



République Tunisienne
Ministère de l'Enseignement Supérieur et
de la Recherche Scientifique
Université de Kairouan
Institut Supérieur d'Informatique et de Gestion
de Kairouan



Devoir de Synthèse

Dans le cadre du module

Systèmes Multi-Agents

**Parcours : Master Recherche en Système d'Information
Intelligent**

Projet 1 : Explicabilité dans les Systèmes Multi-Agents Hybrides

Elaboré par :

Maaoui Asma

Année Universitaire 2025-2026

Table des matières

1	Introduction	3
2	Contexte Théorique	4
2.1	Systèmes Multi-Agents	4
2.2	Architectures Multi-Agents Hybrides	4
2.3	Intelligence Artificielle Explicable	5
3	État de l'Art	6
3.1	Explicabilité dans les Systèmes d'Intelligence Artificielle	6
3.2	Explicabilité dans les Systèmes Multi-Agents	6
3.3	Méthodes d'explication des décisions collectives	7
3.4	Agents Neuro-Symboliques	7
3.5	Comparaison des approches d'explicabilité	9
3.6	Limites des Approches Existantes	10
4	Formulation du Problème de Recherche	11
4.1	Introduction	11
4.2	Définition des Concepts Clés	11
4.3	Décision Collective	12
4.4	Objectifs de l'Explicabilité	12
4.5	Formalisation de la Contribution des Agents	12
4.6	Problématique Synthétique	12
5	Cadre Conceptuel Proposé	13
5.1	Introduction	13
5.2	Architecture du Cadre Conceptuel	13
5.3	Description des Composants	14
5.4	Fonctionnement du Cadre	14
5.5	Avantages du Cadre Proposé	14
5.6	Perspectives d'Évaluation	15
6	PROTOTYPE	16
6.1	Introduction	16
6.2	Architecture du prototype	16
6.3	Description des Composants	17
6.4	Implémentation des Agents	18
6.4.1	Agent ML Classique	18
6.4.2	Agent Neuro-Symbolique	18
6.4.3	Module d'Agrégation des Explications	19
6.4.4	Trace d'Exécution - Explicabilité Démontrée	19
6.5	Conclusion du Chapitre	20
7	Discussion	21

8 Perspectives de Recherche	22
9 Conclusion	23
10 Références	24

Chapitre 1

Introduction

Au cours des dernières années, les systèmes d'intelligence artificielle ont connu une évolution remarquable et sont désormais intégrés dans de nombreux domaines tels que la santé, les transports intelligents et les systèmes d'aide à la décision. Parmi ces approches, les systèmes multi-agents occupent une place importante, car ils permettent de modéliser des environnements complexes à travers l'interaction d'entités autonomes capables de percevoir, raisonner et agir de manière distribuée.

Parallèlement à ces avancées, la complexité croissante des modèles d'intelligence artificielle, en particulier ceux basés sur l'apprentissage automatique, soulève des préoccupations importantes concernant la transparence et la compréhension des décisions produites par ces systèmes. Dans ce contexte, le domaine de l'intelligence artificielle explicable vise à rendre les décisions des systèmes intelligents plus compréhensibles pour les utilisateurs humains.

Toutefois, le problème de l'explicabilité devient encore plus complexe dans le cadre des systèmes multi-agents hybrides, où différents types d'agents peuvent combiner des approches symboliques et des techniques d'apprentissage automatique. Dans de tels environnements distribués, les décisions émergent souvent d'interactions multiples entre agents, ce qui rend difficile l'identification des causes et la production d'explications cohérentes au niveau global.

Dans ce contexte, ce travail vise à analyser les enjeux liés à l'explicabilité dans les systèmes multi-agents hybrides. Plus précisément, il propose d'examiner l'état de l'art des approches existantes, d'identifier leurs principales limites et de proposer un cadre conceptuel permettant de mieux comprendre et structurer les mécanismes d'explication dans les décisions collectives.

Le reste de ce rapport est organisé comme suit : la section suivante présente le contexte théorique relatif aux systèmes multi-agents et à l'intelligence artificielle explicable. La troisième section analyse les principales approches proposées dans la littérature. Ensuite, nous formulons le problème de recherche et proposons un cadre conceptuel pour l'explication des décisions collectives dans les systèmes multi-agents hybrides. Enfin, nous discutons les perspectives de recherche et concluons ce travail.

Chapitre 2

Contexte Théorique

2.1 Systèmes Multi-Agents

Les systèmes multi-agents constituent une approche fondamentale dans le domaine de l'intelligence artificielle distribuée. Un système multi-agents est généralement défini comme un ensemble d'entités autonomes, appelées agents, capables de percevoir leur environnement, de prendre des décisions et d'interagir avec d'autres agents afin d'atteindre des objectifs individuels ou collectifs.

