

ATTERIR SUR LA LUNE ENDORMANT

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

MAEGHT Loan



SOMMAIRE

■ APPRENTISSAGE PAR RENFORCEMENT

DESCRIPTION
EXPLICATION
UTILISATION

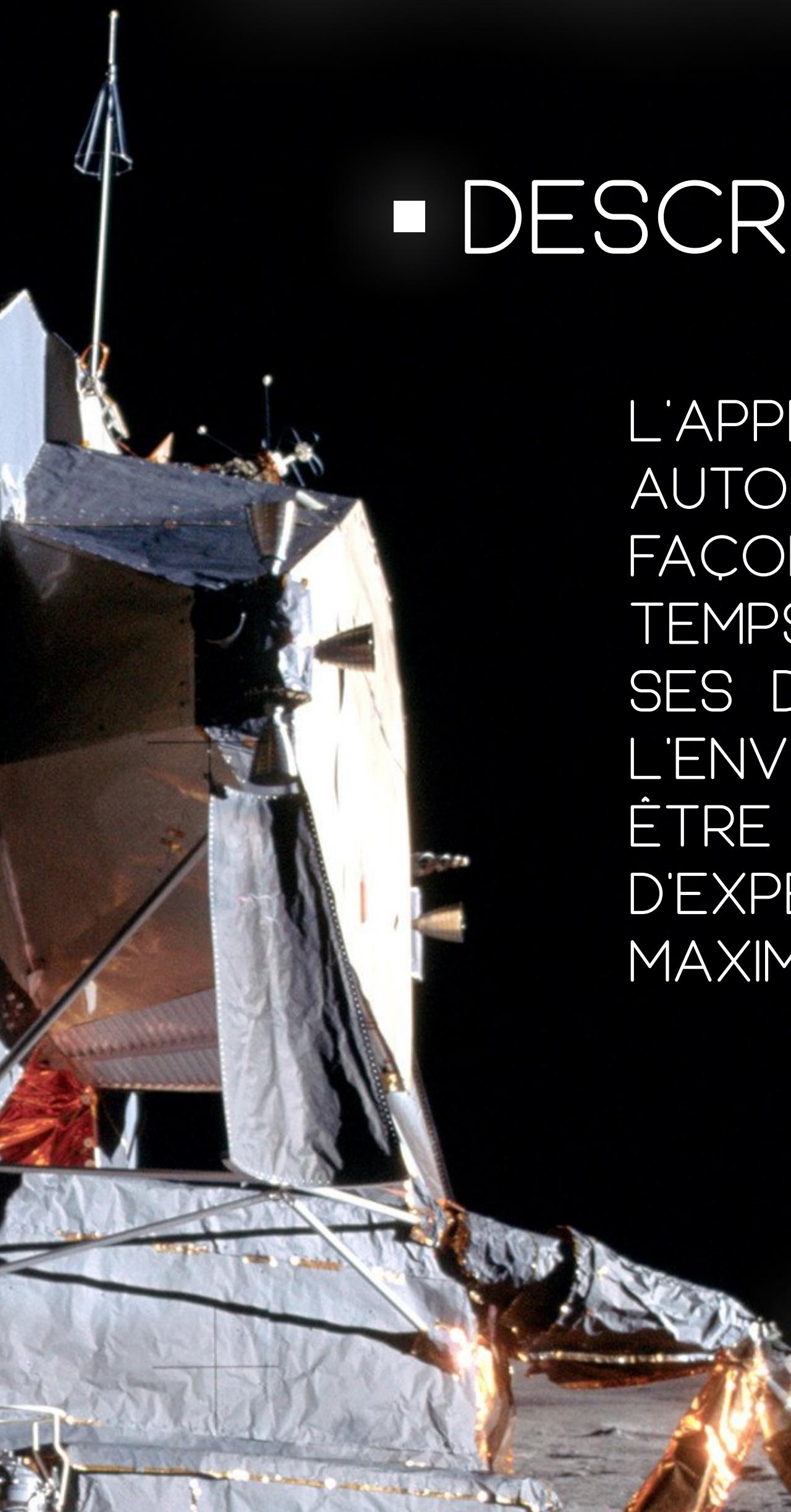
■ L'ATTERRISSEUR LUNAIRE

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 OPTIMAL, EN CE SENS QU'IL MAXIMISE LA SOMME DES RÉCOMPENSES AU COURS DU TEMPS.

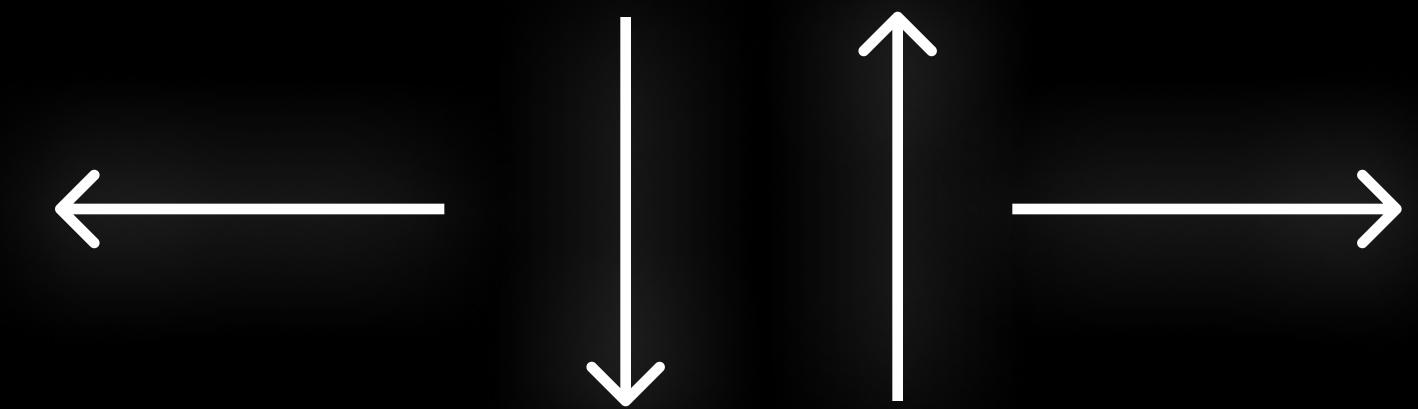
APPRENTISSAGE PAR RENFORCEMENT

■ EXPLICATION

L'AGENT POSSEDE TROIS ELEMENTS IMPORTANTS.

