

# Contrôle d'un agent par apprentissage par renforcement profond

Présentation de mon stage à THALES dans l'équipe AS&BSIM

---

Nizam Makdoud

29 août 2017

THALES

Objectifs et Défis technologiques

Apprentissage Profond

Architecture classique en apprentissage profond

Module de curiosité appliqué au contrôle

# Objectifs et Défis technologiques

---

## **Sujet du stage**

Contrôler un agent dans un environnement simulé à partir de la vision de celui-ci.

Contrôler un agent dans un environnement simulé à partir de la vision de celui-ci.

## Contrainte sur notre contrôle

- Etre pertinent pour toutes entrées (images, graphes, ...) .
- Etre approprié pour tous buts.
- Gérer des environnements complexes et inconnues.

## Contraintes sur notre contrôle

- Être pertinent pour toutes entrées (images, graphes, ...) .
- Être approprié pour tous buts.
- Gérer des environnements complexes et inconnues.

## Technologiee utilisées pour répondre à ce défi

- **L'apprentissage profond** pour la résilience fâce aux entrées.
- **L'apprentissage par renforcement** pour s'adapter aux buts.
- **Module de curiosité** pour explorer.

## **Conclusion sur notre contribution**

Proposition d'un contrôle basé sur de l'apprentissage par renforcement profond utilisant un module de curiosité.

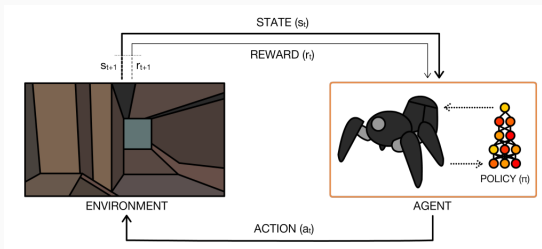
# Apprentissage Profond

---



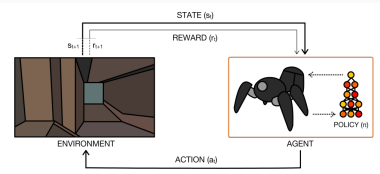
Formalisme **La théorie des processus décisionnels de Markov** :

$$\mathcal{M} = \left\{ \mathcal{X}, \mathcal{A}, \mathcal{R}, \mathcal{P}, \gamma \right\}$$



Formalisme **La théorie des processus décisionnels de Markov** :

$$\mathcal{M} = \left\{ \mathcal{X}, \mathcal{A}, \mathcal{R}, \mathcal{P}, \gamma \right\}$$



- **Politique**

$$\pi, a_t \sim \pi(\cdot, x_t)$$

- **Dynamique**

$$x_{t+1} \sim P(\cdot, x_t, a_t)$$

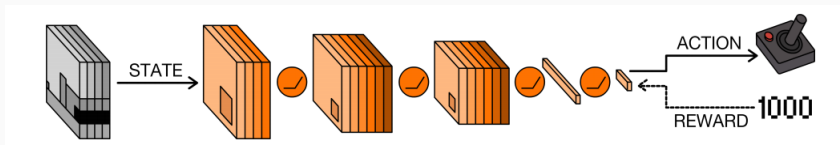
- **Fonction Etat-Action**

$$Q^\pi(x, a) = \mathbb{E}_{P, \pi} \left[ \sum_t \gamma^t r_t(x_t, a_t) \mid x_0, a_0 = x, a \right]$$

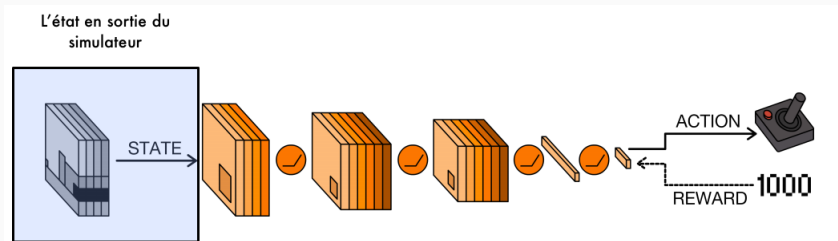
# **Architecture classique en apprentissage profond**

