

Atelier IAC : The Symbol Grounding Problem

ASRI Rida, BLUMET Thomas, EDHIB Awatef

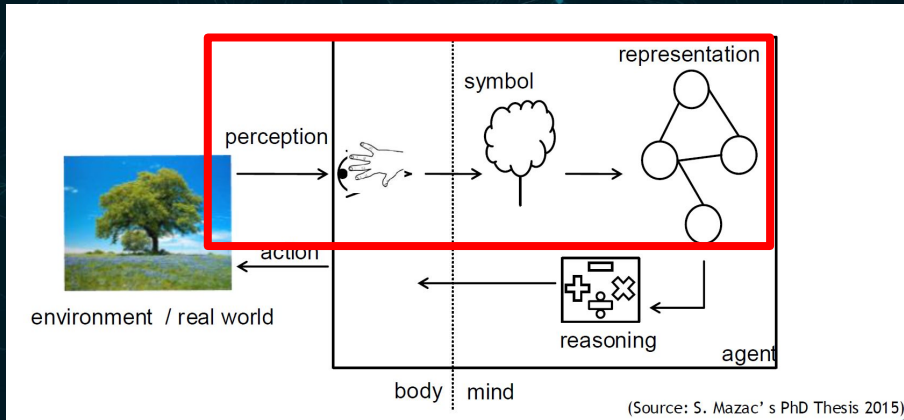
M2IA - Polytech Lyon

Sommaire

- Contexte
- The Symbol Grounding Problem
- L'argument de la "Chinese Room"
- Les solutions au SGP ?
- Débat & Questions
- Conclusion
- Bibliographie

Contexte

Système cognitiviste



Malgré des succès → problème de fossé sémantique !

Qu'est-ce qu'un symbole ?

→ Du grec "sumbolon" : "comparer", "mettre (des choses) ensemble"

→ Représentation de l'abstrait, de ce qui manque ; quelque chose qui représente autre chose ; dont le sens ne se révèle pas immédiatement

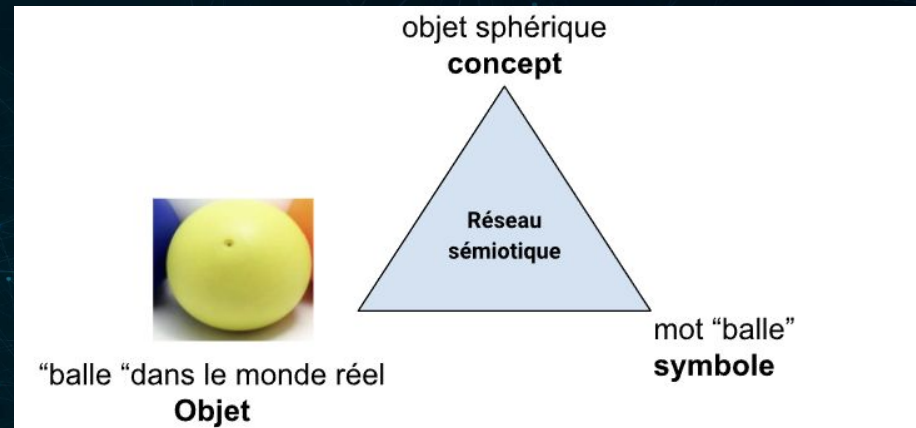
→ Un symbole est ancré $\Leftrightarrow \exists$ méthode associant le symbole à l'objet

→ Notion de réseau sémiotique associé

→ Existence de symboles fondés vs non-fondés :

