



**UNIVERSITÉ
DE GENÈVE**

CENTRE UNIVERSITAIRE
D'INFORMATIQUE

Intelligence artificielle et éthique

L'éthique de l'IA selon la recommandation de l'UNESCO : que peut-on en déduire pour les institutions d'héritage culturel ?

Dre Viola Krebs

Assemblée annuelle du 10 mars 2025
Zentral- und Hochschulbibliothek ZHB, Luzern

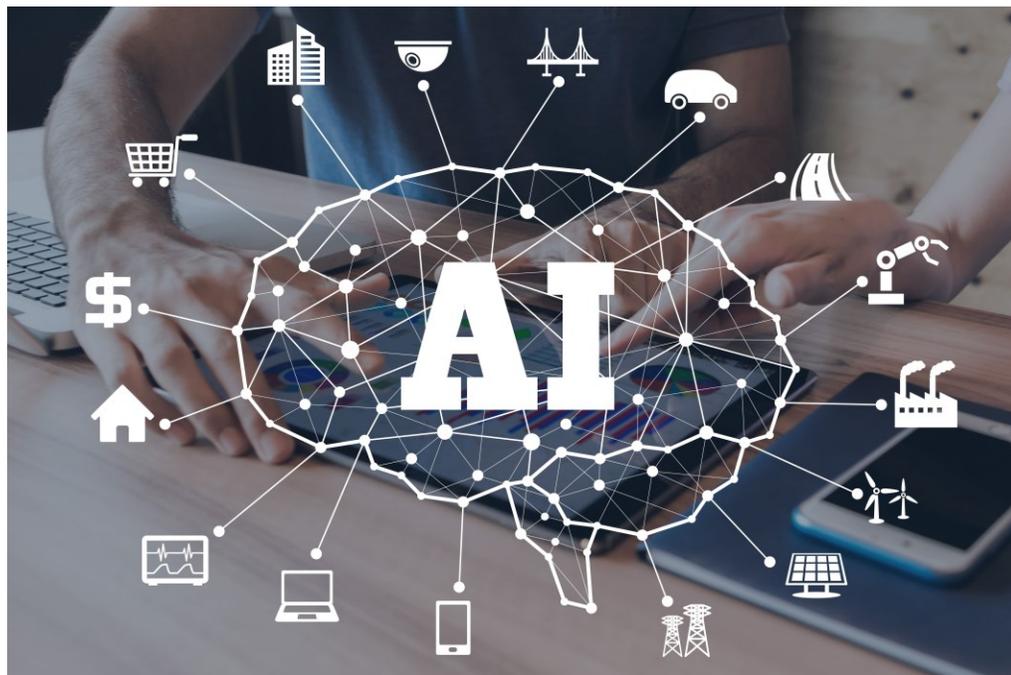
Dre Viola Krebs

- Professeur et formatrice dans différents établissements d'enseignement supérieur en technologies numériques et IA
- Collaboratrice scientifique au Centre Universitaire d'Informatique (CUI), Université de Genève
- Membre associé du LISEC et membre de la Chaire UNESCO de l'Université de Strasbourg
- Auditrice pour différentes normes de qualité dans le domaine de l'éducation (QSC 2.0, eduQua 2021, ISO 21001, ISO 9001, In-Qualis)
- Fondateur d'ICV, une organisation à but non lucratif œuvrant dans le domaine de la communication
- Docteur en Sciences de l'Information et de la Communication, Université de Strasbourg



Email : viola.krebs@unige.ch

Définition de l'Intelligence Artificielle



L'IA – ou Intelligence Artificielle – fait référence au développement de systèmes informatiques et d'algorithmes capables d'effectuer des tâches nécessitant généralement l'intelligence humaine, telles que l'apprentissage, la résolution de problèmes et la prise de décision.

Définition de l'éthique



Du latin **ethicus** « moral »,
latin **ethica** « morale »,
lui-même issu du grec
ancien ἠθικός, **êthikós**

Définition de l'éthique



L'éthique fait référence à ce qui est acceptable ou inacceptable, au sein d'une société ou d'une communauté, en matière de comportement – pas obligatoirement d'un point de vue juridique mais d'un point de vue humain ou culturel.

-- *UNESCO*

<https://www.unesco.org/fr/artificial-intelligence/recommandation-ethics>

Un peu d'histoire...

“What matters is not
always what you can
obtain from society,
but how you
participate in it”

*Tim Berners-Lee, créateur
du World Wide Web*



Un peu d'histoire...

« Ce qui compte, ce n'est pas toujours ce qu'on peut obtenir de la société, mais comment on y participe »

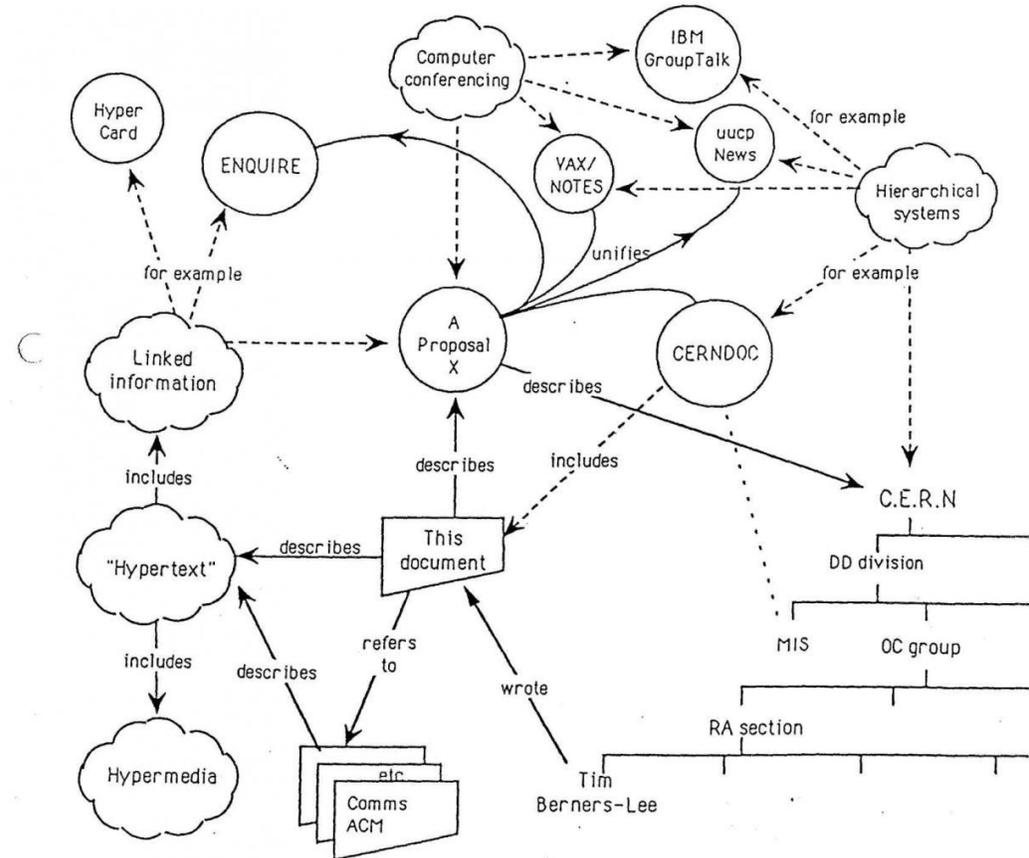
*Tim Berners-Lee, créateur
du World Wide Web*



http et le World Wide Web

- La proposition http a été publiée en 1989
- Le WorldWideWeb est encore une invention relativement récente
- Depuis, elle a transformé le monde d'une manière que ses fondateurs n'auraient pas pu prévoir ou imaginer.

<https://home.cern/science/computing/birth-web/short-history-web>



http et le World Wide Web



LIZARD GLOBAL

web 1.0

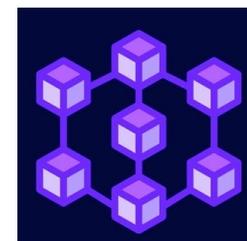
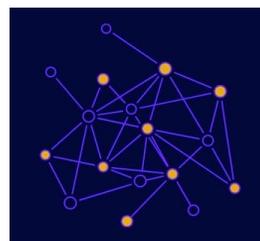
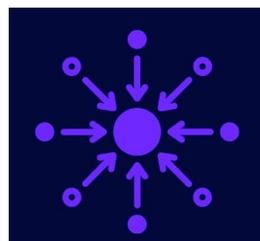
read-only
decentralized

web 2.0

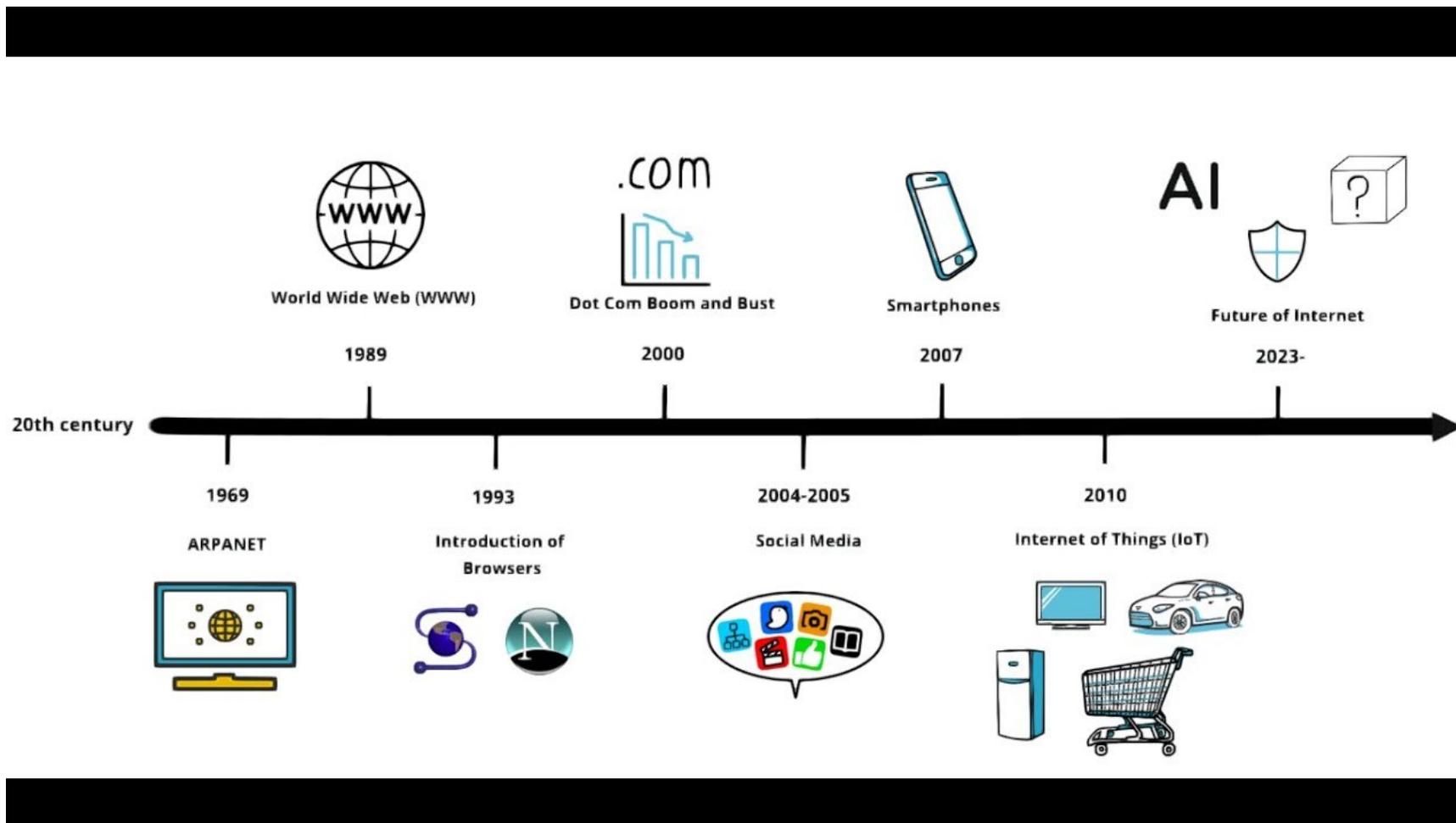
participatory
centralized

web 3.0

no intermediary
decentralized



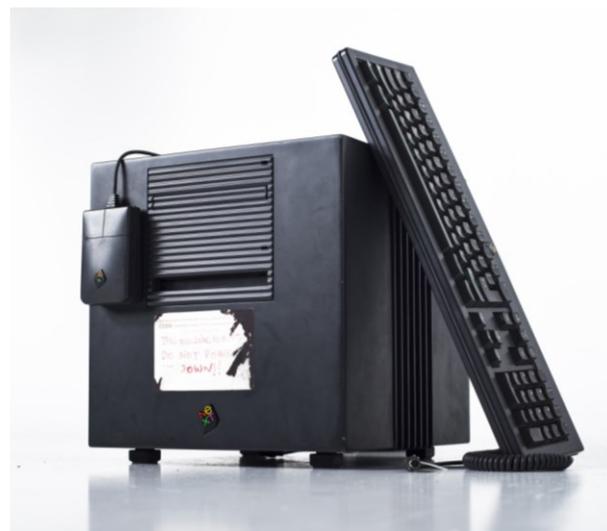
Nombre d'utilisateurs du Web



Evolution des technologies (hardware)



1969 : UNIX™ Ken
Thompson and Dennis
Ritchie @ Bell Labs



1990 : NeXT machine used
by Tim Berners-Lee to
develop and run the first
WWW server, multimedia
browser and web editor



Première image non scientifique sur le Web



Cette image des Horribles Cernettes a été la première image photographique publiée sur le World Wide Web en 1992.

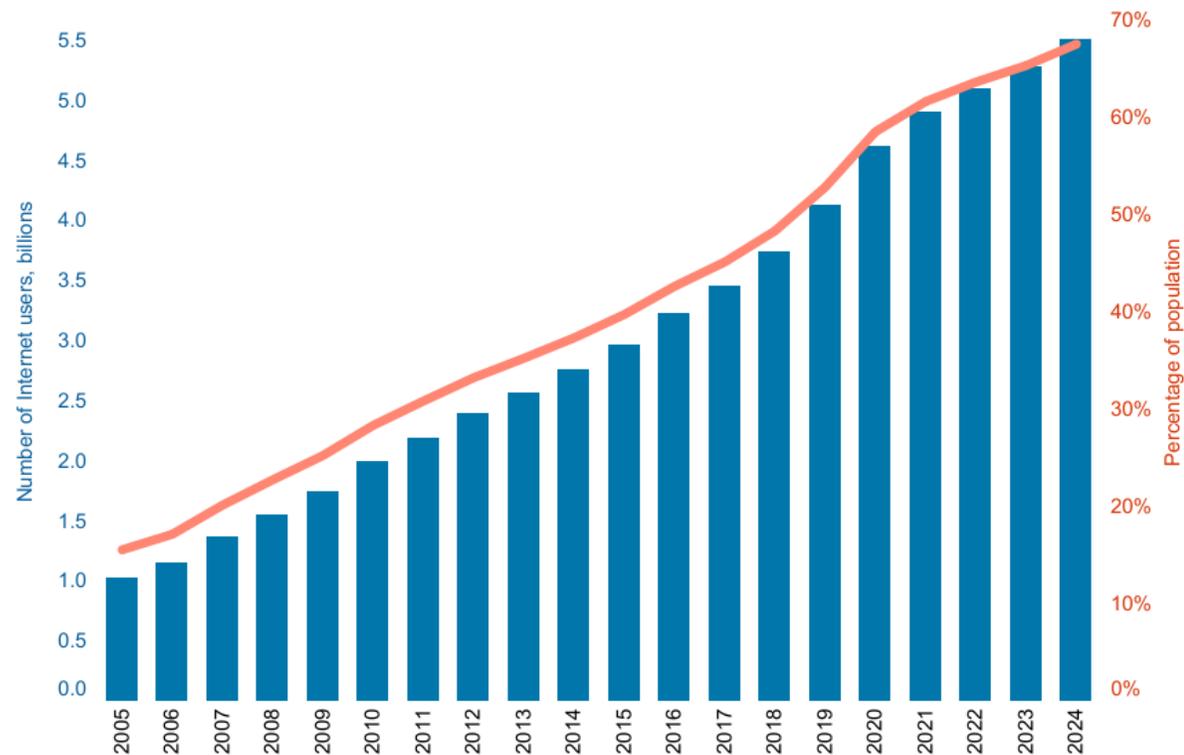


20 ans après... Photo prise au Hardronic Festival au CERN en 2012.

