Introduction. Après les simulations de l'agent en utilisant le modèle PPO avec un acteur-critique basé sur un réseau de neurones imité de celui utilisé en TDQN pour la politique, celui-ci est clairement moins performant que le modèle TDQN, bien qu'on ait un temps d'apprentissage moins important, mais les résultats produits par la méthode PPO sont de loin d'être satisfaisants. En effet l'agent sous TDQN arrivait quand même à générer des bénéfices, mais sous PPO, il perdait toutes les ressources. Par conséquent, pour conserver les avantages du TDQN et pour introduire une fonctionnalité mémoire à long terme dans l'apprentissage afin de permettre à l'agent d'être capable de se souvenir et d'accorder un poids important à l'éventuel saisonnalité du comportement des cours boursiers selon les périodes de l'année et entre les années, un réseau de neurones LSTM est introduit à TDQN. De ce fait, dans ce qui suit, nous allons discuter de la méthode LSTM sur TDQN en général, de ses applications en *trading* TDRQN, des avantages et des inconvénients par rapport à un réseau de neurones classique, puis nous présenterons les stratégies d'améliorations ainsi que les analyses sur les résultats obtenus. Le contenu à cette étape est composé de plusieurs extensions que nous avons séparées en sujet avancé dont nous allons en discuter dans ce rapport de manière concise et dont les détails seront en annexes (rapport étape 4 : sujet avancé produit en parallèle). Les extensions appliquées sont les suivantes, l'introduction d'un réseau GRU dans un modèle TDQN, l'introduction d'un réseau LSTM dans un modèle PPO, l'introduction d'un réseau convolutif dans TDQN, l'introduction d'un réseau GRU dans un modèle PPO, l'intégration d'un contexte (par exemple prix de l'or, du pétrole, etc.) dans l'environnement.

L'état de l'art. Les réseaux récurrents sont une classe de réseaux de neurones qui ont la capacité de prendre en compte des séquences de données de taille variable, en utilisant des neurones qui se rappellent de leur état précédent. La clé de ce fonctionnement est que le réseau récurrent utilise le même ensemble de poids pour chaque étape de la séquence. Cela permet au réseau de généraliser et de traiter des séquences de longueurs différentes. De plus, le réseau récurrent peut apprendre à reconnaître des motifs à travers la séquence, qui peuvent être utilisés pour prédire la prochaine lettre.

Cependant, les réseaux récurrents traditionnels ont des limites, notamment en termes de problèmes de vanishing gradient, où les gradients de la fonction de coût diminuent à mesure que l'on remonte dans le temps, ce qui rend l'apprentissage difficile. Les réseaux récurrents, tels que les LSTM et les GRU, ont été développés pour résoudre ces problèmes.

Les LSTM sont une variante des réseaux récurrents qui ont été développés par Hochreiter et al. pour résoudre les problèmes de vanishing gradient et pour améliorer la capacité de mémorisation des réseaux récurrents.

Ils fonctionnent en utilisant des cellules de mémoire qui permettent au réseau de stocker des informations à long terme. Chaque cellule de mémoire est contrôlée par des portes, qui sont des vecteurs de poids qui déterminent quelle information est stockée, effacée ou récupérée de la cellule de mémoire. Les portes sont calculées à partir des entrées du réseau, ainsi que de l'état caché précédent.

Le fonctionnement des LSTM peut être divisé en trois étapes : l'étape d'oubli, l'étape d'entrée et l'étape de sortie. Pendant l'étape d'oubli, la porte d'oubli détermine quelle information doit être effacée de la cellule de mémoire. Pendant l'étape d'entrée, la porte d'entrée détermine quelle information doit être ajoutée à la cellule de mémoire. Enfin, pendant l'étape de sortie, la porte de sortie détermine quelle information doit être transmise en tant qu'état caché.

Ils ont été utilisés avec succès dans de nombreux domaines, notamment la reconnaissance de la parole, la traduction automatique, la génération de texte et le trading. En *trading*, les LSTM ont été utilisés pour prédire les tendances du marché, ainsi que pour détecter les anomalies et les changements de régime. Ils peuvent être utilisés pour prédire les tendances du marché en utilisant les séquences temporelles précédentes pour prédire la séquence suivante. Les prédictions peuvent être basées sur la séquence de sortie, qui peut être la valeur du marché dans le futur, ou sur une séquence binaire qui indique si le marché va augmenter ou diminuer. Enfin, les LSTM peuvent être combinées avec d'autres techniques d'apprentissage automatique, comme les réseaux de neurones convolutifs ou les réseaux de neurones profonds, pour améliorer la précision des prévisions.

LSTM se sont révélés efficaces pour modéliser des données séquentielles telles que les cours des actions, ce qui les rend bien adaptés aux tâches d'analyse et de prédiction de séries chronologiques telles que les prévisions boursières. Cela signifie qu'ils peuvent être utilisés pour traiter non seulement les données de prix historiques d'une action, mais également d'autres facteurs dépendant du temps tels que le volume des transactions, les taux d'intérêt, l'écart de swap sur défaillance de crédit et d'autres indicateurs financiers.

L'utilisation de LSTM avec des algorithmes de RL tels que PPO et TDQN peut aider à améliorer les performances de *trading*. Ils peuvent aider à capturer les informations temporelles dans les données de marché, telles que les tendances à long terme, et prendre en compte l'historique des données de marché et pour prendre des décisions plus informées. En outre, les algorithmes de RL peuvent apprendre à ajuster les paramètres de trading en fonction des conditions de marché, ce qui peut aider à améliorer les performances en temps réel. En incorporant LSTM dans un modèle d'apprentissage par renforcement, le système peut apprendre à prédire les cours futurs des actions en fonction des données passées, puis utiliser ces informations pour prendre des décisions. Le système peut potentiellement acquérir une compréhension plus complète du marché.

