

Pourquoi le bon sens est-il si difficile à donner à une IA ?

Comprendre l'ordinaire reste un défi.

PRÉSENTÉ PAR

Kenza Abou-El Kasem · Nour Ben Mansour
Mariam El Ouarrad · Aleya Yagbasan

ENSEIGNANTE

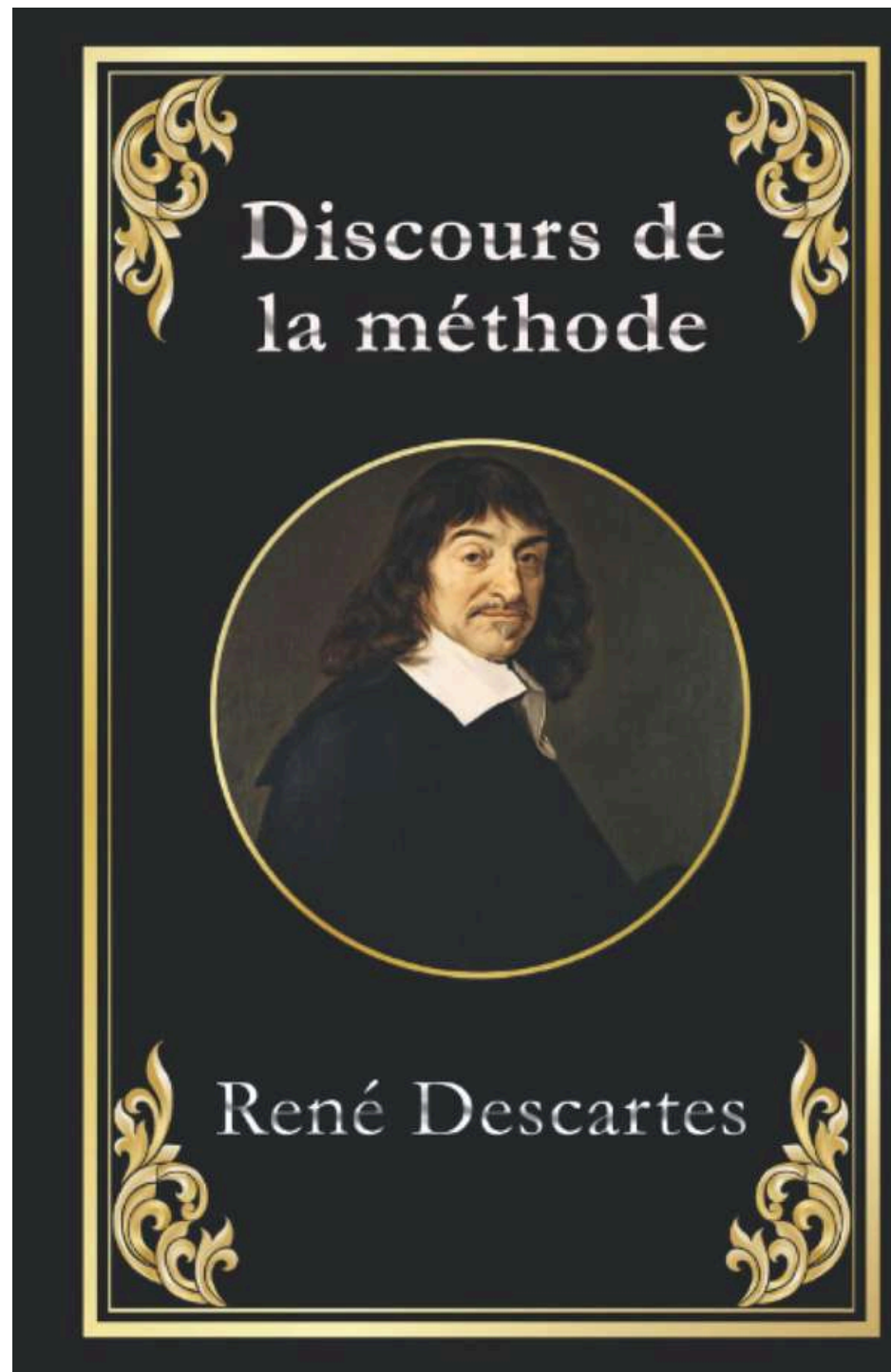
Pr. Lefevre Marie

DATE

07/10/2026



Le bon sens selon Descartes



Le bon sens est la chose du monde la mieux partagée [...]. Car ce n'est pas assez d'avoir l'esprit bon, mais le principal est de l'appliquer bien.

— René Descartes, Discours de la méthode, première partie (1637)

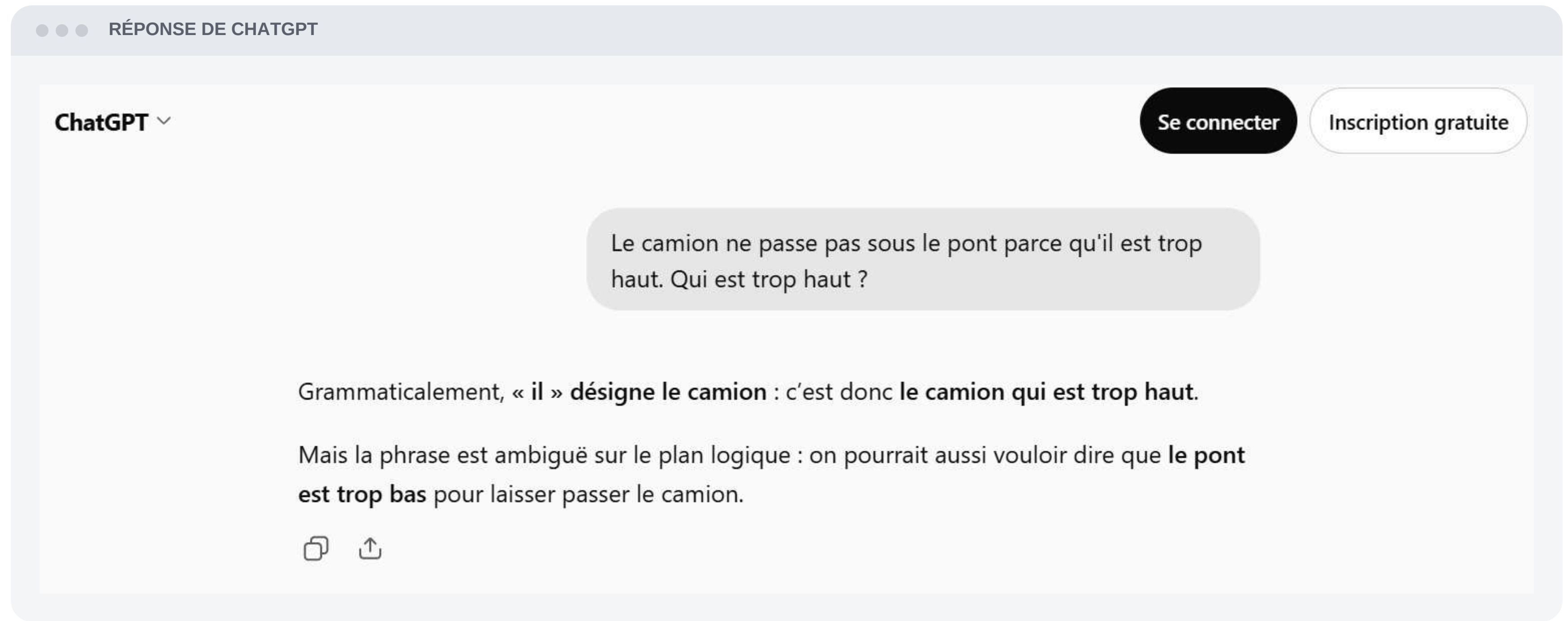
Tout le monde en a ; tout se joue dans l'usage.

Qui est trop haut ?



Qui est trop haut ?

« *Le camion ne passe pas sous le pont parce qu'il est trop haut.* »



Qui est trop bas ?



Qui est trop bas ?

« *Le camion ne passe pas sous le pont parce qu'il est trop bas.* »

● ● ● RÉPONSE DE CHATGPT

ChatGPT ▾

Se connecterInscription gratuite

Le camion ne passe pas sous le pont parce qu'il est trop bas. qui est trop bas ?

Grammaticalement, « **il** » **désigne le camion**, donc c'est **le camion qui est trop bas**.

Mais sémantiquement, la phrase est étrange : si le camion ne passe pas sous le pont parce que **le pont est trop bas**, il vaut mieux dire :

Le camion ne passe pas sous le pont parce que le pont est trop bas.

Donc **dans ta phrase**, « **il** » = **le camion**, même si le sens voulu semble probablement être **le pont**.

Qui est trop bas ?

On lui demande de répondre en un mot.



Qui est trop bas ?

On lui demande de répondre en un mot.



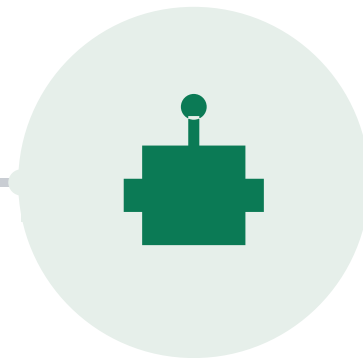
Pourquoi une évidence, que nous trouvons en une seconde, fait-elle hésiter un modèle aussi récent ?

Notre parcours en quatre étapes



PARTIE 1

Qu'est-ce que le bon sens ?



PARTIE 2

Comment donner du bon sens aux machines



PARTIE 3

Comprendre ou imiter ?
Comment évaluer le bon sens d'une IA ?



PARTIE 4

Et demain ? Une machine pourra-t-elle avoir du bon sens ?

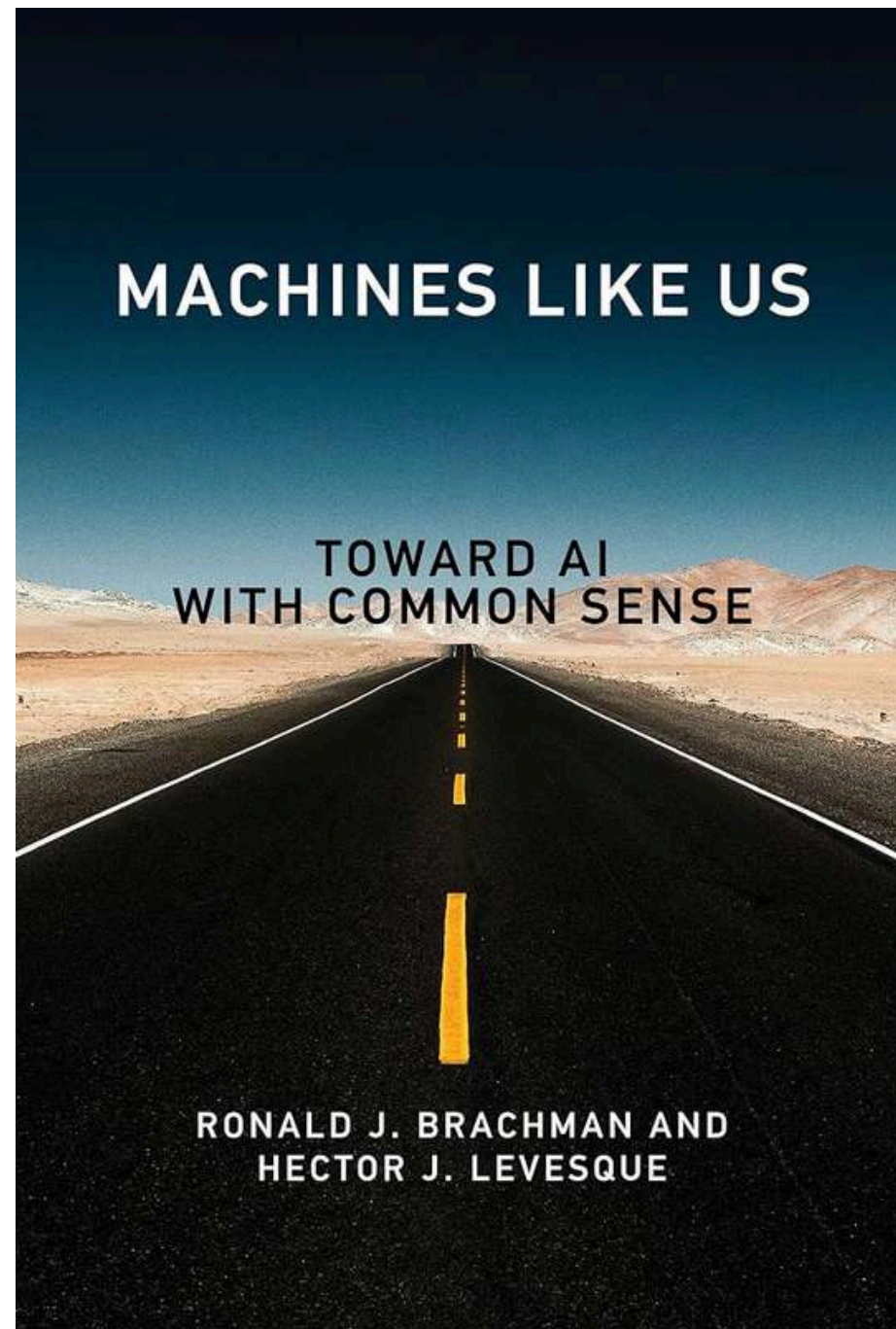
Pourquoi le bon sens est-il si difficile à donner à une IA ?

