
会! 流利地
Yes! fluently

Chinese Room - L'IA Peut-Elle Vraiment Comprendre ?

Une exploration philosophique de la conscience artificielle à travers l'expérience de pensée la plus célèbre de John Searle : la Chambre Chinoise.

Réalisé par

Najid Mohamed **IDBOUSSADEL Abdessamad**

ITJI Amine **BOUAMAMA Youssef**

Sommaire

Contexte

1. Fin des années 1970
2. L'Intelligence Artificielle symbolique est en plein essor.
3. Les chercheurs pensent qu'un programme pourrait « penser » s'il manipule correctement des symboles.
4. John Searle publie en 1980 son article "Minds, Brains and Programs" pour critiquer cette idée.

Système expert ex (MYCIN)

Jerry Fodor "The Language of Thought"

Chinese room

Objectif :

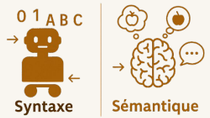
➔ simuler la compréhension ≠ comprendre réellement.

Problématique

La manipulation de symboles suffit-elle à produire une véritable compréhension, ou n’en donne-t-elle qu’une illusion ?

- Les IA modernes peuvent dialoguer, répondre à des questions, résoudre des problèmes…
- Mais “comprendre”, est-ce simplement exécuter un programme ?
- La pensée humaine peut-elle être réduite à une suite de règles formelles ?

Concepts clés



Syntaxe vs Sémantique

Appliquer des regles

Donne un sens

2+2 = 4

IA Forte

Exécution d'un algorithme

Compréhension

Machine pensante

Naturalisme biologique

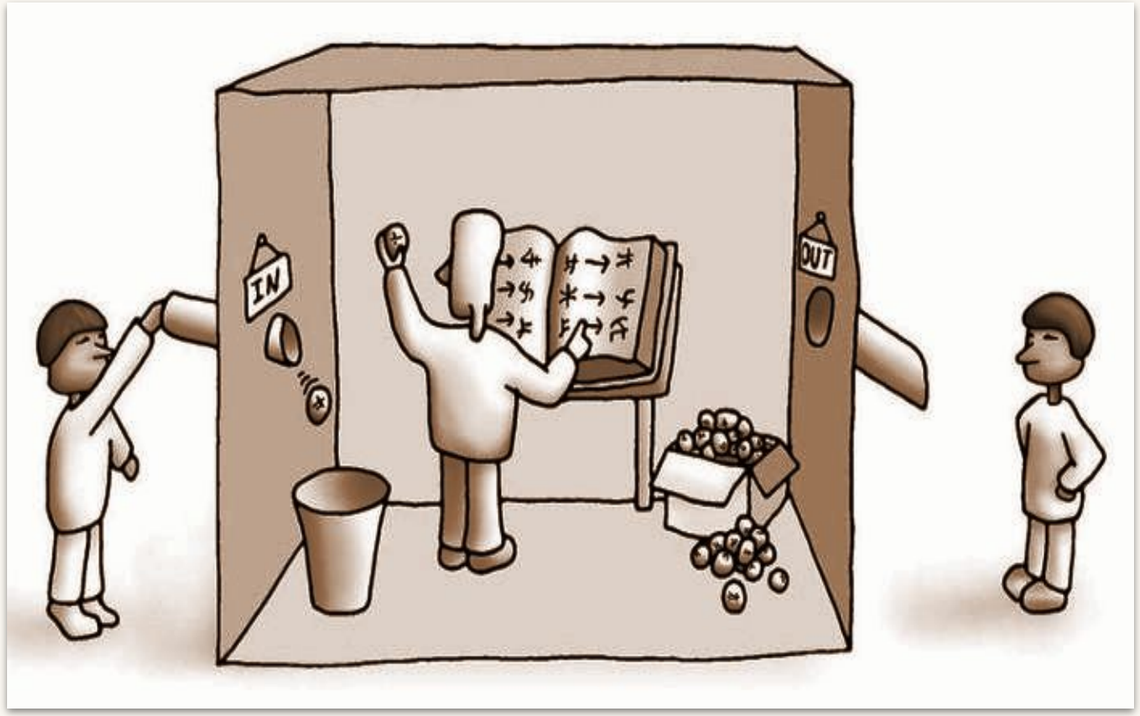
Pouvoirs causaux → Condition suffisante

Intentionnalité

L’expérience de pensée : La Chinese Room

Scenarios de l’experience :

1. Une personne ne parlant pas chinois
2. Elle reçoit des symboles chinois par une fente.
3. Elle possède un manuel (règles) qui lui dit quoi répondre.
4. Elle renvoie les bonnes réponses.
5. De l’extérieur, on croit qu’elle comprend le chinois.
6. En réalité, elle manipule des symboles, elle ne comprend pas.