UNE LISTE D'ACTION POSSIBLE NOTÉ : Δ

EXEMPLE UNE DIRECTION DE DEPLACEMENT



UNE LISTE D'ETAT NOTÉ : S

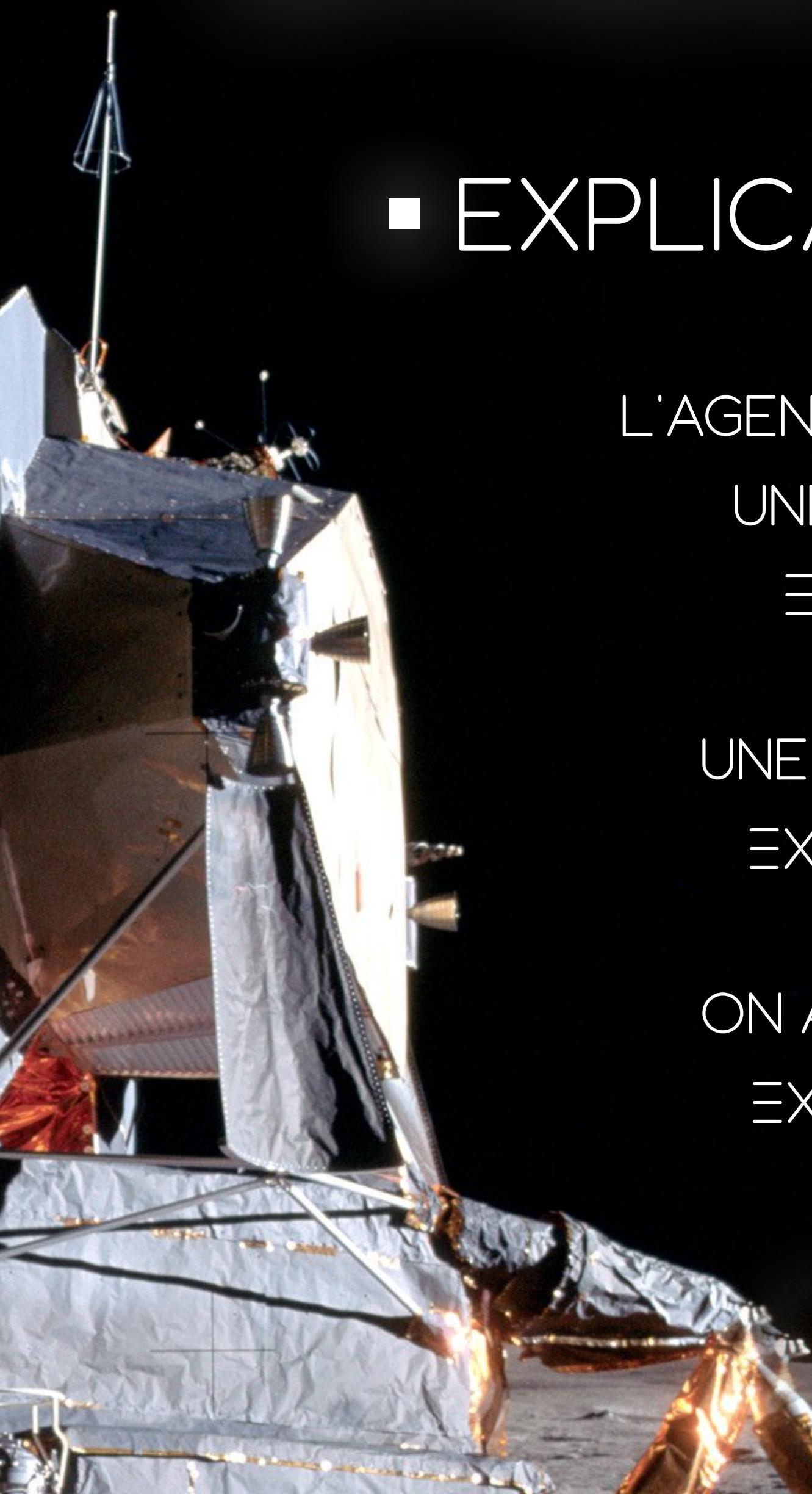
EXEMPLE COORDONÉE, VITESSE



ON ASSOCIE A CHAQUE TUPLE UN SCORE NOTÉ : R

EXEMPLE RALEMENTIR AVANT DE TOUCHER LE SOL

+ 1



APPRENTISSAGE PAR RENFORCEMENT

■ EXPLICATION

L'AGENT POSSEDE DEUX FONCTIONS :

