

# ATTERIR SUR LA LUNE ENDORMANT

COMMENT L'UTILISATION DE L'APPRENTISSAGE  
PAR RENFORCEMENT PERMET DE RESOUDRE  
DES PROBLEMES

MAEGHT Loan



# SOMMAIRE

## ■ APPRENTISSAGE PAR RENFORCEMENT

DESCRIPTION  
EXPLICATION  
UTILISATION

## ■ L'ATTERRISSEUR LUNAIRE

ORIGINE  
OBSTACLE  
RÉSOLUTION



# APPRENTISSAGE PAR RENFORCEMENT

## ■ DESCRIPTION

L'APPRENTISSAGE PAR RENFORCEMENT CONSISTE, POUR UN AGENT AUTONOME À APPRENDRE DES ACTIONS, À PARTIR D'EXPÉRIENCES, DE FAÇON À OPTIMISER UNE RÉCOMPENSE QUANTITATIVE AU COURS DU TEMPS.

L'AGENT EST PLONGÉ AU SEIN D'UN ENVIRONNEMENT, ET PREND SES DÉCISIONS EN FONCTION DE SON ÉTAT COURANT.

EN RETOUR, L'ENVIRONNEMENT PROCURE À L'AGENT UNE RÉCOMPENSE, QUI PEUT ÊTRE POSITIVE OU NÉGATIVE.

L'AGENT CHERCHE, AU TRAVERS D'EXPÉRIENCES ITÉRÉES, UNE STRATÉGIE OPTIMALE, QUI MAXIMISE LA SOMME DES RÉCOMPENSES AU COURS DU TEMPS.

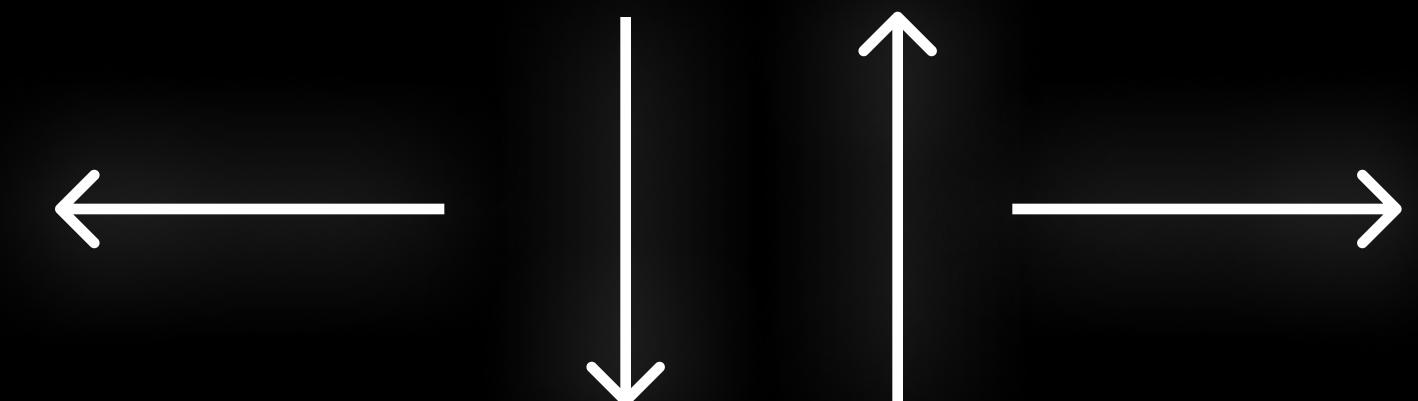
# APPRENTISSAGE PAR RENFORCEMENT

## ■ EXPLICATION

L'AGENT POSSEDE TROIS ELEMENTS IMPORTANTS.

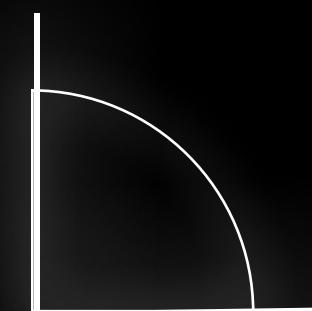
UNE LISTE D'ACTIONS POSSIBLES NOTÉE :  $\Delta$

EXEMPLE UNE DIRECTION DE DEPLACEMENT



UNE LISTE D'ETAT NOTÉE :  $S$

EXEMPLE COORDONÉES, VITESSE



ON ASSOCIE A CHAQUE TUPLET UN SCORE NOTÉ :  $R$

EXEMPLE RALEMENTIR AVANT DE TOUCHER LE SOL

+ 1



# APPRENTISSAGE PAR RENFORCEMENT

## ■ EXPLICATION

L'AGENT POSSEDE DEUX FONCTIONS :

FONCTION EPSILON NOTÉE :  $\epsilon$

CORRESPOND AU TAUX D'EXPLORATION CONTRE EXPLOITATION

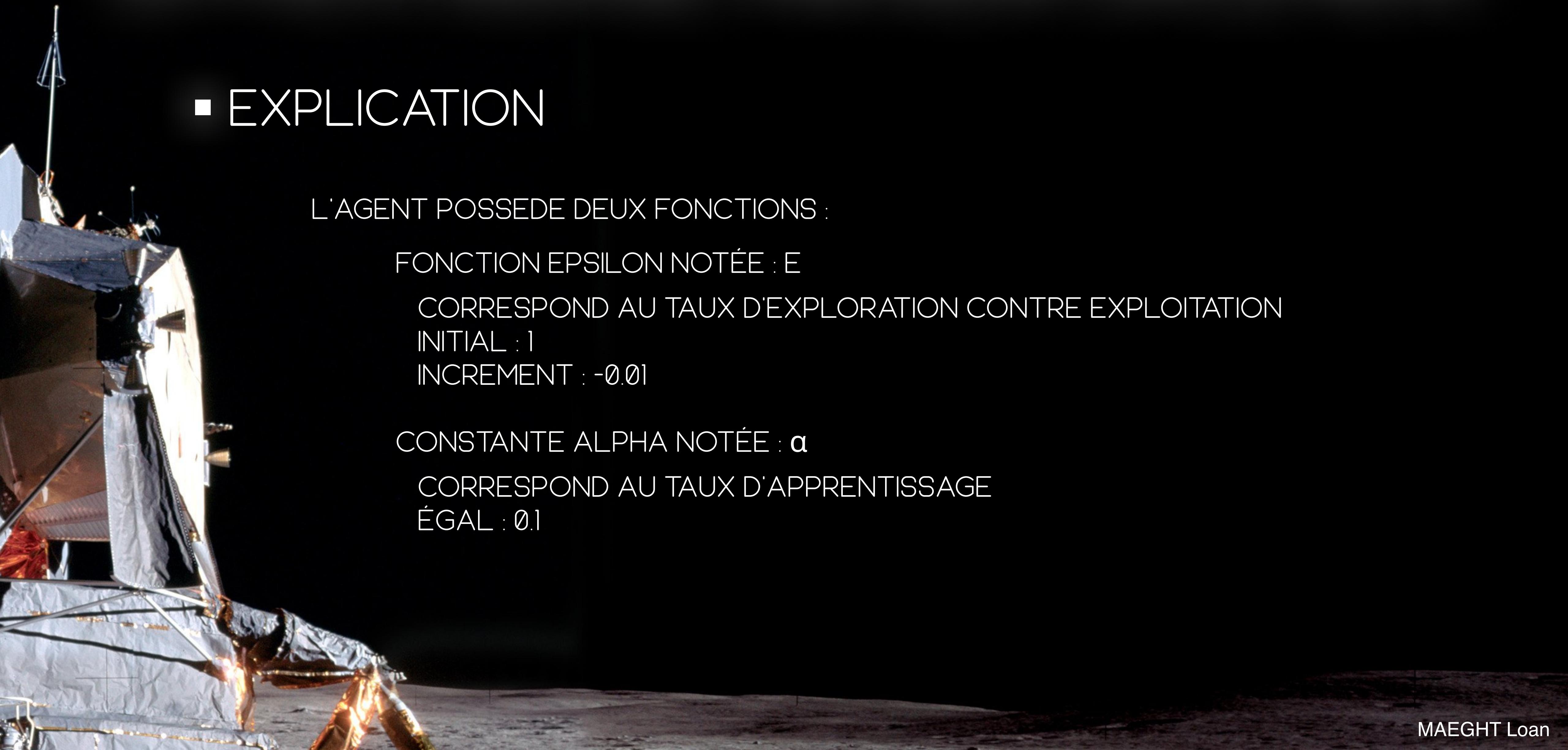
INITIAL : 1

INCREMENT : -0.01

CONSTANTE ALPHA NOTÉE :  $\alpha$

CORRESPOND AU TAUX D'APPRENTISSAGE

ÉGAL : 0.1



# APPRENTISSAGE PAR RENFORCEMENT

## ■ EXPLICATION

L'AGENT POSSEDE UN TABLEAU APPELÉ Q TABLEAU QUI RECENSE POUR CHAQUE CONFIGURATION LE SCORE DES ACTIONS POSSIBLES

EXEMPLE DE TABLEAU :  
STATUT → COORDONNÉES  
ACTION → DÉPLACEMENT 2D

	←	↓	→	↑
(0 . 0)	0.12	0.34	0.17	0.45
(0 . 1)	0.15	0.23	0.36	0.81
(1 . 0)	0.11	0.65	0.87	0.74
(1 . 1)	0.8	0.32	0.01	0.95

# APPRENTISSAGE PAR RENFORCEMENT

## ■ EXPLICATION

LE CHOIX DE L'ACTION DÉPEND DE DEUX ÉLÉMENTS ::

1.

LE CHOIX ENTRE EXPLORATION ET EXPLOITATION

CE QUI SIGNIFIE CHOISIR ENTRE EXPLORER C'EST A DIRE FAIRE DES CHOIX ALÉATOIRE OU EXPLOITER SES CONNAISSANCES

ON PREND UN NOMBRE ALÉATOIRE COMPRIS ENTRE 0 ET 1

SI LE NOMBRE EST PLUS PETIT QUE EPSILON ALORS ON CHOISIT UNE ACTION ALÉATOIRE

	←	↓	→	↑
(0 . 0)	0.12	0.34	0.17	0.45
(0 . 1)	0.15	0.23	0.36	0.81
(1 . 0)	0.11	0.65	0.87	0.74
(1 . 1)	0.8	0.32	0.01	0.95



# APPRENTISSAGE PAR RENFORCEMENT

## ■ EXPLICATION

LE CHOIX DE L'ACTION DÉPEND DE DEUX ÉLÉMENTS ::

2.

