



UNIVERSITÉ  
CÔTE D'AZUR

EFELIA  
ÉCOLE FRANÇAISE DE  
L'INTELLIGENCE ARTIFICIELLE  
CÔTE D'AZUR

# Compétence transversale – L1

## Grands défis sociétaux : Intelligence Artificielle

Pr. Lucile Sassatelli

Professeure des Universités en Informatique, UniCA

Directrice scientifique de EFELIA Côte d'Azur

Image by Alan Warburton / © BBC / Better Images of AI / Nature / CC-BY 4.0





# Ma recherche

- IA pour l'analyse multimédia



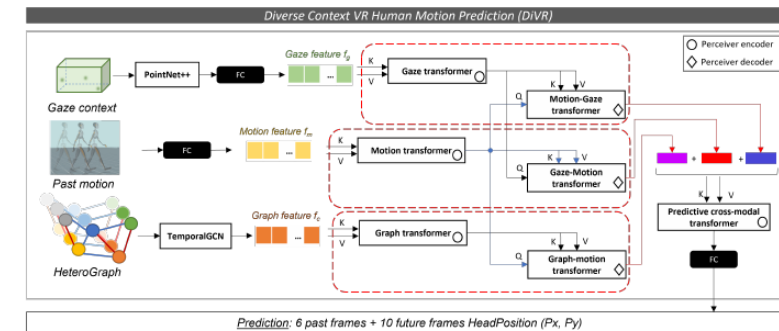
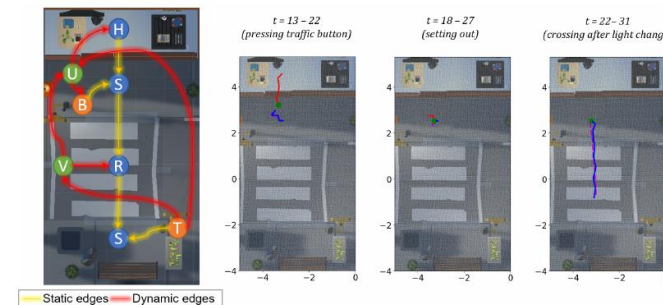
Figure 1 : (A) unequal gaze (B) Nudity and submissive postures (C) animalisation or infantilisation (D) transparent clothing, camera framing, domestic gender roles, and voyeurism

Test Train	EN vs. S		(EN U HN) vs. S	
	EN vs. S	HN vs. S	EN vs. S	HN vs. S
ViT-B/16	0.53 (0.18)	0.62 (0.13)	0.54 (0.24)	0.73 (0.1)
X-CLIP	<b>0.79</b> (0.05)	0.71 (0.05)	0.66 (0.05)	<b>0.82</b> (0.03)
Random	0.32		0.28	
All positive	0.37		0.33	
PCBM-DT	0.68	0.44	0.58	0.38
PCBM-LR	0.64	0.43	0.50	0.37

F1-score on the binary task of objectification detection

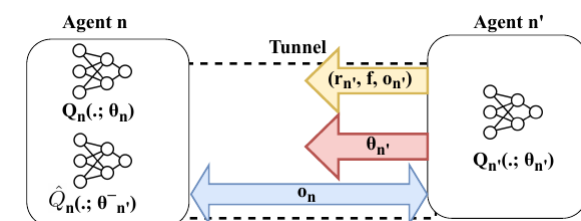
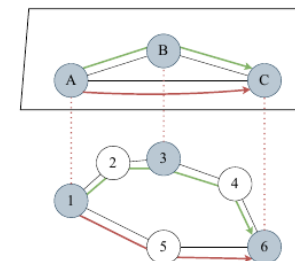
© Julie Tores

- IA pour la prédiction de mouvement



© Franz Franco Gallo

- IA pour l'optimisation des réseaux



© Redha Alliche



# Modalités d'apprentissage

10 à 12h de travail personnel

Pour chacun des 4 chapitres :

- un support pdf
- plusieurs vidéos
- des questions sous la forme de QCM au cours du chapitre
  - pour que vous sachiez si vous avez bien compris les notions importantes que vous venez d'aborder, non noté
  - permet de réviser les QCM notés constituant votre note finale
- des manipulations pratiques de systèmes d'IA
  - n'impliquant aucune programmation, pour vous permettre de mieux comprendre
- un forum pour poser vos questions et échanger :
  - dans lequel vous pourrez interagir un ou une enseignante.



# Modalités d'évaluation

- QCM 1 après le chapitre 2 :
  - coefficient 0.5, temps limité (quelques minutes), ouverture du 10 au 14 novembre
- QCM 2 final après chapitre 4 (en décembre) :
  - coefficient 0.5, temps limité (quelques minutes), ouverture du 8 au 12 décembre



# Objectifs du module

- A la fin de ce module :
  - vous connaîtrez ce que peuvent et ne peuvent pas faire les outils d'IA actuels, leurs impacts sociétaux et environnementaux
  - vous disposerez des éléments scientifiques connus à ce jour pour décider par vous-même si, quand, et comment utiliser des outils d'IA dans votre processus d'apprentissage personnel
  - vous aurez une vision générale de l'impact des outils d'IA pour votre discipline (comment éventuellement ils aident, modifient les pratiques, ou posent de nouveaux problèmes, et quel est l'état des connaissances sur ces questions)



# Plan du module

Chapitre	Titre	Contenu	Date d'ouverture	Date QCM
1	Qu'est-ce que l'IA ?	<ul style="list-style-type: none"><li>• Domaine et familles de méthodes</li><li>• Applications actuelles</li><li>• Définition d'un modèle d'IA, choix humains dans la conception</li><li>• Mythe de l'AGI et confrontation à la réalité des neurosciences</li></ul>		
2	Qu'est-ce que les grands modèles de langue (LLM) ?	<ul style="list-style-type: none"><li>• Principe des modèles génératifs de texte</li><li>• Choix simplificateurs et conséquences : faillibilité, biais</li></ul>		<ul style="list-style-type: none"><li>• QCM 1 noté 3-7/11</li></ul>
3	Quels impacts sur la société ? Quels impacts sur moi ?	<ul style="list-style-type: none"><li>• Exploitation humaine et emplois</li><li>• Exploitation environnementale</li><li>• Apprentissage à l'université</li></ul>		
4	Quels impacts sur ma discipline ?	<ul style="list-style-type: none"><li>• Les intérêts des systèmes d'IA</li><li>• Les enjeux des systèmes d'IA pour son domaine</li></ul>		<ul style="list-style-type: none"><li>• QCM 2 note 8-12/12</li></ul>



# Ressources et vocabulaire

## • Lectures

- AI Myths. <https://www.aimyths.org/>
- S. d'Acoli. [Comprendre la révolution de l'intelligence artificielle](#). 2020
- Kate Crawford. [Contre-atlas de l'intelligence artificielle](#). Zulma éditions, 2022.

(disponibles dans vos BU)

## • Vocabulaire

- ~~Une IA~~
- Un système d'IA (SIA)<sup>1</sup>
- Un modèle d'IA
- Outils basés IA

<sup>1</sup> <https://declarationmontreal-iaresponsable.com/la-declaration/>

## COMPRENDRE L'INTELLIGENCE ARTIFICIELLE

FR <

### Sommaire

- Préambule
- Pour comprendre la diversité des méthodes d'IA, leurs grands principes et leurs évolutions historiques
- La révolution des années 2010 : l'apprentissage de représentations avec le Deep Learning
- L'IA générative
- Les limites des modèles de ML et les problèmes de biais
- Les stéréotypes sociaux reproduits par les modèles d'IA
- Besoin d'une réflexion éthique par toutes et tous ?
- Quelques ressources pour dépasser le discours dominant et amorcer un regard critique

Préambule : mieux comprendre l'intelligence artificielle

+

Pour comprendre la diversité des méthodes d'IA, leurs grands principes et leurs évolutions historiques

+

La révolution des années 2010 : l'apprentissage de représentations avec le Deep Learning

+

L'IA générative

+

Les limites des modèles de ML et les problèmes de biais

+

Quelques ressources pour dépasser le discours dominant et amorcer un regard critique

