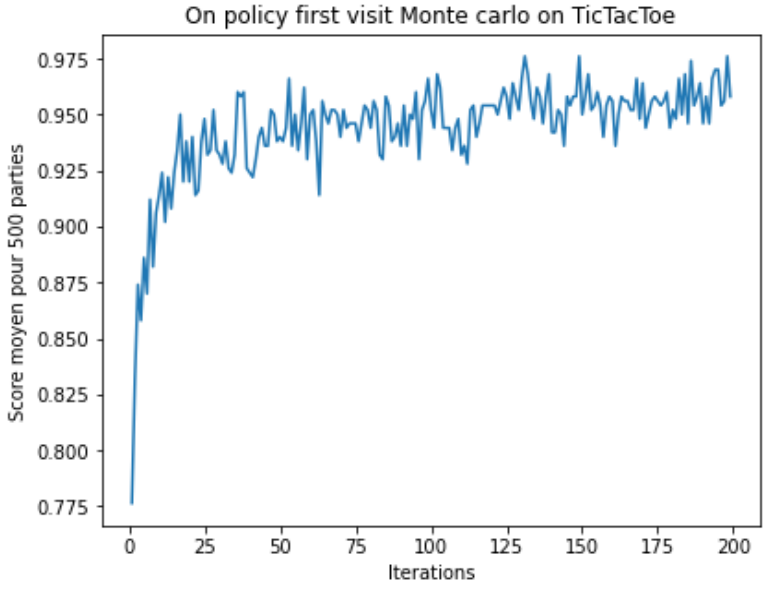
Nous avons testé L'algo On policy first visit monte carlo control sur l'env TIC TAC TOE avec un gamma de 0.99 et un epsilon de 0.1

Nous avons obtenu les résultats suivants pour 100 000 parties jouées :

- **94.4% de parties gagnées**
- **5.6% de parties perdues ou égalités**

Voici les graphiques :





Off policy monte carlo control

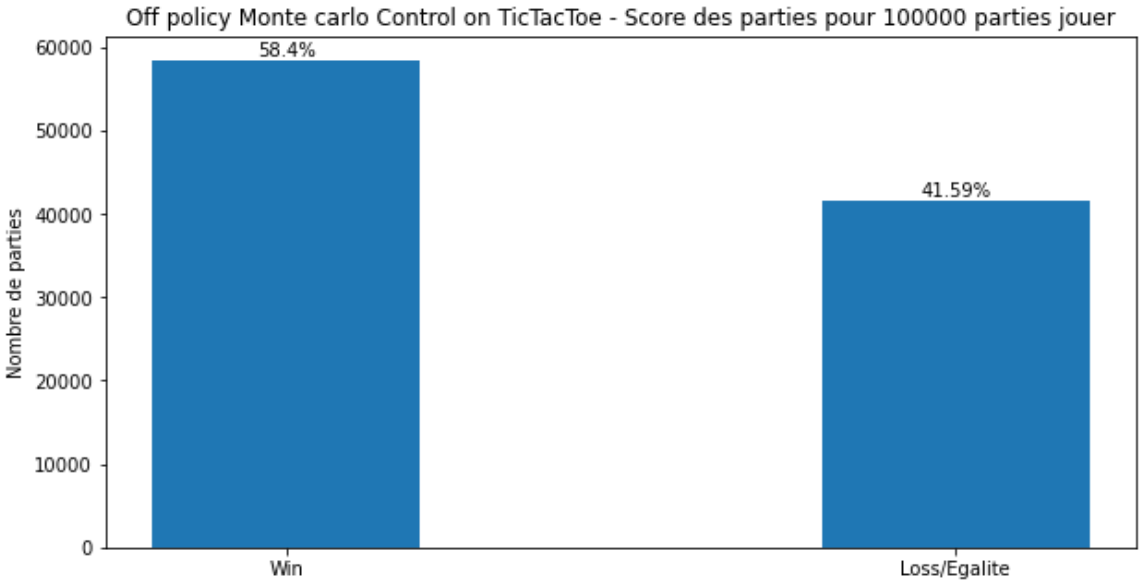
Nous avons testé l'algo Off policy monte carlo control sur l'env TIC TAC TOE avec un gamma de 0.99

