THALES
Building a future we can all trust

État de l'art de l'apprentissage par renforcement de politiques déclaratives

Ivelina Stoyanova - Romain Caillière, Nicolas Museux

Strasbourg – 07/07/2023

Les produits Thales sont la plupart du temps des systèmes critiques dont certains répondent à des problèmes de décisions séquentielles

Systèmes critiques avec problèmes de décision en réaction à un environnement

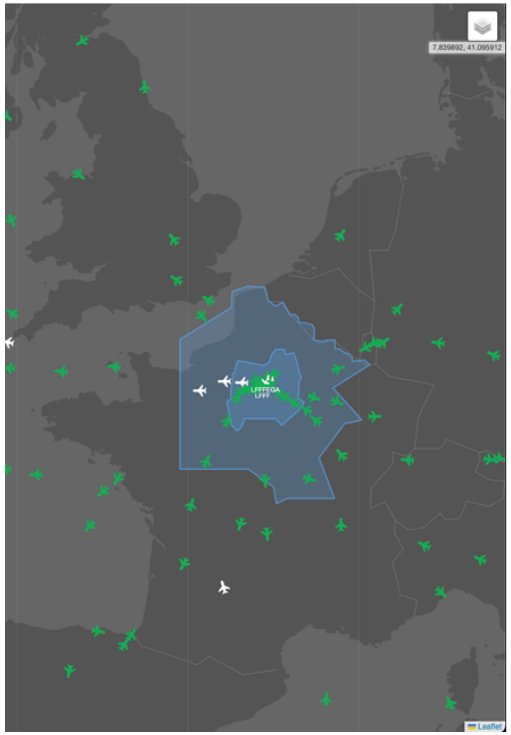
- Contrôle pour les systèmes autonomes,
- Reconnaissance d'activité avec mitigation des actions pour la protection des systèmes,
- Sélection d'actions tactiques pour la défense d'un théâtre d'opérations,
- Et plein d'autres.

Les techniques les plus performantes actuellement reposent sur des modèles « boite noire » non validables/certifiables

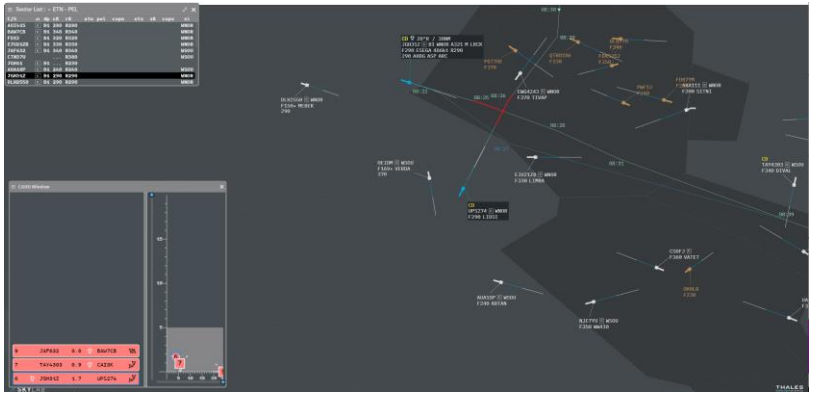
A l'heure actuelle, les algorithmes de Deep Reinforcement Learning sont les techniques SOTA pour les problèmes de RL

- Explication compliquée de la décision du modèle a posteriori : **Explicabilité**
  - Il est compliqué de comprendre le lien entre les entrées et les sorties dans les systèmes « boite noire »
- Impossibilité de lire et de comprendre le modèle pour un humain: **Interprétabilité**
  - les systèmes « boite noire » ne sont pas composés de symboles auxquels une sémantique est associée
- Difficulté importante à **valider/certifier** de tels modèles

La gestion du trafic aérien comme exemple concret du besoin en IA validables/certifiables



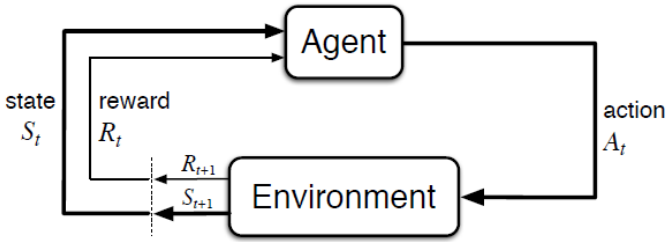
Conflict Detection & Resolution @ Thales



