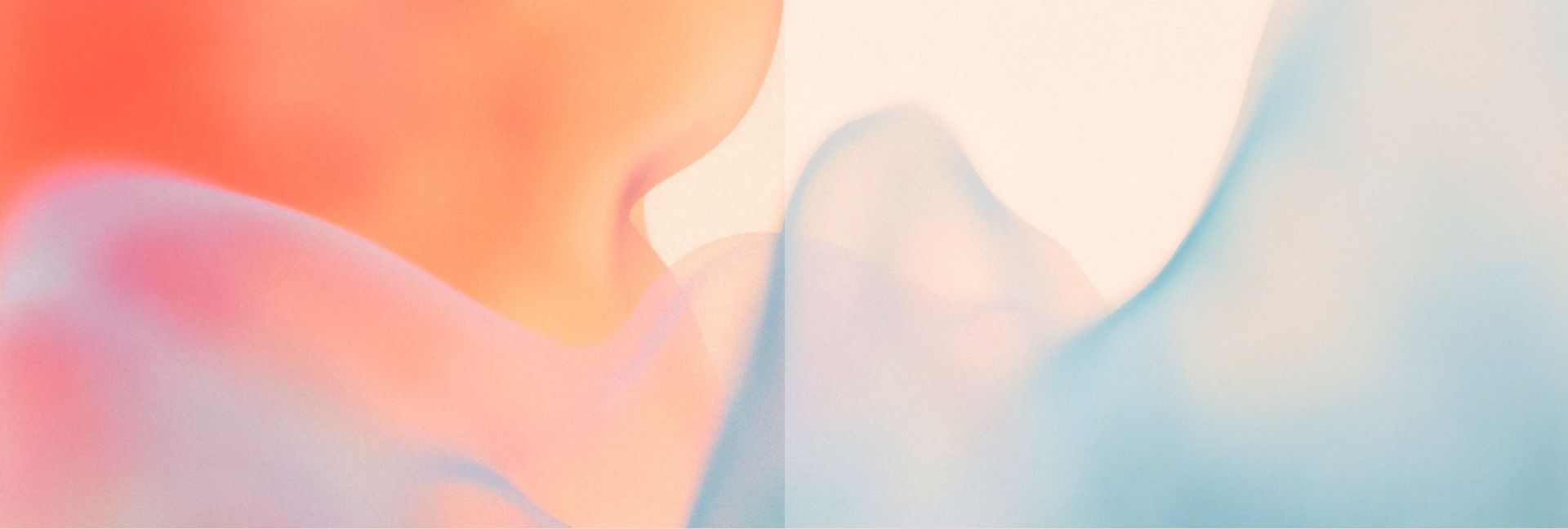




The Frame Problem



00 Introduction

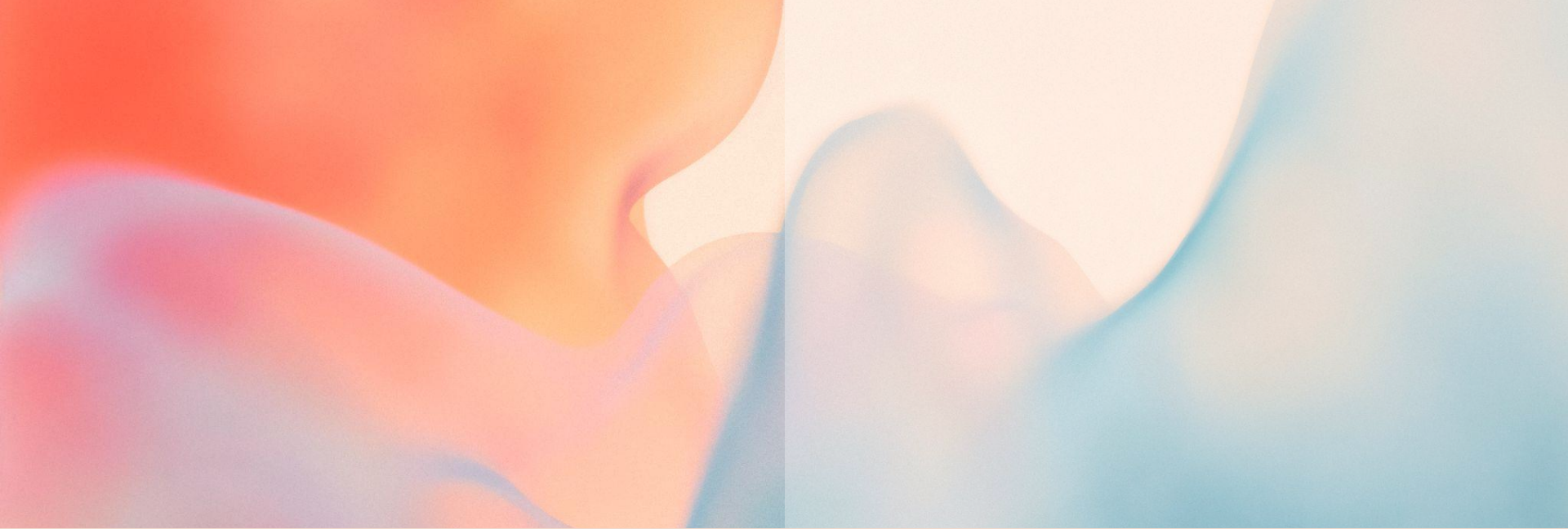
- Le Frame Problem = difficulté à représenter le changement sans tout redéfinir.
- Apparaît dans les années 60, IA symbolique.
- Toujours pertinent avec l'essor des IA modernes.
- Nous allons parler de :