RECHERCHE DU MEILLEUR RESULTAT

POUR UN ETAT DONNÉ ON CHERCHE  
L'ACTION QUI DONNERA LE MEILLEUR  
RÉSULTAT

DANS LE CAS OU IL Y PLUSIEURS VALEURS  
ÉGALÉ, ON PREND LA PREMIÈRE

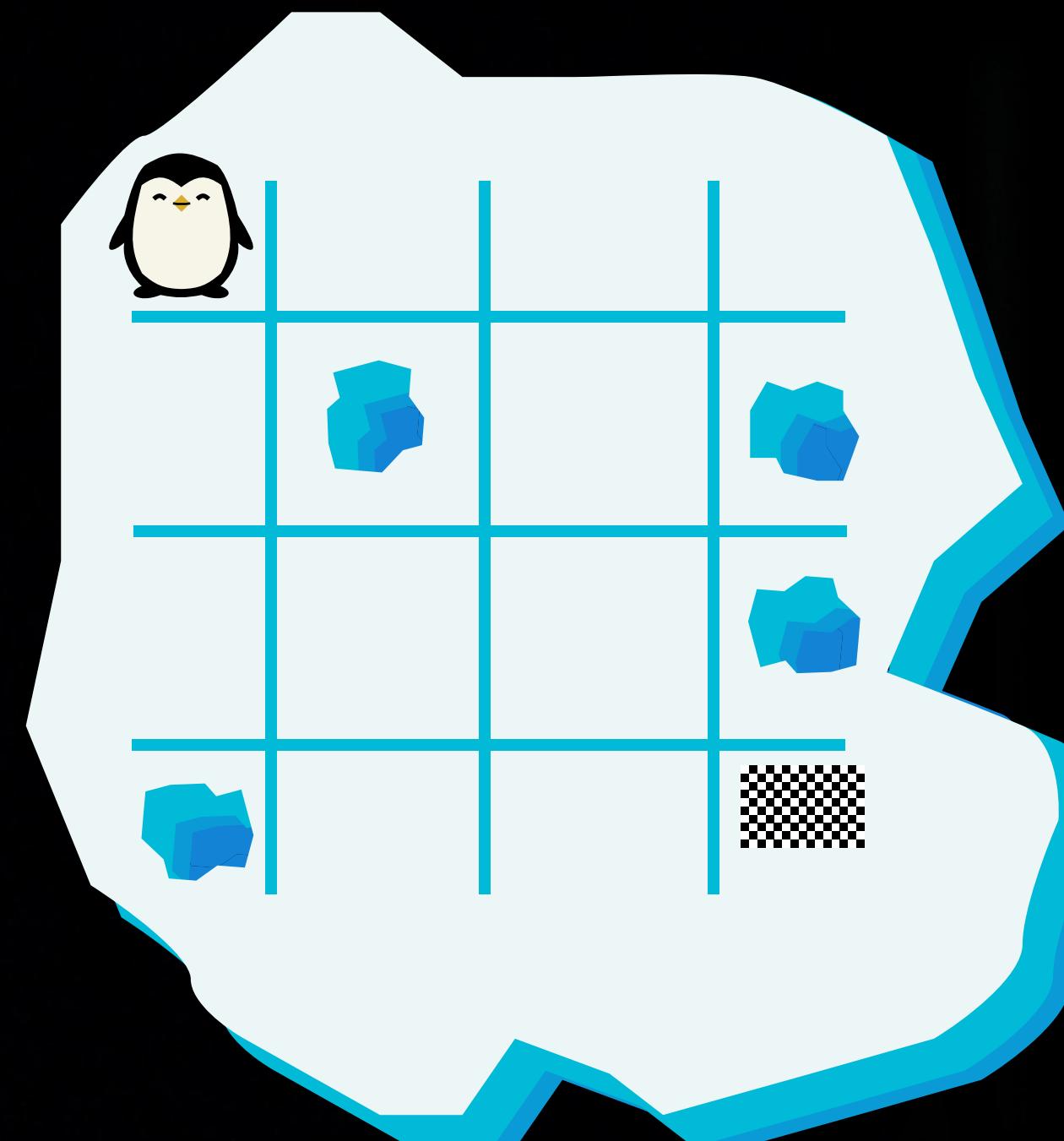
EXEMPLE:

0. 0 --> 0.45 DEPLACEMENT VERS LE HAUT

	←	↓	→	↑
(0 . 0)	0.12	0.34	0.17	0.45
(0 . 1)	0.15	0.23	0.36	0.81
(1 . 0)	0.11	0.65	0.87	0.74
(1 . 1)	0.8	0.32	0.01	0.95

# APPRENTISSAGE PAR RENFORCEMENT

## ■ EXPLICATION MISE EN APPLICATION



NOUS UTILISONS COMME EXEMPLE L'ENVIRONNEMENT APPELÉ "FROZENLAKE-V0"

LE PINGOUIN A POUR OBJECTIF DE REJOINDRE L'ARRIVÉ EN ÉVITANT LES TROUS

LA LARGEUR DU TABLEAU CORRESPOND AU NOMBRE D'ACTION POSSIBLE  
DANS NOTRE CAS 4

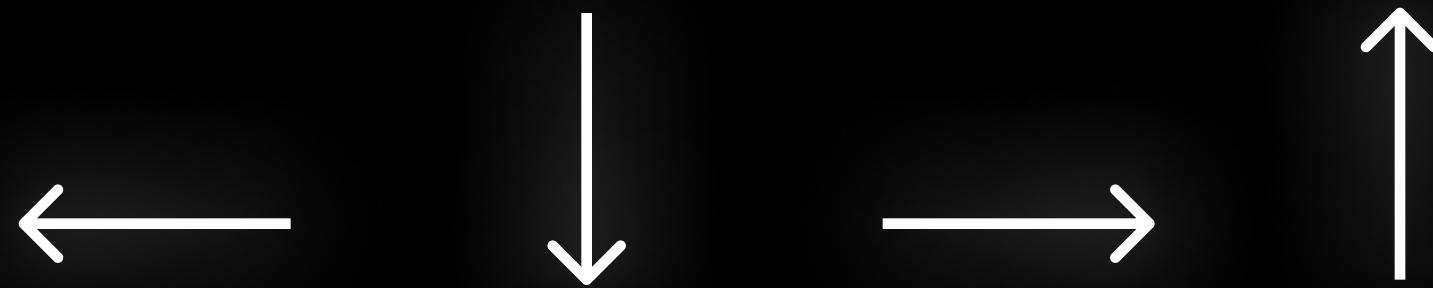
LA LONGUEUR DU TABLEAU CORRESPOND AU NOMBRE DE SITUATION POSSIBLE  
DANS NOTRE CAS 4X4 SOIT 16

LE PINGOUIN REÇOIT UNE RÉCOMPENSE DE 1 S'IL SE TROUVE SUR L'OBJECTIF SINON 0

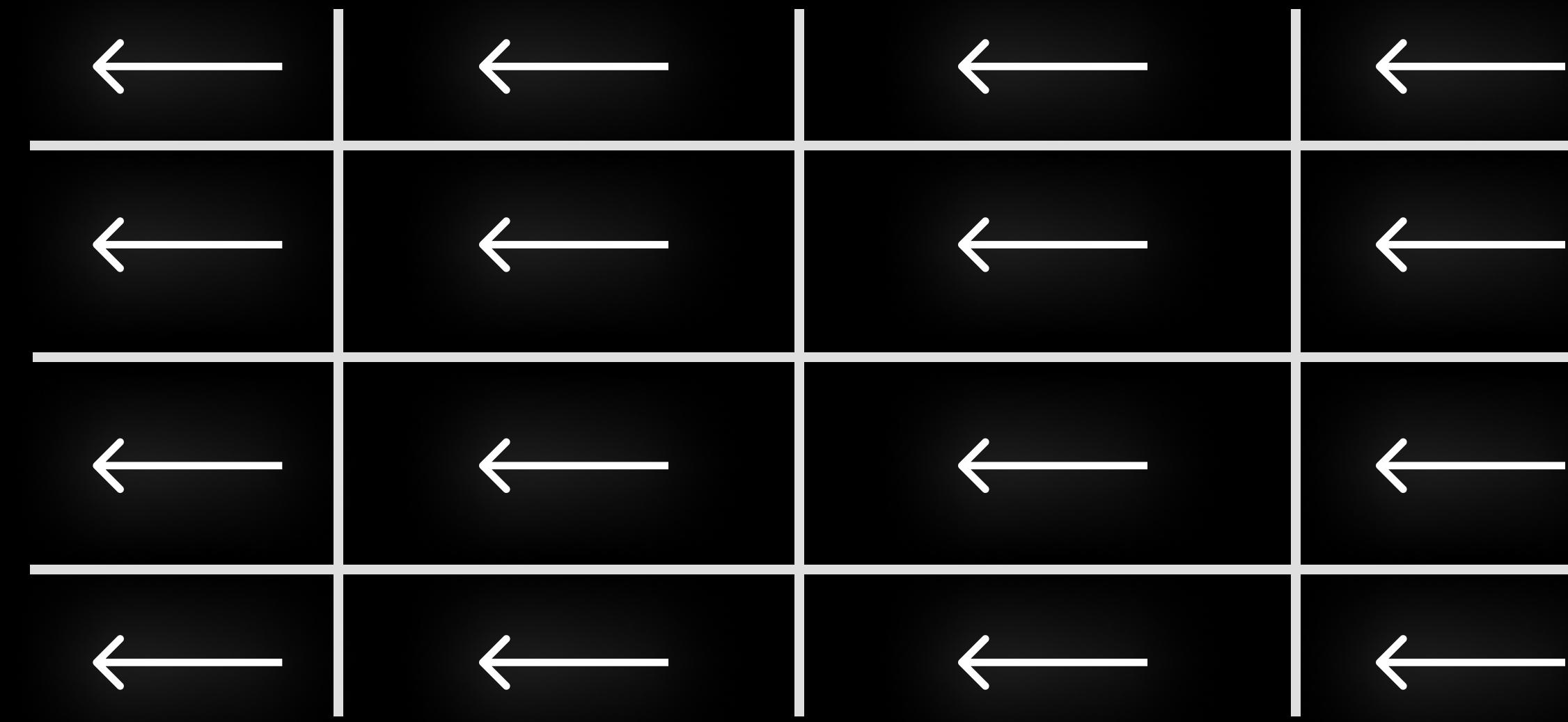
# APPRENTISSAGE PAR RENFORCEMENT

## ■ EXPLICATION

### MISE EN APPLICATION



AU COURS DES PREMIÈRES ÉTAPES, LE PINGOUIN SE DÉPLACE ALÉATOIUREMENT.  
LA PROBABILITÉ QUE LE PINGOUIN TOMBE DANS UN TROU OU DÉPASSE LE NOMBRE DE DÉPLASSEMENTS LIMITÉ EST TRÈS FORTE  
LE TABLEAU EST INITIALISÉ À ZERO DONC L'ACTION PAR DÉFAUT EST UN DÉPLACEMENT PAR LA GAUCHE