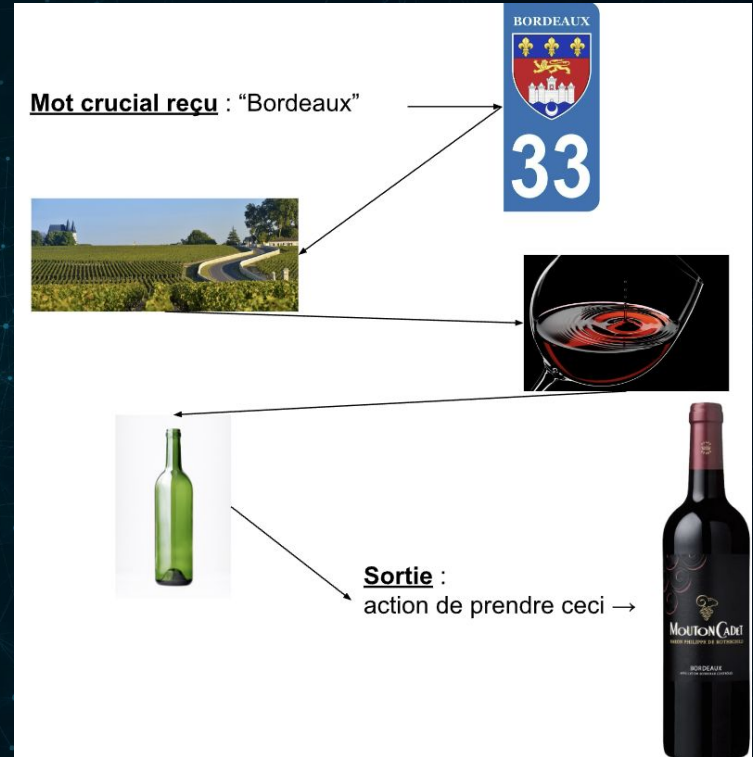
"bourdonnement", "crash" : fondé car relié aux sons produits

"sérendipité", "démocratie" : non-fondés



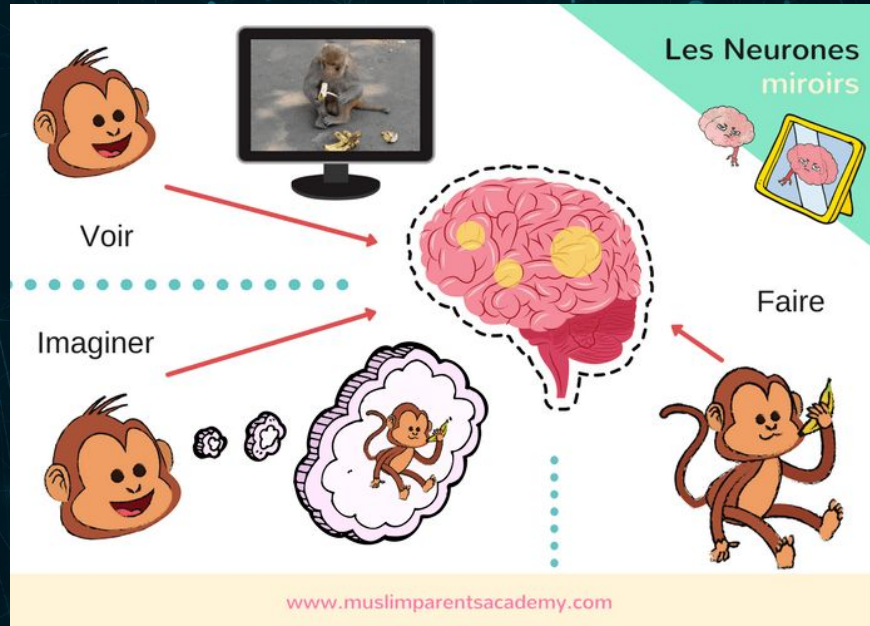
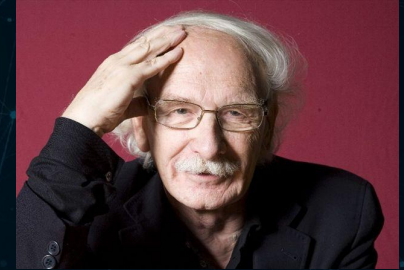
Réseau sémiotique

- Chaque individu se le construit et le maintient à jour
- Ex : “Peux-tu me passer le Bordeaux posé sur la table ?”
- Au fil des expériences : coordination de notre réseau à celui d'autres individus (mais jamais identité)
- nommé “semiotic landscape”
(Steels, *The SGP has been solved so what's next*, 2008)
- différent selon les cultures (ex: couleur rouge)



