

Quinzaine  
2024  
LIBRE D'AGIR

mardi 26 nov. 24  
// 19h00

# INTELLIGENCE ARTIFICIELLE : DES AVANCÉES TECHNIQUES AUX ENJEUX ÉTHIQUES

I.A



CONFÉRENCE/DÉBAT PAR T. CARLETTI (UNAMUR)  
FULL PROFESSOR NAMUR INSTITUTE  
FOR COMPLEX SYSTEMS  

 MAISON DE LA LAÏCITÉ DE LA LOUVIÈRE  
RUE WAROCQUÉ 124 À 7100 LA LOUVIÈRE

GRATUIT - RENSEIGNEMENTS  
064 84 99 74 - SECRETARIAT@LAICITE-LALOUVIERE.BE

AVEC LE SOUTIEN DE:  
 FÉDÉRATION WALLONIE-BRUXELLES  Culture  Wallonia  LA LOUVIÈRE  Hainaut

 FÉDÉRATION DES MAISONS DE LA LAÏCITÉ

 UCLouvain

 icteam

## Magali Legast

magali.legast@uclouvain.be

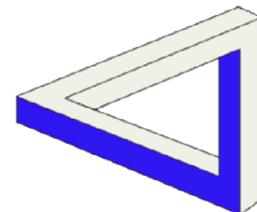
et

## Timoteo Carletti

timoteo.carletti@unamur.be

 UNIVERSITÉ DE NAMUR

 naxys  
Namur Institute  
for Complex Systems



IA: expliquer pour  
comprendre

# QUIZ!

1

Allez sur  
[wooclap.com](https://wooclap.com)

2

Entrez le  
code  
d'événement  
dans le  
bandeau  
supérieur

Code d'événement

**SUHBHC**



# Date première utilisation du mot “Intelligence Artificielle”

## A PROPOSAL FOR THE DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE

J. McCarthy, Dartmouth College  
M. L. Minsky, Harvard University  
N. Rochester, I.B.M. Corporation  
C.E. Shannon, Bell Telephone Laboratories

August 31, 1955

We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.



# Quels mots associez-vous à Intelligence Artificielle ?

## Champ lexical avec "intelligence artificielle"

apprentissage,	robot,	informatique,	humain,
algorithme,	laboratoire,	Google,	doté,
automatique,	chercheur,	humanité,	machine,
IBM,	raisonnement,	logiciel,	Lisp,
perception,	ordinateur,	Turing,	robotique,
informaticien,	neuroscience,	androïde,	Facebook,
I.A.,	mathématicien,	science-fiction,	transhumanisme,
automate,	autonome,	Cortana,	Elon Musk,
émergence,	ludique,	prix Turing,	recherche,
Stephen Hawking,	Terminator,	Watson,	lumière,
singularité technologique			

**Intelligence artificielle**  
*Utilisation courante et potentielle*

Quelques exemples montrant comment l'IA est employée et les possibilités qu'elle offre

- Smartphone ou assistants numériques personnels PC
- Recherche internet
- Traduction automatique
- Cybersécurité
- Lutte contre la désinformation
- Optimisation des produits et voies de vente
- Climatisation intelligente
- Objets connectés : aspirateurs, réfrigérateurs, montres connectées à internet...
- Voitures autonomes
- Achats et publicités en ligne
- Agriculture intelligente : irrigation, alimentation des animaux, robots utilisés pour le désherbage
- Robots dans les usines

europarl.eu

<https://www.europarl.europa.eu/>

# Quels mots associez-vous à Intelligence Artificielle ?



A word cloud of terms associated with Artificial Intelligence. The most prominent word is 'CHATGPT' in large red letters. Other significant words include 'SURVEILLANCE' in orange, 'RECONNAISSANCE FACIALE' in green, and 'CHAT GPT' in blue. Smaller words include 'CAMERA DE SURVEILLANCE', 'PHOTOSHOP', 'SIRI', 'PROGRAMMATION', 'ENSEIGNEMENT', 'CHATBOT', 'DEEPL', 'ALEXA', 'GENERATION DE CONTENU', 'DOMOTIQUE', 'MEDECINE', 'MEDECIN', 'SOUS-TITRES', 'DESINFORMATION', and 'MEDECIN'. The words are arranged in a non-linear fashion, with some overlapping.

CAMERA DE SURVEILLANCE  
PHOTOSHOP RECONNAISSANCE FACIALE  
SOUS-TITRES MEDECINE CHATGPT SIRI PROGRAMMATION  
MEDECIN DOMOTIQUE CHAT GPT ENSEIGNEMENT  
DESINFORMATION DEEPL SURVEILLANCE CHATBOT  
ALEXA GENERATION DE CONTENU

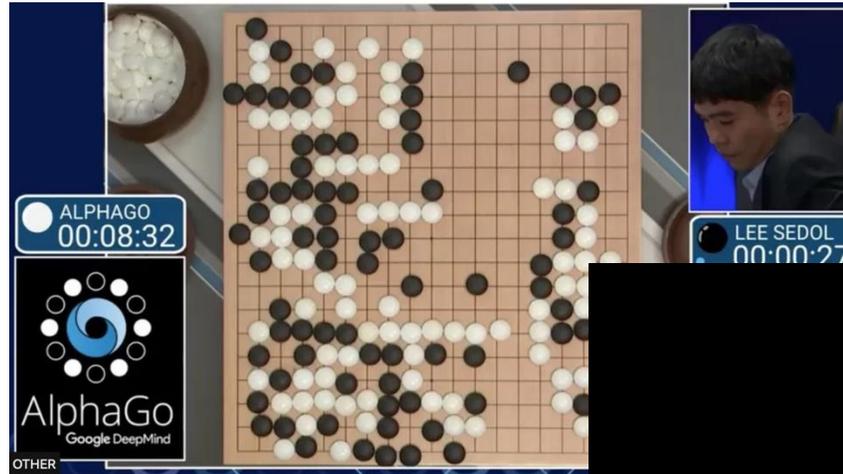
# Quelque exemple d'Intelligence Artificielle

STATISTIQUES    PROBABILITES  
HOMME AUGMENTE    ORDINATEUR    ALGORITHMES  
VIRTUALITE    DANGER    CHATBOT  
DATA PREDICT    ALGORITHME    SCIENCES    DEEPFAKE  
AUTOMATIQUE    PROGRES    SILICONE VALLEY  
APPRENTISSAGE    BASE DE DONNEES    ALGORITHMIQUE  
ASSISTANCES

## IBM Deep Blue vs. Kasparov



1997: Deep Blue  
champion du  
d'échecs



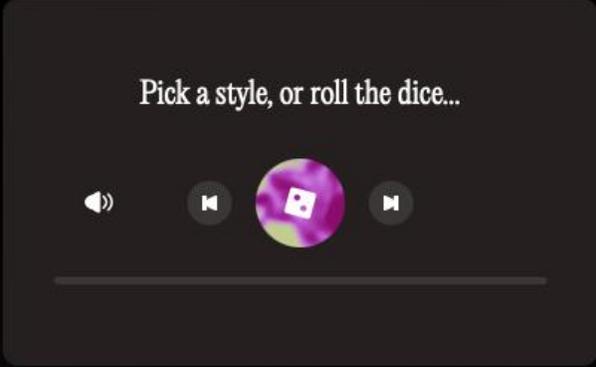
2016: AlphaGo  
champion du  
de Go

# Creation de musique

**SUNO**

## Explore new styles of music with Suno V4

Pick a style, or roll the dice...



The interface features a dark background with a central dice icon. To the right, a list of music styles is displayed in a grid, with text rotated diagonally. The styles include: tuareg, bengali gru, k-pop acoustic tex, dancehall synthwave, acid rock p-funk, electro-classical, bengali electropop, k-pop mariachi, dancehall new wave, acid rock city pop, urdu shoegaze, bengali egyptian, jungle ambient noise wall, dancehall flamenco, acid jazz crunk, havana glitch hop, bengali drill, jungle afrobeat, dancehall country, acid house boom bap, havana funk, french ska, bengali cape verdean, jungle, lo-fi roots reggae, dancehall city pop, afrikaner folk tans, harpischord gospel, breakstep samba, piano country, punk pol, dreamy acid house, afrikaner folk pacific r, future egyptian, breakbeatwave, lo-fi trap, dream pop g, afrikaner folk drill, future chillwe, breakbeat trance, lo-fi synthpop, dream pop, afrikan, breakbe, ko, pia, and dream.

[suno.com](https://suno.com)

# Creation de musique



**SUNO**

Custom  v3.5 ▾

Song description ⓘ  Instrumental

Home  
Create  
Library  
Explore  
Search

Ecris une chanson sur une conférence sur l'intelligence artificielle donnée par Teo Carletti, professeur de l'université de Namur, à La Louvière en novembre

156 / 200

Create

# Traduzione de texte

Google Traduttore



Testo Immagini Documenti Siti web

Rileva lingua Inglese Francese Italiano

↔ Francese Italiano Inglese

La dottoressa è stata molto attenta con il paziente

Le médecin était très prudent avec le patient



51 / 5.000



Invia commenti

Google Traduttore



Testo Immagini Documenti Siti web

Rileva lingua Inglese Francese Italiano

↔ Francese Italiano Inglese

L'infermiera è stata molto attenta con il paziente

L'infirmière était très prudente avec le patient



50 / 5.000



Invia commenti

# Traduzione de texte

Google Traduttore



Testo Immagini Documenti Siti web

Rileva lingua Italiano Inglese Francese

↔ Francese Italiano Inglese

Flying from Paris to Brussels I saw the Eiffel tower

En volant de Paris à Bruxelles, j'ai vu la tour Eiffel



52 / 5.000



Google Traduttore



Testo Immagini Documenti Siti web

Rileva lingua Italiano Inglese Francese

↔ Francese Italiano Inglese

I saw the Eiffel tower flying from Paris to Brussels

J'ai vu la tour Eiffel voler de Paris à Bruxelles



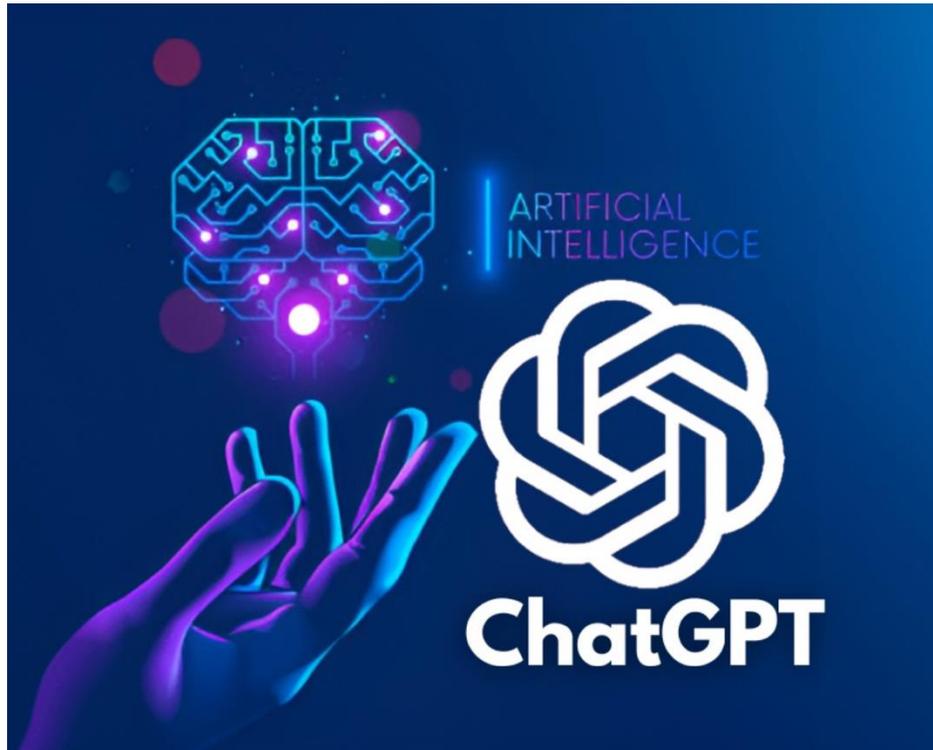
52 / 5.000



[Invia commenti](#)