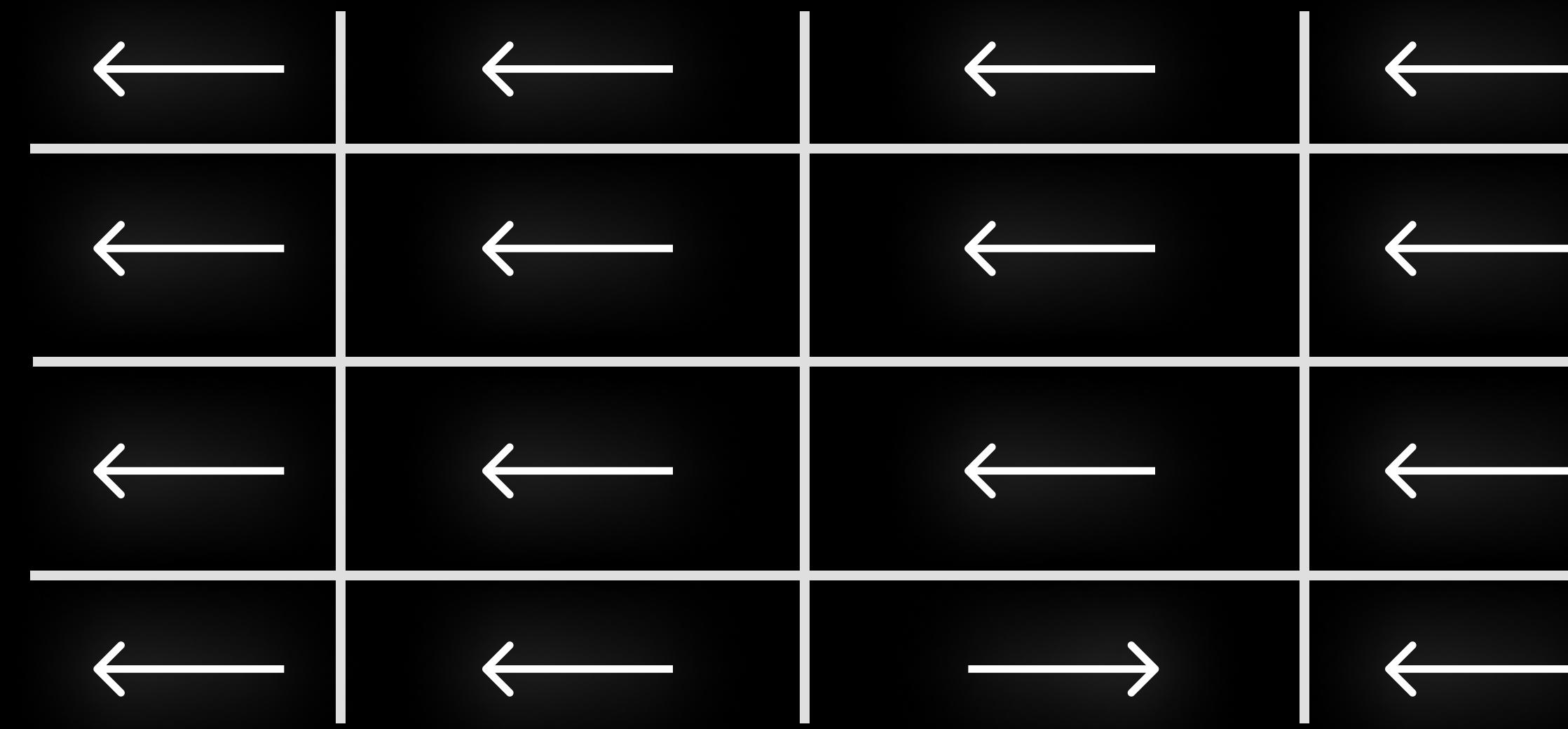
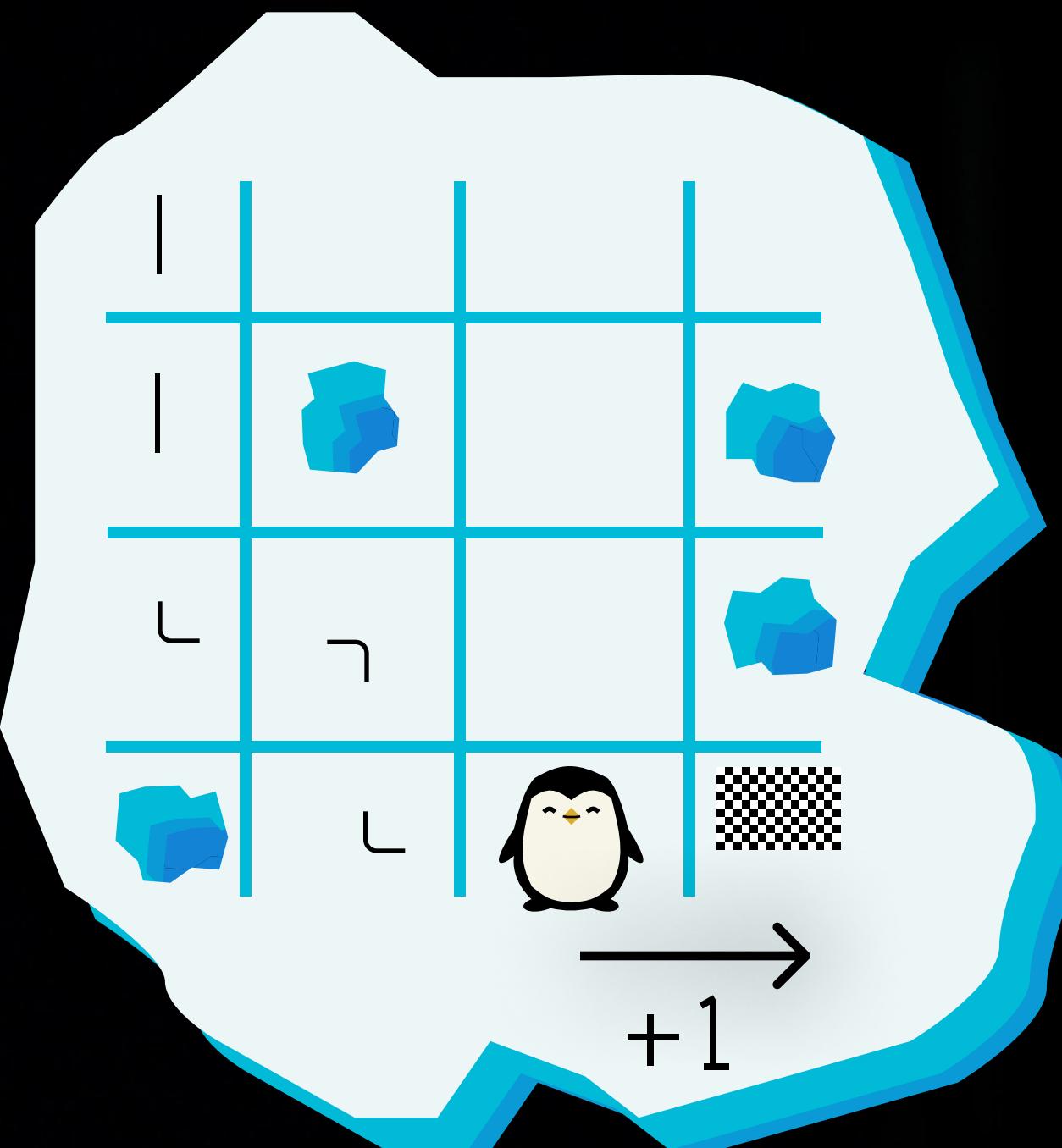
VALEUR EN DETAIL:  
0, 0, 0, 0

