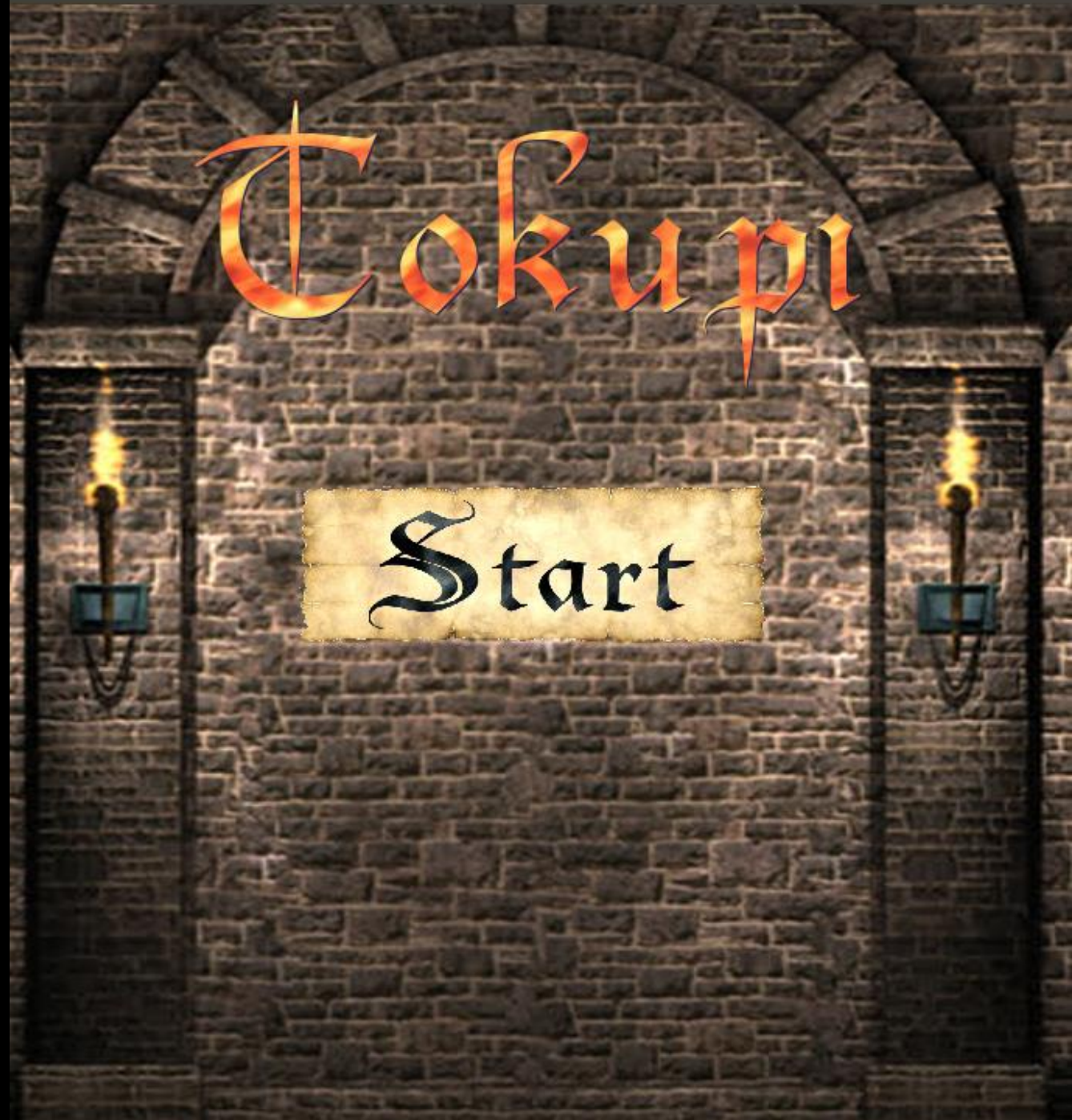


Projet de Majeure

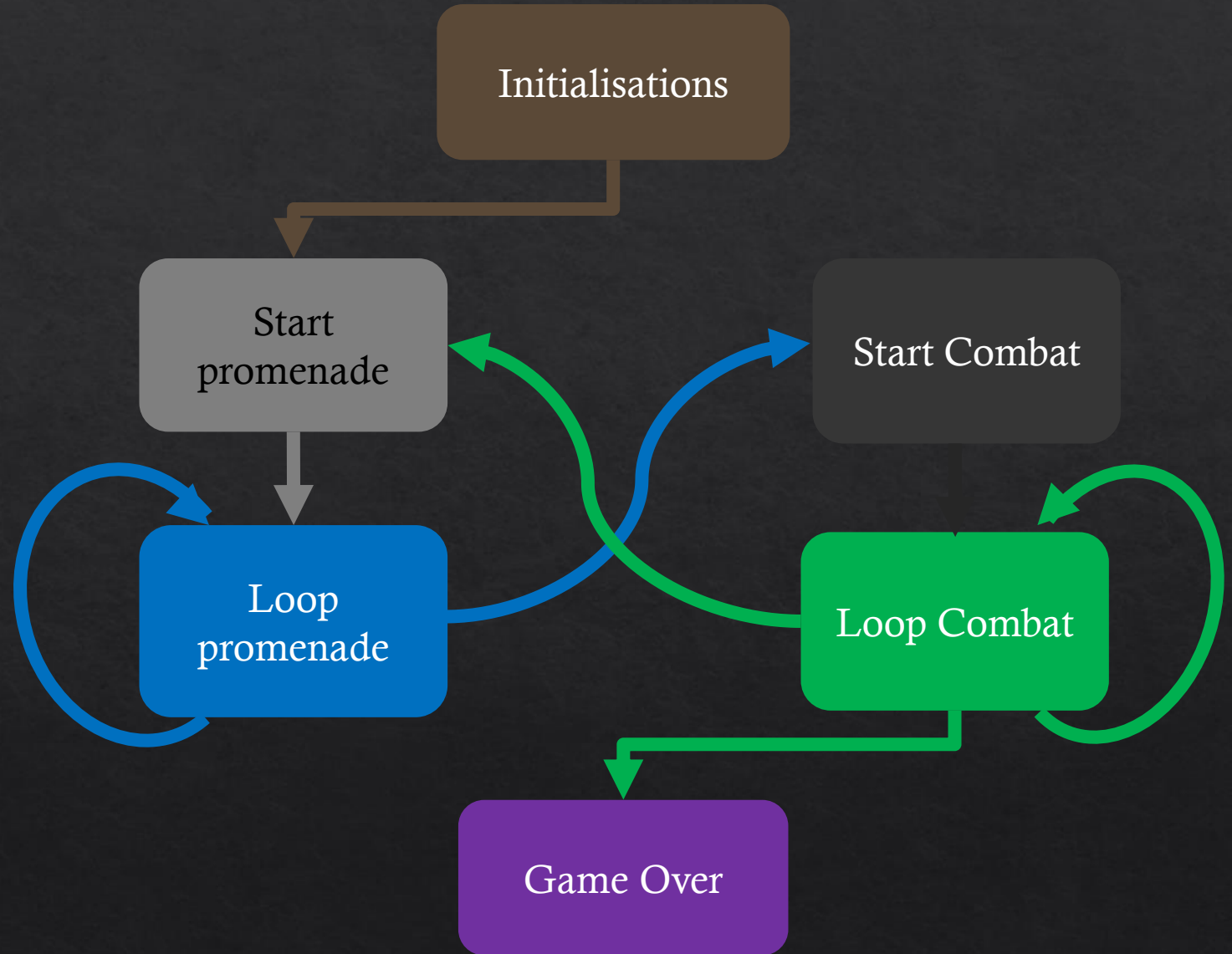
