

Q-LEARNING... DE LA THÉORIE À LA PRATIQUE



Ali Amine Ghazali

2025

<https://github.com/confooca/2025>



NOS EXPERTS AU PROGRAMME

ConFoo.CA
DEVELOPER CONFERENCE

Q LEARNING... DE LA
THÉORIE À LA PRATIQUE

26 février @ 15 h

MLOPS EST UN MYTHE! PIPELINE
E2E DE TEST, PACK ET
VERSIONNER

28 février @ 11 h

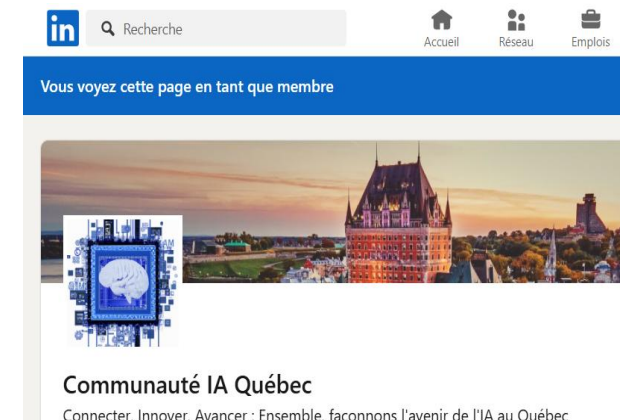
COFOMO



Ali Amine Ghazali

Directeur, Centre
d'excellence en intelligence
artificielle

<https://www.linkedin.com/in/aliamine-ghazali/>



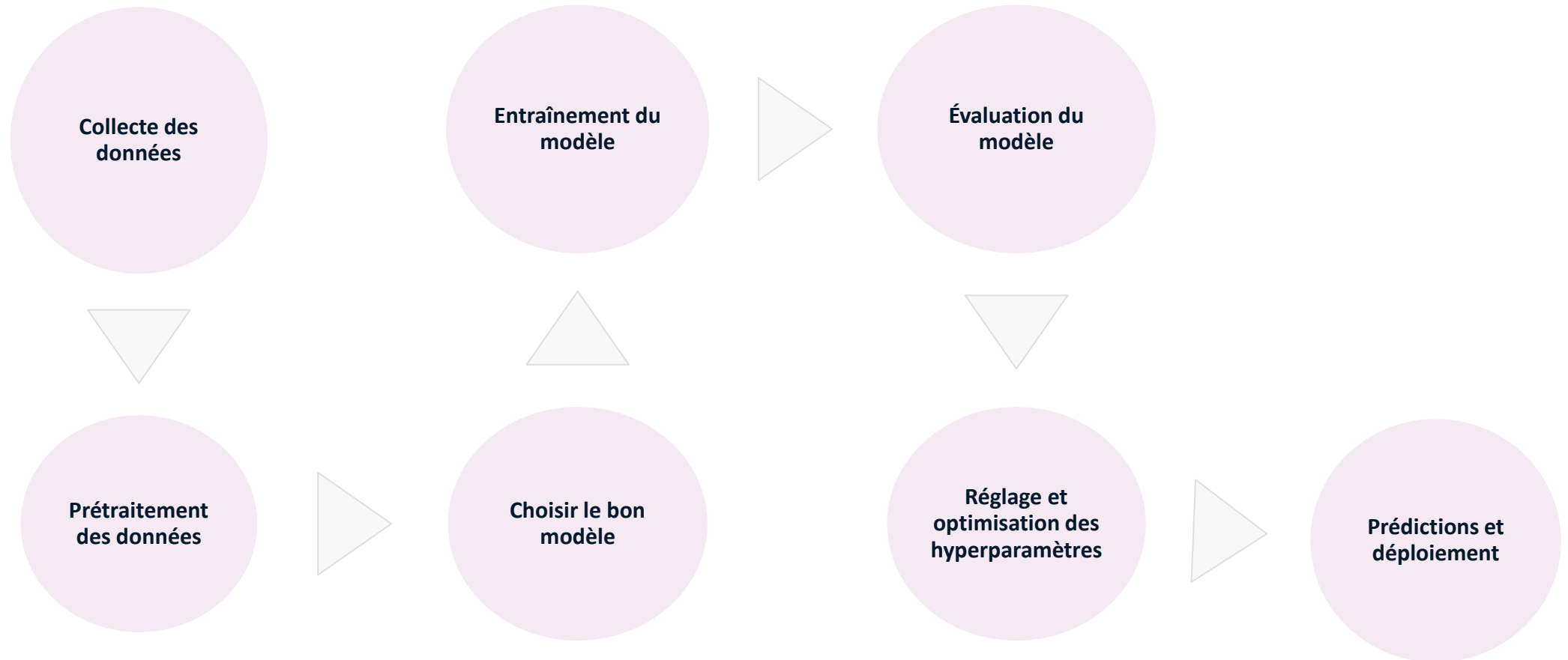
Membre du conseil canadien des normes – comité ISO/IEC – 42001 Technologies de l'information — Intelligence artificielle — Système de management