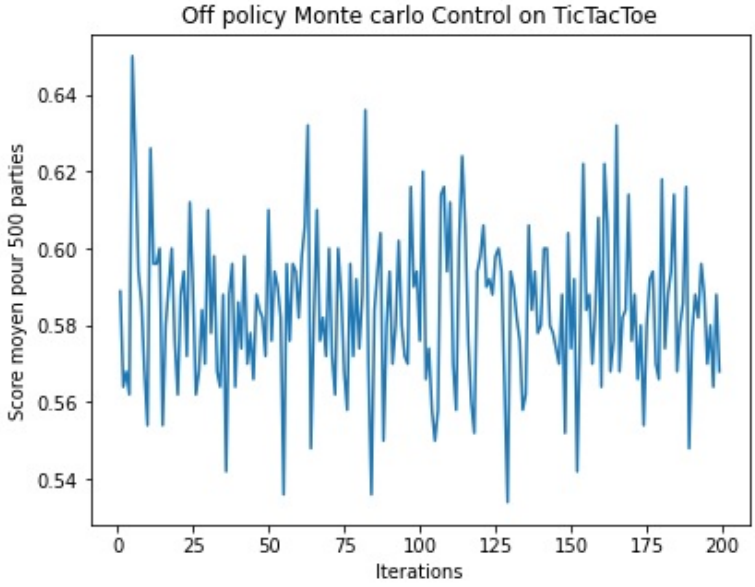
Nous avons obtenu les résultats suivants pour 100 000 parties jouées :

- **58.4% de parties gagnées**

- **41.59% de parties perdues ou egalites**

Voici les graphiques :





SECRET ENV 2

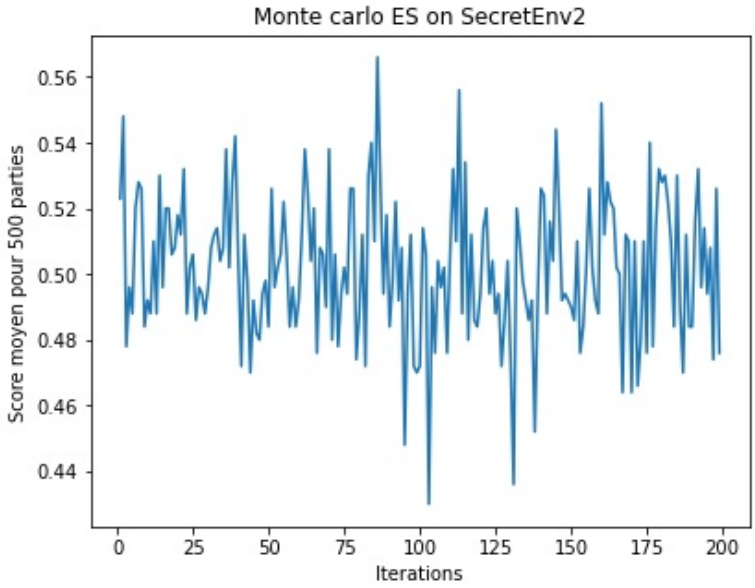
MONTE CARLO ES

Nous avons testé L'algo Monte Carlo ES sur l'env Secret ENV 2 avec un gamma de 0.99

Nous avons obtenu les résultats suivants pour 100 000 parties jouées :

- **50.2% de parties gagnées**
- **49.8% de parties perdues ou égalités**

Voici les graphiques :



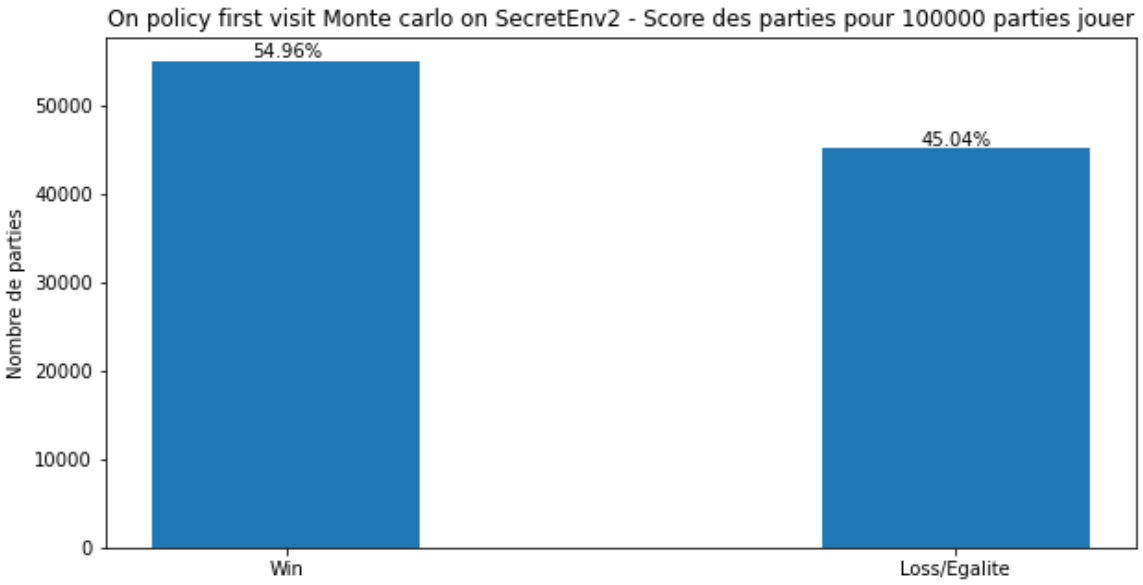
On policy first visit monte carlo control

