



## RL : ENVIRONNEMENTS AVEC OPENAI GYM

Cliquez pour ajouter du texte



En apprentissage par renforcement, le modèle, l'IA, l'algo, ... est souvent appelé **agent**.

Le "monde extérieur", le jeu, le plateau, ... est appelé **environnement**.

La décision, la prédiction, ... de l'agent est appelée **action**.

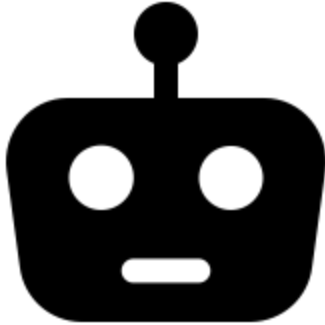
- Car elle change l'environnement
- Si elle ne changeait pas l'environnement, on utiliserait en pratique du supervisé/non supervisé !

L'ensemble des variables qui permettent de décrire l'environnement est appelé **état**.

Ce que perçoit l'agent de l'environnement est (souvent) appelé **observations**.

- **Attention : observations != Etat**
- **Attention : action != récompense**

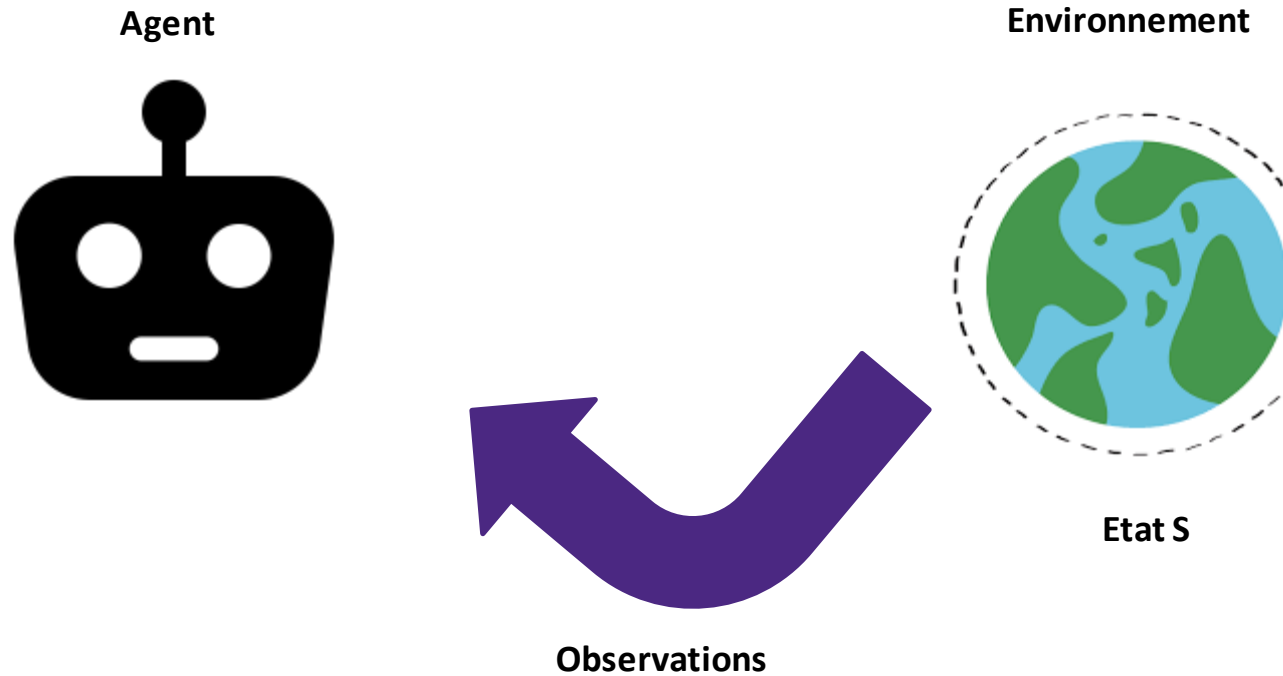
**Agent**

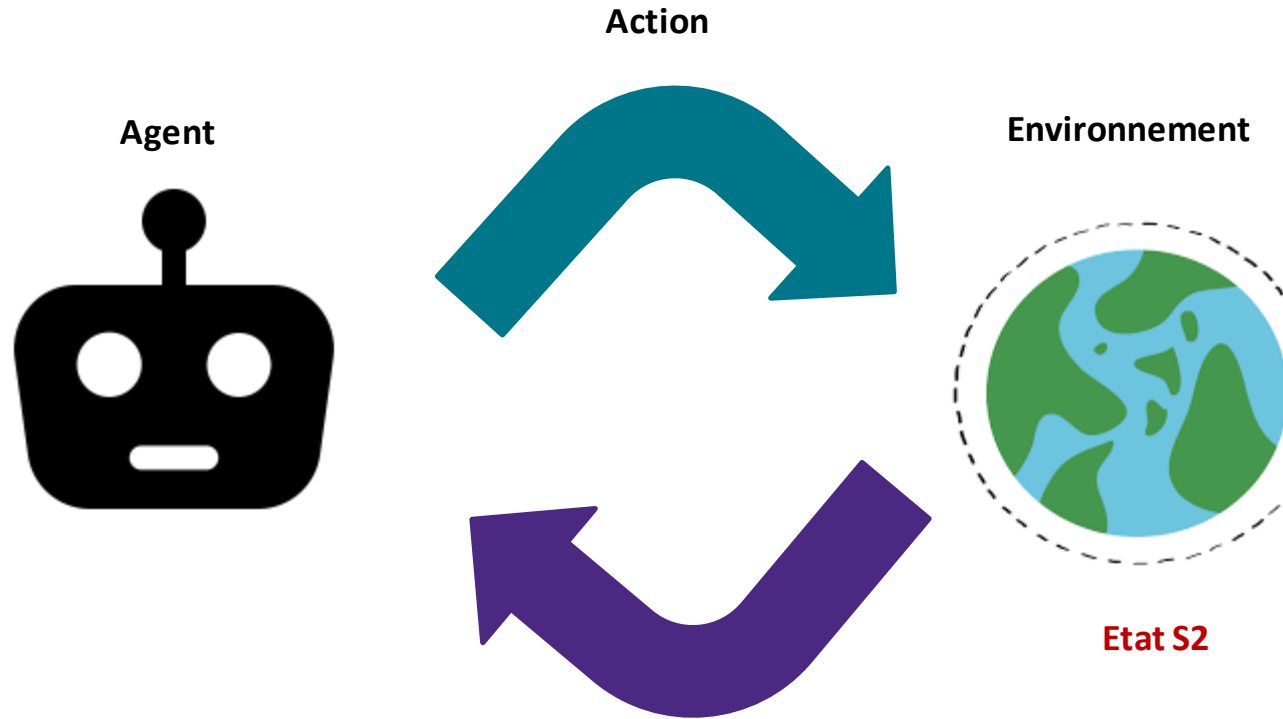


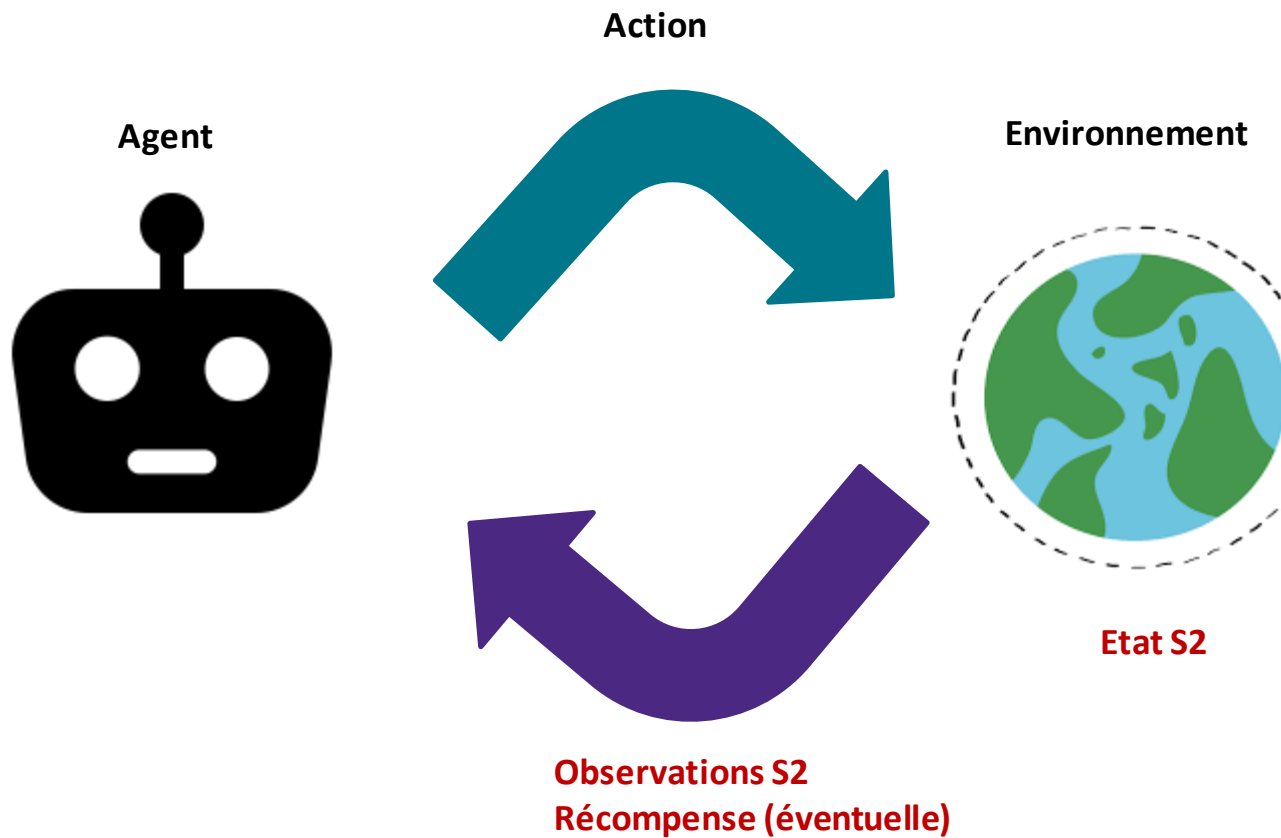
**Environnement**



**Etat S**







Quelques notions sur l'environnement :

- **Discret** : l'espace d'action est discret
- **Continu** : l'espace d'action est continu

	Discret	Continu
Echec		
Go		
Voiture autonome		
Optimisation de charge des datacenters		

Quelques notions sur l'environnement :

- **Discret** : l'espace d'action est discret
- **Continu** : l'espace d'action est continu

	Discret	Continu
Echec	✗	
Go	✗	
Voiture autonome		✗
Optimisation de charge des datacenters	✗	✗



## Quelques notions sur l'environnement :

- **Observable** : les observations sont suffisantes pour **reconstituer complètement l'état de l'environnement**
- **Partiellement observables** : les observations **ne permettent pas de reconstituer avec certitude l'intégralité de l'environnement.**

	Complètement observable	Partiellement observable
Echecs		
Go		
Labyrinthe		
Poker		

## Quelques notions sur l'environnement :

- **Observable** : les observations sont suffisantes pour **reconstituer complètement l'état de l'environnement**
- **Partiellement observables** : les observations **ne permettent pas de reconstituer avec certitude l'intégralité de l'environnement.**