Chaque agent possède un certain degré d'autonomie et peut disposer de capacités de raisonnement, d'apprentissage et de communication. L'interaction entre les agents permet de résoudre des problèmes complexes qui seraient difficiles à traiter par un système centralisé. Dans ce type d'architecture, les décisions émergent souvent d'un processus de coordination ou de négociation entre agents.

Les systèmes multi-agents sont largement utilisés pour modéliser des environnements distribués et dynamiques, tels que les systèmes de transport intelligents, les réseaux de capteurs ou encore les systèmes d'aide à la décision. Leur capacité à représenter des interactions multiples et des comportements adaptatifs en fait une approche particulièrement pertinente pour l'étude de systèmes complexes.

2.2 Architectures Multi-Agents Hybrides

Dans de nombreux systèmes intelligents modernes, les approches purement symboliques ou purement basées sur l'apprentissage automatique ne sont pas toujours suffisantes pour répondre aux exigences de complexité et d'adaptabilité. C'est dans ce contexte qu'apparaissent les systèmes multi-agents hybrides, qui combinent différentes formes de raisonnement au sein d'une même architecture.

Dans un système hybride, certains agents peuvent s'appuyer sur des méthodes de raisonnement symbolique fondées sur des règles ou des représentations logiques, tandis que d'autres utilisent des techniques d'apprentissage automatique pour analyser des données et adapter leur comportement. Cette combinaison permet de tirer parti des avantages de chaque approche : la capacité d'interprétation et de

structuration du raisonnement symbolique, ainsi que la puissance d'apprentissage des modèles statistiques.

Cependant, cette hétérogénéité introduit également de nouveaux défis, notamment en termes de coordination entre agents et de compréhension globale des décisions produites par le système. Les décisions collectives résultant de l'interaction entre différents types d'agents peuvent devenir difficiles à analyser, ce qui soulève des questions importantes concernant la transparence et l'explicabilité des systèmes.

2.3 Intelligence Artificielle Explicable

L'intelligence artificielle explicable vise à rendre les décisions des systèmes d'intelligence artificielle compréhensibles et interprétables pour les utilisateurs humains. Avec l'utilisation croissante de modèles complexes, en particulier dans le domaine de l'apprentissage automatique, la capacité à expliquer les décisions d'un système devient un enjeu majeur pour la confiance, la transparence et la responsabilité des systèmes intelligents.

Dans ce contexte, une explication peut être définie comme un ensemble d'informations permettant de comprendre les raisons ayant conduit à une décision donnée. Les approches d'explicabilité peuvent prendre différentes formes, telles que l'identification des facteurs influençant une décision, la visualisation du processus de raisonnement ou encore la génération d'explications textuelles compréhensibles.

Lorsque ces mécanismes sont appliqués aux systèmes multi-agents, la question de l'explicabilité devient encore plus complexe. En effet, les décisions résultent souvent de l'interaction entre plusieurs agents autonomes, chacun disposant de ses propres informations et de ses propres mécanismes de raisonnement. Par conséquent, produire une explication globale cohérente pour une décision collective représente un défi important pour la recherche actuelle.

Chapitre 3

État de l'Art

3.1 Explicabilité dans les Systèmes d'Intelligence Artificielle

Avec la croissance rapide des systèmes d'intelligence artificielle basés sur des modèles d'apprentissage automatique, la question de la transparence des décisions devient un enjeu majeur. Les modèles complexes, en particulier les réseaux neuronaux profonds, sont souvent considérés comme des « boîtes noires », ce qui rend difficile la compréhension des mécanismes qui conduisent à une décision donnée. Dans ce contexte, le domaine de l'intelligence artificielle explicable vise à développer des méthodes permettant d'interpréter et de justifier les décisions produites par ces systèmes.

Plusieurs approches ont été proposées dans la littérature pour améliorer l'interprétabilité des modèles d'intelligence artificielle. Ces approches incluent notamment les méthodes d'explication post-hoc, telles que les techniques d'attribution de caractéristiques ou les méthodes basées sur l'analyse des contributions des variables dans la décision finale. Ces techniques permettent de fournir des explications compréhensibles pour les utilisateurs humains tout en conservant la performance des modèles complexes.

Cependant, la plupart de ces approches ont été initialement conçues pour des systèmes d'intelligence artificielle centralisés ou pour des modèles opérant de manière individuelle. Leur application dans des environnements distribués, tels que les systèmes multi-agents, soulève de nouveaux défis liés à la nature collective et interactive des décisions.

3.2 Explicabilité dans les Systèmes Multi-Agents

Dans les systèmes multi-agents, les décisions émergent souvent de l'interaction entre plusieurs agents autonomes disposant chacun de leurs propres informations et mécanismes de raisonnement. Cette nature distribuée complique considérablement l'analyse et l'explication des décisions produites par le système.