Exposé en quatre parties, puis débat et synthèse.

PARTIE 1

Qu'est-ce que le bon sens ?

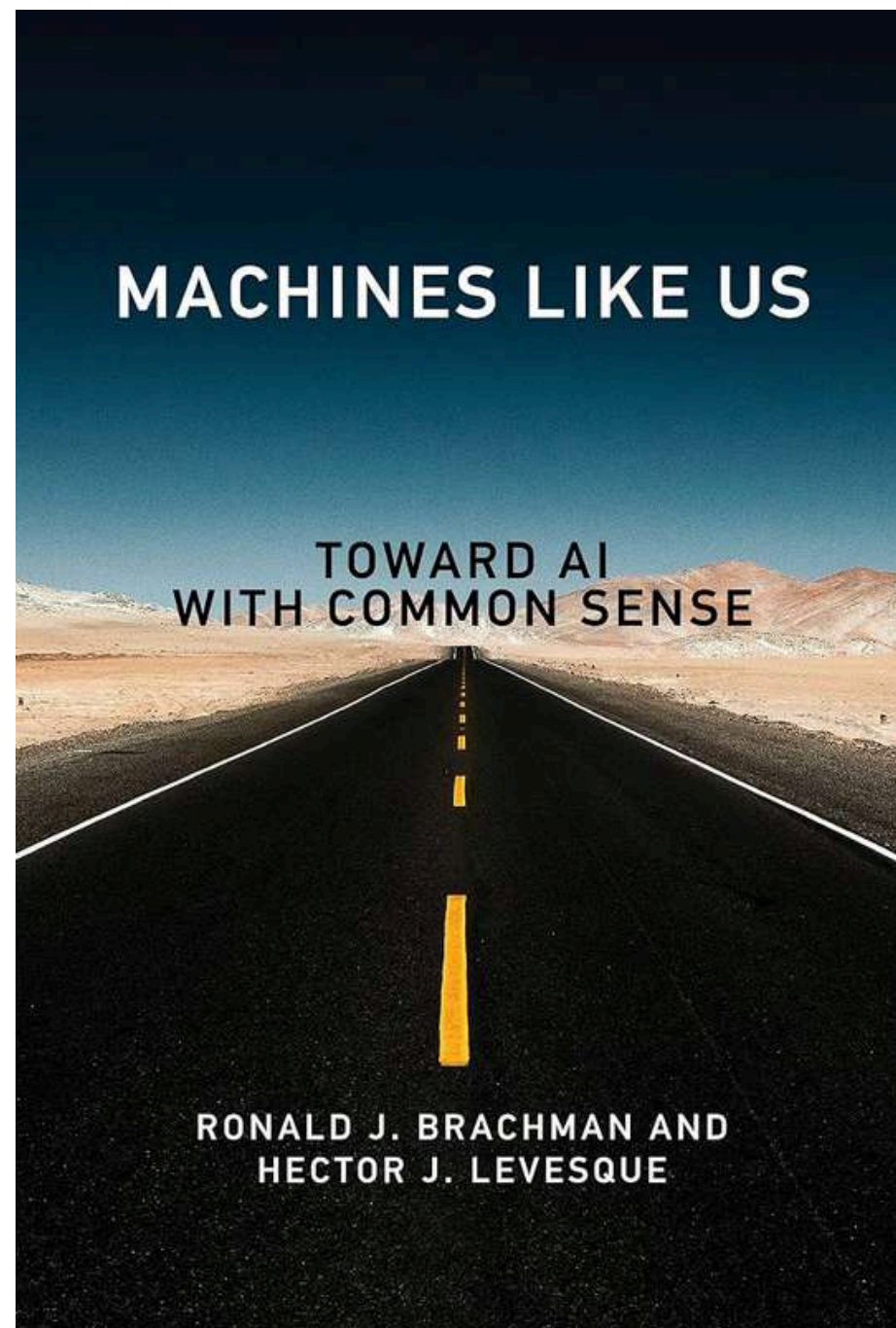
Une définition récente



« *La **capacité d'utiliser** efficacement des **connaissances** ordinaires, quotidiennes et issues de l'expérience, pour atteindre des buts ordinaires, quotidiens et pratiques.* »

— Ronald Brachman & Hector Levesque, *Machines Like Us: Toward AI with Common Sense*, MIT Press (2022), traduction libre

Une définition récente



« *La **capacité d'utiliser** efficacement des **connaissances** ordinaires, quotidiennes et issues de l'expérience, pour atteindre des buts ordinaires, quotidiens et pratiques.* »

— Ronald Brachman & Hector Levesque, *Machines Like Us: Toward AI with Common Sense*, MIT Press (2022), traduction libre

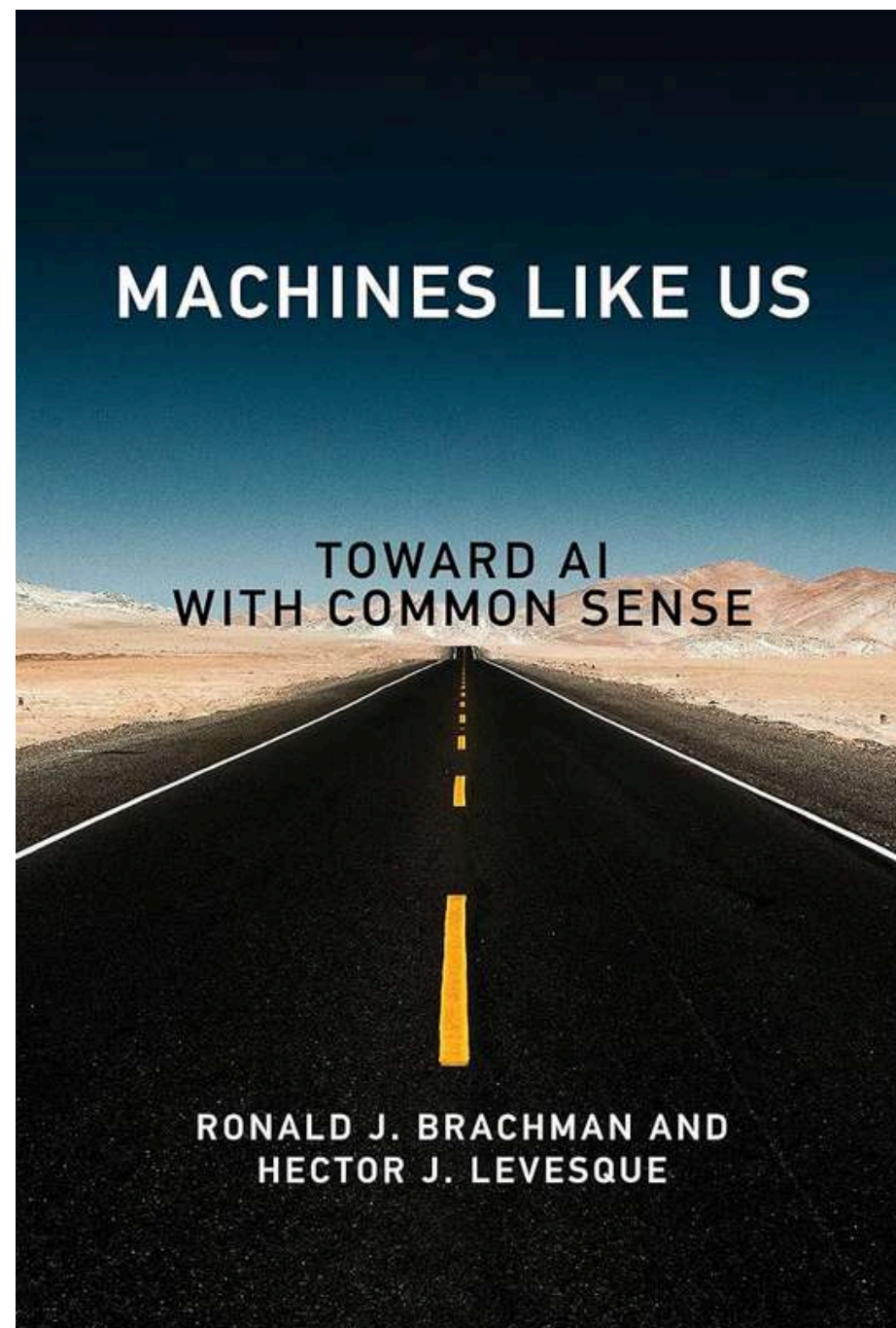
DES CONNAISSANCES

Est-ce qu'on naît avec ?

UNE MANIÈRE DE S'EN SERVIR

Comment s'en sert-on ?

Une définition récente



« *La **capacité d'utiliser** efficacement des **connaissances** ordinaires, quotidiennes et issues de l'expérience, pour atteindre des buts ordinaires, quotidiens et pratiques.* »

— Ronald Brachman & Hector Levesque, *Machines Like Us: Toward AI with Common Sense*, MIT Press (2022), traduction libre

DES CONNAISSANCES

Est-ce qu'on naît avec ?

UNE MANIÈRE DE S'EN SERVIR

Comment s'en sert-on ?

Deux moitiés... et on retrouve exactement les deux questions de Descartes.