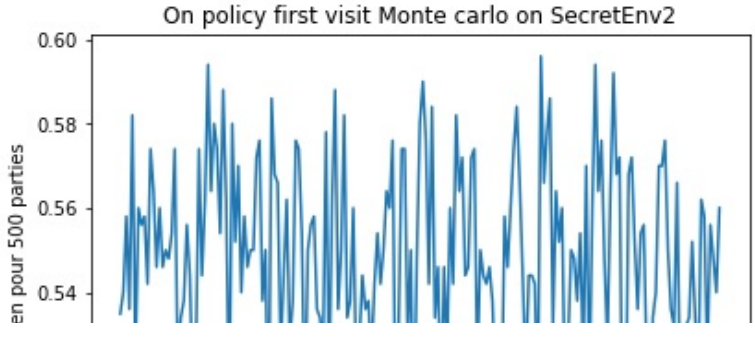
Nous avons testé L'algo On policy first visit monte carlo control sur l'env SECRET ENV 2 avec un gamma de 0.99 et un epsilon de 0.1

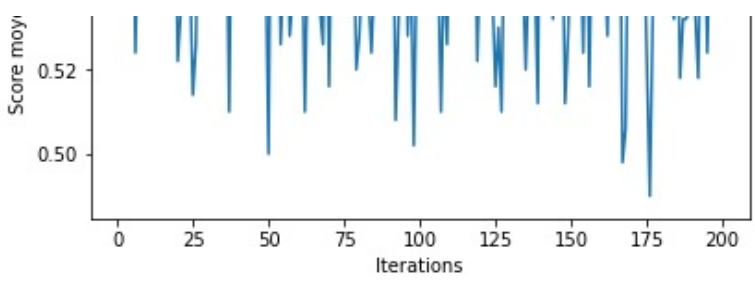
Nous avons obtenu les résultats suivants pour 100 000 parties jouées :

- **54.96% de parties gagnées**
- **45.04% de parties perdues ou égalités**

Voici les graphiques :