1 - Origine et Fondements

2 - Enjeux pratiques et solution

3 - Implications philosophiques & IA moderne





01

Origine et Fondements



Un peu d'histoire

Contexte Historique

Années 1960

Émergence de l'**IA symbolique**. Logique du prédicat, de l'inférence. Recherches qui aspirent à "l'intelligence logique" (raisonnement déductif, planification)

Objectif

Construire des systèmes capables de modéliser les **effets de leurs actions** dans des environnements dynamiques

1969

McCarthy & Hayes → formalisent Frame Problem dans le contexte de la logique du premier ordre.

"Some philosophical problems from the standpoint of artificial intelligence."

Le problème posé

Ce qui change

Modéliser les **effets directs** de chaque action effectuée par l'Agent.

Ce qui ne change pas

Représenter formellement que tout le reste reste inchangé sans énumération explicite : **inertie commonsensique**
Notion de “bon sens”, qui nécessite l’ajout d’axiomes en logique.



Exemple : le robot de Denett

01

Situation initiale :

Robot dans une pièce avec une **batterie** à récupérer et une **bombe à retardement**

02

Action souhaitée :

Prendre la batterie sans déclencher la bombe

03

Défi logique :

Raisonner sur ce qui doit changer (batterie) et sur **ce qui ne doit pas changer** (déclenchement bombe)

04

Problème technique :

Filtrage contextuel de la pertinence : il y a **trop d'axiomes** à gérer (R1D1), et les examiner et les trier est **trop long** (R2D1)



Les “Frame Axioms”

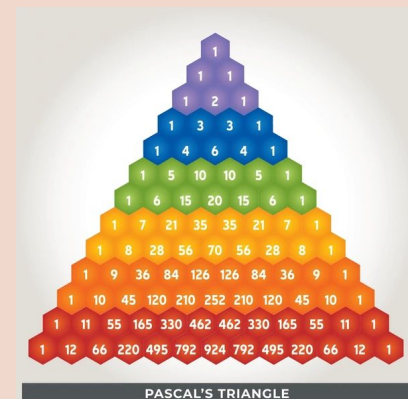
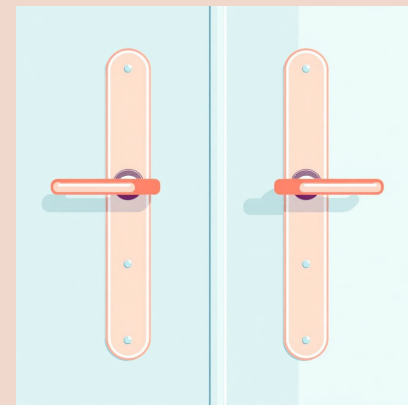
Origine du nom : “The frame problem :

Les “Frame Axioms” sont des règles logiques stipulant les propriétés du monde (symbolique)

Restent constantes après une action, sauf indication contraire

Le problème :

Dans monde complexe (comme le monde réel), l'approche devient **ingérable** : nombre d'axiomes explose, et problème de **qualification et de ramifications**



Exemple logique



Monde simple : block world

Prédicats : $\text{On}(x, y, t)$ signifie
"x est sur y au temps t"

Action

"Déplacer bloc x de y vers z"
 $\text{Move}(x, y, z)$

Conséquences

$\neg \text{On}(x, y, t+1), \text{On}(x, z, t+1)$

Mais quid de $\text{On}(a, b, t+1)$? Ou de $\text{Color}(x, \text{red}, t)$?