Dylan **TOSTI**

Cédric **KUASSIVI**

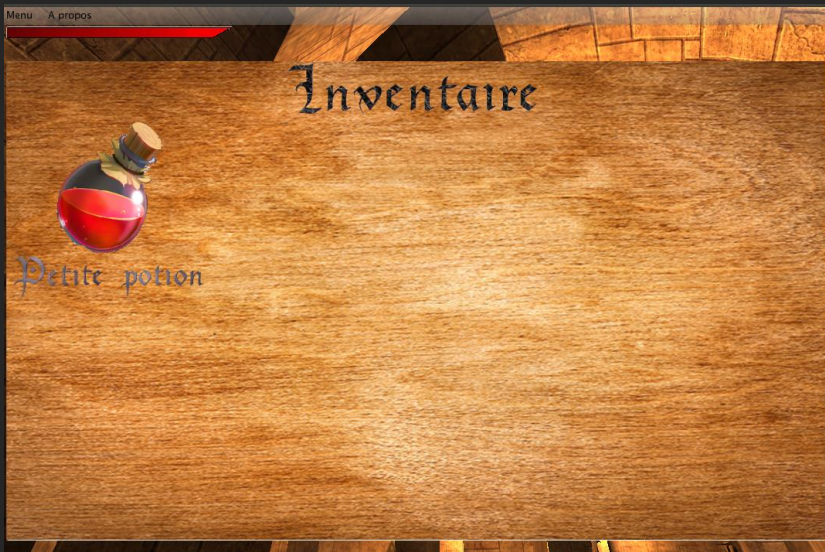