Politiques de décisions déclaratives et normalisées au niveau international (ICAO Doc 4444)

L'apprentissage par renforcement consiste à apprendre une politique/un comportement optimal au sein d'un environnement

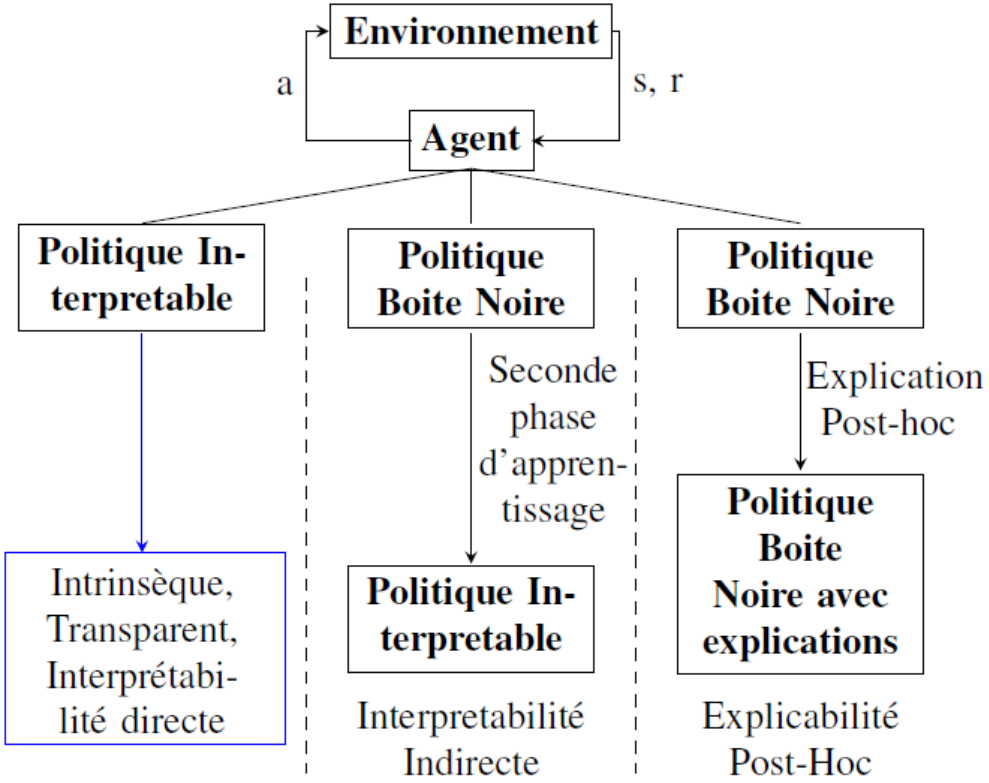
Un apprentissage itératif



- Au fur et à mesure de l'apprentissage, l'agent apprend la meilleure séquence d'associations état-action

- Le comportement appris est optimal au regard :
  - De la **modélisation** des **états**
  - Des **actions** possibles
  - De la **fonction** de **récompense**

Les politiques apprises peuvent être classées selon trois catégories

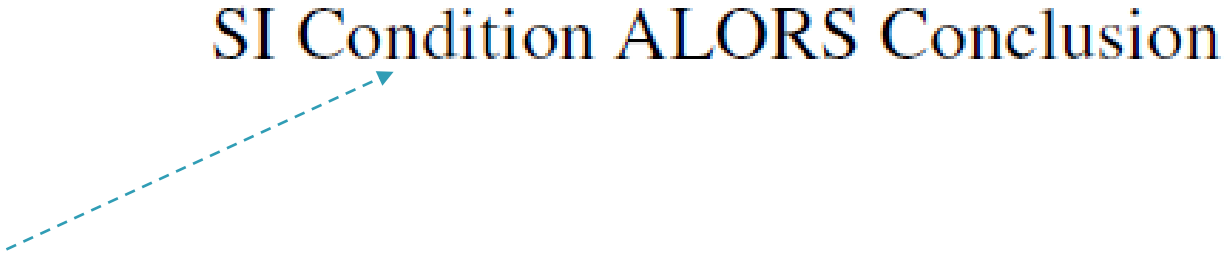


[Puiutta 2020]
[Alharin 2020]
[Heuillet 2021]
[Glanois 2021]