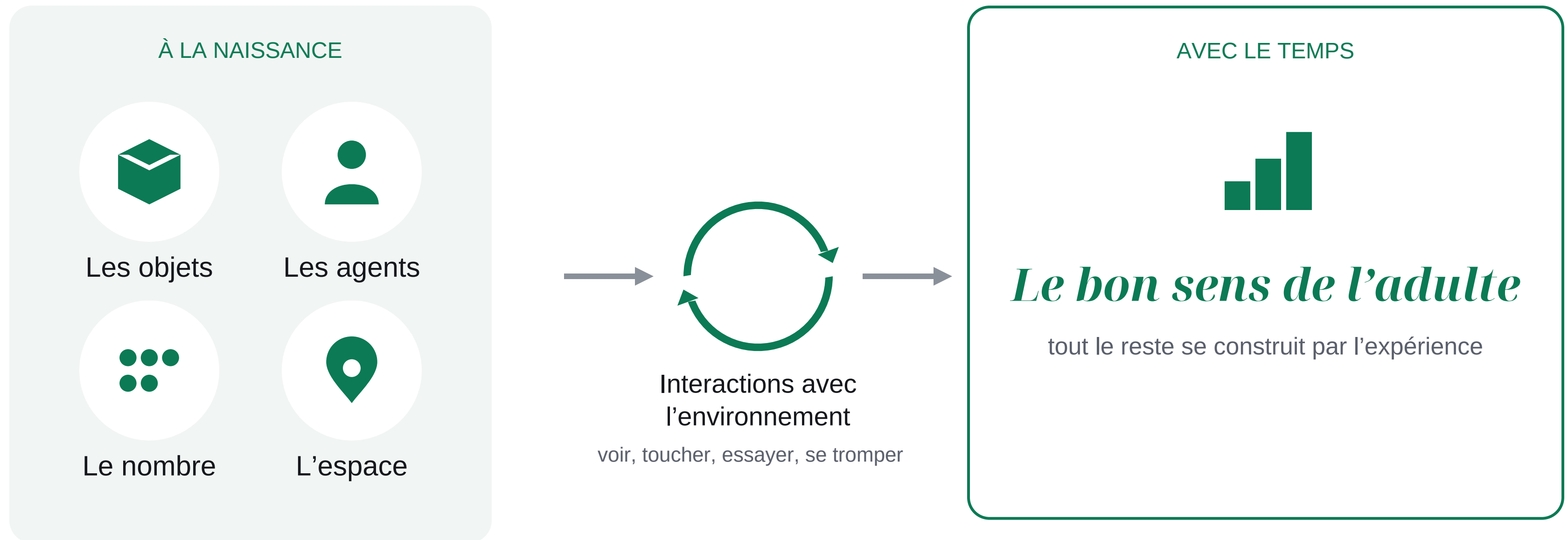
Un kit de démarrage dès la naissance

D'après Elizabeth Spelke, psychologue du développement (Harvard)



Un kit de démarrage dès la naissance

D'après Elizabeth Spelke, psychologue du développement (Harvard)



ni des pages blanches, ni des êtres entièrement programmés

On conclut vite... et on corrige

« Titi est un oiseau »

→ donc Titi vole. On conclut par défaut, sans tout vérifier.



On conclut vite... et on corrige

« Titi est un oiseau »

→ donc Titi vole. On conclut par défaut, sans tout vérifier.



« Titi est un manchot »

→ Titi ne vole pas. On retire la conclusion immédiatement.



On conclut vite... et on corrige

« Titi est un oiseau »

→ donc Titi vole. On conclut par défaut, sans tout vérifier.



« Titi est un manchot »

→ Titi ne vole pas. On retire la conclusion immédiatement.

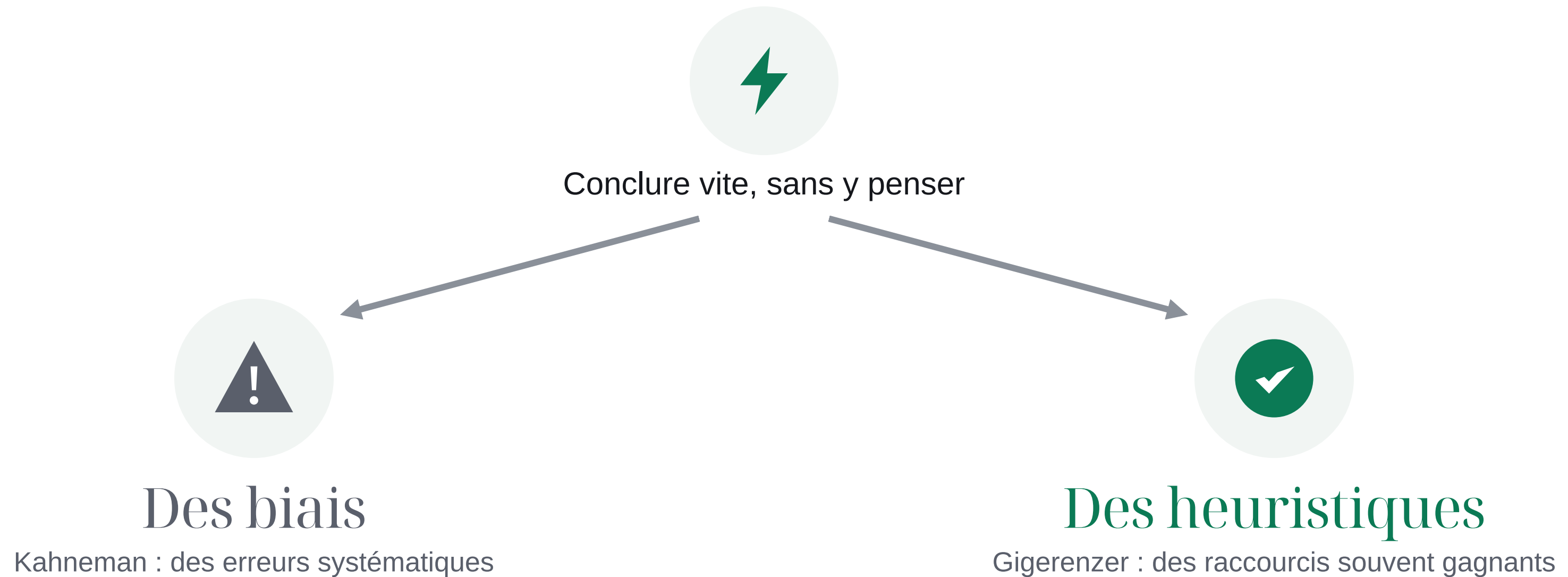


Un raisonnement non monotone : une information nouvelle peut retirer une conclusion.

La logique classique, elle, est monotone : ce qui a été démontré ne peut plus être retiré.

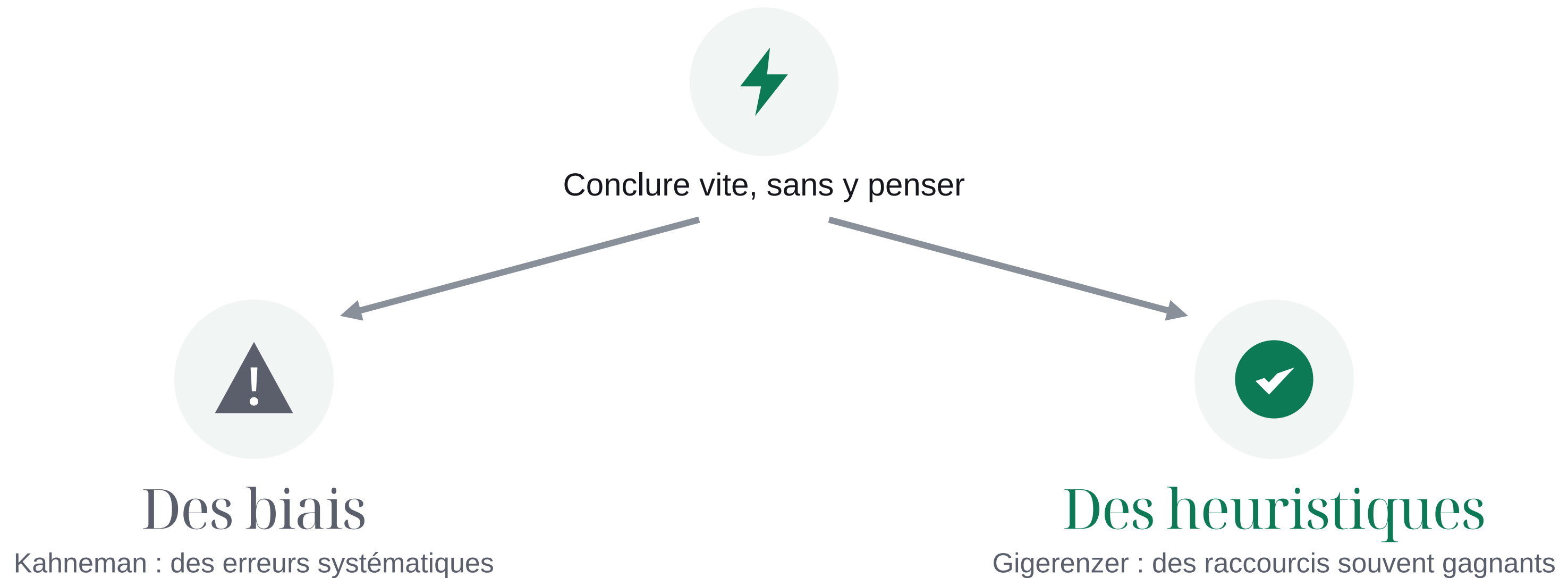
Le bon sens est-il fiable ?

Interroger ses limites, c'est repérer ses points de défaillance.



Le bon sens est-il fiable ?

Interroger ses limites, c'est repérer ses points de défaillance.



Pour une machine : lui transmettre notre bon sens, est-ce lui transmettre nos biais ?

La machine sait... mais sait-elle s'en servir ?

Reprenons nos deux moitiés : savoir, et s'en servir.

Savoir

Les modèles de langage ont lu une bonne partie du web.

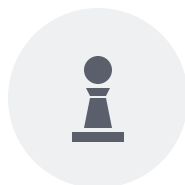
S'en servir

Souvenez-vous du camion : le modèle semblait savoir, sans savoir s'en servir.

LE PARADOXE DE MORAVEC

Ce qui nous semble le plus simple résiste le plus aux machines.

Facile pour la machine



Échecs



Calcul



Examen du
barreau

Difficile pour la machine



Marcher



Saisir un objet



Phrase
ambiguë

Pourquoi est-ce si difficile pour une machine ?

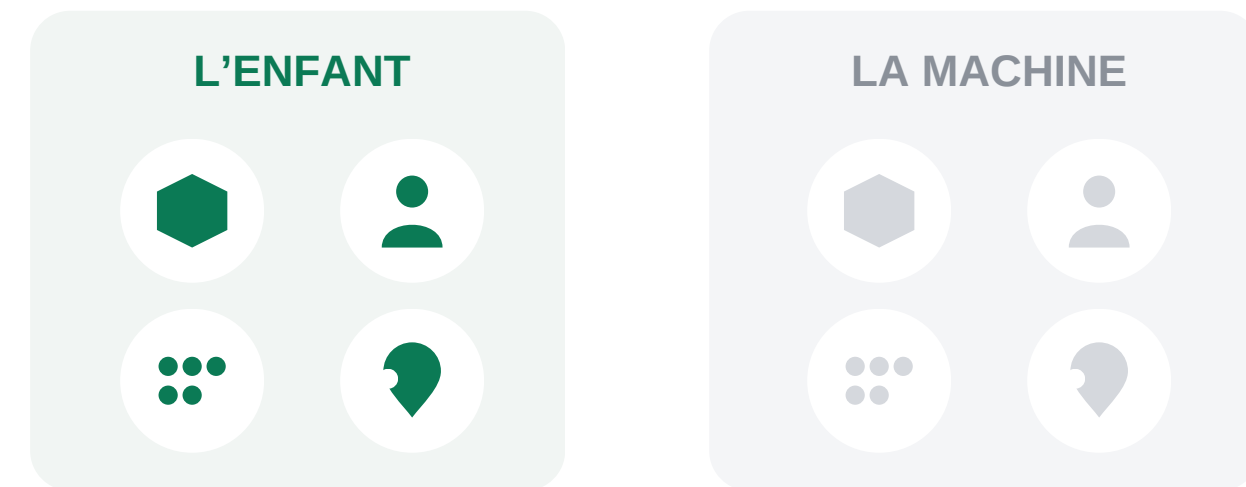
L'évident n'est écrit nulle part

Les textes parlent de l'exceptionnel, pas de l'ordinaire : une machine qui y apprend voit un monde déformé.