FONCTION EPSILON NOTÉ : ϵ

CORRESPOND AU TAUX D'EXPLORATION CONTRE EXPLOITATION

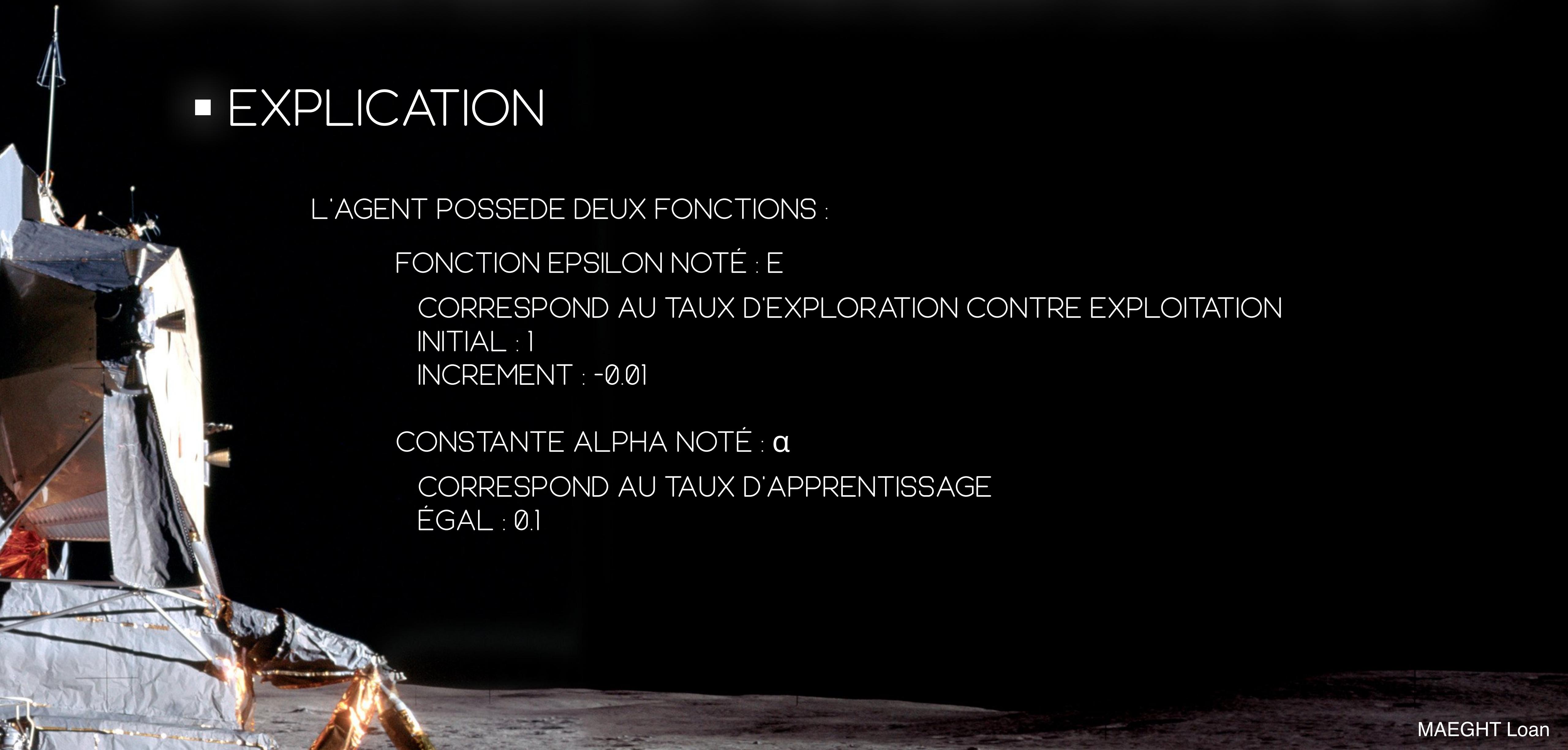
INITIAL : 1

INCREMENT : -0.01

CONSTANTE ALPHA NOTÉ : α

CORRESPOND AU TAUX D'APPRENTISSAGE

ÉGAL : 0.1



APPRENTISSAGE PAR RENFORCEMENT

■ EXPLICATION

L'AGENT POSSEDE UN TABLEAU QUI MET EN RELATION LE TUPLET STATU ET ACTIONS POUR DONNER LEUR SCORE

EXEMPLE DE TABLEAU :
STATU → COORDONNÉ
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 :

LE PREMIER : LE CHOIX ENTRE EXPLORATION
ET EXPLOITATION

ON PREND UN NOMBRE ALÉATOIRE
COMPRIS ENTRE 0 ET 1

SI LE NOMBRE EST PLUS PETIT QUE
 ϵ ALORS ON CHOISI UNE ACTION
ALEATOIRE

	←	↓	↑	→
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 :

LE DEUXIÈME : RECHERCHE DU MEILLEUR
RESULTAT

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

EXEMPLE:

0.0 --> 0.45 DEPLACEMENT A GAUCHE

	←	↓	↑	→
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

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

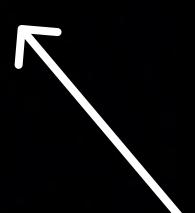
CALCULE DE RÉSULTATS

AU DEBUT DE L'APPRENTISSAGE LE TABLEAU EST INITIALISER A 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	0.0	0.0	0.0
↑	0.0	0.0	0.0	0.0
→	0.0	0.0	0.0	0.0

APPRENTISSAGE PAR RENFORCEMENT

■ EXPLICATION

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

CALCULE DE RÉSULTATS EXEMPLE

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

ON CALCULE CE QUE L'ON GARDE DE L'ANCIEN VALEUR

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

	←	↓	↑	→
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

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

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

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

$$0.207 + 0.125$$

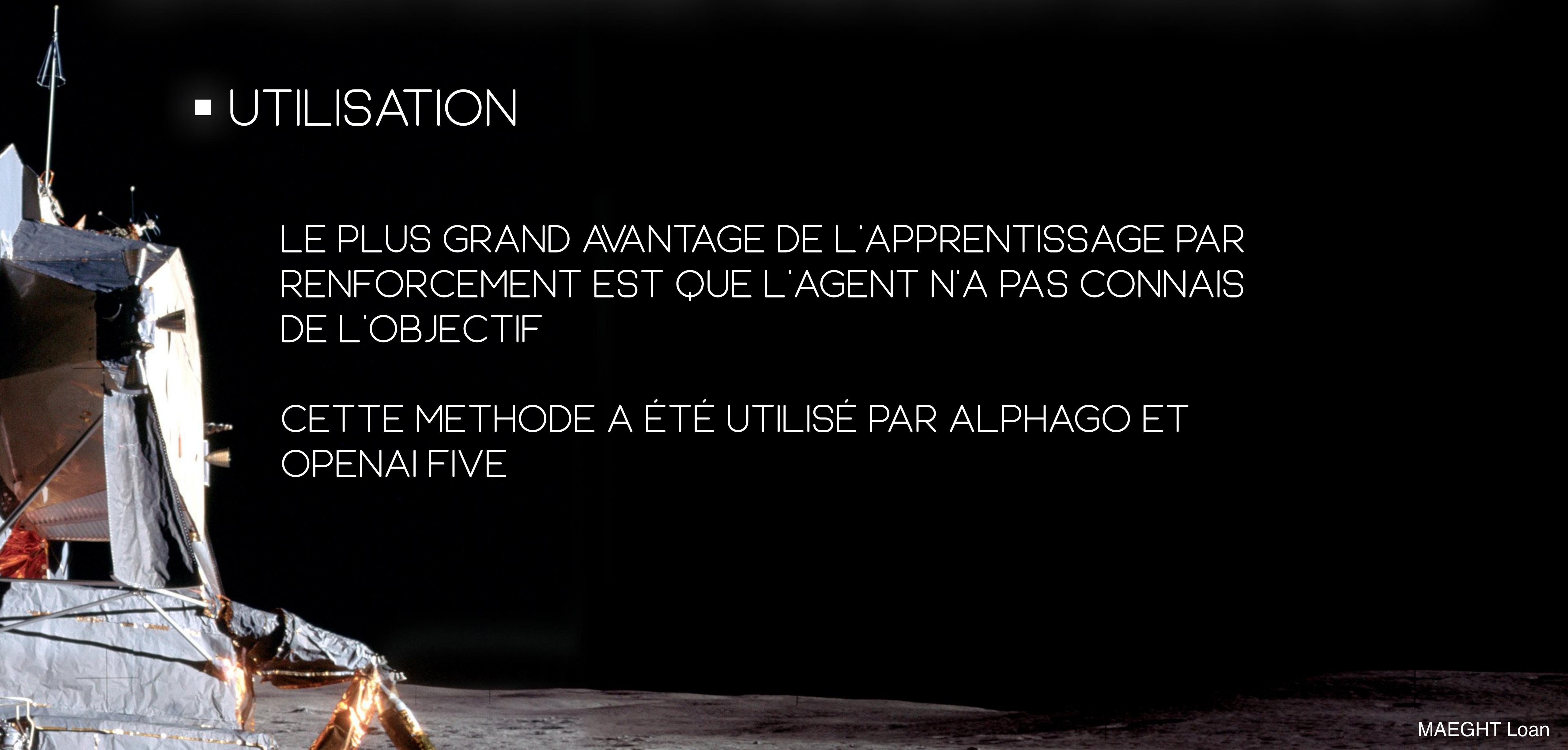
	←	↓	↑	→
0 . 0	0.12	0.34	0.17	0.45
0 . 1	0.15	0.332	0.36	0.81
1 . 0	0.11	0.65	0.87	0.74
1 . 1	0.8	0.32	0.01	0.95