Par exemple, des volumes de transactions élevés peuvent suggérer une activité et une volatilité accrues du marché, tandis que les taux d'intérêt peuvent affecter les coûts d'emprunt et avoir un impact sur la performance de certaines industries ou secteurs. Cela permet au système de prendre en compte les dépendances temporelles des données et d'ajuster ses actions en fonction de l'évolution des conditions du marché.

Méthodologie. Afin d'intégrer de nouvelles méthodes d'apprentissage plus simplement et rendre notre code ajustable plus simplement, nous avons restructuré les classes de notre projet à partir du code initial. Un énorme travail de restructuration des classes a été réalisé, ce travail est nécessaire et inévitable car la structure du code source du projet initial n'était pas adaptée à ajout méthodique des nouveaux modèles. En particulier, une classe mère **DRLAgent** a été implémentée, toutes les classes filles d'un modèle d'apprentissage tels que DQN, PPO, etc. héritent de cette classe. Par la suite, elle facilitera l'ajout de TDRQN (*Trading Deep Recursive Q-Learning Network*) et TDCQN (*Trading Deep Convolutional Q-Learning Network*). Un dossier **Configurations** regroupant toutes les paramètres des modèles a été ajouté pour simplifier l'ajustement des hyperparamètres lors de l'apprentissage. Des fonctionnalités qui introduisent un contexte supplémentaire dans l'environnement ont été implémentés, en conséquence les classes ont été adaptées (cette partie sera traitée plus en détail dans le sujet avancé en annexes).

Pour introduire un réseau LSTM dans le modèle TDQN, nous avons repris la structure initiale à l'étape 1 du projet, et nous avons enrichi le double DQN d'un bloc LSTM à la fin, donnant la classe **TDRQN**. Ceci permet, comme expliqué précédemment, l'ajout d'une propriété de mémoire à long terme et à court terme dans l'apprentissage de la politique en jouant avec les prédictions des Q-values en se basant sur les informations de l'environnement et les estimations du futur. De même, ce réseau LSTM est appliqué au début de notre **ActorCritic**, comme PPO a donné des résultats peu satisfaisants, en essayant d'appliquer LSTM au début, on aurait éventuellement une meilleure extraction des caractéristiques, cela ajouterait la particularité de mémorisation des informations avant d'entrer dans notre acteur-critique.

Dans le but d'étendre notre projet vers des sujets avancés, nous avons introduit divers réseaux et fonctionnalités comme :

- Bloc GRU, un réseau récurrent, dans DQN
- Bloc Linéaire, un réseau linéaire pleinement connecté, dans DQN
- Bloc Attention, un réseau basé sur l'attention, dans DQN
- Bloc GRU, un réseau récurrent dans PPO
- Contexte supplémentaire (par exemple : l'ajout de prix de l'or, de ressources premières, etc.)

Ces extensions seront détaillées dans un annexe pour sujet avancée en complément de ce rapport.

Dans le but de simplifier l'exécution des simulations et l'affichage des résultats, nous avons produit un JupyterNotebook pour chaque type d'apprentissage dans la branche principale *main* notre dépôt git de projet, en particulier, l'application de méthode classique, de TDQN, de PPO, de TDRQN, de TDCQN, respectivement avec et sans contexte.

Expérimentations. Le *bot* a été lancé dans les mêmes conditions que l'étape précédente, sur les actions d'Apple et de Tesla en considérant 30 jours pour la taille des états sur 6 ans d'apprentissage et 2 ans de simulation en phase test.

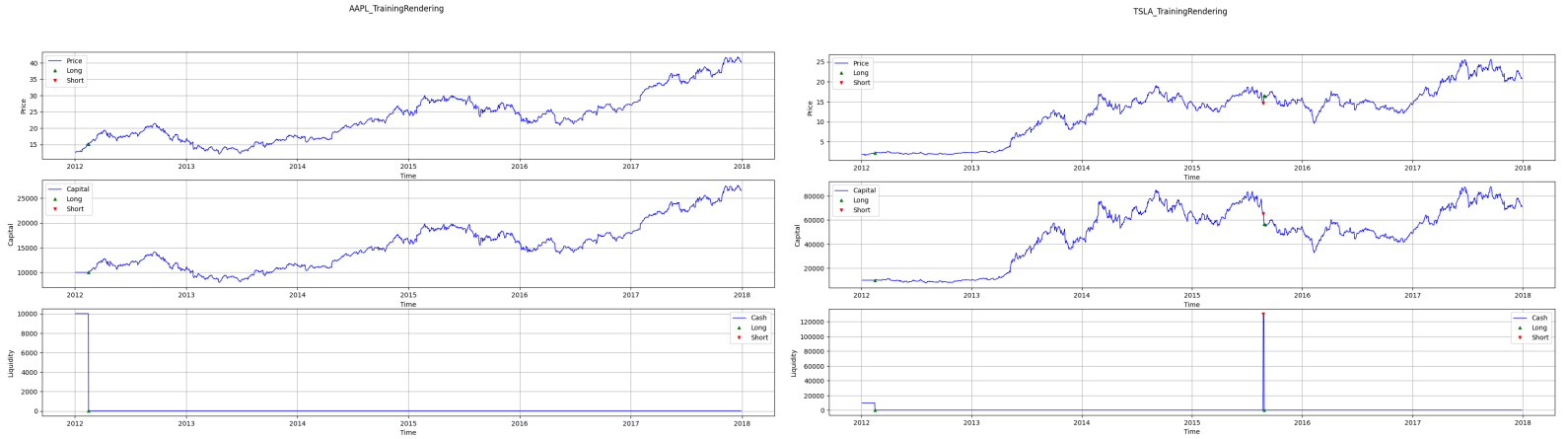

Figure 1 - Apprentissage de l'agent RL (LSTM + TDQN) entre 2012 et 2017 sur les cours boursiers d'Apple (gauche) et de Tesla (droite)