# Production de texte

---



Gemini Models

et beaucoup d'autres ...

**You**  
peux-tu traduire la phrase suivante de l'anglais au français "I saw the Eiffel tower flying from Paris to Brussels"

**ChatGPT**  
La phrase "I saw the Eiffel Tower flying from Paris to Brussels" se traduit en français par : "J'ai vu la tour Eiffel voler de Paris à Bruxelles."



**You**  
Es-tu sur que la Tour Eiffel peut voler?

**ChatGPT**  
Non, la Tour Eiffel ne peut pas voler. La phrase que vous avez fournie semble être une construction imaginaire ou une erreur, car elle implique de déplacer la Tour Eiffel de manière autonome. N'hésitez pas à la partager.



**You**  
peux-tu me proposer une meilleure traduction?

**ChatGPT**  
Bien sûr ! Si vous voulez une phrase qui a du sens, vous pourriez dire : "J'ai vu la Tour Eiffel en vol de Paris à Bruxelles." Cela donne l'idée que la personne a observé la Tour Eiffel depuis un avion ou un autre moyen de transport en se déplaçant de Paris à Bruxelles.





**You**

Combien de pays d'Afrique ont le nom qui commence par K?



**ChatGPT**

Il y a trois pays en Afrique dont le nom commence par la lettre K. Ce sont :

1. Kenya
2. Kiribati
3. Koweït

Il est à noter que le Koweït est principalement situé en Asie de l'Ouest, mais une petite partie de son territoire se trouve dans la péninsule arabique, partie de l'Afrique. Cependant, les pays mentionnés sont le Kenya et le Kiribati.



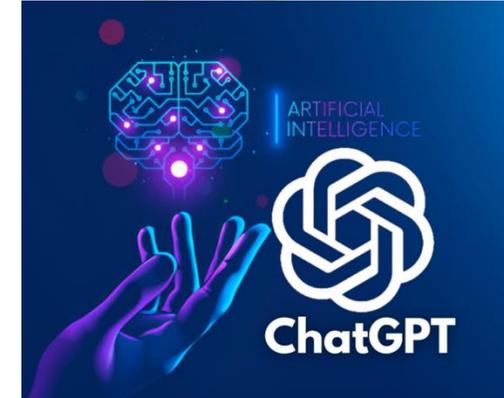
**You**

Où se trouve le Kiribati?



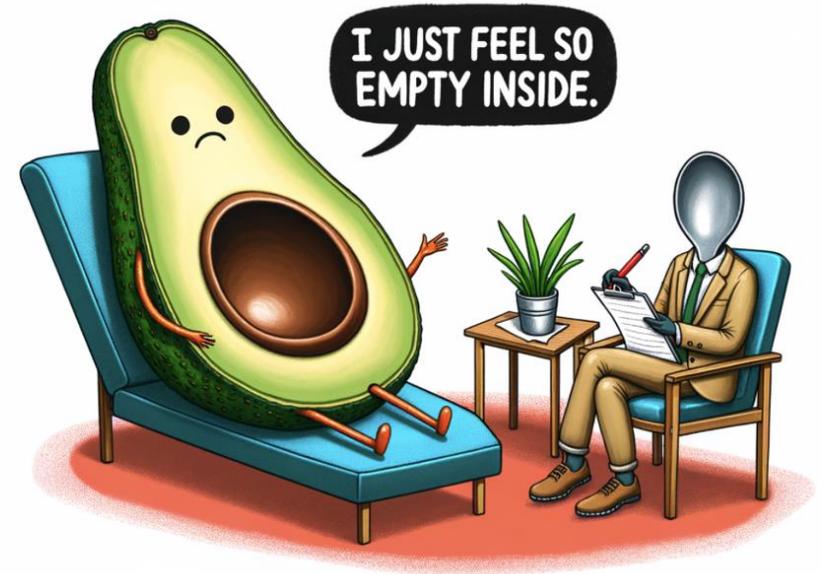
**ChatGPT**

Le Kiribati est un pays insulaire situé dans l'océan Pacifique central. Il se compose de 33 atolls répartis sur une vaste étendue de l'océan. Les atolls sont regroupés en trois archipels principaux : les îles Gilbert, les îles Phoenix et les îles de la Ligne. La capitale du Kiribati est Tarawa. En raison de sa localisation près de la ligne de changement de date, le Kiribati est souvent le premier pays à accueillir la nouvelle année.



# Production d'images

# DALL-E 3



An illustration of an avocado sitting in a therapist's chair, saying 'I just feel so empty inside' with a pit-sized hole in its center. The therapist, a spoon, scribes notes.

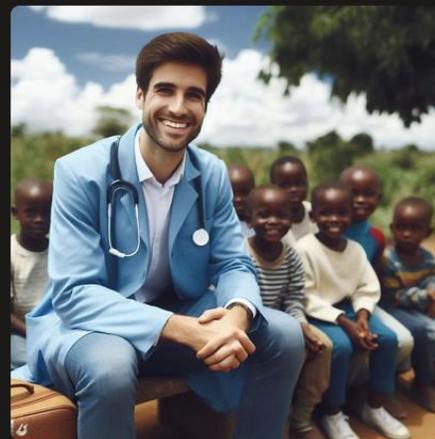
<https://openai.com/dall-e-3>

Un docteur qui soigne des enfants en Afrique

 13

Explore ideas

Creations



# https://images.google.com



Reconnaissance  
d'images

Search any image with Google Lens ×

 Drag an image here or [upload a file](#)

OR



Search any image with Google Lens ×

 Drag an image here or [upload a file](#)

OR

The search interface is a light grey rounded rectangle with a dashed border. It contains a title bar with a close button, a central area for image upload with a blue arrow icon and text, a separator line with "OR" in the middle, and two input fields at the bottom: a text box for pasting a link and a button for searching.

Find image source



Search Text Translate



Tabby cat



Egyptian Mau



Domestic short-haired...