APPRENTISSAGE PAR RENFORCEMENT

■ UTILISATION

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

CETTE METHODE A ÉTÉ UTILISÉ PAR ALPHAGO ET OPENAI FIVE



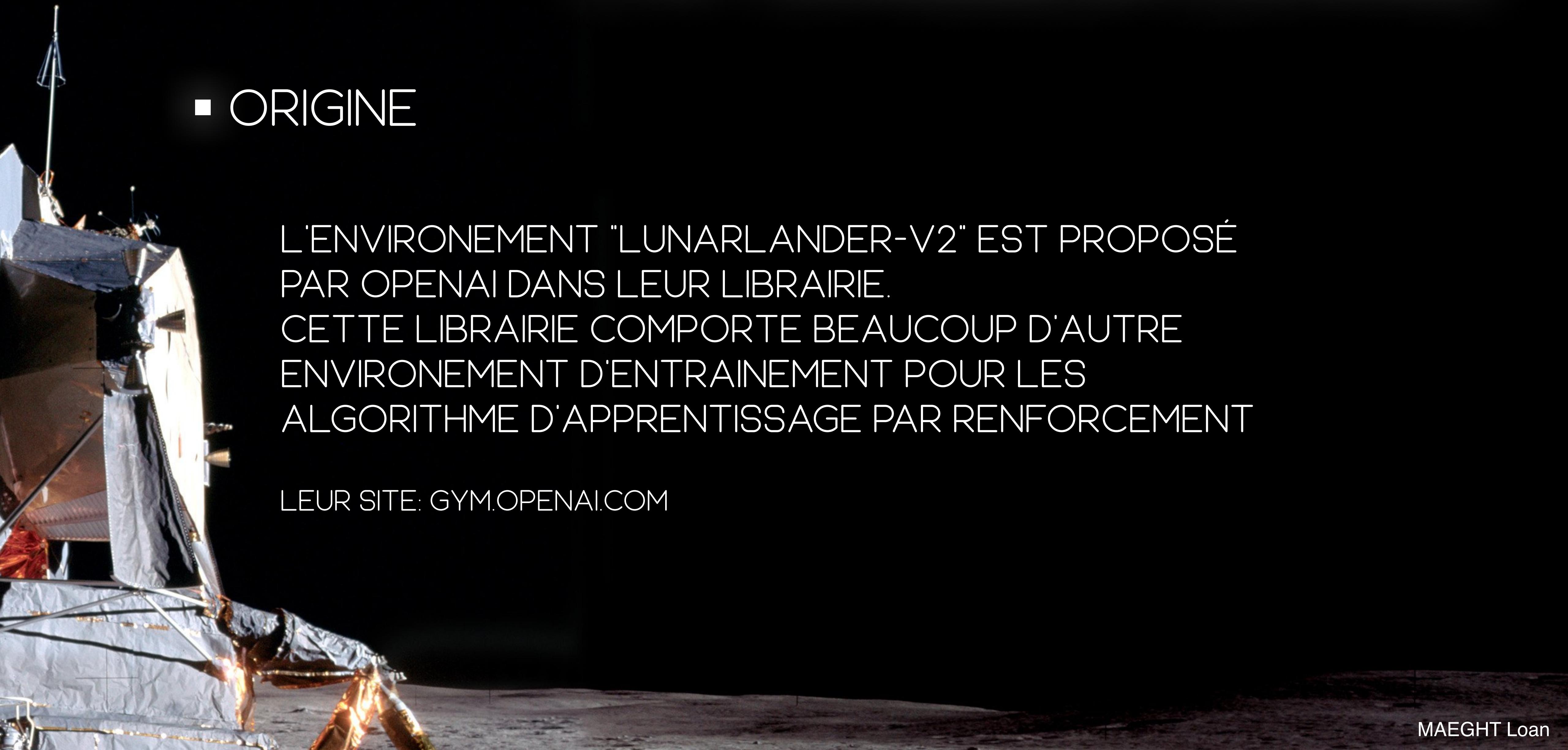
L'ATERRISSEUR LUNAIRE

■ ORIGINE

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

CETTE LIBRAIRIE COMPORTE BEAUCOUP D'AUTRE ENVIRONEMENT D'ENTRAINEMENT POUR LES ALGORITHME D'APPRENTISSAGE PAR RENFORCEMENT