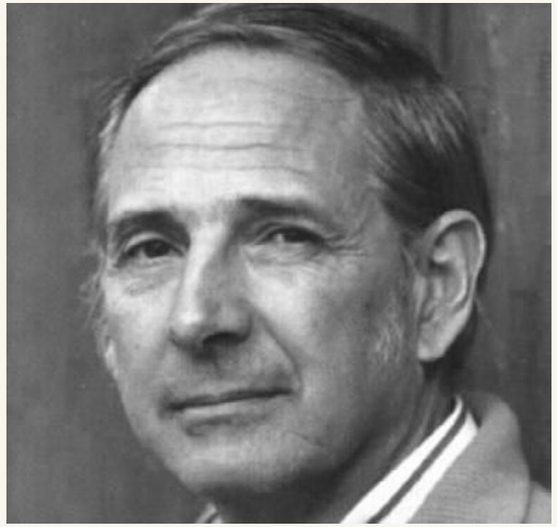


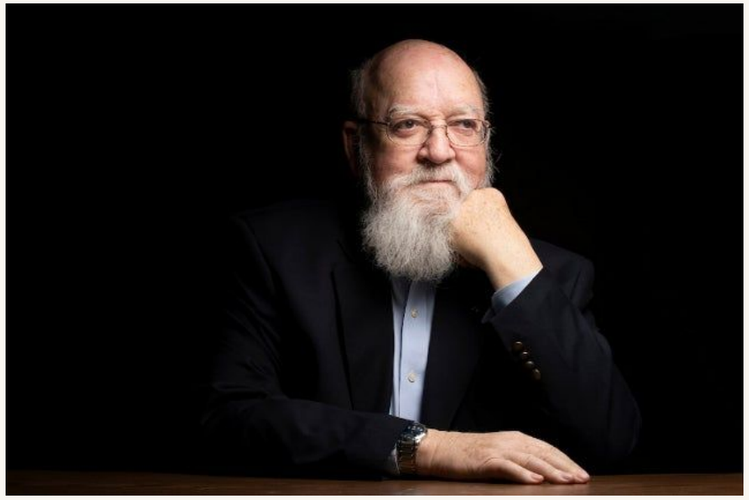
> Si la machine ne comprend pas, peut-on dire que “le système” dans son ensemble comprend ?

The Systems Reply

Les Contre-Arguments à la Chinese Room

La grande fracture philosophique post-1980





Systems Reply : La compréhension émergente

Denett (1987), Block (1978)

- **Le Mauvais Agent :** L’opérateur n’est que le CPU
- **Thèse Fonctionnaliste :** La compréhension est une propriété émergente du système global (Opérateur + Manuel + Boites)
- **La Réplique de Searle :** L’internalisation de toutes les règles ne crée pas l’intentionnalité

Robot Reply : L’Ancrage des Symboles

Harnad (1990), Brooks (1991)

- **Problème :** Le système est désincarné et isolé du monde réel
- **Le SGP (Symbol Grounding Problem) :** Les symboles sont “non ancrés”. Ils n’ont de sens qu’en référence à d’autres symboles
- **La solution :** Le système doit être incarné (robot) avec des capteurs et effecteurs.

Brain Simulator Reply : Le Substrat

Churchland (1989), Chalmers (1996)

- **La Critique :** Searl ne critique que l'IA Symbolique
- **Thèse Neuro-computationaliste :** Simuler les propriétés causales neurobiologiques d'un générerait la compréhension.
- **Contre-Argument de Searle : Simuler = Produire.** La simulation reproduit la forme mais pas le pouvoir causal réel.

Other Minds Reply : Le Problème de l’inférence

Clarks & Chalmers (1998)

- **Problème :** Nous n’avons pas accès à la conscience d’autrui.
- **Principe de parité :** Si l’IA passe le test de Turing alors…
- **Théorie Associée : Extended Mind** La cognition s’étend aux outils. Le manuel pourrait être une extension cognitive du système.