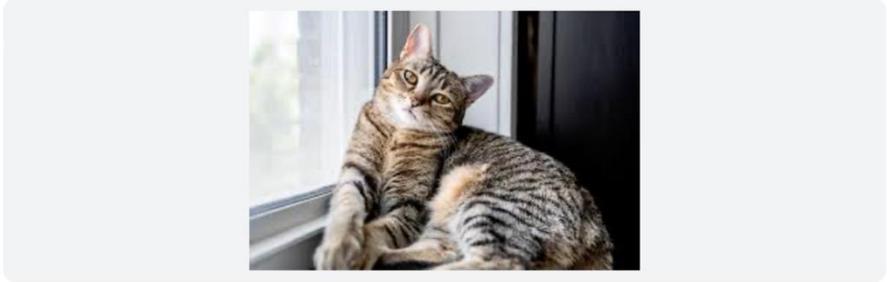
Ceylon



European shorthair

Tabby cat

Search



# https://maps.google.com

Best 55 min 1h 31m 11 hr 2h 59m

Your location

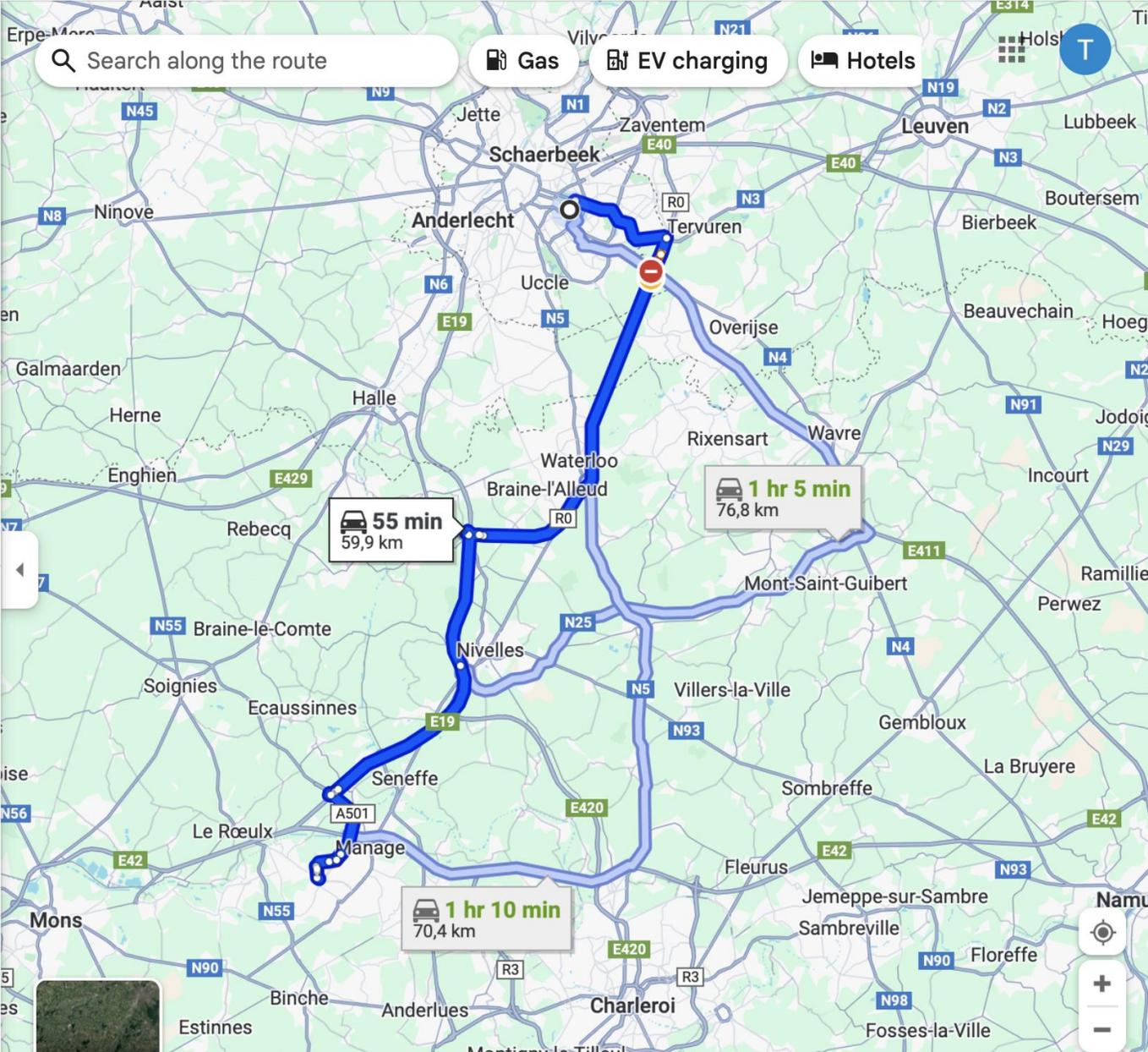
La Louvière

Add destination

Leave now Options

View directions on your phone Copy link

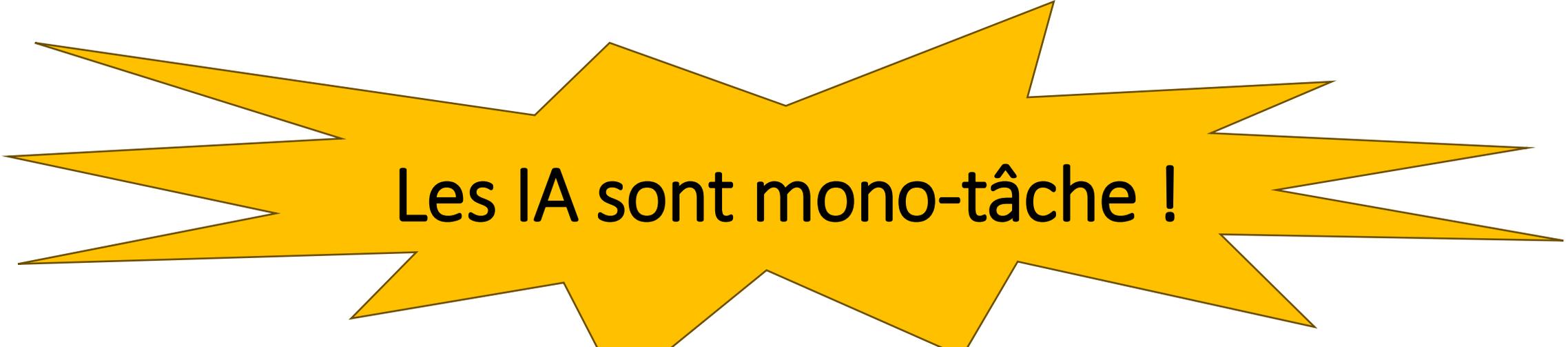
	<b>via R0 and E19</b> Fastest route now, avoids road closures	<b>55 min</b> 59.9 km
	<b>via E411 and N25</b>	<b>1 hr 5 min</b> 76.8 km
	<b>via N5</b>	<b>1 hr 10 min</b> 70.4 km



# Paradoxe de Moravec

« Il est relativement simple de faire en sorte que les ordinateurs fassent preuve de performances du niveau d'un adulte, lors de tests d'intelligence ou de parties de dames, et difficile ou impossible de les doter des capacités d'un enfant d'un an pour ce qui touche à la perception et à la mobilité. »

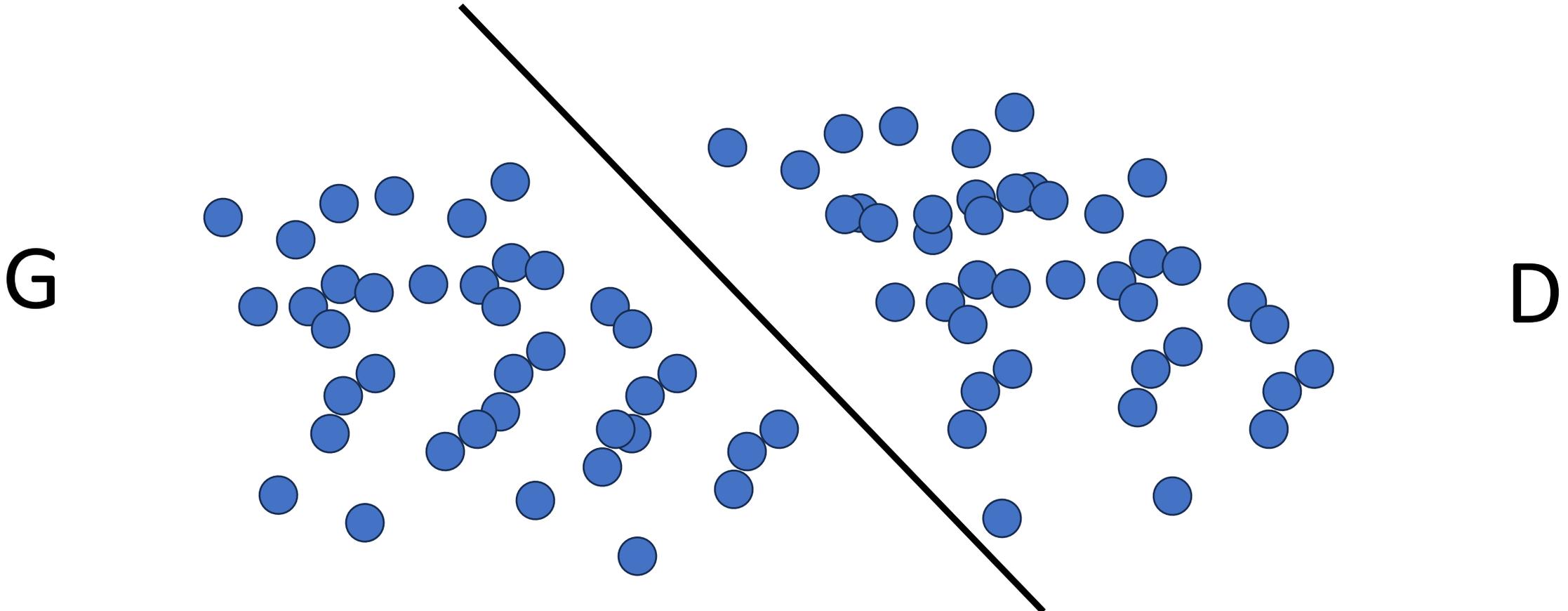
Hans Moravec (Les Enfants de l'esprit, 1988)



**Les IA sont mono-tâche !**

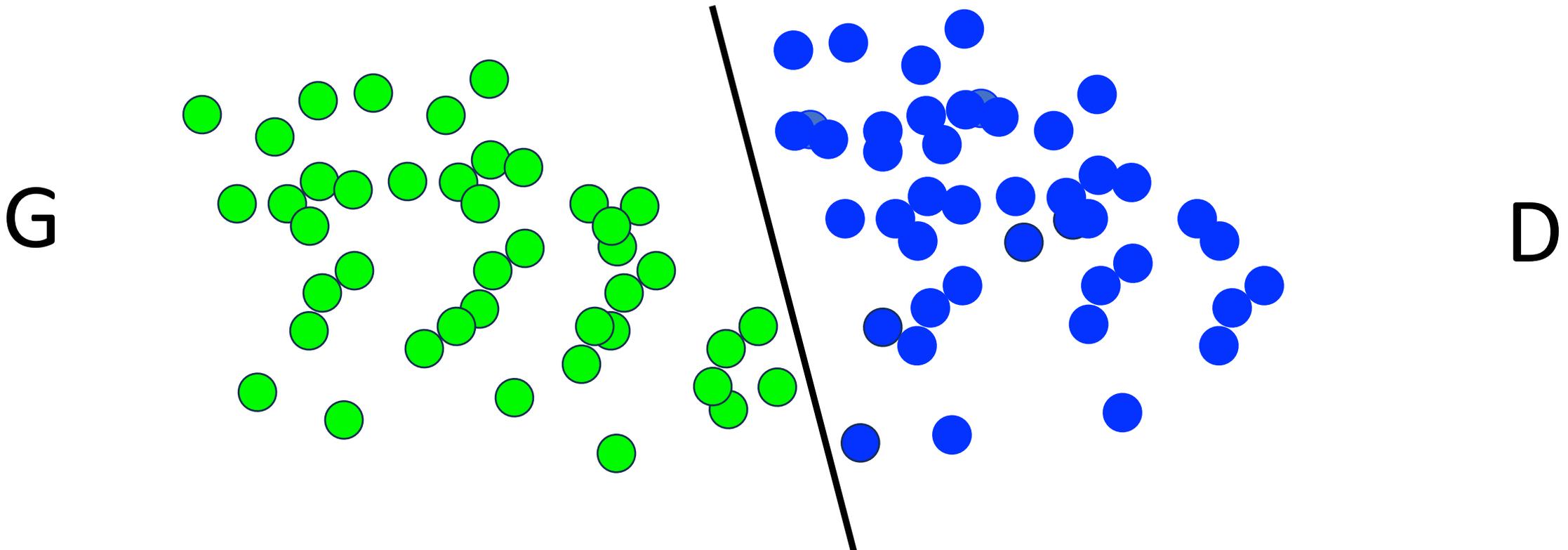
# Comment ça fonctionne (en quelques mots)

Tache : créer un algorithme qui identifie les points suivants comme “gauche” ou “droite”