<https://www.bbc.com/future/article/20160224-the-unlikely-photo-that-kickstarted-the-social-internet>

Nombre d'utilisateurs du Web

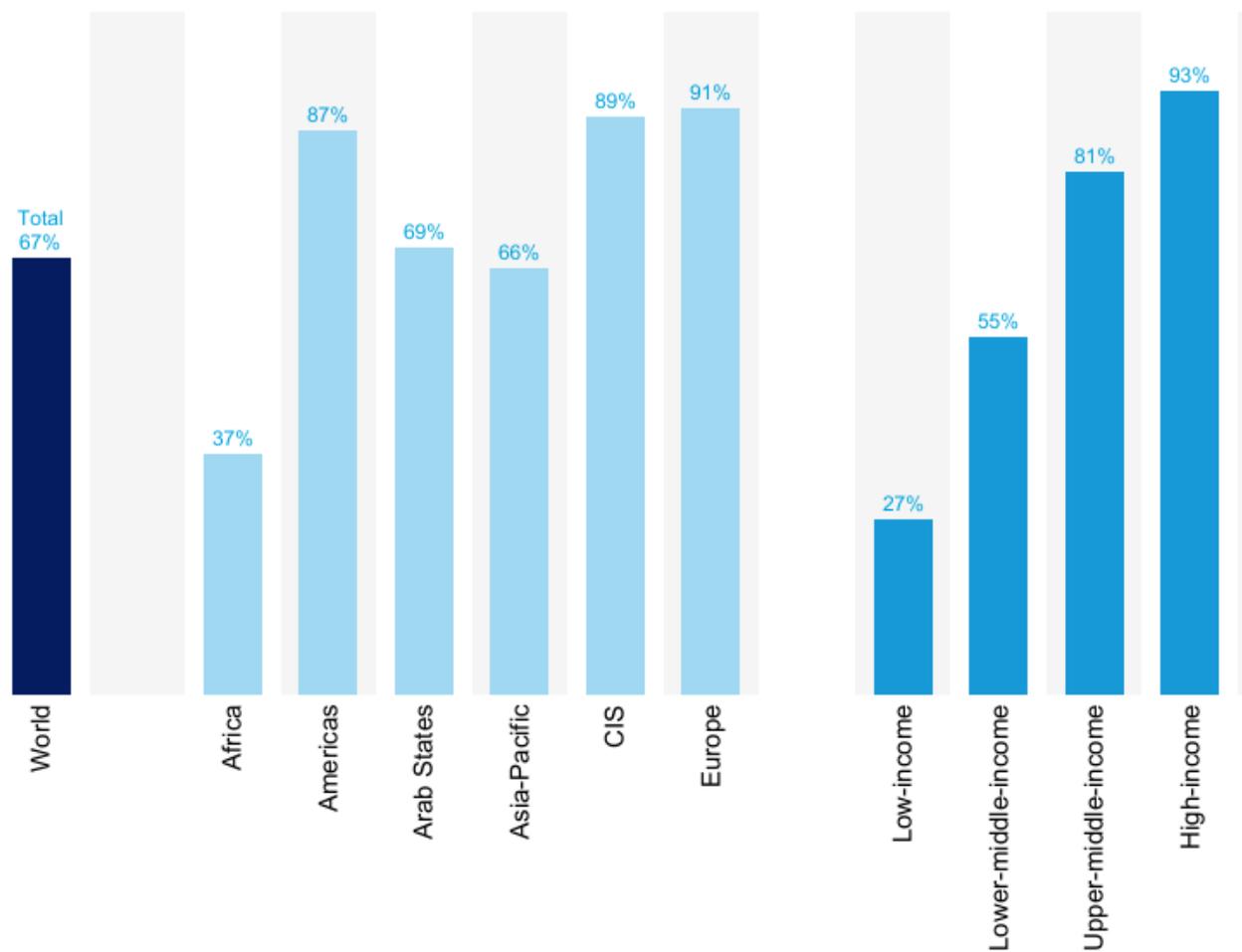
Individuals using the Internet



Source: ITU

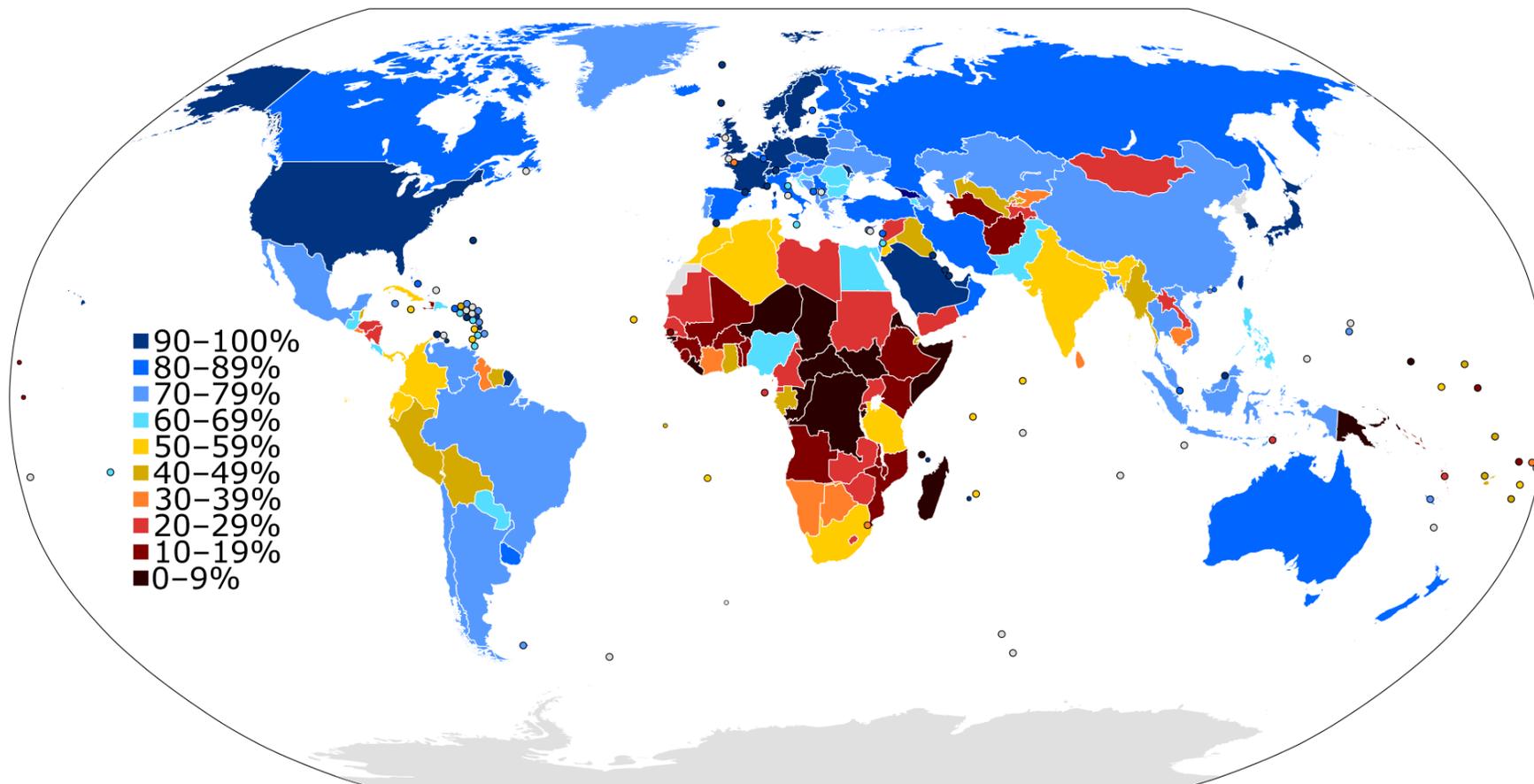
Nombre d'utilisateurs du Web

Percentage of individuals using the Internet by region, 2023



Source: ITU

Nombre d'utilisateurs du Web

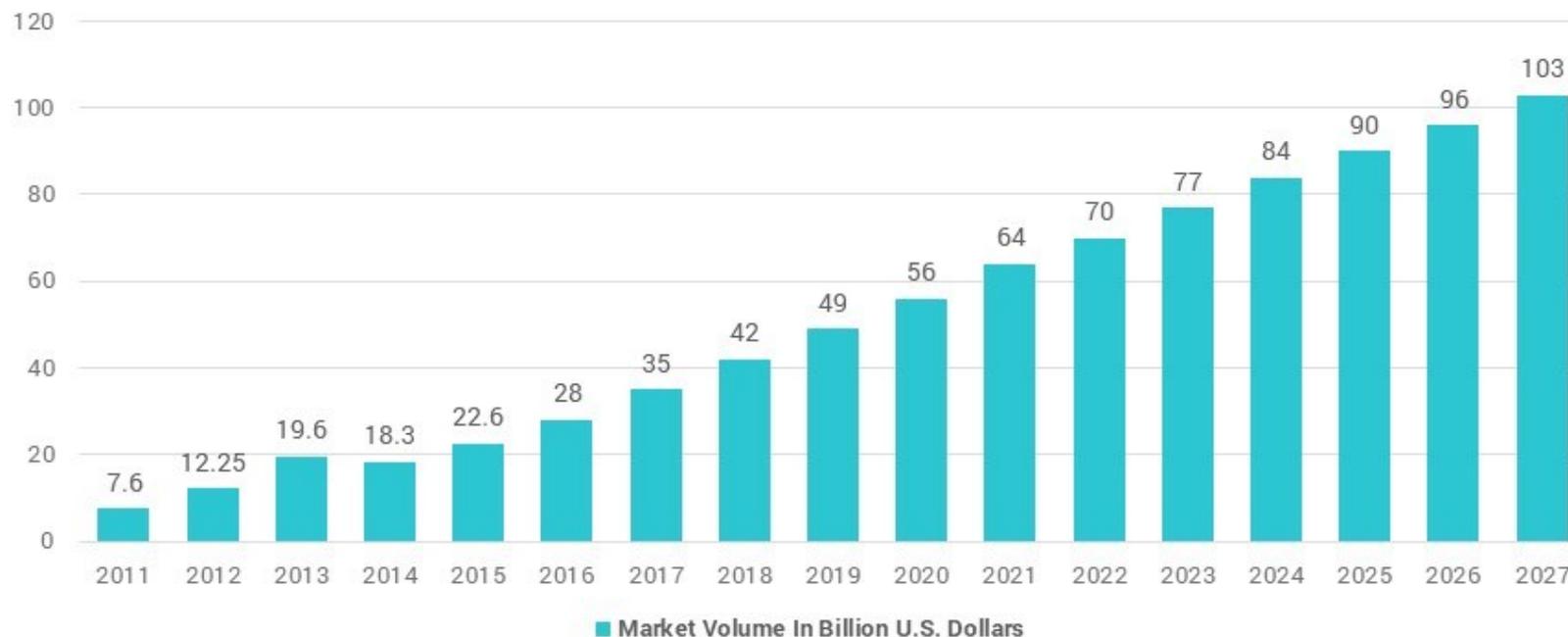


Source: 2023, International Telecommunication Union

30 ans après la création de http

Big Data Market Size Revenue Forecast Worldwide

Big data market size revenue forecast worldwide from 2011 to 2027 (In Billion U.S. Dollars)



30 ans après la création de http



2025 : la révolution numérique s'accélère, avec un monde où le Web et l'Intelligence Artificielle sont devenus des domaines de géopolitique stratégique, à la fois à des fins civiles et militaires.

Concepts anciens (de l'Antiquité au XXe siècle) :

- L'idée de créer des êtres artificiels dotés d'une intelligence semblable à celle de l'humain a des racines anciennes, apparaissant dans les mythes et les histoires.
- Des philosophes et des scientifiques, comme **Aristote** et **Léonard de Vinci**, ont envisagé la possibilité de créer des machines capables d'imiter les fonctions humaines.



Bref historique de l'IA...

Concepts anciens (de l'Antiquité au XXe siècle) :

Aristote

Fondements de l'éthique aristotélicienne

- Le bien suprême : eudaimonia (bonheur)
- Les vertus morales et intellectuelles

Pertinence pour l'IA

- Développement de machines « vertueuses »
- Accent sur l'apprentissage et la formation des intelligences artificielles



Bref historique de l'IA...

Concepts anciens (de l'Antiquité au XXe siècle) :

Léonard de Vinci et l'Éthique de la Création

Approche humaniste de Léonard de Vinci

- L'union de l'art et de la science
- L'observation et l'imitation de la nature

Pertinence pour l'IA

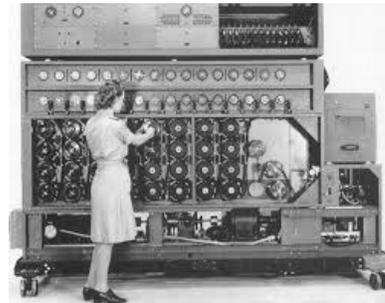
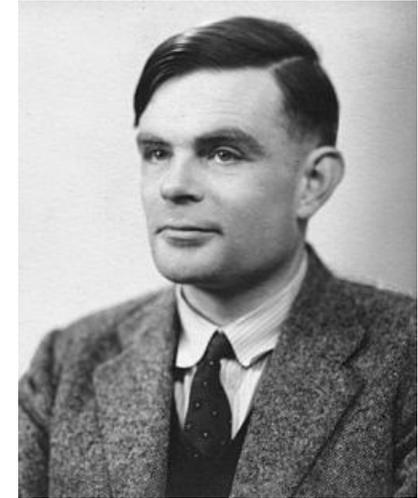
- Création d'IA inspirées de la nature
- Importance de l'esthétique et de l'harmonie dans les interfaces



Bref historique de l'IA...

Alan Turing (1912-1954, UK)

- Mathématicien anglais
- Informaticien
- Père de l'IA
- Fondements théoriques de l'informatique
- Travaillé pour déchiffrer les codes allemands pendant WW2



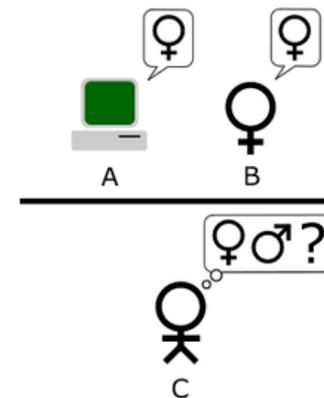
Imitation Game – Alan Turing



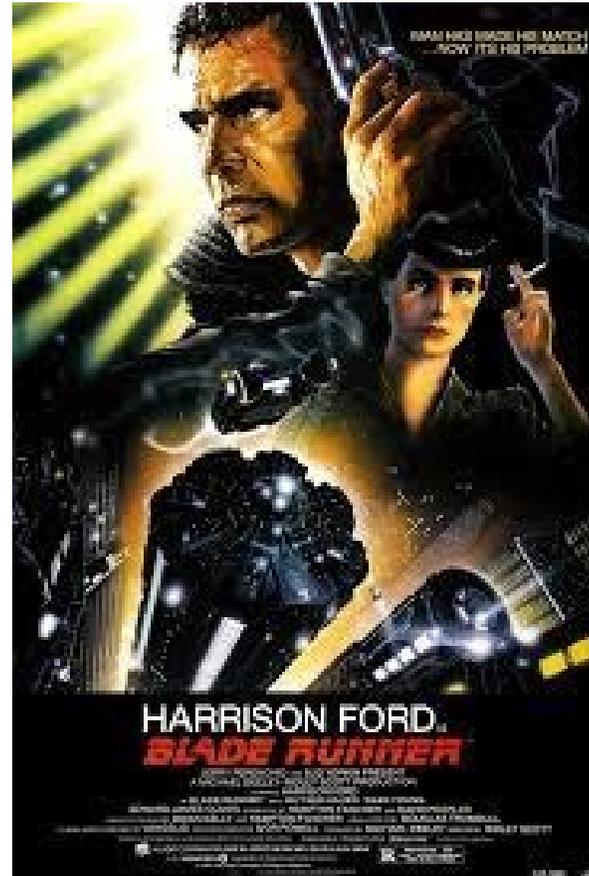
<https://www.youtube.com/watch?v=nuPZUUED5uk>

«Considérons un “digital computer D” avec une mémoire de taille adéquate, des calculs rapides et un programme adapté.»

Est-il possible pour D de jouer le rôle de A dans le jeu d'imitation de manière satisfaisante?»



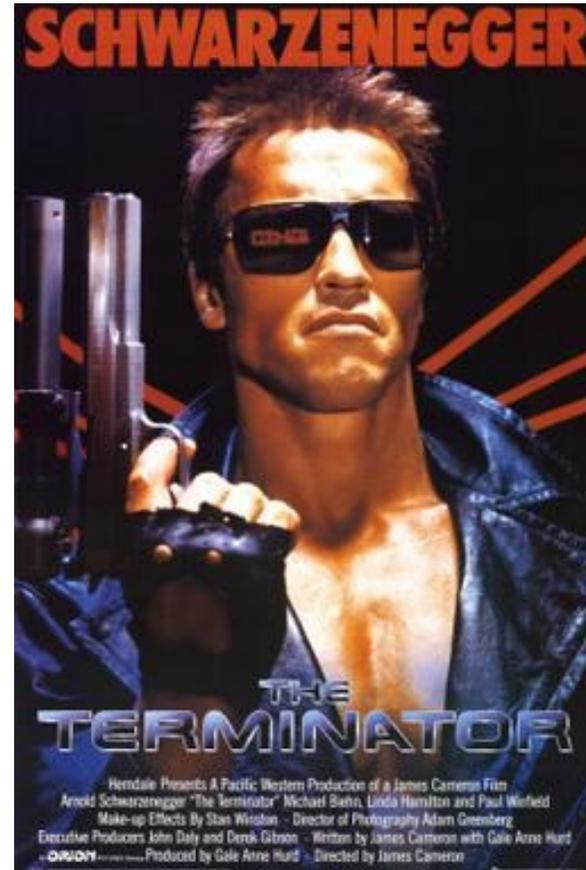
Bref historique de l'IA...



Blade Runner :

- Film de science-fiction réalisé par Ridley Scott, sorti en 1982
- Vaguement basé sur le roman de Philip K. Dick « Les androïdes rêvent-ils de moutons électriques ? »
- Blade Runner 2049 : lancé en 2019

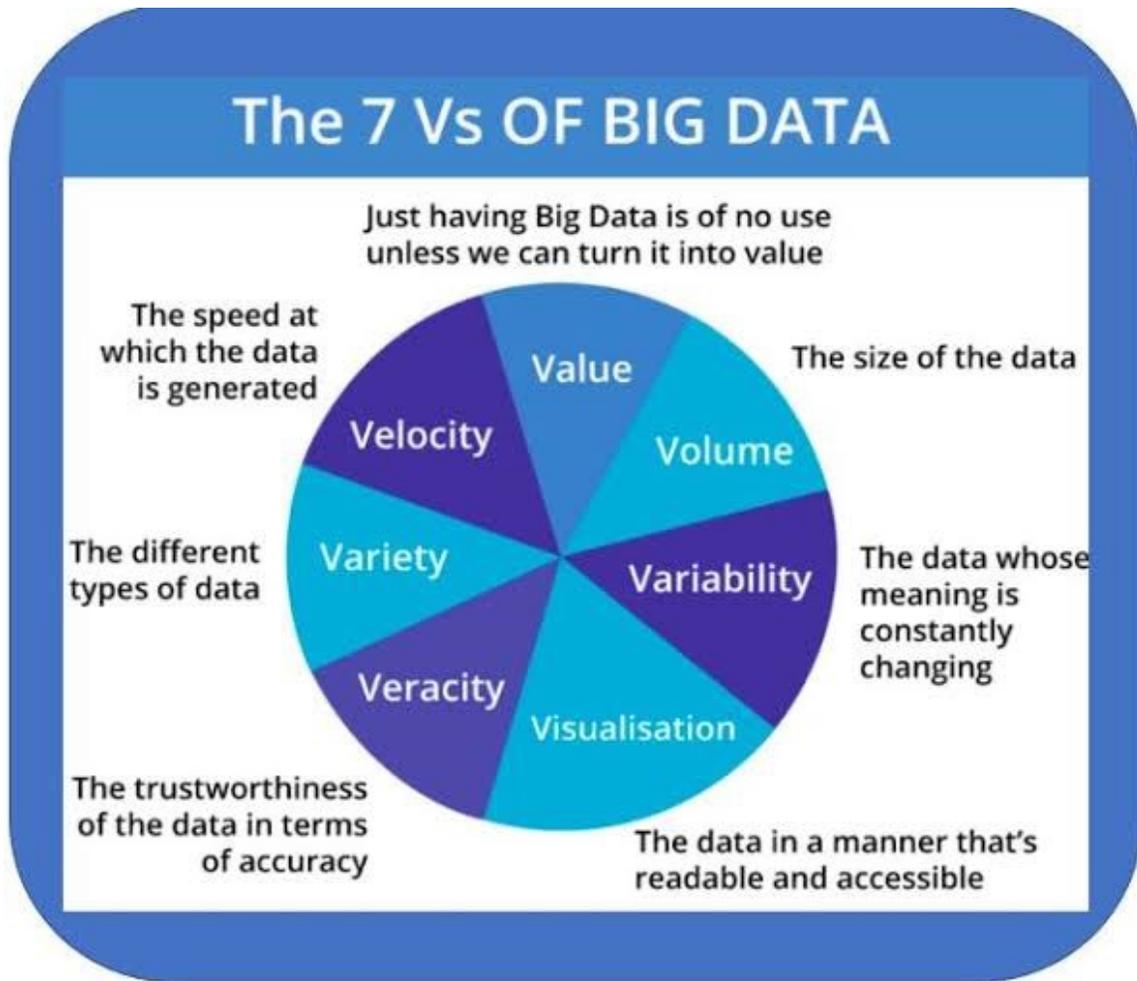
Bref historique de l'IA...