L'APPRENTISSAGE AUTOMATIQUE

- L'apprentissage automatique (Machine Learning) permet aux ordinateurs d'apprendre à partir de données et de prendre des décisions ou des prédictions sans être explicitement programmés pour le faire.

[Qu'est-ce que l'apprentissage automatique ? Définition, types, outils et plus | DataCamp](#)

COMMENT FONCTIONNE L'APPRENTISSAGE AUTOMATIQUE ?



DISTINCTION DES NOTIONS - IA - ML - DL

IA

- Fait référence au développement de programmes qui se comportent intelligemment et imitent l'intelligence humaine grâce à un ensemble d'algorithmes.

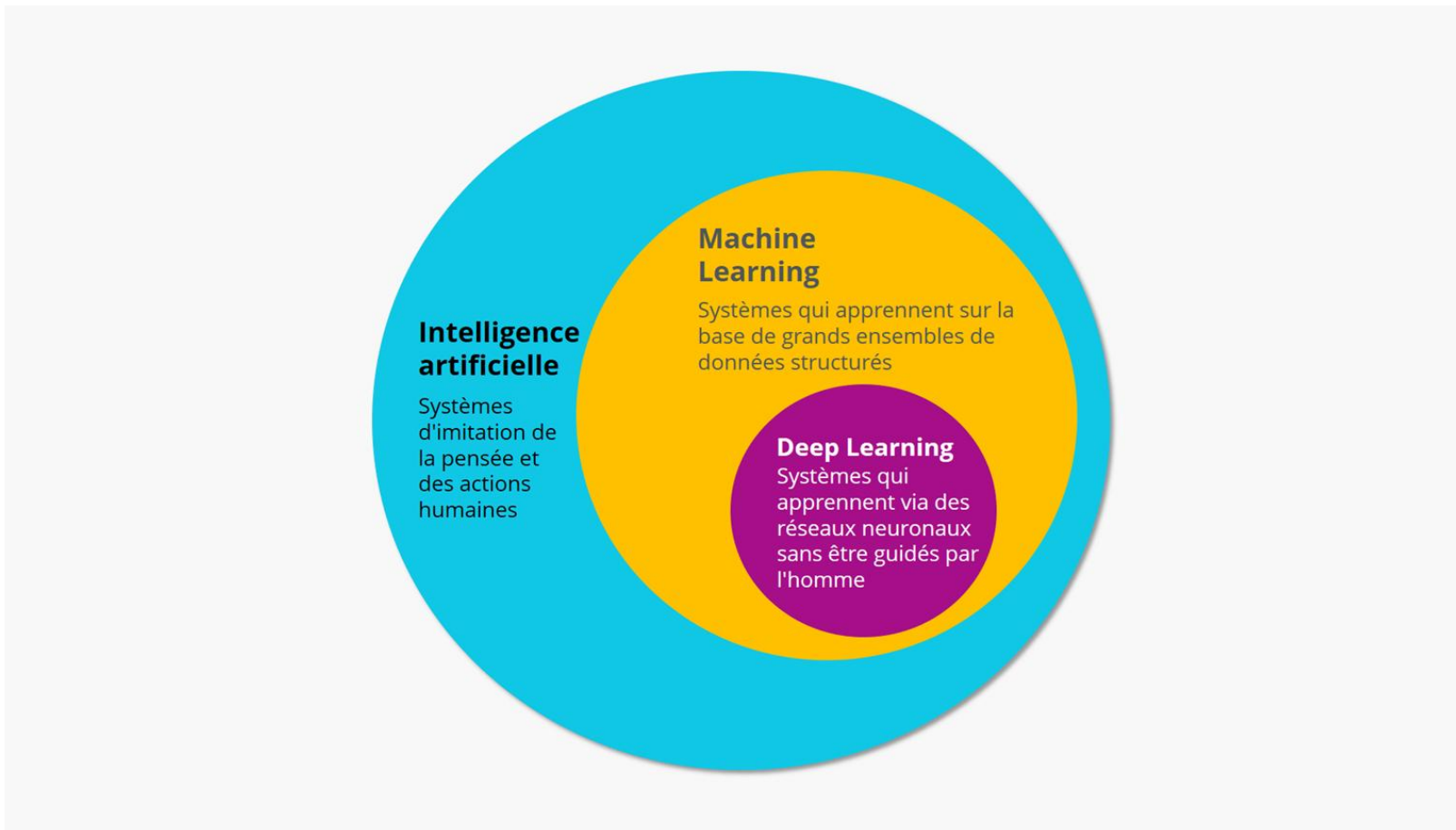
DL

- L'apprentissage profond (Deep Learning), quant à lui, est un sous-domaine du ML traitant d'algorithmes basés essentiellement sur des réseaux de neurones artificiels (RNA) multicouches qui s'inspirent de la structure du cerveau humain.

ML

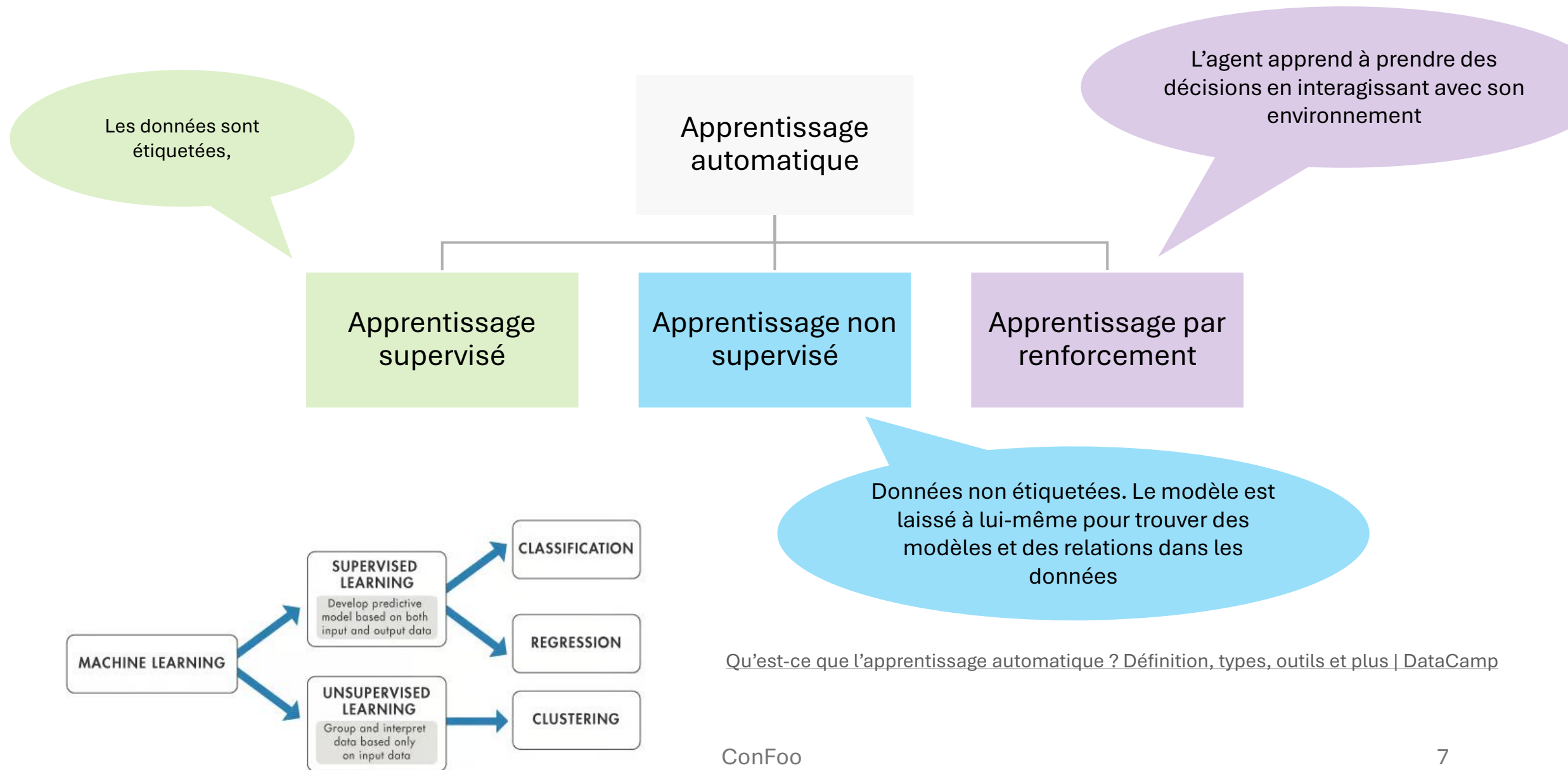
- L'apprentissage automatique est un sous-ensemble de l'IA, qui utilise des algorithmes qui apprennent à partir de données pour faire des prédictions.