Un puissant moyen d'apprentissage : les neurones miroirs

- Découvert en 1990 par Giacomo Rizzolatti chez le singe
- Permet d'apprendre via l'imitation et l'observation



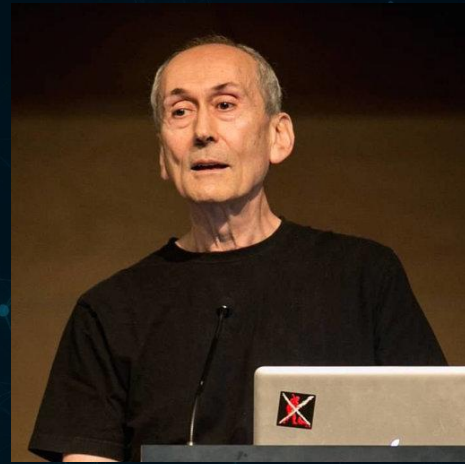
The Symbol Grounding Problem

De quoi s'agit-il ?

Stevan Harnad, *The Symbol Grounding Problem*, 1990

→ pose la Q° suivante : “**Comment les systèmes de symboles formels peuvent interpréter les symboles d'un point de vue sémantique ?**”

Un exemple d'illustration : **Shakey le robot**



Plus de détails ...

Propriétés d'un système de symbole :

- ensemble de “jetons physique arbitraires”
- de 2 types : atomique et composite
- manipulé via un ensemble de règles explicites
- manipulation basé sur la forme et non la signification des jetons
- de la **manipulation syntaxique** naît une **interprétation sémantique**
→ **Le système est sémantiquement interprétable dans son ensemble !**

Ex: $a+2=b$

- symboles atomique: $a, b, +, =, 2$
- chaîne composite: $a+2$
- règle de manipulation: $+$ combine a et 2 pour donner une nouvelle valeur
- l'interprétation sémantique: la valeur de a augmentée par 2 est assignée à la valeur b

S'inspirer des capacités humaines

1. **Capacité de discrimination** : juger si deux entrées sont identiques ou différentes
2. **Capacité d'identification** : attribuer un "nom" à une classe d'entrées considérées comme équivalentes dans certains aspects

Exemple : Des chevaux et un âne



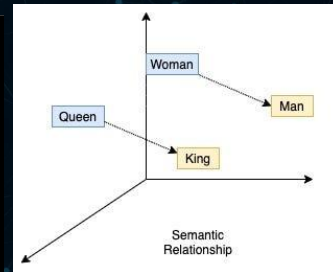
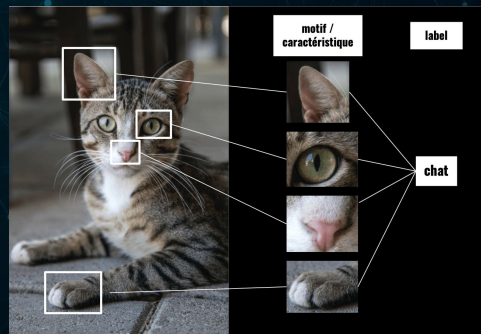
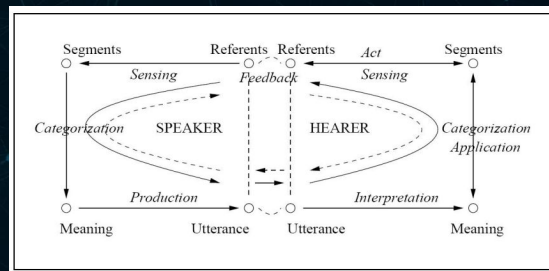
D'où 2 types de représentations (Harnad):

1. **iconique** : Copies analogiques des projections sensorielles servant la discrimination
2. **catégorielle** : filtrage des caractéristiques distinctives de ces projections servant l'identification

Mais la représentation symbolique dans tout ça ?