Pertinence contextuelle

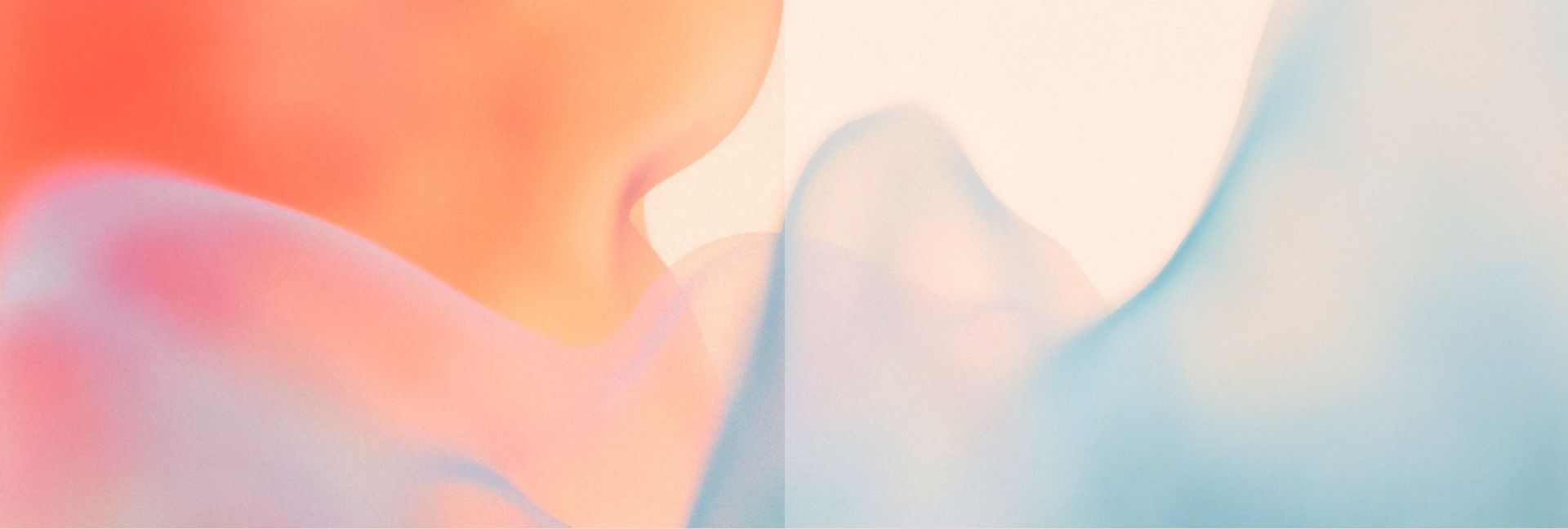
Qu'est-ce qui est véritablement important dans une situation donnée ?

Focalisation de l'attention

Comment déterminer où porter son attention de manière efficace ?

Scalabilité des systèmes symboliques

Comment gérer la complexité croissante des **systèmes symboliques** ?