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



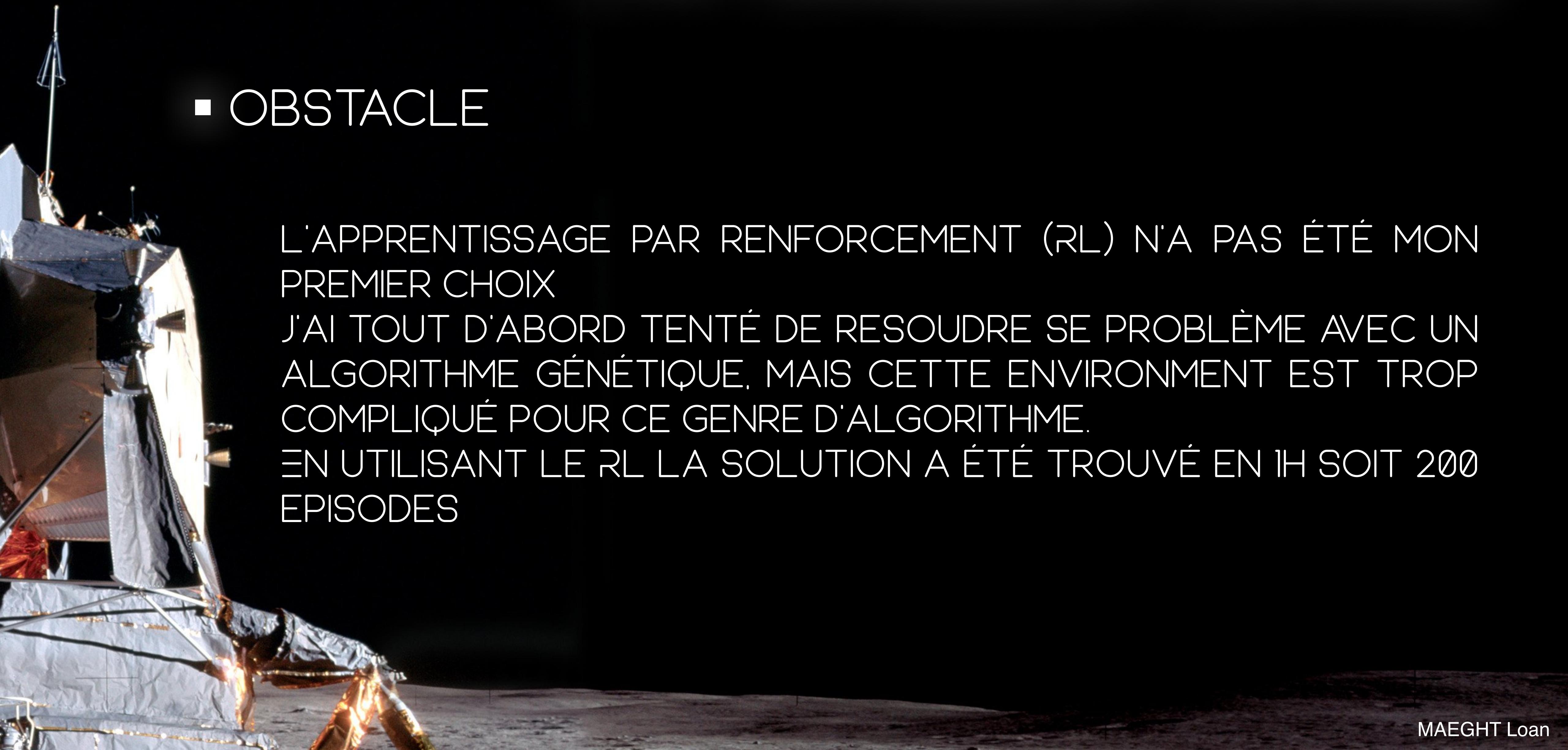
L'ATTERRISSEUR LUNAIRE

■ OBSTACLE

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

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

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



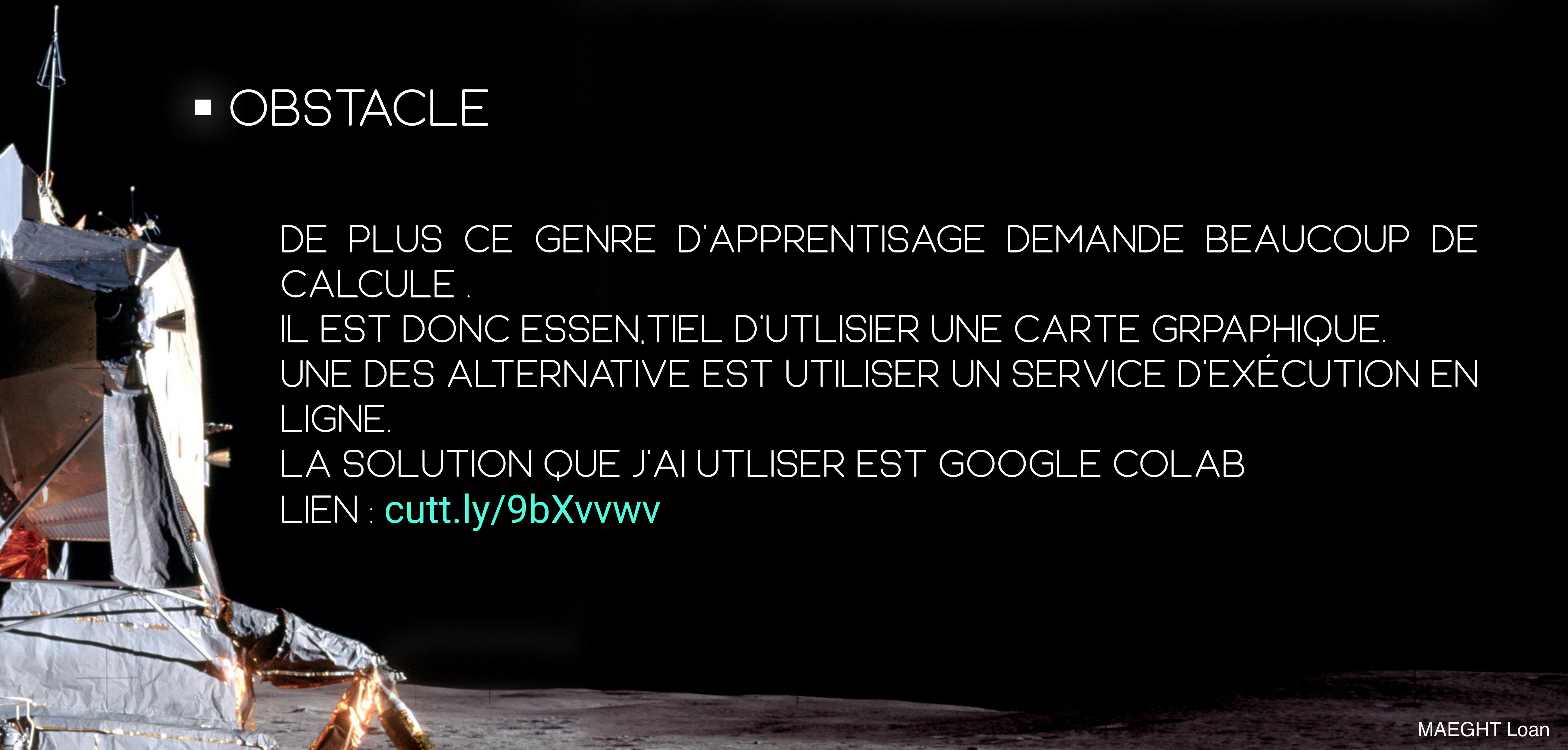
L'ATERRISSEUR LUNAIRE

■ OBSTACLE

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

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