Plusieurs travaux récents se sont intéressés à l'intégration de mécanismes d'explicabilité dans les systèmes multi-agents. Par exemple, certains travaux proposent l'introduction d'agents spécialisés dans la génération d'explications afin d'améliorer la transparence du système global. Dans une étude portant sur un système de recommandation énergétique basé sur une architecture multi-agents, un agent d'explication a été introduit afin de fournir des justifications compréhensibles pour les recommandations produites par le système. Les résultats montrent que l'intégration de mécanismes d'explicabilité permet non seulement d'améliorer la compréhension des décisions, mais également d'augmenter la confiance des utilisateurs dans le système.[1]

D'autres recherches ont exploré l'utilisation de techniques d'apprentissage multi-agents explicables afin de rendre les politiques de décision plus transparentes. Par exemple, certaines approches basées sur l'apprentissage par renforcement multi-agents cherchent à expliquer les stratégies coopératives développées par les agents à travers l'analyse de leur contribution individuelle à la décision collective.[2]

3.3 Méthodes d'explication des décisions collectives

Plusieurs approches ont été proposées pour expliquer les décisions collectives dans les systèmes multi-agents. Parmi celles-ci, certaines méthodes s'appuient sur des concepts issus de la théorie des jeux, tels que les valeurs de Shapley, afin de mesurer la contribution de chaque agent dans une décision collective. Cette approche permet d'identifier l'influence relative de chaque agent dans le résultat global et de mieux comprendre les dynamiques de coopération entre agents.[2]

D'autres travaux récents proposent des méthodes basées sur l'analyse causale afin d'expliquer les comportements collectifs dans les systèmes multi-agents. Par exemple, certains modèles utilisent des modèles causaux structurels et des simulations contrefactuelles afin d'identifier les relations de cause à effet entre les actions des agents et les résultats observés dans le système.[3]

Par ailleurs, plusieurs études ont également exploré l'intégration de techniques d'apprentissage profond et de simulation pour produire des explications compréhensibles dans des environnements multi-agents complexes. Ces approches permettent d'analyser les interactions entre agents et d'identifier les facteurs ayant conduit à une décision collective donnée.[4]

3.4 Agents Neuro-Symboliques

Les agents neuro-symboliques représentent une approche hybride combinant les capacités d'apprentissage des réseaux neuronaux avec les structures logiques et symboliques du raisonnement symbolique. L'objectif principal de ces agents est

de tirer parti à la fois de la puissance prédictive des modèles neuronaux et de l'explicabilité offerte par la représentation symbolique.

Dans un contexte multi-agents, l'intégration de composants neuro-symboliques permet de :

1. **Rendre certaines décisions plus interprétables** : les règles symboliques peuvent être utilisées pour générer des explications lisibles pour l'utilisateur final.
2. **Améliorer la coordination entre agents hétérogènes** : les agents hybrides peuvent partager des connaissances symboliques tout en conservant des modèles neuronaux pour l'adaptation dynamique aux données.
3. **Combiner performance et interprétabilité** : les réseaux neuronaux offrent des performances élevées, tandis que la partie symbolique permet de justifier les décisions.

Plusieurs études récentes mettent en évidence l'intérêt des agents neuro-symboliques pour l'explicabilité. Par exemple :

- **Garcez et al. (2019)** proposent des architectures neuro-symboliques permettant de représenter des connaissances sous forme de règles logiques tout en utilisant des réseaux neuronaux pour l'apprentissage de représentations complexes. Cette approche améliore la traçabilité des décisions dans des systèmes complexes.[5]
- **Marra et al. (2020)** montrent que les agents hybrides neuro-symboliques peuvent générer des explications globales cohérentes même dans des environnements multi-agents distribués, en combinant reasoning symbolique et apprentissage automatique.

Les agents neuro-symboliques apparaissent ainsi comme une solution prometteuse pour résoudre le compromis classique entre performance et explicabilité dans les systèmes multi-agents hybrides. Cependant, plusieurs défis persistent, notamment la scalabilité et la formalisation des explications collectives, qui restent des sujets ouverts pour la recherche actuelle.

Critère	Approches classiques (ML seul en MAS)	Agents Neuro-Symboliques
Base de raisonnement	Modèles statistiques / Deep Learning	Combinaison logique + réseaux neuronaux
Performance	Très élevée sur données complexes	Élevée, mais dépend de l'intégration
Explicabilité	Souvent limitée (boîte noire)	Meilleure grâce aux règles symboliques
Traçabilité des décisions	Faible	Plus structurée
Adaptabilité	Forte (apprentissage automatique)	Forte + structure logique
Complexité	Peut être élevée	Généralement plus complexe à concevoir

TABLE 3.1 – Analyse comparative des approches d’explicabilité dans les systèmes multi-agents

3.5 Comparaison des approches d’explicabilité

Afin de synthétiser les principales approches étudiées dans la littérature, le tableau suivant présente une comparaison des méthodes d’explicabilité appliquées aux systèmes multi-agents, en mettant en évidence leurs principes, leurs niveaux d’explication et leurs principales limites.