DISTINCTION DES NOTIONS - IA - ML - DL



Deep learning vs Machine learning : quelle est la différence ? - IONOS

TYPES D'APPRENTISSAGE AUTOMATIQUE



L'APPRENTISSAGE PAR RENFORCEMENT

L'apprentissage par renforcement (RL) est la partie de l'écosystème de l'apprentissage automatique où l'agent apprend en interagissant avec l'environnement afin d'obtenir la stratégie optimale pour atteindre les objectifs.

Particularité 1

Très différent des algorithmes d'apprentissage automatique supervisé, pour lesquels nous devons ingérer et traiter ces données. L'apprentissage par renforcement ne nécessite pas de données. Au lieu de cela, il apprend de l'environnement et du système de récompense pour prendre de meilleures décisions.

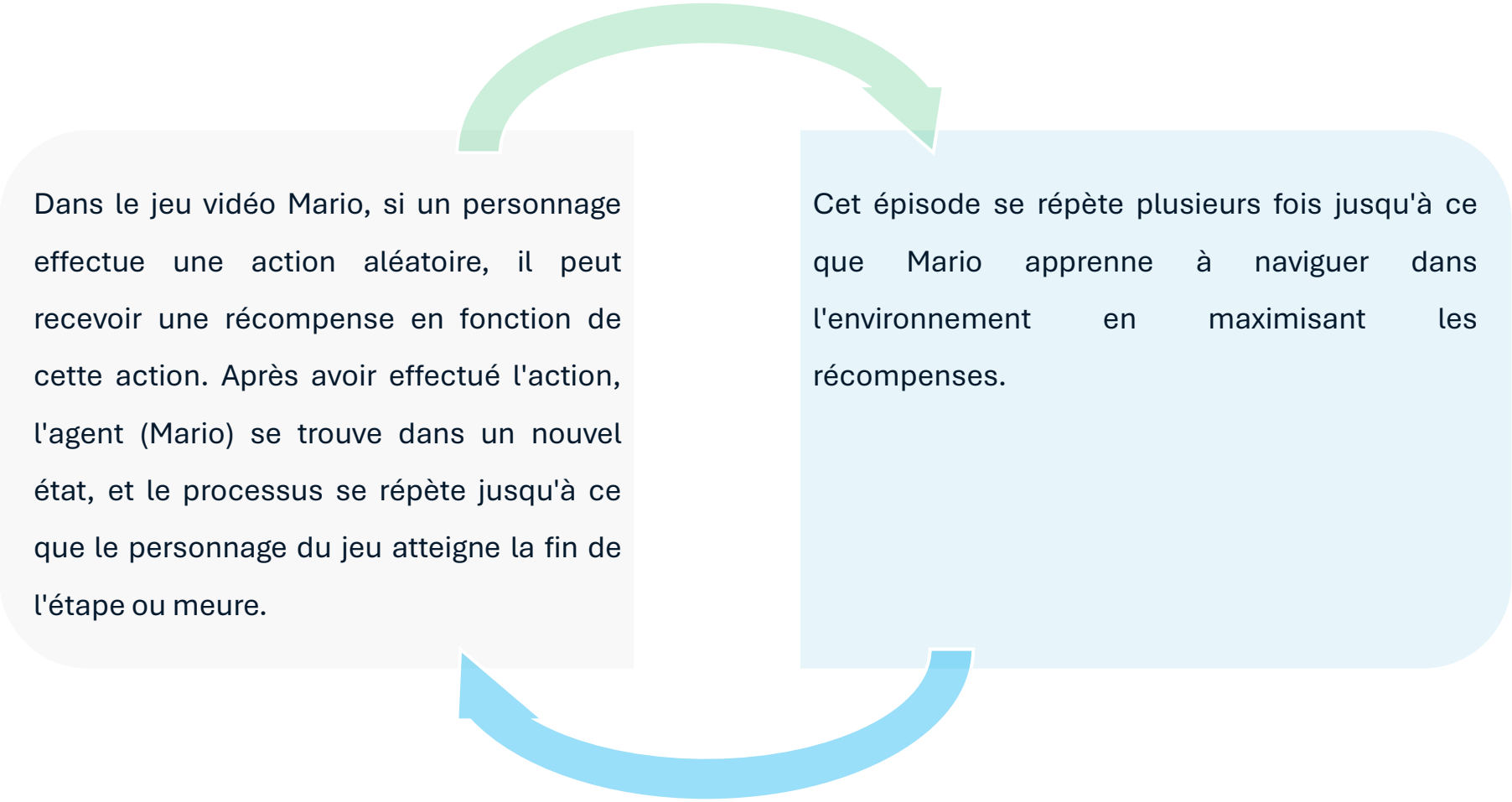
Particularité 2

L'apprentissage par renforcement est adapté aux problèmes où la décision prise à chaque étape peut affecter les résultats futurs, idéal pour les jeux et la robotique par exemple.

Particularité 3

EXEMPLE D'APPRENTISSAGE PAR RENFORCEMENT

MARIO

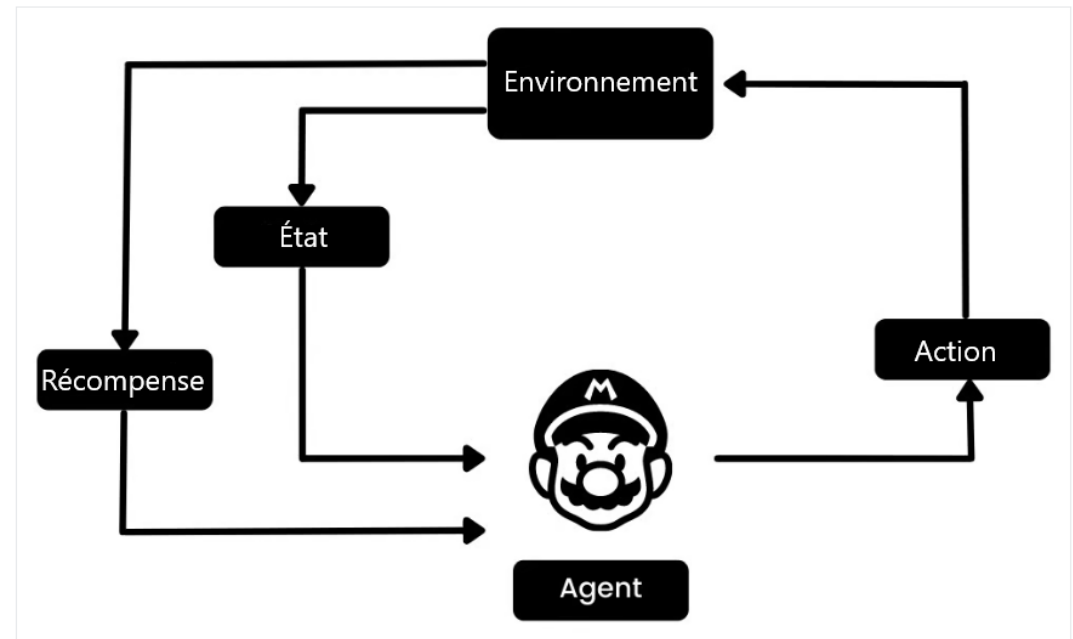


Dans le jeu vidéo Mario, si un personnage effectue une action aléatoire, il peut recevoir une récompense en fonction de cette action. Après avoir effectué l'action, l'agent (Mario) se trouve dans un nouvel état, et le processus se répète jusqu'à ce que le personnage du jeu atteigne la fin de l'étape ou meure.