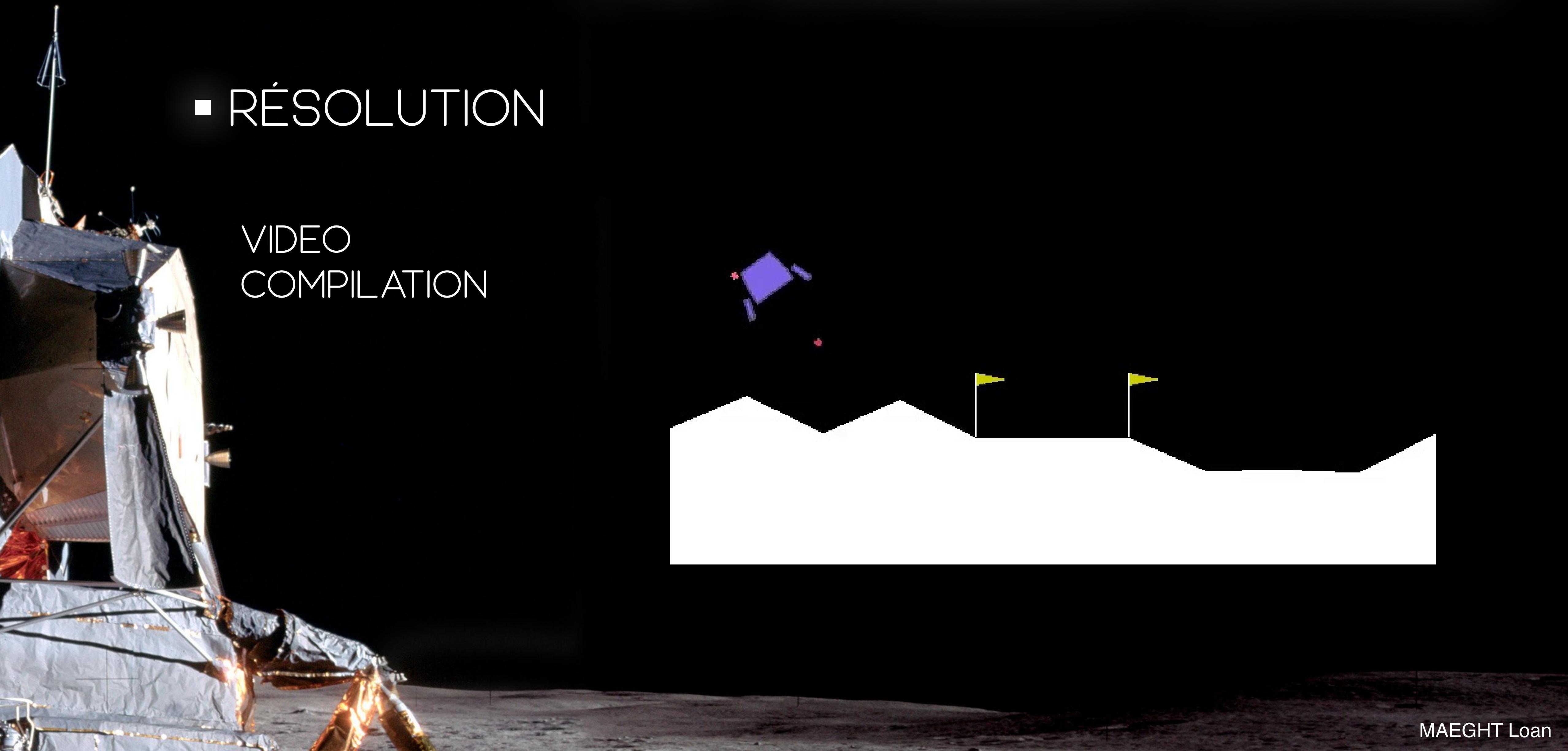
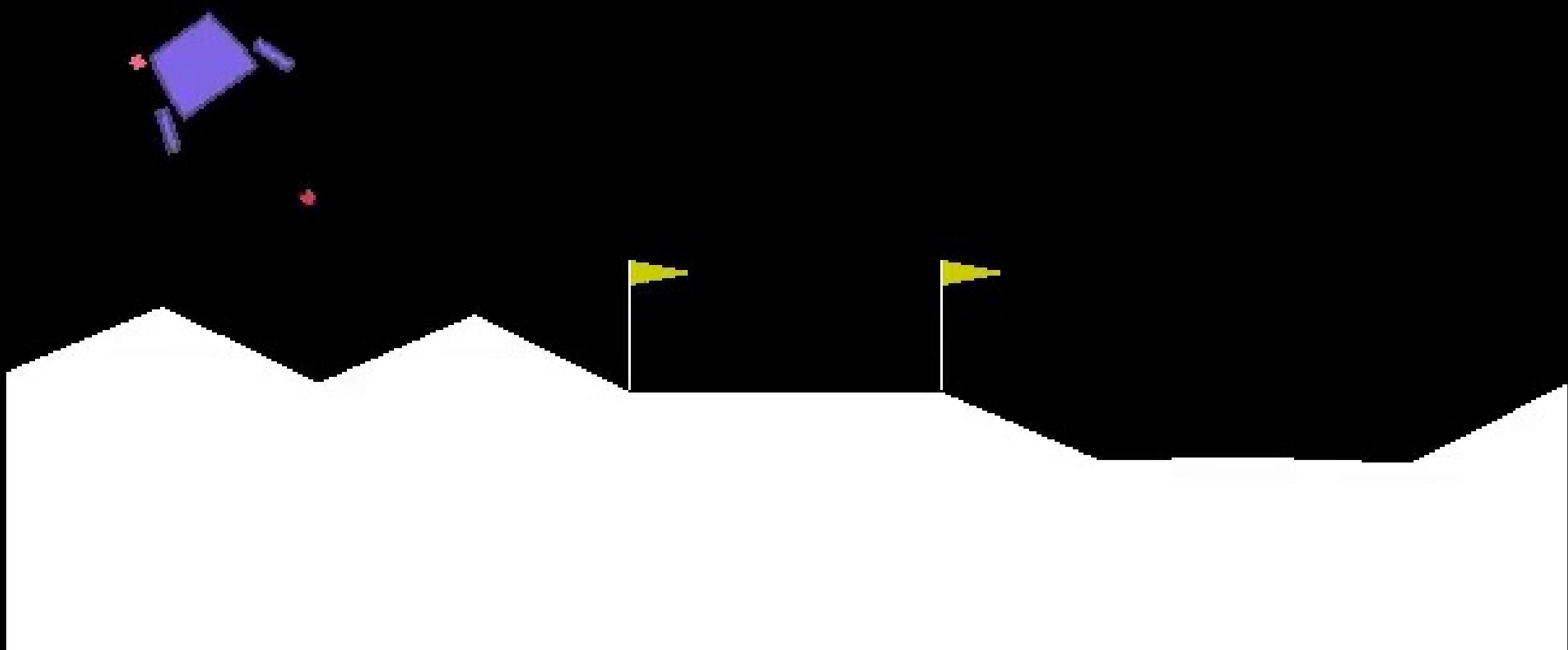
LA SOLUTION QUE J'AI UTILISER EST GOOGLE COLAB
LIEN : cutt.ly/9bXvvwv



L'ATERRISSEUR LUNAIRE

- RÉSOLUTION

VIDEO
COMPIRATION



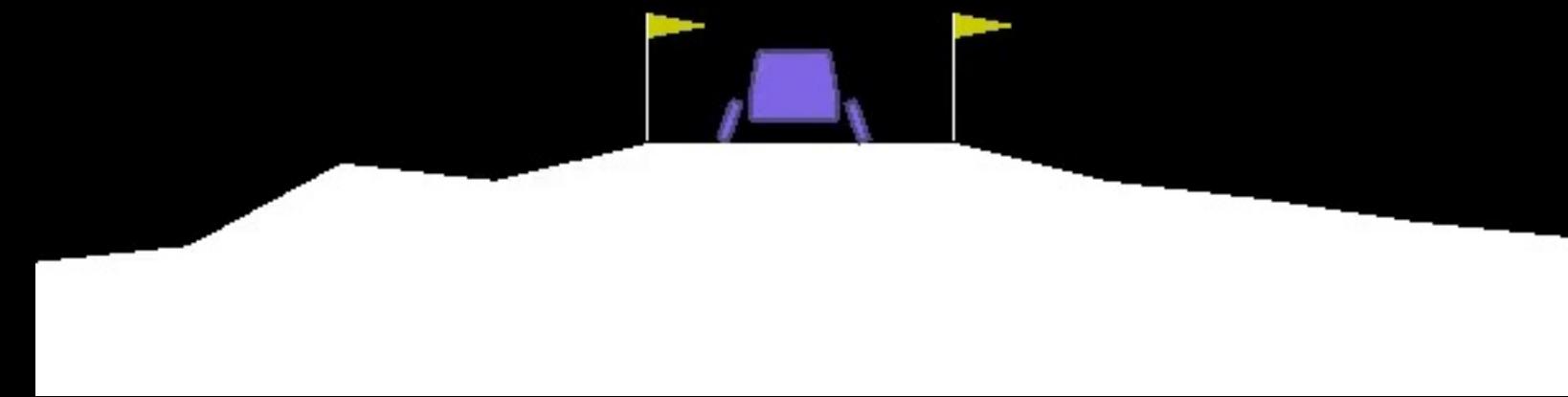
L'ATERRISSEUR LUNAIRE



■ RÉSOLUTION

8 ENTRÉE:

- COORDONNÉ HORIZONTAL
- COORDONNÉ VERTICAL
- VITESSE HORIZONTAL
- VITESSE VERTICAL
- ANGLE
- VITESSE ANGULAIRE
- PREMIER PIED CONTACT
- DEUXIÈME PIED CONTACT



