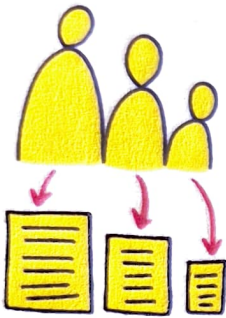


Question n°20

**QUELLES SONT LES QUESTIONS ÉTHIQUES
SOULEVÉES PAR L'IA ?**

LES LIMITATIONS INHÉRENTES DE L'IA

LES PRINCIPES DE FONCTIONNEMENT DES SYSTÈMES D'IA ACTUELS PRODUISENT SOUVENT DES RÉSULTATS CORRECTS MAIS PRÉSENTENT QUATRE GRANDES LIMITATIONS :



LES BIAIS

ILS SURVIENNENT LORSQUE LES DONNÉES D'ENTRAÎNEMENT NE REPRÉSENTENT PAS FIDÈLEMENT (au sens statistique) LA POPULATION OU LE CONTEXTE D'APPLICATION. ILS PEUVENT ÉGALEMENT DÉCOULER DE LA CONCEPTION ET DU PARAMÉTRAGE DU PROCESSUS D'APPRENTISSAGE.

AVOCAT



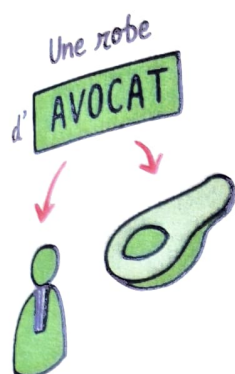
LES CORRÉLATIONS FALLACIEUSES OU INFONDÉES

LES MODÈLES D'IA REPOSENT SUR DES CORRÉLATIONS STATISTIQUES (entre une entrée et une sortie) QUI PEUVENT ÊTRE INFONDÉES, CONDUISANT À DES RÉSULTATS ERRONÉS MALGRÉ UNE PRÉCISION ANNONCÉE ÉLEVÉE. EN IA GÉNÉRATIVE, ON APPELLE PARFOIS CES ERREURS (à tort) "HALLUCINATIONS".



L'ABSENCE DE RAISONNEMENT CAUSAL

LES SYSTÈMES D'IA ACTUELS N'INTÈGRENT AUCUN RAISONNEMENT CAUSAL NI MÉCANISME DÉDUCTIF. CELA PEUT ENTRAÎNER DES CONTRADICTIONS ET RENDRE IMPOSSIBLE L'EXPLICATION DU RAISONNEMENT MENANT À UN RÉSULTAT, PUISQU'IL N'EXISTE PAS RÉELLEMENT (ce qu'on appelle l'effet "boîte noire").



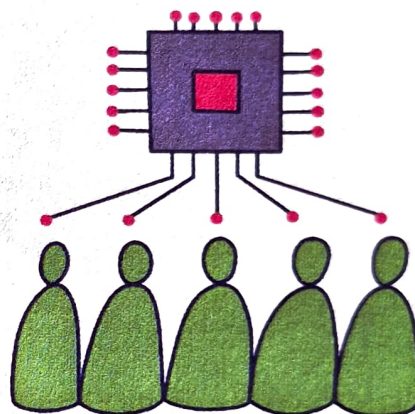
L'ABSENCE DE SÉMANTIQUE

EN APPRENTISSAGE SUPERVISÉ, LES CLASSES SONT DÉFINIES PAR DES ÉTIQUETTES HUMAINES, QUI ONT UN SENS, TANDIS QU'EN APPRENTISSAGE NON SUPERVISÉ, ELLES SONT REGROUPÉES PAR SIMILARITÉ SANS LIEN AVEC LEUR SIGNIFICATION RÉELLE.

POURQUOI PARLE-T-ON D'ENJEUX ÉTHIQUES DE L'IA ?

LES DÉFIS POSÉS PAR L'IA PROVIENNENT PRINCIPALEMENT DE SES MÉTHODES TECHNIQUES, DES CHOIX DE CONCEPTION, DES DONNÉES ET DES PROCESSUS D'ENTRAÎNEMENT.

CES DÉFIS, BIEN QU'ILS SEMBLERENT RELEVER DE LA ROBUSTESSE TECHNIQUE, DEVIENNENT ÉTHIQUES LORSQU'ILS CONCERNENT DES DÉCISIONS LIÉES AUX ÊTRES HUMAINS, COMME DANS LA RECONNAISSANCE D'IMAGES OU DES DOMAINES SENSIBLES TELS QUE LE RECRUTEMENT, LA MÉDECINE ET LA JUSTICE.