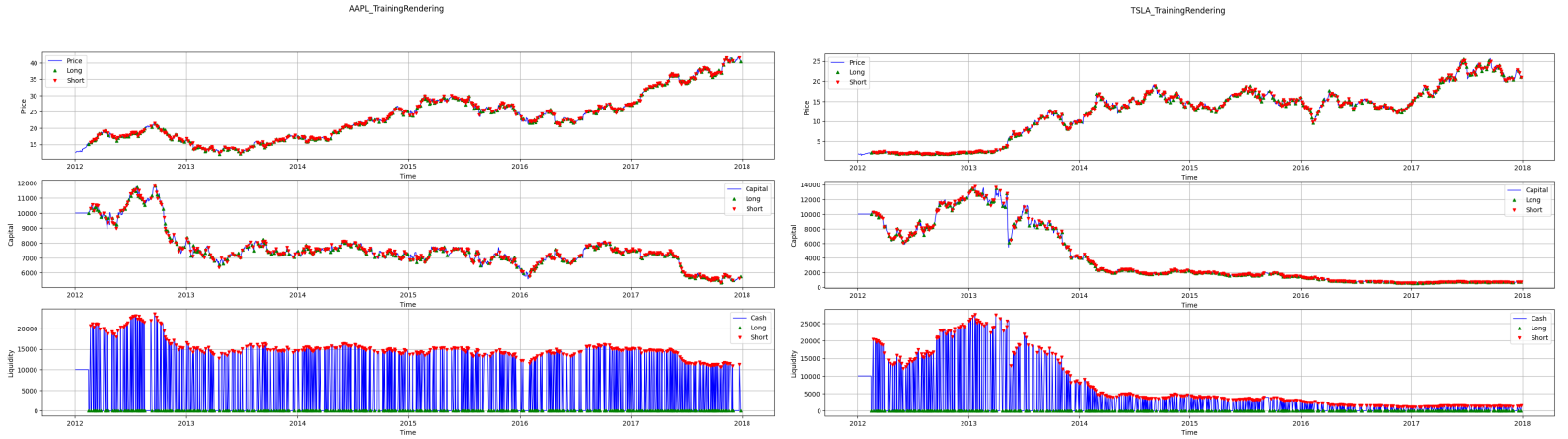

Figure 2 - Apprentissage de l'agent RL (LSTM + PPO) entre 2012 et 2017 sur les cours boursiers d'Apple (gauche) et de Tesla (droite)

En figures 1 et 2, nous avons respectivement les courbes d'apprentissages de LSTM dans TDRQN et LSTM dans PPO du *bot* sur les actions d'Apple et de Tesla. L'apprentissage diffère beaucoup, on remarque qu'étrangement le modèle TDRQN avec un réseau LSTM ne prend qu'une décision alors que le modèle PPO avec un réseau LSTM reste sur une prise de décision assez fréquente, mais la décision reste assez correcte vis-à-vis du gain obtenu à la fin. En figure 3 et 4, nous avons les métriques de performance, en particulier le ratio de Sharpe, en phase d'entraînement et de test, celui-ci fluctue pas mal quelque soit la phase, et nous remarquons que le Sharpe ratio en modèle LSTM dans TDRQN, en entraînement est en moyenne inférieur à celui de test, alors que pour le modèle LSTM dans PPO, le ratio en phase *train* et test se rapproche plus, ceci est probablement dû à la stabilité du modèle PPO sur les types DQN.

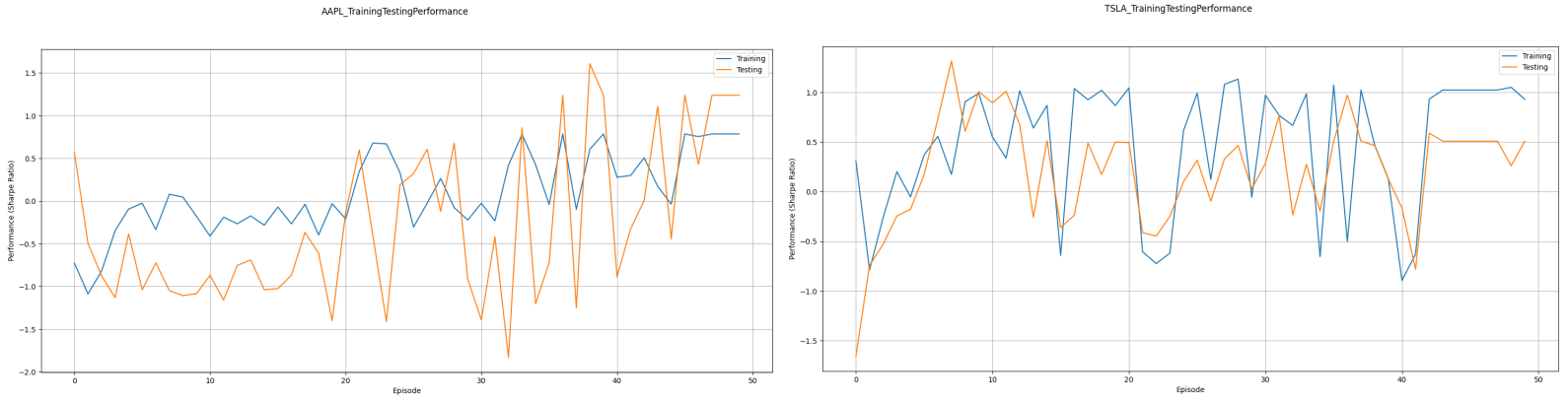

Figure 3 - Courbes d'apprentissage de l'agent (TDQN + LSTM) sur les *stocks* d'Apple (gauche) et de Tesla (droite), *training* (bleu) et *testing* (jaune)