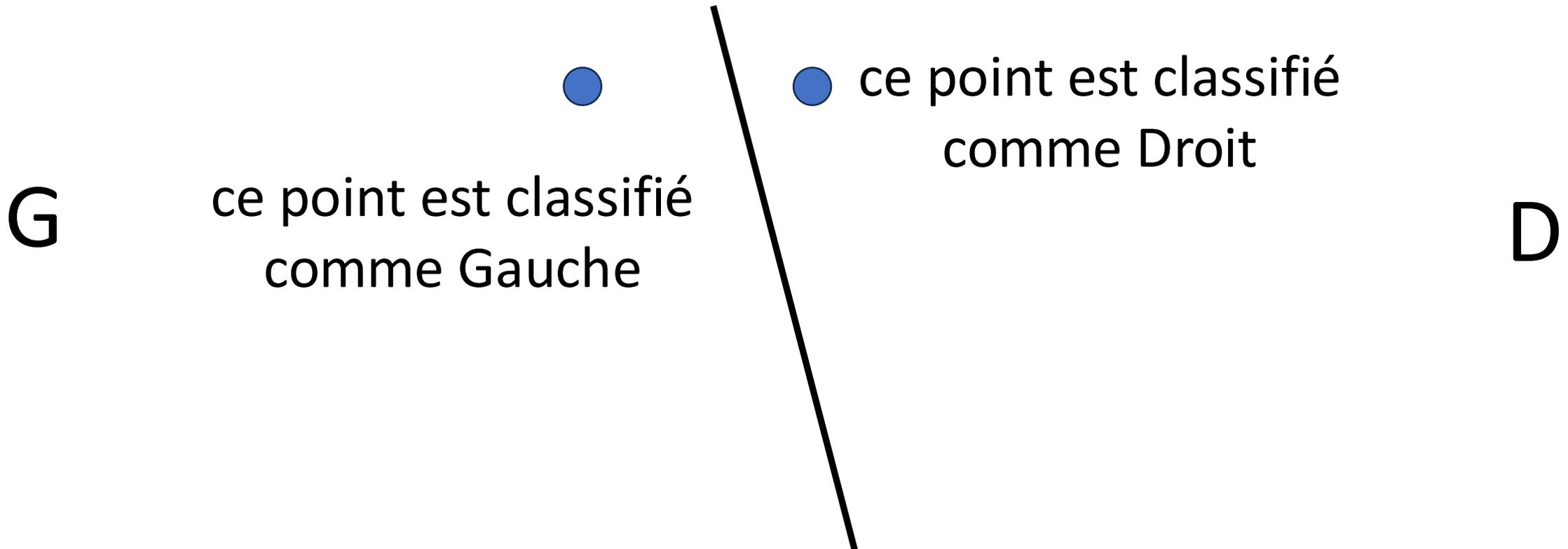
# Comment ça fonctionne (en quelques mots)

Entraînement : on utilise des points  
avec des "étiquettes"



# Comment ça fonctionne (en quelques mots)

Validation : on utilise des nouveaux points,  
la ligne ne bouge plus



# Classification : Données annotées avec catégories définies

Apprendre :



Données d'entraînement  
**Résultats = catégories définies**  
(classe)



Algorithme d'apprentissage



Modèle de **classification**  
(algorithme)

Prédire :



Nouvelle donnée

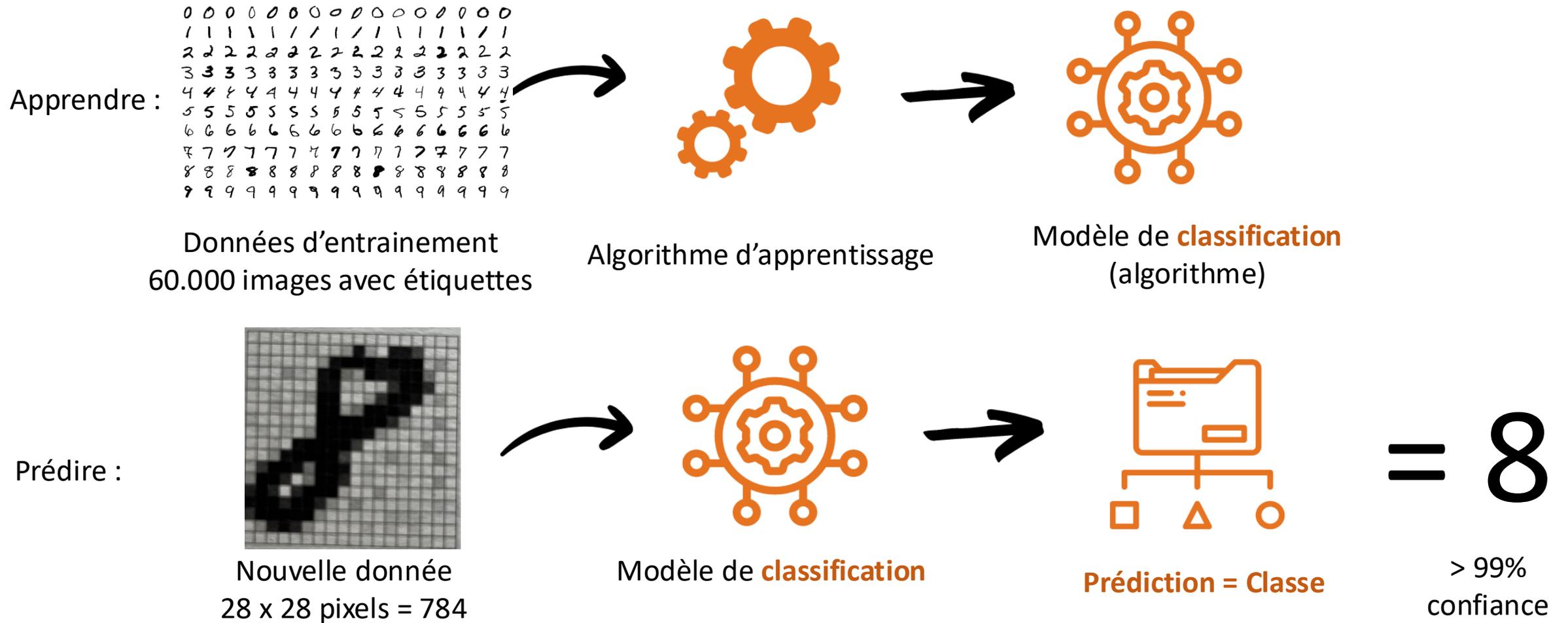


Modèle de **classification**



**Prédiction = Classe**

# Reconnaissance des chiffres manuscrites (MNIST)



- ✓ besoin beaucoup de données annotées;
- ✓ problèmes éthiques (main d'oeuvre);
- ✓ problèmes écologiques (énergie, eau) ;

# Un cout humain pour l'annotation

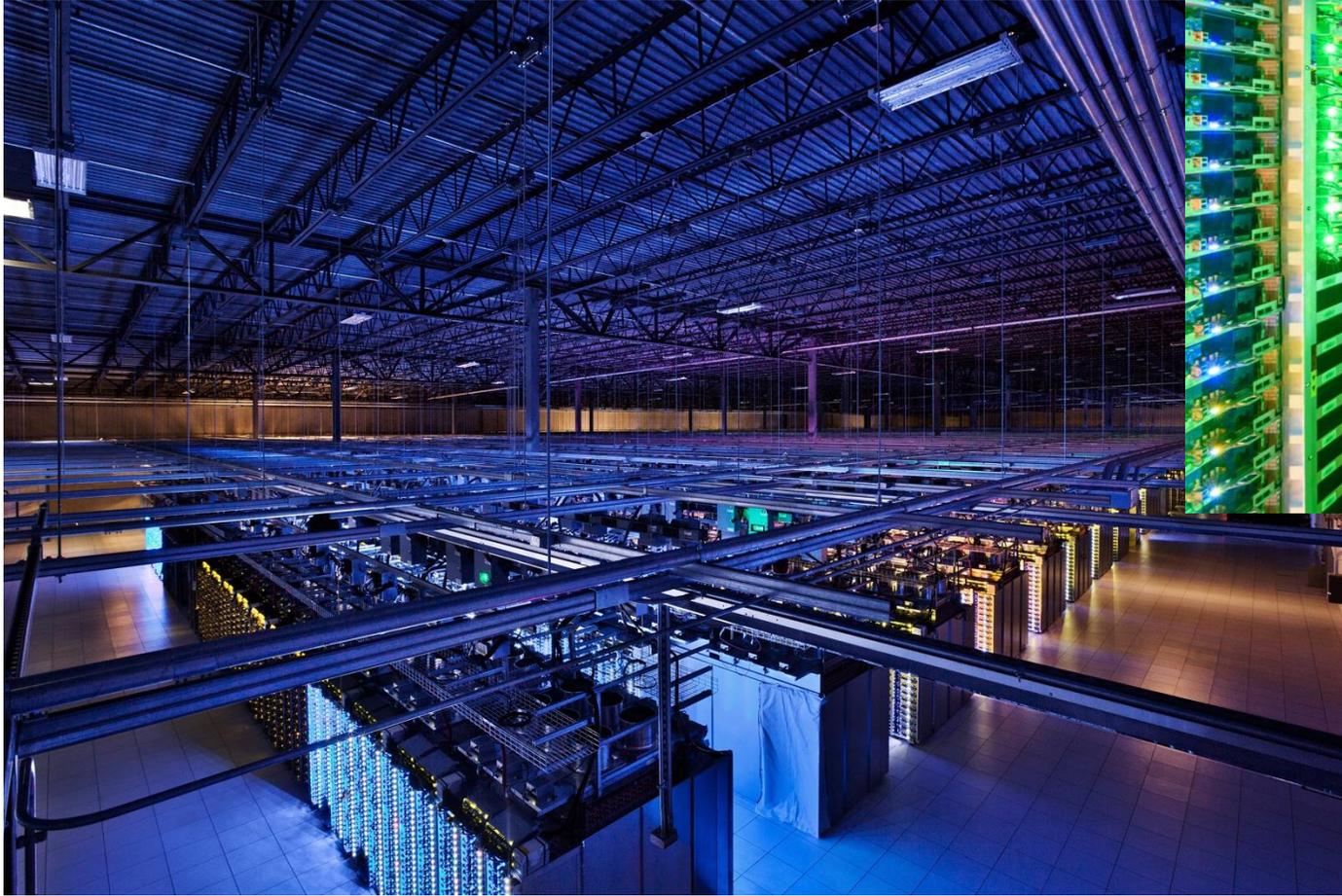


Illustration du Time générée par Dall-e avec la description :

*A seemingly endless view of African workers at desks in front of computer screens in a printmaking style*

Une vue apparemment sans fin de travailleurs africains assis à des bureaux devant des écrans d'ordinateur dans un style de gravure