Pedro **FOLETTTO PIMENTA**

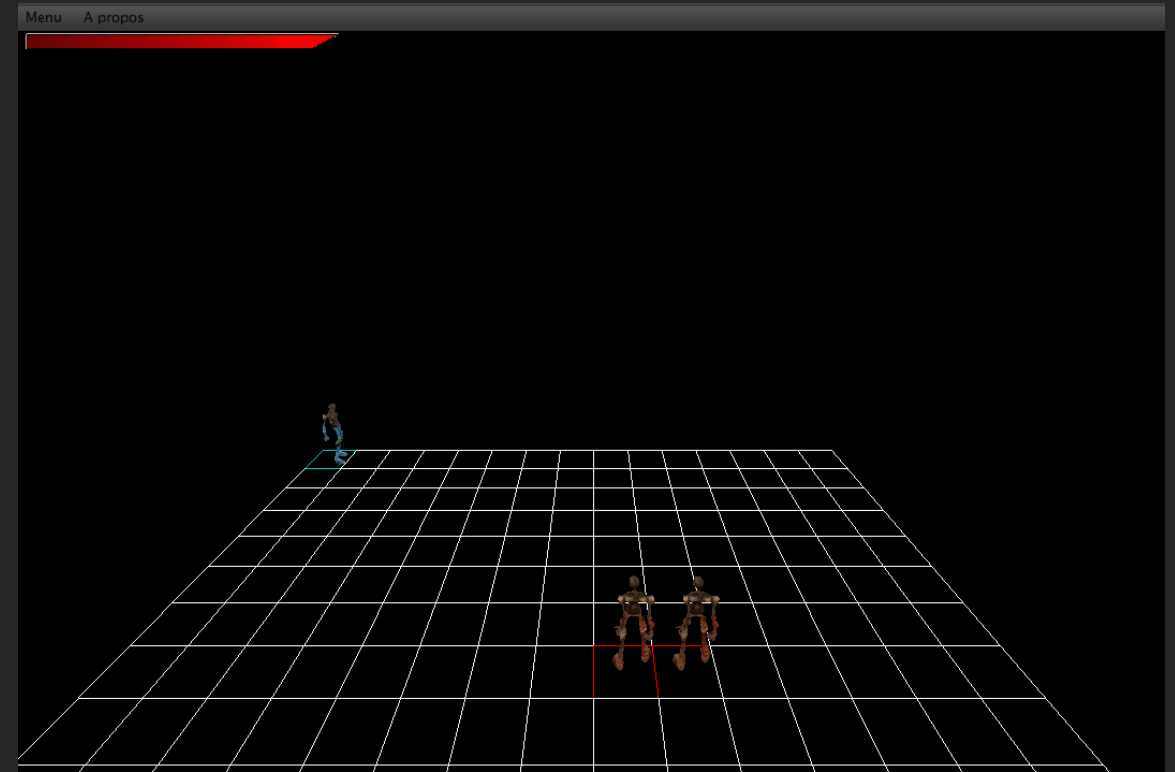
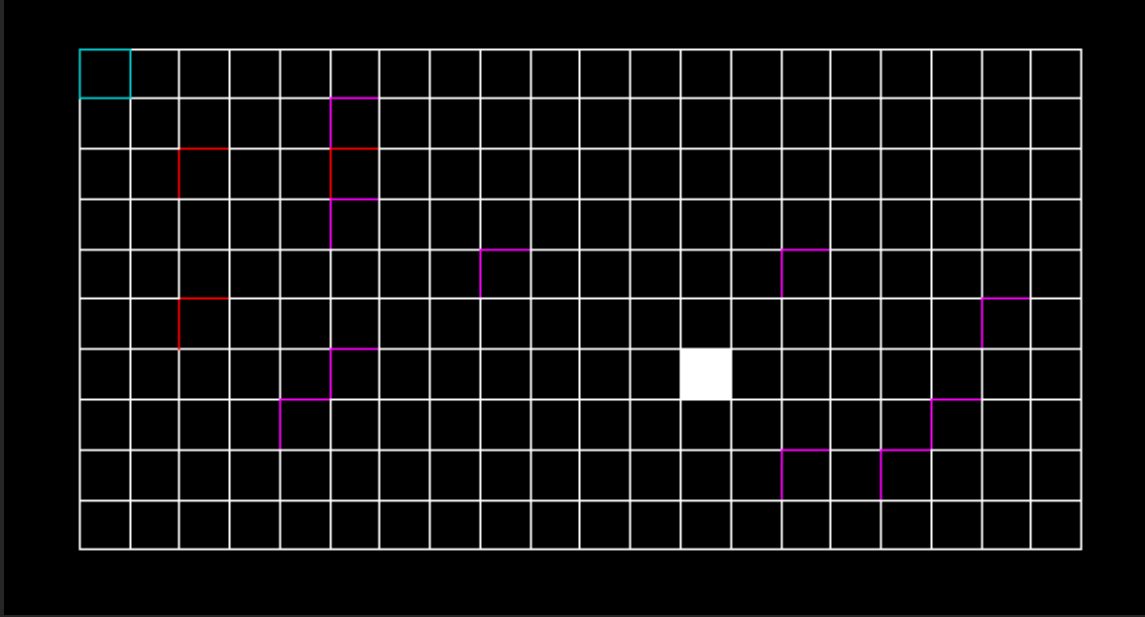