Cet épisode se répète plusieurs fois jusqu'à ce que Mario apprenne à naviguer dans l'environnement en maximisant les récompenses.

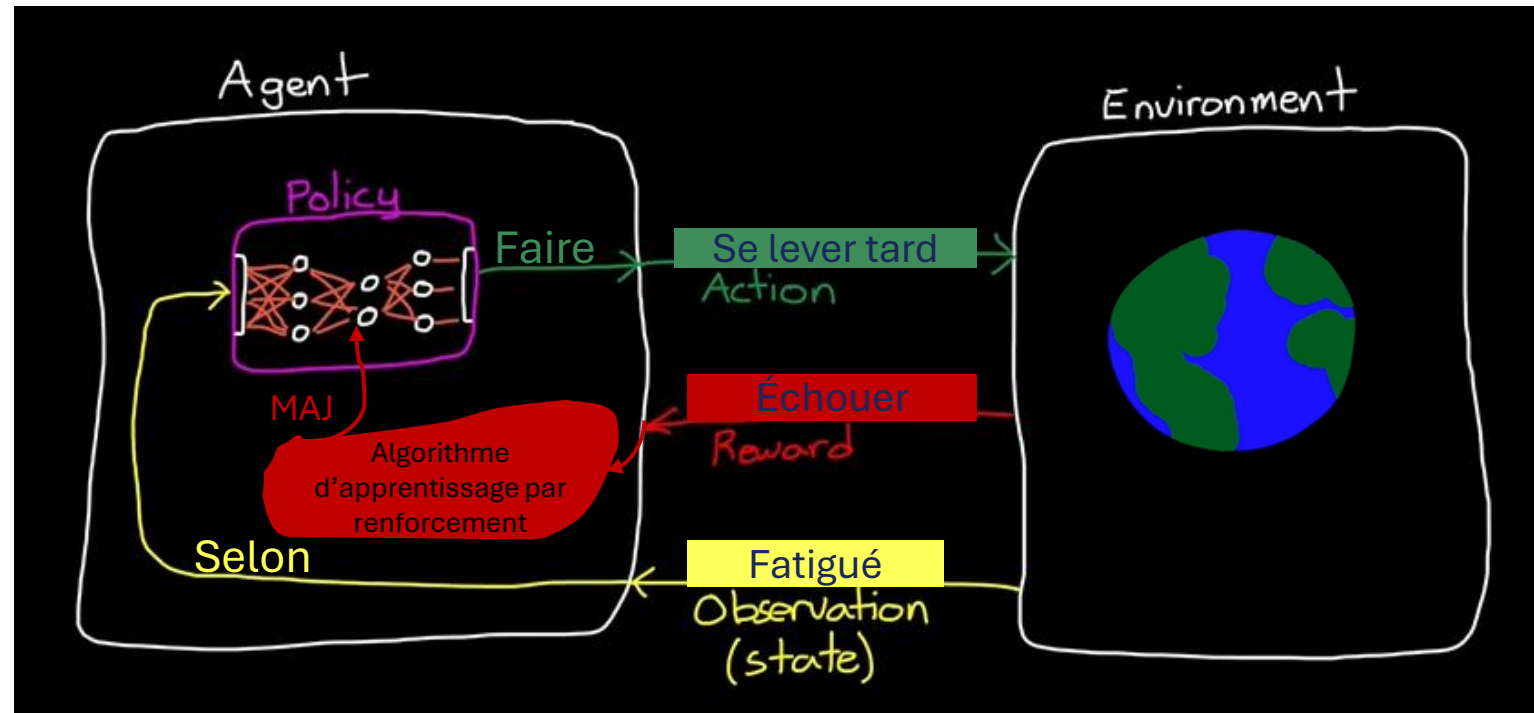
L'APPRENTISSAGE PAR RENFORCEMENT EN 5 ÉTAPES

1. L'agent se trouve à l'état zéro dans un environnement.
2. Il prendra une mesure basée sur une stratégie (policie) spécifique.
3. Il recevra une récompense ou une punition en fonction de cette action.
4. En tirant les leçons des actions précédentes et en optimisant la stratégie.
5. Le processus se répète jusqu'à ce qu'une stratégie optimale soit trouvée



L'APPRENTISSAGE PAR RENFORCEMENT EN 5 ÉTAPES

Exemple de passage d'examen



APPRENTISSAGE PAR RENFORCEMENT AVEC LE Q-LEARNING

- ✓ L'apprentissage Q est un algorithme sans modèle, qui trouvera la meilleure série d'actions en fonction de l'état actuel de l'agent.
- ✓ Le "Q" signifie qualité. La qualité représente la valeur de l'action pour maximiser les récompenses futures.