# APPRENTISSAGE PAR RENFORCEMENT

## ■ EXPLICATION

MISE EN APPLICATION  
EPISODE 10

NOUS PASSONS DIRECTEMENT À LA PREMIÈRE ÉTAPE QUI VA METTRE  
À JOUR LE TABLEAU

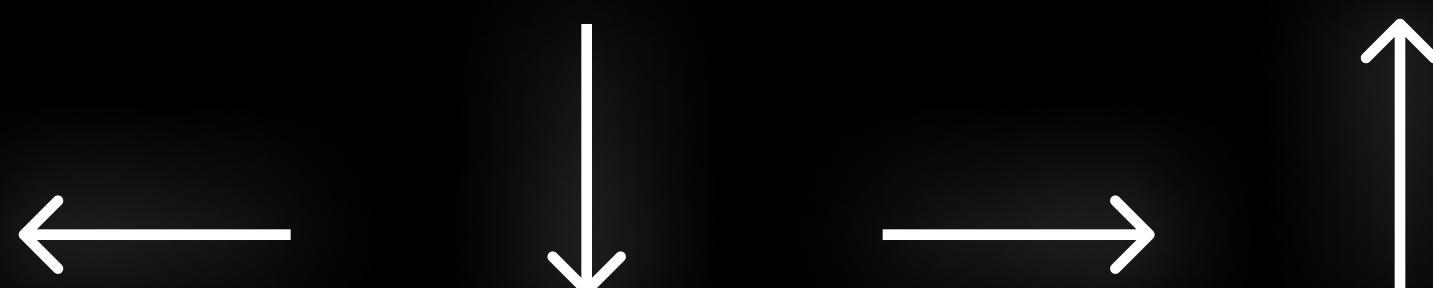


VALEUR EN DÉTAIL:  
0. 0. 0. 1 0

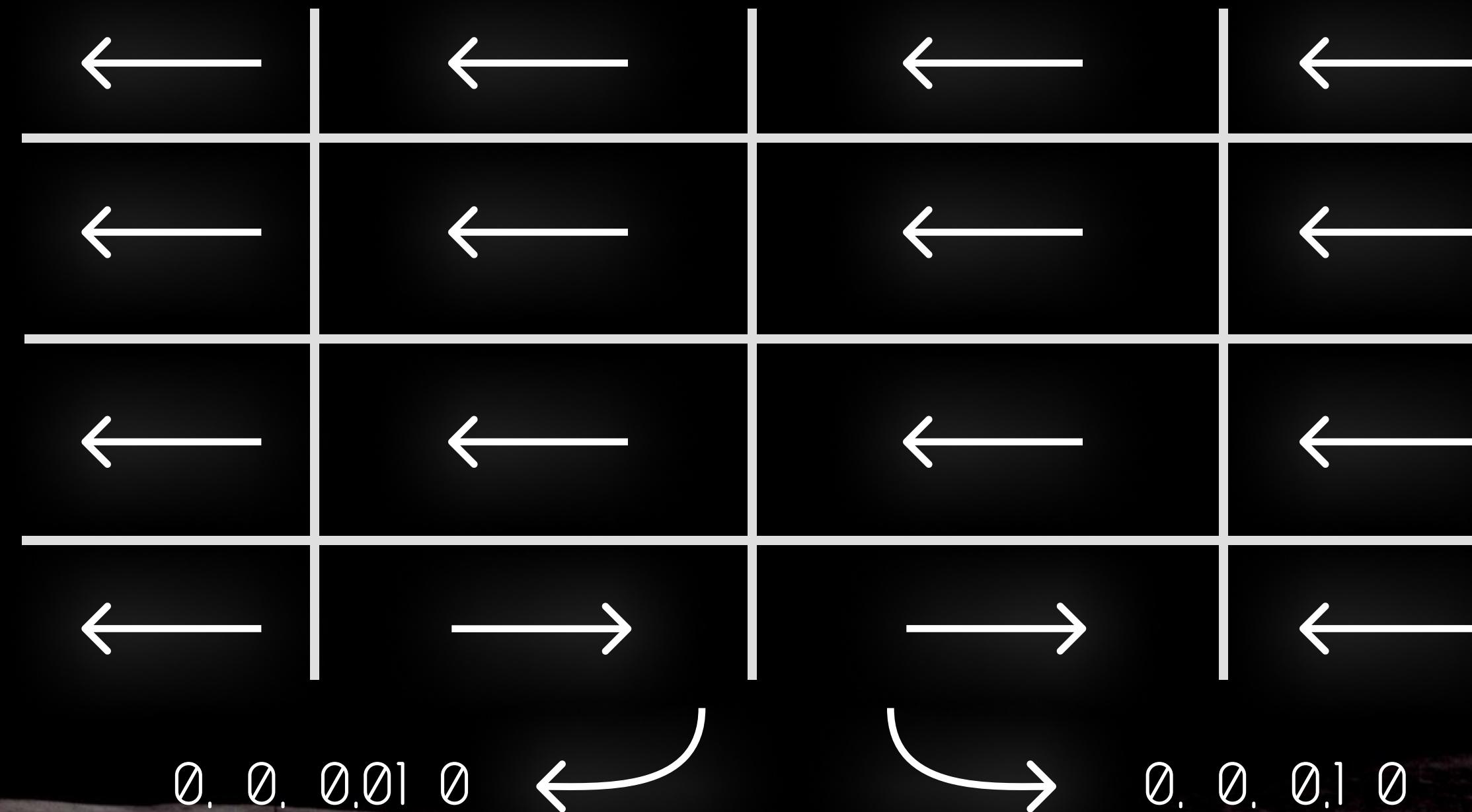
# APPRENTISSAGE PAR RENFORCEMENT

## ■ EXPLICATION

MISE EN APPLICATION  
EPISODE 15



APRES CHAQUE ACTION DU PINGUIN, SA CASE DE DEPART SE MET À JOUR EN CALCULANT UNE NOTE POUR L'ACTION REALISÉE ET UNE NOTE POUR SA NOUVELLE SITUATION.. CETTE METHODE RÉPÉTÉE DE MULTIPLES FOIS PERMET DE GUIIDER PROGRESSIVEMENT LE PINGUIN VERS SON BUT.

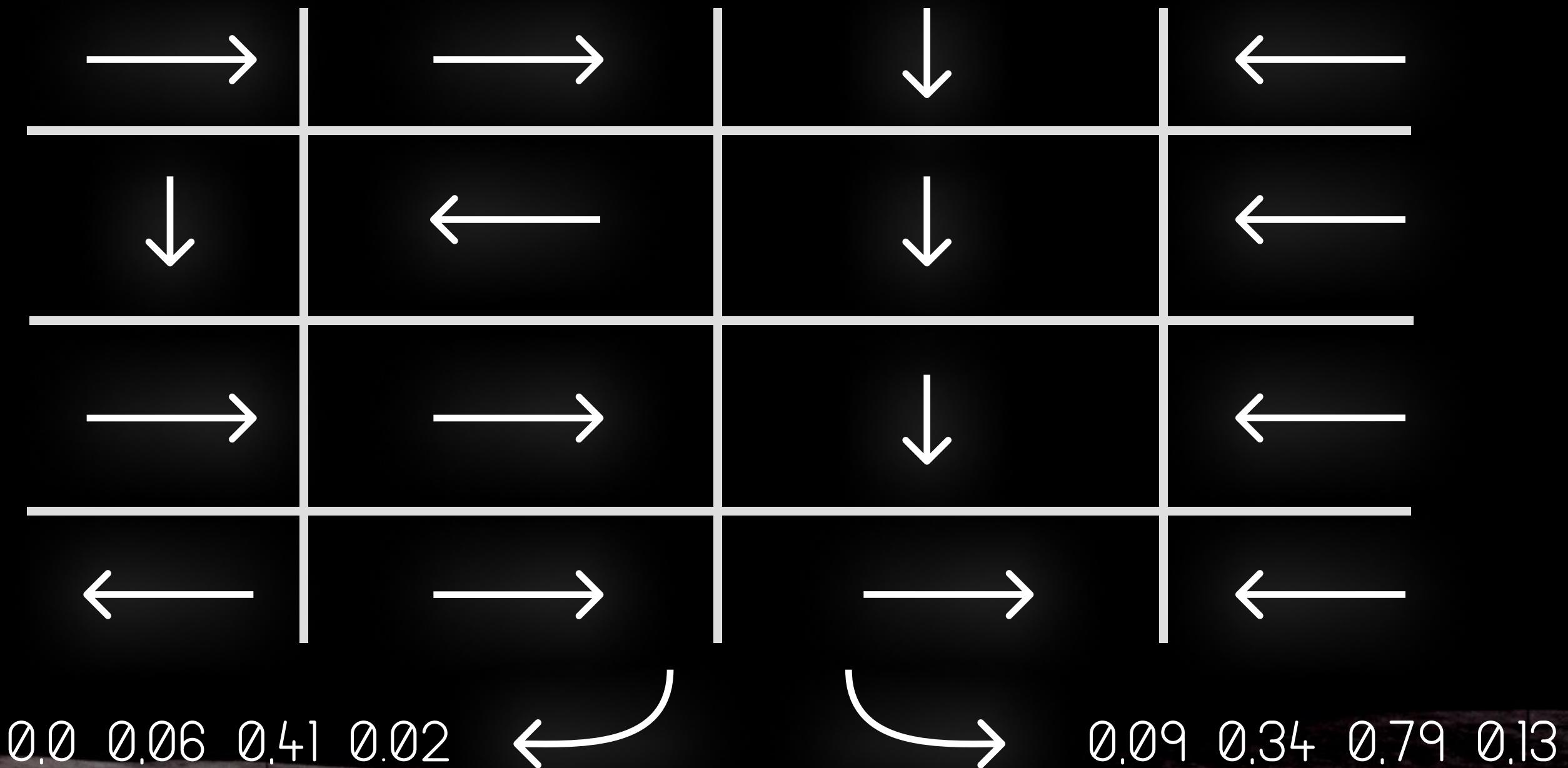
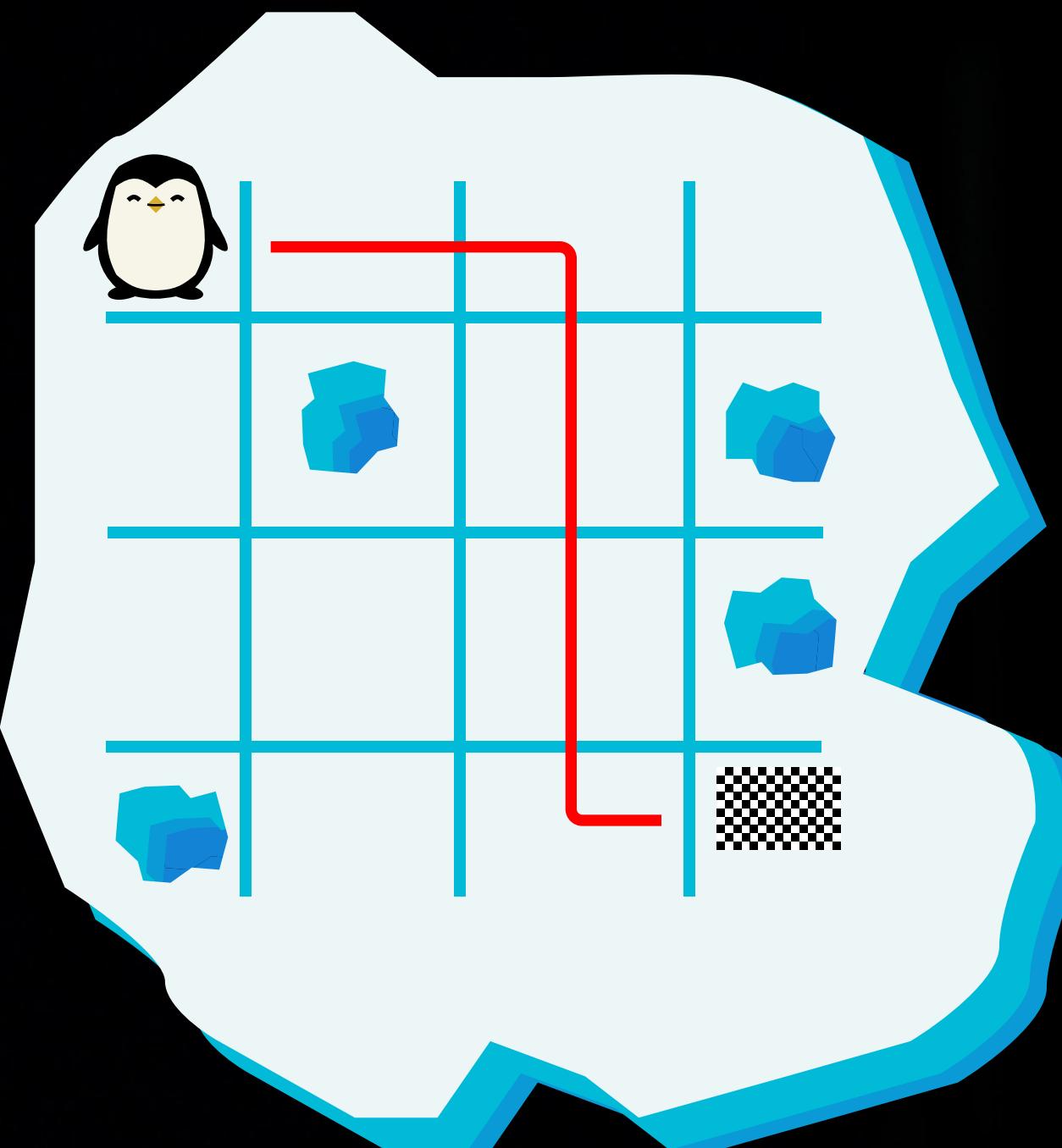