Architecture du code



Mode Jeu Libre





Mode Combat

Q-Learning avec une Q-table

$Q =$

| | 0 | 1 | 2 | 3 | 4 | 5 |
|---|----|----|----|----|----|-----|
| 0 | 0 | 0 | 0 | 0 | 80 | 0 |
| 1 | 0 | 0 | 0 | 64 | 0 | 100 |
| 2 | 0 | 0 | 0 | 64 | 0 | 0 |
| 3 | 0 | 80 | 51 | 0 | 80 | 0 |
| 4 | 64 | 0 | 0 | 64 | 0 | 100 |
| 5 | 0 | 80 | 0 | 0 | 80 | 100 |



Q learning avec une Q-table

| State | Action | | | | | |
|-------|--------|----|----|----|----|-----|
| | 0 | 1 | 2 | 3 | 4 | 5 |
| 0 | -1 | -1 | -1 | -1 | 0 | -1 |
| 1 | -1 | -1 | -1 | 0 | -1 | 100 |
| 2 | -1 | -1 | -1 | 0 | -1 | -1 |
| 3 | -1 | 0 | 0 | -1 | 0 | -1 |
| 4 | 0 | -1 | -1 | 0 | -1 | 100 |
| 5 | -1 | 0 | -1 | -1 | 0 | 100 |

Objectif: apprendre quelle est la meilleure action à prendre dans chaque état (situation) possible.

Comment: un tableau où on calcule le maximum des récompenses futurs attendus, pour chaque état, pour chaque action.

États

Chaque état est un nombre entier calculé à partir de:

- Distance pour l'ennemi en X
- Distance pour l'ennemi en Y
- HP de l'ennemi
- HP de soi même