+



<https://univ-cotedazur.fr/efelia-cote-dazur/comprendre-lia>

# Plan



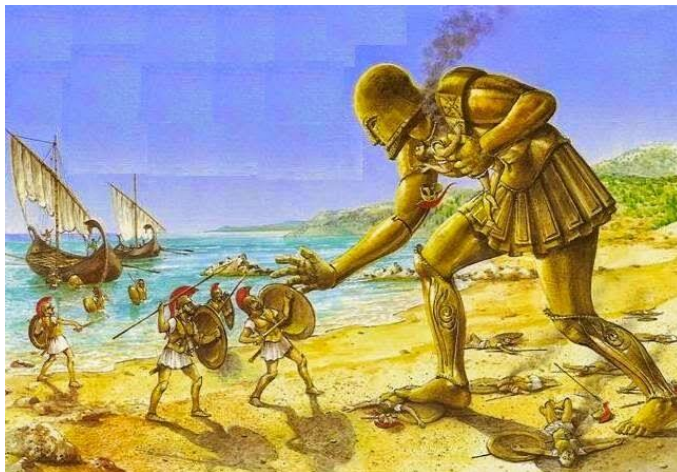
1. Domaine et familles de méthodes :
  - IA symbolique (basée connaissance)
  - IA statistique (basée données)
2. Exemples de systèmes d'IA :
  - Pour différents domaines : langue, vision, profilage
  - Pour différentes tâches : classification, régression, génération
3. Principe d'un modèle d'IA/ML
  - Base : un modèle de ML est une fonction paramétrée
  - Principe de l'entraînement : trouver les paramètres à partir de données
  - Processus de création d'un modèle avec les différents choix humains
4. IA étroite et IA générale (*Artificial General Intelligence – AGI*)
  - IA étroite vers IA générale : un acte de foi ?
  - Mais que nous disent les dernier progrès en psychologie et neurosciences sur « l'intelligence » ?
  - Quelles racines idéologiques à l'IA générale, quels acteurs, intérêts, et quelles conséquence du discours actuel ?





# L'IA et la mythologie

- La notion (pas le terme) est présente depuis l'antiquité (*Automates et créatures artificielles d'Héphaïstos*, ou le *Golem* dans la mythologie juive, etc).



(sources: Wikipedia, <https://www.greeklegendsandmyths.com/automatons.html> ,  
[http://www.alanturing.net/turing\\_archive/pages/Reference%20Articles/what\\_is\\_AI/What%20is%20AI02.html](http://www.alanturing.net/turing_archive/pages/Reference%20Articles/what_is_AI/What%20is%20AI02.html)  
Stanford Encyclopedia of Philosophy: <https://plato.stanford.edu/entries/artificial-intelligence/>)



© Frédéric Precioso

# 1952-1956 : La naissance de l'IA et la cybernétique

- 1930-50 **neurologie** : le cerveau est un réseau électrique de neurones
- **Cybernétique** (Norbert Wiener) : contrôle et stabilité des réseaux électriques
- **Théorie de l'information** (Claude Shannon) : signal et communications
- **Théorie du calcul** (Alan Turing) : manipulation de symboles pour effectuer un raisonnement

→ **Peut-on créer un cerveau artificiel ?**

→ 1920-30 : Hilbert : est-ce que tout raisonnement mathématique peut être formalisé ?



# 1956 : workshop de Dartmouth

- Organisé par Marvin Minsky, John McCarthy, Claude Shannon, Nathan Rochester, **pensant que :**
  - “Tout aspect de l'apprentissage ou de toute caractéristique de l'intelligence peut être décrit si précisément qu'une machine peut être créée pour le simuler.”
  - “Nous proposons qu'une étude de l'intelligence artificielle soit menée par **10 personnes pendant 2 mois à l'été 1956** au Dartmouth College (NH, USA).”
  - “Nous pensons qu'une avancée significative peut être réalisée [...] si un groupe de scientifiques soigneusement sélectionnés y travaille ensemble pendant un été.”
- L'objectif était clairement trop ambitieux, aussi « soigneusement sélectionnés » ces chercheurs furent-ils : 65 ans après, le but n'est pas atteint.



# Définition de l'IA

- L'IA est un domaine né à l'intersection de l'informatique, des mathématiques et des neurosciences au milieu du XXème siècle.
- L'objectif est la conception d'approches computationnelles (d'algorithmes<sup>1</sup>), c'est-à-dire automatisées avec des ordinateurs, de tâches jusque là uniquement réalisées par des humains.
- Ces tâches peuvent concerner des problèmes cognitifs élaborés et donc jugés à l'époque difficiles à résoudre par des algorithmes, comme démontrer des théorèmes mathématiques, ou des problèmes résolus de façon subconsciente et automatique par les êtres vivants, comme se déplacer, ou marcher pour certains.



## 2 grandes familles en IA

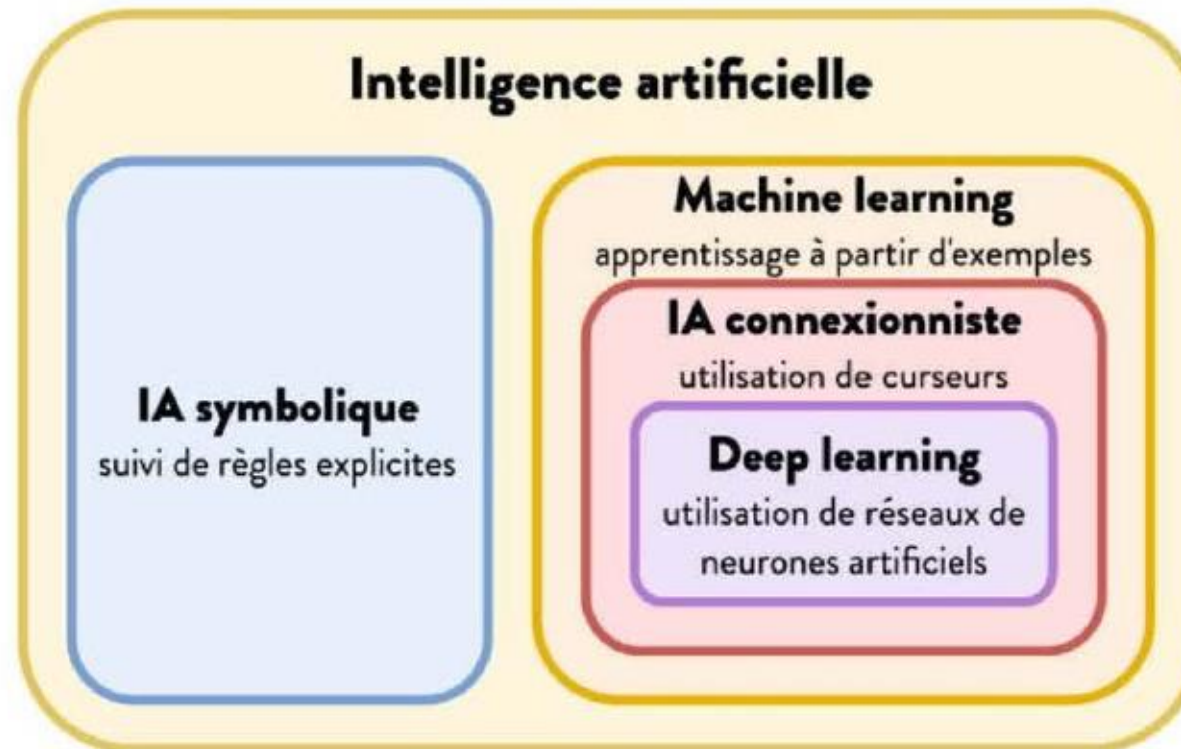
- Ces courants, connus sous le noms de top-down AI et bottom-up AI, ont donné lieu à des familles méthodologiques différentes en IA :
- **l'IA symbolique** :
  - approche dominante entre les années 1950 et 1990
  - Hypothèse d'existence de symboles de haut niveau pour représenter les briques de raisonnement (champs du raisonnement, de la programmation logique, des systèmes de règles), basé connaissance
- **l'IA dite statistique, ou basée données** :
  - conçoit et met en oeuvre des outils statistiques pour identifier des motifs (des associations, ou « corrélations » en statistiques) dans les données, permettant de les relier à une catégorie, ou de les re-générer.
  - famille de méthodes ayant connu le plus de succès depuis les années 2000 : c'est le champ du **machine learning - ou apprentissage automatique**, avec notamment les **réseaux de neurones artificiels**.