Terminator:

- Terminator est un film d'action de science-fiction américain de 1984 réalisé par James Cameron
- Arnold Schwarzenegger dans le rôle de Terminator, un assassin cybernétique envoyé dans le passé de 2029 à 1984 pour tuer Sarah Connor, dont le fils à naître sauvera un jour l'humanité de l'extinction par Skynet, une intelligence artificielle hostile dans un futur post-apocalyptique.

Bref historique de l'IA...



Montée du Big Data (années 2000 à aujourd'hui) :

- La disponibilité de grandes quantités de données, associée à des algorithmes améliorés, a conduit à des avancées significatives dans les applications de l'IA, en particulier dans des domaines tels que la reconnaissance d'images, le traitement du langage naturel et les systèmes de recommandations.

Bref historique de l'IA...

1950

Genèse

De l'IA et du premier cycle de battage médiatique

1983

Réapparition de l'IA entre 1983 et 2010

2010

Apprentissage profond (années 2010 à aujourd'hui) :

L'apprentissage profond obtient des résultats remarquables dans des tâches telles que la reconnaissance d'images et de parole.

Des entreprises comme Google, Facebook et OpenAI ont été à l'avant-garde en repoussant les limites des capacités de l'IA.

2012

Web et service web –

Applications Web App/API : Les applications Web sont de plus en plus axées sur l'expérience utilisateur, avec des interfaces utilisateur riches et réactives.

Services Web : intelligence artificielle (IA) et apprentissage automatique (ML) intégrés dans de nombreux services Web.

Applications mobiles

2018

Avenir

Des cerveaux, des esprits et des machines
2018-2035

2020

Paysage actuel (années 2020) :

L'IA est omniprésente dans la vie de tous les jours. Les discussions sur l'impact sociétal de l'IA se poursuivent.

Bref historique de l'IA...

Réseaux sociaux (2004 à aujourd'hui) :

- **Facebook** lancé en 2004 par Mark Zuckerberg, d'abord destiné aux étudiants de l'Université Harvard.
- **Twitter** (2006) a introduit le concept de microblogging, permettant aux utilisateurs de partager des messages courts en temps réel.
- **Instagram** (2010) a connu un succès rapide en se concentrant sur le partage de photos et de vidéos.
- L'émergence de plateformes comme **TikTok** (2016) a introduit le partage de vidéos courtes et créatives.
- **Facebook** a acquis **Instagram** (2012) et **WhatsApp** (2014), consolidant sa position dominante.

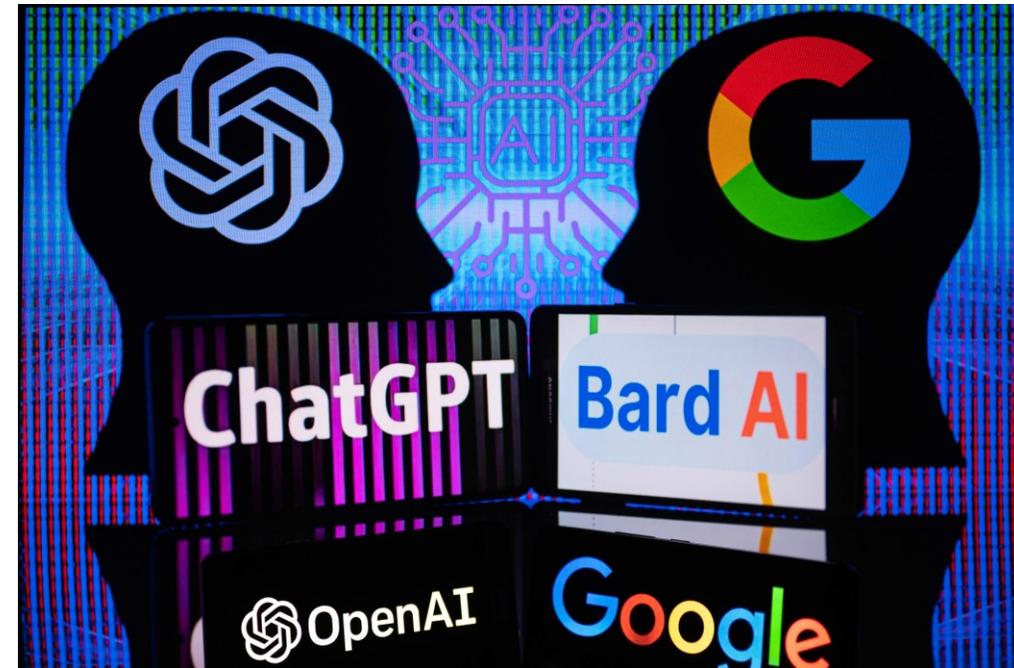


Bref historique de l'IA...

Big Tech...

(années 2010 à aujourd'hui) :

- Des entreprises comme Google, Facebook et OpenAI ont été à l'avant-garde en repoussant les limites des capacités de l'IA.



Bref historique de l'IA...

Sleepwalkers :

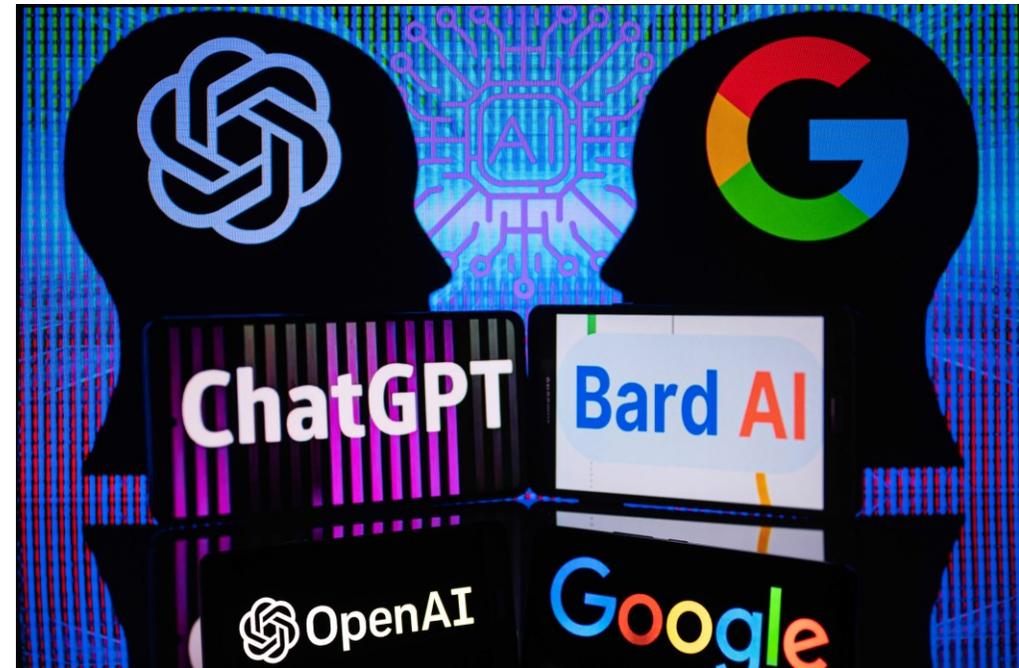
Comment l'IA transforme notre vie sans que nous le réalisons

Sleepwalkers vous emmène dans l'univers fascinant de l'intelligence artificielle, en scrutant son influence sur notre vie de tous les jours.

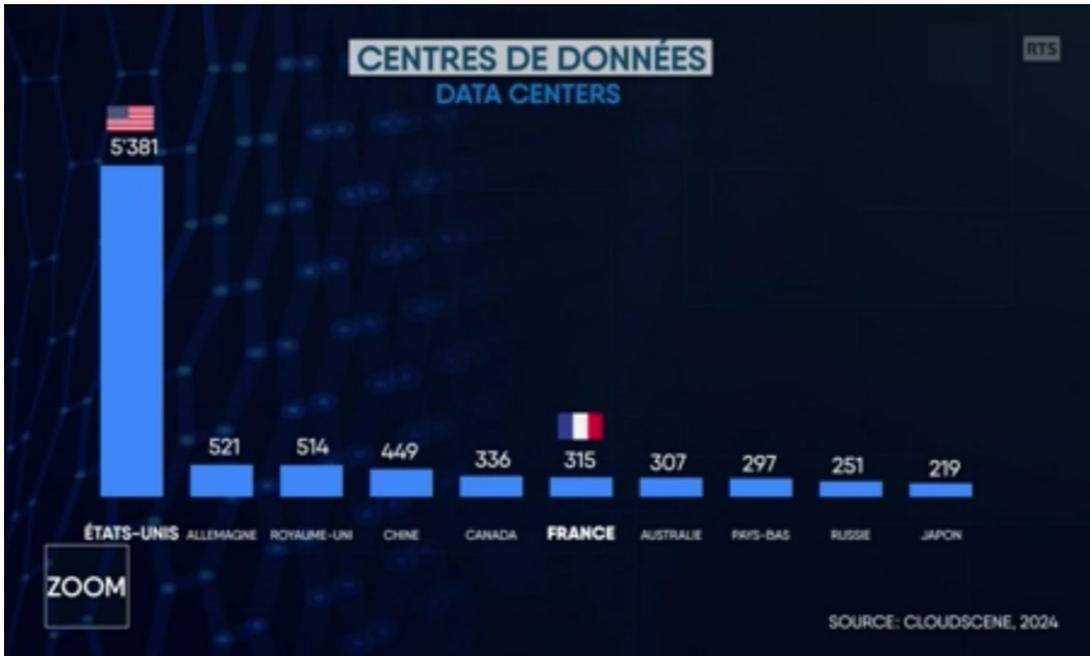
Ce podcast regorge d'histoires fascinantes, d'entretiens inspirants et d'analyses approfondies sur les répercussions de l'IA dans divers domaines tels que la santé, la sécurité, la créativité, l'amour et la politique.

Dimensions à double tranchant de l'IA, capable d'apporter des avancées significatives comme de poser d'importants dilemmes.

<https://oscar-black.com/blog/intelligence-artificielle/podcast-intelligence-artificielle/>



Bref historique de l'IA...



<https://www.rts.ch/emissions/geopolitis/2025/video/ia-la-grande-bataille-28804439.html>

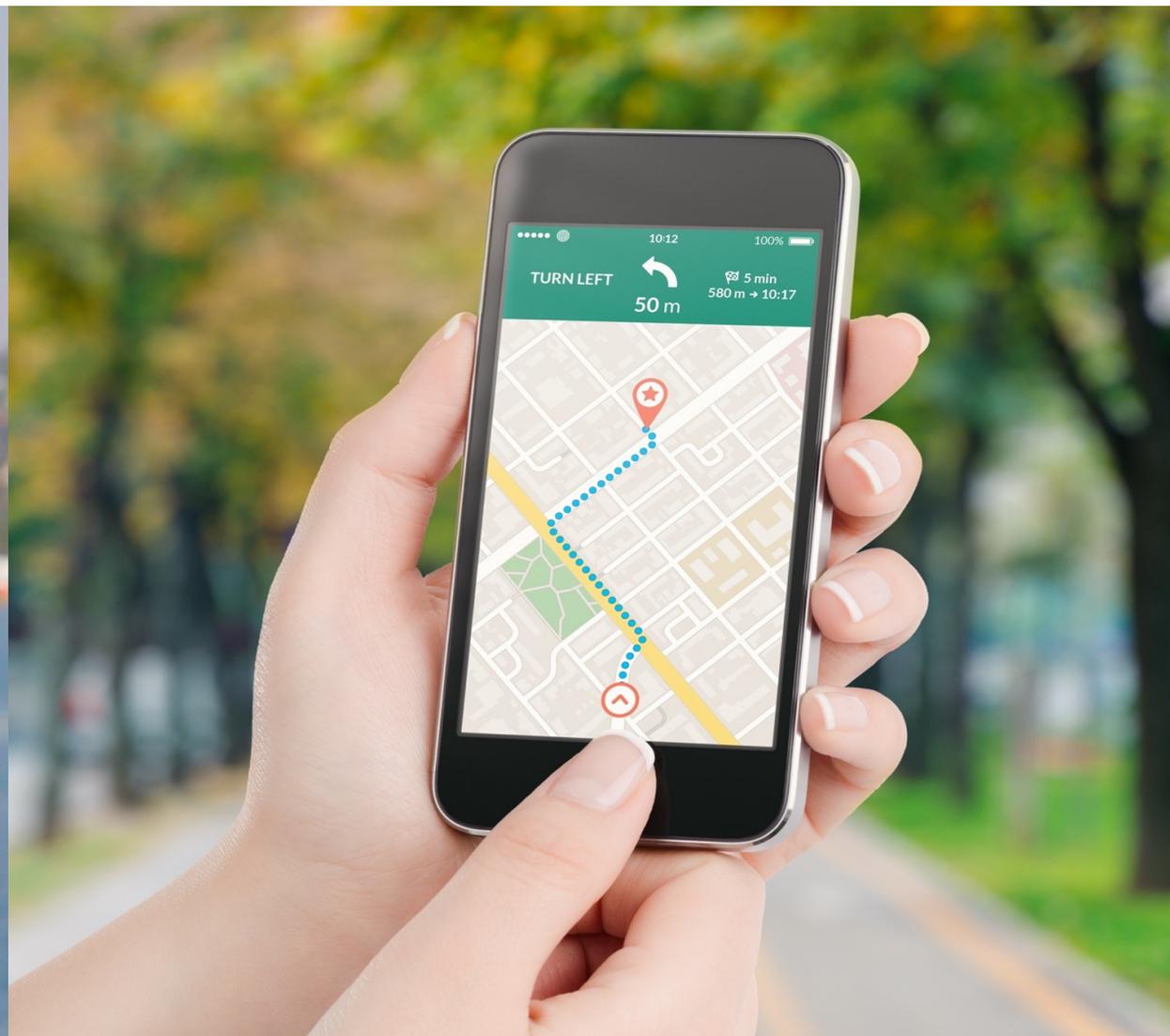
Bref historique de l'IA...



Paysage actuel (2025) :

- La guerre de l'information, info et infotox
- La dépendance au numérique
- L'accès aux ressources technologiques (développement du numérique et puces électroniques -> la Suisse n'est pas considérée comme un pays prioritaire par les Etats-Unis)

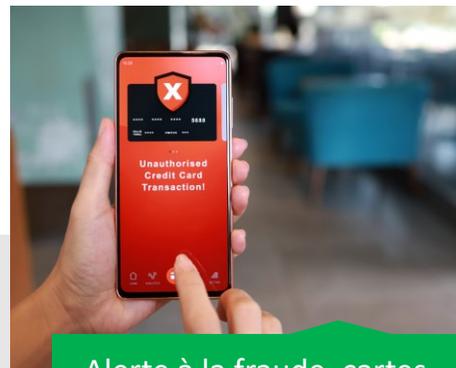
Intelligence humaine et Intelligence artificielle



Exemples de l'IA qui nous entoure



Uber



Alerte à la fraude, cartes
de crédit et bancaires



Online Shopping



Publicité en ligne basée
sur vos clics

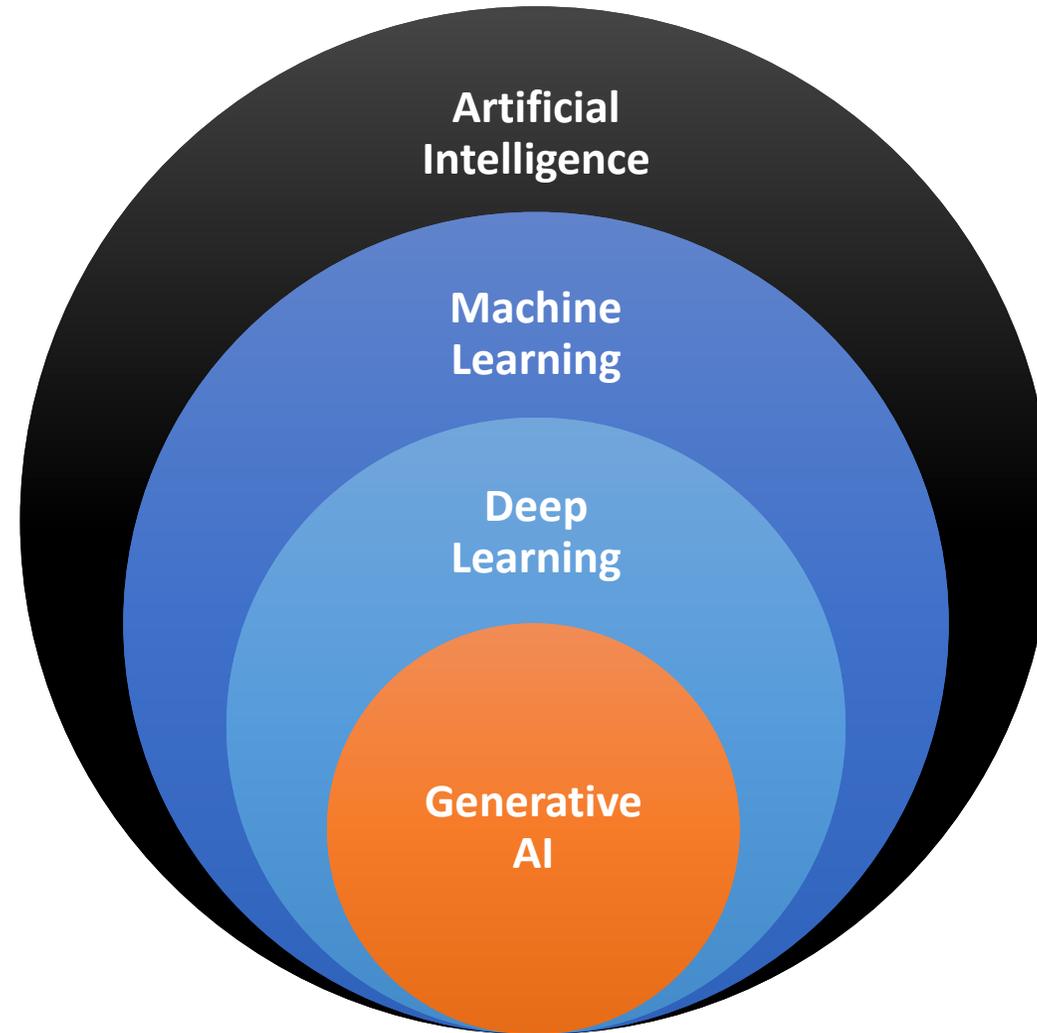


Politique : cartes de
circonscriptions électorales



Jeu d'échecs en ligne

Couches d'intelligence artificielle



Fondamentaux de l'IA

- **Apprentissage automatique** : Apprentissage supervisé, non supervisé et par renforcement ; Applications concrètes de l'apprentissage automatique ; Apprentissage automatique (Machine Learning) versus apprentissage profond (Deep Learning)
- **IA générative** : ChatGPT, génération de textes et des images
- **Réseaux de neurones** : Architectures d'apprentissage profond populaires (par exemple, CNN, RNN, GAN)
- **IA sémantique** : données liées, ontologie, raisonnement sémantique
- **Traitement du langage naturel (NLP)** : Applications de la PNL ; chatbots ; analyse des sentiments ; traduction automatique
- **IA distribuée** : systèmes multi-agents ; modélisation et ingénierie / développement



Bénéfices et Défis de l'IA



- Automatisation
- Disponibilité 24h/24 et 7j/7
- Économies de coûts
- Personnalisation
- Prise de décision améliorée
- Précision améliorée
- Évolutivité
- Analyses prédictives



- Qualité et biais des données
- Problèmes de confidentialité
- Risques de sécurité
- Conformité réglementaire
- Considérations éthiques
- Coûts initiaux élevés
- Manque de talents qualifiés
- Interopérabilité
- Déplacement d'emploi

IA et éthique

Comment distinguer le vrai du faux ?