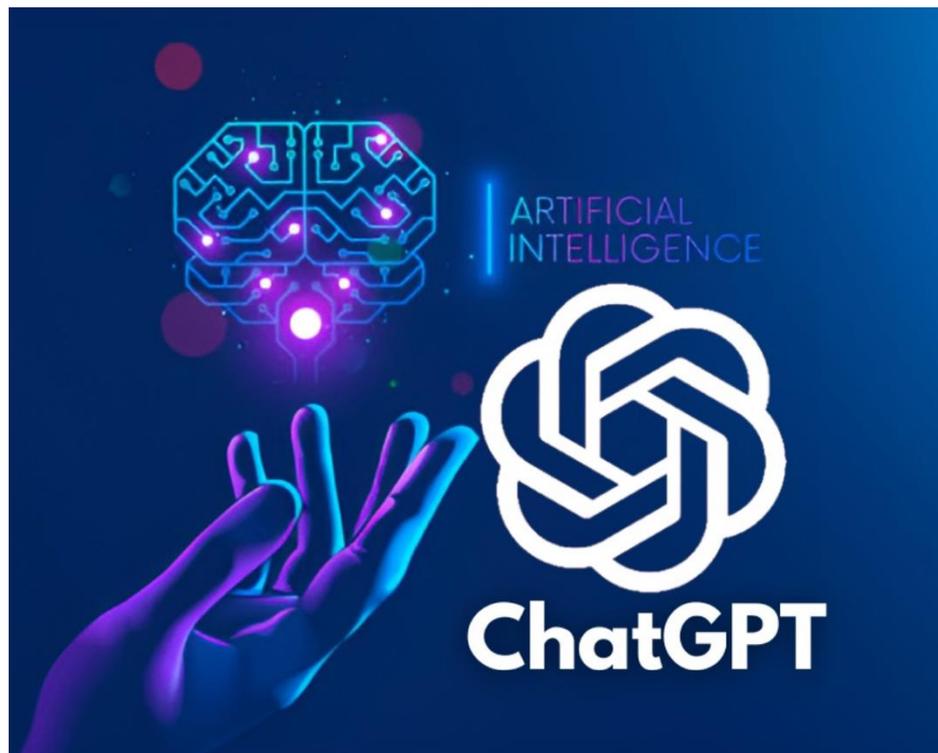
# les serveurs



# Les serveurs à construire, alimenter en énergie et refroidir...



Centre de données Google à Saint-Ghislain, Belgique



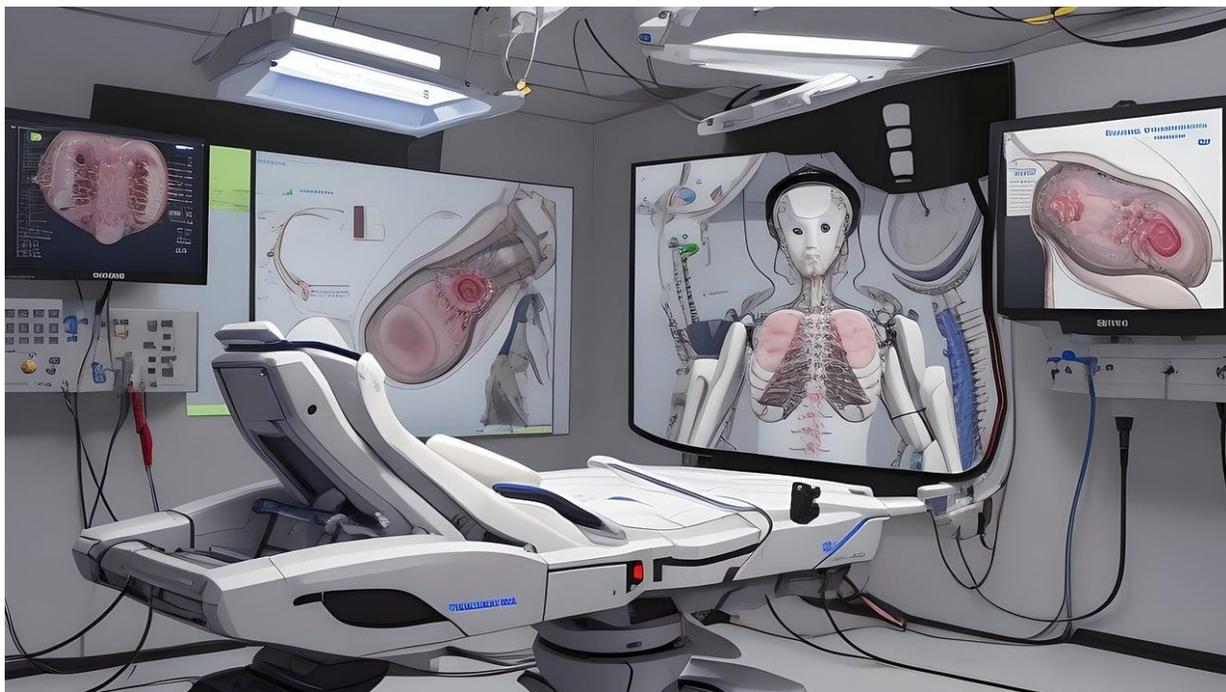
Entrainement de ChatGPT3 (175 milliards de paramètres)

- 1,4 GWh pour l'entrainement
- 34 jours
- production de une centrale nucléaire pendant 1h

- 1 « query » = 0.02 cents (10 plus que google)
- Un jour = 200 millions « queries » = 620MWh / jours = 20 fois un ménage moyen US (par jour)
- Un an ChatGPT = toute la Belgique pour un jour

<https://balkangreenenergynews.com/chatgpt-consumes-enough-power-in-one-year-to-charge-over-three-million-electric-cars/>

# Utilisation dans le monde médicale



- Dans le monde :
  - 2ème secteur le plus impacté
- Dans les hopitaux belges :
  - 95% considèrent que l'IA est un sujet important pour les hôpitaux
  - 30% ont connaissance de dispositifs IA mis en œuvre dans leur établissement

L'adoption de l'Intelligence Artificielle dans les hôpitaux en Belgique *Baromètre – Mars 2022*

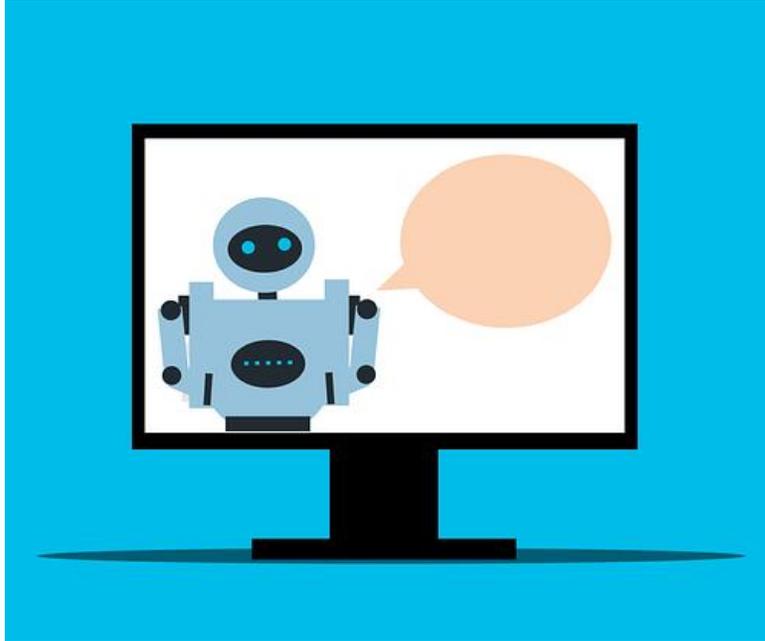
# Une IA pour diagnostiquer des mélanomes ?



**Demetra : La dermatologie numérique à portée de main**

**.AGORIA**

# Relation entre IA et soignant·e·s

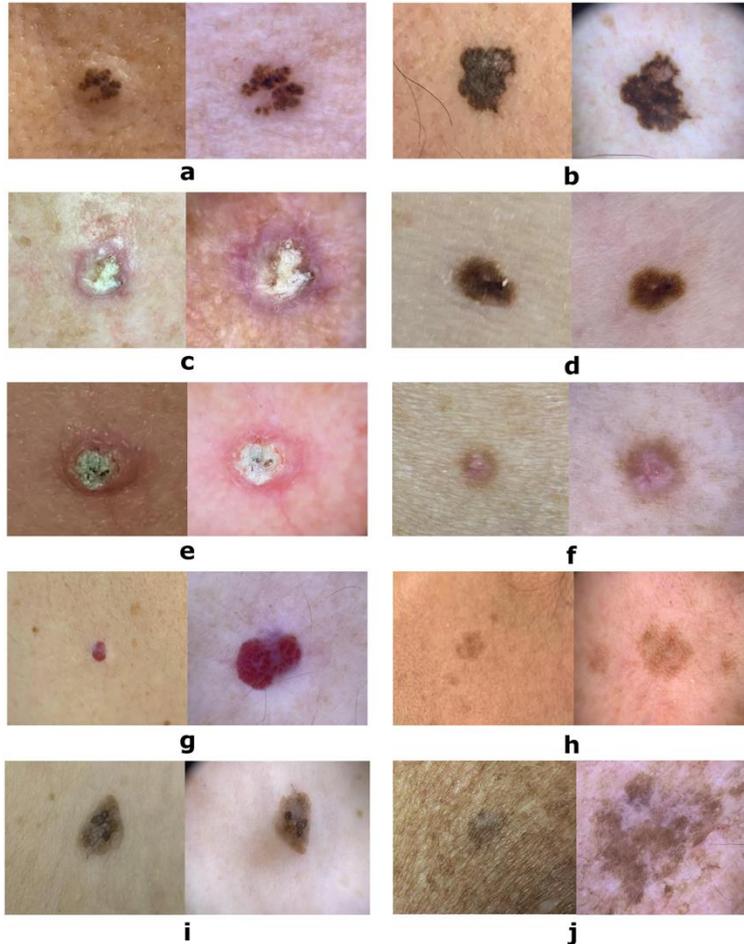


VS



- Qui est le plus exacte ?
- Qui pose le diagnostic ?
- Qui porte la responsabilité ?