	Complètement observable	Partiellement observable
Echec	✗	
Go	✗	
Labyrinthe	✗	✗
Poker		✗

Quelques notions sur l'environnement :

- **Déterministe:** la même action effectuée dans un même état  $S1$  mène toujours au même état  $S2$
- **Stochastique : non déterministe, une même action dans  $S1$  peut mener à des états différents**
  - Souvent quand notre environnement est une approximation imparfaite du monde réel.
  - Ex : voiture autonome et route
  - Ex : robotique et usure

**OpenAI Gym c'est :**

**- une bibliothèque d'environnements classiques d'apprentissage par renforcement**

- Que ce soit fourni par eux ou par la communauté
- `Pip install gym` versus `pip install gym[all]`

**- un framework simple pour utiliser des environnements en python**

**- un standard de syntaxe dédié spécifiquement aux problèmes de RL tels que décrits plus haut**

**~~- une bibliothèque qui s'intègre facilement aux notebooks~~**

`env = gym.make(...)`: **Instancie un nouvel environnement**

`env.reset()` : **(Re)-initialise un environnement**

- Syntaxe justifiée par le besoin d'entraîner des modèles

`env.render()` : **Visualise le contenu d'un environnement**

- Suivant l'environnement, graphiquement, en texte, ou pas du tout

`Observation, reward, done, info = env.step(action)`:

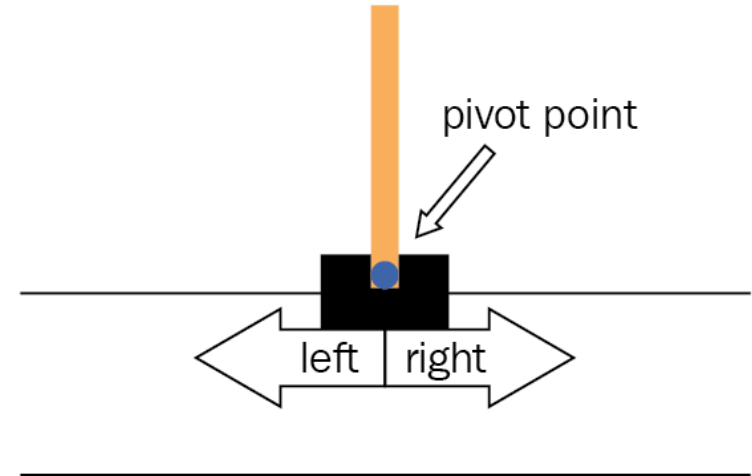
- Envoie une action à l'environnement
- Reçoit les observations du nouvel état, celui après application de l'action
- Done = True si la condition d'arrêt (réussite/échec) du problème est remplie

`env.close()`

# EXEMPLE : CARTPOLE

Un problème classique mais une version simple de problèmes très concrets.

- Il faut maintenir en équilibre une tige maintenue en équilibre sur un chariot.
- On peut uniquement pousser le chariot à gauche ou à droite.
- La tige peut avoir une vitesse/déplacement initial.
- Il peut y avoir du vent sur la tige.



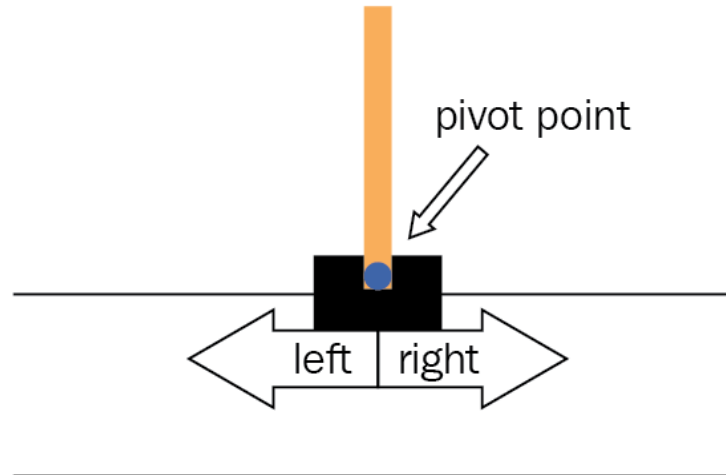
# EXEMPLE : CARTPOLE

Questions :

Cartpole est-il discret ou continu ?