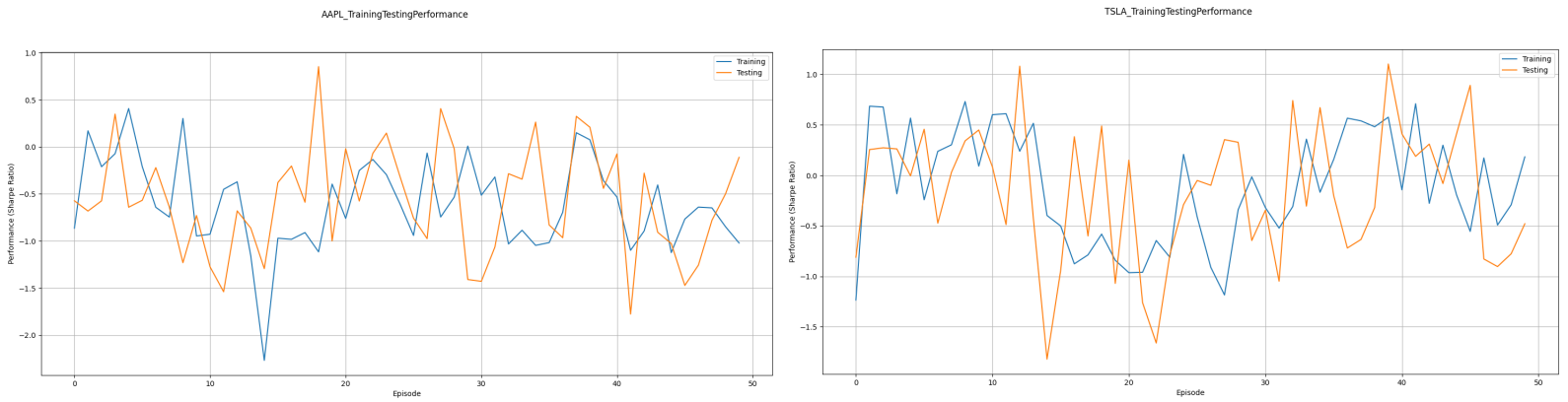

Figure 4 - Courbes d'apprentissage de l'agent (PPO + LSTM) sur les *stocks* d'Apple (gauche) et de Tesla (droite), *training* (bleu) et *testing* (jaune)

En figure 5 et 6 se trouvent les courbes de récompenses du système selon les modèles TDQN avec LSTM et PPO avec LSTM. Pour TDQN + LSTM, cn entraînement, les valeurs tendent vers des récompenses positives croissantes avec quelques fluctuations, cependant, pour PPO + LSTM, elles ont une tendance constante et conservent des fluctuations.

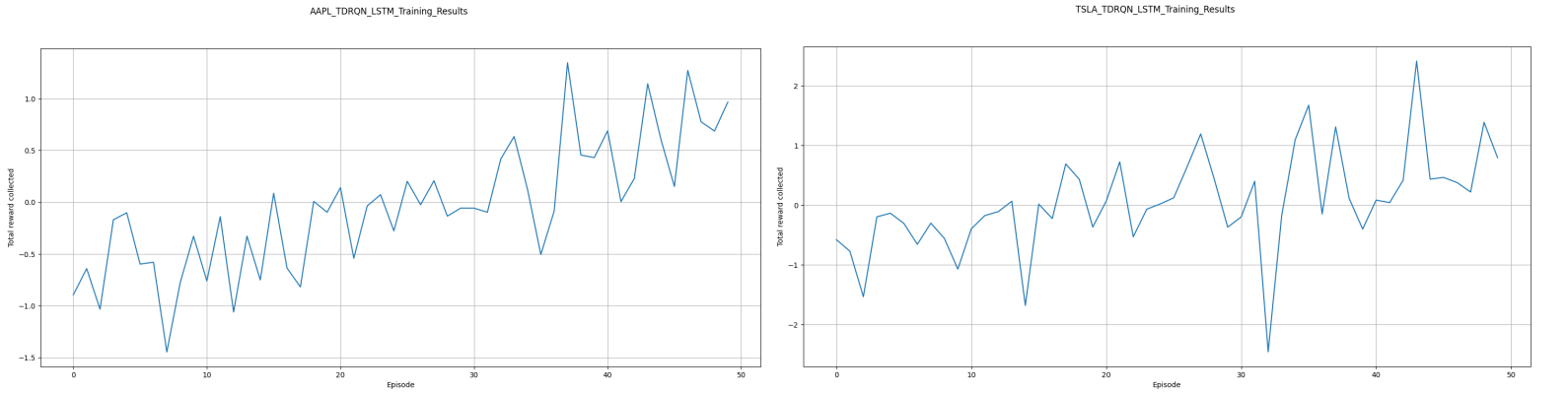

Figure 5 - Courbes de récompenses du système (TDQN + LSTM) sur les *stocks* d'Apple (gauche) et de Tesla (droite)