IA et éthique

Comment distinguer le vrai du faux ?

RTS Info Sport Culture | Proche-Orient TV & Streaming Audio

Sciences-Tech. Publié le 16 janvier 2024 à 10:22 Partager

YouTube interdit les vidéos utilisant l'IA pour redonner vie à des victimes de crimes



YouTube interdit les vidéos utilisant l'IA pour redonner vie à des victimes de crimes / La Matinale / 2 min. / le 16 janvier 2024

<https://www.rts.ch/info/sciences-tech/14627609-youtube-interdit-les-vidéos-utilisant-lia-pour-redonner-vie-a-des-victimes-de-crimes.html>

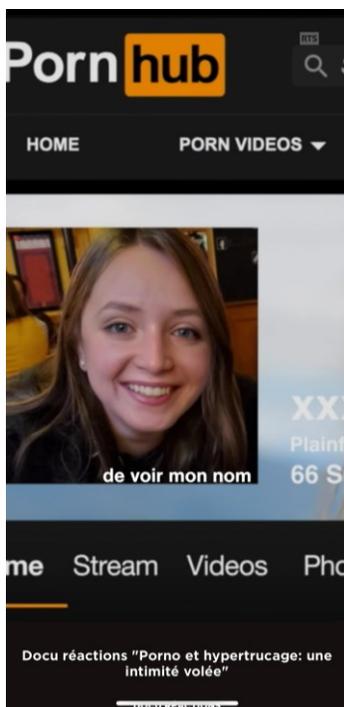
Il est désormais interdit de publier sur YouTube toute vidéo qui, grâce à l'intelligence artificielle (IA), redonne vie à des victimes d'affaires criminelles très médiatisées. La mesure, adoptée par la plateforme américaine, est entrée en vigueur mardi.

Les contenus qui font parler les morts cumulent des millions de vues sur TikTok et YouTube. Réalisées grâce à l'intelligence artificielle, ces vidéos peuvent être produites très facilement avec les dérives que cela entraîne, puisque même des enfants sont mis en scène.

C'est notamment le cas du petit Grégory qui, face caméra, raconte dans une vidéo les détails de sa mort à la première personne. "Un jour, j'ai disparu, ma maman m'a cherché partout, mais c'est dans la Vologne, un petit ruisseau près de chez moi, qu'ils m'ont retrouvé, sans vie", raconte ainsi une reconstitution du petit Grégory, décédé en 1984 dans des circonstances mystérieuses qui sont encore commentées aujourd'hui régulièrement dans les médias.

IA et éthique

Comment distinguer le vrai du faux ?



<https://www.rts.ch/play/tv/docu-reactions/video/docu-reactions-porno-et-hypertrucage-une-intimite-volee?urn=urn:rts:video:14908404>

Conséquences d'un Deepfake

News / World

British engineering giant Arup revealed as \$37 million deepfake scam victim

CNN By CNN | 12:25am May 18, 2024

[Tweet](#)

[Facebook](#)

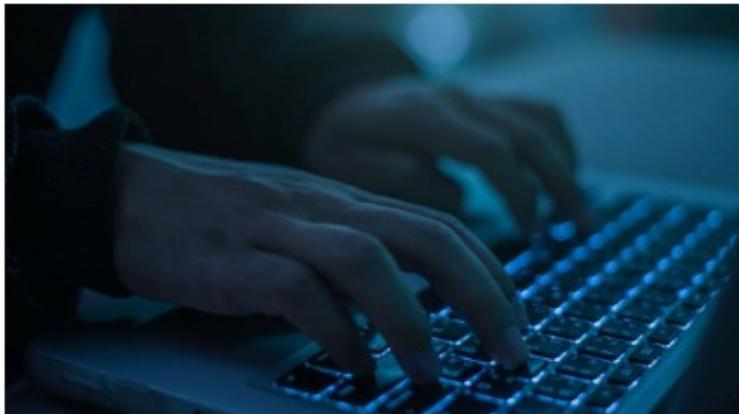
[Mail](#)

A British multinational design and engineering company behind world-famous buildings such as the Sydney Opera House has confirmed that it was the target of a deepfake scam that led to one of its Hong Kong employees paying out \$37 million (\$US25 million) to fraudsters.

A spokesperson for London-based Arup told CNN on Friday that it notified Hong Kong police in January about the fraud incident, and confirmed that fake voices and images were used.

"Unfortunately, we can't go into details at this stage as the incident is still the subject of an ongoing investigation. However, we can confirm that fake voices and images were used," the spokesperson said in an emailed statement.

[READ MORE: Family of US airman shot by police claim deputy went to wrong home](#)



British company Arup has confirmed it was the victim of a deepfake scam in Hong Kong boonchai/wednesday/Moment P/P/Getty Images/CNN (boonchai/wednesday/Moment P/P/Getty Images/CNN)

- Un comptable d'Arup, société qui a conçu l'Opera House de Sydney

IA

Accélérationnistes VS Doomers

☰ **DAZED** +

**Doomer vs
Accelerationist: the
two tribes fighting
for the future of AI**



Jude Law in A.I. Artificial Intelligence, 2001 (Film Still)

LIFE & CULTURE - FEATURE

**Chaos at OpenAI shines a spotlight
on a fierce debate about AI, and the
probability it will spell doom for the
human race**

24th November 2023

Text Thom Waite

20:54

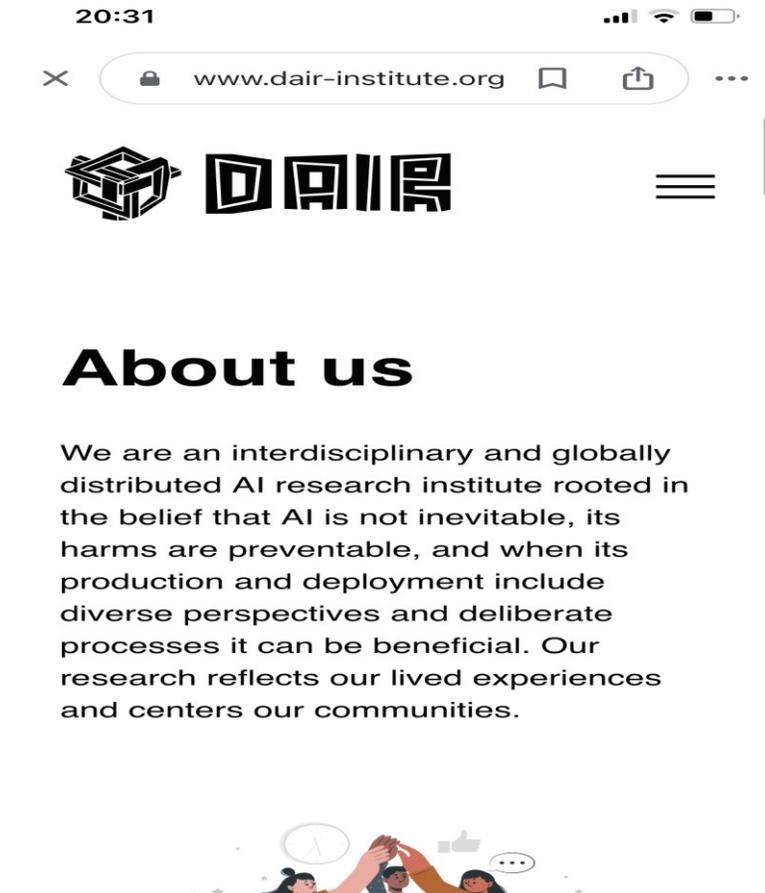


× activeacceleration.tech 📖 📄 ⋮

Effective Acceleration means
accepting the future.

The force of technocapitalistic progress is
inevitable. Technocapital is an inexorable
physical process.

Alternatives de résistance



Biais Culturels dans l'IA

- Préjugés ou distorsions dans les données d'entraînement
- Réflexion des stéréotypes et normes culturelles
- Impact sur les décisions et les résultats de l'IA



Types de Biais Culturels



Types de Biais Culturels



Biais de représentation

- Sous-représentation de certaines populations dans les données
- Sur-représentation de groupes dominants

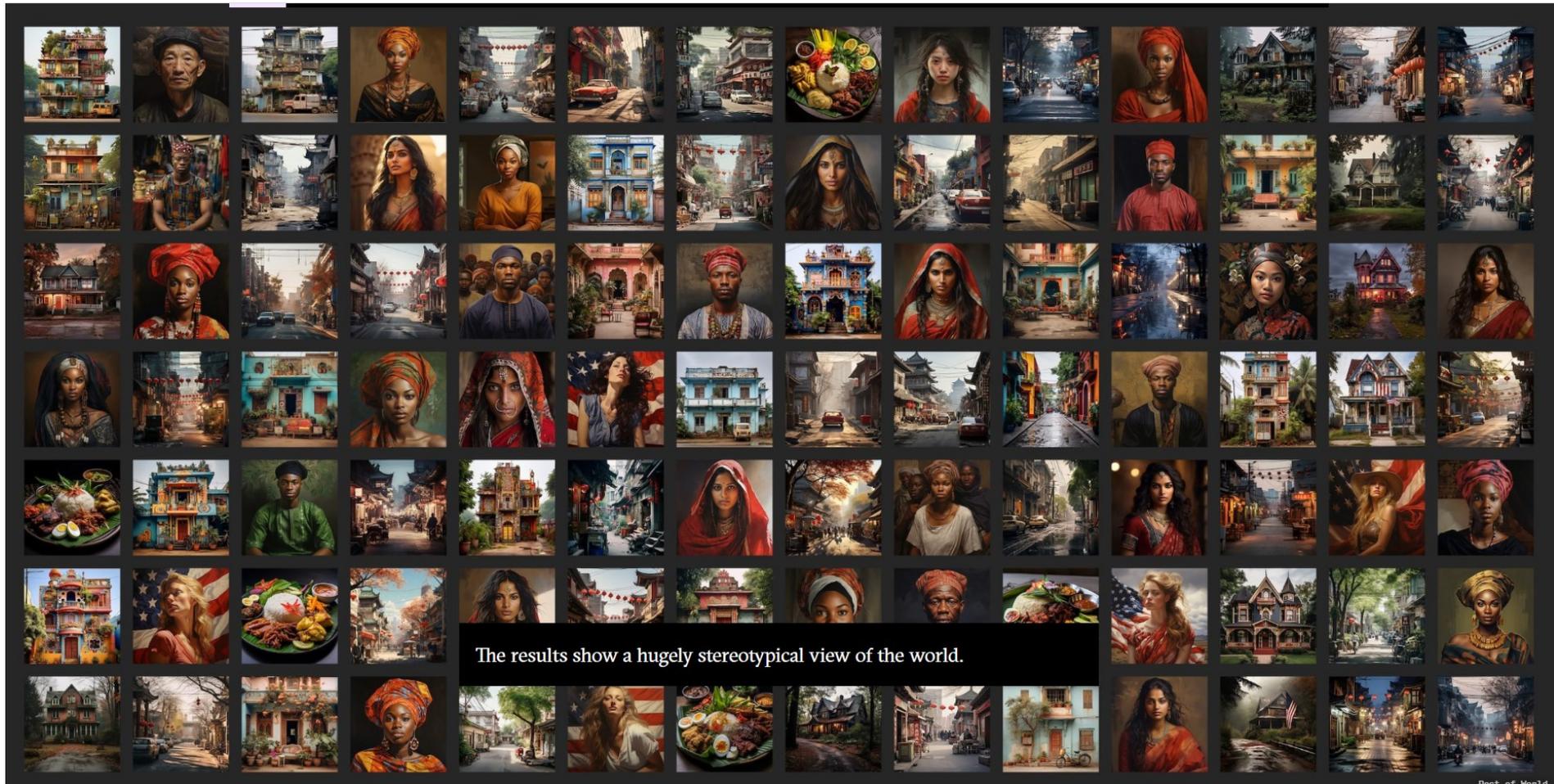
Biais de traitement

- Décisions discriminatoires basées sur des caractéristiques culturelles
- Prédicteurs influencés par des normes culturelles spécifiques

Biais d'évaluation

- Normes de performance basées sur des critères culturels
- Évaluation injuste des individus en fonction de leur origine culturelle

Types de Biases Culturels



<https://restofworld.org/2023/ai-image-stereotypes/>

Exemples de Biais Culturels

Biais dans les systèmes de recrutement

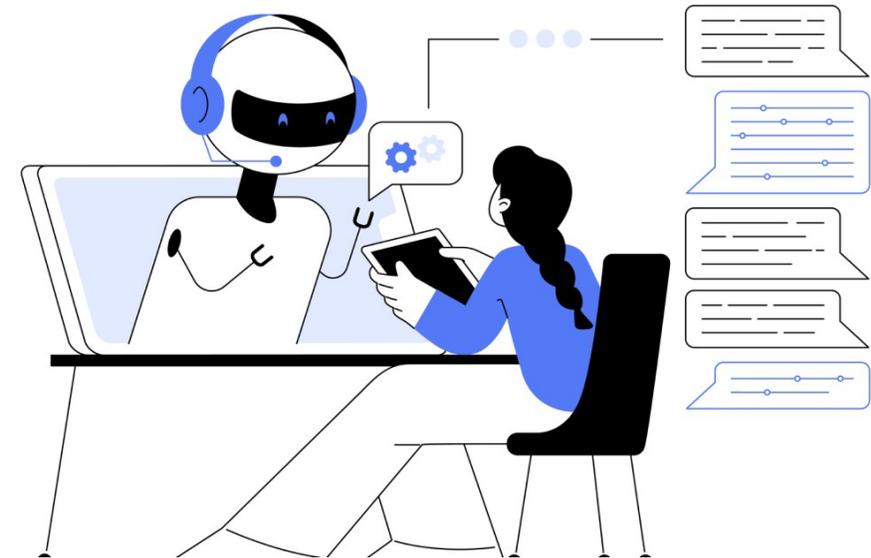
- Favoriser les candidats masculins
- Ignorer les noms culturellement différents

Biais dans les systèmes de prédiction criminelle

- Ciblage disproportionné de certaines communautés
- Prédiction basée sur des données policières biaisées

Biais dans les systèmes de soins de santé

- Sous-estimation de la gravité des symptômes chez certaines populations
- Manque de représentation des groupes minoritaires dans les données de recherche



Causes et Conséquences des Biais Culturels

Causes

- Données d'entraînement non représentatives
- Présence de stéréotypes dans les données
- Biais inconscients des concepteurs d'algorithmes



Conséquences

- Renforcement des inégalités sociales et culturelles
- Discrimination systémique et injustices
- Perte de confiance dans les systèmes d'IA

Stratégies pour Atténuer les Biais Culturels

- Diversification des ensembles de données
- Validation et vérification des modèles d'IA
- Sensibilisation et formation des développeurs et des utilisateurs
- Conception de modèles explicables pour identifier les biais

IA en traduction et interprétation

Il existe environ 7'111 langues dans le monde.

Ce chiffre varie sensiblement en fonction des critères utilisés pour définir une langue distincte.

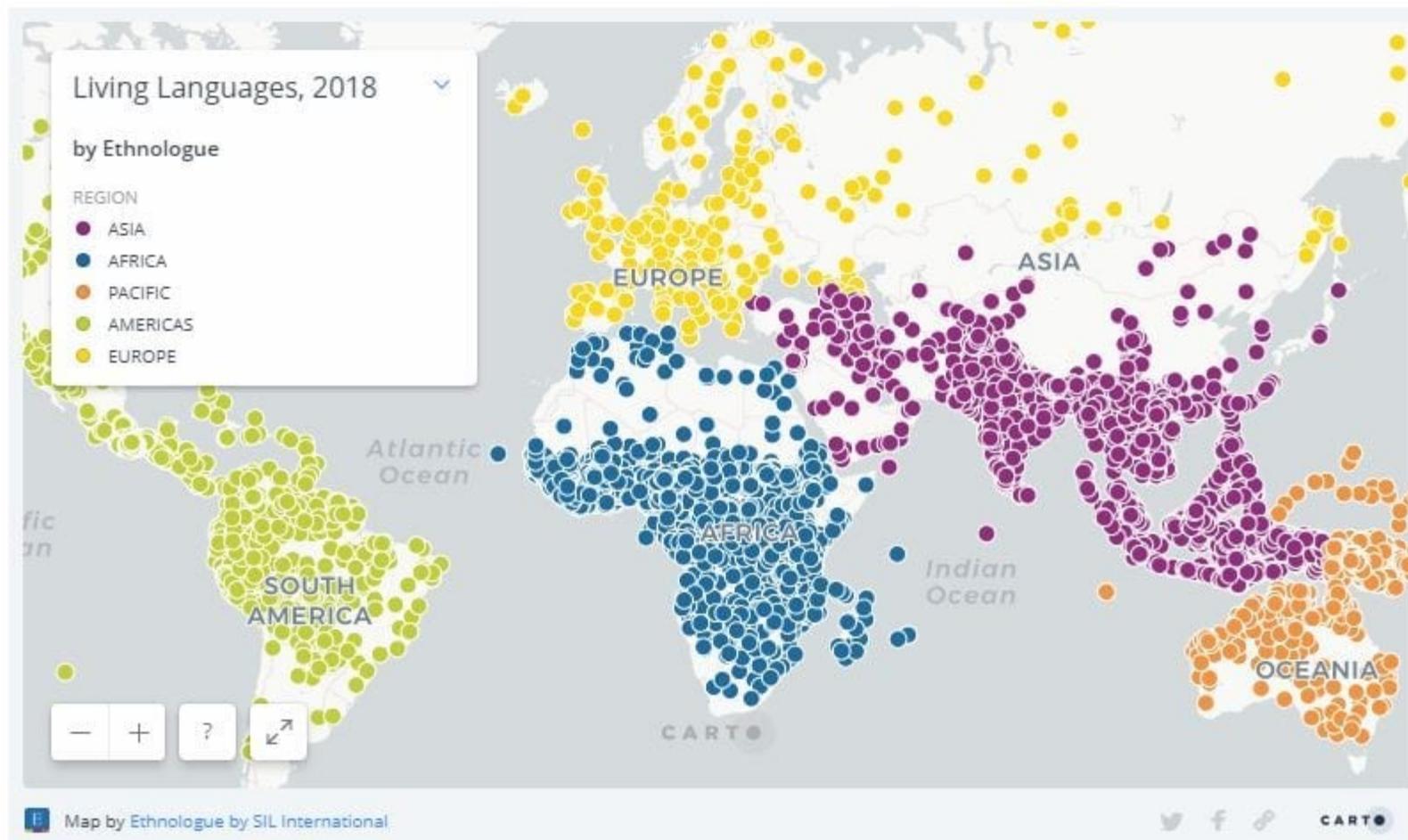
Ces langues appartiennent à différentes familles linguistiques, regroupant des langues partageant des origines communes.