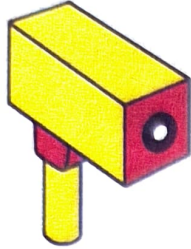


L'IA VISE À SIMULER UNE INTELLIGENCE HUMAINE, CAPABLE DE RÉSOUDRE DES PROBLÈMES COMPLEXES ET D'UTILISER LE LANGAGE. CET OBJECTIF SOULÈVE DES QUESTIONS PHILOSOPHIQUES SUR LA NATURE DE L'INTELLIGENCE HUMAINE ET DES DILEMMES ÉTHIQUES SUR L'AUTONOMIE ET LA DIGNITÉ HUMAINES :

Si l'intelligence humaine peut être réduite à des calculs algorithmiques, cela remet-il en cause l'identité humaine ?

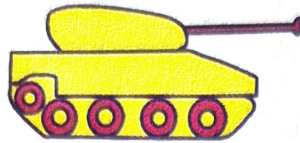
Les robots indiscernables des humains devraient-ils bénéficier d'un statut moral équivalent ?

OUTRE CES RÉFLEXIONS THÉORIQUES, L'IA SOULÈVE DES QUESTIONS ÉTHIQUES CONCRÈTES LIÉES À SON UTILISATION :



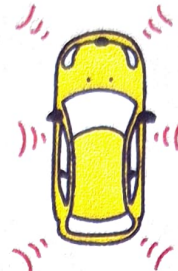
SURVEILLANCE DE MASSE

Atteinte à la démocratie et aux droits humains



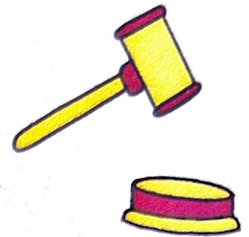
ARMES AUTONOMES

Incompatibilité de l'IA avec le droit humanitaire international



VÉHICULES AUTONOMES

IA encore immature



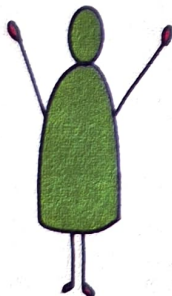
JUSTICE

Manque de discernement moral problématique

ENFIN, L'IMPACT DE L'IA SUR L'EMPLOI RESTE UNE QUESTION CENTRALE, MAIS IL S'AGIT DAVANTAGE D'UN ENJEU SOCIÉTAL QUE D'UNE QUESTION ÉTHIQUE PURE.

QUELS GRANDS PRINCIPES ÉTHIQUES POUR L'IA ?

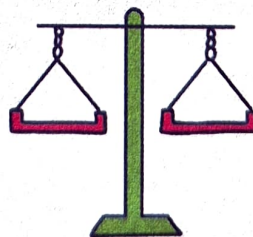
LES PRINCIPES DE LA BIOÉTHIQUE ONT INFLUENCÉ L'ÉTHIQUE DES TECHNOLOGIES, NOTAMMENT EN IA. EN 2019, LE GROUPE D'EXPERTS EN IA DE LA COMMISSION EUROPÉENNE (HLEG-AI) A DÉFINI QUATRE PRINCIPES ÉTHIQUES POUR UNE IA DE CONFIANCE :



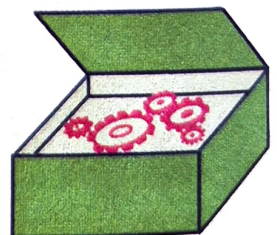
RÉSPÉCT DE L'AUTONOMIE HUMAINE



PRÉVENTION DE TOUTE ATTEINTE



ÉQUITÉ ET JUSTICE



EXPLICABILITÉ

QUELLES SONT LES QUESTIONS ÉTHIQUES SOULEVÉES PAR L'IA ?

CES PRINCIPES SE TRADUISENT DE MANIÈRE PLUS CONCRÈTE PAR SEPT EXIGENCES :

- ✓ Préservation de l'action humaine et du contrôle humain
- ✓ Robustesse technique et sécurité
- ✓ Respect de la vie privée et gouvernance des données
- ✓ Transparence (traçabilité, explicabilité, communication)
- ✓ Diversité, non-discrimination et équité
- ✓ Bien-être sociétal et environnemental
- ✓ Responsabilité et redevabilité (accountability)

LE MOT DE LA FIN

EN 2024, L'UNION EUROPÉENNE A ADOPTÉ UNE RÉGLEMENTATION SUR L'IA QUI REPOSE SUR L'ÉVALUATION DES RISQUES QU'ELLE PRÉSENTE POUR L'INTÉGRITÉ, LA SANTÉ, LA SÉCURITÉ ET LES DROITS FONDAMENTAUX.