[https://www.alanturing.net/turing\\_archive/pages/reference%20articles/what\\_is\\_ai/What%20is%20AI09.html](https://www.alanturing.net/turing_archive/pages/reference%20articles/what_is_ai/What%20is%20AI09.html)

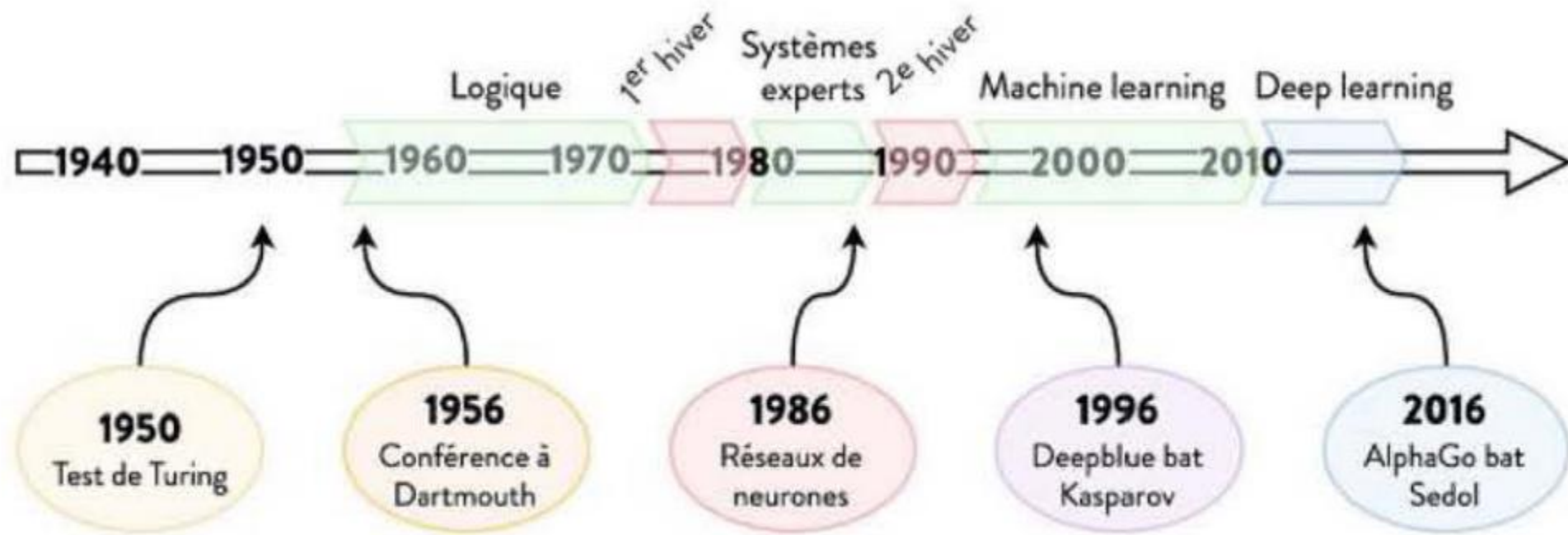


# On récapitule :



L'IA est séparée en deux catégories : les algorithmes à base de règles (IA symbolique) et les algorithmes d'apprentissage (IA connexionniste).

# On récapitule :



Frise chronologique résumant l'histoire de l'intelligence artificielle

# Plan

## 1. Domaine et familles de méthodes :

- IA symbolique (basée connaissance)
- IA statistique (basée données)

## 2. Exemples de systèmes d'IA :

- Pour différents domaines : langue, vision, profilage
- Pour différentes tâches : classification, localisation, régression, génération, classement

## 3. Principe d'un modèle d'IA/ML

- Base : un modèle de ML est une fonction paramétrée
- Principe de l'entraînement : trouver les paramètres à partir de données
- Processus de création d'un modèle avec les différents choix humains

## 4. IA étroite et IA générale (*Artificial General Intelligence* – AGI)

- IA étroite vers IA générale : un acte de foi ?
- Mais que nous disent les dernier progrès en psychologie et neurosciences sur « l'intelligence » ?
- Quelles racines idéologiques à l'IA générale, quels acteurs, intérêts, et quelles conséquence du discours actuel ?





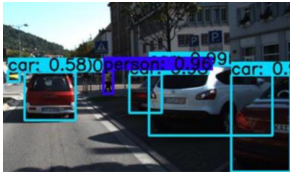


# Les applications actuelles en traitement automatique de la langue naturelle (TALN, *NLP*)

Entrée	Sortie	Nom de la tâche	Domaine	Type de tâche de ML
<b>Avis</b> <i>« Ce que j'ai aimé dans ce film, c'est la clim de la salle de ciné. »</i>	Positif, Négatif, Neutre Négatif	Classification de sentiment	NLP	Classification
<b>Mail</b>	Spam, non-spam	Classification de spam	NLP	Classification
<b>Texte en français</b> <i>« Ce que j'ai aimé dans ce film, c'est la clim de la salle de ciné. »</i>	<b>Texte en anglais</b> <i>"The one thing I enjoyed in this movie was the AC of the teather."</i>	Traduction automatique	NLP	Génération
<b>Instruction</b> <i>Indique si cet avis est positif, négatif ou neutre. Avis : « Ce que j'ai aimé dans ce film, c'est la clim de la salle de ciné. »</i>	<b>Réponse</b> Cet avis est <b>**négatif**</b> . La personne ne met en avant que la climatisation de la salle de cinéma comme aspect positif, ce qui sous-entend qu'elle n'a pas apprécié le film lui-même.	Complétion	NLP	Génération








# Les applications actuelles en vision par ordinateur (CV)

Entrée	Sortie	Nom de la tâche	Domaine	Type de tâche de ML
	chat, chien, voiture	Classification d'image	CV	Classification
Ref:  New: 	Même personne, pas même personne	Identification biométrique	CV	Classification (appariement -- <i>matching</i> )
Image: 	Boîtes englobantes et labels: 	Détection d'objets	CV	Localisation et classification









# Les application en vision-langage

Entrée	Sortie	Nom de la tâche	Domaine	Type de tâche de ML
<p>Texte d'instruction :</p> <p><sup>1</sup><i>Generate an image of an elephant swimming underwater. aesthetic. Fantasy.</i></p>	<p>Image :</p> 	Génération texte à image	CV-NLP	Génération
<p>Texte d'instruction avec image :</p> <p><sup>2</sup><i>What could have happened based on the current scene?</i></p> 	<p>Texte :</p> <p><i>Based on the current scene in the image, it is possible that a hurricane or severe weather event caused significant damage to the buildings and infrastructure in the area.</i></p>	Génération image à texte (Visual Question Answering)	CV-NLP	Génération
<p>Texte d'instruction avec image :</p> <p><sup>3</sup><i>Q: In this image, how many eyes can you see on the animal?</i></p> 	<p>Texte :</p> <p><i>The image shows one eye of the animal. It's a close-up of a bald eagle facing slightly to its right, which presents only one side of its face to the viewer.</i></p>	Génération image à texte (Visual Question Answering)	CV-NLP	Génération
<p>Texte d'instruction avec image :</p> <p><sup>1</sup><i>Render sunset</i></p> 	<p>Image :</p> 	Edition d'image	CV-NLP	Génération



# Les applications en profiling humain

Entrée	Sortie	Nom de la tâche	Domaine	Type de tâche de ML	
<b>Profil utilisateur·rice</b> <i>Écoutes passées, temps d'écoute, connexions sociales, catégorie sociale</i>	<b>Recommandation de musiques</b> Chanson1, chanson2, ...	Recommandations d'écoute	Profilage pour les industries culturelles	Classement (tri)	
<b>Profil utilisateur·rice</b> <i>Matches passés, connexions sociales, catégorie sociale</i>	<b>Recommandation de partenaires</b> Partenaire1, partenaire2, ...	Recommandations de partenaire	Profilage pour les relations sociales	Classement (tri)	
<b>Curriculum vitae</b> <i>Univ, hobbies, sports</i>	<b>Employabilité</b> 0-10	Estimation d'employabilité	Gestion de ressources humaines	Régression	
<b>Profil patient·e</b> <i>Age, sexe, tension, cholestérol, médicaments, assurance</i>	<b>Besoin de suivi médical</b> 0/1	Estimation de besoin médical	Gestion du système de santé	Classification	

A. Moradi, et al., [“Advancing Cultural Inclusivity: Optimizing Embedding Spaces for Balanced Music Recommendations,”](#) arXiv:2405.17607, May 2024.