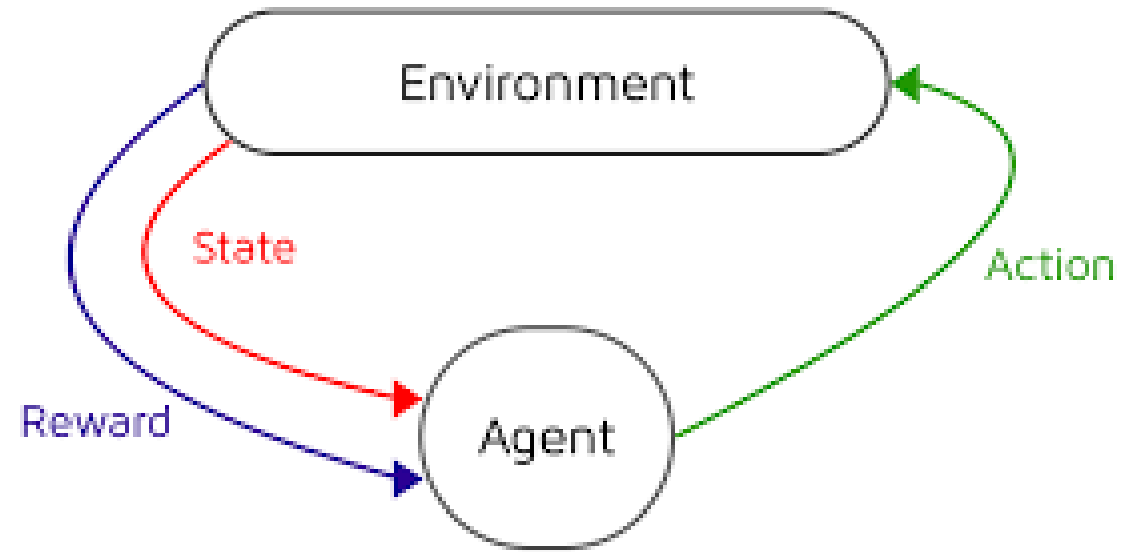
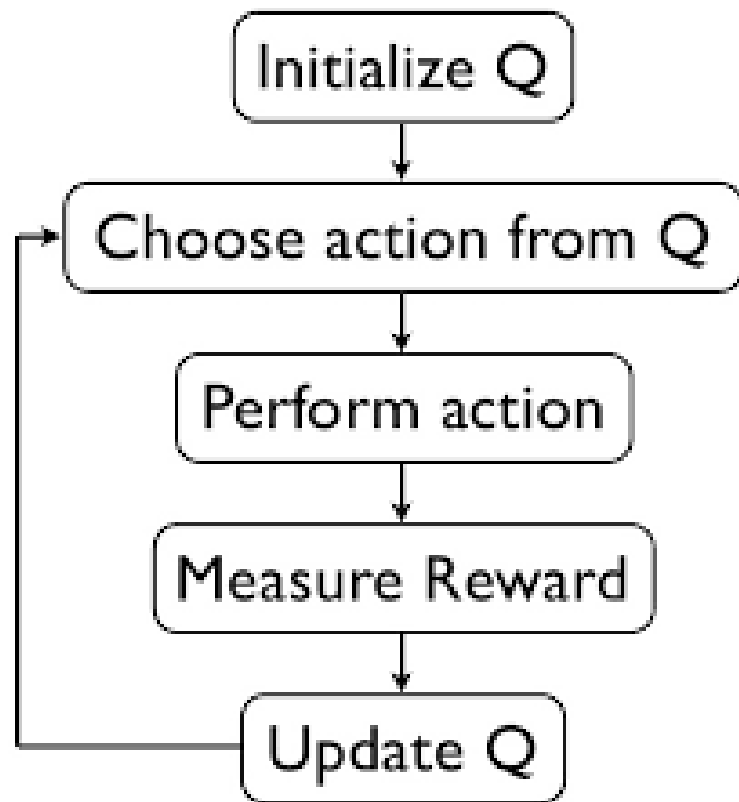
Nombre d'états au total: 100

Actions

- Se déplacer 1 case en haut
- Se déplacer 1 case en bas
- Se déplacer 1 case à gauche
- Se déplacer 1 case à droite
- Attaquer

Nombre d'actions au total: 5

Boucle d'entraînement



Optimisation du tableau


$$Q^{\pi}(s_t, a_t) = \underline{E}[R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \dots | s_t, a_t]$$

Q-Values for the state
given a particular state

Expected discounted
cumulative reward

Given the state and action

Optimisation du tableau

$$Q^{\pi}(s_t, a_t) = \underline{E}[R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \dots | s_t, a_t]$$


$$Q^{new}(s_t, a_t) \leftarrow (1 - \alpha) \cdot \underbrace{Q(s_t, a_t)}_{\text{old value}} + \underbrace{\alpha}_{\text{learning rate}} \cdot \overbrace{\left(\underbrace{r_t}_{\text{reward}} + \underbrace{\gamma}_{\text{discount factor}} \cdot \underbrace{\max_a Q(s_{t+1}, a)}_{\text{estimate of optimal future value}} \right)}^{\text{learned value}}$$

Merci pour votre attention !

Avez-vous des question ?