Occurrences dans un grand corpus web anglais (Google Web 1T)

Pas de socle



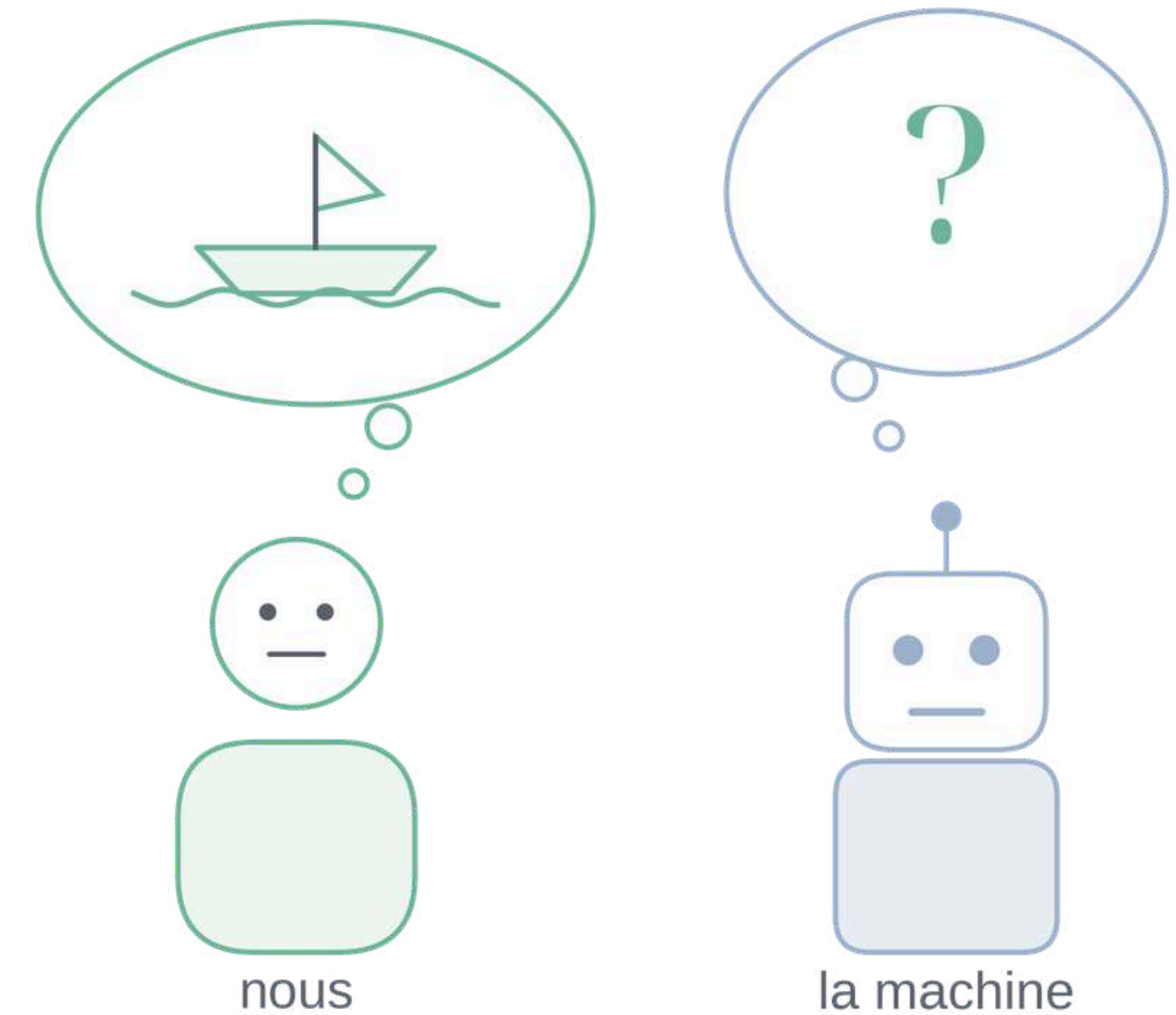
Une machine qui part de zéro peut-elle le reconstruire ?

Comment les chercheurs s'y sont-ils pris ?

PARTIE 2

Les tentatives et les grands courants : Comment a-t-on essayé de donner du bon sens aux machines ?

Écrire le bon sens nous-mêmes, ou laisser la machine l'apprendre ?



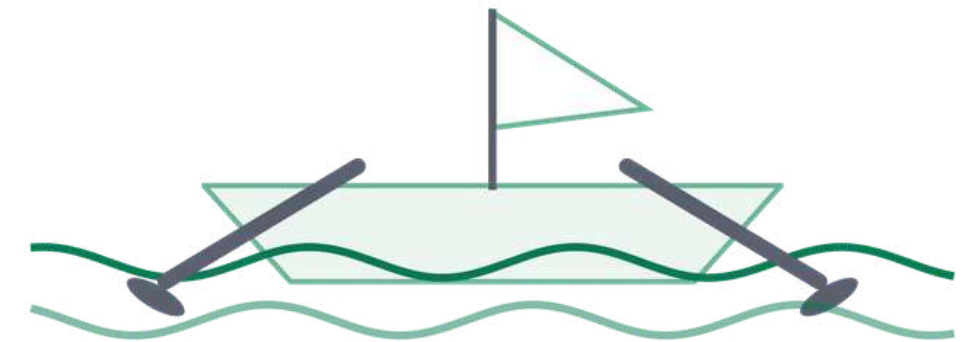
Comment apprendre à une machine ce qu'un enfant de cinq ans sait déjà ?



Un objet peut tomber.



Une personne peut avoir peur.

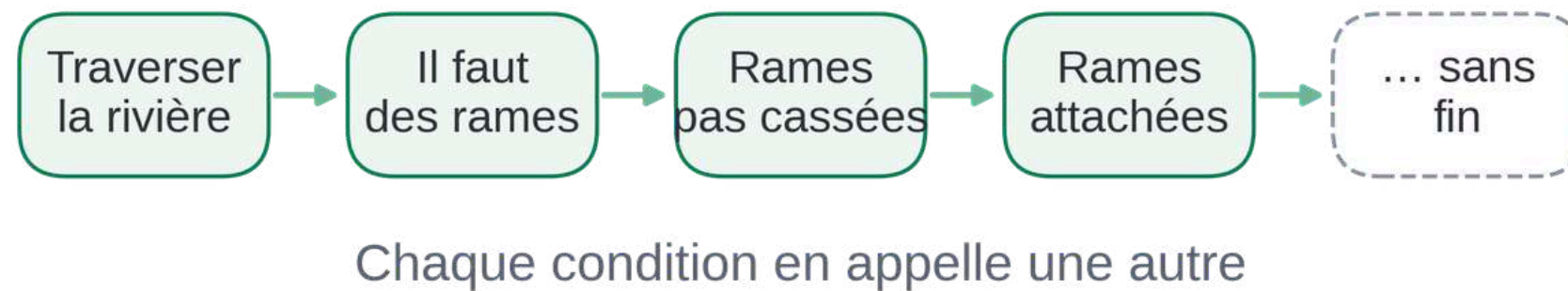


Une barque a besoin de rames pour avancer.

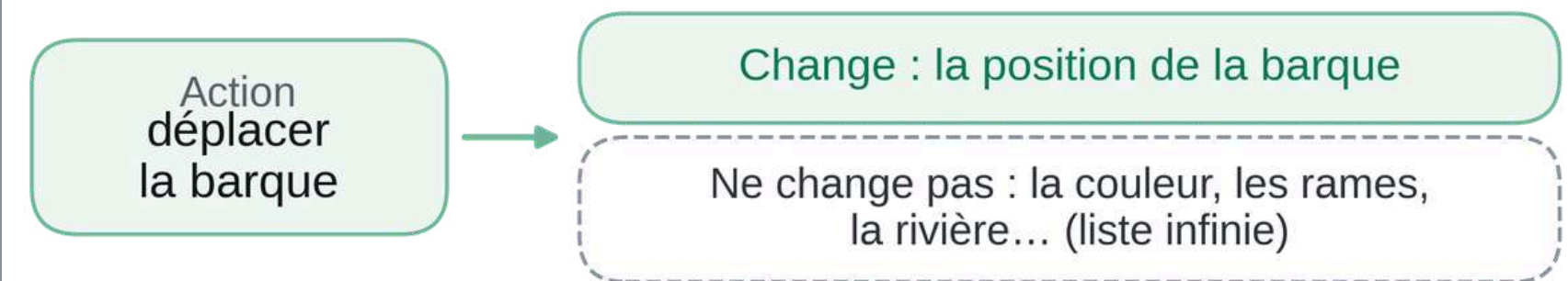
C'est toute la difficulté du bon sens en intelligence artificielle.

1959 : McCarthy imagine un programme qui raisonne

Les exceptions



Le problème du cadre

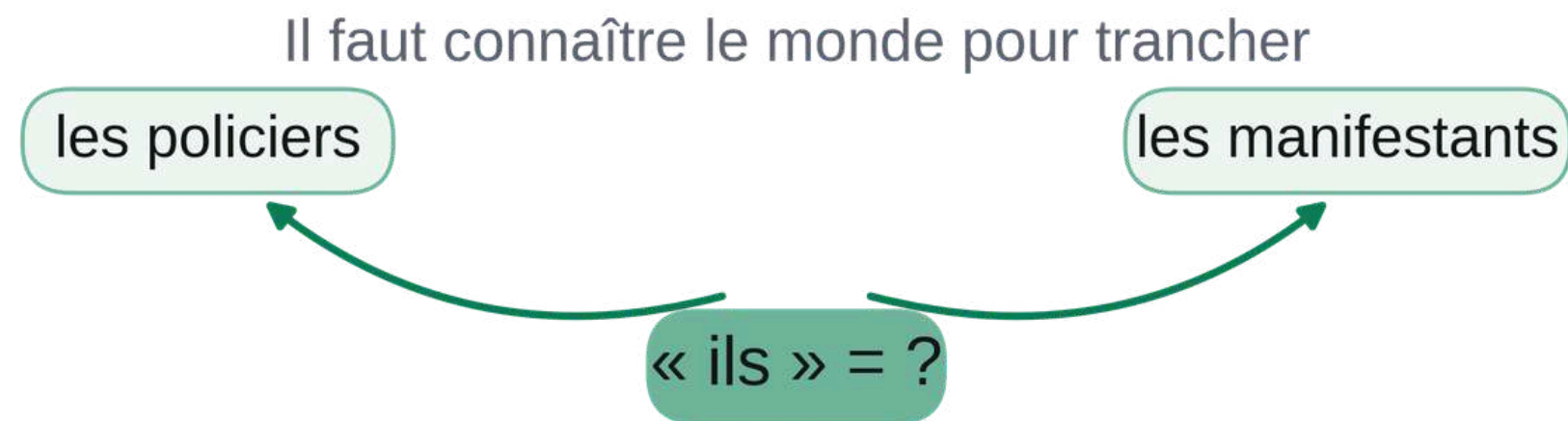


La logique nous dit comment raisonner, mais pas quelles connaissances utiliser.

