Synthèse de la présentation du “Frame Problem”

par Artus Bleton, Guilhem Dupuy, Dimitrios Stephanou

Définition

Le Frame Problem désigne la **difficulté à représenter les effets du changement** dans un système logique, sans devoir tout redéfinir et énumérer tout ce qui ne change pas. Il apparaît dans les années 1960 avec l’émergence de **l’IA symbolique** (logique du premier ordre, raisonnement déductif).
Le problème est toujours **pertinent aujourd’hui**, malgré l’évolution vers des IA plus statistiques.

1. Origine et Fondements

Formulé pour la première fois en 1969 par **McCarthy & Hayes**, dans une tentative d’illustrer les limitations des IAs symboliques de l’époque à fonctionner dans un monde complexe.
Repris notamment par Dennett dans son exemple classique : **le robot de Dennett** doit récupérer une batterie sans déclencher une bombe — il échoue car il ne peut pas facilement distinguer ce qui **change** de ce qui **reste inchangé**, ou bien nécessite un temps de calcul trop long pour le faire. Ils ont également des difficultés à se représenter les **répercussions indirectes** de leurs actions.

Introduction de la nécessité de **"frame axioms"**, ou axiomes de cadre : des règles explicites stipulant que certaines choses et propriétés ne changent pas, sauf indication contraire. Problèmes associés :

- **Qualification** : sous quelles conditions une action modifie l’un de ces axiomes de cadre. Devient très complexe dans un monde complexe. Exemple type : les actions qui impactent la couleur d’un mur incluent le peindre, mais également s’y frotter par erreur, son éclairage, etc
- **Ramification** : quelles sont les conséquences indirectes d’une action ?

2. Enjeux et Solutions

Le frame problem est un enjeu fondamental en Intelligence Artificielle, particulièrement dans l’approche symbolique. Celle-ci doit modéliser explicitement les objets, leurs propriétés et leurs relations dans un monde changeant — une tâche qui devient vite complexe. À l’inverse, les approches connexionnistes (comme les réseaux de neurones) contournent en partie ce problème, car elles n’ont pas besoin de représenter les objets de manière explicite : elles apprennent plutôt des régularités statistiques à partir des données.

1. Enjeux

- **Explosion combinatoire :** décrire toutes les propriétés et effets possibles de chaque action conduit à une croissance exponentielle du nombre de règles.

- **Monde dynamique et incertain :** difficile de maintenir à jour une représentation fidèle quand tout peut changer à tout moment.

- **Coût cognitif et computationnel :** les systèmes symboliques doivent “penser” chaque changement, alors qu’une partie du contexte devrait pouvoir être implicite ou inférée.

2. Solutions proposées

- **Formalismes logiques avancés :**
    - *Fluent Calculus, Situation Calculus, Circonscription logique* : ces cadres tentent de préciser ce qui change et ce qui reste invariant après une action, réduisant la redondance dans les représentations.
    - L'idée est de **raisonner par invariance** plutôt que de re-décrire tout l'état du monde.
- **Approches computationnelles hybrides :**
    - Utilisation de méthodes issues du **machine learning** (clustering, abstraction de caractéristiques) pour identifier les éléments réellement pertinents dans un contexte donné.
    - Ces techniques permettent d'alléger la modélisation symbolique en automatisant la sélection des variables importantes.
- **Modèles cognitifs inspirés de la psychologie :**
    - *Raisonnement défaisable* : accepter qu'un raisonnement puisse être révisé à la lumière de nouvelles informations.
    - *Perception contextuelle* et *implication causale* (ex. travaux de Pollock) : simuler la façon dont les humains filtrent l'information pertinente sans tout formaliser.

3. **Synthèse**

En IA symbolique, l'enjeu n'est pas seulement de représenter le monde, mais de le mettre à jour efficacement.
L'objectif est de concevoir des systèmes capables de raisonner de manière souple et économique — un compromis entre rigueur logique et plasticité cognitive.
Les approches hybrides (symbolique + connexionniste) apparaissent aujourd'hui comme une voie prometteuse pour dépasser les limites classiques du frame problem.