Approche	Principe	Niveau	Adaptation aux SMA	Limites
Explainable Machine Learning	Utilisation de méthodes post-hoc pour interpréter les décisions des modèles d’apprentissage automatique	Locale	Applicable mais principalement pour des agents individuels	Difficulté à expliquer les interactions entre plusieurs agents
Valeurs de Shapley	Basé sur la théorie des jeux pour mesurer la contribution de chaque agent dans une décision collective	Locale / Globale	Pertinent pour analyser la contribution des agents	Coût computationnel élevé dans les systèmes multi-agents complexes
Analyse causale	Identification des relations de cause à effet entre actions et résultats	Globale	Permet de comprendre les dynamiques collectives	Modélisation complexe et dépendance à des hypothèses causales
Agents neuro-symboliques	Combinaison de raisonnement symbolique et apprentissage automatique	Locale / Globale	Très adapté aux systèmes multi-agents hybrides	Complexité d’intégration et problèmes de scalabilité

TABLE 3.2 – Comparaison des approches d’explicabilité dans les systèmes multi-agents

3.6 Limites des Approches Existantes

Malgré les avancées importantes dans le domaine de l'intelligence artificielle explicable appliquée aux systèmes multi-agents, plusieurs défis demeurent. Tout d'abord, la plupart des méthodes existantes se concentrent sur l'explication des décisions individuelles plutôt que sur l'explication des comportements collectifs émergents dans les systèmes multi-agents.

Ensuite, les interactions complexes entre agents rendent difficile la production d'explications cohérentes et globales. Dans de nombreux cas, les explications générées restent locales et ne permettent pas de comprendre pleinement les mécanismes de coordination entre agents.

Enfin, il existe encore un manque de cadres théoriques unifiés permettant de formaliser l'explicabilité dans les systèmes multi-agents hybrides combinant raisonnement symbolique et apprentissage automatique. Cette limitation ouvre la voie à de nouvelles perspectives de recherche visant à développer des modèles capables d'expliquer de manière cohérente les décisions collectives dans des environnements multi-agents complexes.

Parmi les limites les plus critiques, la gestion des conflits d'interprétation entre agents hétérogènes reste un problème ouvert : lorsqu'un agent classique basé sur l'apprentissage automatique et un agent neuro-symbolique produisent des explications contradictoires pour une même décision collective, aucun mécanisme formel n'existe actuellement pour arbitrer ces conflits et garantir une explication globale cohérente.

Chapitre 4

Formulation du Problème de Recherche

4.1 Introduction

Dans les systèmes multi-agents hybrides, les décisions collectives résultent souvent de l'interaction entre plusieurs agents hétérogènes utilisant différentes méthodes de raisonnement, telles que le raisonnement symbolique et l'apprentissage automatique. Cette complexité rend difficile l'interprétation des décisions et soulève des questions cruciales en matière de transparence, traçabilité et explicabilité.

Le problème de recherche vise à formaliser ces défis et à proposer un cadre permettant de générer des explications cohérentes pour les décisions collectives.

4.2 Définition des Concepts Clés

Pour formaliser le problème, nous définissons les éléments suivants :

1. Agent ($Agent_i$) : entité autonome capable de percevoir son environnement, de raisonner et de prendre des décisions. Chaque agent est défini comme :

$$\mathbf{Agent}_i = (\mathbf{Perception}_i, \mathbf{Connaissance}_i, \mathbf{Décision}_i, \mathbf{Explication}_i) [6]$$

- $Perception_i$: informations reçues de l'environnement et des autres agents
- $Connaissance_i$: base de connaissances et modèles internes
- $Dcision_i$: mécanisme de prise de décision
- $Explication_i$: module permettant de générer des justifications pour les décisions individuelles

2. **Système Multi-Agents Hybride (SMAH)** : ensemble de n agents hétérogènes

$$\mathbf{SMAH} = \{\mathbf{Agent}_1, \mathbf{Agent}_2, \dots, \mathbf{Agent}_n\}$$

4.3 Décision Collective

La décision globale du système, notée *Décision_globale*, résulte de l'agrégation des décisions individuelles de chaque agent :

$$Décision_{globale} = f(Décision_1, Décision_2, \dots, Décision_n) [2]$$

où f représente le mécanisme d'agrégation (vote, négociation, consensus, etc.) qui combine les décisions locales pour produire un résultat collectif.