# APPRENTISSAGE PAR RENFORCEMENT

## ■ EXPLICATION

MISE EN APPLICATION  
EPISODE 400

LE TABLEAU DE QUALITÉ FINI PAR SE STABILISER. VOICI LE CHEMIN  
TROUVÉ PAR L'ALGORITHME POUR MAXIMISER LA RÉCOMPENSE



# APPRENTISSAGE PAR RENFORCEMENT

## ■ EXPLICATION

ANCIENNE VALEUR :  $s'$   
VALEUR ACTUEL :  $s$   
NOUVELLE VALEUR :  $s^*$

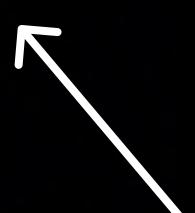
### CALCUL DE RÉSULTATS

AU DEBUT DE L'APPRENTISSAGE LE TABLEAU EST INITIALISÉ À ZERO

FORMULE :

$$Q(s, a) =$$

$$(1-\alpha)Q''(s', a') + \alpha(R + \text{MAX}(Q(s^*:)))$$



TAUX D'APPRENTISSAGE

	←	↓	→	↑
(0, 0)	0.0	0.0	0.0	0.0
(0, 1)	0.0	0.0	0.0	0.0
(1, 0)	0.0	0.0	0.0	0.0
(1, 1)	0.0	0.0	0.0	0.0

# APPRENTISSAGE PAR RENFORCEMENT

## ■ EXPLICATION

$$Q(s, a) = (1-\alpha)Q''(s', a') + \alpha(R + \text{MAX}(Q(s^*)))$$

CALCUL DES RÉSULTATS EXEMPLE

$$s = 0, 1 \quad a = 1 \rightarrow s^* = 0, 0$$

ON PONDÈRE L'ANCIENNE  
VALEUR AVEC LE TAUX  
D'APPRENTISSAGE

$$(1-\alpha)Q''(s', a') = 0.9 * 0.23  
= 0.207$$

	←	↓	→	↑
(0, 0)	0.12	0.34	0.17	0.45
(0, 1)	0.15	0.23	0.36	0.81
(1, 0)	0.11	0.65	0.87	0.74
(1, 1)	0.8	0.32	0.01	0.95

# APPRENTISSAGE PAR RENFORCEMENT

## ■ EXPLICATION

CALCULE DE RÉSULTATS EXEMPLE

$$s = 0.1 \quad a = 1 \rightarrow s^* = 0.0$$

$$R = 0.8$$

ON CALCULE LE RESULTAT THÉORIQUE  
TOTAL EN AJOUTANT LE RÉSULTAT  
THÉORIQUE DE L'ACTION SUIVANTE

$$\begin{aligned} a(R + \text{MAX}(Q(s^*:))) &= 0.1(0.8 + 0.45) \\ &= 0.125 \end{aligned}$$

$$Q(s, a) = (1-a)Q''(s', a') + a(R + \text{MAX}(Q(s^*:)))$$

$$0.207 + 0.125$$

	←	↓	→	↑
(0, 0)	0.12	0.34	0.17	0.45
(0, 1)	0.15	0.23	0.36	0.81
(1, 0)	0.11	0.65	0.87	0.74
(1, 1)	0.8	0.32	0.01	0.95

# APPRENTISSAGE PAR RENFORCEMENT

## ■ EXPLICATION

CALCULE DE RÉSULTATS EXEMPLE

$$s = 0, | a = 1 \rightarrow s^* = 0, 0$$

$$Q(s.a) = 0.332$$

$$Q(s.a) = (1-\alpha)Q''(s'.a') + \alpha(R + \text{MAX}(Q(s^*:)))$$

$$0.207 + 0.125$$

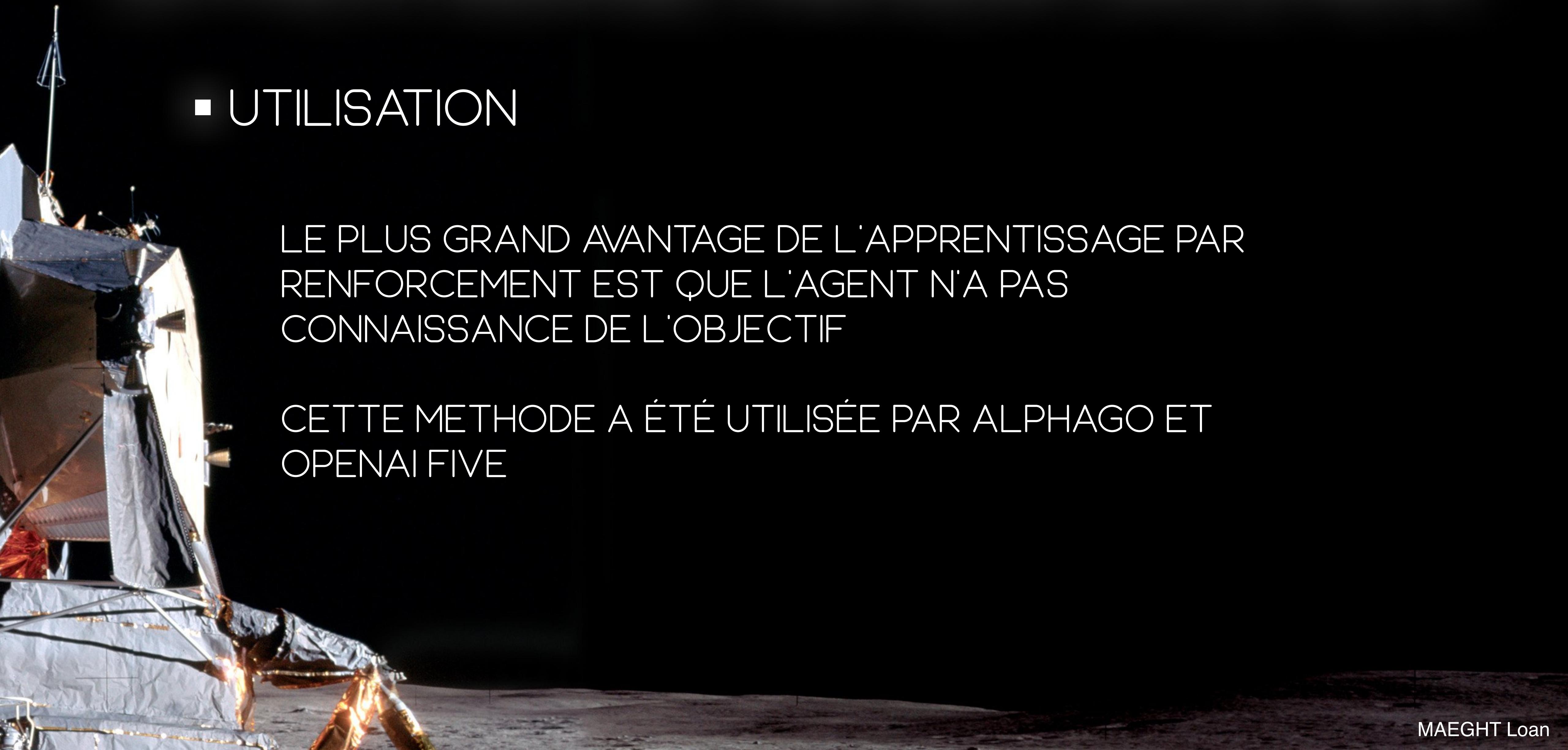
	←	↓	→	↑
(0, 0)	0.12	0.34	0.17	0.45
(0, 1)	0.15	0.332	0.36	0.81
(1, 0)	0.11	0.65	0.87	0.74
(1, 1)	0.8	0.32	0.01	0.95

# APPRENTISSAGE PAR RENFORCEMENT

## ■ UTILISATION

LE PLUS GRAND AVANTAGE DE L'APPRENTISSAGE PAR RENFORCEMENT EST QUE L'AGENT N'A PAS CONNAISSANCE DE L'OBJECTIF

CETTE METHODE A ÉTÉ UTILISÉE PAR ALPHAGO ET OPENAI FIVE



# L'ATERRISSEUR LUNAIRE

## ■ ORIGINE

L'ENVIRONEMENT "LUNARLANDER-V2" EST PROPOSÉ PAR OPENAI DANS LEUR LIBRAIRIE.