02

Enjeux Pratiques et Solutions



Enjeux Pratiques et Solutions

1 - Applications

2 - Explosion combinatoire

3 - Solutions

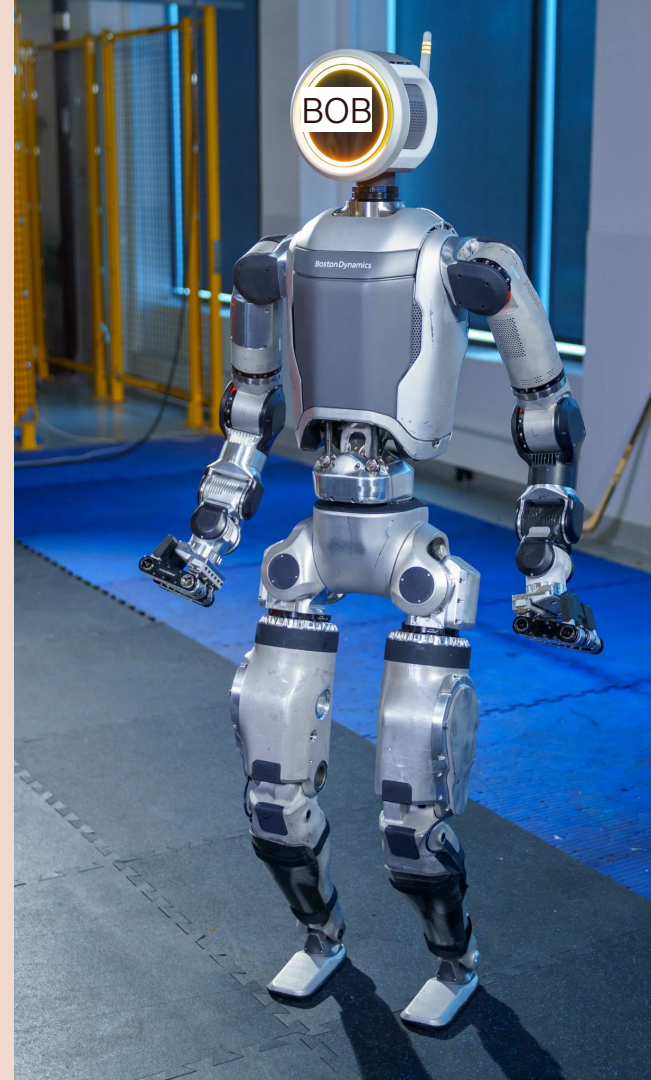
4 - Conclusion



Enjeux Pratiques et Solutions

1 - Applications

- Agents Intelligents
- Tâches de planification
- Systèmes experts



Enjeux Pratiques et Solutions

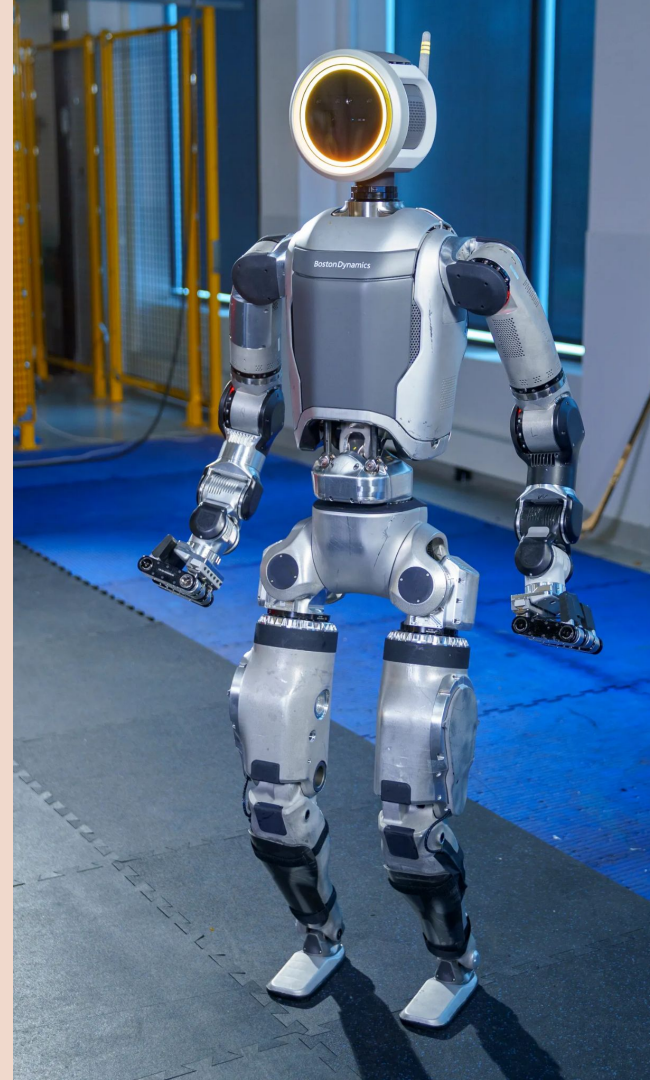
2 - Explosion combinatoire

- Ambition de créer des Agents autonomes **complexes**
- Quête de l'intelligence **humaine**

N objets avec M propriétés $\rightarrow N \times M$ actions possibles.

Combien d'objets avec combien de propriétés pour chaque objet du monde pouvant être modélisés ?

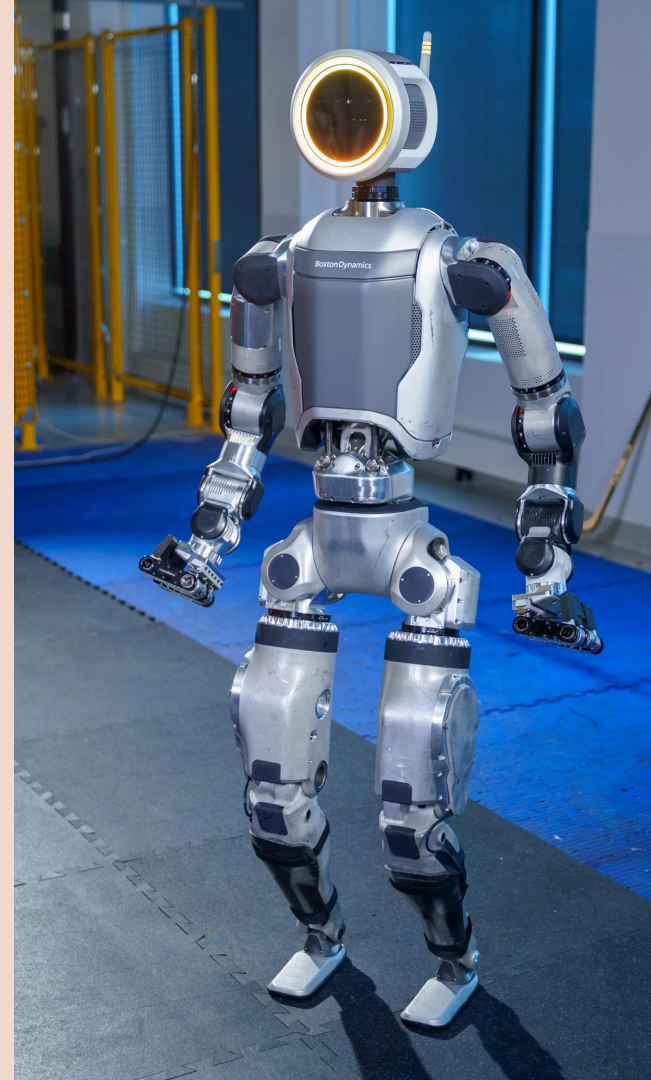
vous avez 4h



Enjeux Pratiques et Solutions

3 - Solutions

- **formalismes logiques** (e.g. Fluent Calculus, circumscription),
- **solutions computationnelles** plus modernes (e.g. clustering, systèmes multi-agents).
- **modèles cognitifs** (e.g. raisonnement défaisable),