Les systèmes à base de règles sont un type de modèle déclaratif reposant sur des symboles auxquels sont associés une sémantique

SI Condition ALORS Conclusion

Les systèmes à base de règles sont un type de modèle déclaratif reposant sur des symboles auxquels sont associés une sémantique



Prémisse : formule logique de termes vrais ou faux
conjonction / disjonction / négation

Les systèmes à base de règles sont un type de modèle déclaratif reposant sur des symboles auxquels sont associés une sémantique

SI Condition ALORS Conclusion

Logique
- **Propositionelle**: constante, OR, AND, NOT
- **Predicat**: + quantifieurs, fonctions, variables
- **Temporelle**: + relations temporelles

Domaine
- **Symbolique**, **discret**: Logique Booléenne
- **Continue**, **mixte**: Logique floue
- **Evènementiel**: Logique temporelle

Règles
- **Decisionelle**: *SI condition vraie alors faire action*
- **Deductive**: *Si formule vraie, alors déduction vraie*
- **Descriptive**: *Si propriétés vraies, alors classe est C*

THALES
Building a future we can all trust

Apprentissage de règles pour des domaines discrets

Logiques à valeurs de vérité Booléennes

Apprentissage de politique à base de règles basées sur la logique Booléenne

RL tabulaire : Q-Learning, SARSA, dérivés

- Table = base de règles : SI état = s_i ALORS action = argmax_a(Q(s_i, a))

CLARION : architecture cognitive [Sun, 2001]

- RL tabulaire = bas niveau | Règles = haut niveau (généralisation/spécialisation)

RL relationnel : description explicite des états et des actions basée sur des prédicats relationnels. [Džeroski, 2001]

Neural Logic RL : Apprentissage de règles à partir de Deep RL [Jiang, 2019, Ma 2021, Payani 2020]

=> Logique propositionnelle et/ou nécessité de l'expertise humaine en amont

L'apprentissage de politiques déclaratives via l'utilisation des machines de Tsetlin

Repose sur les automates de Tsetlin (un automate/terme) [Granmo, 2018]

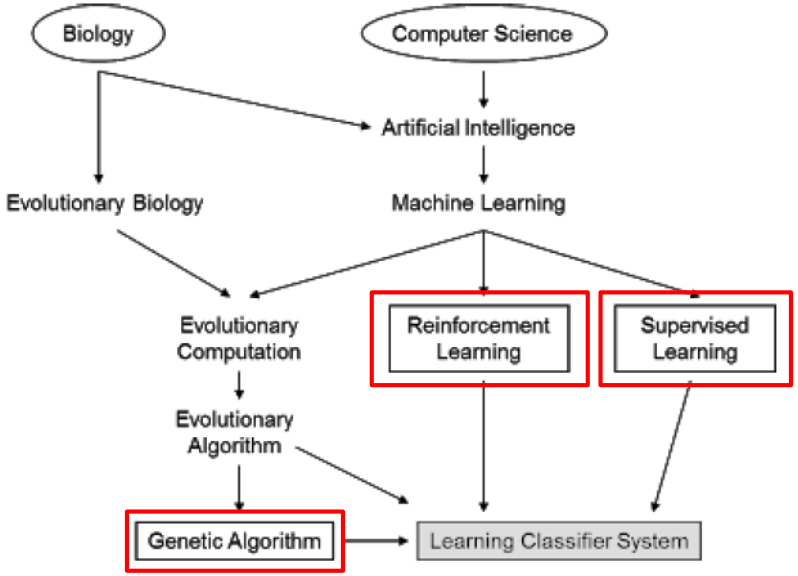
Se déroule en trois étapes répétées durant l'apprentissage

- 1 L'extraction de motifs frequents → Spécialisation
- 2 La discrimination des littéraux → Généralisation
- 3 La choix de l'action (par un vote des règles)

Classification → Régression → Apprentissage par renforcement [Gorji 2021, 2023]

→ Ne s'attaque pas à des environnements complexes pour le moment, difficulté du passage à l'échelle (notamment pour l'interprétabilité)