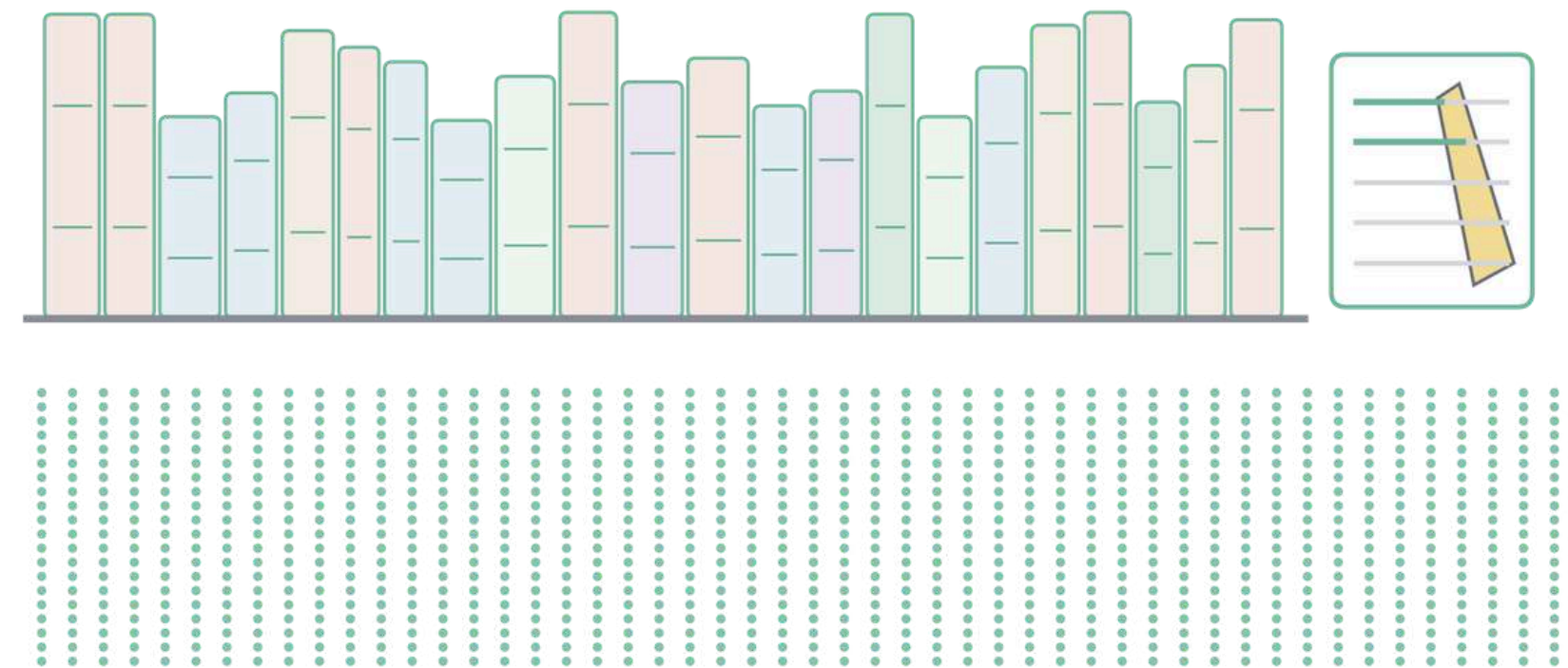
1984 : Lenat lance Cyc, écrire le bon sens à la main

« Les policiers ont arrêté les manifestants parce qu'ils avaient peur de la violence. »

Qui est « ils » ?



1 million de règles annoncées par Lenat.



1 point = 1 000 règles. Tout écrire à la main : énorme travail, et tout n'est pas couvert.

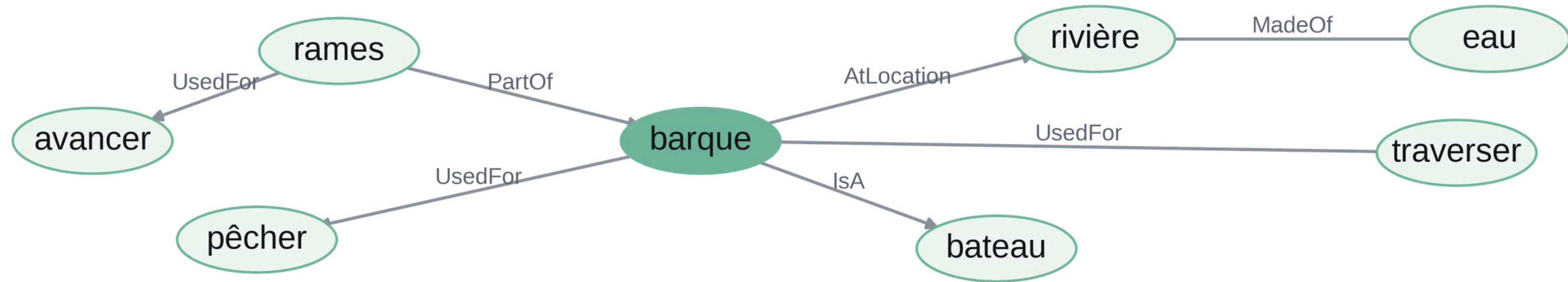
Source : Lenat (1995), « CYC : A Large-Scale Investment in Knowledge Infrastructure », CACM.

Pour réduire le coût : demander à la foule

≈ 14 500
volontaires

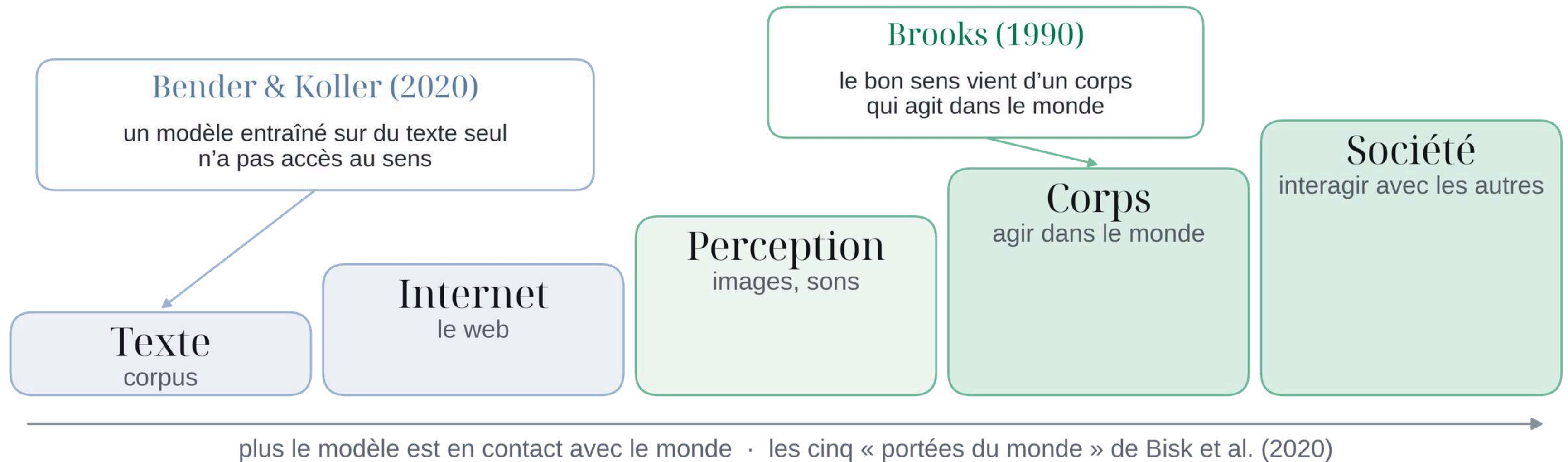
≈ 700 000
phrases écrites

Des millions
de connaissances dans ConceptNet



Beaucoup plus de données, mais moins de contrôle et de précision.