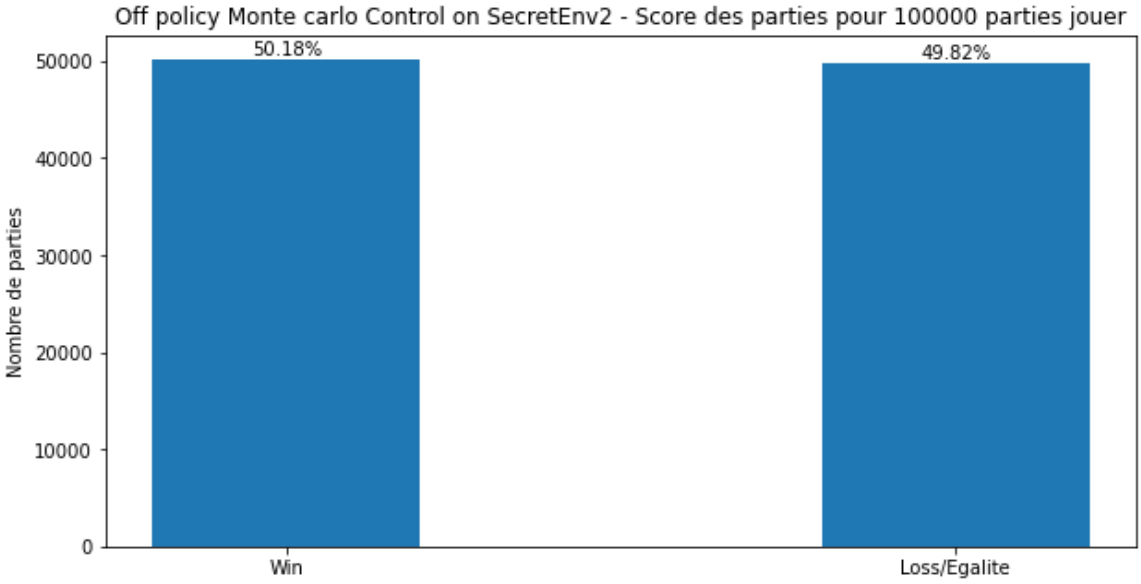
Off policy monte carlo control

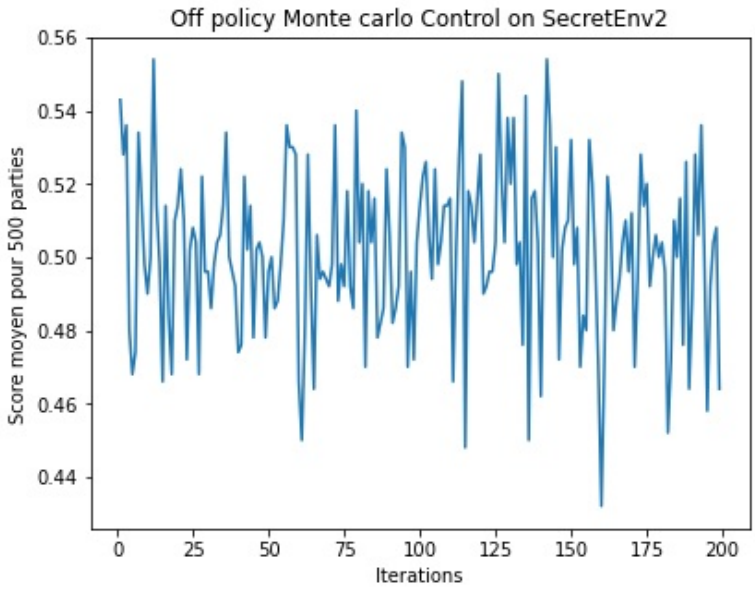
Nous avons testé l'algo Off policy monte carlo control sur l'env SECRET ENV 2 avec un gamma de 0.99

Nous avons obtenu les résultats suivants pour 100 000 parties jouées :

- **50.18% de parties gagnées**
- **49.82% de parties perdues ou égalités**

Voici les graphiques :





Temporal difference learning

TIC TAC TOE

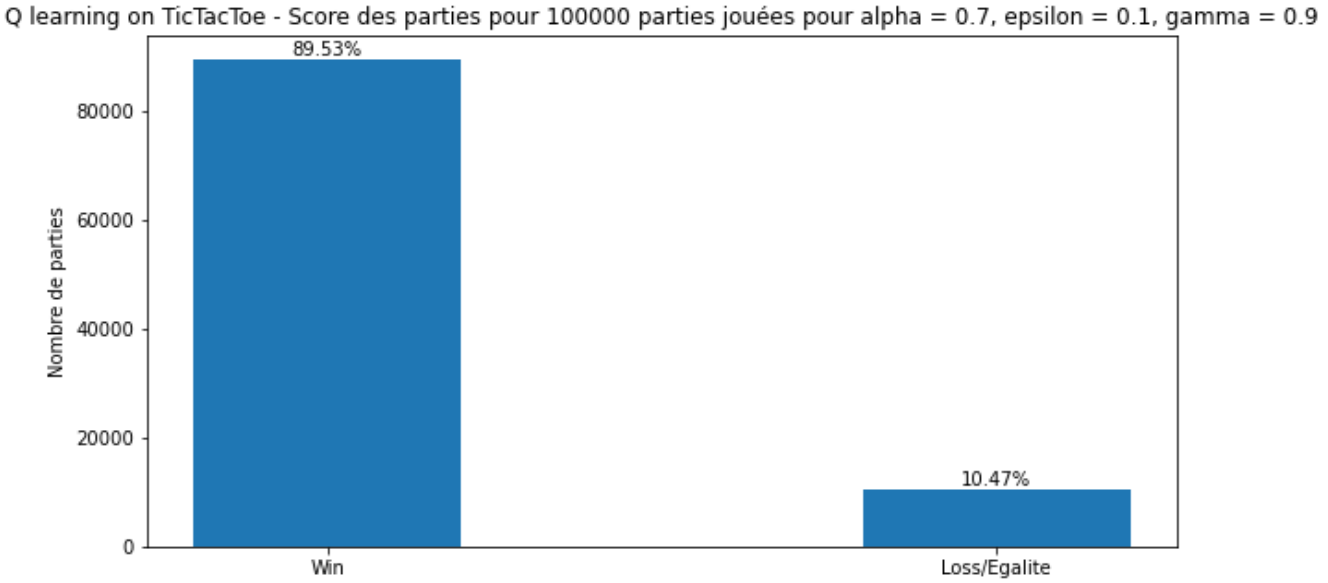
Q_learning

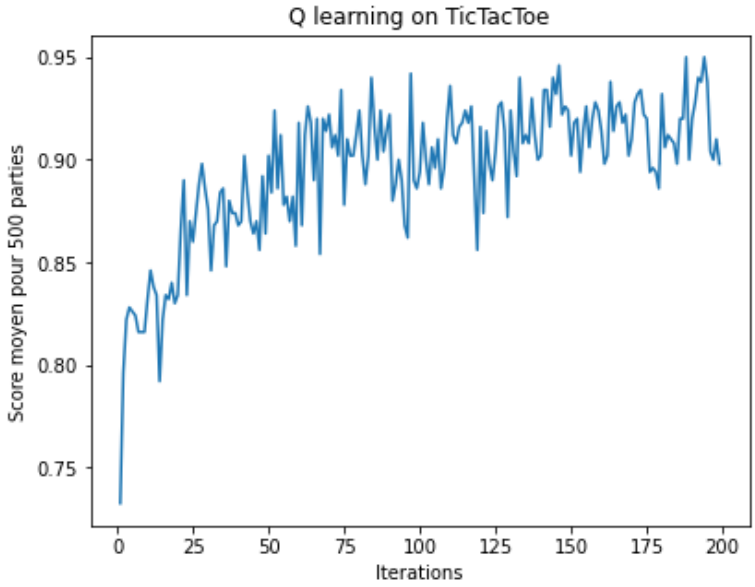
Nous avons testé L'algo Q_learning sur l'env TIC TAC TOE avec un alpha de 0.7, un epsilon de 0.1 et un gamma de 0.9