CETTE LIBRAIRIE COMPORTE BEAUCOUP D'AUTRES ENVIRONEMENTS D'ENTRAÎNEMENT POUR LES ALGORITHMES D'APPRENTISSAGE PAR RENFORCEMENT

LEUR SITE: [GYM.OPENAI.COM](https://gym.openai.com)



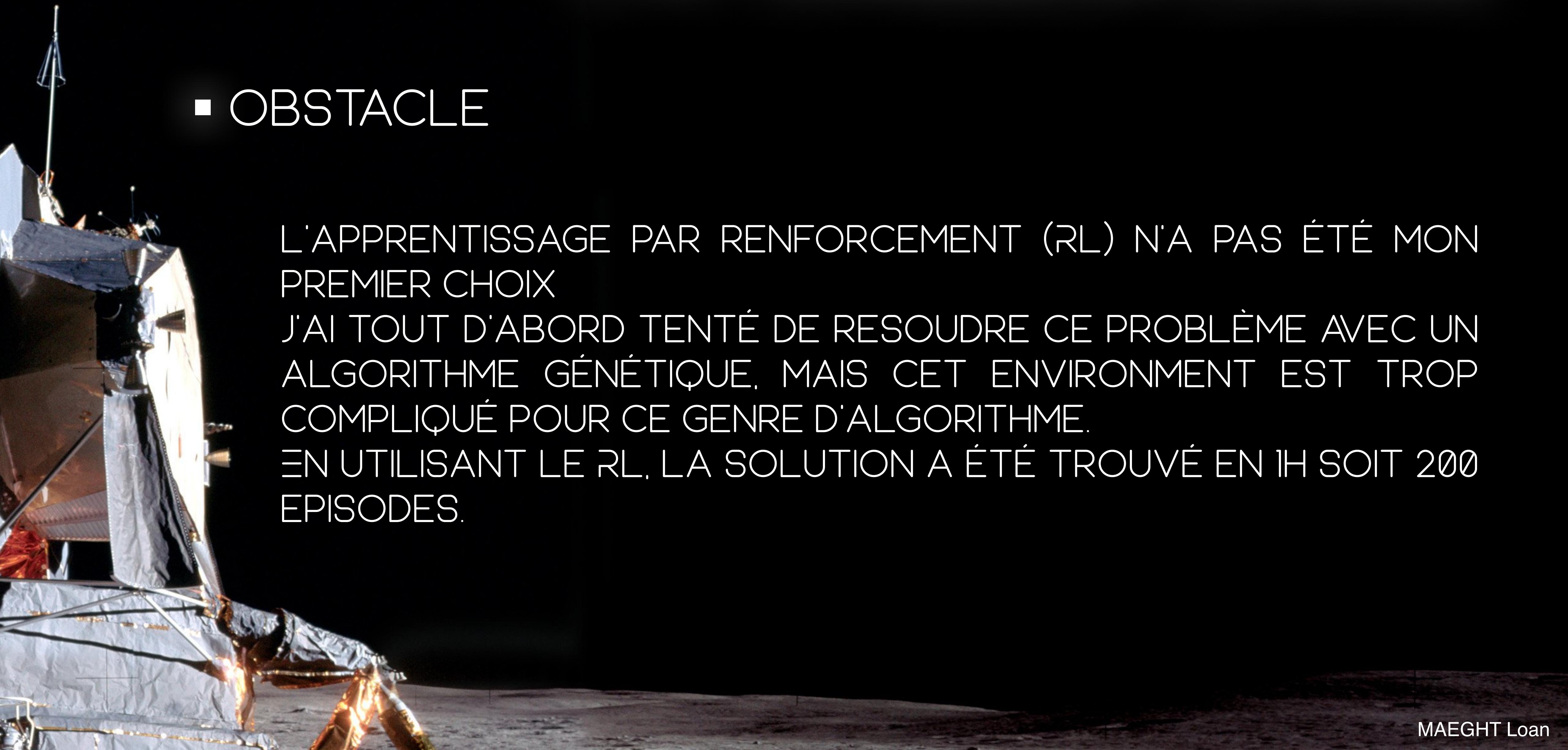
# L'ATTERRISSEUR LUNAIRE

## ■ OBSTACLE

L'APPRENTISSAGE PAR RENFORCEMENT (RL) N'A PAS ÉTÉ MON PREMIER CHOIX

J'AI TOUT D'ABORD TENTÉ DE RESOUDRE CE PROBLÈME AVEC UN ALGORITHME GÉNÉTIQUE. MAIS CET ENVIRONMENT EST TROP COMPLIQUÉ POUR CE GENRE D'ALGORITHME.

EN UTILISANT LE RL, LA SOLUTION A ÉTÉ TROUVÉ EN 1H SOIT 200 EPISODES.



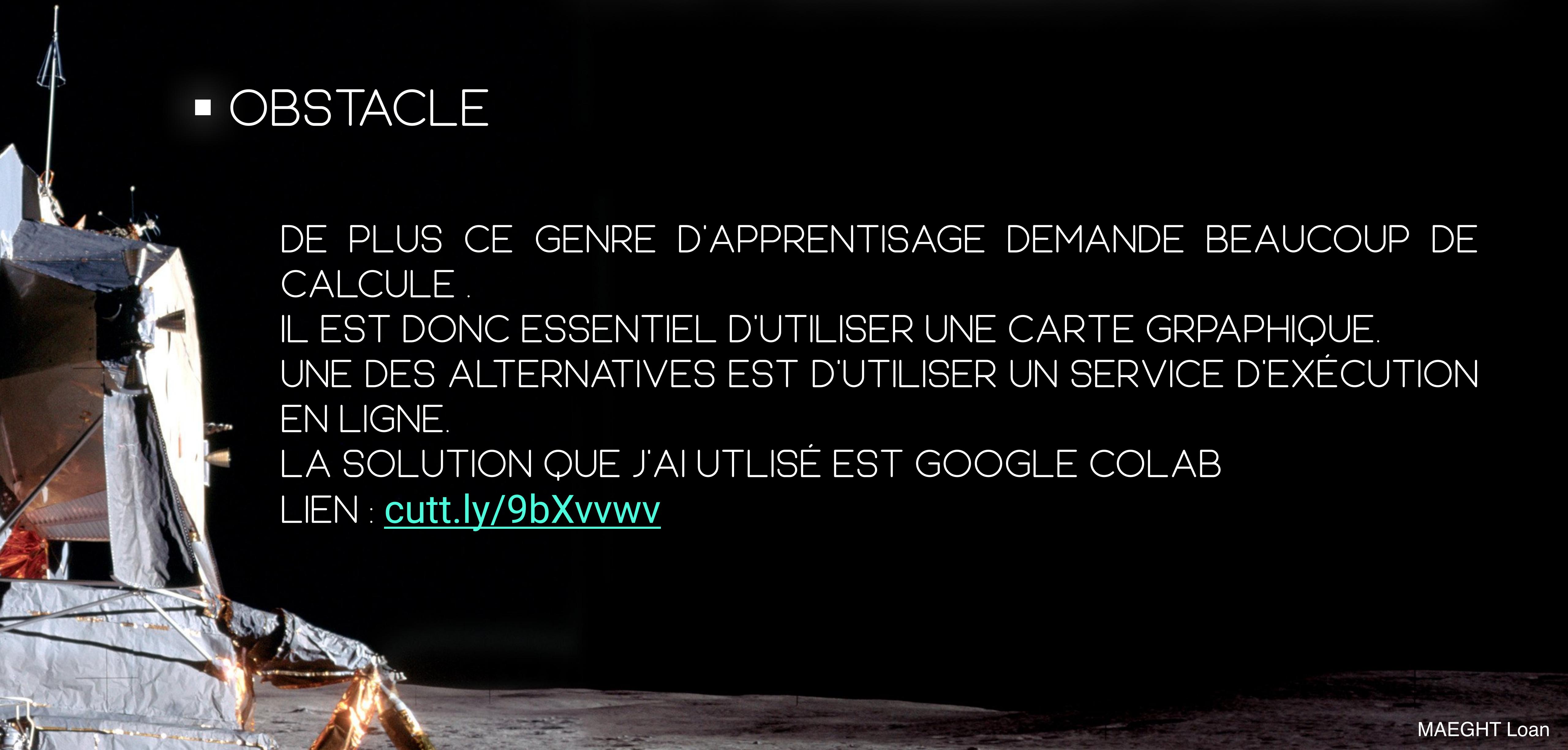
# L'ATTERRISSEUR LUNAIRE

## ■ OBSTACLE

DE PLUS CE GENRE D'APPRENTISAGE DEMANDE BEAUCOUP DE CALCULE .

IL EST DONC ESSENTIEL D'UTILISER UNE CARTE GRAPHIQUE.  
UNE DES ALTERNATIVES EST D'UTILISER UN SERVICE D'EXÉCUTION EN LIGNE.

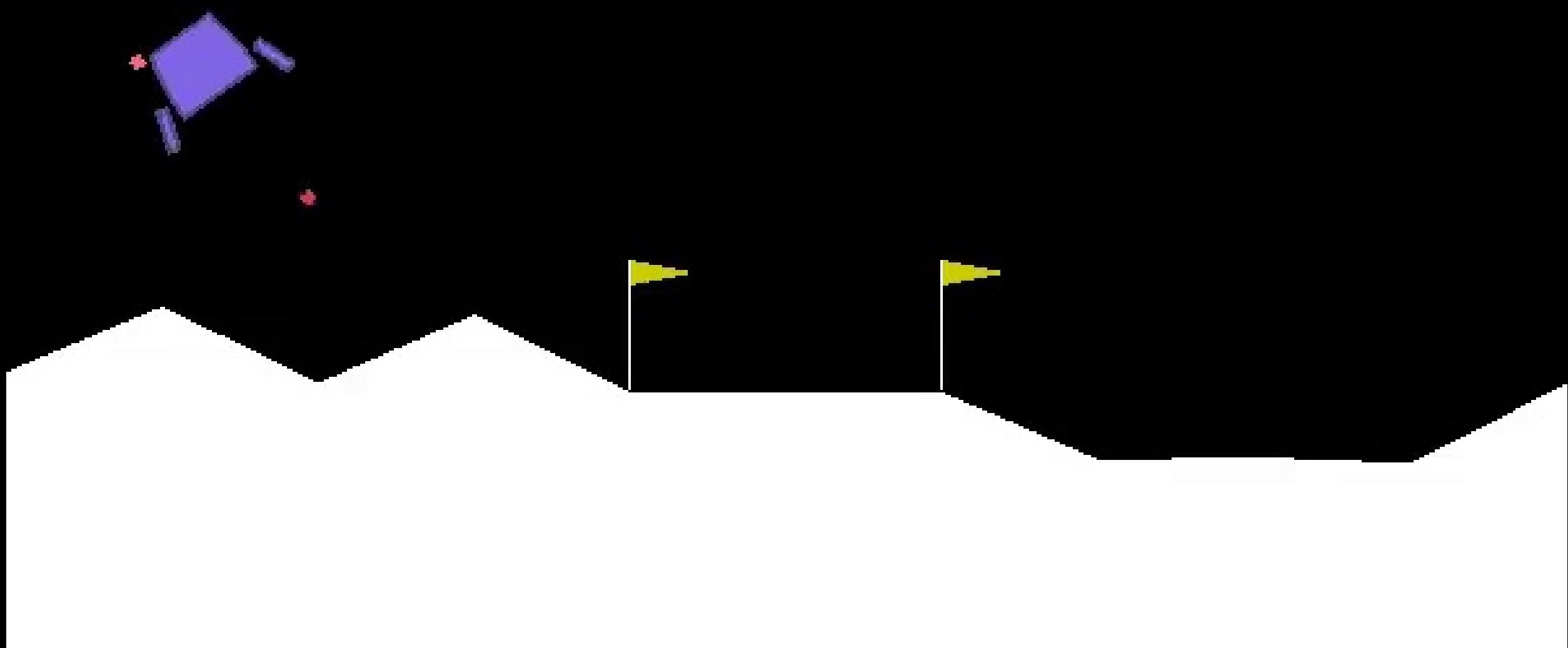
LA SOLUTION QUE J'AI UTILISÉ EST GOOGLE COLAB  
LIEN : [cutt.ly/9bXvvwv](https://cutt.ly/9bXvvwv)