ENVIRON
7000
LANGUES

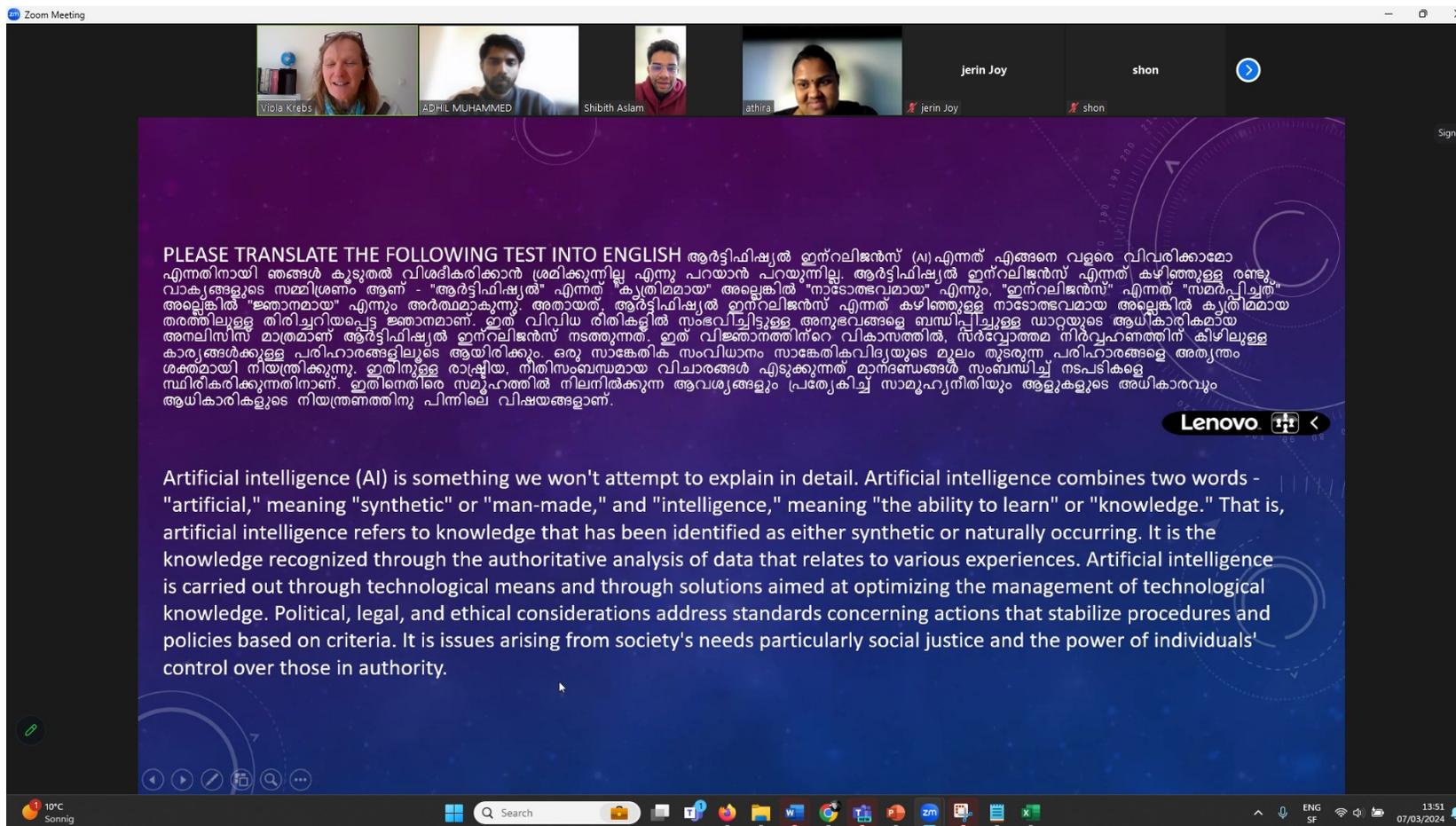
DIFFÉRENTES
SONT PARLÉES
DANS LE MONDE !

IA en traduction et interprétation



- Complexité de la syntaxe
- Diversité des écritures
- Utilisation des caractères spéciaux

IA en traduction

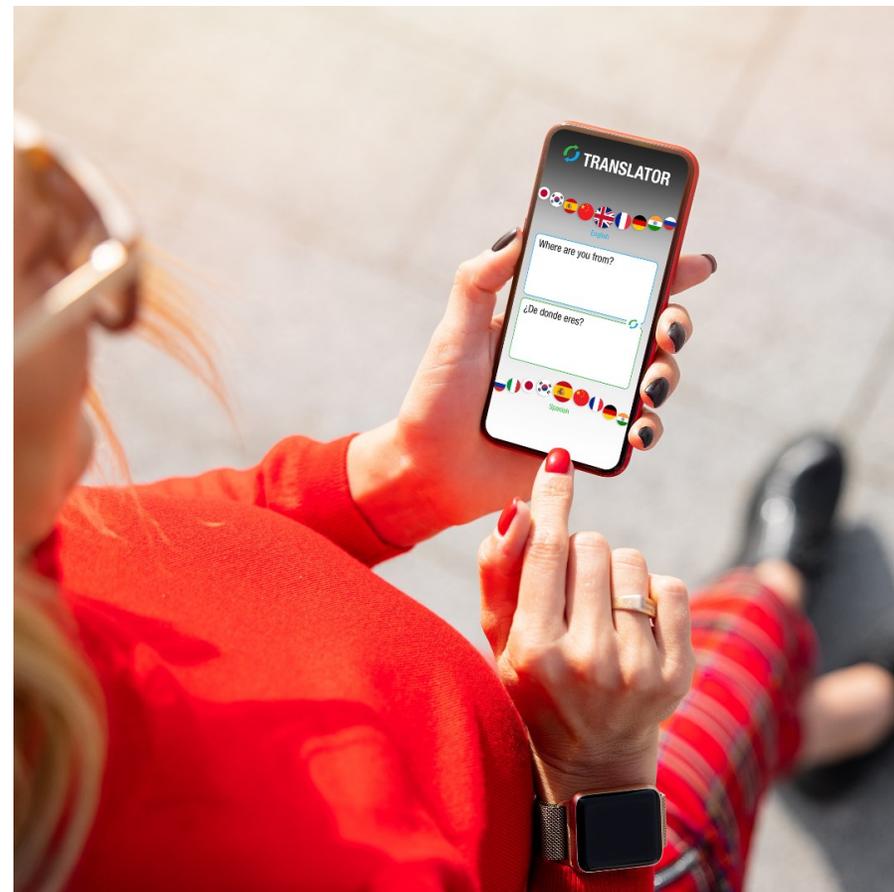


PLEASE TRANSLATE THE FOLLOWING TEST INTO ENGLISH ആർട്ടിഫിഷ്യൽ ഇൻറലിജൻസ് (AI) എന്നത് എങ്ങനെ വളരെ വിവരിക്കാമോ എന്നതിനായി ഞങ്ങൾ കൂടുതൽ വിശദീകരിക്കാൻ ശ്രമിക്കുന്നില്ല എന്നു പറയാൻ പറയുന്നില്ല. ആർട്ടിഫിഷ്യൽ ഇൻറലിജൻസ് എന്നത് കഴിഞ്ഞുള്ള രണ്ടു വാക്യങ്ങളുടെ സമ്മിശ്രണം ആണ് - "ആർട്ടിഫിഷ്യൽ" എന്നത് കൃത്രിമമായ അല്ലെങ്കിൽ "നാടോടവുമായ" എന്നും, "ഇൻറലിജൻസ്" എന്നത് "സമർപ്പിച്ചത് അല്ലെങ്കിൽ "ജ്ഞാനമായ" എന്നും അർത്ഥമാകുന്നു. അതായത്, ആർട്ടിഫിഷ്യൽ ഇൻറലിജൻസ് എന്നത് കഴിഞ്ഞുള്ള നാടോടവുമായ അല്ലെങ്കിൽ കൃത്രിമമായ തരത്തിലുള്ള തിരിച്ചറിയപ്പെട്ട ജ്ഞാനമാണ്. ഇത് വിവിധ രീതികളിൽ സംഭവിച്ചിട്ടുള്ള അനുഭവങ്ങളെ ബന്ധിപ്പിച്ചുള്ള ധാരാളം ആധികാരികമായ അനലിസിസ് മാത്രമാണ് ആർട്ടിഫിഷ്യൽ ഇൻറലിജൻസ് നടത്തുന്നത്. ഇത് വിജ്ഞാനത്തിന്റെ വികാസത്തിൽ, സർവ്വോത്തമ നിർവ്വഹണത്തിന് കഴിഞ്ഞുള്ള കാര്യങ്ങൾക്കുള്ള പരിഹാരങ്ങളിലൂടെ ആയിരിക്കും. ഒരു സാങ്കേതിക സംവിധാനം സാങ്കേതികവിദ്യയുടെ മൂലം തുടരുന്ന പരിഹാരങ്ങളെ അത്യന്തം ശക്തമായി നിയന്ത്രിക്കുന്നു. ഇതിനുള്ള രാഷ്ട്രീയ, നീതിസംബന്ധമായ വിചാരങ്ങൾ എടുക്കുന്നത് മാനദണ്ഡങ്ങൾ സംബന്ധിച്ച് നടപടികളെ സ്ഥിരീകരിക്കുന്നതിനാണ്. ഇതിനൊര സമൂഹത്തിൽ നിലനിൽക്കുന്ന ആവശ്യങ്ങളും പ്രത്യേകിച്ച് സാമൂഹ്യനീതിയും ആളുകളുടെ അധികാരവും ആധികാരികളുടെ നിയന്ത്രണത്തിനു പിന്നിലെ വിഷയങ്ങളാണ്.

Artificial intelligence (AI) is something we won't attempt to explain in detail. Artificial intelligence combines two words - "artificial," meaning "synthetic" or "man-made," and "intelligence," meaning "the ability to learn" or "knowledge." That is, artificial intelligence refers to knowledge that has been identified as either synthetic or naturally occurring. It is the knowledge recognized through the authoritative analysis of data that relates to various experiences. Artificial intelligence is carried out through technological means and through solutions aimed at optimizing the management of technological knowledge. Political, legal, and ethical considerations address standards concerning actions that stabilize procedures and policies based on criteria. It is issues arising from society's needs particularly social justice and the power of individuals' control over those in authority.

IA en interprétation

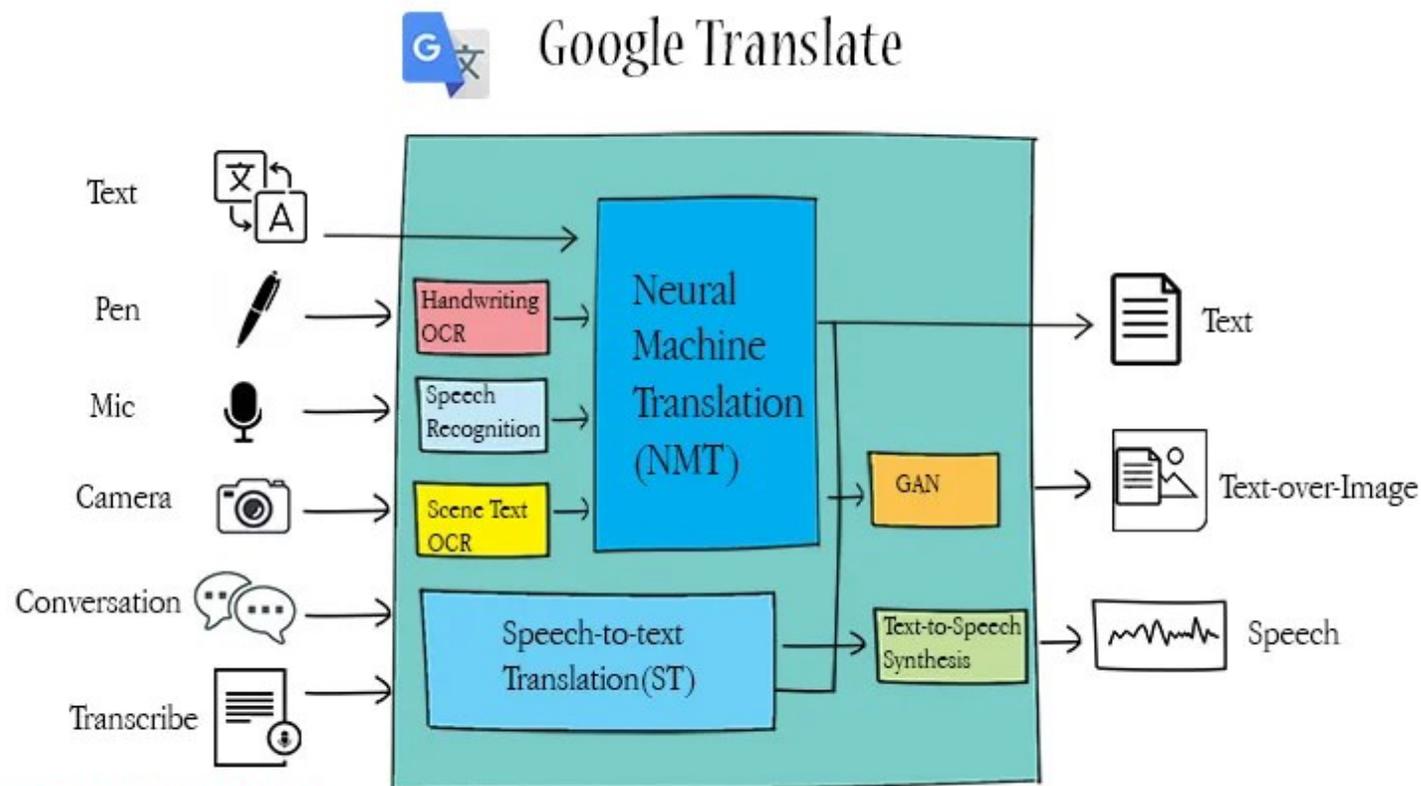
- La traduction basée sur l'IA s'est principalement concentrée sur les langues écrites
- Mais près de 3'500 langues vivantes sont principalement parlées et ne disposent pas d'un système d'écriture largement utilisé
- Cela rend impossible la création d'outils de traduction automatique à l'aide de techniques standard, qui nécessitent de grandes quantités de texte écrit pour entraîner un modèle d'IA.



<https://about.fb.com/news/2022/10/hokkien-ai-speech-translation/>

IA en traduction et interprétation

Google Neural Machine Translation (Depuis 2016)



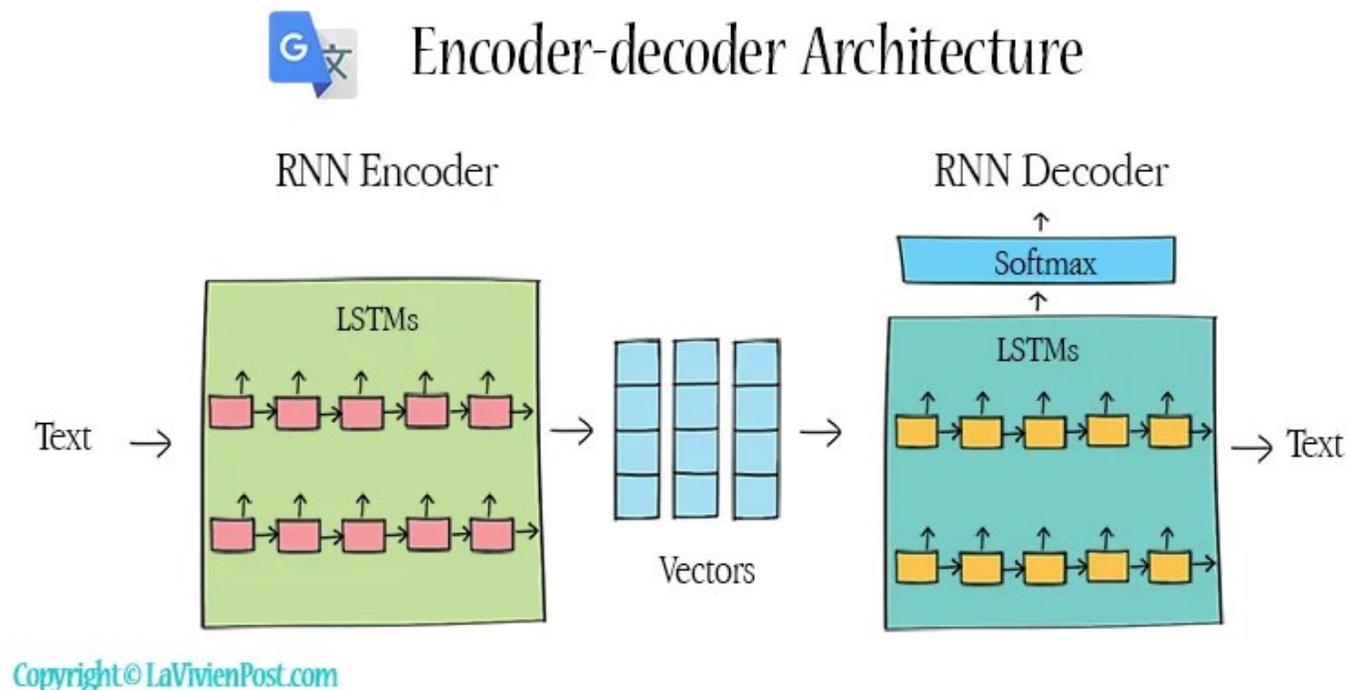
Copyright © LaVivienPost.com

Le NLP (*Natural Language Processing*) est une technologie permettant aux machines de comprendre et de générer la **parole** ou le **texte** humain.