J. A. Hutson, et al., [“Debiasing Desire: Addressing Bias & Discrimination on Intimate Platforms,”](#) *Proc. ACM Hum.-Comput. Interact.*, vol. 2, no. CSCW, Nov. 2018.

Jeffrey Dastin, [“Amazon scraps secret AI recruiting tool that showed bias against women,”](#) Reuters, 2018.

Ziad Obermeyer, et al., [“Dissecting racial bias in an algorithm used to manage the health of populations,”](#) *Science*, vol. 366, no. 6464, pp. 447–453, 2019.




# Plan

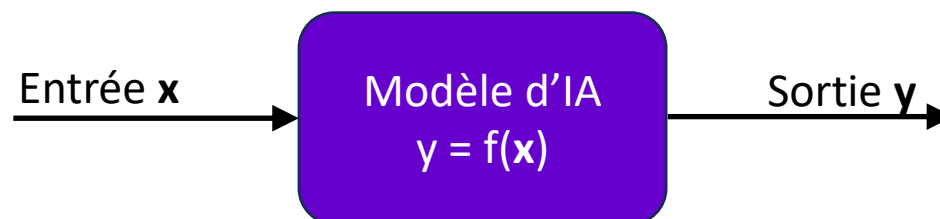
1. Domaine et familles de méthodes :
  - IA symbolique (basée connaissance)
  - IA statistique (basée données)
2. Exemples de systèmes d'IA :
  - Pour différents domaines : langue, vision, profilage
  - Pour différentes tâches : classification, localisation, régression, génération, classement
-  3. Principe d'un modèle d'IA/ML
  - Base : un modèle de ML est une fonction paramétrée
  - Principe de l'entraînement : trouver les paramètres à partir de données
  - Processus de création d'un modèle avec les différents choix humains
4. IA étroite et IA générale (*Artificial General Intelligence – AGI*)
  - IA étroite vers IA générale : un acte de foi ?
  - Mais que nous disent les dernier progrès en psychologie et neurosciences sur « l'intelligence » ?
  - Quelles racines idéologiques à l'IA générale, quels acteurs, intérêts, et quelles conséquence du discours actuel ?



# Un modèle de machine learning/Modèle d'IA

Entrée	Fonction mathématique Sortie $y = f(x)$
	chat, chien, voiture

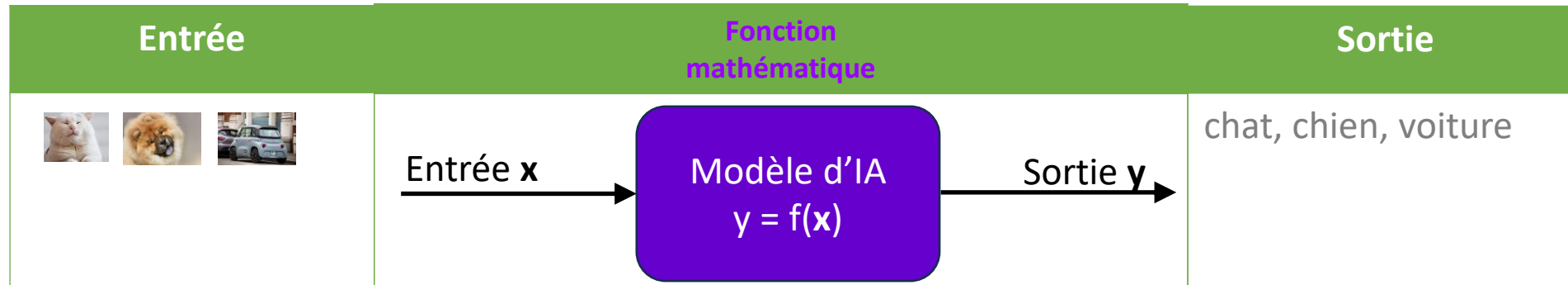
- Un modèle d'IA est une fonction mathématique qui associe une entrée à une sortie, permettant d'automatiser cette association



- On peut décider plusieurs formes pour cette fonction, mais ce n'est pas le modèle d'un cerveau humain.

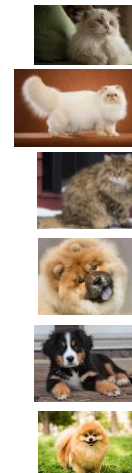


# Comment trouver ce modèle d'IA / cette fonction ?



## Un jeu de données d'entraînement

- On considère souvent qu'il existe des **paires d'exemples entrée-sortie où on connaît la sortie idéale désirée**.
  - On peut prendre **exemple sur ces paires  $(x,y)$  pour trouver la fonction mathématique  $f(.)$**  les reliant/interpolant au mieux : **apprentissage supervisé**
  - Le terme « apprentissage » fait référence ici à la recherche d'une fonction mathématique, pas à un processus biologique ou humain.

**x****y (« label »)**  
(0: chat, 1: chien)

0

0

0

1

1

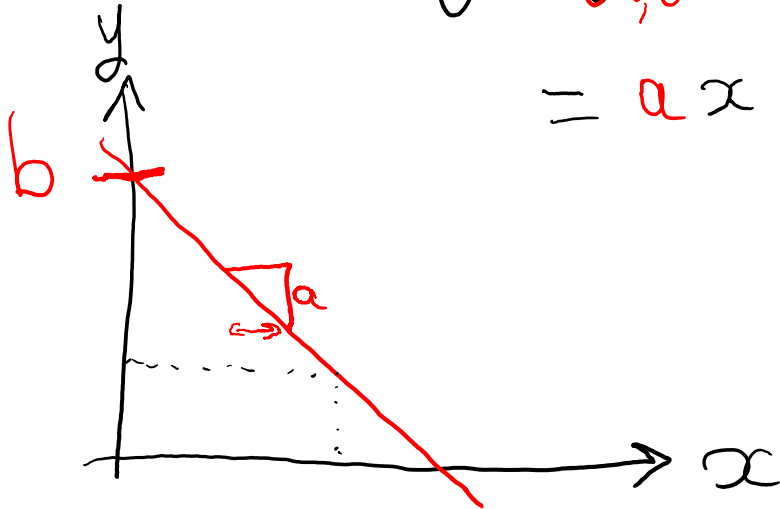
1





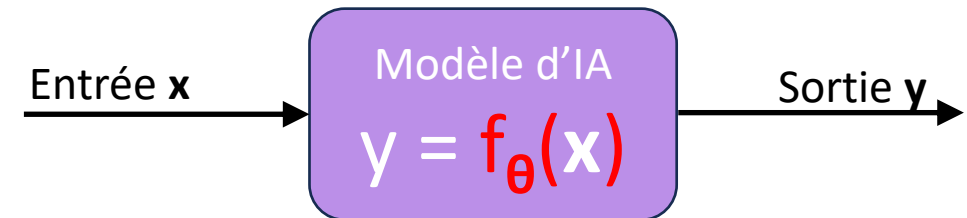
# Rappel : une fonction paramétrée

$$y = f_{a,b}(x) \\ = ax + b$$

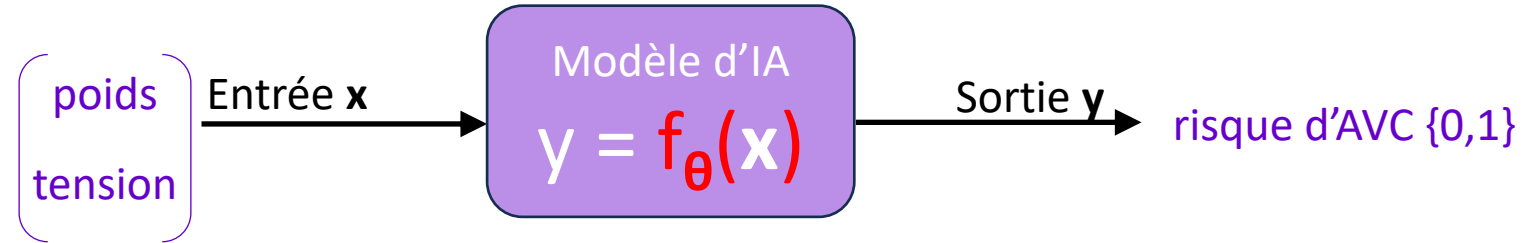


- Et plus généralement :