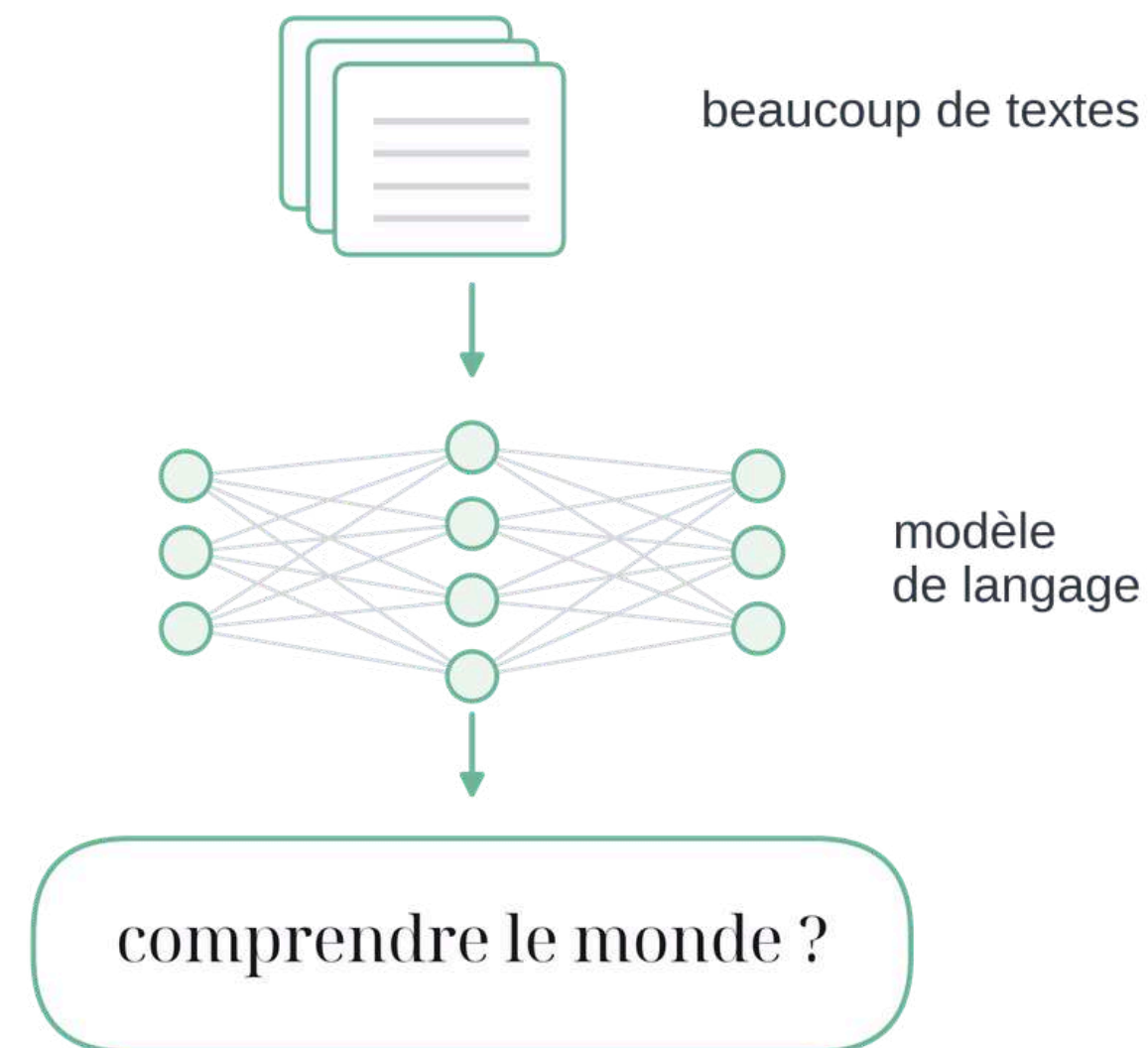
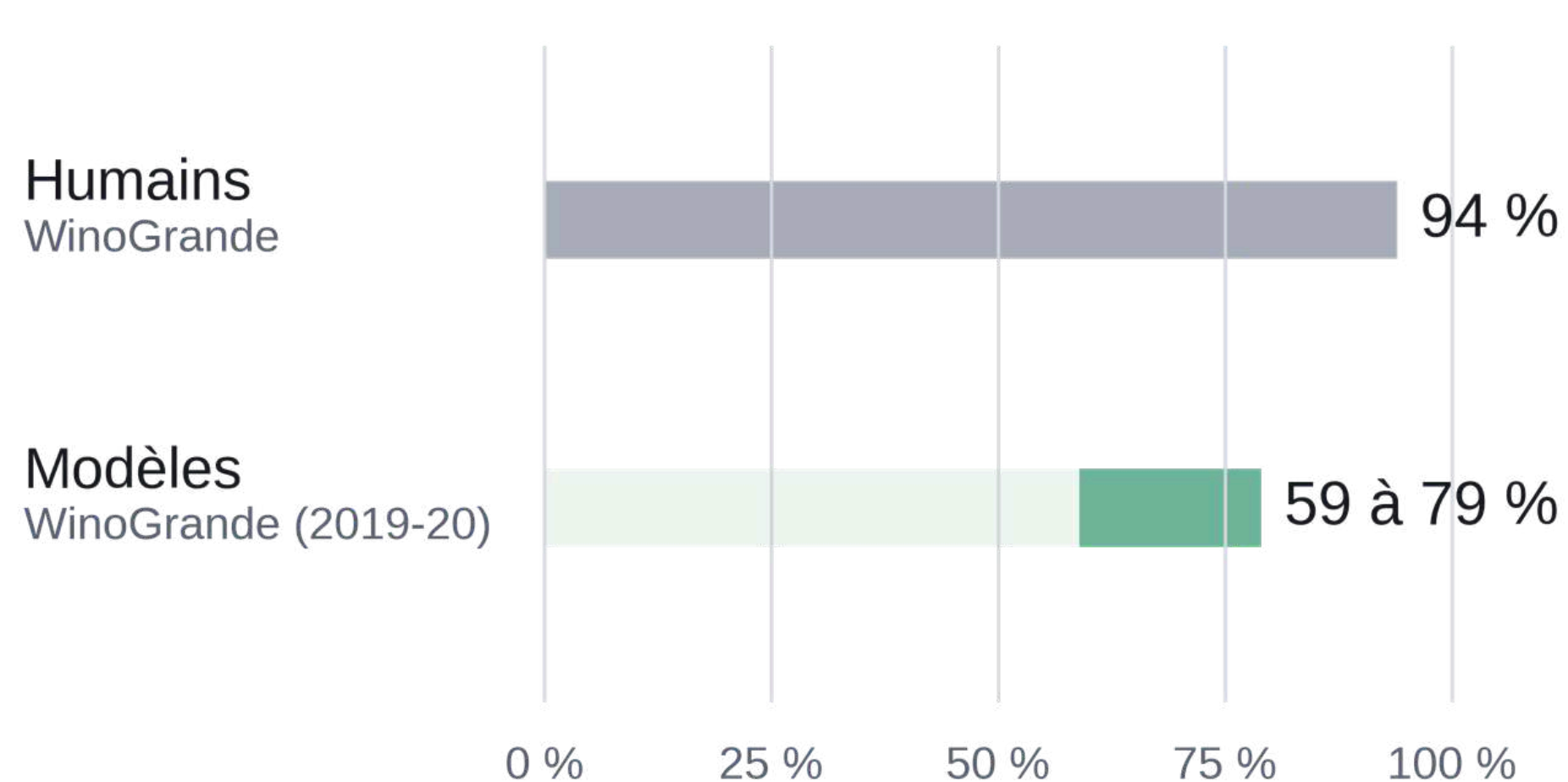
Un autre changement de perspective : le corps



Le lien avec notre cours sur la cognition incarnée.

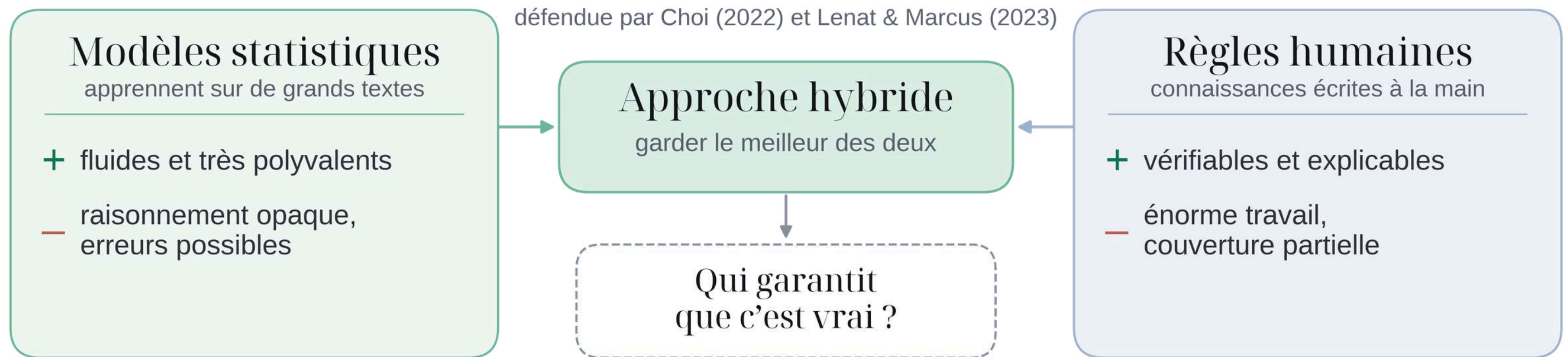
Brooks (1990) · Bisk et al. (2020) · Bender & Koller (2020)

Le tournant statistique



Scores de l'époque : Sakaguchi et al. (2020), WinoGrande.

Combiner les approches



Chaque approche a apporté quelque chose



Mais aucune ne semble suffisante à elle seule.

Les modèles ont d'excellents scores. Ont-ils vraiment du bon sens ?

Et comment le savoir ?

PARTIE 3

Comprendre ou imiter :
comment savoir si une IA
a du bon sens ?

Les différents types de benchmarks

TEXTE

Example 1: A man is sitting on a roof. He

Continuation:

A. starts pulling up roofing on a roof. *

B. is using wrap to wrap a pair of skis.

C. is ripping level tiles off.

D. is holding a Rubik's cube.

Les différents types de benchmarks

IMAGE



 Q1: Are the kids about the same age?

 A1: No, Conf: 58.5%

 Q2: How many kids are there?

 A2: 3, Conf: 95.0%

 Q3: Is the kid in man's arm youngest?

 A3: No, Conf: 68.4%

 A3: Yes



 Q1: How many horses are there?

 A1: 2, Conf: 100.0%

 Q2: Is the white horse on the left?

