

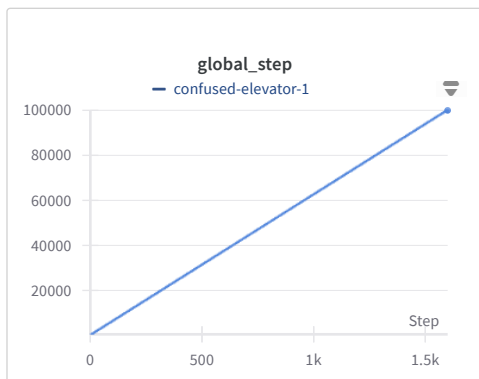


CartPole training report

Add a description...

Marouane BENFATTOUM

Section 1

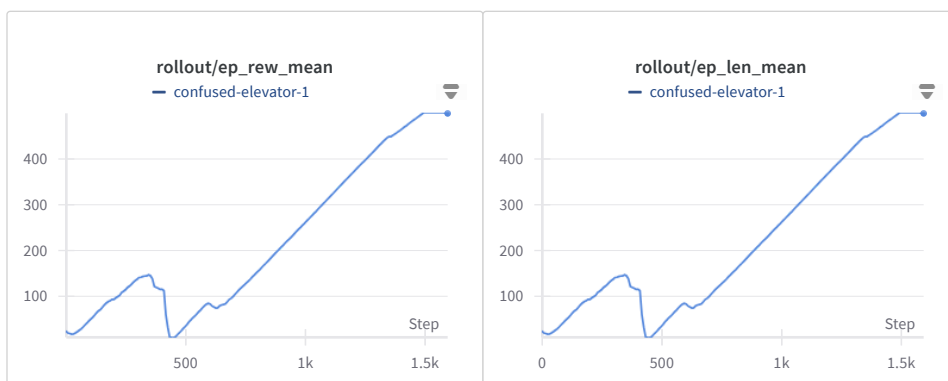


Import panel

Add panel



Ce graphique illustre l'évolution progressive du nombre d'itérations au fur et à mesure que l'apprentissage avance. On remarque une augmentation régulière des étapes d'entraînement, ce qui est logique puisque l'agent s'entraîne sur un grand nombre d'épisodes pour améliorer ses performances. Cela montre que l'apprentissage suit un déroulement constant sans interruption



Import panel

Add panel



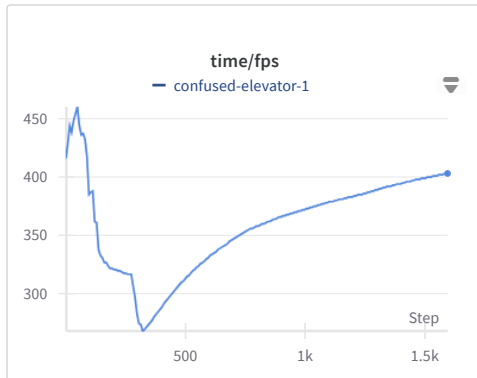
2. rollout/ep_rew_mean

Ce graphe représente la récompense moyenne obtenue par épisode durant l'apprentissage. On constate que la courbe monte de façon générale, ce qui veut dire que l'agent apprend à mieux se comporter dans