4.4 Objectifs de l'Explicabilité

Le problème de recherche consiste à définir un cadre capable de :

1. Fournir des **explications compréhensibles** pour la décision globale
2. Garantir la **cohérence des explications** entre agents hétérogènes
3. Maintenir un **compromis entre performance et interprétabilité**
4. Permettre la **traçabilité des décisions individuelles et collectives**

4.5 Formalisation de la Contribution des Agents

Pour analyser la contribution de chaque agent à la décision globale, on peut introduire une fonction de contribution :

$$Contribution_i = g(Decision_i, Interaction_i)$$

où g mesure l'impact de l'agent i dans le résultat collectif. Cette formalisation permet de :

- Identifier les agents influents
- Produire des explications locales et globales
- Comparer les approches classiques et neuro-symboliques

4.6 Problématique Synthétique

En résumé, le problème de recherche peut être formulé ainsi :

Comment définir un cadre formel d'explicabilité pour les systèmes multi-agents hybrides, capable de générer des explications cohérentes pour les décisions collectives, en tenant compte des interactions entre agents hétérogènes et de la combinaison de raisonnement symbolique et apprentissage automatique ?

Cette formulation servira de base pour proposer un cadre conceptuel dans le chapitre suivant.

Chapitre 5

Cadre Conceptuel Proposé

5.1 Introduction

Suite à la formalisation du problème, nous proposons un cadre conceptuel pour améliorer l'explicabilité dans les systèmes multi-agents hybrides. Ce cadre vise à :

- Fournir des explications cohérentes pour les décisions collectives
- Intégrer les agents neuro-symboliques et les agents classiques
- Assurer un compromis entre performance et interprétabilité

Le cadre repose sur trois couches principales : agents hybrides, explications locales, et agrégation des explications pour produire une explication globale.

5.2 Architecture du Cadre Conceptuel

Voici un diagramme conceptuel illustrant le cadre proposé :

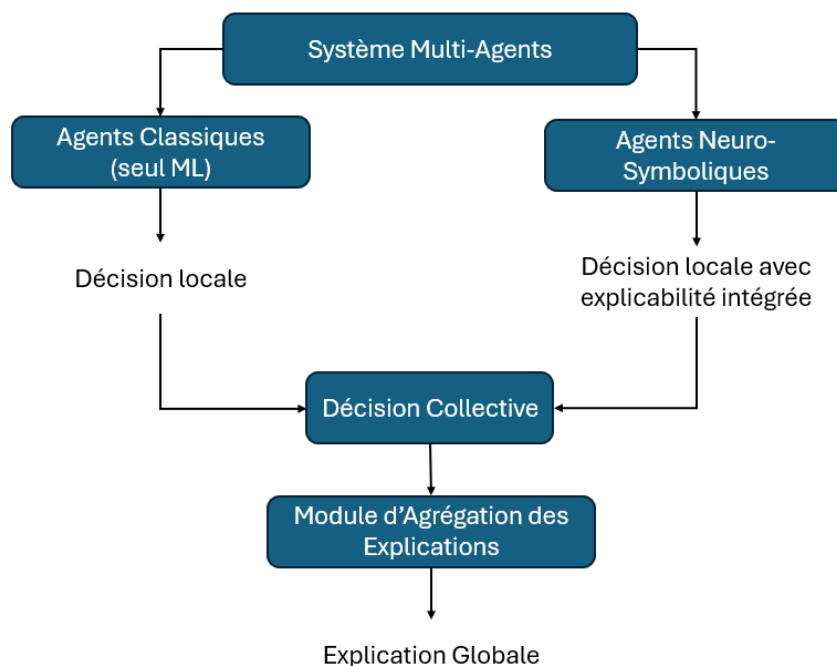


FIGURE 5.1 – Diagramme conceptuel des agents hybrides et de l'agrégation des explications

5.3 Description des Composants

1. Agents Classiques (ML seul)

- Prise de décision basée uniquement sur l'apprentissage automatique
- Fournit une décision locale mais souvent peu explicable

2. Agents Neuro-Symboliques

- Combinaison de raisonnement symbolique et de modèles neuronaux
- Capables de générer des explications locales lisibles et structurées

3. Module d'Agrégation des Explications

- Collecte les explications locales de tous les agents
- Génère une explication globale cohérente pour la décision collective
- Permet de détecter les conflits et d'assurer la cohérence

4. Décision Collective et Explication Globale

- Résultat final du système multi-agents
- Explication structurée permettant à l'utilisateur de comprendre les décisions
- Traçabilité complète : chaque décision locale peut être reliée à l'explication globale

5.4 Fonctionnement du Cadre

1. Chaque agent observe l'environnement et prend une décision locale.
2. Les agents neuro-symboliques produisent explications locales associées à leur décision.
3. Les décisions et explications locales sont transmises au module d'agrégation.
4. Le module produit la décision collective et l'explication globale pour l'utilisateur.
5. Si nécessaire, le système peut analyser la contribution de chaque agent pour améliorer la transparence.