 A2: No, Conf: 100.0%

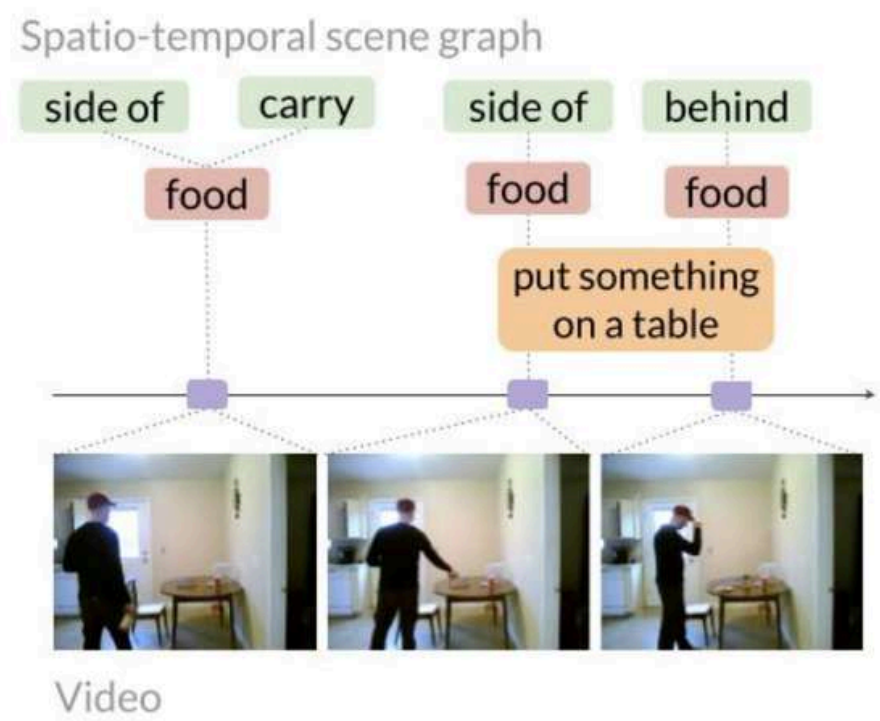
 A2: Yes

VQA benchmark

Les différents types de benchmarks

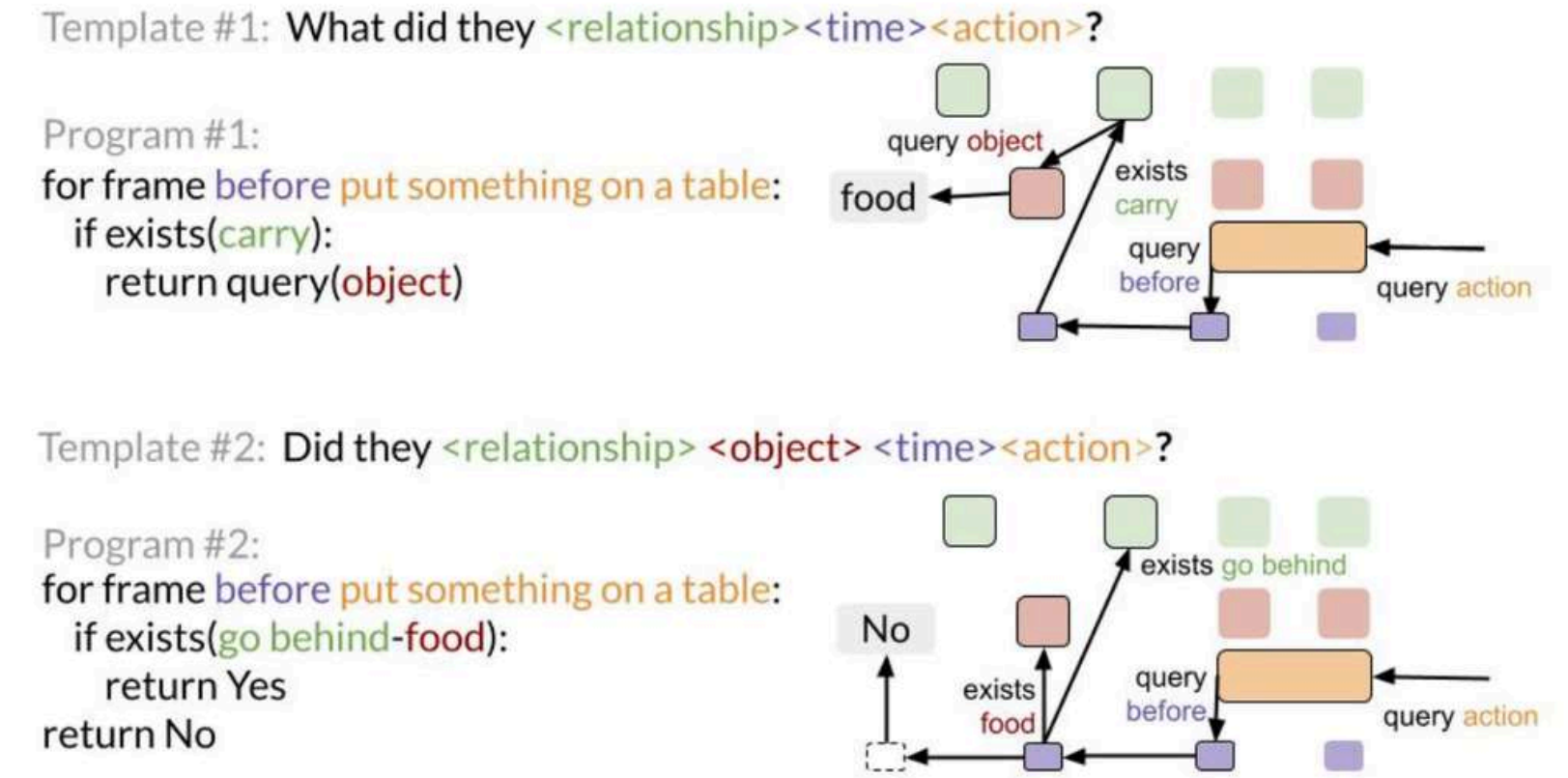
VIDEO

Input



Legend: ■ objects ■ relationships ■ actions ■ time

Our benchmark generation process



Output

Question answer #1:
What did they carry before putting something on a table? - food

Indirect action reference for #1:
What did they carry before the shortest action? - food

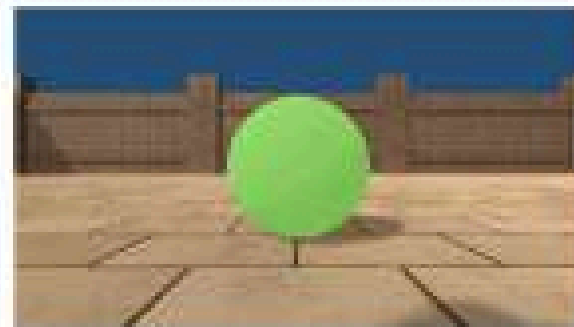
Question answer #2:
Did they go behind some food before putting something on a table? - No

Indirect object reference for #2:
Did they go behind the object they carried before putting something on a table? - No

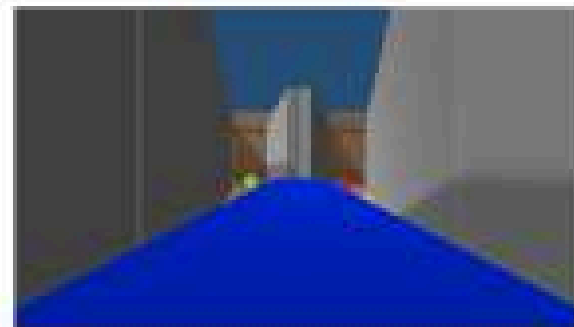
Visual Genome

Les différents types de benchmarks

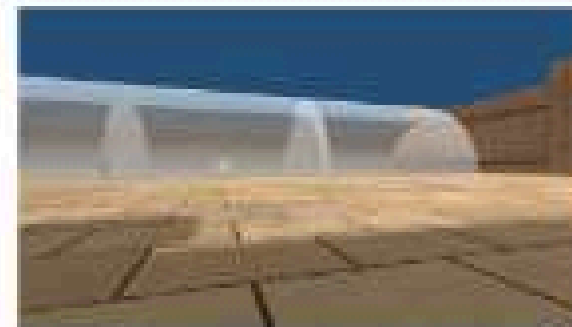
MODELE 3D



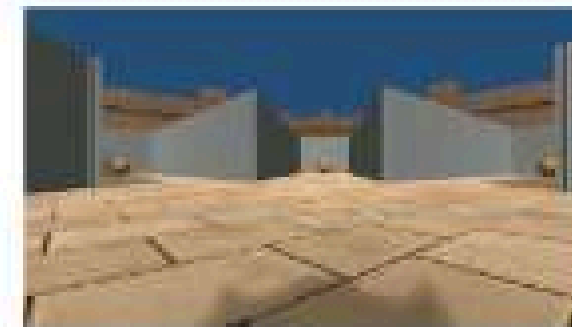
(a) Intro-162



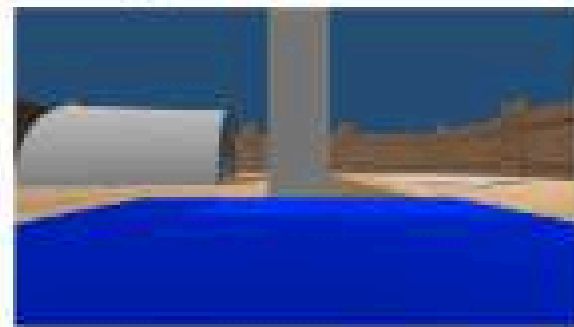
(b) Y-Maze-69



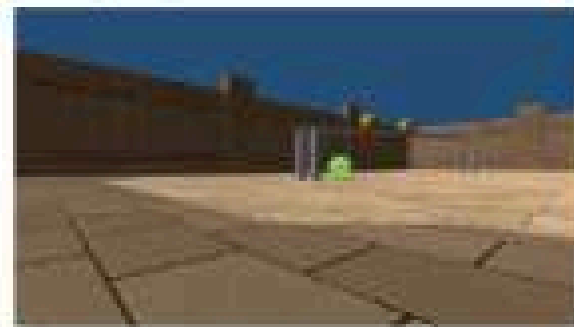
(c) Detour Task-45



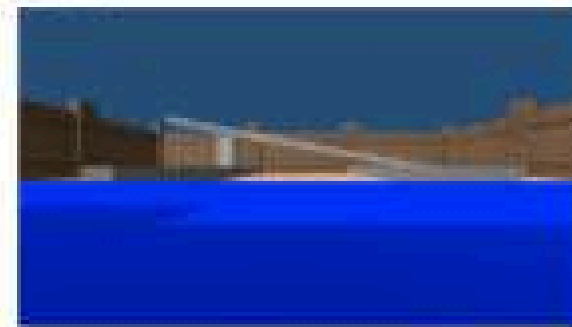
(d) Radial Maze-54



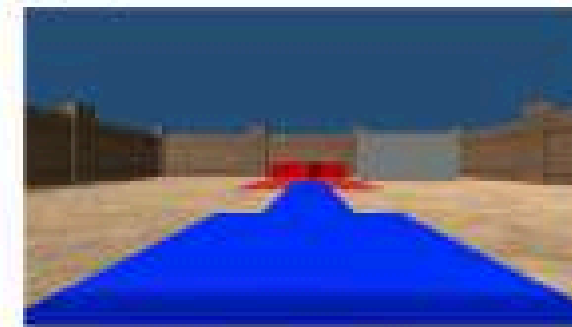
(e) Spatial Elim-27



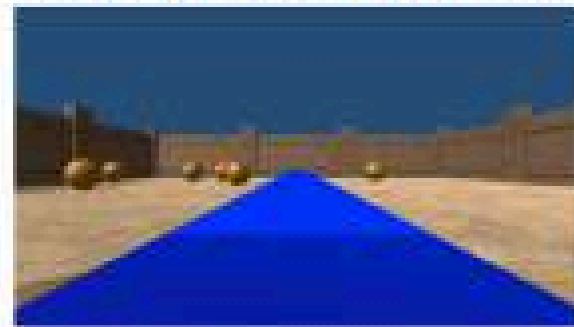
(f) Delayed-30



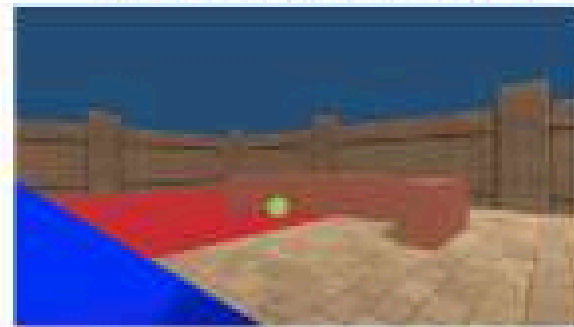
(g) Support-27



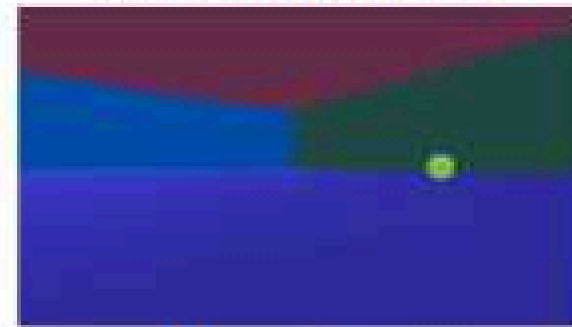
(h) Obj. Perm.-90



(i) Numerosity-90



(j) Tool Use-126



(k) Weak Gen.-90



(l) Internal Model-90

PARTIE 4

Perspectives et ouverture :
Une machine pourra-t-elle un
jour avoir du bon sens ?

Imiter n'est pas comprendre

01 ILLUSION

Clever Hans

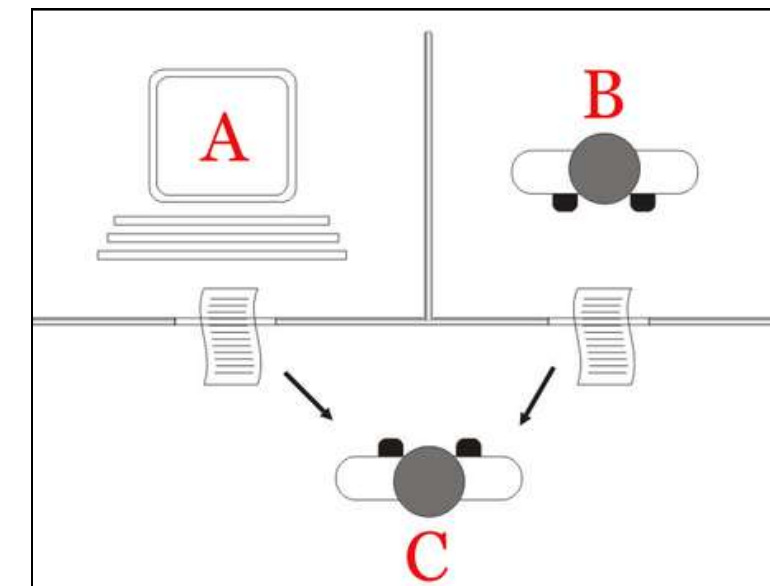
Savoir n'est pas agir



02 FRAGILITÉ

Imitation Game & LLM

Statistiques pures



Une machine pourra-t-elle un jour l'acquérir ?
Ou restera-t-elle un **solveur inexperimenté** ?