<https://www.lavivienpost.com/google-translate-and-transformer-model/>

IA en traduction

- **Encodeur** : convertit les mots d'entrée en vecteurs
- **Décodeur** : convertit les vecteurs en mots dans une autre langue
- Dans ce modèle, l'encodeur et le décodeur utilisent tous deux des **réseaux neuronaux récurrents (RNN)**.
- Le RNN est un réseau neuronal qui utilise des connexions de rétroaction pour propager les informations d'une étape à l'autre.
- **Le RNN** : multicouche empilée de mémoire à long terme (LSTM). Le LSTM est un type de RNN capable d'effectuer un transfert de séquence à séquence.

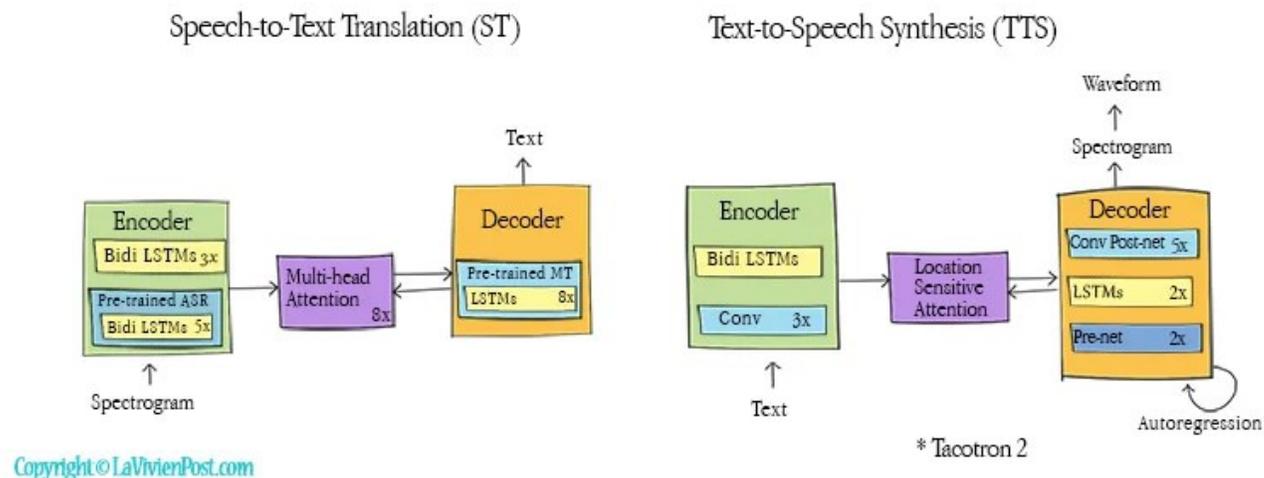


IA en traduction et interprétation

- **Encodeur** : convertit les mots d'entrée en vecteurs
- **Décodeur** : convertit les vecteurs en mots dans une autre langue
- Dans ce modèle, l'encodeur et le décodeur utilisent tous deux des **réseaux neuronaux récurrents (RNN)**.
- Le RNN est un réseau neuronal qui utilise des connexions de rétroaction pour propager les informations d'une étape à l'autre.
- Le **RNN** : multicouche empilée de mémoire à long terme (LSTM). Le LSTM est un type de RNN capable d'effectuer un transfert de séquence à séquence.



Speech-to-Speech Translation (S2ST) Cascade System

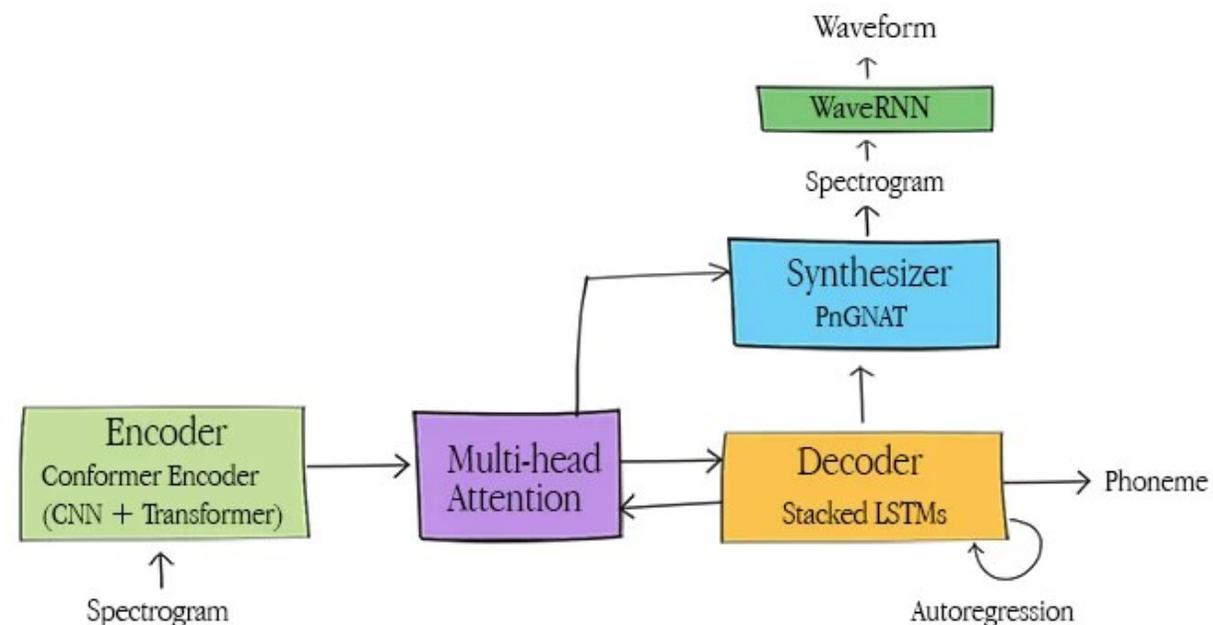


IA en interprétation

- **Encodeur** : convertit les mots d'entrée en vecteurs
- **Décodeur** : convertit les vecteurs en mots dans une autre langue
- Dans ce modèle, l'encodeur et le décodeur utilisent tous deux des **réseaux neuronaux récurrents (RNN)**.
- Le RNN est un réseau neuronal qui utilise des connexions de rétroaction pour propager les informations d'une étape à l'autre.
- **Le RNN** : multicouche empilée de mémoire à long terme (LSTM). Le LSTM est un type de RNN capable d'effectuer un transfert de séquence à séquence.



Speech-to-Speech Translation (Translatotron 2)



Copyright © LaVivienPost.com

WORLD
ECONOMIC
FORUM

Open AI CEO Sam Altman
(Photo: Fabrice Coffrini / AFP)





OpenAI

OpenAI

Structure déclarée en tant qu'association à but non lucratif
le 11 décembre 2015.

Les sacrifiés

☰ **TIME** **SUBSCRIBE**

Stay on Top of Cybersecurity...
sponsored by: IT Security Solutions

BUSINESS • TECHNOLOGY

Exclusive: OpenAI Used Kenyan Workers on Less Than \$2 Per Hour to Make ChatGPT Less Toxic

15 MINUTE READ



This image was generated by OpenAI's image-generation software, Dall-E 2. The prompt was: "A seemingly endless view of African workers at desks in front of computer screens in a printmaking style." TIME does not typically use AI-generated art to illustrate its stories, but chose to in this instance in order to draw attention to the power of OpenAI's

✕ venturebeat.com 🔖 📄 ☰

VentureBeat 🔍 ☰

Sam Altman wants up to \$7 trillion for AI chips. The natural resources required would be 'mind boggling'

Sharon Goldman

@sharongoldman

February 9, 2024 11:20 AM

f X in



Biais politiques de l'IA

- Distorsions et préjugés liés à des objectifs politiques ou idéologiques
- Réflexion des agendas politiques dans la conception et l'utilisation de l'IA
- Impact sur les décisions, les politiques publiques et les processus démocratiques

Accueil > News > Facebook

Christopher Wylie : l'homme qui peut faire exploser en vol Palantir et Cambridge Analytica (et Facebook ?)

Technologie : Ce n'est pour l'heure qu'une piste. Mais le lien entre les partisans de Donald Trump et l'entreprise Cambridge Analytica, spécialisée dans le Big Data et l'analyse comportementale, vient d'être établi par un lanceur d'alerte britannique.



Par Guillaume Series | Publié le mercredi 28 mars 2018 à 17:35 - Modifié le jeudi 29 mars 2018 à 13:05

Reactions 2 Post Partager plus

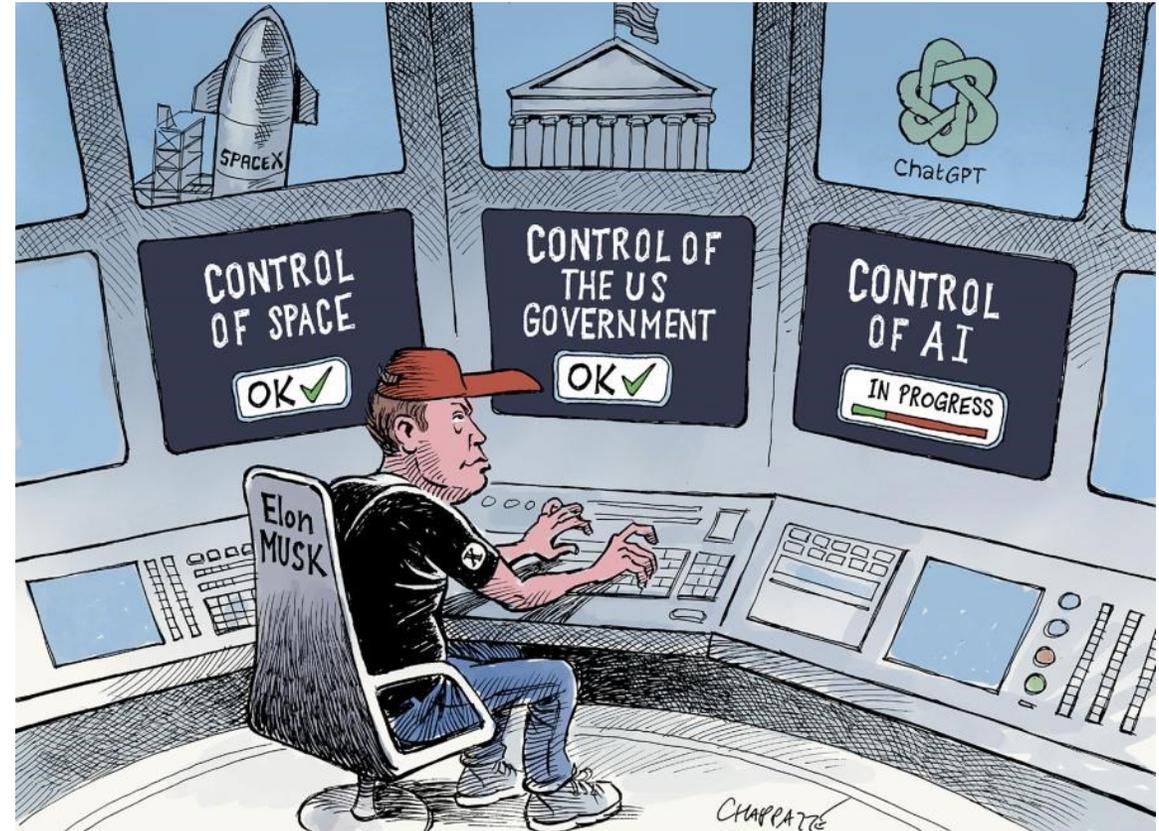
Le moins que l'on puisse dire, c'est que quand Chris Wylie s'exprime, ça dépote ! Le lanceur d'alerte aux cheveux roses a fait une série de déclarations explosives devant les parlementaires britanniques dans le cadre du dossier Cambridge Analytica [rapporté The Register](#). Sa motivation ? Il justifie ses révélations par le fait qu'il ne pense pas que les opérations de propagande "soient propices au processus démocratique, qu'il s'agisse d'une présidentielle américaine ou d'une élection locale".



Cambridge Analytica

Biais politiques de l'IA

- Distorsions et préjugés liés à des objectifs politiques ou idéologiques
- Réflexion des agendas politiques dans la conception et l'utilisation de l'IA
- Impact sur les décisions, les politiques publiques et les processus démocratiques

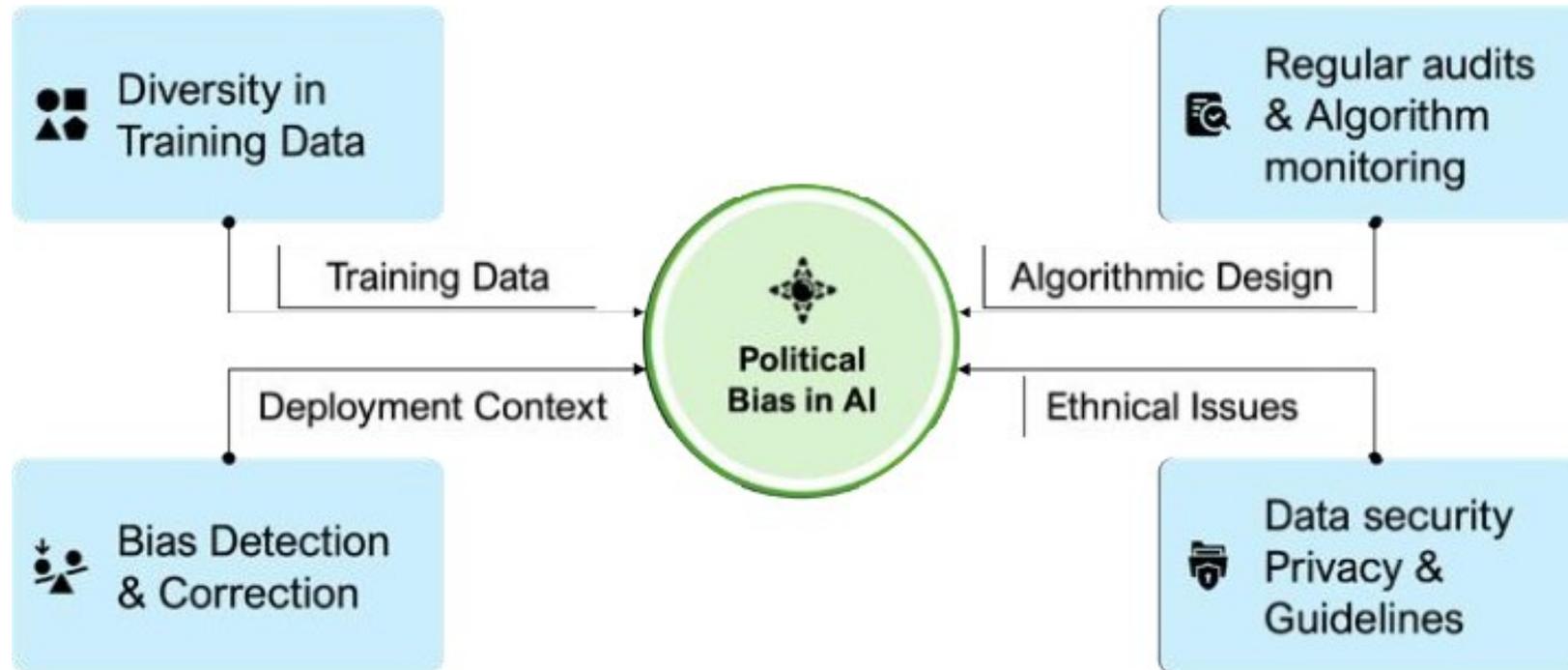


Biais politiques de l'IA



- Assistant et non protagoniste
- Esprit critique
- Nécessité de former
- Besoins des connaissances métiers
- Probabilité
- Nécessité de données de qualité et en grande quantité
- Explicabilité
- Services externes
- Transparence des algorithmes

Biais politiques de l'IA



https://www.researchgate.net/figure/Illustration-of-the-Mitigation-Strategies-for-Political-Bias-in-AI_fig2_382322175

- Besoin de directives éthiques pour l'utilisation de l'IA
- Appels à la transparence dans les processus décisionnels
- Surveillance continue et évaluation des impacts sur les communautés concernées

<https://www.wedemain.fr/dechiffrer/la-une-machine-a-biais-comment-lintelligence-artificielle-renforce-les-stereotypes/>



Aspects Juridiques de l'IA

Principales mesures contenues dans les textes de régulation de l'IA

		   				
		CdE / CAHAI	CE / GEHN IA	Règlement CE	OCDE	UNESCO
		CONTRAINANT ?	NON-CONTRAINANT	CONTRAINANT	NON-CONTRAINANT	NON-CONTRAINANT
Droits de l'homme	5					
Responsabilité	5					
Transparence	5					
Équité, non-discrimination	4					
Intervention et supervision humaines	4					
Sécurité et robustesse	4					
Dignité humaine	3					
Diversité	3					
Protection des données et de la vie privée	3					
Bien-être	2					
Données et gouvernance des données	2					
Démocratie	1					
Documentation technique	1					
Durabilité	1					
Environnement	1					
État de droit	1					
Gestion des risques	1					
Gouvernance et collaboration multipartites et adaptatives	1					
Proportionnalité et innocuité	1					
Sensibilisation et éducation	1					
Sociétés pacifiques, justes et interconnectées	1					
Mesures juridiques fondées sur les droits de l'homme	6	4	2	3	6	
Mesures politiques et sociétales	1	2	0	1	8	
Mesures techniques	2	4	6	2	3	

https://lestempselectriques.net/ANALYSE_IA.pdf

Mise en œuvre de l'Explicabilité

Les principes éthiques de l'IA



Transparence



Responsabilité



Confidentialité
& Sécurité



Atténuation
des biais



Gain
d'autonomie



Contrôle
humain

Réglementation et Normes

- Appels à des lignes directrices éthiques pour l'utilisation de l'IA dans le langage
- Exigences de transparence dans les algorithmes de traitement du langage
- Surveillance continue des impacts sociolinguistiques des systèmes d'IA



<https://www.unesco.org/en/artificial-intelligence/recommendation-ethics>

Réglementation et Normes

- Appels à une gouvernance éthique de l'IA dans les contextes politiques
- Nécessité de réglementations claires sur l'utilisation de l'IA dans les campagnes électorales
- Exigences de transparence et de redevabilité pour les organisations et les gouvernements utilisant l'IA