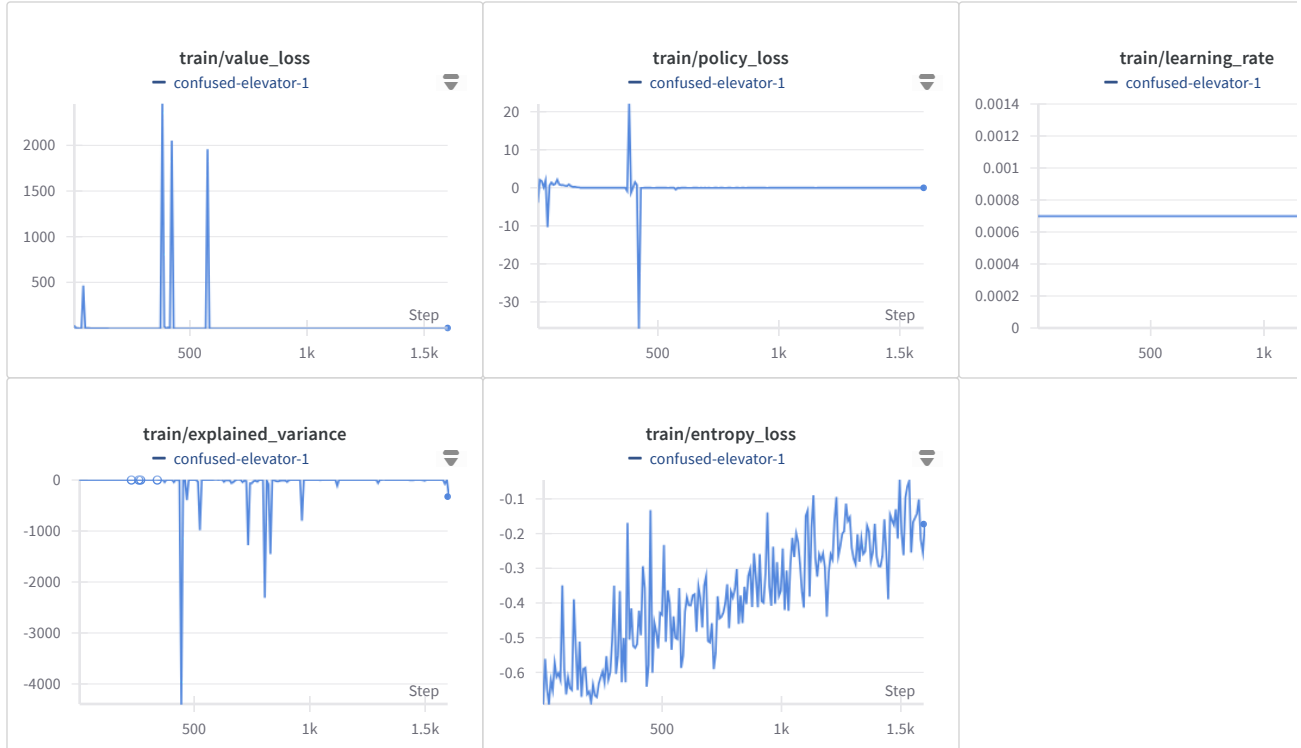
l'environnement. Il y a une petite baisse vers les 450 étapes, mais ensuite les performances s'améliorent à nouveau, ce qui est assez courant dans les phases d'apprentissage.

3. rollout/ep_len_mean

Ce graphique montre la durée moyenne des épisodes au fil du temps. On voit une tendance à l'augmentation, ce qui indique que l'agent arrive à rester plus longtemps sans échouer. C'est cohérent avec l'amélioration de la récompense moyenne, puisque plus l'agent reste stable, plus il maximise ses gains dans l'environnement CartPole.

[Import panel](#)[Add panel](#)

Le graphique montre le nombre d'images par seconde (FPS) durant l'entraînement. Au départ, la vitesse est élevée, puis elle chute fortement, probablement à cause d'une surcharge temporaire. Ensuite, le FPS remonte progressivement, signe que le modèle devient plus stable et efficace avec le temps.

[Import panel](#)[Add panel](#)

train/value_loss

Au début de l'entraînement, on observe des pics importants de la perte de valeur, atteignant parfois des niveaux très élevés. Cela montre que le modèle avait du mal à estimer correctement la valeur des états au départ. Ensuite, cette perte diminue progressivement pour se stabiliser à des valeurs proches de zéro, ce qui indique une meilleure estimation au fil de l'apprentissage.

train/policy_loss

La perte de politique présente des fluctuations marquées, avec une chute importante autour de 500 étapes, suivie d'une stabilisation. Cette évolution traduit des ajustements conséquents de la politique d'action du modèle avant qu'elle ne se régule.

train/learning_rate

Le taux d'apprentissage reste constant à environ 0.0007 durant tout l'entraînement. Cela signifie qu'il n'y a pas eu de modification dynamique du learning rate, ce qui permet de garder une certaine stabilité dans les mises à jour des paramètres.

train/explained_variance

La variance expliquée est très instable, avec des valeurs fortement négatives, atteignant parfois des extrêmes vers -4000. Cela révèle une grande difficulté du modèle à expliquer la variance des retours pendant certaines phases de l'apprentissage.

train/entropy_loss

L'entropie augmente de façon irrégulière mais globale au fil du temps. Cela montre que le modèle cherche à diversifier ses choix d'actions, favorisant l'exploration. Une entropie croissante est généralement un signe que l'agent évite de devenir trop prévisible.