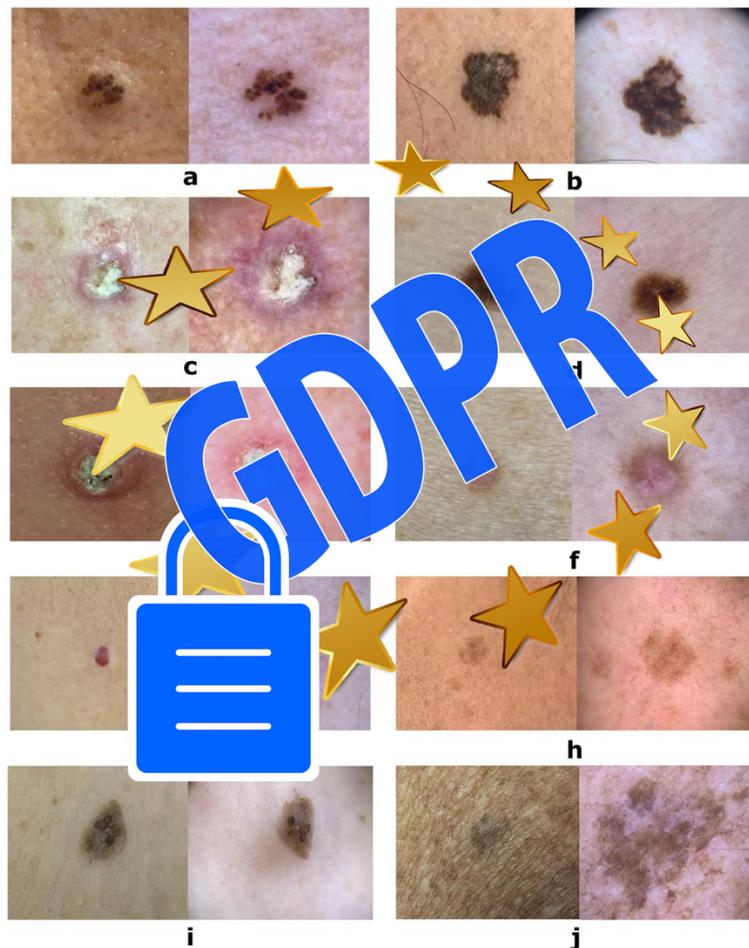
# Des données d'entraînement



*A dataset of skin lesion images collected in argentina for the evaluation of AI tools in this population*

María agustina Ricci Lara, María VictoriaRodríguez Kowalczuk, Maite Lisa Eliceche, María Guillermina Ferraresso, Daniel Roberto Luna, Sonia Elizabeth Benitez & Luis Daniel Mazzuocolo

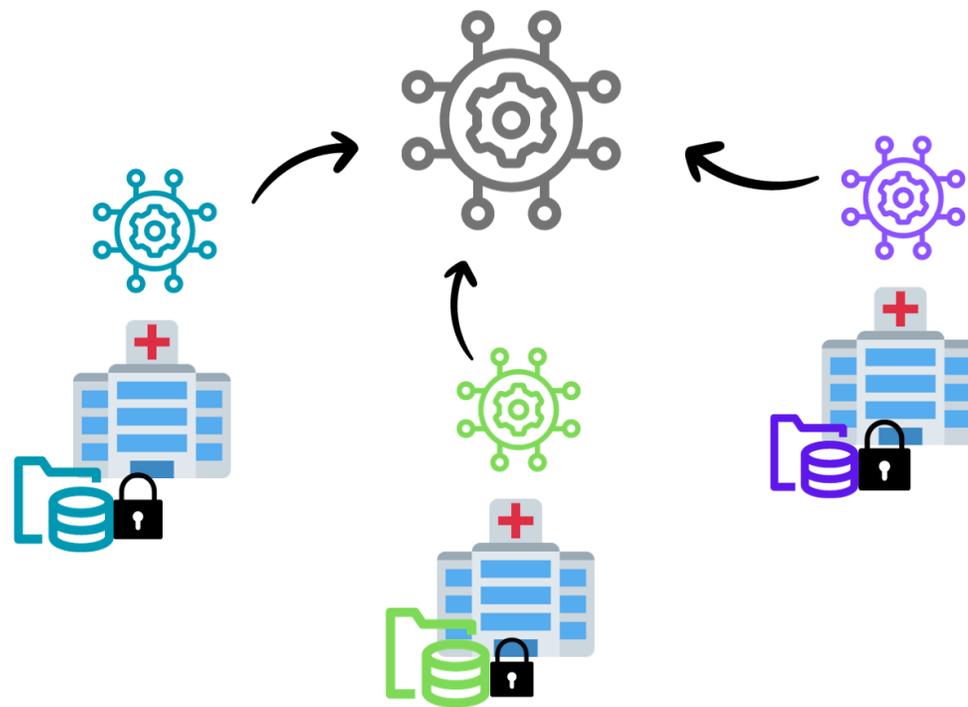
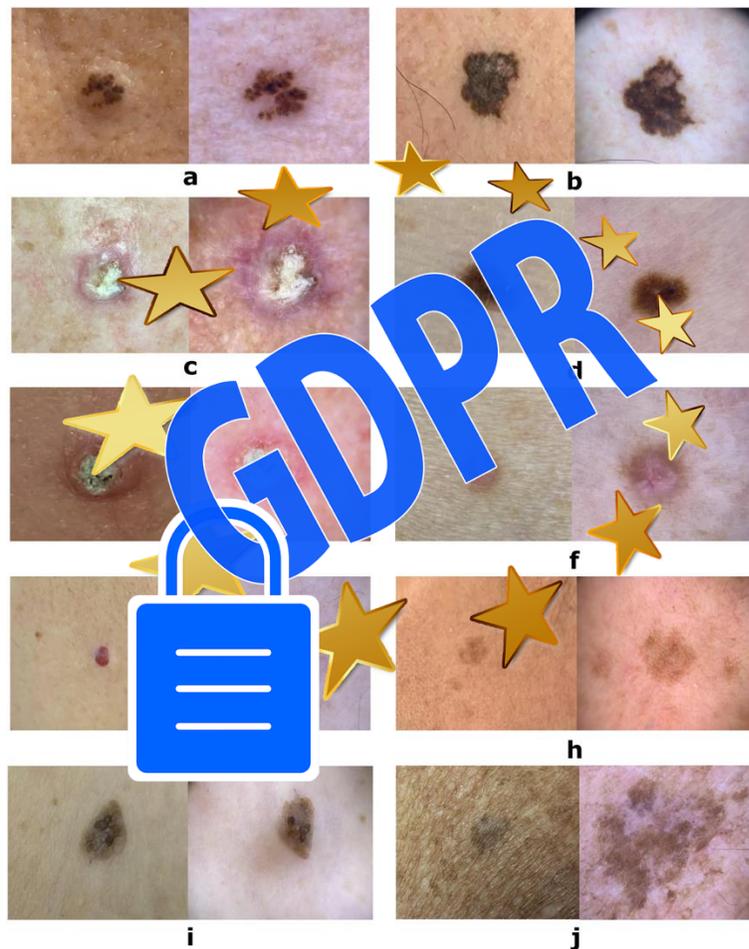
# Des données d'entrainement... et la vie privée



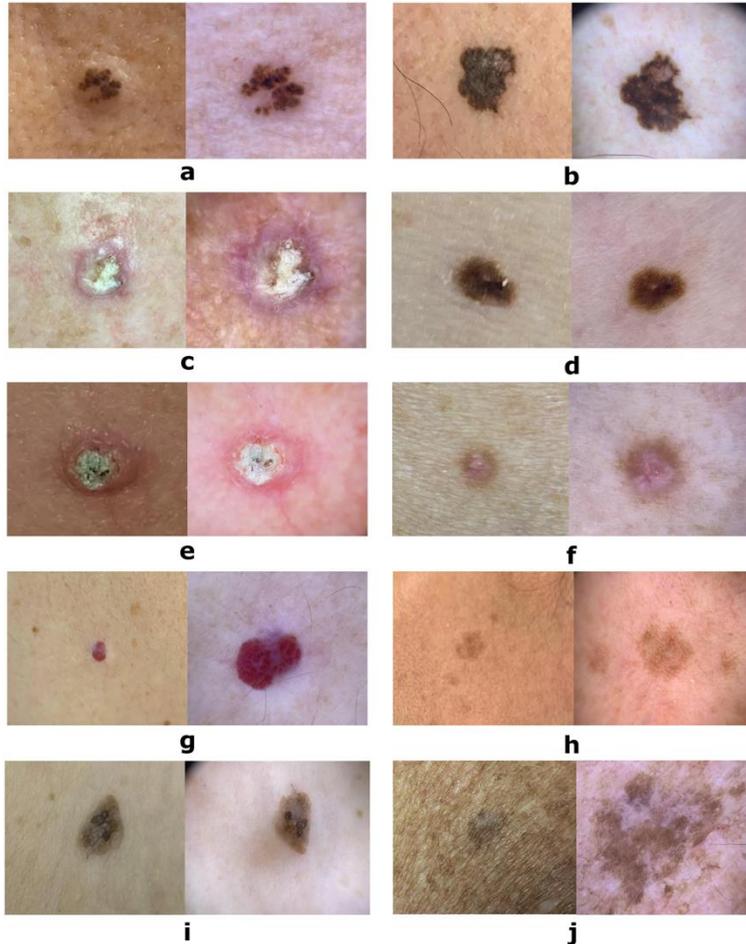
*A dataset of skin lesion images collected in argentina for the evaluation of AI tools in this population*

María agustina Ricci Lara, María VictoriaRodríguez Kowalczuk, Maite Lisa Eliceche, María Guillermina Ferraresso, Daniel Roberto Luna, Sonia Elizabeth Benitez & Luis Daniel Mazzuocolo