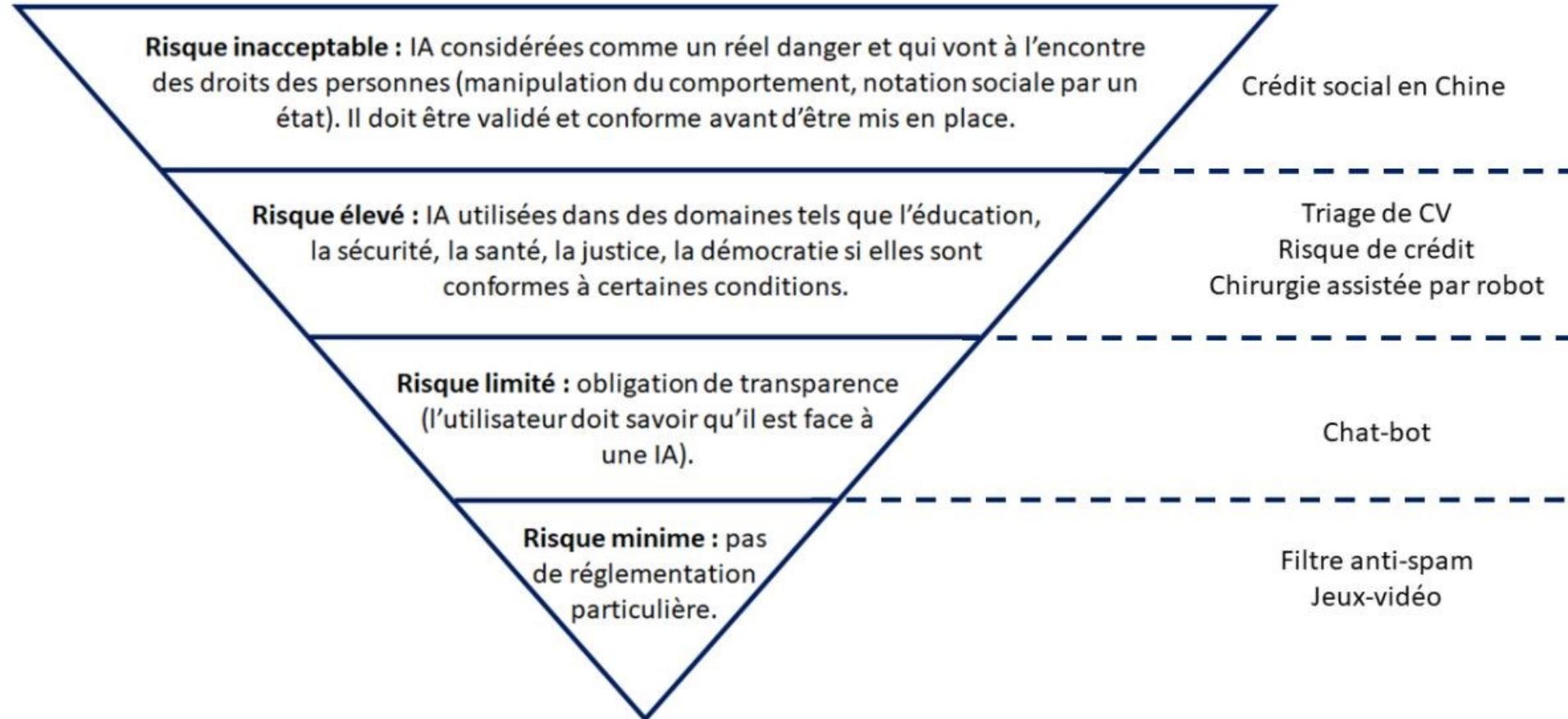
Aspects Juridiques de l'IA : EU AI Act



European AI Act

<https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>

Biais sociolinguistiques



Régulation de l'IA

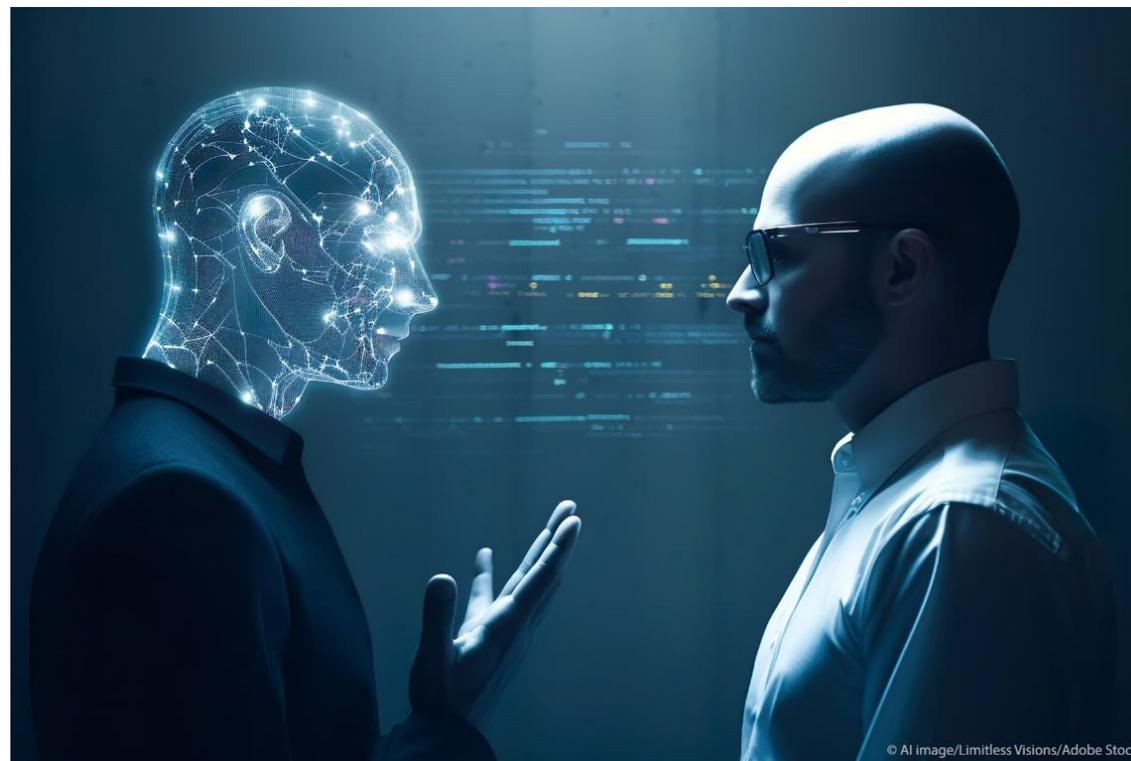
Loi sur l'IA de l'UE : première réglementation de l'intelligence artificielle

<https://www.europarl.europa.eu/news/fr/headlines/society/20230601STO93804/loi-sur-l-ia-de-l-ue-premiere-reglementation-de-l-intelligence-artificielle>

<https://www.weforum.org/agenda/2023/06/european-union-ai-act-explained/>

<https://artificialintelligenceact.eu/the-act/>

<https://www.ceps.eu/wp-content/uploads/2021/04/AI-Presentation-CEPS-Webinar-L.-Sioli-23.4.21.pdf?>



Norme ISO/IEC 42001:2023

Pourquoi la norme ISO/IEC 42001 est-elle importante ?

ISO/IEC 42001 est la première norme mondiale relative aux systèmes de gestion de l'IA, fournissant des conseils précieux pour ce domaine technologique en évolution rapide. Il aborde les défis uniques posés par l'IA, tels que les considérations éthiques, la transparence et l'apprentissage continu. Pour les organisations, il présente une manière structurée de gérer les risques et les opportunités associés à l'IA, en équilibrant l'innovation et la gouvernance.

Avantages

- Cadre de gestion des risques et des opportunités
- Démontrer une utilisation responsable de l'IA
- Traçabilité, transparence et fiabilité
- Économies de coûts et gains d'efficacité

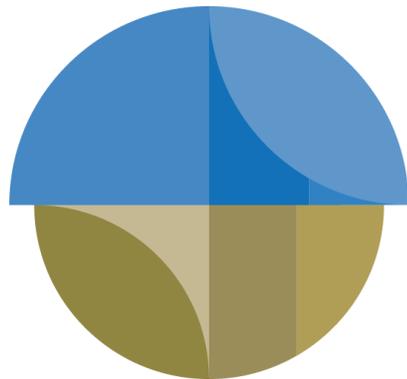


ISO/IEC 42001:2023
Information technology
Artificial intelligence
Management system

L'IA et l'avenir du marché du travail



97 million



85 million

Growing job demand:

1. Data Analysts and Scientists
2. AI and Machine Learning Specialists
3. Big Data Specialists
4. Digital Marketing and Strategy Specialists
5. Process Automation Specialists
6. Business Development Professionals
7. Digital Transformation Specialists
8. Information Security Analysts
9. Software and Applications Developers
10. Internet of Things Specialists

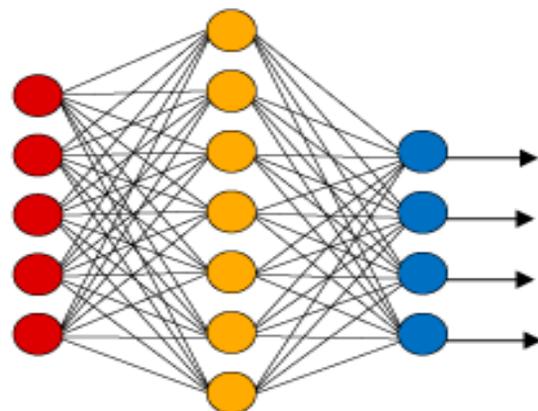
Decreasing job demand:

1. Data Entry Clerks
2. Administrative and Executive Secretaries
3. Accounting, Bookkeeping and Payroll Clerks
4. Accountants and Auditors
5. Assembly and Factory Workers
6. Business Services and Administration Managers
7. Client Information and Customer Service Workers
8. General and Operations Managers
9. Mechanics and Machinery Repairers
10. Material-Recording and Stock-Keeping Clerks

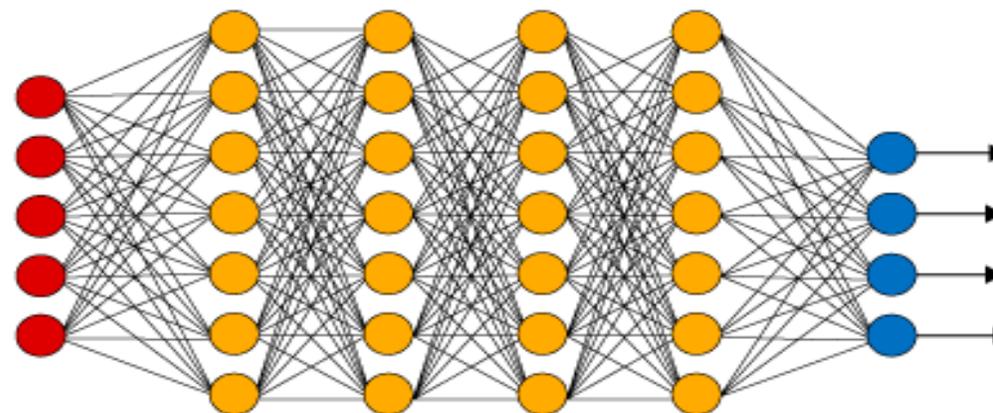
Source: Future of Jobs Report 2020, World Economic Forum.

Explicabilité de l'IA

Simple Neural Network



Deep Learning Neural Network

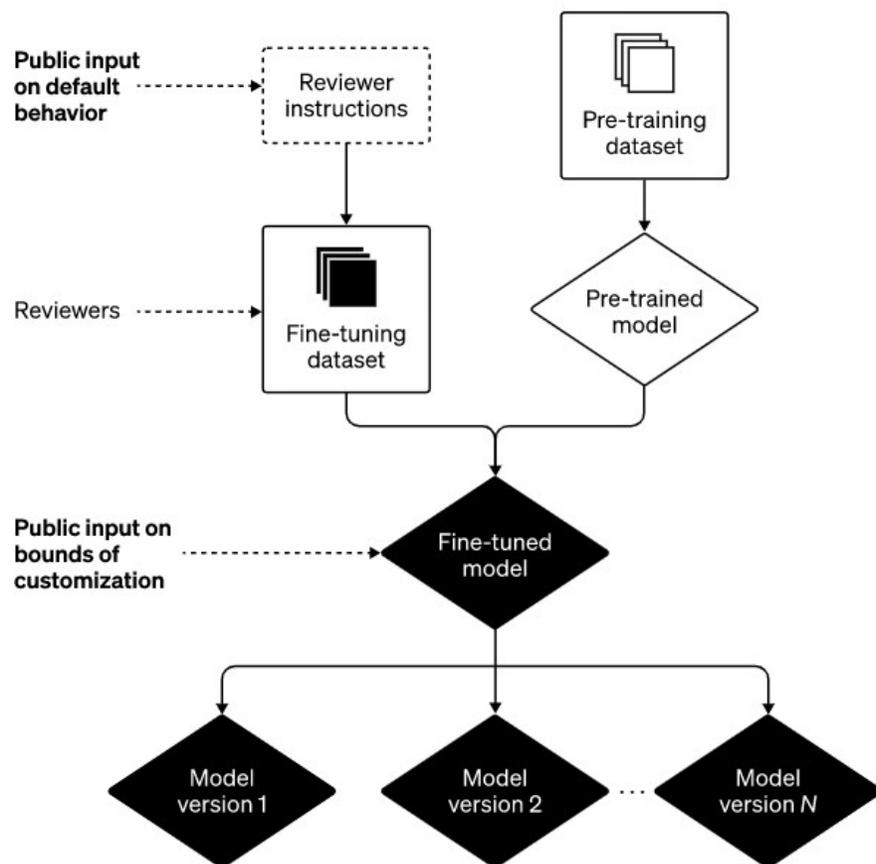


● Input Layer ● Hidden Layer ● Output Layer

Source: <https://towardsdatascience.com/mnist-vs-mnist-how-i-was-able-to-speed-up-my-deep-learning-11c0787e6935>

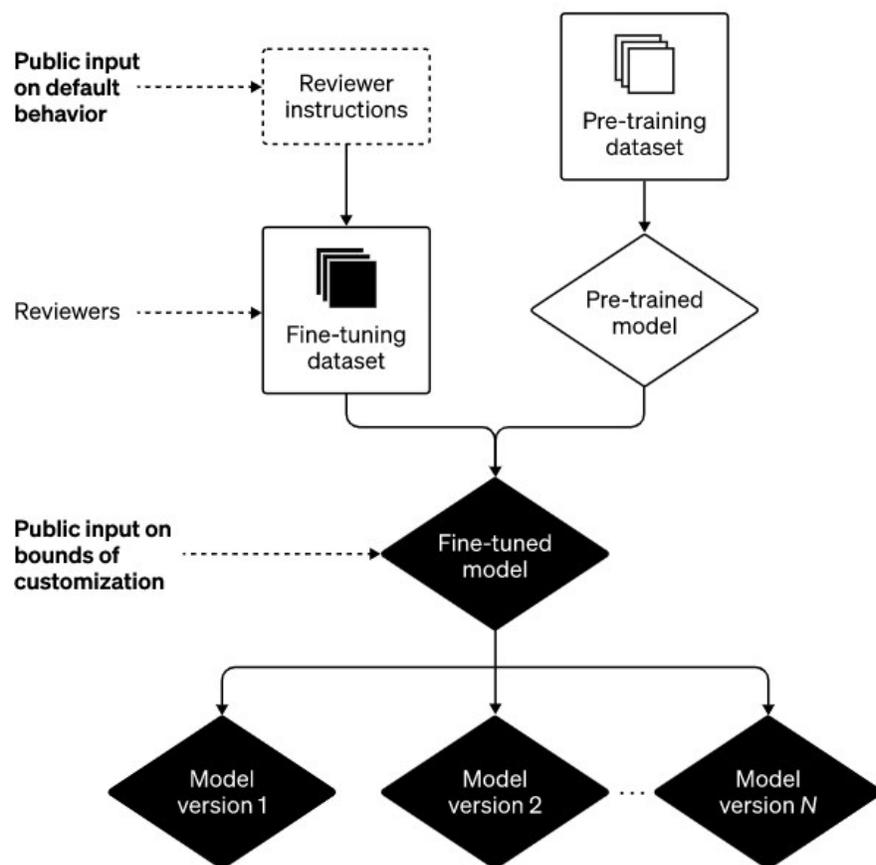
Voir aussi: <https://www.youtube.com/watch?v=aircAruvnKk>

Explicabilité de l'IA



- **Renforcer la confiance** : Les utilisateurs ont besoin de comprendre pourquoi les décisions sont prises par les systèmes d'IA pour leur faire confiance.
- **Responsabilité** : Identifier les erreurs ou les biais dans les décisions prises par les modèles d'IA et permettre aux concepteurs d'en être responsables.
- **Conformité réglementaire** : De nombreuses réglementations exigent maintenant que les systèmes d'IA soient explicables pour garantir la transparence et la conformité.

Explicabilité de l'IA



Documentation Complète :

- Fournir une documentation détaillée sur la conception, l'entraînement et le déploiement du modèle d'IA.
- Inclure des descriptions claires des données utilisées, des méthodes d'entraînement, des métriques de performance, etc.

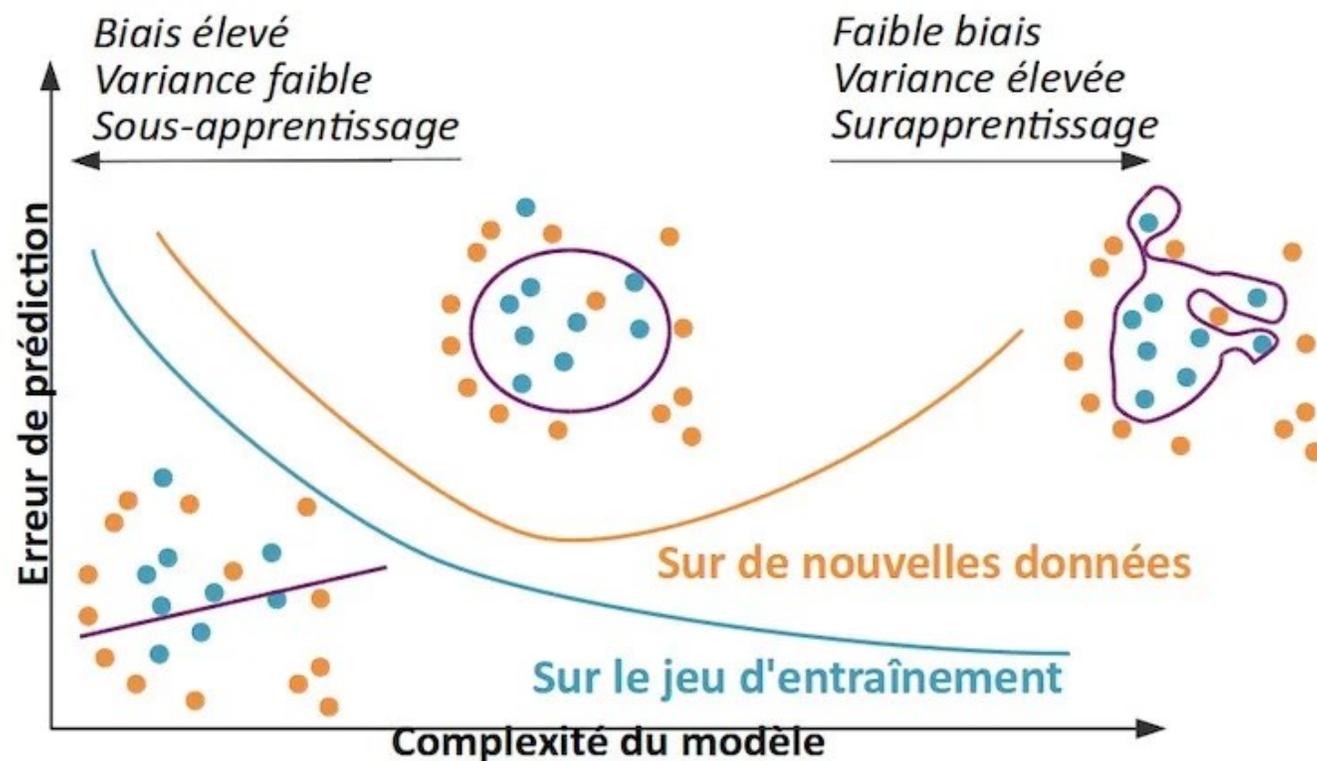
Méthodes d'Interprétabilité :

- Utiliser des techniques telles que la cartographie des caractéristiques (*feature mapping*) pour montrer quelles caractéristiques influencent les prédictions.
- Fournir des explications sur la manière dont le modèle a pris une décision spécifique.