Nous avons obtenu les résultats suivants pour 100 000 parties jouées :

- **89.53% de parties gagnées**
- **10.47% de parties perdues ou égalités**

Voici les graphiques :





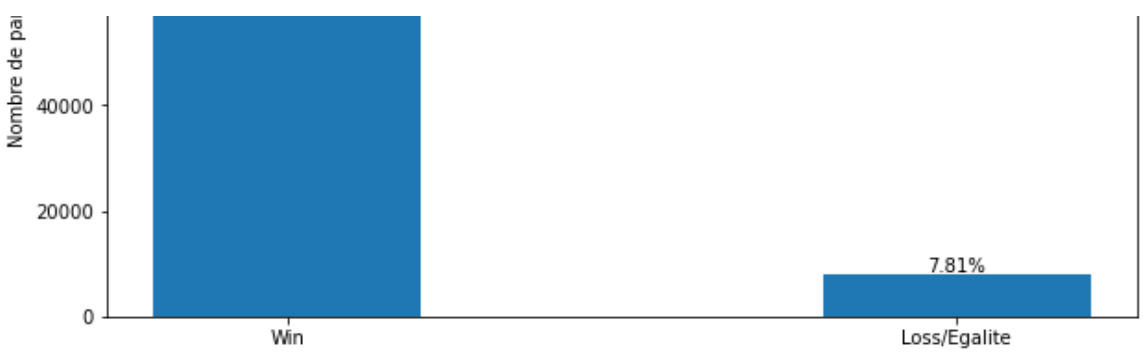
Expected Sarsa

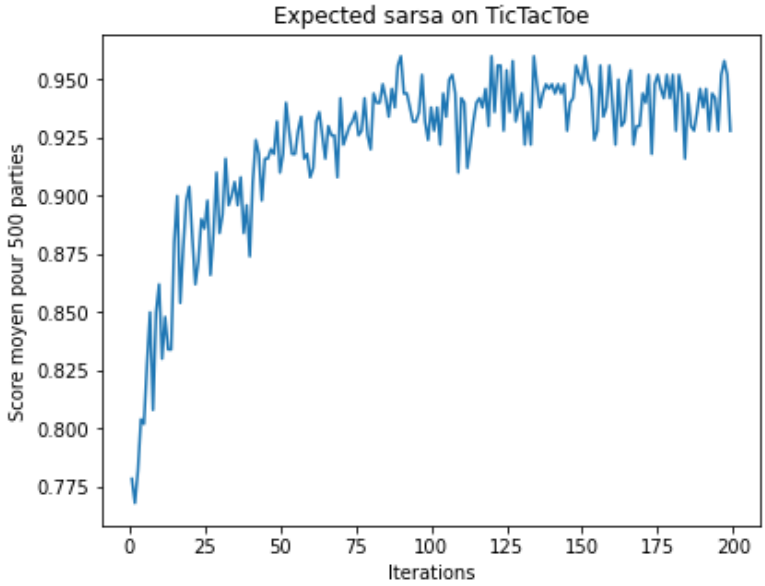
Nous avons testé L'algo Expected Sarsa sur l'env TIC TAC TOE avec un alpha de 0.7, un epsilon de 0.1 et un gamma de 0.9

Nous avons obtenu les résultats suivants pour 100 000 parties jouées :

- **92.18% de parties gagnées**
- **7.82% de parties perdues ou égalités**

Voici les graphiques :