LES SYSTÈMES SONT CLASSÉS EN QUATRE NIVEAUX DE RISQUE, DÉTERMINANT LES EXIGENCES RÉGLEMENTAIRES :

BAS RISQUE = SOUMIS À AUCUNE EXIGENCE PARTICULIÈRE	RISQUE MOYEN = DOIVENT GARANTIR LA TRANSPARENCE	HAUT RISQUE = NÉCESSITENT UNE CONFORMITÉ ET UNE CERTIFICATION	RISQUE INACCEPTABLE = INTERDITS
---	--	--	--

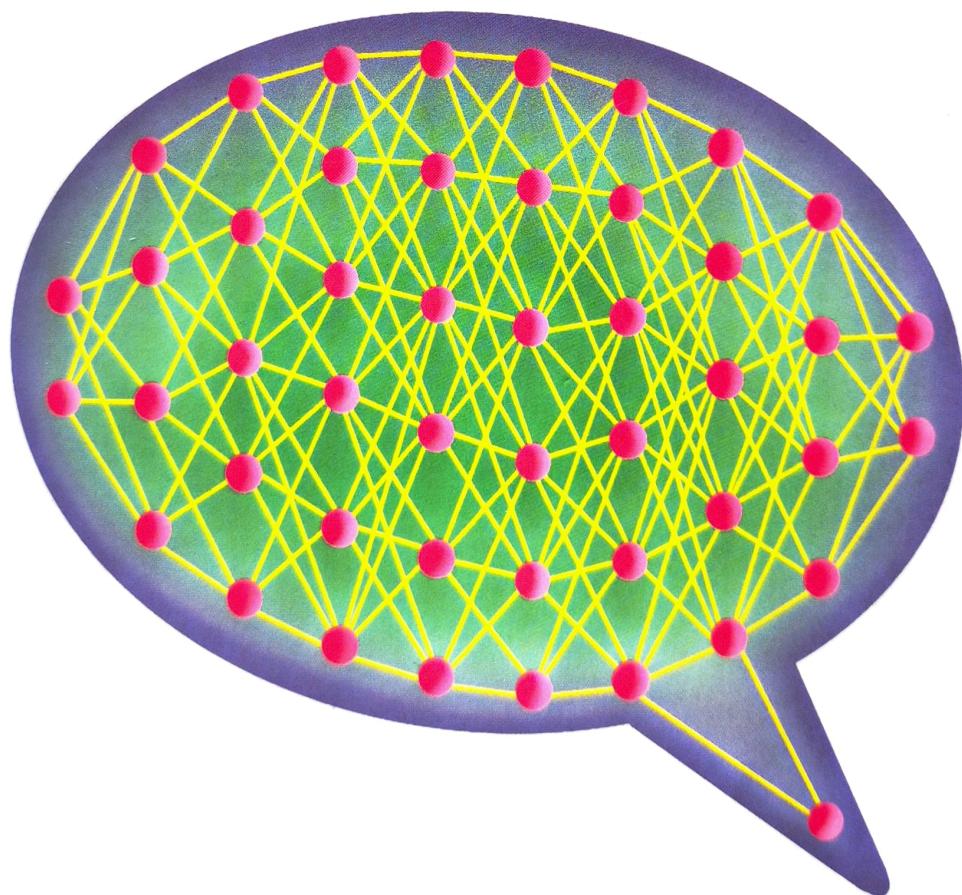
OLIVIER CAPPÉ

CLAIRE MARC

TOUT COMPRENDRE

(OU PRESQUE)

SUR L'INTELLIGENCE ARTIFICIELLE



CNRS EDITIONS

L'intelligence artificielle (IA) est aujourd'hui évoquée partout, tout le temps, tantôt comme notre planche de salut, tantôt comme une créature menaçant l'humanité. On a en tout cas le sentiment d'avoir déjà tout entendu sur l'IA. Mais cette impression est trompeuse, parce qu'elle laisse de côté la technologie sous-jacente. La comprendre est pourtant essentiel, pour se représenter clairement ce que l'on peut attendre de ces nouveaux outils, désormais largement diffusés et employés dans divers domaines, comme l'analyse et la génération d'images, le contrôle de robots ou la détection précoce de facteurs de risque dans le domaine de la santé.

Acteur majeur de la recherche en France, le CNRS a souhaité partager les connaissances de chercheurs et chercheuses concernant les concepts et techniques utilisés en intelligence artificielle. D'où vient l'IA et en quoi consiste-t-elle ? Quelles sont les méthodes mobilisées et les ressources nécessaires pour la faire fonctionner ? Quelles en sont les multiples utilisations ? Quelles sont ses limites, ses risques, et les défis qu'elle soulève ?