Visualisation Intuitive :

- Utiliser des outils de visualisation pour représenter graphiquement les résultats et les prédictions du modèle.
- Cela permet aux utilisateurs de comprendre visuellement le fonctionnement du modèle d'IA.

Défis de l'Explicabilité de l'IA



<https://blog.enioka.com/2022/03/14/churn-modelling-prevenir-la-perte-de-clients-avec-de-lia-2-3/>

- **Complexité des modèles :** Certains modèles d'IA, en particulier les réseaux de neurones profonds, peuvent être difficiles à expliquer en raison de leur complexité.
- **Trade-off avec la Performance :** Parfois, rendre un modèle plus explicatif peut réduire sa performance.
- **Protection des Informations Sensibles :** L'explication du modèle peut parfois révéler des informations sensibles contenues dans les données.

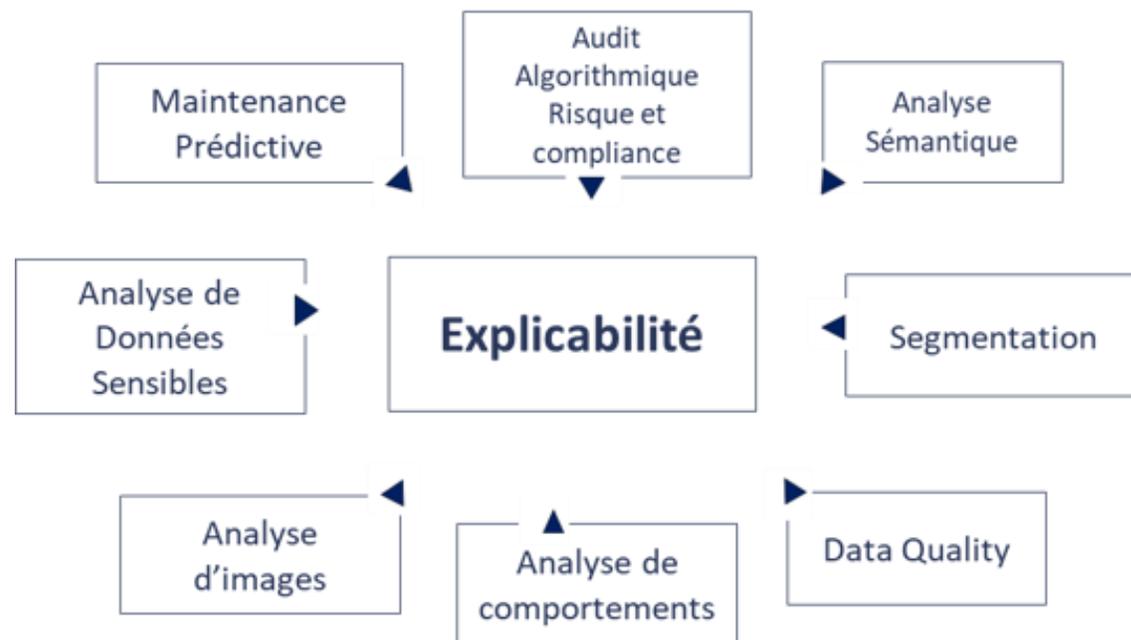
Exemples de mise en œuvre de l'Explicabilité

- **Médecine** : Explication des diagnostics médicaux pour permettre aux médecins de comprendre les raisons derrière les recommandations du modèle.
- **Finance** : Explication des prévisions financières pour aider les analystes à comprendre les facteurs qui influencent les prédictions.
- **Sécurité** : Explication des décisions prises par les systèmes de détection d'intrusion pour aider les analystes à comprendre les menaces potentielles.



Mise en œuvre de l'Explicabilité

- Intégration continue de l'explicabilité dans le cycle de vie de développement de l'IA.
- Développement de normes et de meilleures pratiques pour l'explicabilité de l'IA.
- Recherche sur des techniques plus avancées d'explication de modèles pour les modèles complexes.



Ethical Hackers

Les « ethical hackers » pour l'Intelligence Artificielle (IA) jouent un rôle crucial dans l'identification des vulnérabilités et des risques potentiels des systèmes d'IA.

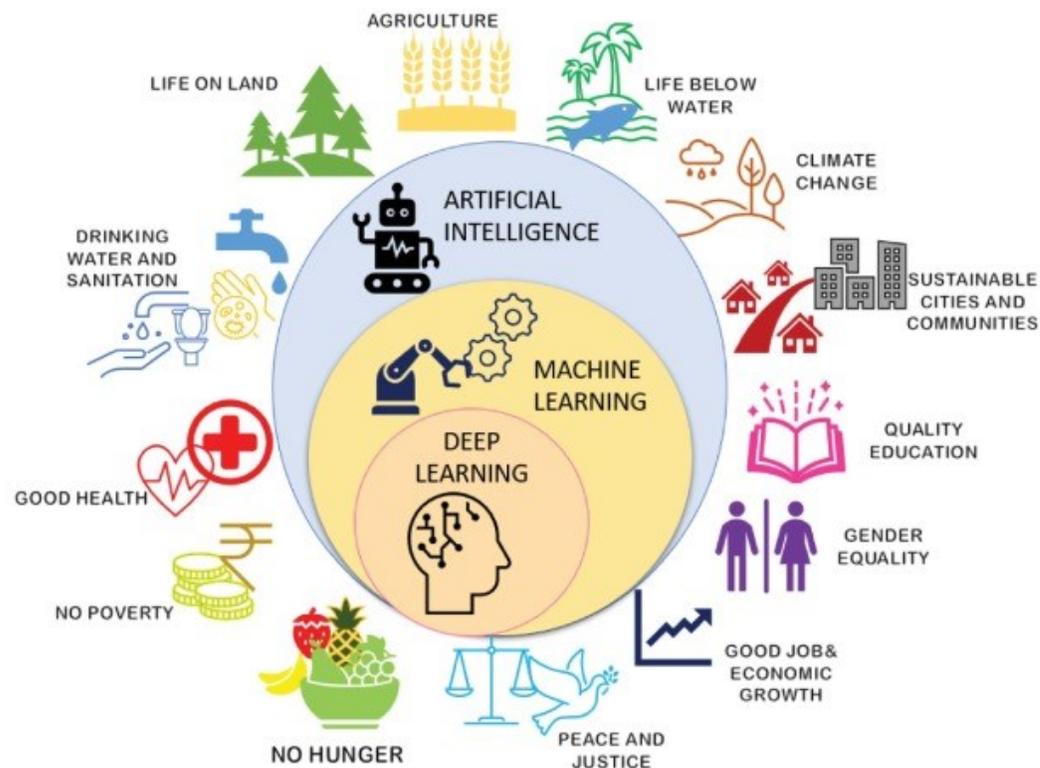


IA et durabilité



Les ordinateurs consomment de l'énergie, beaucoup d'énergie...

Impact écologique des logiciels IA

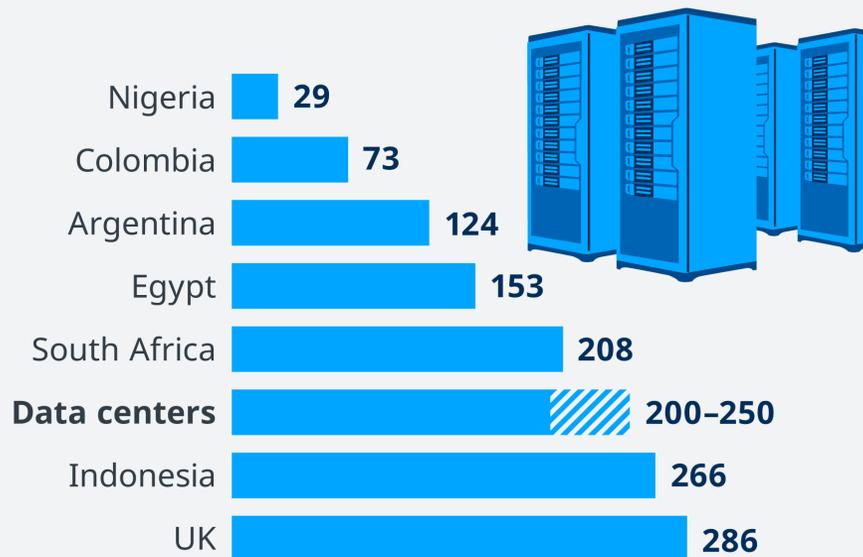


Les ordinateurs consomment de l'énergie, beaucoup d'énergie...

IA et durabilité

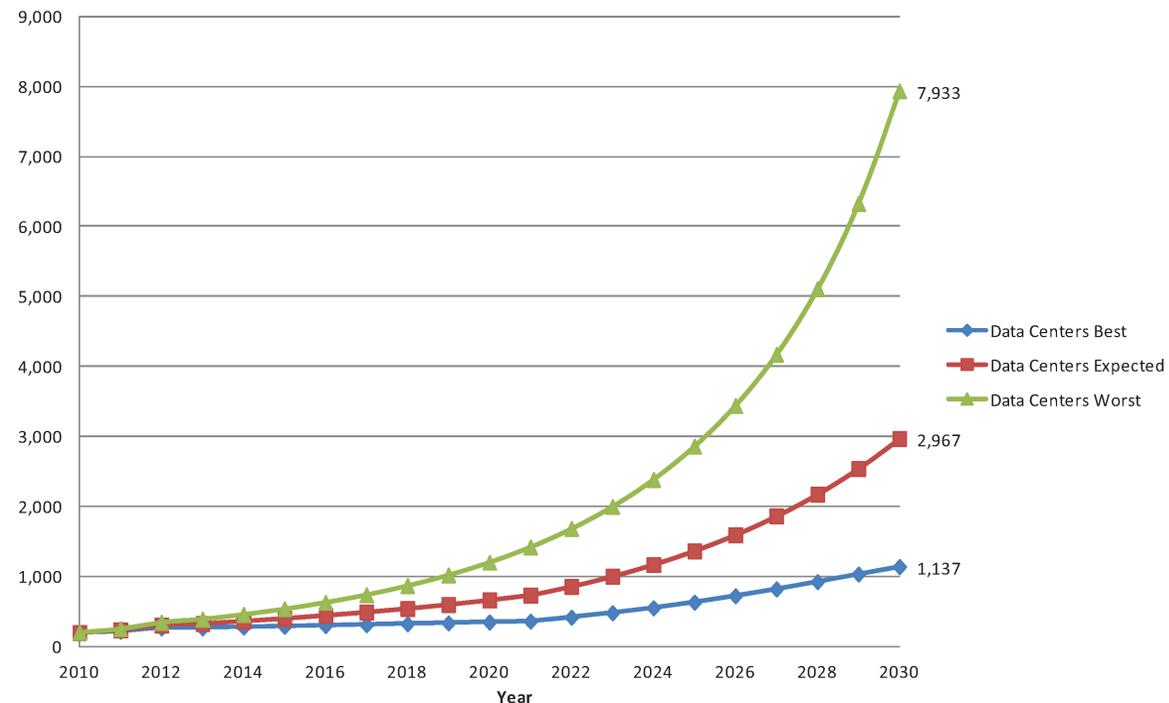
Data centers use more electricity than entire countries

Domestic electricity consumption of selected countries vs. data centers in 2020 in TWh



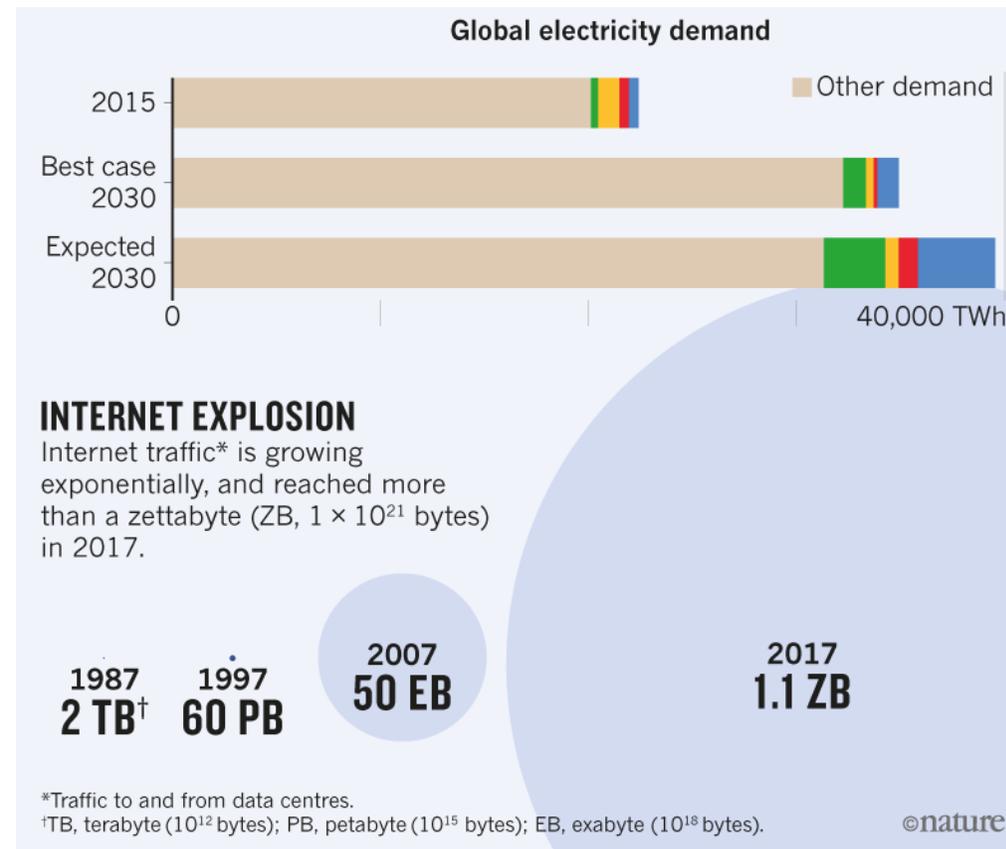
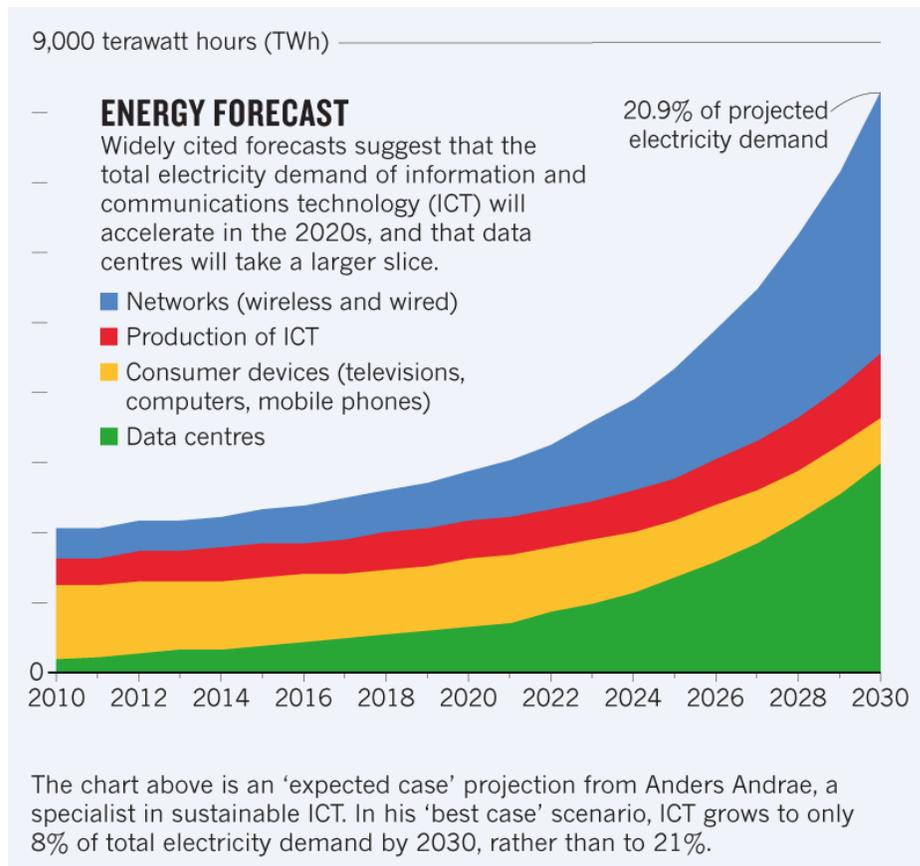
Source: Enerdata, IEA

Electricity usage (TWh) of Data Centers 2010-2030



On Global Electricity Usage of Communication Technology: Trends to 2030 by Anders S. G. Andrae, ScilitPreprints.org *ORCID andTomas Edler

IA et durabilité



IA et durabilité

La mondialisation crée également des marchés mondiaux pour le déversement des déchets électroniques :

L'exemple d'**Agbogbloshie Ghana**

Décharge électronique du Ghana : les écologistes s'inquiètent du déversement de déchets électroniques dans

la capitale, Accra, CCTV Afrique,

<https://www.youtube.com/watch?v=WzR79Wv2oJA>

Toxicity : la vie à Agbogbloshie, la plus grande décharge de déchets électroniques au monde au Ghana, documentaire RT,

<https://www.youtube.com/watch?v=mleQVO1Vd1I>



Photo de Andrew McConnell,
<https://www.andrewmcconnell.com/Biography/1>

Formations continues IA

- Générale pour non-spécialistes
 - Education
 - Ethique
-
- Possibilité d'empiler les crédits --> CAS personnalisé

<https://cui.unige.ch/fr/pin/clinique-de-l-ia/>