3. Implications Philosophiques

Le Frame Problem et la cognition humaine

Le Frame Problem ne concerne pas seulement les machines : il se manifeste aussi chez l'être humain. Lorsque nous agissons ou raisonnons, nous devons en permanence déterminer quelles informations sont pertinentes et quelles autres peuvent être ignorées. Cette sélection contextuelle, que notre cerveau réalise de manière intuitive, reste extrêmement difficile à formaliser dans un système logique ou algorithmique.

Interprétations philosophiques

Même si le terme "Frame Problem" n'apparaît pas directement dans leurs écrits, plusieurs

philosophes ont proposé des réflexions qui éclairent ses enjeux.

Jerry Fodor, dans sa théorie du langage de la pensée (*The Language of Thought*, 1975), défend une vision symbolique et computationnelle de l'esprit : la pensée s'effectue par manipulation de représentations mentales structurées. Dans cette perspective, le Frame Problem illustre simplement la complexité inhérente à ce type de raisonnement formel.

Daniel Dennett, lui, est l'un des premiers à relier explicitement le Frame Problem à la philosophie de l'esprit. À travers l'exemple de son robot, il montre que ce problème met en évidence la difficulté pour une intelligence, artificielle ou non, de décider rapidement ce qui est pertinent pour agir.

Enfin, **Hubert Dreyfus**, critique du cognitivisme, s'appuie sur la phénoménologie pour souligner que la compréhension humaine repose sur un savoir implicite, corporel et contextuel, que les approches symboliques ne peuvent pas capturer. Le Frame Problem illustre ainsi, selon lui, la limite fondamentale des modèles de raisonnement fondés sur des règles explicites.

4. Discussion & Débat

Quelles IA sont concernées aujourd'hui ?

- **Concernées** : Robots autonomes, systèmes de planification symbolique, voitures autonomes (certaines méthodes)
- **Non concernées** : LLMs, IA statistiques, systèmes de recommandation. voitures autonomes (méthodes deep learning par exemple

Est-ce que seules les IAs symboliques sont concernées par le Frame problem ?

En théorie, oui. C'est dans ce contexte qu'a initialement été posé le problème. Les IAs non-symboliques (réseaux de neurones, modèles probabilistes…) n'ont pas ce souci parce qu'elles n'utilisent pas de représentations logiques explicites du monde.

Cependant, des formes implicites analogues du problème peuvent se poser en IA non symbolique :

- **Pertinence contextuelle** pour des réseaux de neurones, qui peuvent avoir du mal à distinguer ce qui est important dans une situation
- Calcul des impacts de leurs actions pour des modèles de langue : problématique des LLMs qui doivent être ajustés pour éviter certains biais dangereux (racisme, violence). Cependant, ils sont aujourd'hui corrigés de manière artificielle, pas directement par le modèle lui-même car certains défendent qu'il n'a pas de représentation explicites du monde.

Le sujet n'est actuellement pas tranché parmi les scientifiques.

Comment le cerveau humain résout / contourne le problème ?

Via des mécanismes neuronaux conscients de focalisation de l'attention. On filtre les informations en choisissant où on regarde, sur quoi on se concentre, etc. Et via des mécanismes inconscients, de filtre des informations jugées utiles.

Beaucoup de théories existent sur l'émergence de ces comportements, mais pas de consensus.

Faut-il abandonner la logique symbolique pour s'affranchir du Frame Problem ?

Pas forcément, des approches hybrides tentent aujourd'hui d'y répondre. Certaines approches d'apprentissage par renforcement permettent à des modèles d'inférer sur leurs propres règles (à partir d'un set défini au préalable), leur permettant de trouver leur propre réponse au problème.

Conclusion

« Le Frame Problem cessera d'être un problème pour l'IA dès qu'on arrêtera d'attendre d'elle qu'elle le résolve parfaitement — ce que même l'intelligence humaine ne fait pas. »

Il révèle **les limites de la modélisation symbolique**, et ne fait pas consensus au sein de la communauté scientifique. Il motive certaines approches **hybrides**, se voulant plus proches du fonctionnement humain lui-même partiellement connu.