# L'ATERRISSEUR LUNAIRE

## ■ RÉSOLUTION

VIDEO  
COMPIRATION



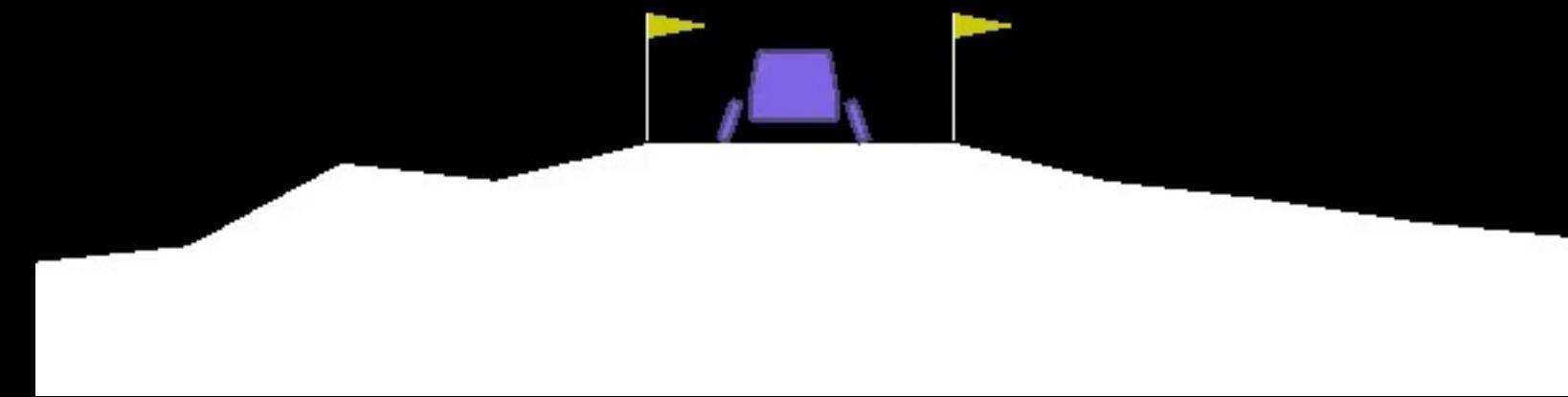
CLIQUEZ SUR L'IMAGE POUR OUVRIR LE LIEN

# L'ATTERRISSEUR LUNAIRE

## ■ RÉSOLUTION

8 ENTRÉES

COORDONNÉE HORIZONTALE  
COORDONNÉE VERTICALE  
VITESSE HORIZONTALE  
VITESSE VERTICALE  
ANGLE  
VITESSE ANGULAIRE  
PREMIER PIED CONTACT  
DEUXIÈME PIED CONTACT



4 SORTIES / ACTION

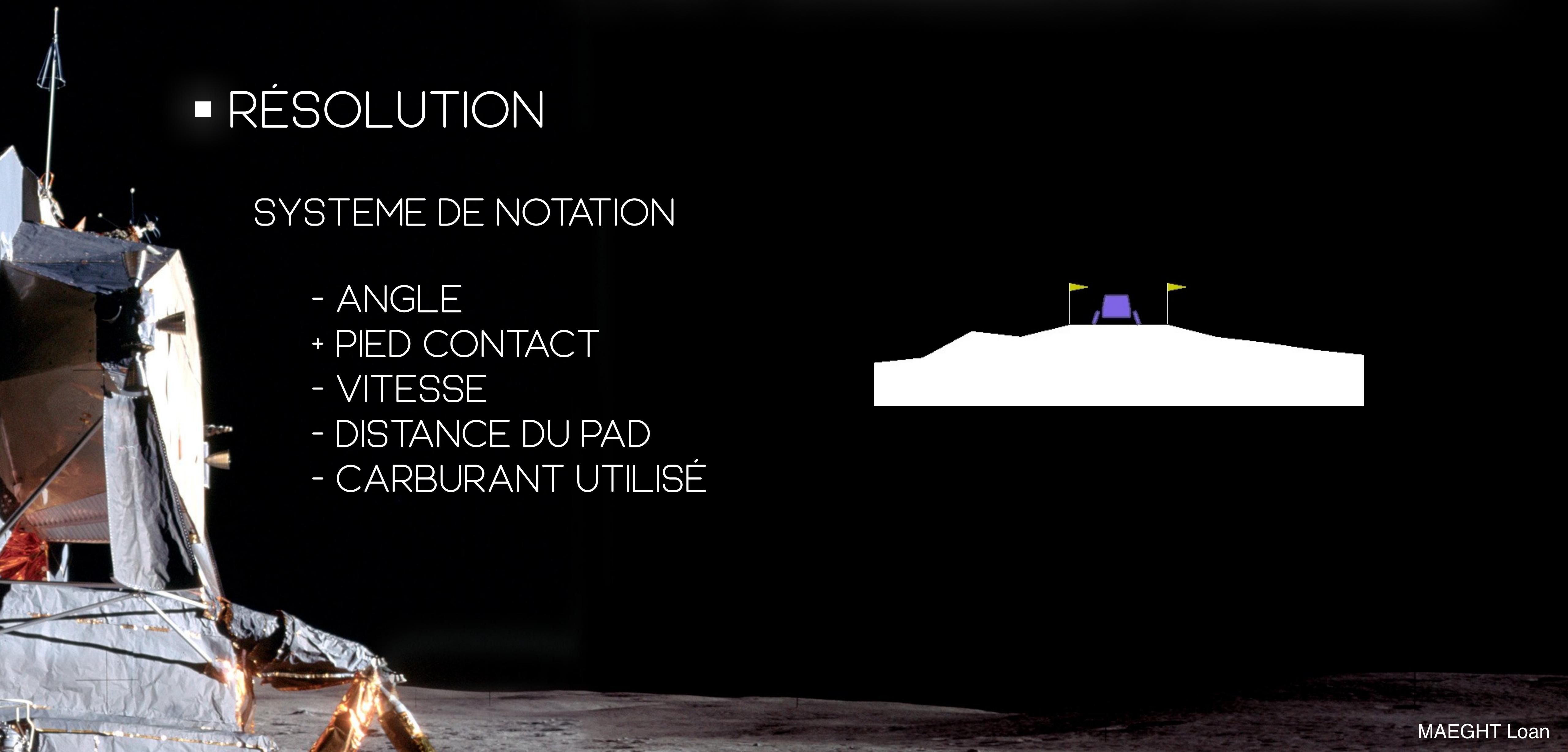
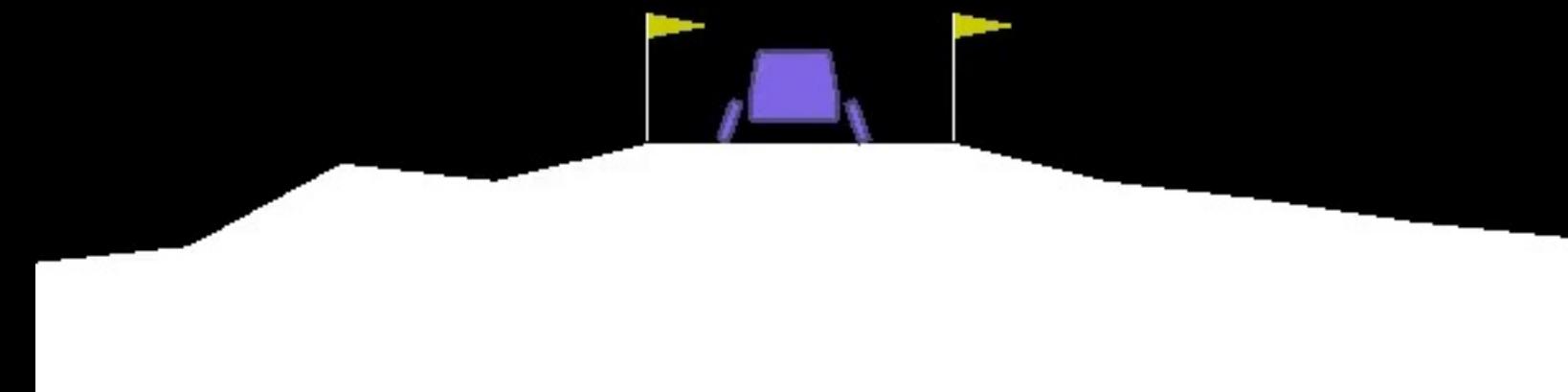
NE RIEN FAIRE  
MOTEUR PRINCIPAL  
MOTEUR SECONDAIRE GAUCHE  
MOTEUR SECONDAIRE DROIT