Revue sur les différents types d'ancrage

- **Sensorimoteur** (via l'embodiment e.g. ASIMO de Honda)
- **Communicatif** (e.g. The Guessing Game)
- **Épistémique** : connexion entre symboles et source externe d'informations
- **Relationnel** (e.g. Word2Vec)
- **Référentiel** (relié symbole à un objet concret)



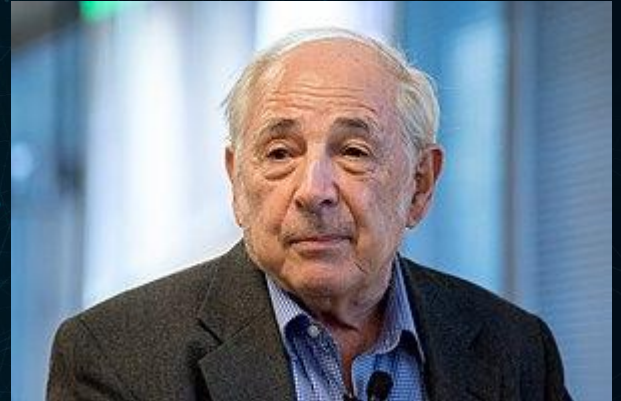
L'argument de la “Chinese Room”

De quoi parle-t-on ?

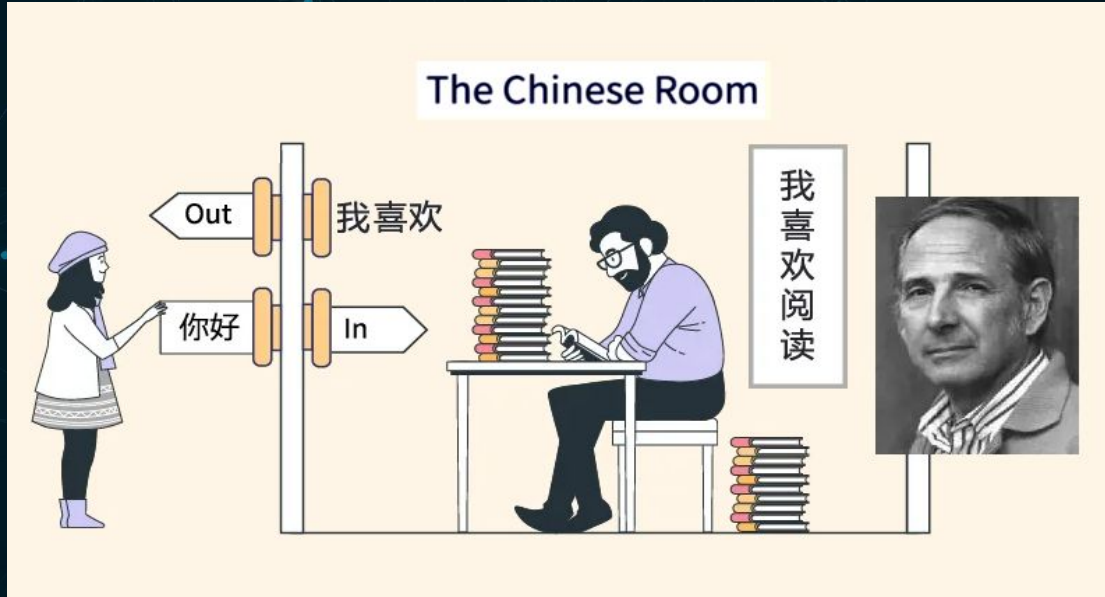
- John Searle, "*Minds, Brains and Programs*", 1980
→ Est-ce qu'un ordinateur pourrait comprendre vraiment ce qu'il fait ?
- Lien avec le SGP ?



→ Harnad ne répond pas directement à l'argument de Searle !