Introduction des Learning Classifier Systems

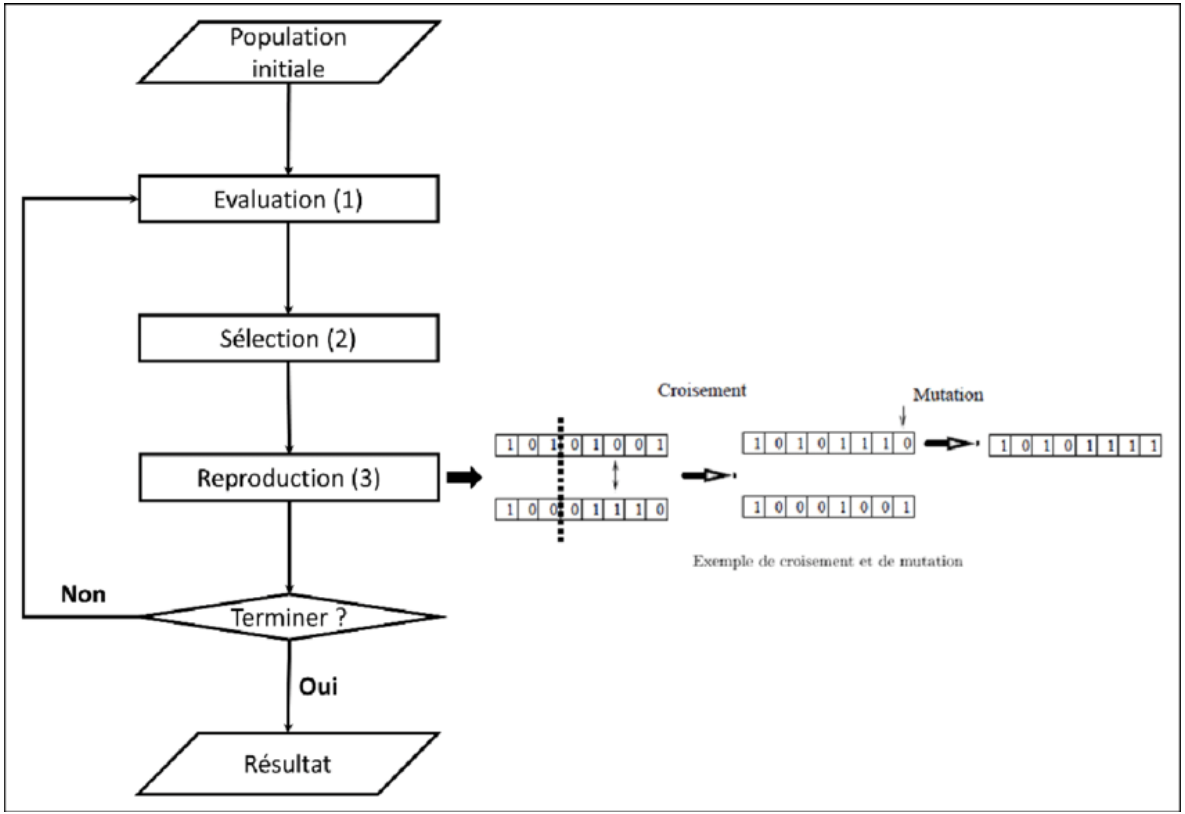


Proposé par John H. Holland en 1986

... a learning paradigm in which an agent learns to perform a task by evolving a population of condition-action rules (i.e. the classifiers) through temporal difference learning and genetic algorithms.

P.L. Lanzi 2005

Les algorithmes génétiques sont une méthode d'optimisation pour obtenir une solution approchée en un temps raisonnable