4 SORTIE / ACTION

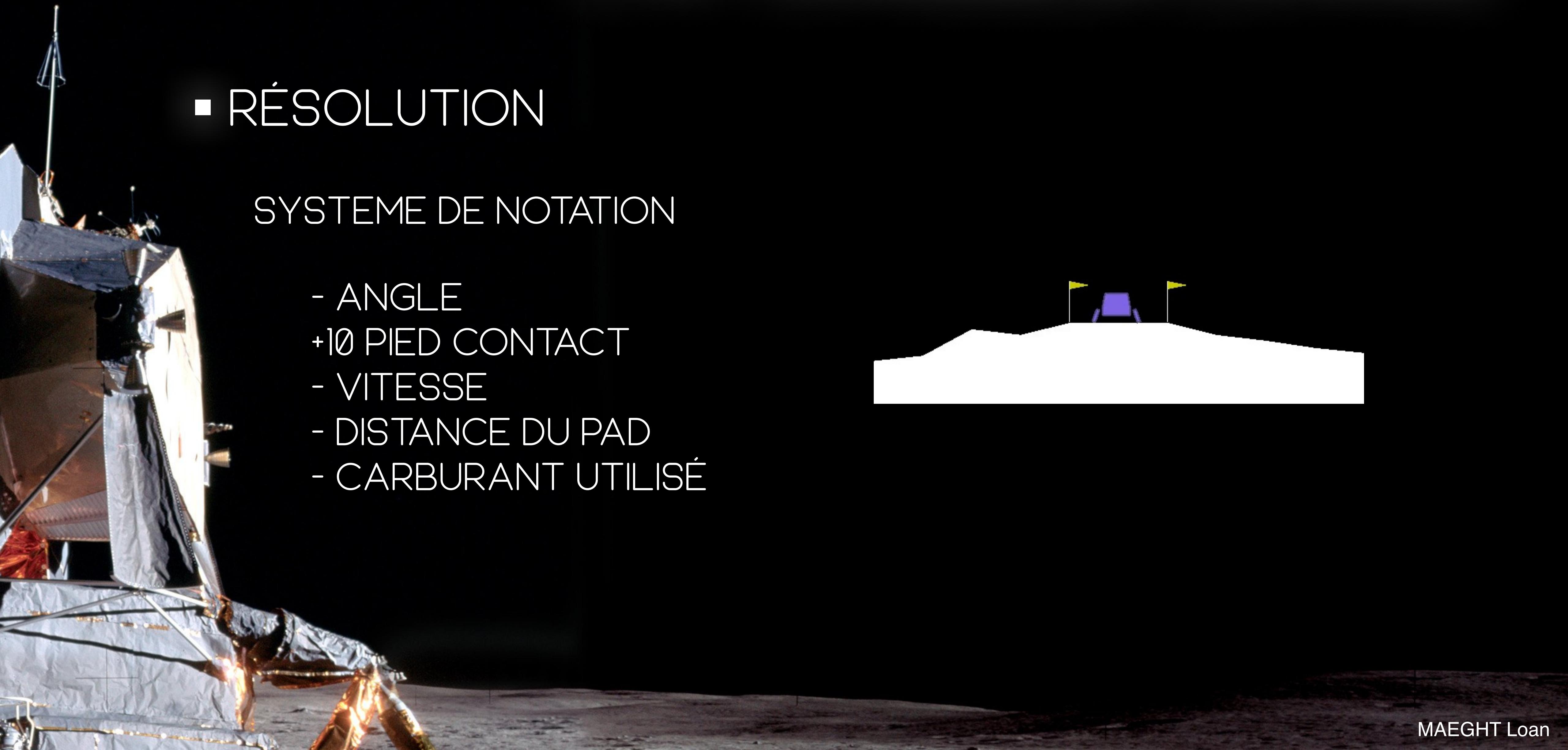
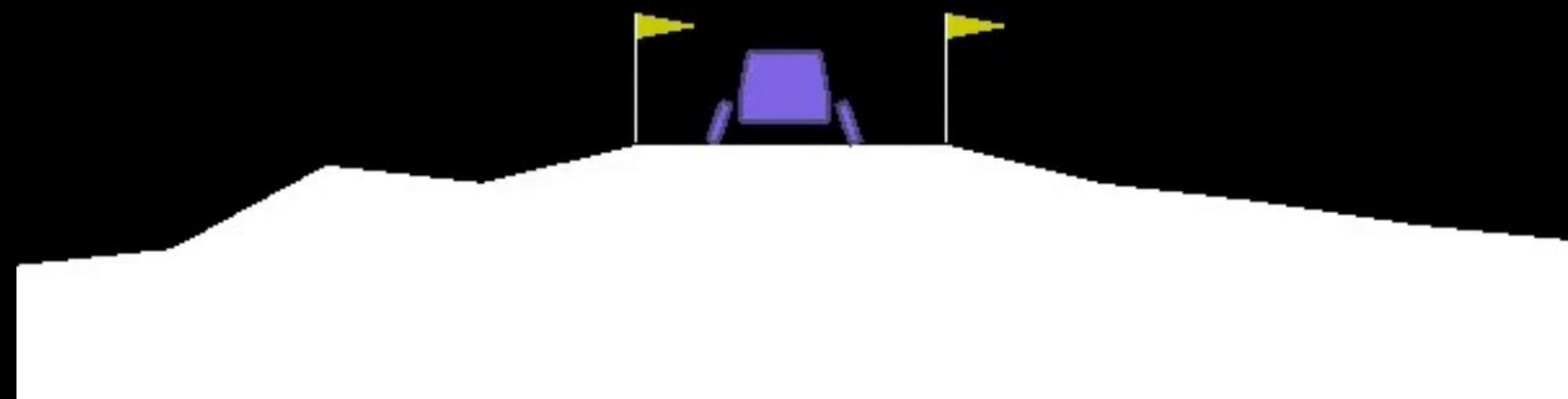
- NE RIEN FAIRE
- MOTEUR PRINCIPAL
- MOTEUR SECONDAIRE GAUCHE
- MOTEUR SECONDAIRE DROIT