5.5 Avantages du Cadre Proposé

1. Transparence améliorée : toutes les décisions sont traçables et expliquées
2. Compatibilité avec les systèmes hybrides : intégration facile d'agents ML et neuro-symboliques
3. Cohérence des explications : module d'agrégation réduit les contradictions entre agents
4. Flexibilité : cadre applicable à différents types de SMA (recommandation, planification, gestion de ressources)

5.6 Perspectives d'Évaluation

Pour valider ce cadre, on peut envisager :

1. **Mesures de compréhension utilisateur** : est-ce que l'explication est claire et utile ?
2. **Mesures de cohérence globale** : toutes les explications locales concordent avec la décision globale
3. **Mesures de performance** : temps de calcul et qualité des décisions collectives

Chapitre 6

PROTOTYPE

6.1 Introduction

Afin d'illustrer concrètement le cadre conceptuel proposé au Chapitre 5, nous avons développé un prototype fonctionnel d'un système multi-agents hybride (SMAH) appliqué à un cas d'usage réel : la planification intelligente des révisions académiques orientée par objectif de moyenne. Ce prototype constitue une validation empirique du cadre théorique défini précédemment. Il démontre que les quatre composants du cadre Agents Classiques (ML), Agents Neuro-Symboliques, Module d'Agrégation et Explication Globale peuvent être implémentés de manière fonctionnelle, produisant des décisions à la fois performantes et entièrement explicables. Le prototype prend en entrée le profil académique de l'étudiante (moyenne précédente, moyenne cible, état physique, niveau de compréhension par matière) et génère en sortie un plan de révision hebdomadaire personnalisé, accompagné d'une documentation complète de traçabilité décisionnelle.

6.2 Architecture du prototype

L'architecture du prototype suit fidèlement le modèle proposé en Section 5.2 (Figure 5.1). Elle repose sur un pipeline décisionnel structuré en quatre niveaux successifs, depuis la perception des données jusqu'à la génération de l'explication globale. Comme illustré dans la Figure 6.1, le flux de traitement suit quatre étapes principales :

1. Perception des données : collecte des entrées utilisateur (profil académique, état physique, compréhension).
2. Traitement par les agents hybrides : Agent ML pour le scoring, Agent NS pour l'ajustement contextuel.
3. Agrégation : fusion des décisions locales et résolution des conflits potentiels.
4. Génération des sorties : plan de révision + trace d'exécution + explication globale.



FIGURE 6.1 – Architecture du prototype SMAH : pipeline décisionnel du Study Planner

6.3 Description des Composants

Le tableau suivant décrit les quatre composants principaux du prototype en correspondance avec le cadre conceptuel du Chapitre 5 :

Composant	Rôle	Entrées	Sorties
Module Objectif	Calcule la note cible par matière selon la moyenne visée	moyenne_préc., moy. Cible, difficulté, coeff	note_cible, gap, effort
Agent ML Classique	Score de priorité multi-facteurs (décision locale)	diff., coeff., compréhension, gap	score numérique trié
Agent Neuro-Symbolique	Ajustement contextuel par règles SI/ALORS	score ML, état physique, jour, EDT	ordre ajusté + explications locales
Module d'Agrégation	Fusion décisions + traçabilité + résolution conflits	Décision_ML, Décision_NS, contraintes temps	Plan final + Trace complète

TABLE 6.1 – Composants du prototype SMAH et leur correspondance avec le Chapitre 5

6.4 Implémentation des Agents

6.4.1 Agent ML Classique

L'Agent ML calcule un score de priorité numérique pour chaque matière selon la formule suivante :

$$score = difficulté \times coeff \times facteur_compréhension \times facteur_gap$$

où les facteurs sont définis comme suit :

- $facteur_compréhension = 1 + (compréhension - 1) / 10 \rightarrow$ varie de 1.0 (tout compris) à 1.9 (rien compris)
- $facteur_gap = 1 + gap / 20 \rightarrow$ matière avec grand écart par rapport à l'objectif reçoit un score plus élevé
- $gap = \max(0, note_cible - moyenne_précédente) \rightarrow$ points manquants pour atteindre l'objectif

Cet agent fournit une décision locale basée sur des données objectives. Il correspond au composant « Agents Classiques (ML seul) » du Chapitre 5.3, qui produit une décision locale mais peu explicable en isolation.