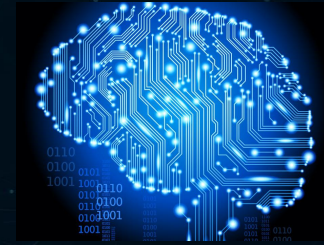
Description de l'expérience de pensée



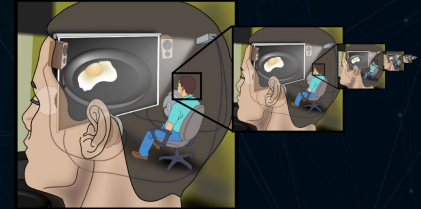
- **Conclusion** : si remplace humain par un ordinateur → même problème !
- Manque d'intentionnalité abordé par Searle (côté IA) et Brentano (côté philosophie esprit)



L'avis de Searle sur l'approche computationnelle de l'esprit



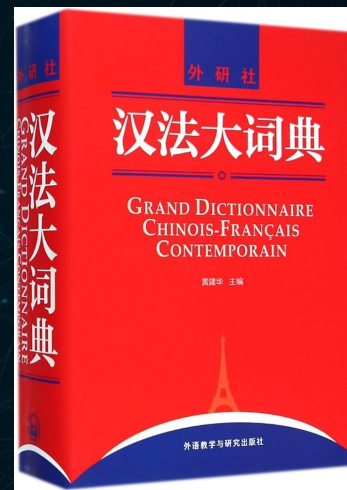
- Searle, “*Is the Brain a Digital Computer*”, 1990
 - Q° : Est-ce que le cerveau humain fonctionne comme un ordinateur numérique ?
- Le problème de l'explication causale
→ Pas de "niveau programmatique" indépendant responsable de la cognition
- Critique du cognitivisme → néglige l'importance des mécanismes physiques réels du cerveau
- Problème de l'homoncule → le cerveau ne suit donc aucun “programme”
- Conclusion :



Le point de vue d'Harnad sur Searle

- **Le tour du dictionnaire chinois/chinois :**
→ apprendre le chinois avec un dictionnaire chinois/chinois
- D'où 2 versions de la Chinese Room :
 - Difficile : le chinois comme 2nd langue

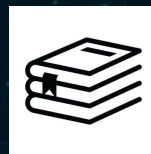
=> Créer un cycle infini des symboles
 - Impossible : le chinois comme 1ère langue



Les solutions au SGP ?