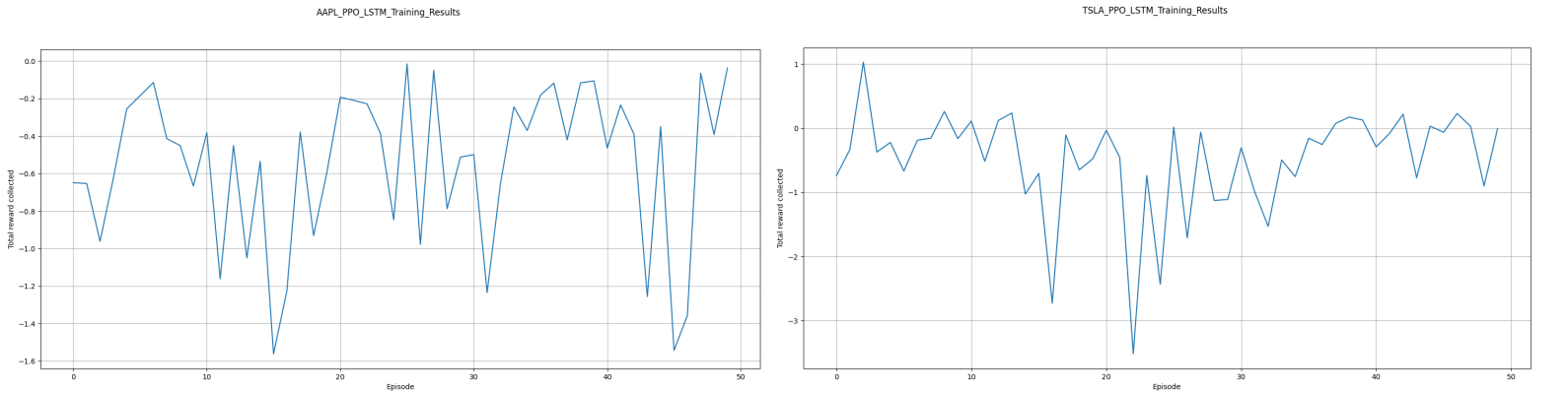

Figure 6 - Courbes de récompenses du système (PPO + LSTM) sur les *stocks* d'Apple (gauche) et de Tesla (droite)

Performance Indicator	TDRQN_LSTM (Training)
Profit & Loss (P&L)	16517
Annualized Return	13.69%
Annualized Volatility	24.58%
Sharpe Ratio	0.786
Sortino Ratio	1.074
Maximum Drawdown	43.78%
Maximum Drawdown Duration	144 days
Profitability	100.00%
Ratio Average Profit/Loss	inf
Skewness	-0.204

Performance Indicator	PPO_LSTM (Training)
Profit & Loss (P&L)	-4285
Annualized Return	-7.63%
Annualized Volatility	24.66%
Sharpe Ratio	-0.256
Sortino Ratio	-0.379
Maximum Drawdown	54.79%
Maximum Drawdown Duration	1279 days
Profitability	45.94%
Ratio Average Profit/Loss	1.060
Skewness	0.525

Performance Indicator	TDRQN_LSTM (Training)
Profit & Loss (P&L)	61055
Annualized Return	23.94%
Annualized Volatility	47.10%
Sharpe Ratio	0.928
Sortino Ratio	1.405
Maximum Drawdown	61.44%
Maximum Drawdown Duration	361 days
Profitability	66.67%
Ratio Average Profit/Loss	4.062
Skewness	0.713

Performance Indicator	PPO_LSTM (Training)
Profit & Loss (P&L)	-9254
Annualized Return	-100.00%
Annualized Volatility	49.43%
Sharpe Ratio	-0.621
Sortino Ratio	-0.762
Maximum Drawdown	95.96%
Maximum Drawdown Duration	973 days
Profitability	49.27%
Ratio Average Profit/Loss	0.845
Skewness	-0.989

Figure 7 - Tableaux des métriques de performance en phase d'entraînement du *bot* (LSTM dans TDQN et PPO) sur les *stocks* d'Apple (gauche) et de Tesla (droite)

En figure 7, nous avons les métriques concernant les phases d'entraînement de nos modèles LSTM avec TDQN et PPO. On remarque clairement qu'en entraînement l'agent s'en sort bien avec le modèle TDQN + LSTM comparé à PPO + LSTM dont on voit beaucoup plus de perte. On note également que la volatilité de Tesla (les deux tableaux à droite) permet au modèle de soit gagner davantage, soit de perdre davantage, selon les performances d'apprentissage. En figure 8 et 9, nous avons les courbes de valeurs du portfolio de l'agent, comme en phase d'entraînement, pour le modèle TDQN + LSTM, l'agent ne prend qu'une décision sur toute la période d'observation alors que le modèle PPO + LSTM en prend de manière plus fréquente, ce comportement n'a pas pu être expliqué.

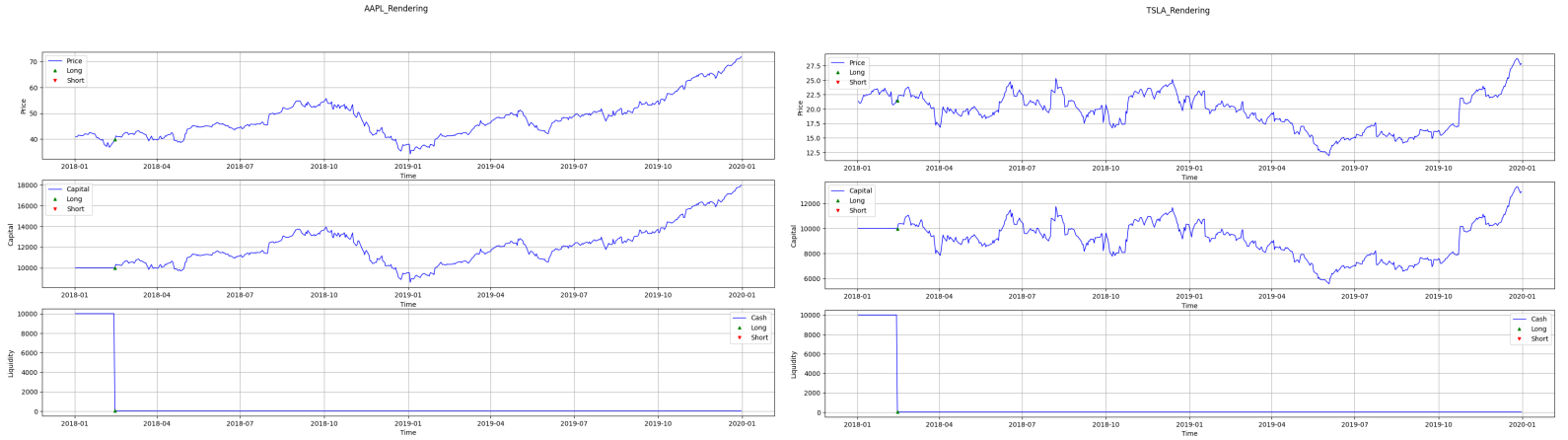

Figure 8 - Application de l'agent RL (TDQN + LSTM) entre 2018 et 2019 sur les cours boursiers d'Apple (gauche) et de Tesla (droite)