Enjeux Pratiques et Solutions

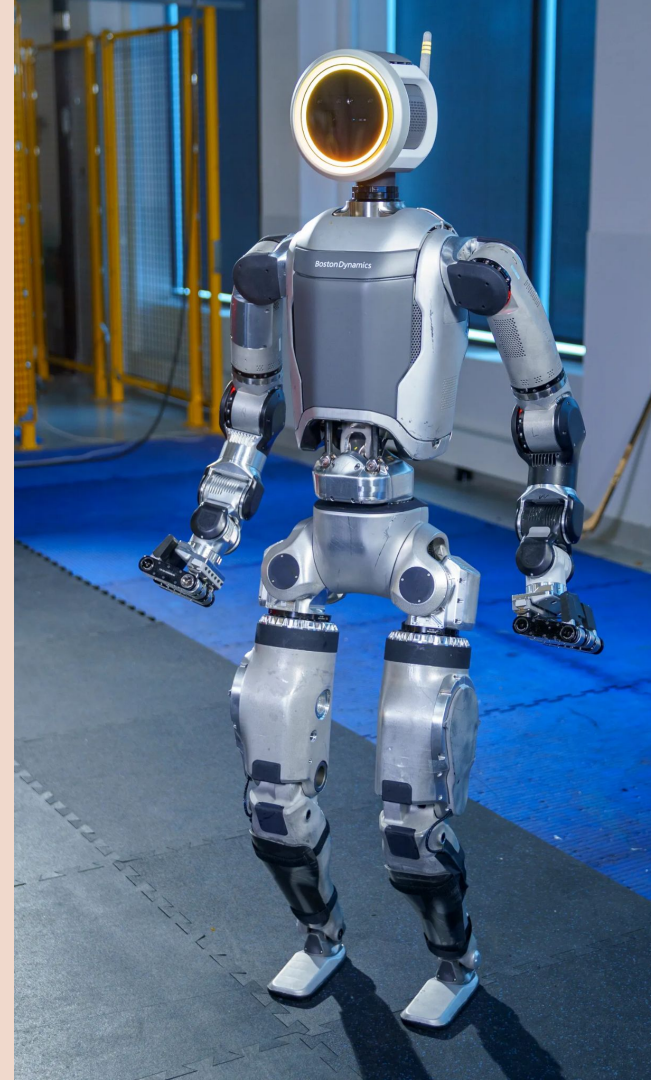
3.1 Formalismes logiques

Axiomes de mise a jour d'état,

- Specifier explicitement les limites de chaque actions.

The concurrent, continuous Fluent Calculus, (Thielscher, 2001)

- Ajoute certains fluents (ce qui change),
- Supprime certains fluents (ce qui n'est plus vrai),
- Et tout le reste reste implicitement inchangé.



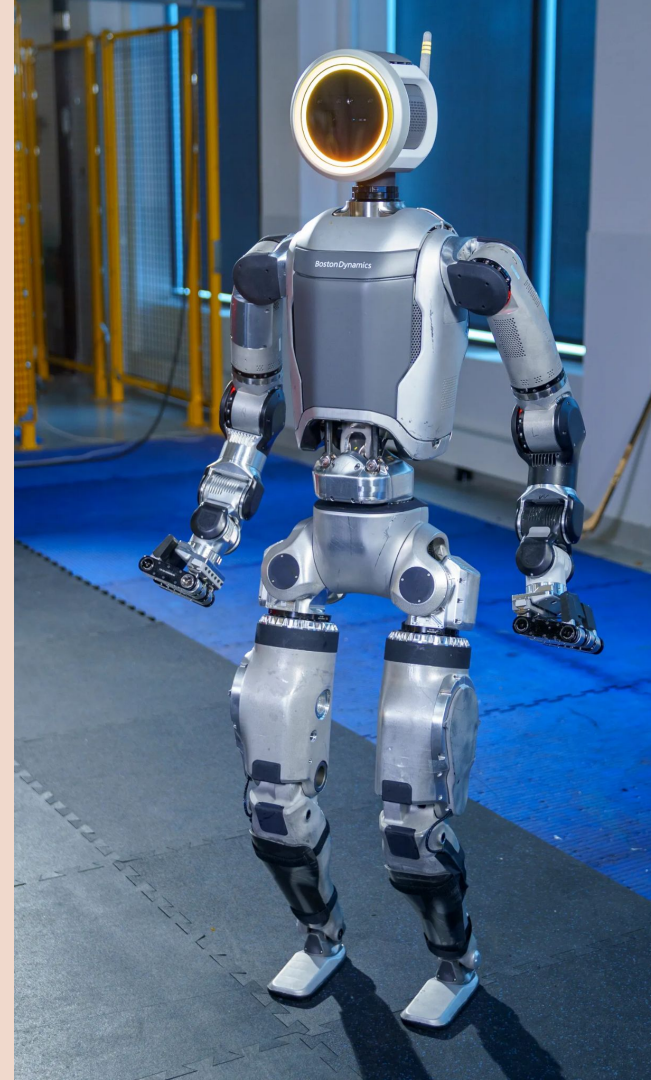
Enjeux Pratiques et Solutions

3.2 Solutions Computationnelles

The Generalist Approach to Frame Problems ([Hidaka & Kashyap, 2014](#))

Clustering dimensionnel :