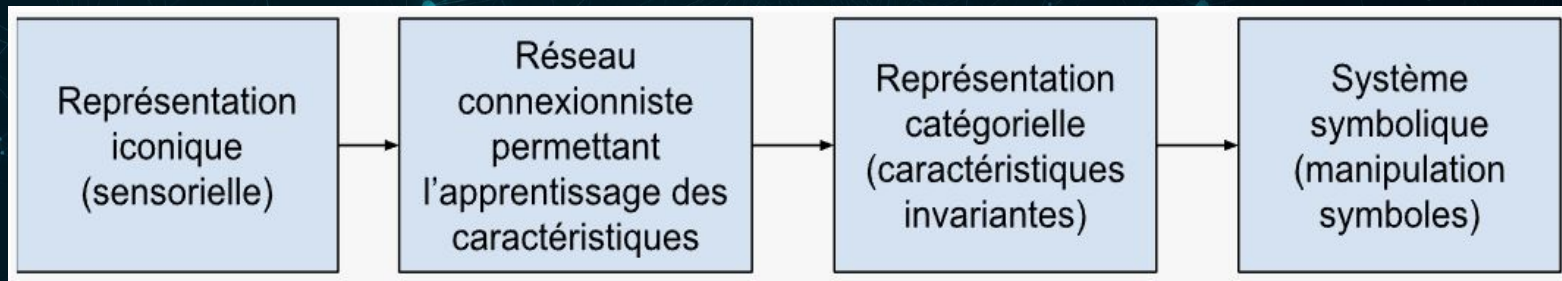
Les approches intéressantes

- concept de **Minimum Grounding Sets**



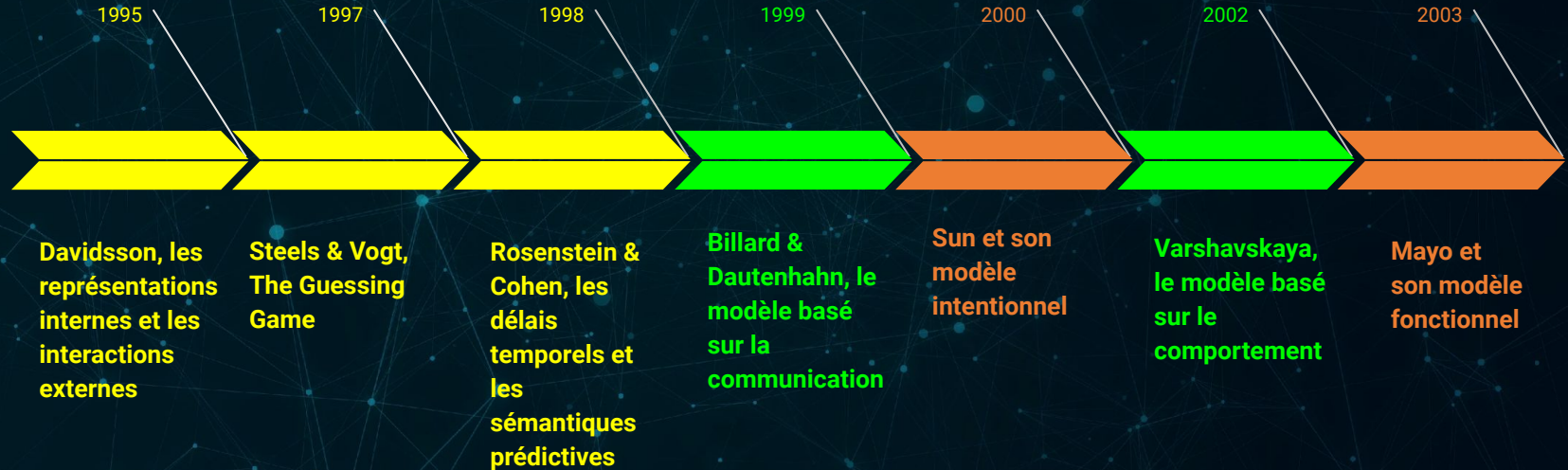
- **Modèle hybride de Harnad :**

→ lié les symboles élémentaires à des représentations via des réseaux connexionnistes



Les autres solutions proposées sans réussite

Approche **représentationnaliste**
semi-représentationnaliste
non-représentationnaliste



Pourquoi ? Non respect de la “Z-condition”

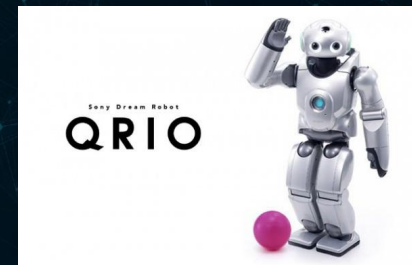
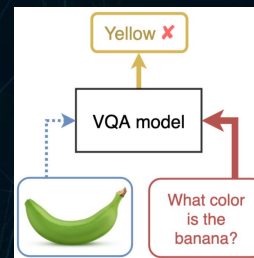
Taddeo & Floridi, *Solving the SGP : a critical Review of 15 years of Research* (2005)

Les principales approches actuelles et perspectives

- **Ancrage neuro-symbolique** : mix symbolisme / connexionnisme (manipulation symboles + réseaux neurones)
Exemple : VQA
- **Association symbole-symbole** : de la combinaison entre symboles naît leur signification
Exemple : modèle Word2Vec en NLP
- **Cognition incarnée** : les symboles s'ancre grâce aux interactions de l'agent avec le monde réel
Exemple : robots interactifs

Perspectives :

- Utilisation de jumeaux numériques → améliorer l'ancrage avec représentation entités physique





Merci de votre attention !