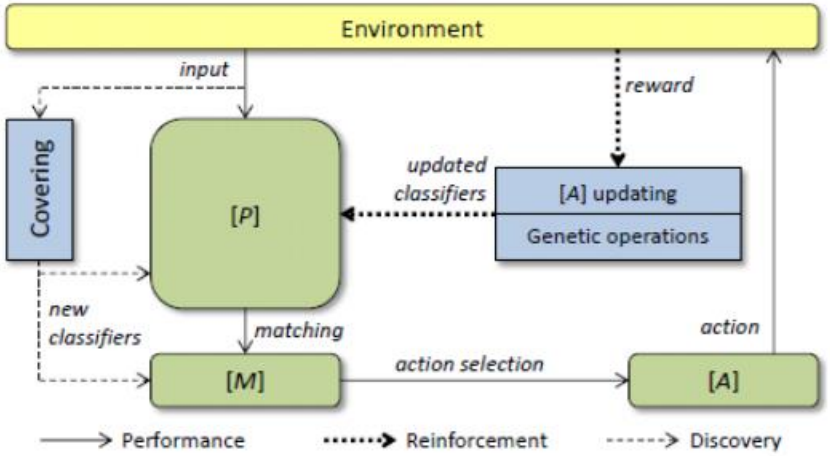


Walid Bedhiafi 2017

{OPEN}

Learning Classifier System: une combinaison Apprentissage par Renforcement/Algorithme Génétique/ Système à Base de Règles

- **Les individus sont :**
  - Des règles (Michigan style)
  - Des jeux de règles (Pittsburgh style)
- **Le jeu des algorithmes génétiques fait évoluer les règles**
- **eXtended Classifier Systems** [Butz 2000, Wilson 1995]
  - SI état ALORS action prédit **p**
- **Gestion des dimensions continues** [Stone 2003]
- **XCS-RC** [Fredivianus 2015] : **remplacement de l'AG par le principe de l'induction**



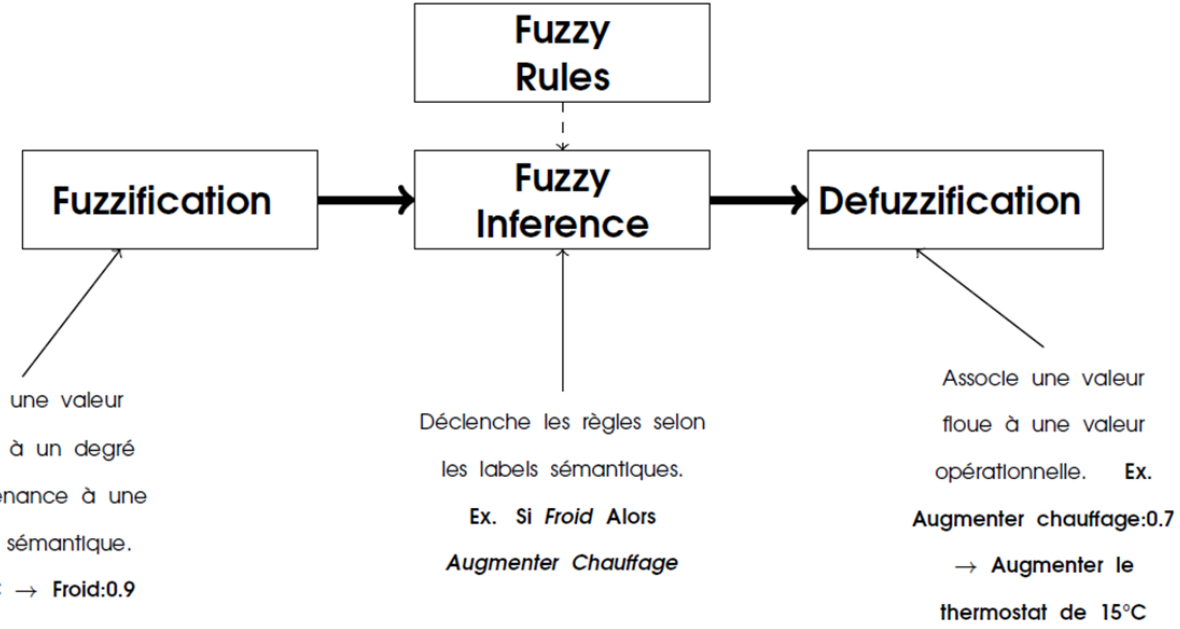
THALES
Building a future we can all trust

Apprentissage de règles pour des domaines continus

Logiques à valeurs de vérité floues

La logique floue étend la logique booléenne classique avec des valeurs de vérités partielles

$$\chi_A(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{si } x \notin A \end{cases} \quad \Rightarrow \quad \mu_{\underset{\sim}{A}}(x) \in [0,1]$$



Fuzzy Learning Classifier Systems : les LCS reposant sur le logique floue

Trois différences principales avec les LCS [Bonarini 1999, 2007, Casillas 2007]

- Le déclenchement de la règle, grâce à l'utilisation de la fonction d'appartenance,
  - Toutes les règles se déclenchent avec un certain degré