- regroupement des inputs par clustering
- Fonctionne sur des données diverses

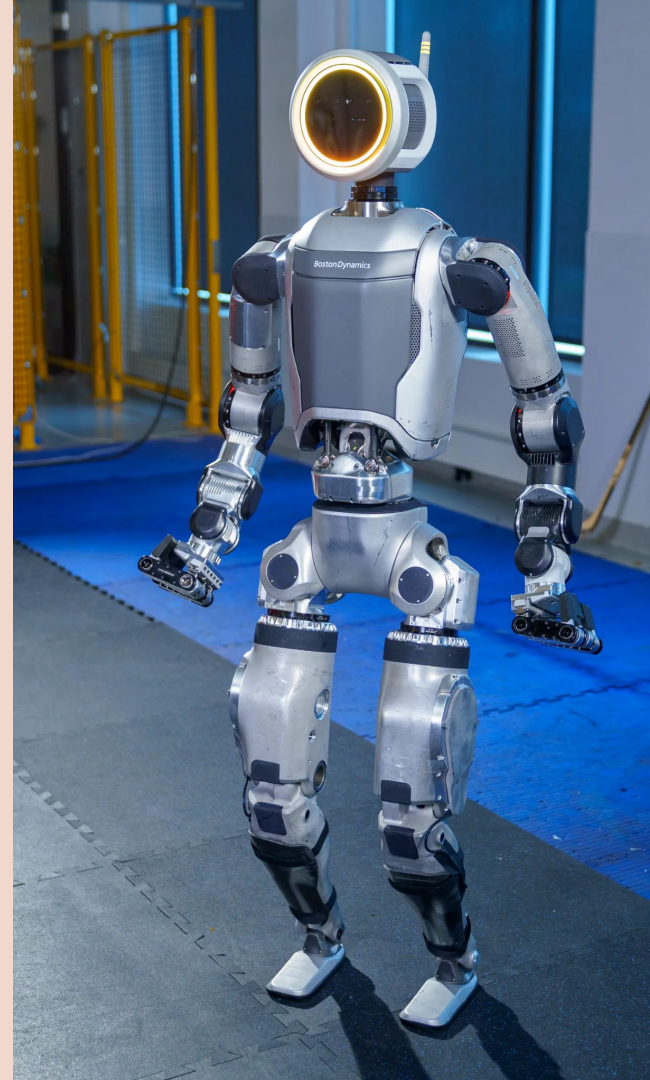


Enjeux Pratiques et Solutions

3.3 Modèles cognitifs

Pollock (1997), Reasoning about change and persistence.

- **Fiabilité de la perception**
- **Projection temporelle**
- **Implication causale**