Débat & Questions

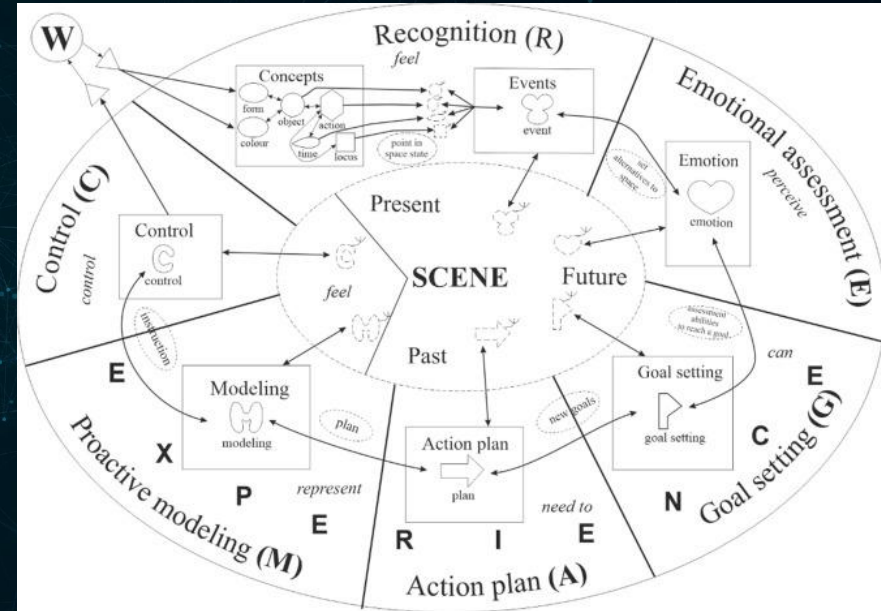
Liste de questions

- L'IA pourra-t-elle un jour vraiment comprendre, ou sera-t-elle toujours une impressionnante imitatrice ? (exemple de ChatGPT ou Copilot)
- Lorsque vous demandez à une IA générative des informations sur un sujet que vous ne connaissez pas, contentez-vous seulement de les “recopier” ou bien faites-vous l'effort de comprendre, en vous assurant d'avoir compris ce qui est dit ?
- Que penser du fonctionnement avec des processus quantifiés dans le cerveau (Roger Penrose) ? (Argument selon lequel certains aspects de la conscience ne peuvent pas être expliqués par des processus algorithmiques classiques et nécessitent des phénomènes quantiques non-computables)

En résumé ...

Le SGP est-il considéré comme résolu ?

- Plusieurs tentatives infructueuses
- Diverses approches testées actuellement
- Dépendance vis-à-vis des progrès technologique
- La question de départ a évolué : comment connecter les symboles au monde réel pour qu'ils acquièrent leur signification ?
- But de réaliser à l'avenir une AGI → nécessaire de résoudre SGP
- Selon Steels oui mais pour d'autres non !
- Au final recherches toujours en cours !



Bibliographie

Stevan Harnad, *The Symbol Grounding Problem* (1990) , <https://arxiv.org/html/cs/9906002>

Luc Steels, *The SGP has been solved so what's next* (2008) , <https://ai.vub.ac.be/sites/default/files/steels-08e.pdf>

John Searle, *Is the brain a digital computer* (1990), <https://courses.cs.umbc.edu/471/papers/searle.htm>

Stevan Harnad, *Language Writ Large: LLMs, ChatGPT, Grounding, Meaning, and Understanding* (2024) , <https://philarchive.org/rec/HARLWL-4>

Mariarosaria Taddeo et Luciano Floridi, *Solving the Symbol Grounding Problem: a Critical Review of Fifteen Years of Research* (2005) , https://www.cs.ox.ac.uk/activities/ieq/research_reports/ieq_rr210605.pdf

Jianhui Li , *The Difficulties in Symbol Grounding Problem and the Direction for Solving It* (2022) , <https://www.mdpi.com/2409-9287/7/5/108>

Goonmeet Bajaj, *Grounding From an AI and Cognitive Science Lens* (2024), <https://arxiv.org/pdf/2402.13290>

Maher ,Ventura & Magerko, *The Grounding Problem: An Approach to the Integration of Cognitive and Generative Models* (2023), <https://ojs.aaai.org/index.php/AAAI-SS/article/view/27695/27468>

Minao Kukita, *Has the symbol grounding problem been solved or not ?* , <https://www.is.nagoya-u.ac.jp/dep-ss/phil/kukita/works/Kyoto-NUS-WS2014.pdf>