LES TERMES FONDAMENTAUX DU Q-LEARNING

Termes utiles pour comprendre les principes fondamentaux du Q-Learning.

- ✓ **État ou situation = (s)**: position actuelle de l'agent dans l'environnement.
- ✓ **Action = (a)**: une mesure prise par l'agent dans un état particulier.
- ✓ **Récompenses = (r)**: pour chaque action, l'agent reçoit une récompense ou une pénalité.
- ✓ **Episodes**: la fin de la phase, où les agents ne peuvent plus prendre de nouvelles mesures. Il se produit lorsque l'agent a atteint l'objectif ou a échoué.
- ✓ **$Q(s_t, a_t)$** : il s'agit de l'estimation actuelle de $Q(s_{t+1}, a)$.
- ✓ **Table Q**: l'agent tient à jour la table Q des ensembles d'états et d'actions.

COMMENT FONCTIONNE Q-LEARNING ?

Prenant l'exemple d'un lac gelé. Dans cet environnement, l'agent doit traverser le lac gelé du point de départ au point d'arrivée, sans tomber dans les trous.

Dans cet exemple nous allons utiliser :

- OpenAI Gym (gymnasium)
- Environnement FrozenLake-v1
- Table Q
- Fonction Q







[Deep_reinforcement_learning_Course/Q learning/FrozenLake/Q Learning avec FrozenLake.ipynb](#) chez master · [simoninithomas/Deep_reinforcement_learning_Course](#) · [Lien avec GitHub](#)

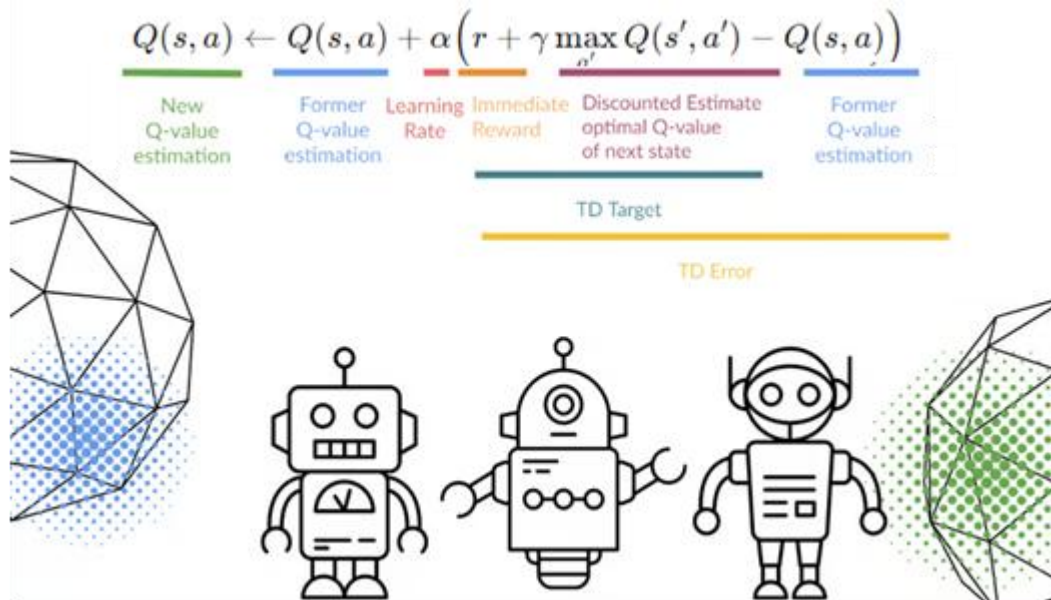
COMMENT FONCTIONNE LA TABLE Q

Dans notre exemple, le personnage peut se déplacer vers le haut, le bas, la gauche et la droite. Nous avons quatre actions possibles et quatre états (**début**, **inactif**, **mauvais chemin** et **fin**).

On peut également considérer que le mauvais chemin est à l'origine de la chute dans le trou. Nous allons initialiser le tableau Q avec des valeurs à 0.