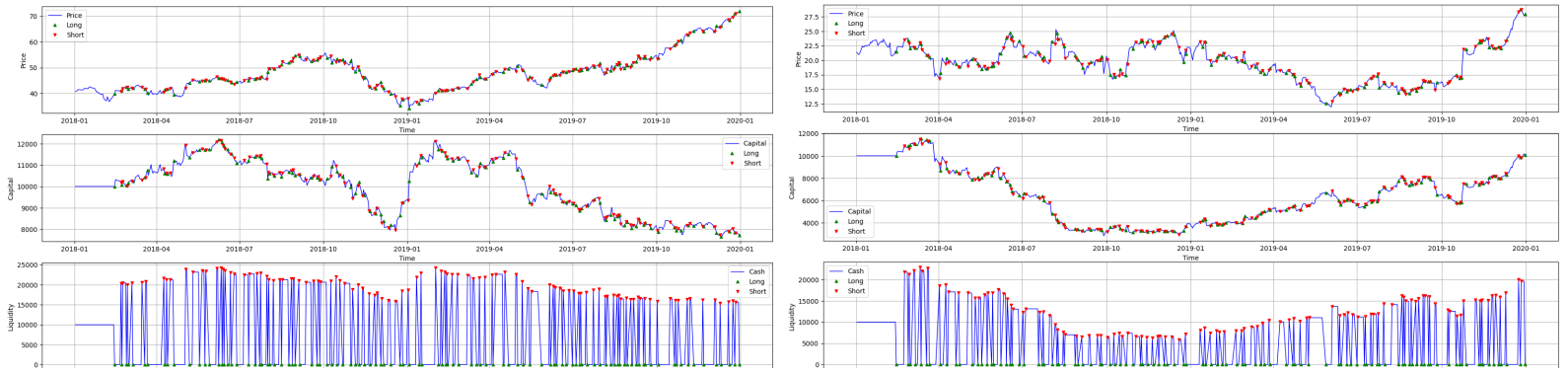

Figure 9 - Application de l'agent RL (PPO + LSTM) entre 2018 et 2019 sur les *stocks* d'Apple (gauche) et de Tesla (droite)

En comparant les résultats d'entraînement en figure 7 et les résultats de test en figure 10, nous obtenons à peu près les mêmes conclusions que l'analyse en phase d'entraînement, en effet le modèle TDQN + LSTM performe mieux que PPO + LSTM que cela soit sur les cours volatiles ou non. Notons toutefois qu'en phase de test le modèle PPO + LSTM a moins perdu que durant la phase d'entraînement. Sinon, dans la généralité, les deux modèles ont respectivement moins gagné et moins perdu, ceci est certainement dû à la durée de la période de simulation, entraînement sur six années et application sur deux années. De plus, nous avons les mêmes conclusions que l'étape 1 sur le modèle TDQN, le TDQN + LSTM est beaucoup moins risqué sur Apple que sur Tesla, Apple étant moins volatile. On le voit en examinant les ratios de Sharpe (1.239 contre 0.507) et de Shortino (1.558 et 0.741) qui prend en compte le facteur de variance des cours boursiers. On remarque qu'il y a une profitabilité de 100% peu importe le cours boursier pour le modèle TDQN + LSTM.

Performance Indicator	TDRQN_LSTM (Testing)
Profit & Loss (P&L)	7971
Annualized Return	28.82%
Annualized Volatility	26.58%
Sharpe Ratio	1.239
Sortino Ratio	1.558
Maximum Drawdown	38.46%
Maximum Drawdown Duration	62 days
Profitability	100.00%
Ratio Average Profit/Loss	inf
Skewness	-0.475

Performance Indicator	PPO_LSTM (Testing)
Profit & Loss (P&L)	-2291
Annualized Return	-9.98%
Annualized Volatility	26.74%
Sharpe Ratio	-0.354
Sortino Ratio	-0.508
Maximum Drawdown	37.26%
Maximum Drawdown Duration	380 days
Profitability	45.54%
Ratio Average Profit/Loss	1.054
Skewness	0.288

Performance Indicator	TDRQN_LSTM (Testing)
Profit & Loss (P&L)	2960
Annualized Return	24.06%
Annualized Volatility	53.05%
Sharpe Ratio	0.507
Sortino Ratio	0.741
Maximum Drawdown	52.76%
Maximum Drawdown Duration	205 days
Profitability	100.00%
Ratio Average Profit/Loss	inf
Skewness	0.542

Performance Indicator	PPO_LSTM (Testing)
Profit & Loss (P&L)	69
Annualized Return	13.43%
Annualized Volatility	52.93%
Sharpe Ratio	0.270
Sortino Ratio	0.373
Maximum Drawdown	75.60%
Maximum Drawdown Duration	139 days
Profitability	52.11%
Ratio Average Profit/Loss	0.922
Skewness	0.204

Figure 10 - Tableaux des métriques de performance en phase de test du *bot* (LSTM avec TDQN et PPO) sur les *stocks* d'Apple (gauche) et de Tesla (droite)