SECRET ENV 3

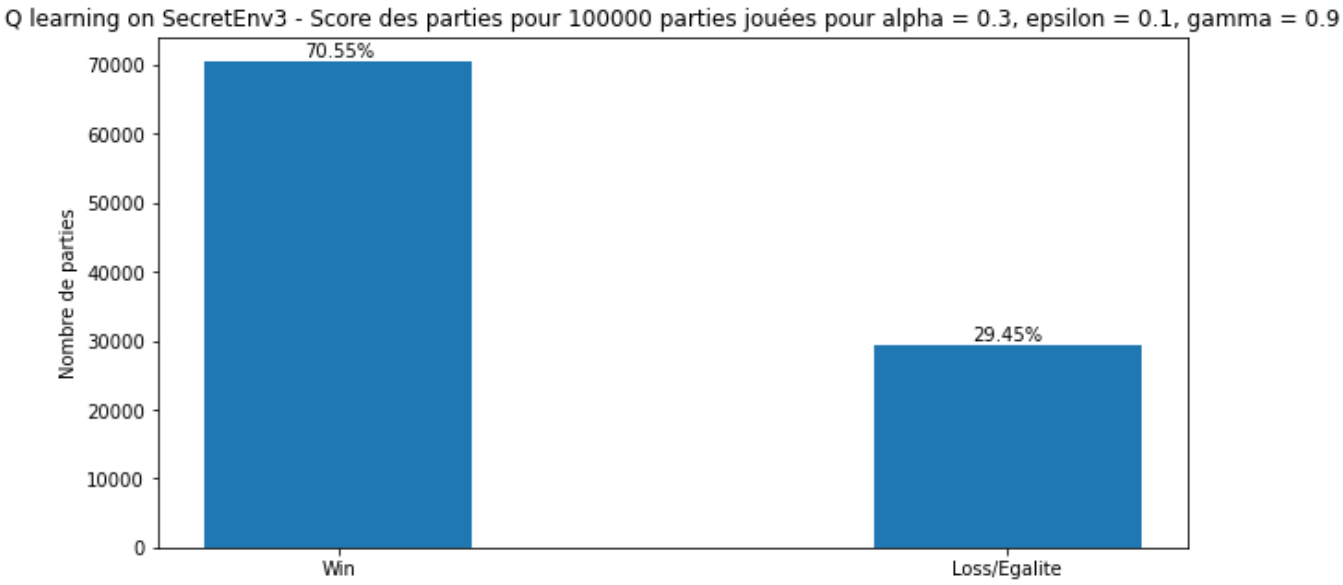
Q_learning

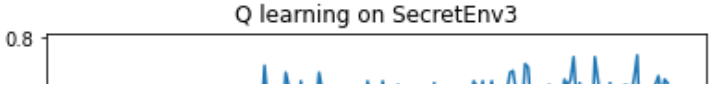
Nous avons testé L'algo Q_learning sur l'env Secret env 3 avec un alpha de 0.3, un epsilon de 0.1 et un gamma de 0.9

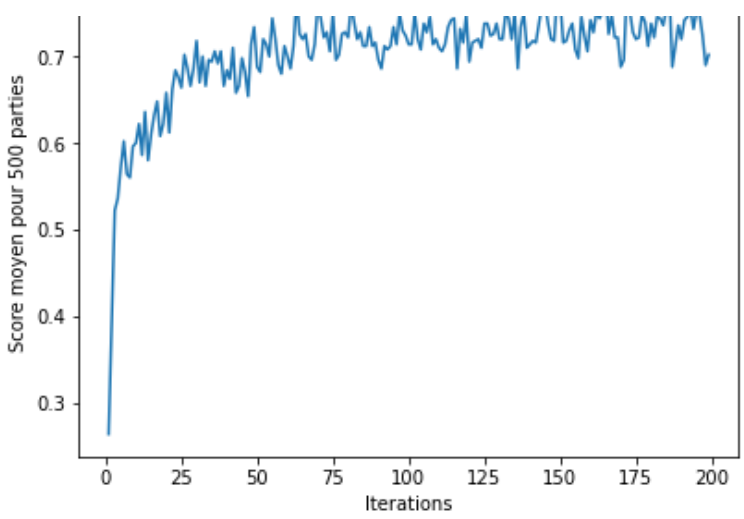
Nous avons obtenu les résultats suivants pour 100 000 parties jouées :

- **70.55% de parties gagnées**
- **29.45% de parties perdues ou égalités**

Voici les graphiques :