				
Start	0	0	0	0
Idle	0	0	0	0
Hole	0	0	0	0
End	0	0	0	0

COMMENT METTRE À JOUR LA TABLE Q



	→	←	↑	↓
Start	0	1	0	0
Idle	2	0	0	3
Hole	0	2	0	0
End	1	0	0	0

MISE À JOUR DU TABLEAU Q

La formule de mise à jour classique du Q-Learning est la suivante :

$$Q(s,a) \leftarrow Q(s,a) + \alpha (r + \gamma \max_{a'} Q(s',a') - Q(s,a))$$

Section	Description
$Q(s,a)$	Il s'agit de l'estimation actuelle de la qualité de l'action « a » dans l'état « s ». C'est notre « hypothèse » ou notre croyance actuelle quant à l'efficacité de cette action.
r	C'est la récompense immédiate reçue après avoir effectué l'action « a » dans l'état « s ». Elle représente le retour instantané de l'environnement (par exemple, un point positif ou négatif).
s' et $\max_{a'} Q(s',a')$	Après avoir pris l'action « a », l'agent se retrouve dans un nouvel état « s' ». Parmi toutes les actions possibles dans « s' », $\max_{a'} Q(s',a')$ correspond à la meilleure estimation de la qualité (la valeur la plus élevée) pour cet état. Cette partie représente l'estimation des récompenses futures .
γ	Ce paramètre, appelé facteur d'actualisation , est un nombre entre 0 et 1. Il détermine l'importance accordée aux récompenses futures par rapport à la récompense immédiate. Un « γ » proche de 1 signifie que l'agent considère fortement les gains futurs, tandis qu'un « γ » faible lui fait privilégier les gains immédiats.
α	Le taux d'apprentissage , qui contrôle l'ampleur de la mise à jour. Plus « α » est élevé, plus la nouvelle information (la différence entre ce qu'on attendait et ce qui est réellement observé) va modifier l'estimation actuelle. À l'inverse, un « α » bas signifie que l'agent ajuste lentement sa Q-table.
$(r + \gamma \max_{a'} Q(s',a') - Q(s,a))$	Représente l'erreur de différence temporelle, ou TD error . Elle mesure l'écart entre : <ul style="list-style-type: none">• Ce que l'agent attendait initialement ($Q(s,a)$),• Et ce qu'il constate réellement, c'est-à-dire la somme de la récompense immédiate « r » et de la meilleure récompense future possible $\gamma \max_{a'} Q(s',a')$.

The background is a vibrant blue field filled with a complex network of glowing lines and dots. The lines are thin and vary in brightness, some appearing as sharp streaks while others are more diffuse. The dots are small, bright blue spheres scattered throughout the composition. The overall effect is one of dynamic energy and digital connectivity.

DÉMONSTRATION

<https://github.com/confooca/2025>



NOS EXPERTS AU PROGRAMME

ConFoo.CA
DEVELOPER CONFERENCE

Q LEARNING... DE LA
THÉORIE À LA PRATIQUE

26 février @ 15 h

MLOPS EST UN MYTHE! PIPELINE
E2E DE TEST, PACK ET
VERSIONNER

28 février @ 11 h

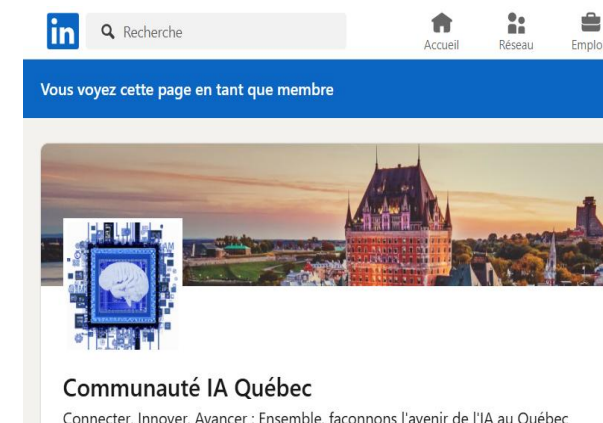
COFOMO



Ali Amine Ghazali

Directeur, Centre
d'excellence en intelligence
artificielle

<https://www.linkedin.com/in/aliamine-ghazali/>



Membre du conseil canadien des normes – comité ISO/IEC – 42001 Technologies de l'information — Intelligence artificielle — Système de management