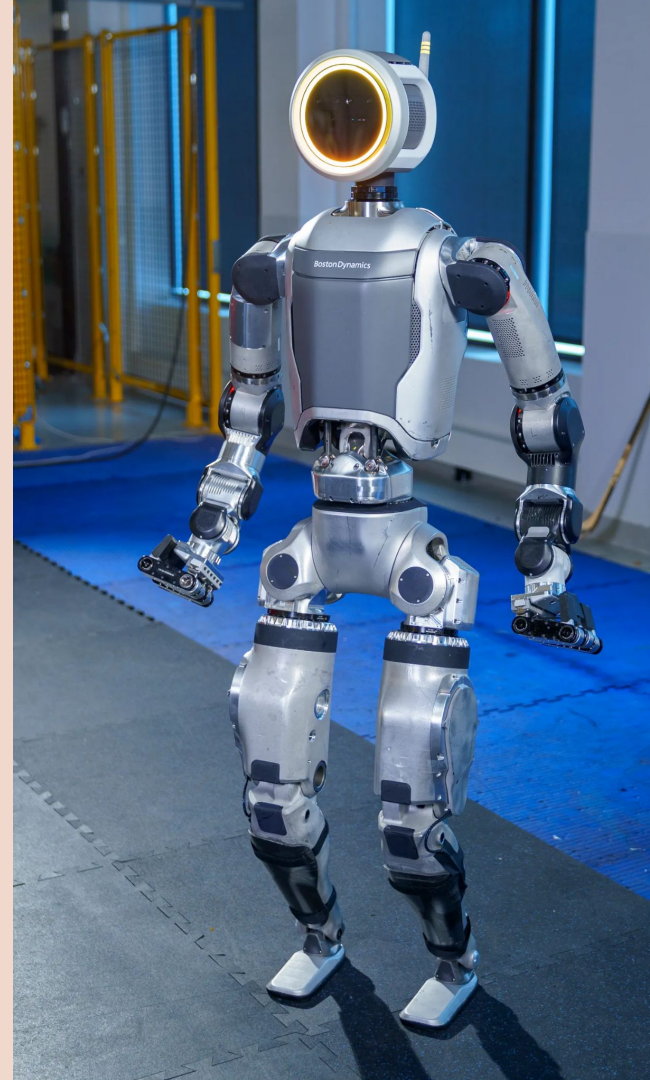


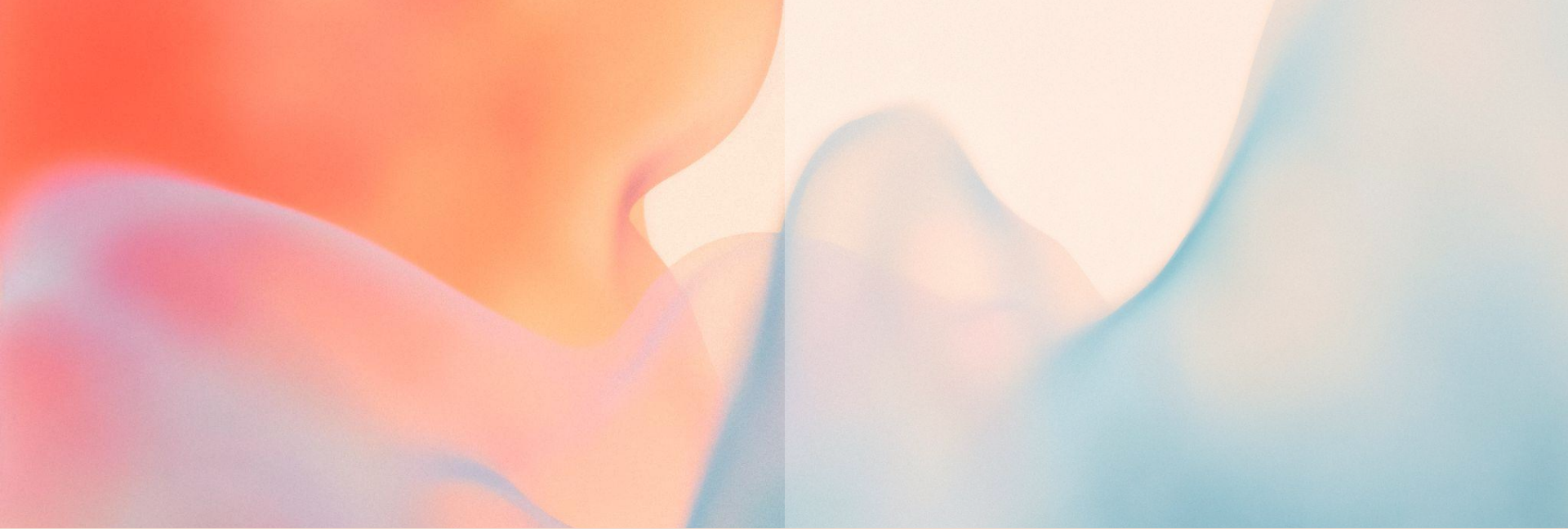
Enjeux Pratiques et Solutions

4 - Conclusion

- **Très** difficile de modéliser le monde.
 - Chaos
 - Effet de bords
- Incompréhension du fonctionnement humain :
 - Raisonnement non logique ?
 - Biais

<< Le frame problem cessera d'être un problème pour l'IA dès qu'on arrêtera d'attendre d'elle qu'elle le résolve de manière parfaite — ce que même l'intelligence humaine ne fait pas . >>





03 Implications philosophiques



You can enter a subtitle here if you need it

Le problème de la pertinence chez l'être humain

- Le cerveau humain traite une immense quantité d'informations
- Seules les plus pertinentes sont retenues pour l'action
- Ce filtrage automatique reste un mystère de la cognition





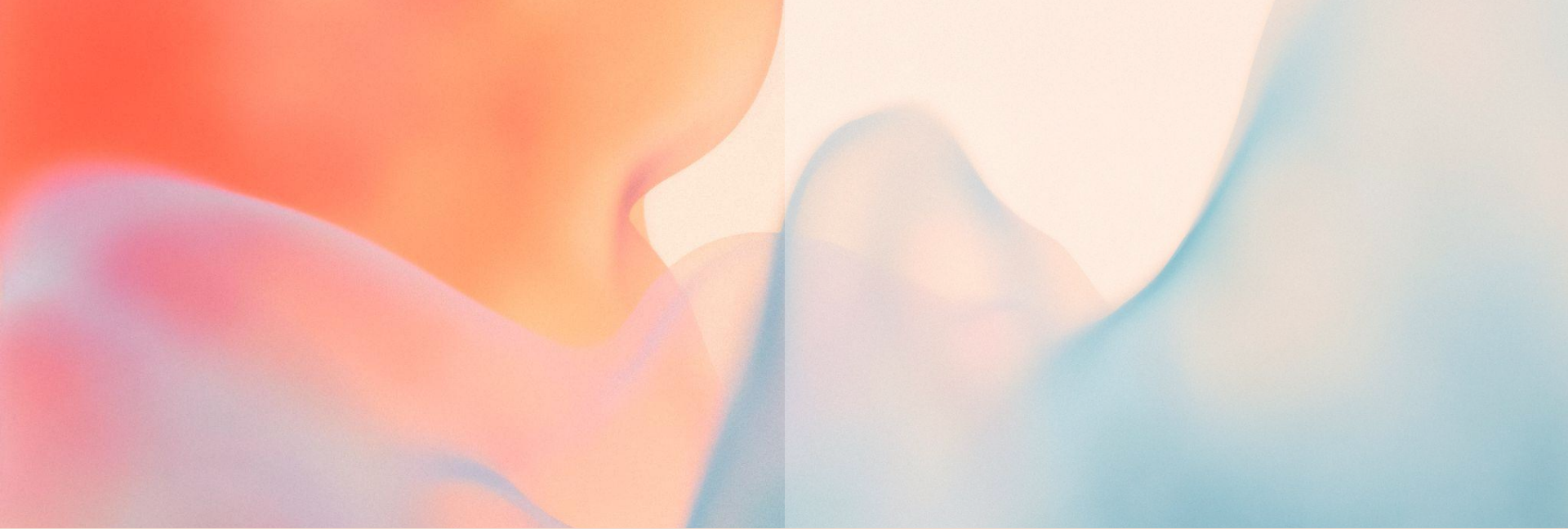
Le sandwich de Dennett

- Mission : faire un **sandwich** et verser un verre de **bière**
- Le cerveau se concentre:
 - ouvrir le frigo,
 - étaler la mayonnaise,
 - porter l'assiette
 - verser la bière
 - etc.
- Il ne pense pas à :
 - la couleur du frigo
 - la température de la pièce
 - le nombre d'assiettes dans le placard

Comment le cerveau a t-il cette capacité de filtrage?