- La manière dont les actions sont choisies
  - La meilleure prédiction moyenne

- Au niveau de l'inférence floue et de la défuzzification
  - L'action à appliquer est calculée selon les principes de la défuzzification

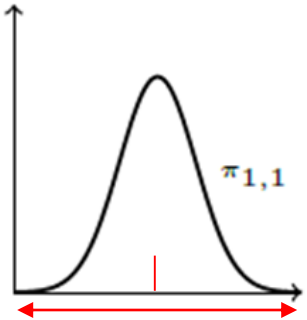
Apprentissage de politique à base de règles basées sur la logique floue

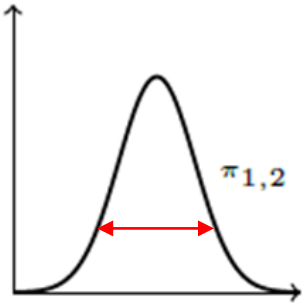
Modification de la conclusion des règles

- FQL : FIS pour permettre au Q-Learning d'appliquer des actions continues [Glorennec 1997]
- Actor-Critic : l'acteur est modélisé par un FIS [Lee 1989]

Modification des prémisses des règles

Modification des centres (moyenne)





Modification des largeurs (écart-type)

[Berenji 1992, Desouky 2011, Joo Er 2004, Lin 1994, 1996, Wang 2007]

Capacité de suppression des règles inutiles [Hosoya 2012]

Le problème de l'interprétabilité des systèmes à base de règles

Règles ≠ Interprétable par définition

Lipton [C. Lipton, 2017] : Interprétabilité =

- Simulable - capacité pour un humain de saisir immédiatement le sens du modèle
- Décomposable - impliquant que les entrées, les paramètres et le calcul soient facilement appréhensibles
- Algorithmiquement transparent - fournissant une garantie sur le résultat du modèle sur des données inconnues

[Miller 1956]: La règle des 7 plus ou moins 2.

- Un nombre maitrisé de termes dans la prémisse
- Un nombre maitrisé de règles dans la politique

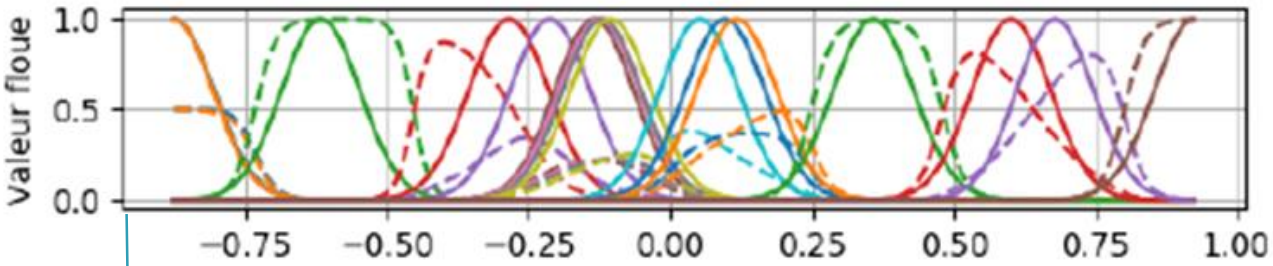
L'enchainement des règles [Glanois 2022, Trillo 2020]

REF TRT-Fr/STI/LRASC/RCA/23,0064 – 07-07-2023
Thales Research & Technology France
Modèle trtp version 8,0,5 / Modèle : 87211168-DOC-GRP-FR-006

{OPEN}

Illustration via l'interprétabilité des règles apprises avec le Dynamic Fuzzy Q-Learning

Association symbole-sémantique : un symbole par fonction d'appartenance



Distinguabilité : capacité à repérer le terme discriminant

Perspectives et travaux futurs

Réflexion sur les métriques d'interprétabilité

- Définir formellement les critères d'interprétabilité
- Intégrer ces critères à l'apprentissage par renforcement
- Apprendre un jeu de règles interprétables par construction

Validation des modèles à base de règles floues

- Utilisation des méthodes formelles pour valider le modèle à base de règles appris

Règles évènementielles

- Mise au point d'un environnement spécifique
- Réflexion sur un algorithme capable d'apprendre dans ce contexte