6.4.2 Agent Neuro-Symbolique

L'Agent Neuro-Symbolique ajuste l'ordre et la durée des sessions selon six règles contextuelles SI/ALORS, intégrant des connaissances cognitives et organisationnelles que l'Agent ML ne peut pas capturer :

Règle	Condition (SI)	Action (ALORS)	Justification cognitive
R1	état ∈ {fatigué, normal, énergétique}	Ordre et durée adaptés (35–75 min)	Adaptation à la capacité cognitive disponible
R2	compréhension ≥ 8/10	+15 min sur la session	Faible compréhension → besoin de plus de temps
R3	gap > 3 ET diff ≥ 12, ou gap > 5 ET diff ≥ 8	Remontée en priorité absolue	Matière difficile avec grand écart → risque fort
R4	matière ∈ cours du jour	Boost ×1.4 sur le score	Consolidation mémorielle : réviser le jour même
R5	matière == 'Anglais Spécifique'	Session courte 20 min, placée en premier	Répétition espacée : efficace pour les langues
R6	coefficient ≥ 2.0	Marquage priorité haute ☆	Impact direct sur la moyenne générale

TABLE 6.2 – Règles SI/ALORS de l'Agent Neuro-Symbolique

Ces règles correspondent au composant « Agents Neuro-Symboliques » du Chapitre 5.3, qui génèrent des explications locales lisibles et structurées, combinant raisonnement symbolique et adaptation dynamique au contexte.

6.4.3 Module d'Agrégation des Explications

Le Module d'Agrégation implémente directement la formalisation proposée en Section 4.3 :

$$Décision_{globale} = f(Décision_1, Décision_2, \dots, Décision_n)$$

$$Contribution_i = g(Decision_i, Interaction_i)$$

Ce module collecte les décisions locales des deux agents, gère les contraintes temporelles (heures disponibles par jour), résout les conflits potentiels entre la durée souhaitée par l'Agent NS et le temps réellement disponible, et produit le plan final. Chaque décision est étiquetée selon sa source :

- ML -> décision issue uniquement de l'Agent ML
- NS -> décision issue uniquement de l'Agent Neuro-Symbolique (ex : Anglais R5)
- ML+NS -> décision combinant les deux agents
- AGG -> décision contrainte par le Module d'Agrégation (conflit temps)

6.4.4 Trace d'Exécution - Explicabilité Démontrée

Le prototype génère automatiquement une documentation complète de traçabilité décisionnelle, accessible directement dans l'interface. Cette trace constitue la validation concrète de l'objectif d'explicabilité défini en Section 4.4 :

Objectif (Section 4.4)	Implémentation dans le prototype	Preuve dans la trace
Explications compréhensibles	Formule ML affichée avec toutes les valeurs numériques	score = diff × coeff × ~~fc~~ × fg détaillé par matière
Cohérence entre agents	Module d'agrégation détecte et résout les conflits	Badge source (ML/NS/ML+NS/AGG) sur chaque décision
Compromis performance / interprétabilité	Agent ML : performance ; Agent NS : interprétabilité	Les deux décisions sont visibles et comparables
Traçabilité complète	Règles NS : déclenchées ou non déclenchées affichées	Toutes les règles évaluées sont documentées par jour

TABLE 6.3 – Correspondance entre objectifs d'explicabilité (Chapitre 4) et implémentation

À titre d'exemple, pour un profil avec moyenne précédente de 11/20 et moyenne cible de 14/20, la trace d'exécution produit pour le lundi :

Agent ML :

```
Parallélisme : 13 × 2.0 × 1.4 × 1.02 = 37.1 (rank #1)
Modélisation : 18 × 1.5 × 1.4 × 1.0 = 37.8 (rank #2)
```

Agent NS (Lundi) :

- ☑ R1 – État normal → ordre moyen→difficile→facile, 55 min
- ☑ R4 – Cours du jour → boost ×1.4 appliqué à Modélisation
- ☑ R5 – Anglais → 20 min, placé en premier
- ☑ R3 – Non déclenchée : aucune matière difficile avec gap critique

Module d'Agrégation (Lundi) :

```
[NS] Anglais Spécifique    20 min → Règle R5 placement fixe
[ML+NS] Modélisation & Graphes 55 min → ML rank#2 + NS boost R4
[ML+NS] Parallélisme        55 min → ML rank#1 + NS badge R6
```

6.5 Conclusion du Chapitre

Ce chapitre a présenté un prototype fonctionnel qui implémente fidèlement l'architecture SMAH proposée au Chapitre 5. Le prototype démontre qu'un système multi-agents hybride peut être à la fois performant grâce à l'Agent ML qui optimise les scores de priorité et entièrement explicable grâce à l'Agent Neuro-Symbolique dont chaque règle est tracée et justifiée, et au Module d'Agrégation qui documente la source de chaque décision finale.

Ce prototype démontre qu'un SMAH peut être à la fois performant et explicable chaque décision est traçable, justifiable, et compréhensible par l'utilisateur final. Il constitue ainsi une réponse concrète à la problématique posée en Section 4.6.