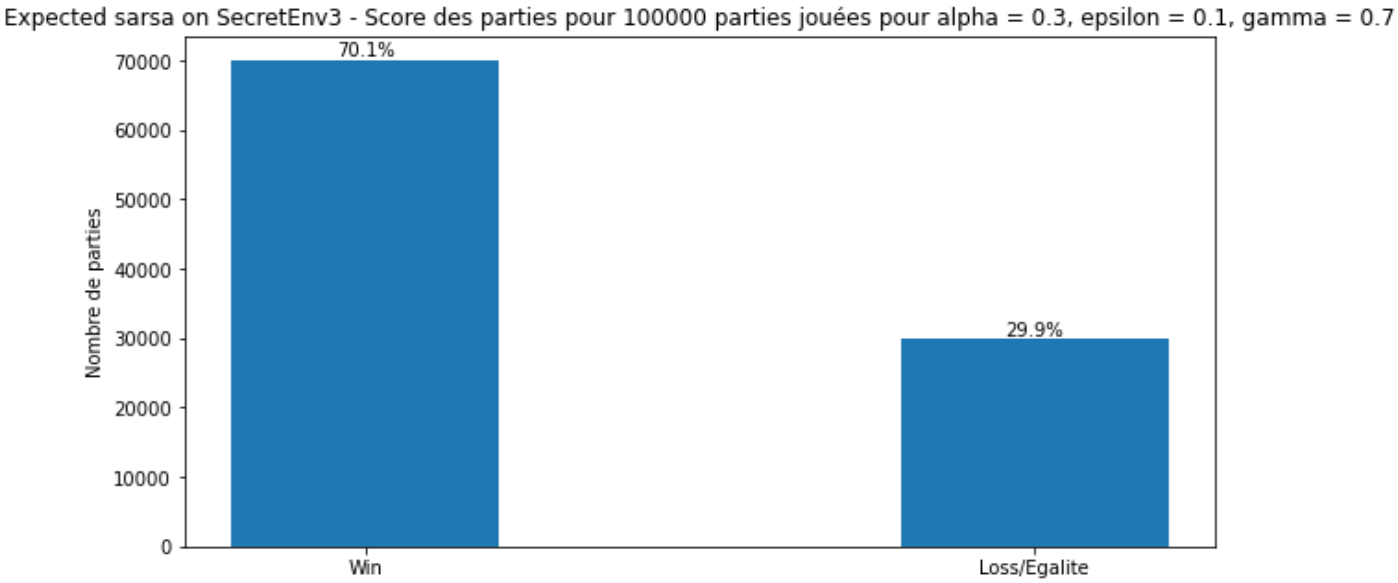
Expected Sarsa

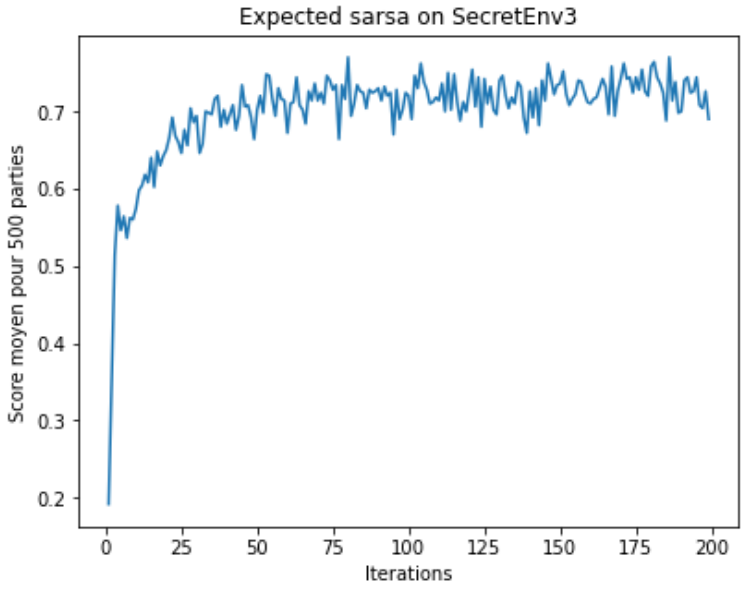
Nous avons testé L'algo Expected Sarsa sur l'env SecretEnv3 avec un alpha de 0.3, un epsilon de 0.1 et un gamma de 0.7

Nous avons obtenu les résultats suivants pour 100 000 parties jouées :

- **70.1% de parties gagnées**
- **29.9% de parties perdues ou égalités**

Voici les graphiques :





Deep reinforcement learning

TIC TAC TOE

Episodic semi gradient sarsa

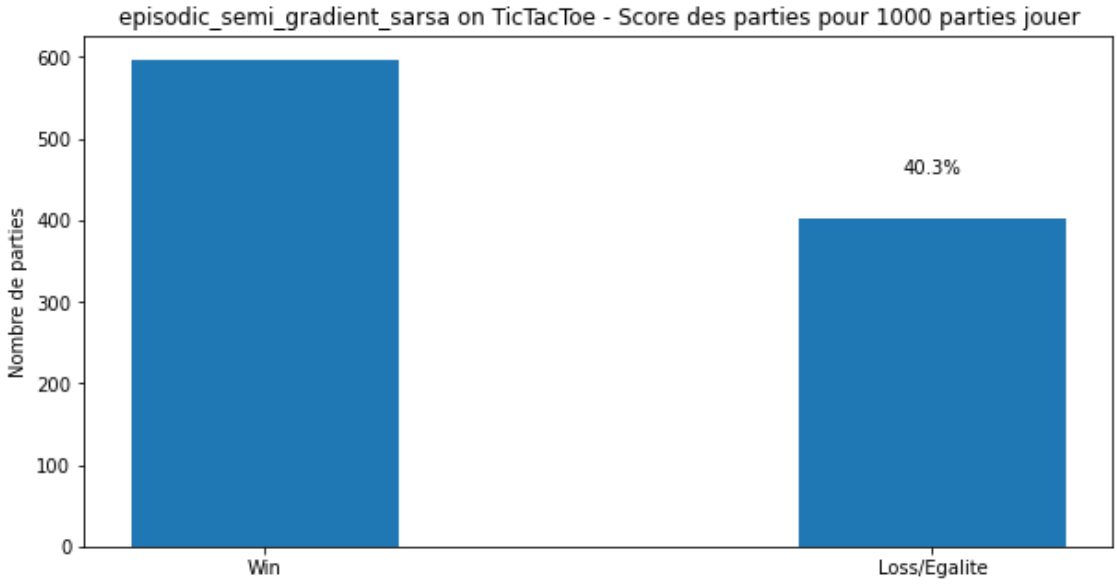
Nous avons testé L'algo Episodic semi gradient sarsa sur l'env TIC TAC TOE avec un epsilon de 0.1 et un gamma de 0.9.