$$y = f_{\theta}(x)$$



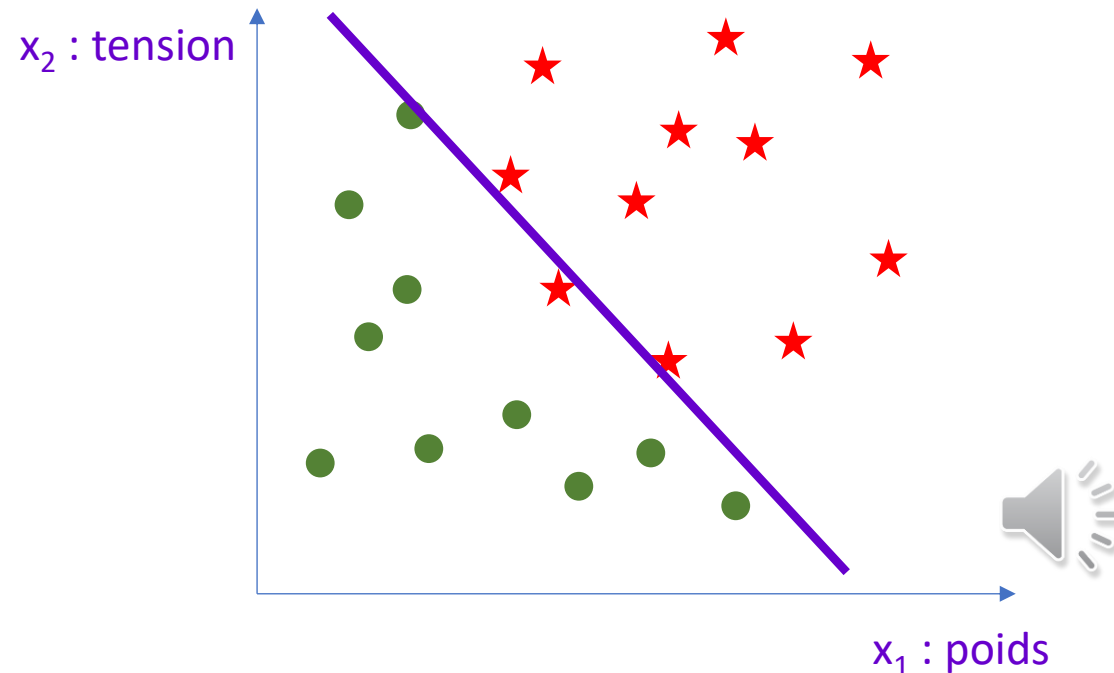
# Qu'est-ce qu'un modèle IA/ML ?



- Une fonction paramétrée pour :

- Classer des échantillons :  
*classification*

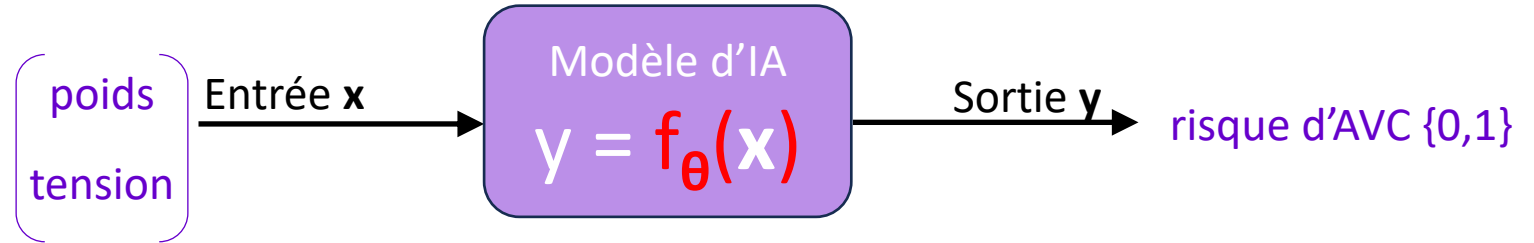
$$y = \text{signe}(w_1 x_1 + b + w_2 x_2) = \text{sgn}(\mathbf{w}^T \mathbf{x} + b)$$



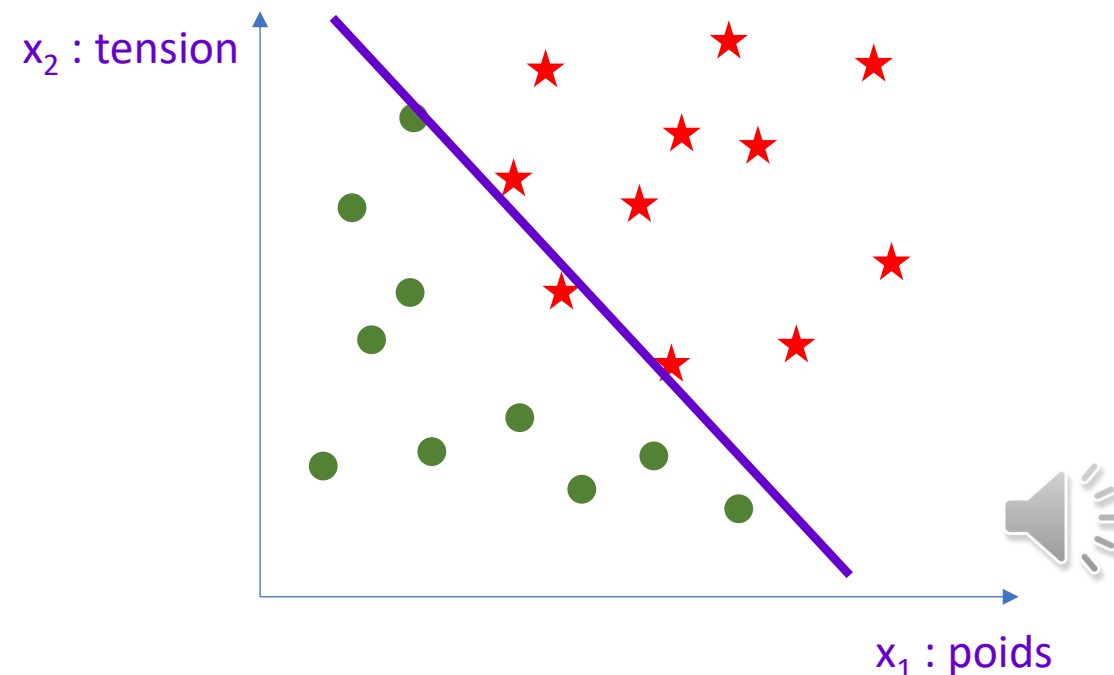
# Trouver les paramètres de la fonction

- **Entraîner** un modèle de ML :

- Trouver les paramètres de la fonction (coefficients  $w$  de la moyenne et seuil  $b$ )
- pour minimiser l'erreur de prédiction sur les exemples d'entraînement
- Réalisé par calcul de dérivée pour minimiser l'erreur