# Des données d'entrainement... et la vie privée



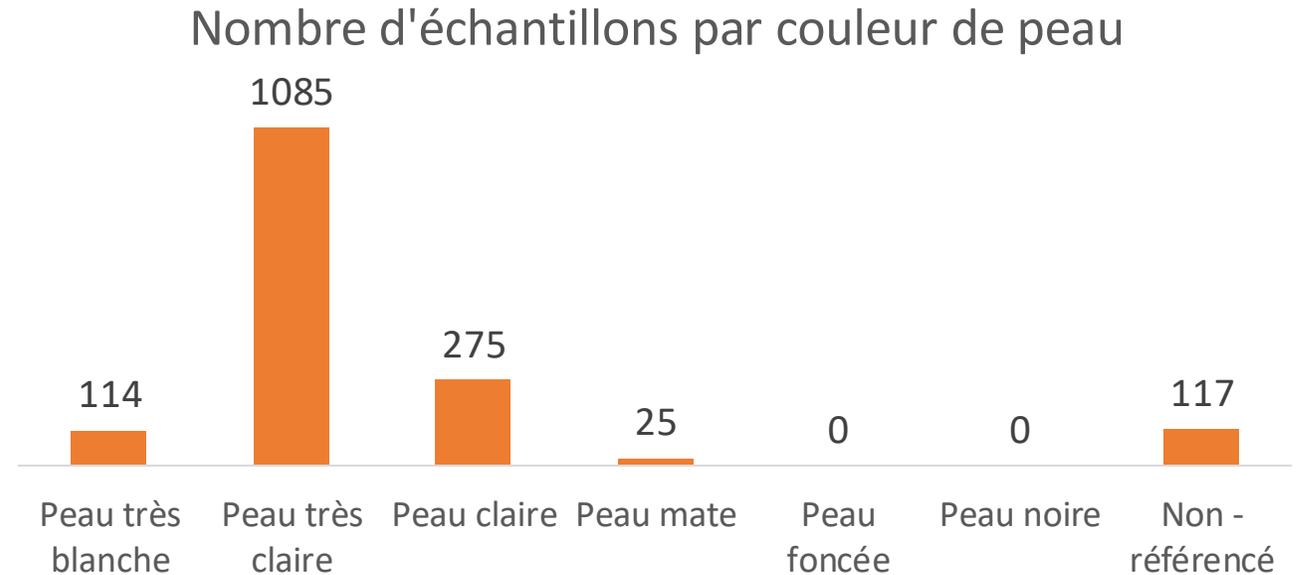
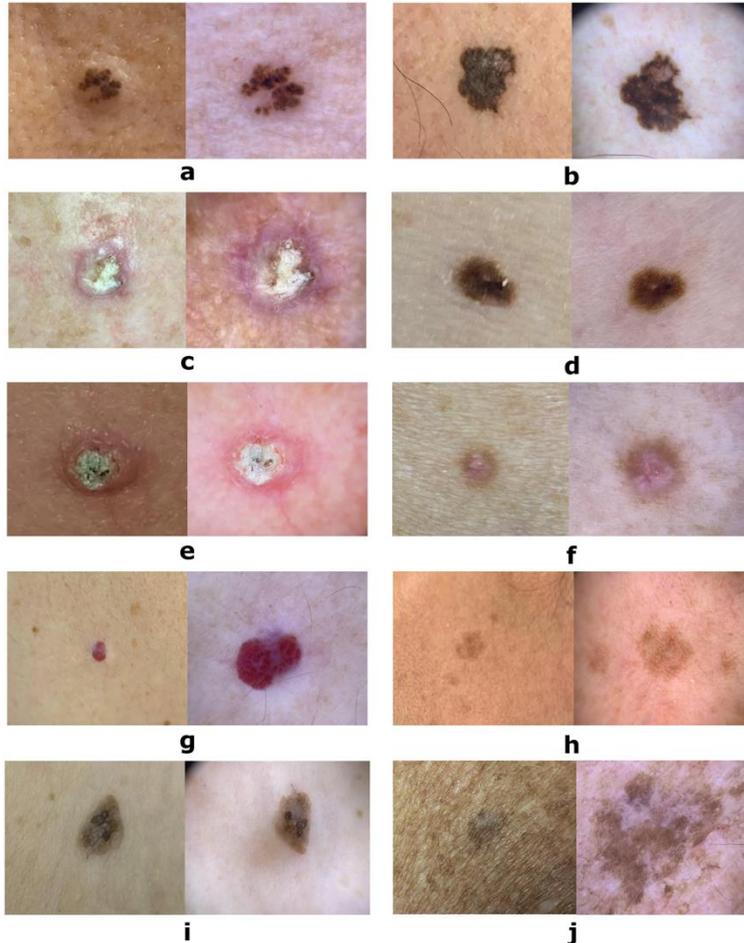
# Des données d'entraînement



*A dataset of skin lesion images collected in argentina for the evaluation of AI tools in this population*

María agustina Ricci Lara, María VictoriaRodríguez Kowalczuk, Maite Lisa Eliceche, María Guillermina Ferraresso, Daniel Roberto Luna, Sonia Elizabeth Benitez & Luis Daniel Mazzuocolo

# Des données d'entraînement... très blanches

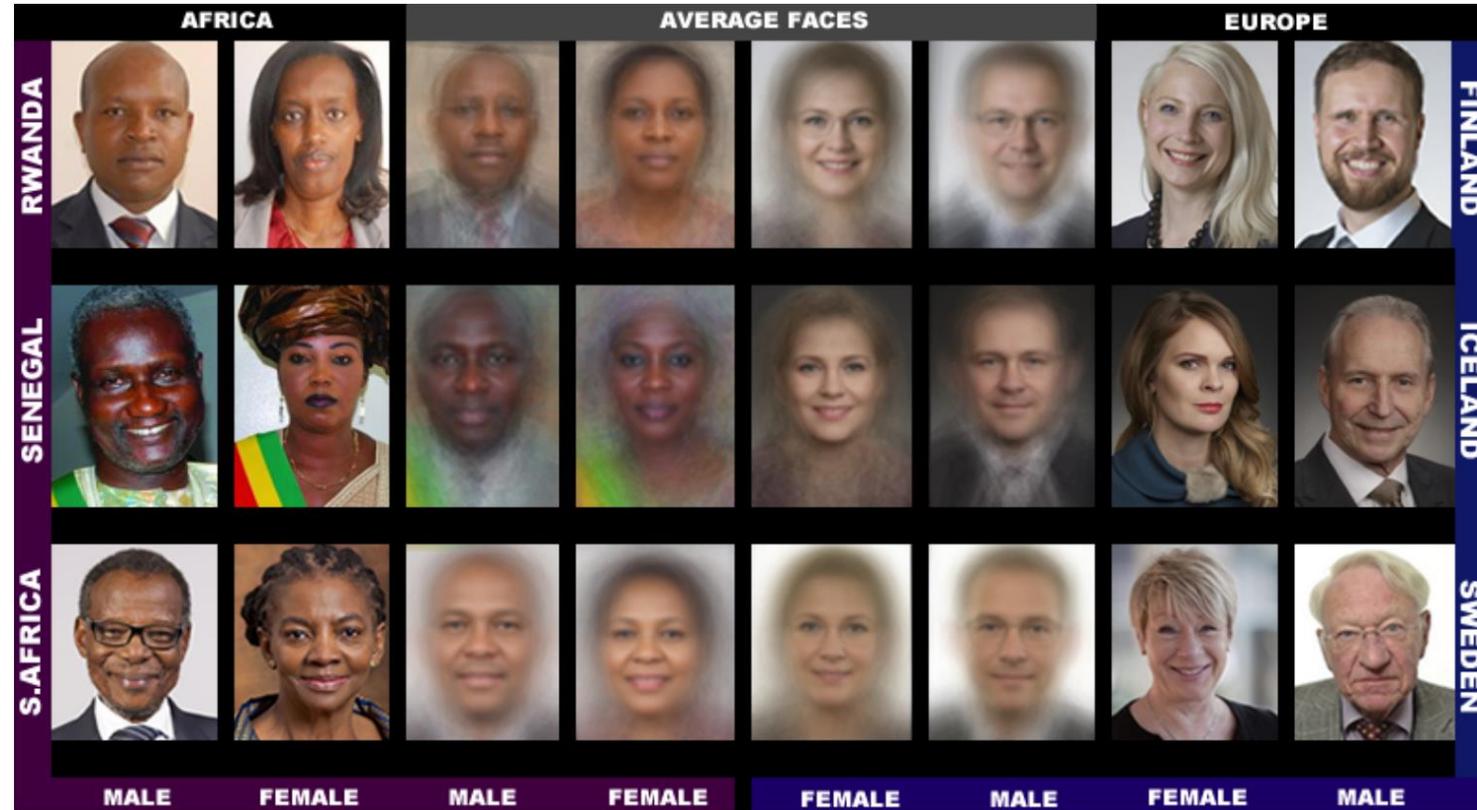


*A dataset of skin lesion images collected in argentina for the evaluation of AI tools in this population*

María agustina Ricci Lara, María Victoria Rodríguez Kowalczyk, Maite Lisa Eliceche, María Guillermina Ferrareso, Daniel Roberto Luna, Sonia Elizabeth Benitez & Luis Daniel Mazzuocolo

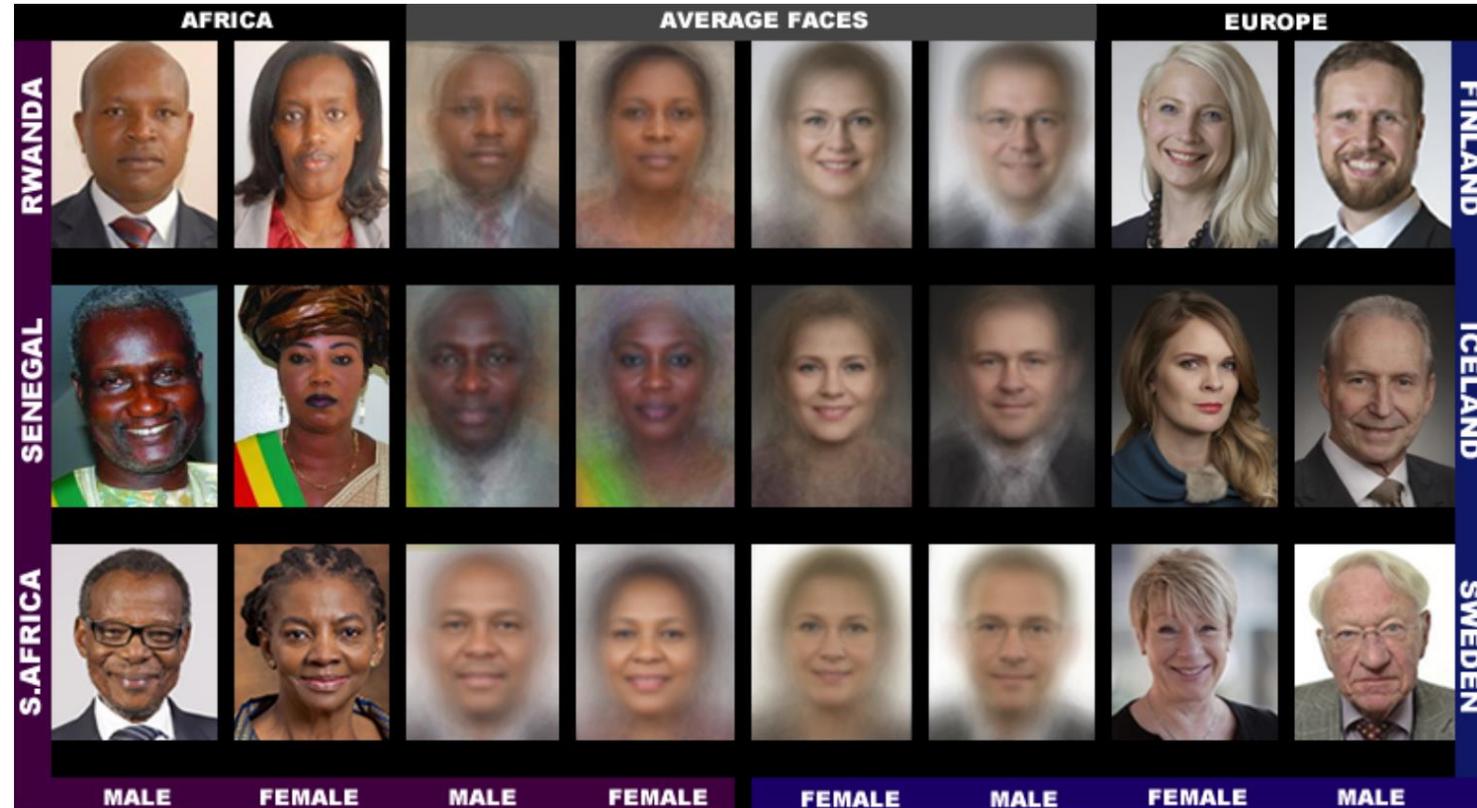
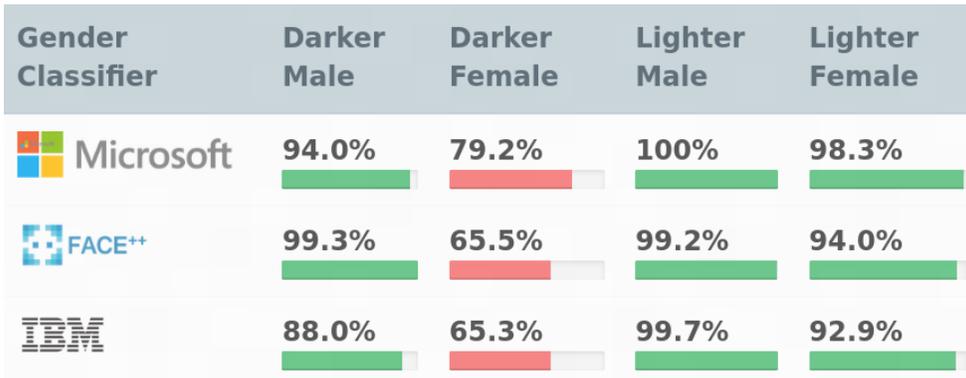
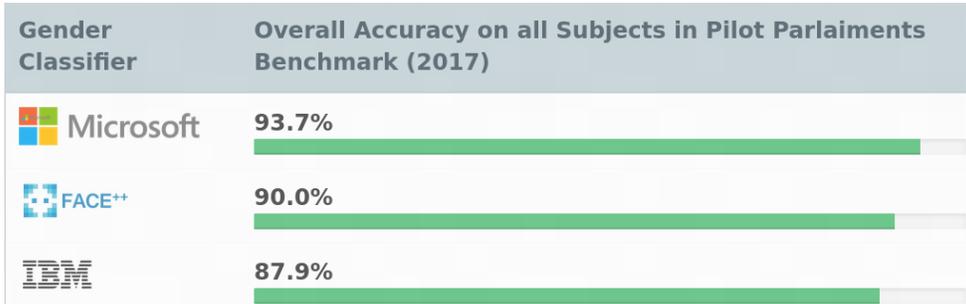
# Tache : Classification du genre