Sous l'apparente légèreté du graphisme, cet ouvrage couvre ces enjeux fondamentaux, et rend compte de l'essentiel des connaissances, afin que chacun et chacune puisse s'informer et participer aux débats.

Olivier Cappé est directeur de recherche CNRS au Département d'informatique de l'ENS-PSL. En 2013, il a reçu le Grand prix de l'Académie des sciences pour ses travaux sur le traitement statistique des signaux.

Claire Marc est médiatrice scientifique et facilitatrice graphique. Elle est co-autrice, à CNRS Éditions, de Tout comprendre (ou presque) sur le climat (2022); Tout comprendre (ou presque) sur la biodiversité (2023); Tout comprendre (ou presque) sur l'eau (2024).

19 € prix valable en France
ISBN : 978-2-271-15424-8



9 782271 154248

dl UN LIVRE A
LE MÊME PRIX
PARTOUT

www.cnrseditions.fr

PRÉSENTATION DES AUTEURS

OLIVIER CAPPÉ - CNRS

DÉPARTEMENT D'INFORMATIQUE DE L'ENS - PSL (DI ENS)

Directeur de recherche CNRS au Département d'informatique de l'ENS - PSL (DI ENS), Olivier Cappé conduit des recherches sur l'apprentissage automatique au sein du Centre sciences des données (CSD) de l'Ecole normale supérieure (ENS - PSL). Il est également Professeur attaché en IA de l'Université Paris Sciences et Lettres (PSL) où il est notamment co-responsable de la seconde année de la formation de master IASD (Intelligence Artificielle, Systèmes, Données). Distingué par un Grand prix de l'Académie des sciences en 2013 pour ses travaux sur le traitement statistique des signaux, il a également exercé les fonctions de Directeur adjoint scientifique de CNRS Sciences informatiques de 2017 à 2023.

CLAIRE MARC - MÉDUSE COMMUNICATION

MÉDIATRICE SCIENTIFIQUE

FACILITATRICE GRAPHIQUE

Titulaire d'un master en océanographie et d'un master de médiation scientifique, Claire Marc a fait le constat qu'un immense fossé séparait les scientifiques du grand public. Elle a donc créé son agence de communication scientifique Méduse Communication afin que chaque citoyen puisse se réapproprier les sciences. Comment ? En accompagnant les professionnels des sciences dans la communication autour de leurs travaux pour valoriser leurs résultats auprès du public. Sa spécialité : réaliser des synthèses visuelles d'articles scientifiques.

SOMMAIRE

AVANT-PROPOS	6
Question 1 : QUI A INVENTÉ L'IA ?	9
Question 2 : EXISTE-T-IL UNE SEULE FAÇON DE FAIRE DE L'IA ?	15
Question 3 : D'OÙ VIENT L'IA QUE L'ONT CONNAÎT ACTUELLEMENT ?	21
Question 4 : QU'EST-CE QUE L'APPRENTISSAGE AUTOMATIQUE ?	27
Question 5 : QU'EST-CE QU'UN RÉSEAU DE NEURONES ?	33
Question 6 : COMMENT FONCTIONNE UN GRAND MODÈLE DE LANGUE ?	39
Question 7 : DE QUEL MATÉRIEL A-T-ON BESOIN POUR FAIRE DE L'IA ?	47
Question 8 : QU'EST-CE QUE L'IA GÉNÉRATIVE ?	53
Question 9 : L'IA VOIT-ELLE DÉSORMAIS MIEUX QUE NOUS ?	59
Question 10 : L'IA PEUT-ELLE CONTRIBUER À LA CRÉATION ARTISTIQUE ?	65
Question 11 : L'IA COMPREND-ELLE TOUTES LES LANGUES ?	71
Question 12 : L'IA PEUT-ELLE APPRENDRE PAR ELLE-MÊME ?	77
Question 13 : L'IA PEUT-ELLE PILOTER DES ROBOTS ?	83
Question 14 : L'IA PEUT-ELLE AMÉLIORER NOTRE SANTÉ ?	89
Question 15 : L'IA PEUT-ELLE ÊTRE SOCIALE ?	95
Question 16 : L'IA PEUT-ELLE OBTENIR UN PRIX NOBEL ?	101

Question 17 : L'IA PEUT-ELLE ÊTRE INDUITE EN ERREUR ?	107
Question 18 : L'IA EST-ELLE RESPECTUEUSE DE LA VIE PRIVÉE ?	113
Question 19 : L'IA AMPLIFIE-T-ELLE LES BIAIS ?	119
Question 20 : QUELLES SONT LES QUESTIONS ÉTHIQUES SOULEVÉES PAR L'IA ?	125
PRÉSENTATION DES AUTEURS	130
REMERCIEMENTS	132