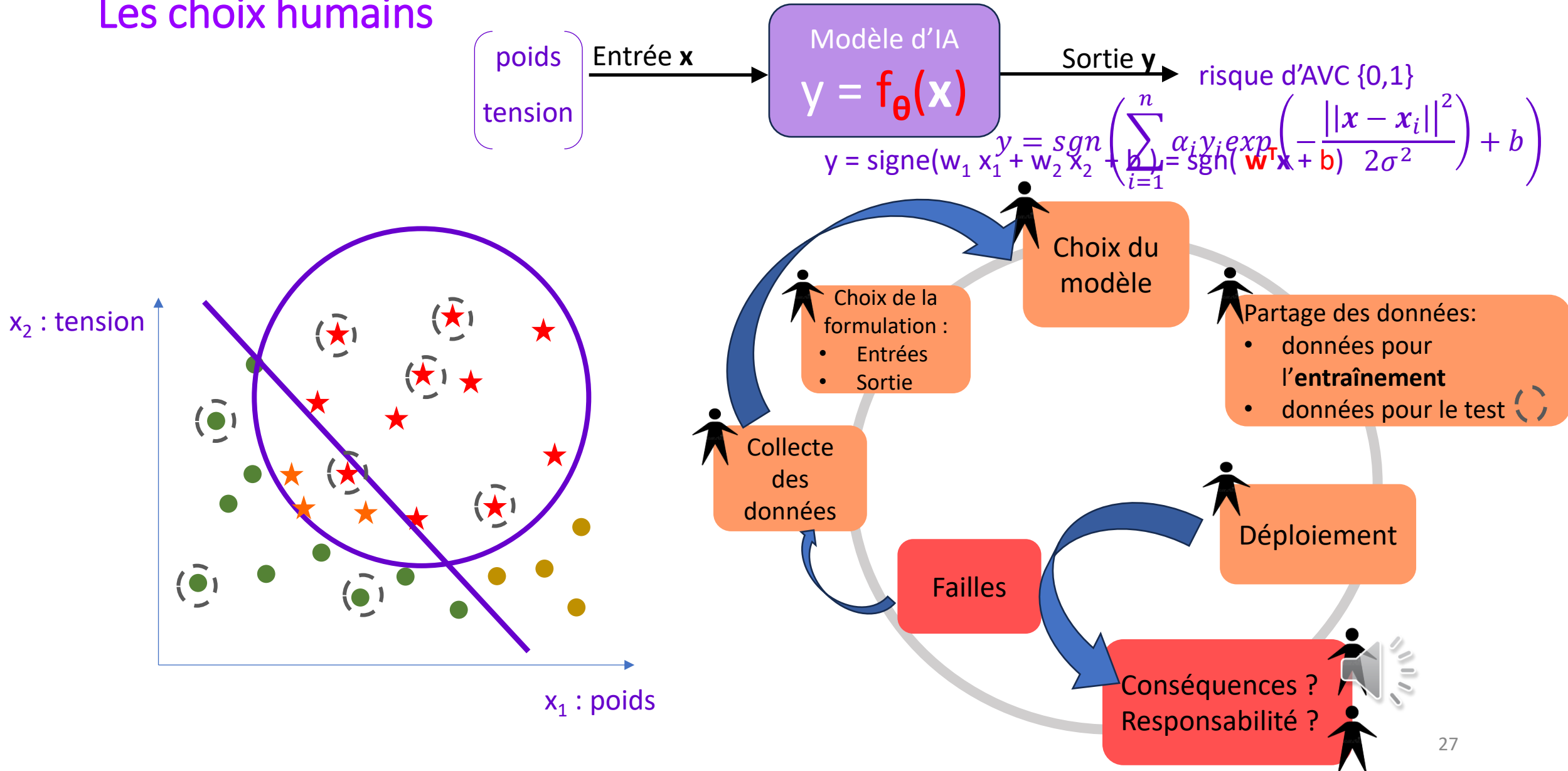


$$y = \text{signe}(w_1 x_1 + b + w_2 x_2) = \text{sgn}(\mathbf{w}^T \mathbf{x} + b)$$

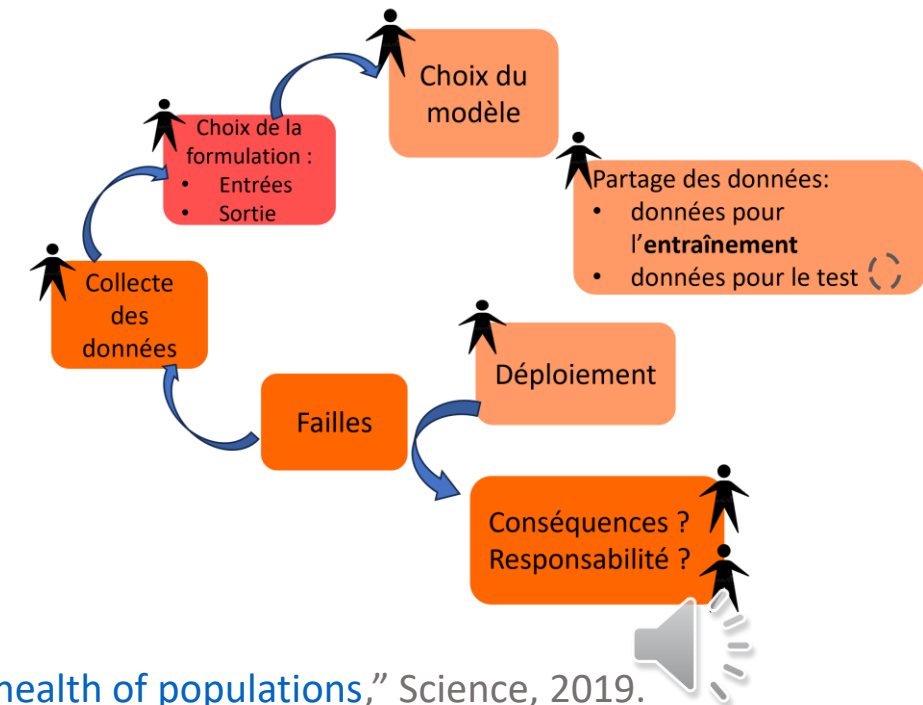
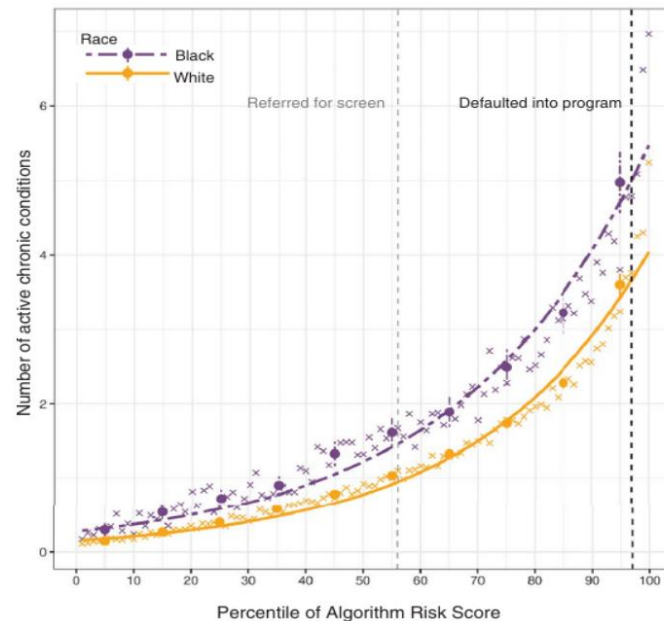
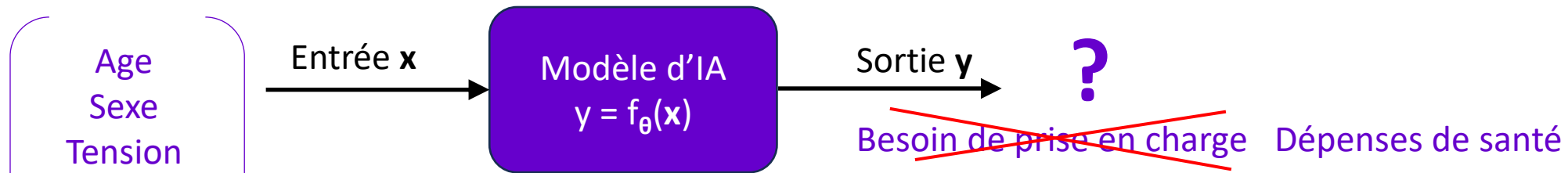


# Qu'est-ce qu'un modèle IA/ML ?

## Les choix humains



# Cas 1 : Optimiser le système de santé ?

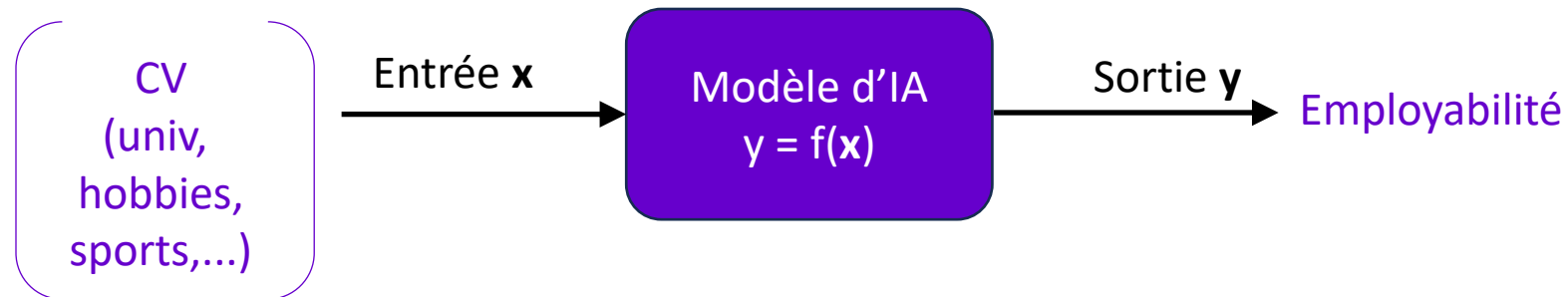


Z. Obermeyer et al. "[Dissecting racial bias in an algorithm used to manage the health of populations](#)," Science, 2019.