Des Arguments Théoriques aux Machines Réelles

- **Pendant des décennies le débat s’est déroulé sur des programmes hypothétiques**
- **Aujourd’hui, nous faisons face aux LLM**

Large Language Models et le retour au débat

Fonctionnement des LLMs

Architecture Transformers (Vaswani 2017)

- Mécanisme d’auto-attention
- Pondération contextuelle des tokens

Entrainement massif

- Plusieurs trillions de tokens
- Objectif : prédiction du token suivant

Capacités Observées

Capacités cognitives

- Raisonnement complexe
- Mémoire contextuelle étendue
- Génération créative (codde, poésie, essais)
- Theory of mind

Capacités émergentes (Wei 2022)

- Apparitions soudaine au-delà de seuils critiques
- Arithmétique complexe
- Résolution de problémes abstraits

Limites Fondamentales

Hallucinations

- Génération d'informations factuellement incorrectes
- Absence de vérification externe
- Confiance apparente élevée

Manque de grounding

- Aucune expérience sensorimotrice
- Symboles linguistiques désincarnés
- Pas d'ancrage perceptuel dans les concepts

Absence d'intentionnalité

- Pas de buts, désirs ou intentions propres
- Calculs probabilistes uniquement
- Pas d'intentionnalité philosophique

"Perroquets stochastiques" (Bender et al., 2021)

Retour à Searle : “Syntaxe ≠ Sémantique”

Principe fondamental de Searle :

> La manipulation syntaxique, aussi sophistiquée soit-elle, ne produit pas de compréhension sémantique.

Reformulation moderne (Shanahan, 2024) :

> Réseau de neurones artificiels ↔ Personne dans la chambre
> Tokens textuels ↔ Symboles chinois
> Opérations mathématiques ↔ Règles du manuel
> Aucun "moment de compréhension"

PERSPECTIVE PHILOSOPHIQUE

Elle se divise en deux positions :

Position réaliste / Searle :

- Une IA ne fait que traiter des symboles, elle ne comprend pas.
- La compréhension est un produit du cerveau biologique.

Position fonctionnaliste / Dennett, Chalmers :

- Si un système agit comme s'il comprenait, il peut être considéré comme comprenant fonctionnellement.
- La compréhension serait une propriété émergente du comportement global.

Les questions qui se posent :

- Comprendre ou simuler la compréhension ?

PERSPECTIVE COGNITIVE

Point de vue de Harnad et Clark :

- Embodied Cognition : le sens naît de l'interaction avec le monde.
- Nécessité d'un corps et d'une expérience réelle : pour comprendre, il ne suffit pas de manipuler des symboles. il faut percevoir, agir et ressentir dans le monde.

Position neurocognitive - McCulloch :

- L'esprit et la compréhension émergent des circuits neuronaux : "The mind is in the head".
- McCulloch renforce l'idée que la compréhension n'émerge pas uniquement de la manipulation symbolique, mais d'un système physique complexe comme le cerveau.

- Faut-il une conscience biologique pour comprendre ?
- Peut-on comprendre sans perception, émotions et expérience vécue ?

PERSPECTIVE ÉTHIQUE

Une IA peut produire des réponses cohérentes sans rien “comprendre”. Cela pose des problèmes de responsabilité, de confiance et de biais.

Point de vue de Floridi :

- Une IA ne comprend pas moralement ses actes : responsabilité portée par les concepteurs et utilisateurs.
- Risque d’illusion de compréhension : hallucinations, décisions automatisées.

Questions ouvertes :

- Peut-on donner de la responsabilité et de la confiance à l’IA dans le monde réel ?
- Faut-il prévoir un cadre légal pour des IA plus avancées ou conscientes ?

Conclusion

Textes Fondamentaux

Searle, J.R. *"Minds, Brains, and Programs"*, 1980

McCulloch, W.S. *"Why the Mind Is in the Head"*

Fodor, J. *"The Language of Thought"*

Dennett, D. *"Consciousness Explained"*

Encyclopédies Philosophiques

- Stanford Encyclopedia of Philosophy : "Chinese Room"
- Internet Encyclopedia of Philosophy : "Chinese Room Argument"
- Philosophy Now : débats contemporains

Ressources Vidéo

- BBC : The Chinese Room Experiment (Marcus du Sautoy)
- "The Chinese Room Is a Dishonest Argument"
- "Can Machine Fool Us?"

QUIZ

https://www.menti.com/al3xdppmts5t