En ajoutant les résultats sur les modèles TDQN et PPO simples, en figure 11 et 12, nous remarquons que le modèle TDQN enrichie du réseau LSTM performe mieux sur le cours boursier Apple comparé au modèle TDQN de base (abordé à l'étape 1), cependant pour le cours boursier Tesla qui est plus volatile, TDQN + LSTM ne devance pas TDQN simple. Pour le modèle basé sur PPO + LSTM, les résultats restent assez médiocres, mais celui-ci serait quand même mieux que PPO simple, en particulier, si nous regardons le *Annualized Return*, nous avons -9.98% (PPO + LSTM) contre -22.07% (PPO simple) pour Apple, et 13.43% (PPO + LSTM) contre -100% (PPO simple). Nous pouvons tirer les mêmes conclusions à partir des chiffres sur le profit. Concernant les ratios de Sharpe et de Sortino, ils sont en général assez bas et dans le négatif, sauf pour PPO + LSTM appliqué à Tesla qui est légèrement positif mais reste loin des performances de TDQN + LSTM.

Performance Indicator	TDQN (Training)
Profit & Loss (P&L)	2609
Annualized Return	5.96%
Annualized Volatility	24.68%
Sharpe Ratio	0.281
Sortino Ratio	0.373
Maximum Drawdown	48.51%
Maximum Drawdown Duration	145 days
Profitability	47.54%
Ratio Average Profit/Loss	1.187
Skewness	-0.335

Performance Indicator	TDQN (Testing)
Profit & Loss (P&L)	6288
Annualized Return	24.91%
Annualized Volatility	26.50%
Sharpe Ratio	1.056
Sortino Ratio	1.323
Maximum Drawdown	44.20%
Maximum Drawdown Duration	62 days
Profitability	57.89%
Ratio Average Profit/Loss	1.338
Skewness	-0.463

Performance Indicator	TDQN (Training)
Profit & Loss (P&L)	378334
Annualized Return	32.10%
Annualized Volatility	46.50%
Sharpe Ratio	1.545
Sortino Ratio	2.419
Maximum Drawdown	24.16%
Maximum Drawdown Duration	15 days
Profitability	47.86%
Ratio Average Profit/Loss	2.750
Skewness	0.818

Performance Indicator	TDQN (Testing)
Profit & Loss (P&L)	6873
Annualized Return	33.89%
Annualized Volatility	51.77%
Sharpe Ratio	0.764
Sortino Ratio	1.084
Maximum Drawdown	40.84%
Maximum Drawdown Duration	90 days
Profitability	50.82%
Ratio Average Profit/Loss	1.396
Skewness	0.252

Figure 11 - Tableaux des métriques de performance en phase d'entraînement et de test du *bot* (TDQN) sur les *stocks* d'Apple (gauche) et de Tesla (droite)

Performance Indicator	PPO (Training)
Profit & Loss (P&L)	-6059
Annualized Return	-20.59%
Annualized Volatility	24.64%
Sharpe Ratio	-0.507
Sortino Ratio	-0.721
Maximum Drawdown	70.73%
Maximum Drawdown Duration	1023 days
Profitability	44.90%
Ratio Average Profit/Loss	0.979
Skewness	-0.042

Performance Indicator	PPO (Testing)
Profit & Loss (P&L)	-3712
Annualized Return	-22.07%
Annualized Volatility	26.71%
Sharpe Ratio	-0.735
Sortino Ratio	-0.874
Maximum Drawdown	54.75%
Maximum Drawdown Duration	250 days
Profitability	51.10%
Ratio Average Profit/Loss	0.727
Skewness	-0.926

Performance Indicator	PPO (Training)
Profit & Loss (P&L)	-9332
Annualized Return	-100.00%
Annualized Volatility	47.33%
Sharpe Ratio	-0.717
Sortino Ratio	-0.970
Maximum Drawdown	94.65%
Maximum Drawdown Duration	1060 days
Profitability	44.42%
Ratio Average Profit/Loss	0.857
Skewness	0.117

Performance Indicator	PPO (Testing)
Profit & Loss (P&L)	-8221
Annualized Return	-100.00%
Annualized Volatility	53.50%
Sharpe Ratio	-1.340
Sortino Ratio	-1.626
Maximum Drawdown	89.20%
Maximum Drawdown Duration	380 days
Profitability	46.25%
Ratio Average Profit/Loss	0.787
Skewness	-0.872

Figure 12 - Tableaux des métriques de performance en phase d'entraînement et de test du *bot* (PPO) sur les *stocks* d'Apple (gauche) et de Tesla (droite)

En intégrant les résultats des modèles classiques, en figure 13 et 14, nous voyons que le modèle TDQN + LSTM est comparable aux performances des stratégies de *trading* classique (hors IA), en particulier B&H et MATF pour le cours d'Apple. De même, pour le cours de Tesla, TDQN + LSTM a des retours de gain équivalents à la stratégie classique B&H et est mieux que les autres stratégies. Pour obtenir un rappel des stratégies classiques et des abréviations utilisées, se référer aux rapports précédents.

Performance Indicator	B&H (Testing)
Profit & Loss (P&L)	7961
Annualized Return	28.80%
Annualized Volatility	26.55%
Sharpe Ratio	1.239
Sortino Ratio	1.559
Maximum Drawdown	38.43%
Maximum Drawdown Duration	62 days
Profitability	100.00%
Ratio Average Profit/Loss	inf
Skewness	-0.476

Performance Indicator	S&H (Testing)
Profit & Loss (P&L)	-7981
Annualized Return	-100.00%
Annualized Volatility	44.17%
Sharpe Ratio	-1.591
Sortino Ratio	-2.204
Maximum Drawdown	82.28%
Maximum Drawdown Duration	250 days
Profitability	0.00%
Ratio Average Profit/Loss	0.000
Skewness	0.145