---

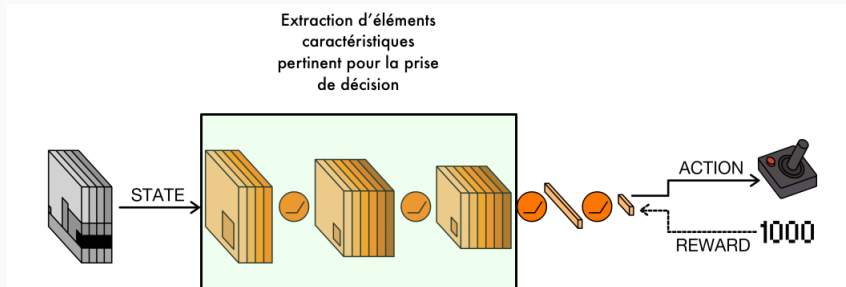
## Architecture classique en apprentissage profond



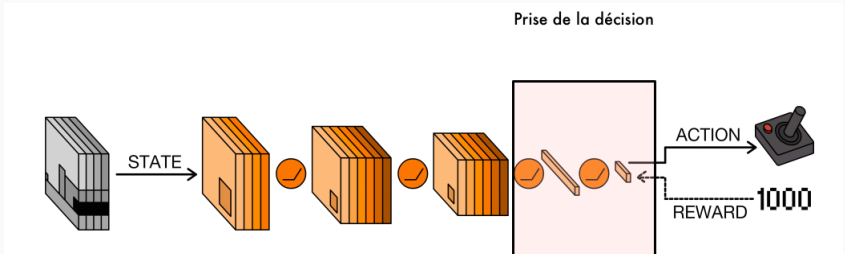
## Gestion des entrées



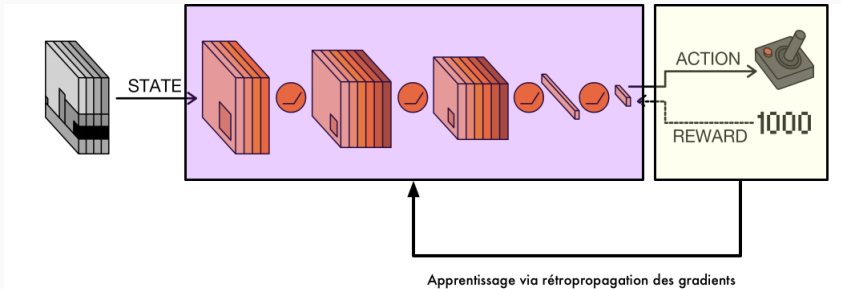
## Acquisitions automatiques d'éléments pertinents à la prise de décisions



## Prise de la décision par l'agent

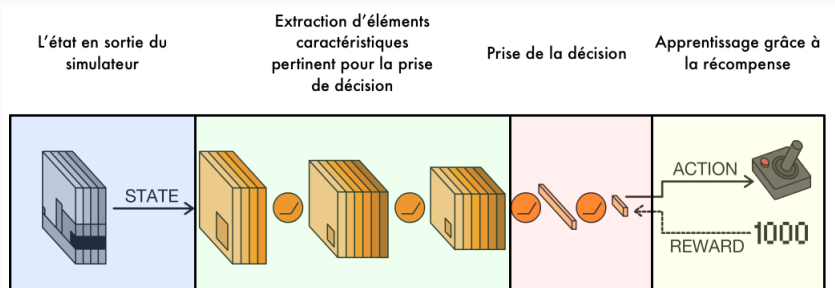


## Apprentissage du réseau





## Résumé de la base utilisée pour le contrôle



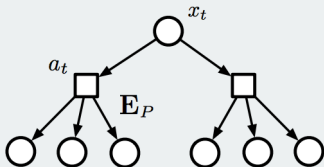
# **Module de curiosité appliqué au contrôle**

---

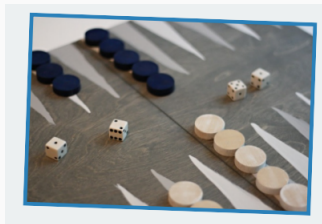
# Nécessité d'exploration d'un environnement

Nécessité d'exploration de l'environnement ?

## Model-based



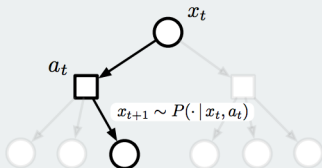
- **Dynamique de l'environnement**
  - Connue
  - Fonction d'état-action : Connue
  - Politique : Connue



# Nécessité d'exploration d'un environnement

Nécessité d'exploration de l'environnement ?

## Model-free



- **Dynamique de l'environnement**

→ Inconnue

→ Fonction d'état-action : Inconnue

→ Politique : Inconnue