Cartpole est-il complètement observable ?

Cartpole est-il déterministe ?



# EXEMPLE : CARTPOLE

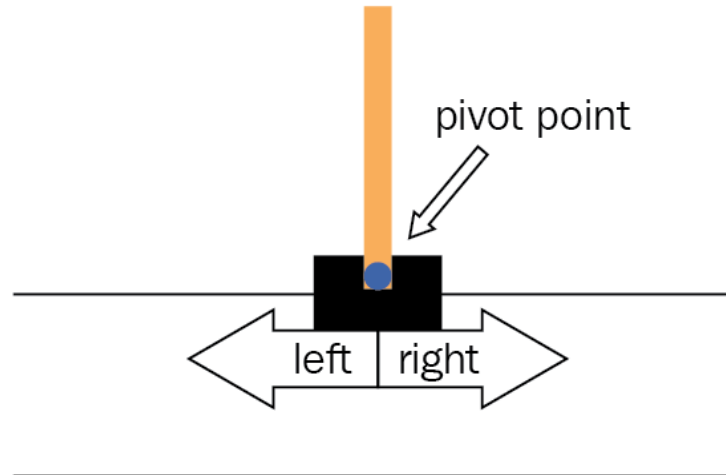
Questions :

**Cartpole est-il discret ou continu ?**

- Discret car l'espace d'action est de 2 : vers la droite ou vers la gauche. On ne contrôle pas l'intensité de la force, juste un choix binaire de direction.

**Cartpole est-il complètement observable ?**

**Cartpole est-il déterministe ?**





# EXEMPLE : CARTPOLE

Questions :

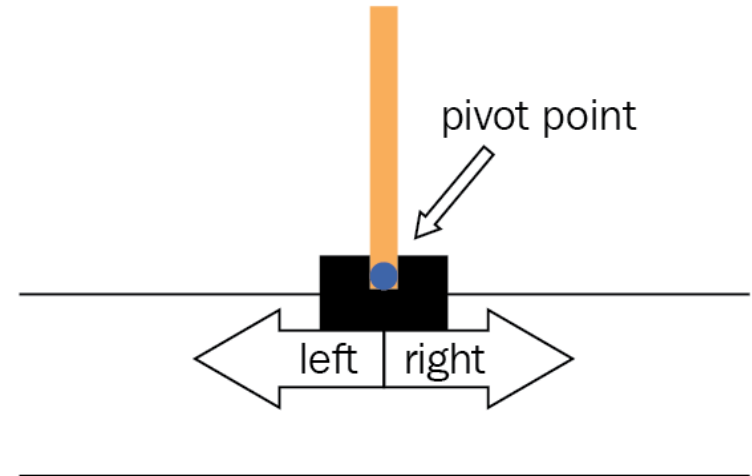
**Cartpole est-il discret ou continu ?**

- Discret car l'espace d'action est de 2 : vers la droite ou vers la gauche. On ne contrôle pas l'intensité de la force, juste un choix binaire de direction.

**Cartpole est-il complètement observable ?**

- Oui

**Cartpole est-il déterministe ?**



Questions :

**Cartpole est-il discret ou continu ?**

- Discret car l'espace d'action est de 2 : vers la droite ou vers la gauche. On ne contrôle pas l'intensité de la force, juste un choix binaire de direction.

**Cartpole est-il complètement observable ?**

- Oui

**Cartpole est-il déterministe ?**

- Sans vent, oui. Le vent est généré aléatoirement, donc non en présence de vent.