Performance Indicator	MATF (Testing)
Profit & Loss (P&L)	6856
Annualized Return	25.92%
Annualized Volatility	24.82%
Sharpe Ratio	1.178
Sortino Ratio	1.801
Maximum Drawdown	14.86%
Maximum Drawdown Duration	20 days
Profitability	42.31%
Ratio Average Profit/Loss	3.181
Skewness	0.404

Performance Indicator	MAMR (Testing)
Profit & Loss (P&L)	-3460
Annualized Return	-19.08%
Annualized Volatility	28.28%
Sharpe Ratio	-0.610
Sortino Ratio	-0.813
Maximum Drawdown	51.09%
Maximum Drawdown Duration	204 days
Profitability	56.67%
Ratio Average Profit/Loss	0.492
Skewness	-0.291

Figure 13 - Evaluation de performance en phase de test des méthodes classiques pour Apple

Performance Indicator	B&H (Testing)
Profit & Loss (P&L)	2960
Annualized Return	24.06%
Annualized Volatility	53.05%
Sharpe Ratio	0.507
Sortino Ratio	0.741
Maximum Drawdown	52.76%
Maximum Drawdown Duration	205 days
Profitability	100.00%
Ratio Average Profit/Loss	inf
Skewness	0.542

Performance Indicator	S&H (Testing)
Profit & Loss (P&L)	-2980
Annualized Return	-7.38%
Annualized Volatility	46.04%
Sharpe Ratio	-0.154
Sortino Ratio	-0.205
Maximum Drawdown	54.04%
Maximum Drawdown Duration	144 days
Profitability	0.00%
Ratio Average Profit/Loss	0.000
Skewness	-0.024

Performance Indicator	MATF (Testing)
Profit & Loss (P&L)	-7327
Annualized Return	-100.00%
Annualized Volatility	52.61%
Sharpe Ratio	-0.989
Sortino Ratio	-1.231
Maximum Drawdown	79.85%
Maximum Drawdown Duration	229 days
Profitability	34.38%
Ratio Average Profit/Loss	0.533
Skewness	-0.309

Performance Indicator	MAMR (Testing)
Profit & Loss (P&L)	854
Annualized Return	18.99%
Annualized Volatility	58.03%
Sharpe Ratio	0.358
Sortino Ratio	0.538
Maximum Drawdown	65.30%
Maximum Drawdown Duration	159 days
Profitability	67.65%
Ratio Average Profit/Loss	0.496
Skewness	0.550

Figure 14 - Evaluation de performance en phase de test des méthodes classiques pour Tesla

La solution d'une amélioration par un réseau LSTM pour intégrer une fonctionnalité mémoire permettant d'ajuster les poids des données lors de l'apprentissage a globalement donné des résultats satisfaisants. Pour un modèle TDQN enrichie de d'un réseau LSTM, les gains du portfolio sont comparables au TDQN simple, avec une amélioration sur les cours boursiers stables (peu volatiles). Le modèle performe de manière comparable voire parfois mieux que les méthodes classiques de *trading*. Cependant, pour le modèle PPO enrichi de d'un réseau LSTM, nous avons une grande amélioration comparé au modèle PPO simple, l'agent amélioré performe globalement mieux comparé aux stratégies classiques, mais n'excelle par le modèle TDQN + LSTM. Nous avons vu que la volatilité pouvait être un avantage pour l'agent s'il est entraîné à profiter de ces tendances, cependant dans certains cas, comme PPO + LSTM, l'agent n'est pas capable d'utiliser cette opportunité, tout comme PPO simple, la volatilité dans ce modèle impacte les performances de l'agent, et empire les résultats, on pourrait croire que PPO en général n'arrive pas à gérer aussi bien la volatilité que TDQN.

Conclusion. L'objectif était d'appliquer le réseau Long Short-Term Memory dans l'apprentissage par *Trading Deep Q-Network* et par *Proximal Policy Optimization*, et de comparer les résultats à ceux obtenus aux étapes précédents, c'est-à-dire TDQN et PPO de base. Le réseau LSTM a l'avantage d'être robuste à la disparition du gradient, donc permet une architecture du réseau plus profonde, par conséquent il réalise une extraction des caractéristiques plus complexe et plus poussé. Il a également l'avantage de distribuer des importances différentes aux entrées des neurones, donc cela

permet l'ajustement du poids des informations optimisant la propagation du gradient, donc le réseau est davantage capable de conserver l'information pertinente pour la prédiction et produit un score plus proche de l'optimal. De ce fait, l'application d'un réseau LSTM pourrait améliorer le modèle existant TDQN et corriger les défauts de PPO. Ces hypothèses ont été confirmées par l'expérience réalisée, en effet lors des simulations, nous obtenons des résultats nettement mieux que les simulations précédentes. En particulier pour le modèle PPO, le réseau LSTM a permis de mémoriser les patterns ou des caractéristiques lointaines et en combinant avec les observations récentes, l'agent décide avec plus de perspective, plus de visibilité, cependant l'amélioration apportée ne devance pas ceux de TDQN + LSTM. Pour apporter plus d'éléments de preuves de nos hypothèses, nous pouvons ajouter un plus grand nombre de simulations appliquées à plusieurs cours boursiers différents et en moyennant les résultats, nous pourrons accepter ou réfuter nos hypothèses. Mais dû à un manque de ressources, en particulier vu l'exigence computationnelle des apprentissages, il faudrait envisager de faire tourner les simulations sur des serveurs à plus hautes capacités de calculs.

En parallèle à ce travail, des extensions ont été ajoutées en guise de sujet avancé (étape 4), pour pouvoir analyser l'influence de l'architecture des modèles (en particulier les réseaux utilisés dans les modèles) sur les performances du *trading* bot. L'ajout d'un réseau récurrent GRU, de réseau convolutif, pleinement connectés, de réseau basé sur l'attention et d'un contexte permettant d'introduire des informations (du marché par exemple) pour aider sur la prise de décision.