- **Fodor:**
 - Le cerveau raisonne à partir d'un langage interne
 - le Frame Problem révèle la limite de cette modélisation.
- **Dreyfus:**
 - La pensée est incarnée et contextuelle
 - L'intelligence vient du corps et de l'expérience vécue ; le sens émerge du contexte plutôt que d'un raisonnement symbolique.
- **Dennett:**
 - Le cerveau n'a jamais eu besoin de résoudre le problème : il agit grâce à une perception sélective façonnée par l'évolution.



04 Discussion



0 – Quelles IAs aujourd'hui sont concernées par le Frame Problem ?

Concernées

- Robots industriels
- Robots autonomes,
- systèmes de planification symbolique,
- voitures autonomes (méthodes symboliques)

Non concernées

- Voitures autonomes (approches de deep learning basées sur NN)
- LLMs
- IA statistiques
- systèmes de recommandation

Débat – Liste de sujets

1. A votre avis, comment l'esprit humain résout / contourne ce problème ?
2. Est-ce que les humains ont eux aussi un Frame Problem ?
3. Faut-il abandonner la logique symbolique pour progresser en IA ?
4. Peut-on avoir une IA "générale" sans résoudre ce problème ?
5. A-t-on aujourd'hui vraiment résolu ce problème, ou juste contourné ?
6. À quoi ressemblerait une vraie solution au Frame Problem ?
7. Est-ce que ce problème prouve qu'on ne peut pas "simuler" l'esprit humain avec des machines ?

Sujet 1 : 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.

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

- Pertinence contextuelle pour des réseaux de neurones - comment distinguer ce qui est important dans une situation ?
- Calcul des impacts de leurs actions pour les LLMs qui doivent être ajustés pour éviter certains biais dangereux (racisme, violence).
- Cependant pour ces LLMs, problèmes aujourd'hui corrigés par intervention humaine (apprentissage supervisé, révision des données d'entraînement, hard-safety)

En Conclusion : Le sujet n'est actuellement pas tranché parmi les scientifiques, et fait toujours débat

Sujet 2 : Comment le cerveau humain résout/contourne le problème ?

L'approche neuro-scientifique :

- 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.
- Via des mécanismes inconscients, de filtre des informations jugées utiles.

Problématique : beaucoup de théories existent sur l'émergence de ces comportements, mais pas de consensus. Comment donc faire émerger ces mécanismes dans des IAs ?

Sujet 3 : Faut-il abandonner la logique symbolique pour s'affranchir du Frame Problem ?

Pas forcément :

- Beaucoup d'approches hybrides émergent aujourd'hui, entre autre pour répondre à la problématique
- Des approches d'apprentissage par renforcement permettent à des modèles d'inférer sur leurs propres règles (axiomes donc)
- Ces approches partent à la base de règles définies par des chercheurs, mais leur permet d'affiner leur propres mécanismes de “tri” des informations et de calcul des impacts de leurs actions => Piste de réponse au problème

Problématique : beaucoup de théories existent sur l'émergence de ces comportements, mais pas de consensus.
Comment donc faire émerger ces mécanismes dans des IAs ?

Littérature

0. McCarthy & Hayes (1969) – “Some Philosophical Problems from the Standpoint of Artificial Intelligence”

- Contexte : IA symbolique, systèmes logiques rigides (General Problem Solver).
- Objectif : généraliser à des situations plus complexes.
- Proposent le **Calcul des situations** (ex : `Result(a, s)` pour état après action).
- Introduisent les **Frame Axioms** : dire ce qui **ne change pas** après une action.
- Parlent aussi de :
 - **Qualification problem** : dans quelles conditions une action est possible ?
 - **Ramification problem** : quelles sont les conséquences indirectes ?
 - **Reification problem** : comment représenter des idées abstraites comme des objets.
- Problème du bon sens : l'IA a du mal à savoir ce qu'il faut ignorer ou garder constant.

1. Daniel C. Dennett – *Cognitive Wheels: The Frame Problem of AI* (1984)

- Exemples des robots R1D1, R2D1, etc
- Son parti pris : Le Frame Problem n'est pas juste un bug technique à résoudre en IA, c'est un révélateur profond de ce qu'est vraiment la cognition.

2. Jarek Gryn – *The Frame Problem in Artificial Intelligence and Philosophy* (2013)

- Tentatives de solutions : logique par défaut (Reiter) ⇔ certaines choses restent vraies sauf preuve contraire
- Holisme : dur de déterminer ce qui est pertinent => la connaissance est interconnectée

Merci

Des questions ?

Envoyez nous vos remarques à basura@univ-lyon1.fr

Dimitrios Stephanou
Artus Bleton
Guilhem Dupuy