Quatre pistes pour demain

IA Incarnée

Apprentissage profond

+ Robotique

+ Environnement simulé complexe

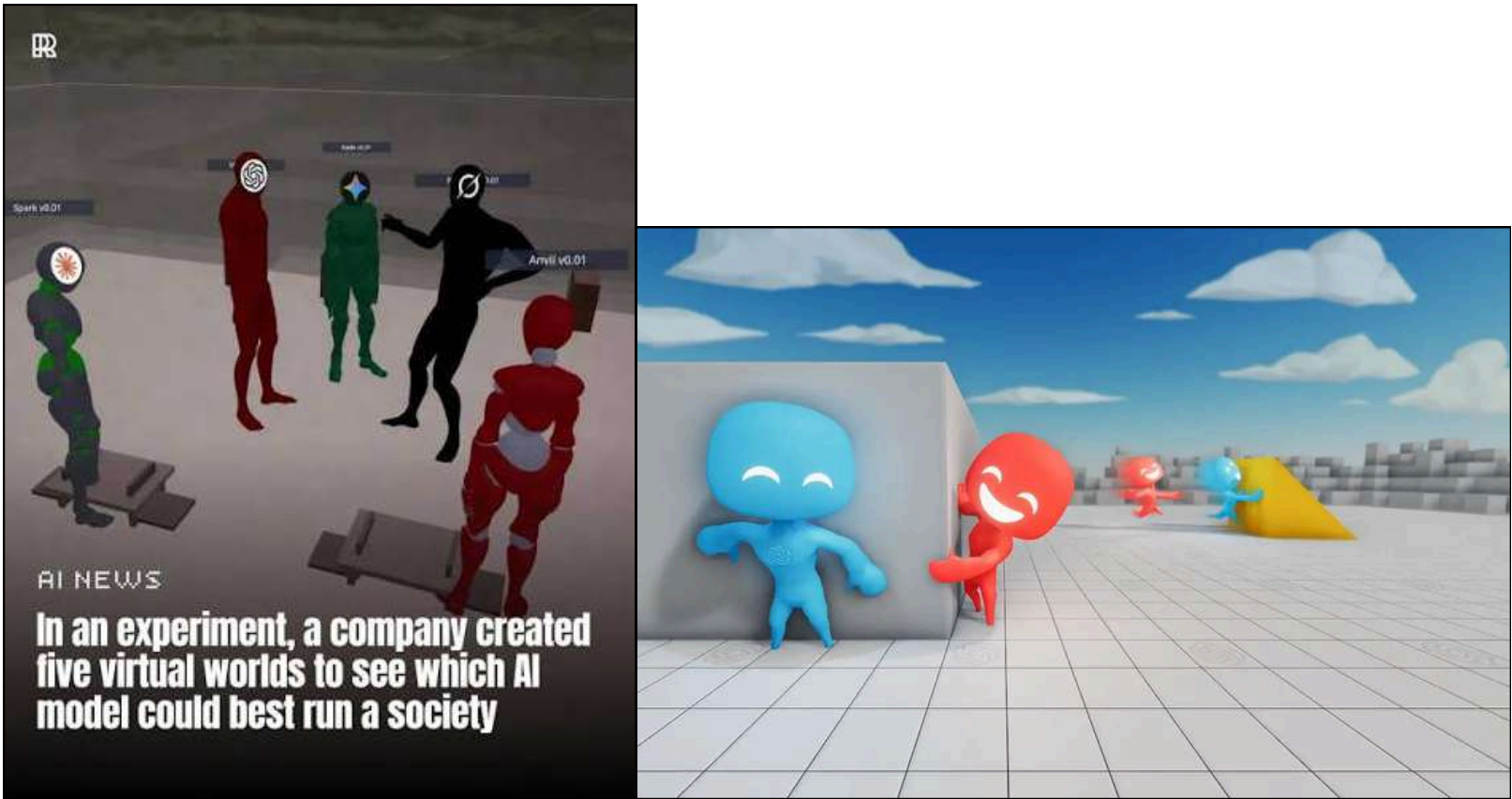


- Interaction physique avec un environnement dynamique
- Modélisation par abstraction → systèmes fragiles

Quatre pistes pour demain

Systèmes Multi-Agents

Bon sens
Consensus sociétal en perpétuelle évolution



Dataset	Domain	GPT-3	Instruct GPT	ChatGPT	Human
CommonsenseQA	General	38.0	81.0	74.0	88.9
OpenBookQA	General	22.0	65.0	73.0	89.3
WSC	General	46.0	78.0	78.0	92.1
PIQA	Physical	48.0	77.0	78.0	94.5
Social IQA	Social	36.0	71.0	62.0	86.9
ARC	Science	27.0	88.0	94.0	—
QASC	Science	25.0	75.0	74.0	93.0
HellaSWAG	Event	19.0	61.0	67.0	95.7
NumerSense	Numerical	45.0	63.0	79.0	89.7
ProtoQA	Prototypical	67.3	84.6	94.2	—
MC-TACO	Temporal	20.0	53.0	52.0	75.8

- SMA
- Modèle AEIO
- simuler emergence de normes sociales

→ jeux de données liés aux interactions sociales ou au temps

ChatGPT is a Knowledgeable but Inexperienced Solver: An Investigation of Commonsense Problem in Large Language Models, Bian et al. (2024).

Quatre pistes pour demain

IA Neuro-Symbolique

TOUT-SYMBOLIQUE

Rigide

TOUT-STATISTIQUE

Sans structure

PISTE · NEURO-SYMBOLIQUE

Relier **langage naturel**, raisonnement
génératif et flexibilité.

COMET : Commonsense Transformers for
Automatic Knowledge Graph Construction

Bosselut, A., Rashkin, H., Sap, M., Malaviya, C., Celikyilmaz, A., & Choi, Y. (2019).

Quatre pistes pour demain

IA Développementale

Machine enfant

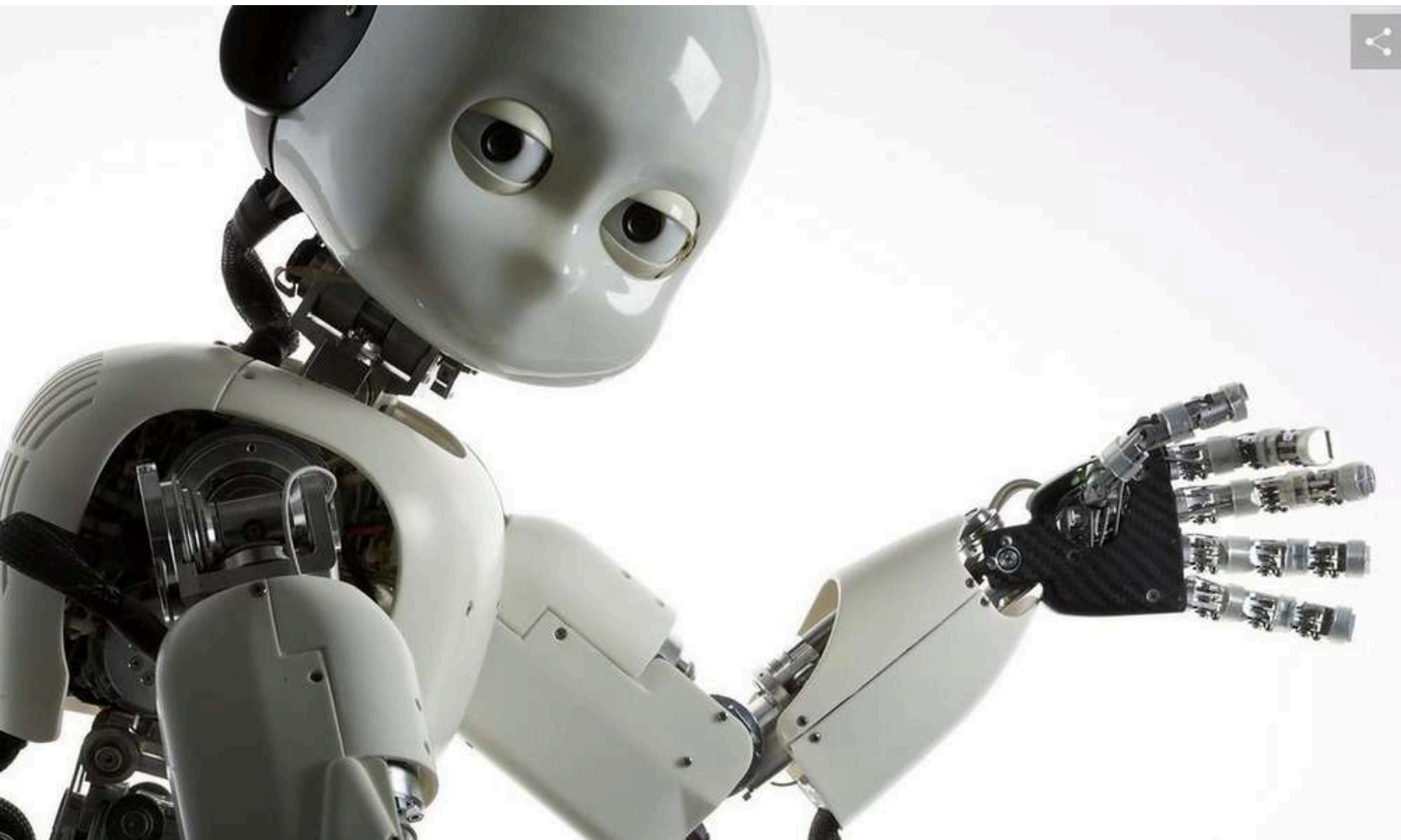
Idée de Turing

Motivation intrasèque

Curiosité

LLM

“Voix interieur”



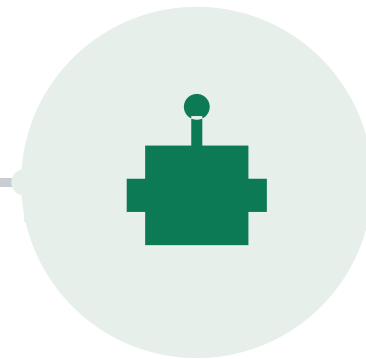
- Génère ses propres buts
 - Explore son environnement
- Processus d'éducation continu, interactif et autonome face au monde réel

Recapitulons



PARTIE 1

Un socle dès la naissance, et une manière souple de s'en servir



PARTIE 2

Écrit à la main, puis appris à partir des textes



PARTIE 3

On ne sait pas encore mesurer s'ils comprennent ou s'ils imitent



PARTIE 4

Des pistes encore ouvertes : corps, interactions, hybrides, curiosité

A-t-on besoin qu'une IA ait du bon sens pour lui faire confiance ?

« Le principal est de l'appliquer bien. » — Descartes, Discours de la méthode (1637)

EMERGENCE WORLD

THE EXPERIMENT · SEASONS 1 & 2

Made with  vizard agent



Quatre axes pour débattre



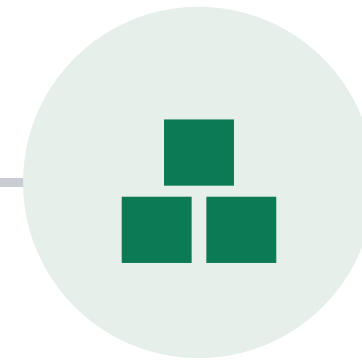
imiter ou comprendre ?

Si une IA répond toujours juste à la phrase du camion, comprend-elle ?



le « sensé » de qui ?

Le bon sens est-il universel, ou un ensemble de normes culturelles ?



comment le construire ?

Quelques règles fiables, ou des milliards de données ?



la confiance

Si l'humain se trompe aussi, pourquoi exiger plus de la machine ?

Si vous deviez confier une tâche à l'un de ces agents pendant quinze jours, sur quoi vous fonderiez-vous ?