L. Seyyed-Kalantari et al., "[Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in underserved patient populations](#)," Nature Medicine, Dec. 2021.

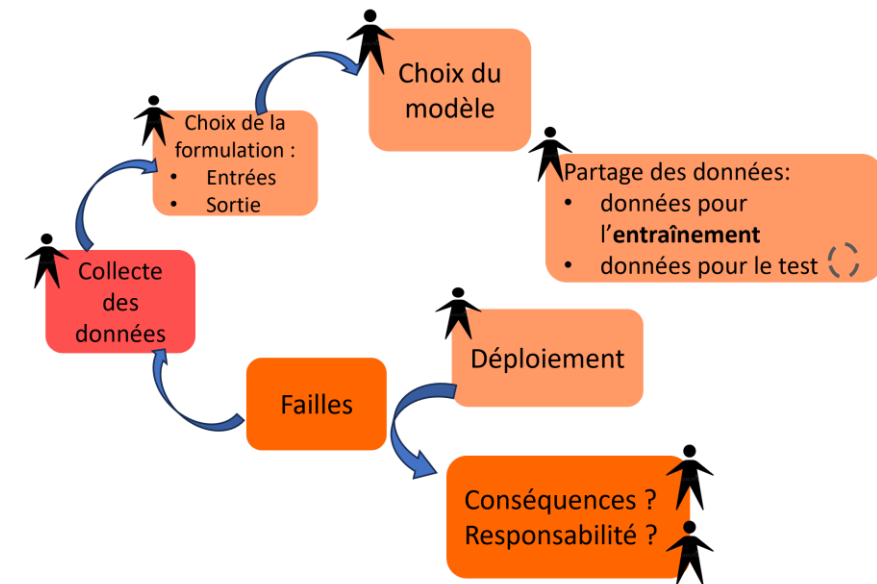


# Cas 4 : Filtrer des CV pour recruter ?




- CV de femmes systématiquement rejetés
- Cause : données historiques de recrutement biaisées

→ Le système reproduit le passé



Jeffrey Dastin, "Amazon scraps secret AI recruiting tool that showed bias against women," Reuters, 2018.

# Plan

1. Domaine et familles de méthodes :
  - IA symbolique (basée connaissance)
  - IA statistique (basée données)
2. Exemples de systèmes d'IA :
  - Pour différents domaines : langue, vision, profilage
  - Pour différentes tâches : classification, localisation, régression, génération, classement
3. Principe d'un modèle d'IA/ML
  - Base : un modèle de ML est une fonction paramétrée
  - Principe de l'entraînement : trouver les paramètres à partir de données
  - Processus de création d'un modèle avec les différents choix humains
-  4. IA étroite et IA générale (*Artificial General Intelligence* – AGI)
  - IA étroite vers IA générale : un acte de foi ?
  - Mais que nous disent les dernier progrès en psychologie et neurosciences sur « l'intelligence » ?
  - Quelles racines idéologiques à l'IA générale, quels acteurs, intérêts, et quelles conséquence du discours actuel ?

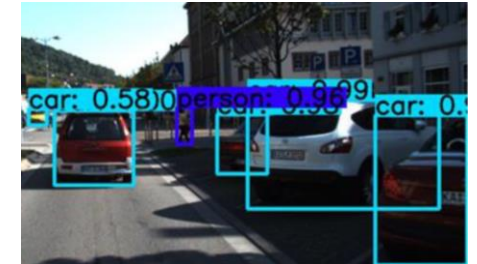


# IA étroite (ANI) et IA générale (AGI)

- Une fonction mathématique/un modèle, spécifique pour chaque tâche
- Systèmes réels qualifiés d'**IA étroite** par opposition à l'**IA générale (AGI)**, introduit en 1997
- Jusqu'en ~2010, chercheur·euses en NLP, CV, ML, prenaient leurs distances avec le terme d'"IA"
- **Le terme de AGI a été popularisé en 2007** (par Goertzel et Pennachin) pour décrire des systèmes d'IA "avec une intelligence de niveau humain voire sur-humain".

Cassio Pennachin and Ben Goertzel (editors). [Artificial general intelligence](#). Berlin: Springer, 2007.

T. Gebru and É. P. Torres, "[The TESCREAL bundle: Eugenics and the promise of utopia through artificial general intelligence](#)," FM, Apr. 2024, doi: 10.5210/fm.v29i4.13636.



Exemple d'IA étroite/Artificial  
Narrow Intelligence (ANI)



Exemple de représentation du  
**fantasme** d'IA générale/Artificial  
General Intelligence (AGI)

# Artificial General Intelligence... Intelligence ?

- Mai 2023 : Microsoft et OpenAI publient “Sparks of Artificial General Intelligence: Early experiments with GPT-4” prétendant que la capacité du modèle GPT-4 à réaliser diverses tâches représente une forme précoce d'IA générale.
- Mais qu'est-ce que l'intelligence ?
- Microsoft reconnaît dans son article cette difficulté et choisit d'utiliser la définition présentée par "un groupe de 52 psychologues s'étant accordé sur une définition large" en 1994, énoncée par une psychologue du nom de Linda Gottfredson :
  - mais pseudoscience raciste, arguant que les personnes noires ont un QI plus faibles



# Critique de la rhétorique actuelle autour de l'IA générale ou pourquoi l'IA générale est un fantasme

1. IA étroite vers IA générale : un acte de foi ?
2. Mais que nous disent les dernier progrès en psychologie et neurosciences sur « l'intelligence » ?
3. Quelles racines idéologiques à l'IA générale, quels acteurs, intérêts, et quelles conséquence du discours actuel ?



# 1. IA étroite vers IA générale : un acte de foi ?

- Progresser dans des tâches d'IA étroite ne donne aucune garantie d'un progrès vers une sorte quelconque d'IA générale (AGI), qui peut imiter ou surpasser l'intelligence humaine dans de multiples domaines.
- Gary Marcus (expert en psychologie cognitive) : « Penser que de tels progrès incrémentaux sur des tâches spécifiques mèneront au final à 'résoudre l'intelligence' est comme penser qu'on peut construire une échelle jusqu'à la Lune. »





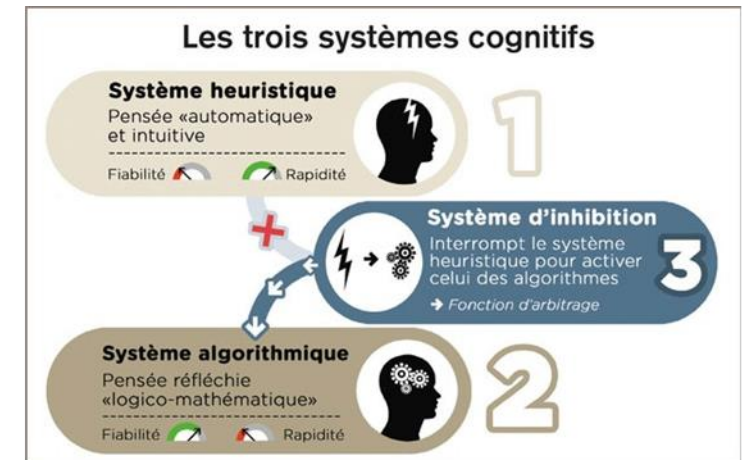
## 2. Définition de l'intelligence : que nous disent les progrès en psychologie et neurosciences ?