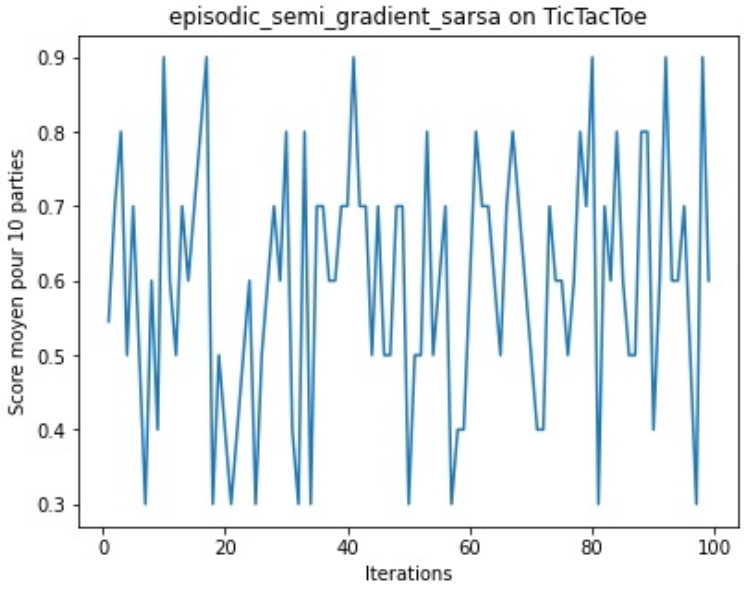
Nous avons créé un modèle avec 6 couches Dense de : 16, 64, 128, 256, 128, 1 couches du modèle

Nous avons obtenu les résultats suivants pour 1000 parties jouées :

- **59.7% de parties gagnées**
- **40.3% de parties perdues ou égalités**

Voici les graphiques :



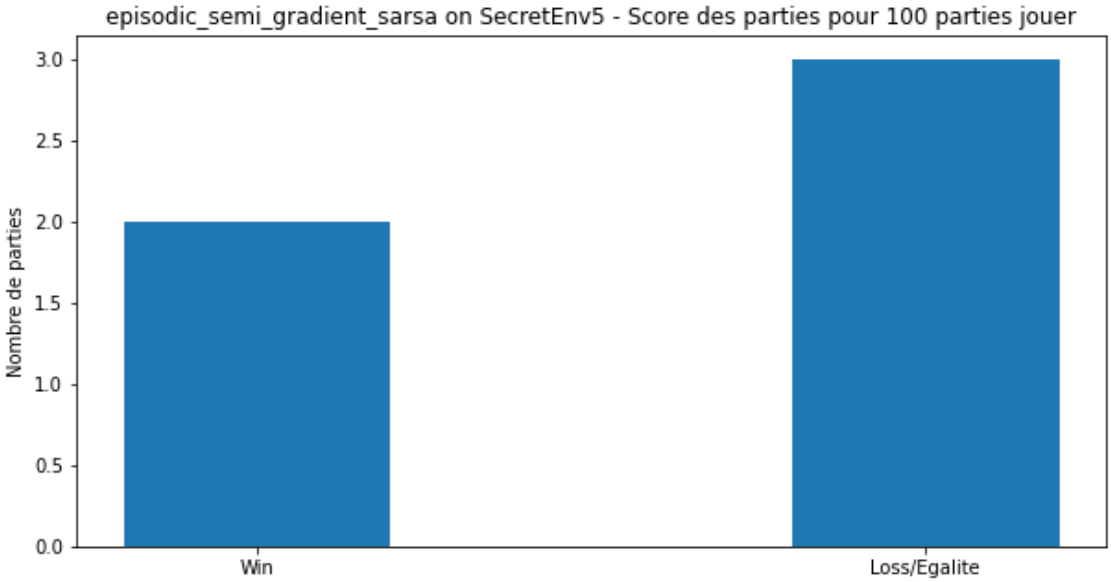


SECRET ENV 5

Episodic semi gradient sarsa

Nous avons testé L'algo Episodic semi gradient sarsa sur l'env TIC TAC TOE avec un epsilon de 0.1 et un gamma de 0.9.

Nous avons créé un modèle avec 6 couches Dense de : 16, 64, 128, 256, 128, 1 couches du modèle


episodic_semi_gradient_sarsa on SecretEnv5 - Score des parties pour 100 parties jouer
Nombre de parties
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Win
Loss/Egalite


episodic_semi_gradient_sarsa on SecretEnv5
Score moyen pour 10 parties
50
40
30
20
10
0
−10
1
2
3
4
5
6
7
8
9
Iterations