L'ATERRISSEUR LUNAIRE

■ RÉSOLUTION

SYSTEME DE NOTATION

- ANGLE
- +10 PIED CONTACT
- VITESSE
- DISTANCE DU PAD
- CARBURANT UTILISÉ

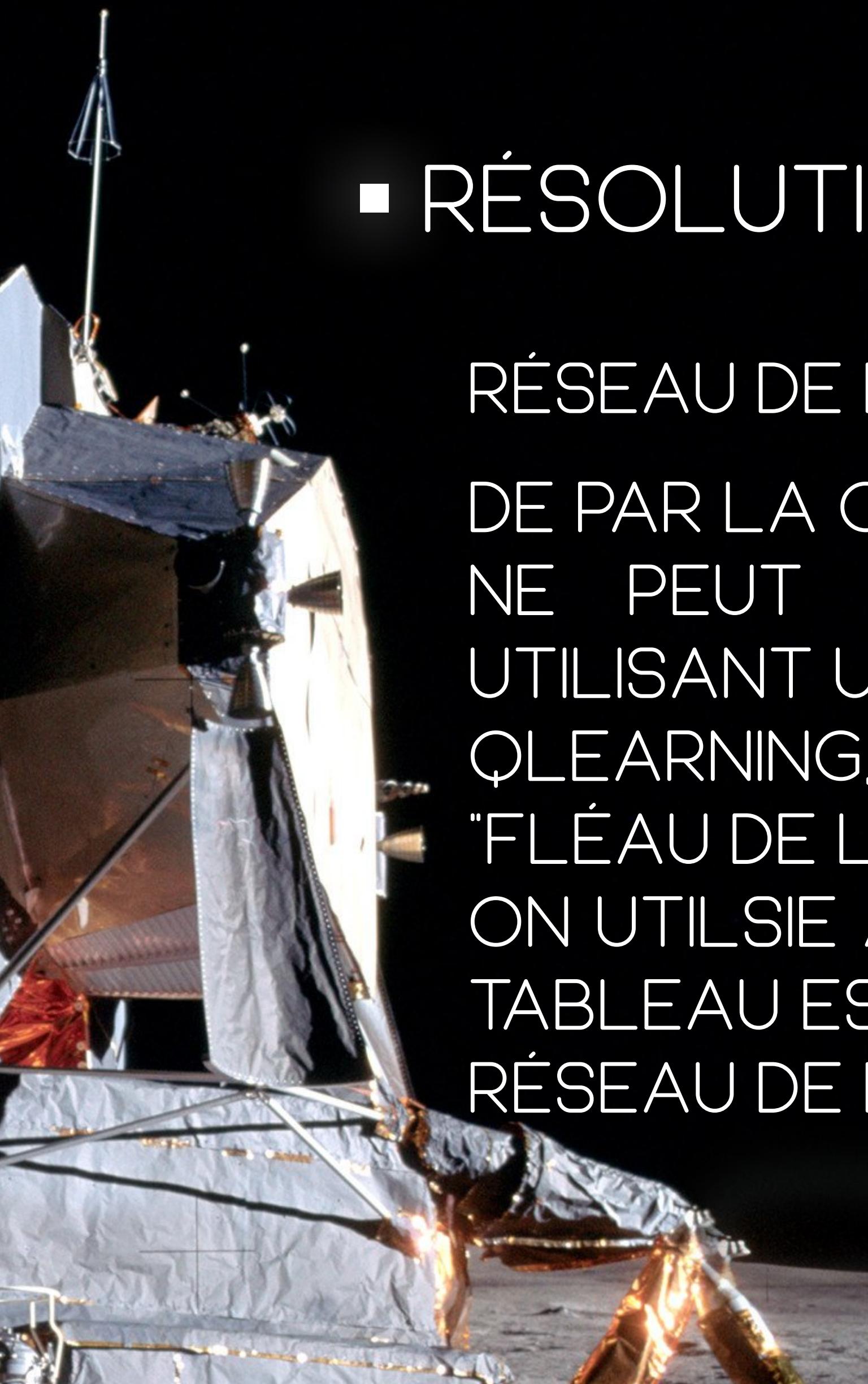
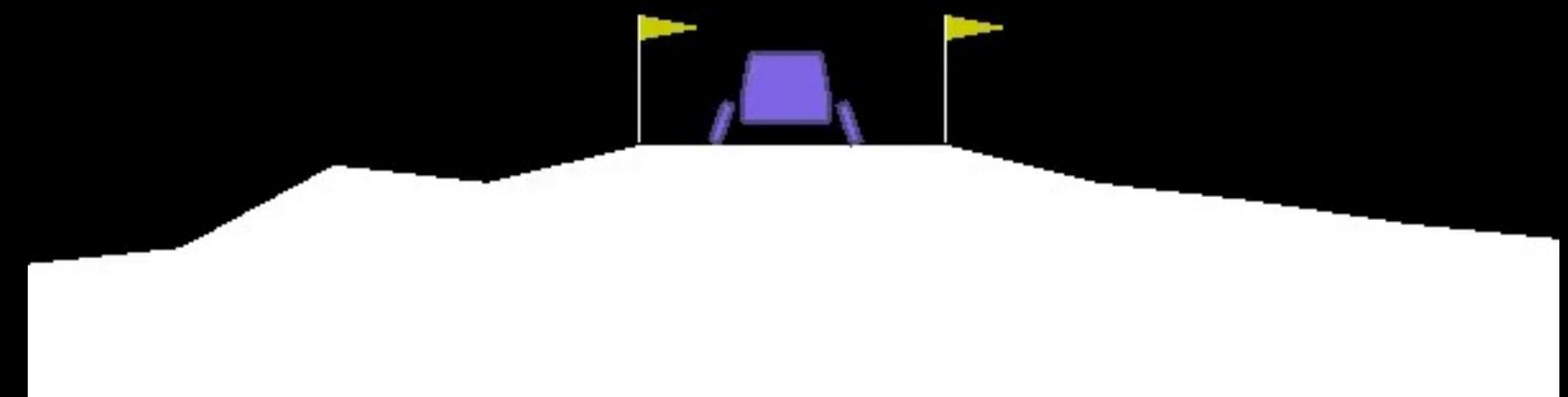


L'ATERRISSEUR LUNAIRE

■ RÉSOLUTION

RÉSEAU DE NEURONE

DE PAR LA COMPLEXITÉ DES ENTRÉS ON
NE PEUT PAS UTILISER UN MODEL
UTILISANT UN TABLEAU AUSSI APPELÉ
QLEARNING. CE PHENOMÈNE EST NOMÉ
"FLEAU DE LA DIMENSION"
ON UTILISE ALORS DEEP QLEARNING. LE
TABLEAU EST ALORS REMPLACÉ PAR UN
RÉSEAU DE NEURONE



L'ATERRISSEUR LUNAIRE

- RÉSOLUTION

TOUT LES FICHIERS SONT DISPONIBLES SUR LE GITHUB:

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



ATTEERRIR SUR LA
LUNE ENDORMANT

MAEGHT LOAN