- O. Houdé : l'intelligence humaine permet de
  - « raisonner sur le monde par essais et erreurs, créer des modèles cognitifs internes (par des jeux d'inhibition/activation d'heuristiques approximatives et d'algorithmes exacts), élaborer une intention, un but, ressentir des émotions, mieux, des sentiments (doute, curiosité, regret, etc.) et entrer en interaction sociale avec autrui : connaître l'autre, le respecter, voire l'aimer, fonder une morale et une éthique pour la vie en société. Voilà ce qu'est, en quelques mots, l'intelligence – tant cognitive que sociale et affective. »



## 2. Mais que nous disent les derniers progrès en psychologie et neurosciences sur l'intelligence ?

- 3 systèmes organisent notre raisonnement :
  - **Système 1** : rapide, réflexes associatifs, heuristiques, pratiques mais souvent fausses dans les cas complexes (croyances, stéréotypes)
  - **Système 2** : logique et déduction, raisonnement algorithmique.
  - **Système 3** : l'arbitre qui peut inhiber la réponse rapide et moins coûteuse en énergie du système 1 pour diriger les informations vers le système 2.
- **Système 3 : fonctions exécutives du cortex préfrontal, en lien avec les émotions et l'autoévaluation :**
  - percevoir les « signaux d'erreur », ressentir la peur de l'erreur et l'envie de réussir, le doute, l'anticipation du regret, la curiosité
  - réagir en fonction de son but propre, de ses intentions, de ses enjeux personnels d'images de soi ou survie
  - pour envisager de se ressaisir, suspendre un circuit court et inhiber ses heuristiques trop tentantes.

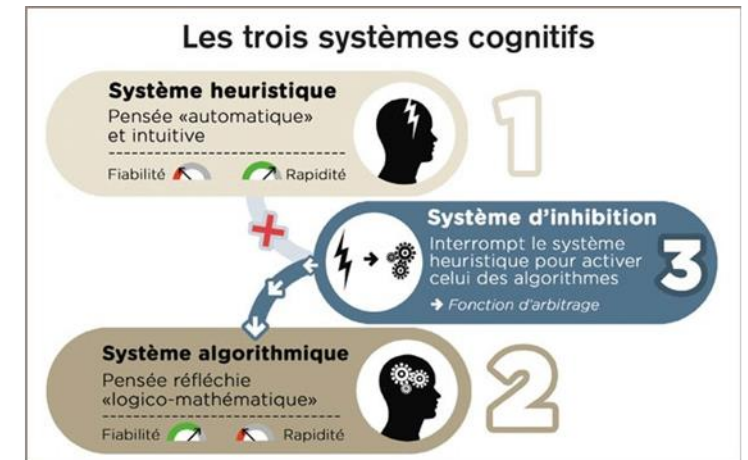


« Réfléchir c'est résister à soi-même » ©O. Houdé



## 2. Mais que nous disent les derniers progrès en psychologie et neurosciences sur l'intelligence ?

- « **Qu'est-ce que l'intelligence ?**
  - **C'est la capacité d'inhibition.**
- Inhiber quoi ?
  - Les circuits courts, trop rapides, trop faciles : les heuristiques lorsqu'elles sont erronées.
- Pour activer quoi ensuite ?
  - Les algorithmes logiques, plus lents et réfléchis, efficaces dans leur exactitude. Ce sont les circuits longs. Mais la vitesse de la pensée ne leur est pas favorable, ni notre paresse cognitive. »
- Pour le cerveau humain, c'est possible, mais **pour un ordinateur, cela paraît difficile** :
  - car il n'a pas peur, pas de but propre, d'enjeux personnels d'image (de soi) ou de survie, pas de corps à préserver qui le guiderait, via ses émotions et sentiments, à prendre les bonnes décisions.»



« Réfléchir c'est résister à soi-même » ©O. Houdé



### 3. Quelles racines idéologiques à la quête vers l'AGI, quels acteurs et quels intérêts ?

- T. Gebru et E. P. Torres montrent que cette quête prend ses racines dans l'eugénisme anglo-américain, à travers le transhumanisme.
- Iels identifient une généalogie d'idéologies interconnectées :
  - "transhumanism", "extropianism", "singularitarianism", "cosmism", "rationalism", "effective altruism", et "long-termism".
- Exemples :
  - Goertzel (2010) à propos du cosmisme : "les humains fusionneront avec la technologie," ce qui inaugurera "une nouvelle phase de l'évolution de notre espèce," "on développera une IA consciente et une technologie pour uploader l'esprit" qui permettra "une durée de vie infinie à ceux qui choisissent de renoncer à la biologie et de s'uploader."
  - "on s'étendra aux étoiles et on sillonnera l'univers," on créera "des réalités synthétiques" (des mondes virtuels), et on "développera l'ingénierie spatio-temporelle et la « magie du futur » scientifique bien au-delà de notre compréhension et de notre imagination actuelles"



### 3. Quelles racines idéologiques à l'IA générale, quels acteurs et quels intérêts ?

- 2024: milliards de dollars investis par les Big Tech dans la course à l'AGI, sous l'influence de personnalités avec ces idéologies :
  - Sam Altman (PDG d'OpenAI) : transhumaniste qui croit que nos cerveaux seront numérisés au cours de sa vie, et promeut des idées proches du cosmisme et du long-termism comme coloniser la galaxie. Il prédit en 2022 : « il est clair qu'on aura intelligence et énergie illimitée avant 2030. »
  - Elon Musk : transhumaniste dont l'entreprise Neuralink vise à « fusionner nos esprits avec l'IA », il a une influence considérable dans la communauté rationaliste, a fondé ou co-fondé plusieurs entreprises visant à créer l'IA générale, et décrit le long-termism comme "proche de sa philosophie".



### 3. Quelles racines idéologiques à l'IA générale, quels acteurs et quels intérêts ?

- « Ces courants accumulent les paradoxes, se contredisent et se réclament pour la plupart de la science, mais ils relèvent le plus souvent d'un mythe. Derrière ces approches souvent entretenues et promues par l'écosystème technologique se retrouve une symbolique associée à des scénarios catastrophiques ou à une forme de religiosité garantissant la promesse d'un salut virtuel infini. »
- Ils considèrent que leur quête de l'IA générale doit à la fois servir une utopie mais risque d'anéantir l'humanité.
- Le discours autour du risque existentiel lié à l'IA générale permet de focaliser l'attention du public et du politique sur des problèmes encore imaginaires et la détourne ainsi des dommages déjà bien réels subis notamment par les mêmes groupes déjà cibles de la première vague d'eugénisme.

L. Langlois et al., [“L'éthique au cœur de l'IA,”](#) Observatoire international sur les impacts sociétaux de l'intelligence artificielle et du numérique, Oct. 2023.

T. Gebru and É. P. Torres, [“The TESCREAL bundle: Eugenics and the promise of utopia through artificial general intelligence,”](#) FM, Apr. 2024, doi: 10.5210/fm.v29i4.13636.



# Vous ?

- Utilisez-vous des outils d'IA générative/chatbots ?
- Si non :
  - Pourquoi ne les utilisez-vous pas ?
- Si oui :
  - Pour quoi faire les utilisez vous ?
  - Comment les percevez-vous ?
    - Avez-vous la sensation qu'ils sont parfois, d'une certaine façon, conscients ?
    - Qu'ils sont souvent plus intelligents que vous ? Ou plus bêtes ?
  - Pour quelle raison les utilisez vous ?
  - Qu'est-ce qui vous gêne ou ne vous satisfait pas quand vous les utilisez pour vos études ? Par exemple en lien avec
    - les raisons pour lesquelles vous les utilisez (par exemple : parce-que tout le monde ou presque les utilise, donc vous les utilisez pour ne pas être désavantagé-e, ou parce-que ce qu'on vous demande ne vous intéresse pas donc vous voulez vite vous en débarrasser, ...)
    - l'impact que cet usage a sur vous, (par exemple : si vous acquérez bien ou mal les connaissances, sur vos notes, et finalement si ça questionne les raisons-mêmes pour lesquelles vous étudiez à l'université, ...)
  - Si vous n'aviez pas de contraintes (pas de notes, pas de surcharge de travail, pas de pression due à l'usage du reste de la promo, etc.) comment préféreriez-vous utiliser, ou ne pas utiliser, les outils d'IA dans vos études ?