# L'ATERRISSEUR LUNAIRE

## ■ RÉSOLUTION

### SYSTEME DE NOTATION

- ANGLE
- + PIED CONTACT
- VITESSE
- DISTANCE DU PAD
- CARBURANT UTILISÉ



# L'ATTERRISSEUR LUNAIRE

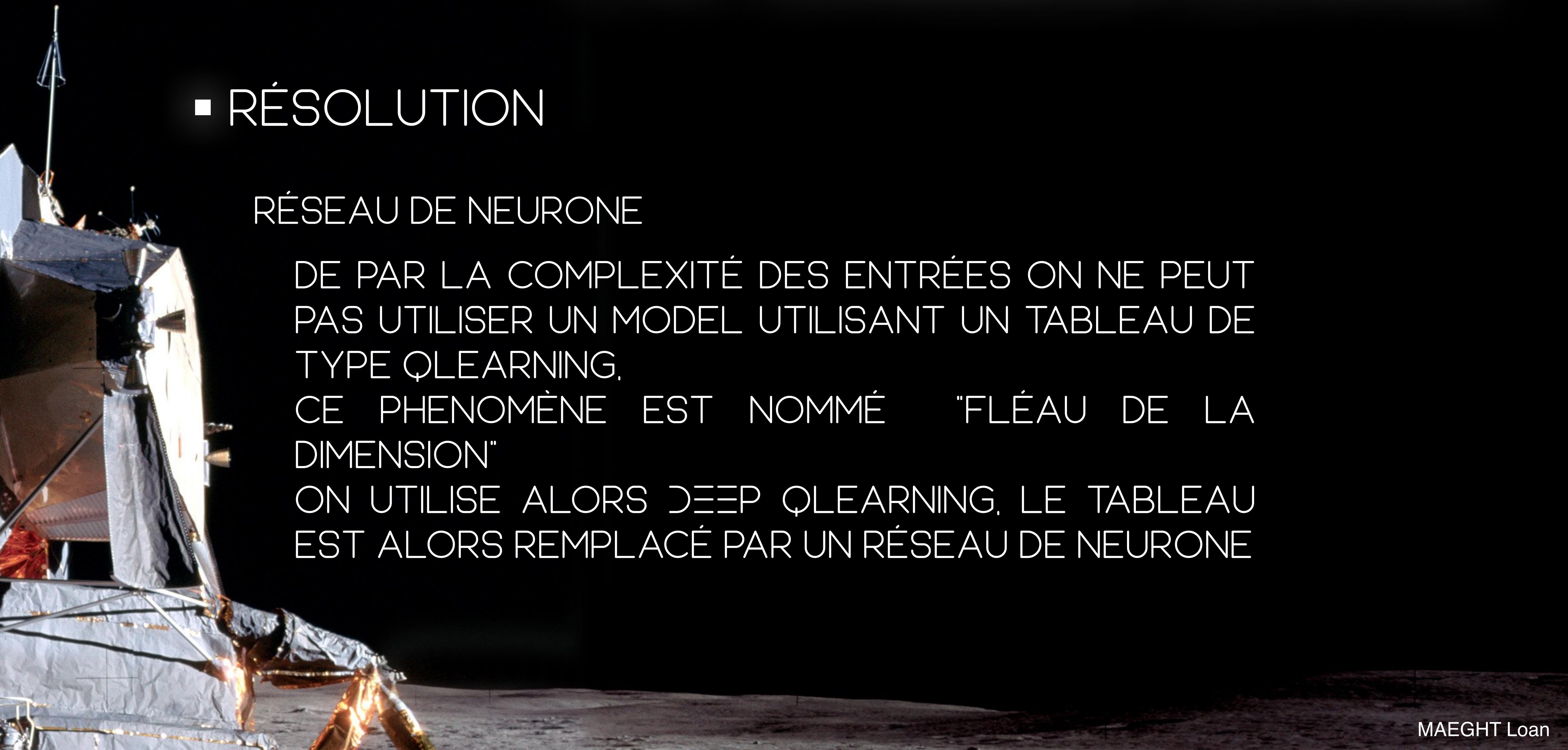
## ■ RÉSOLUTION

### RÉSEAU DE NEURONE

DE PAR LA COMPLEXITÉ DES ENTRÉES ON NE PEUT PAS UTILISER UN MODEL UTILISANT UN TABLEAU DE TYPE QLEARNING.

CE PHENOMÈNE EST NOMMÉ "FLEAU DE LA DIMENSION"

ON UTILISE ALORS DEEP QLEARNING. LE TABLEAU EST ALORS REMPLACÉ PAR UN RÉSEAU DE NEURONE

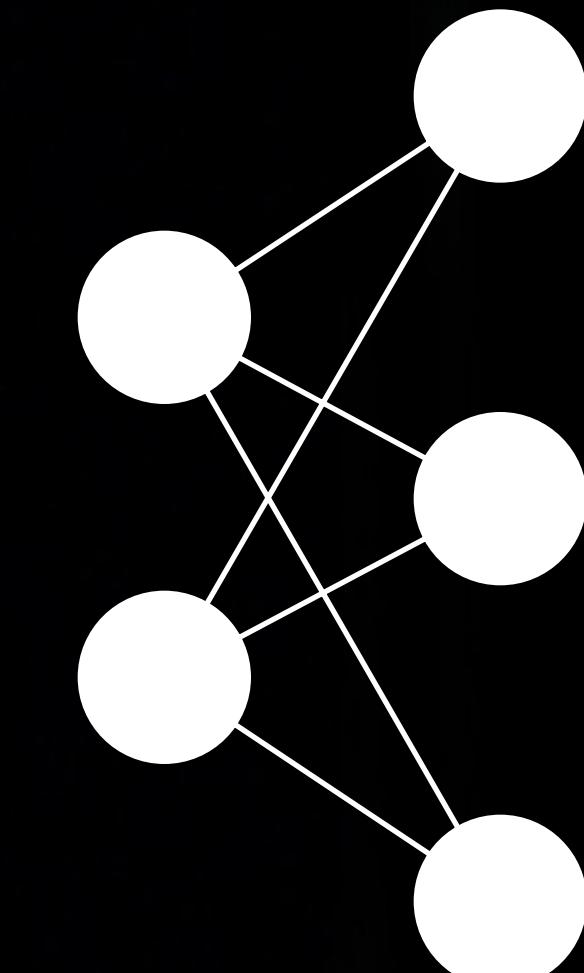


# L'ATERRISSEUR LUNAIRE

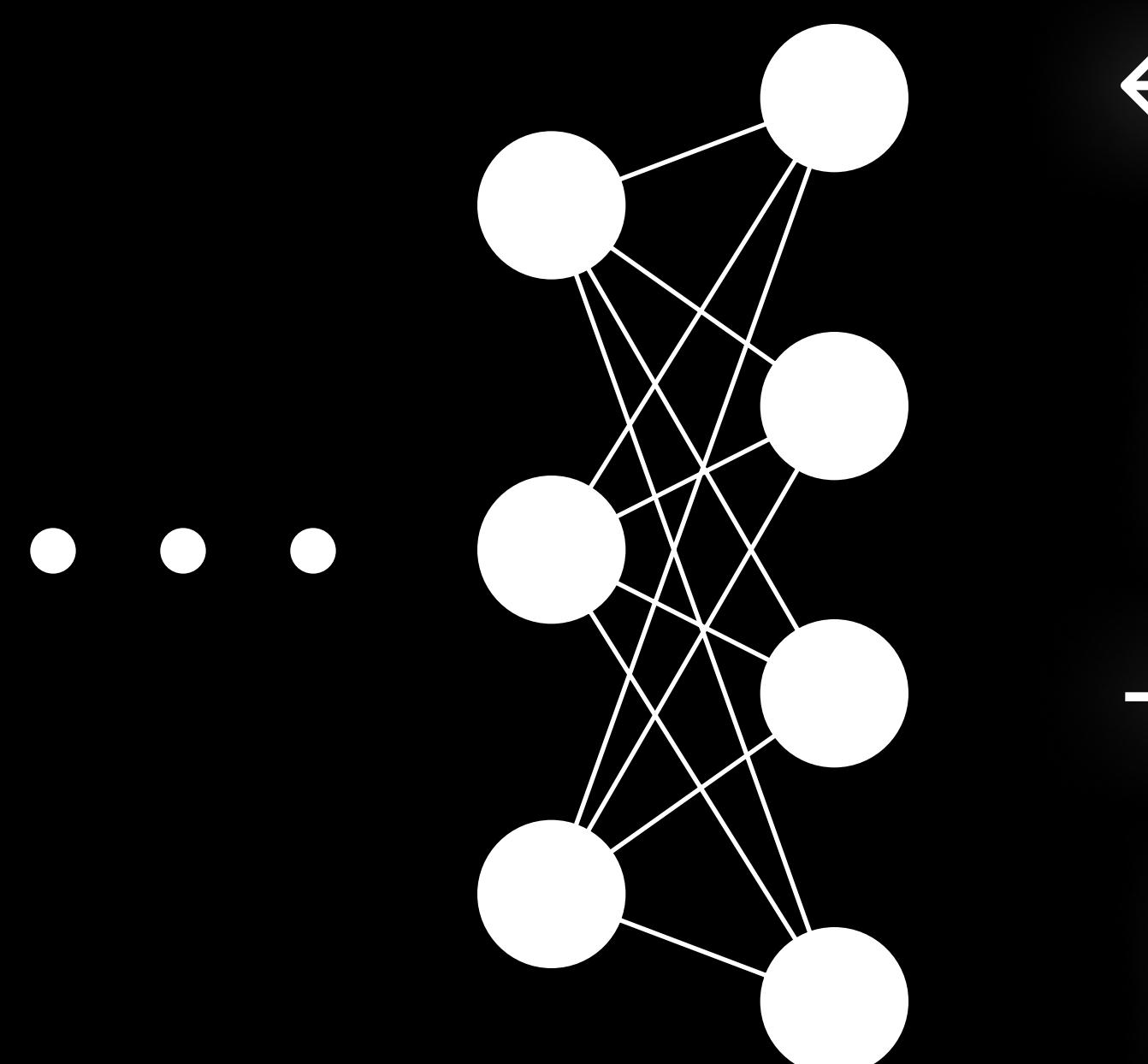
## ■ RÉSOLUTION

RÉSEAU DE NEURONE

COOR X  
4  
COOR Y  
7



FONCTIONNEMENT



← 0.54  
↓ 0.18  
→ 0.91  
↑ 0.46



# L'ATERRISSEUR LUNAIRE

- RÉSOLUTION

TOUT LES FICHIERS SONT DISPONIBLES SUR LE GITHUB:

<https://github.com/qypol342/lunarlander-v2-q-learning>



ATTERIR SUR LA  
LUNE ENDORMANT

MAEGHT LOAN