Chapitre 7

Discussion

La modélisation proposée du problème de décision collective dans un système multi-agents intelligent met en évidence l'importance de la coordination entre agents autonomes. Dans un environnement complexe, chaque agent possède une perception locale de la situation et prend des décisions basées sur ses propres connaissances ou modèles d'apprentissage.

Cependant, ces décisions individuelles peuvent parfois être conflictuelles ou sous-optimales lorsqu'elles sont considérées à l'échelle globale. C'est pourquoi les systèmes multi-agents intègrent généralement des mécanismes d'agrégation, de négociation ou de consensus afin de produire une décision collective cohérente.

Dans le contexte des agents neuro-symboliques, cette problématique devient encore plus intéressante. Les approches neuro-symboliques permettent de combiner la capacité d'apprentissage des réseaux neuronaux avec les mécanismes de raisonnement logique issus de l'intelligence artificielle symbolique. Cette hybridation offre un potentiel important pour améliorer la qualité des décisions collectives dans les systèmes distribués.

Ainsi, la formulation du problème présentée dans ce travail vise à structurer le processus de décision collective en tenant compte à la fois des capacités d'apprentissage et des mécanismes de coordination entre agents.

Un défi particulièrement important mis en évidence par ce travail concerne les conflits d'interprétation entre agents hétérogènes. En effet, lorsqu'un agent classique basé sur l'apprentissage automatique et un agent neuro-symbolique produisent des explications divergentes pour une même décision collective, le module d'agrégation proposé dans notre cadre conceptuel joue un rôle central : il permet de détecter ces conflits, de les arbitrer et de produire une explication globale cohérente et compréhensible pour l'utilisateur final.

Chapitre 8

Perspectives de Recherche

Plusieurs pistes de recherche peuvent être envisagées afin d’approfondir cette problématique.

Tout d’abord, il serait intéressant d’explorer l’intégration de modèles d’apprentissage avancés, tels que l’apprentissage par renforcement ou les modèles de deep learning, afin d’améliorer les capacités d’adaptation des agents dans des environnements dynamiques.

Ensuite, le développement de mécanismes de coordination plus sophistiqués pourrait permettre d’optimiser le processus de décision collective. Par exemple, l’utilisation de protocoles de négociation automatisée ou d’algorithmes de consensus distribués pourrait améliorer la robustesse du système.

Par ailleurs, l’intégration complète d’architectures neuro-symboliques dans les systèmes multi-agents constitue une direction prometteuse. Ces architectures pourraient permettre aux agents de combiner apprentissage automatique, raisonnement logique et partage de connaissances.

Enfin, l’évaluation expérimentale de ces approches dans des scénarios réels ou simulés représenterait une étape importante pour mesurer leur efficacité et leur applicabilité dans des domaines tels que les systèmes autonomes, la robotique collaborative ou les systèmes intelligents distribués.

Chapitre 9

Conclusion

Ce chapitre a présenté la formulation du problème de recherche relatif à la prise de décision collective dans les systèmes multi-agents intelligents. L'objectif principal est de comprendre comment les décisions individuelles des agents peuvent être combinées afin de produire une décision globale cohérente et efficace.

La modélisation proposée met en évidence le rôle central des mécanismes d'agrégation et de coordination dans les environnements multi-agents. Elle souligne également l'intérêt des approches hybrides, notamment les agents neuro-symboliques, qui permettent de combiner les capacités d'apprentissage des modèles neuronaux avec les mécanismes de raisonnement symbolique.

Cette formulation constitue une base conceptuelle pour le développement d'approches plus avancées permettant d'améliorer la qualité des décisions collectives dans les systèmes distribués intelligents.

Chapitre 10

Références

- [1] R. Azoulay et al., "An Explainable Multi-Agent Recommendation System for Energy-Efficient Decision Support in Smart Homes," Environmental Data Science, Cambridge University Press, 2023.
- [2] S. Heuillet, F. Couthouis, and N. Díaz-Rodríguez, "Collective eXplainable AI : Explaining Cooperative Strategies and Agent Contribution in Multiagent Reinforcement Learning with Shapley Values," IEEE Computational Intelligence Magazine, vol. 17, no. 1, pp. 59–71, 2022.
- [3] A. I. Weinberg, "MACIE : Multi-Agent Causal Intelligence Explainer for Collective Behavior Understanding," arXiv preprint arXiv :2511.15716, November 2025.
- [4] Z. Wang et al., "From Individual Decisions to Team Emergence : A Survey on Explainable Cooperative Multi-Agent Reinforcement Learning," arXiv preprint, 2024.
- [5] A. d. S. Garcez and L. C. Lamb, "Neurosymbolic AI : The 3rd Wave," Artificial Intelligence Review, vol. 56, pp. 12387–12406, Springer, 2023.
- [6] M. Wooldridge, An Introduction to MultiAgent Systems, 2nd ed. Chichester, UK : John Wiley & Sons, 2009.