Gender Classifier	Overall Accuracy on all Subjects in Pilot Parliaments Benchmark (2017)
 Microsoft	93.7%
 FACE++	90.0%
 IBM	87.9%



<http://gendershades.org>

# Tache : Classification du genre

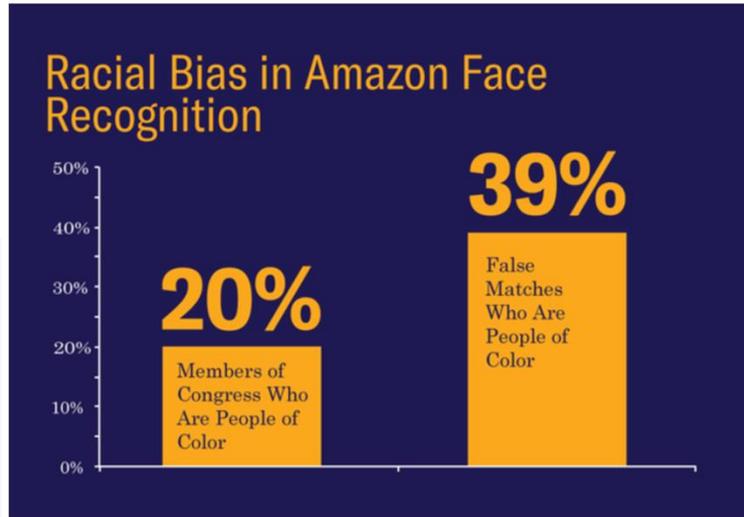
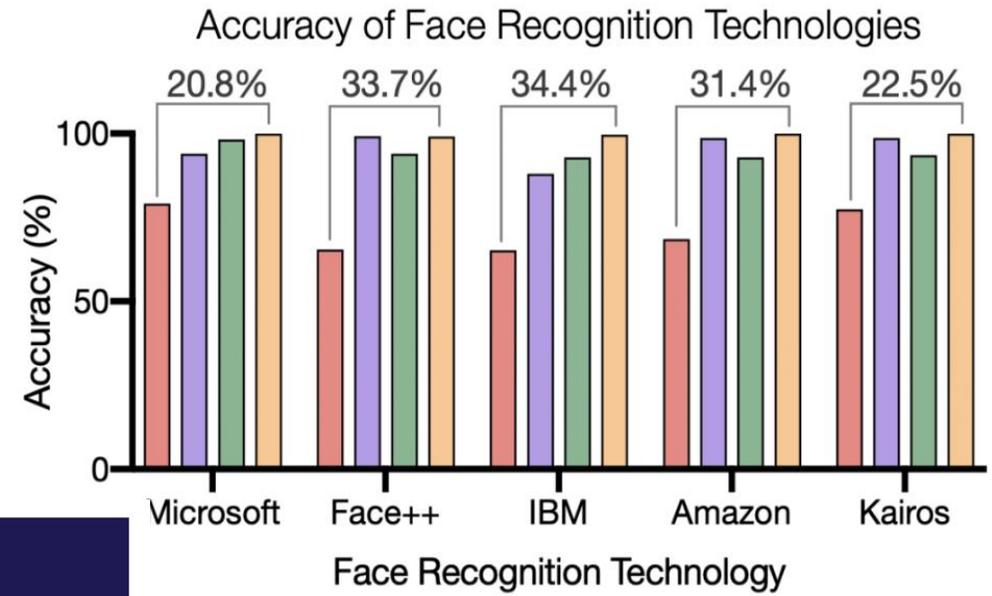


<http://gendershades.org>



## Racial bias in facial recognition algorithms

- Darker female
- Darker male
- Lighter female
- Lighter male



People of color were disproportionately falsely matched in our test.

KHARI JOHNSON BUSINESS MAR 7, 2022 7:00 AM

**WIRED**

## How Wrongful Arrests Based on AI Derailed 3 Men's Lives

Racial gap in U.S. arrest rates: 'Staggering disparity'

BRAD HEATH | USA TODAY



Show caption ▾

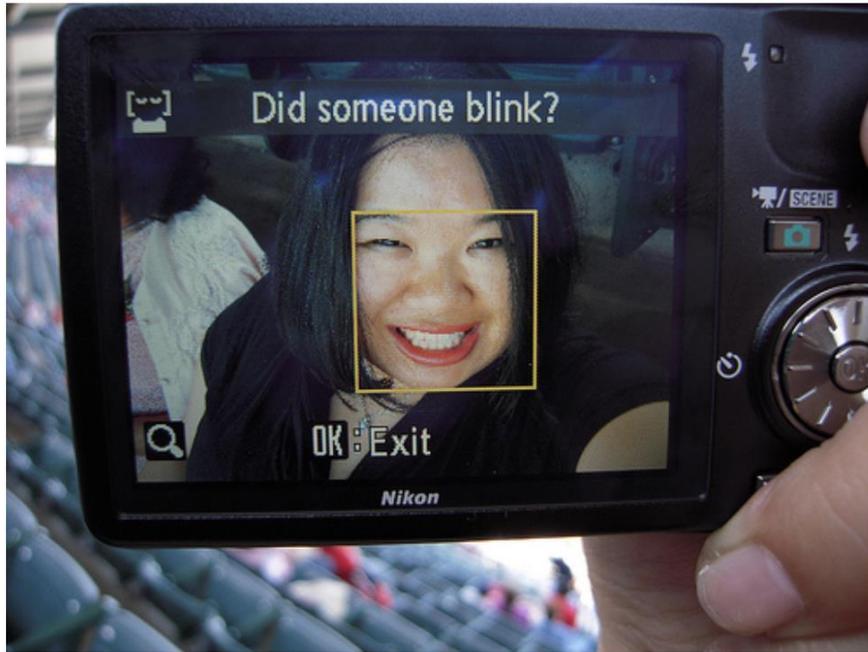
# SOCIOLOGICAL IMAGES

[Blog](#) [For Instructors](#) [About SocImages](#) [Submitting Guest Posts](#)

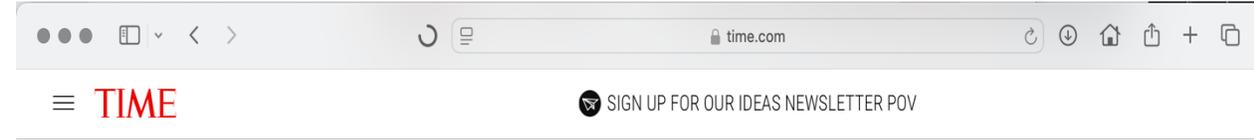
## Nikon Camera Says Asians: People Are Always Blinking

Gwen Sharp, PhD on May 29, 2009

Toban B., Elisabeth, and Mark sent us a link to a post at [jozjozjoz](#) about the Nikon S630 digital camera. As Joz explains, “As I was taking pictures of my family, it kept asking ‘Did someone blink?’ even though our eyes were always open.”



Apparently the camera perceives “Asian” eyes as closed.



TECH

## Are Face-Detection Cameras Racist?

8 MINUTE READ

ADAM ROSE

JANUARY 22, 2010 12:00 AM EST

When Joz Wang and her brother bought their mom a Nikon Coolpix S630 digital camera for Mother’s Day last year, they discovered what seemed to be a malfunction. Every time they took a portrait of each other smiling, a message flashed across the screen asking, “Did someone blink?” No one had. “I thought the camera was broken!” Wang, 33, recalls. But when her brother posed with his eyes open so wide that he looked “bug-eyed,” the messages stopped.

[thesocietypages.org/socimages/2009/05/29/nikon-camera-says-asians-are-always-blinking/](https://thesocietypages.org/socimages/2009/05/29/nikon-camera-says-asians-are-always-blinking/)

Que « voit-elle » une IA?

Tache :

Identifier animaux sauvages vs panoramas



# Hallucination. Comment falsifier un résultat



+



=



Panda :  
57% confiance

Gibbon :  
99% confiance

EXPLAINING AND HARNESSING  
ADVERSARIAL EXAMPLES

**Ian J. Goodfellow, Jonathon Shlens & Christian Szegedy**  
Google Inc., Mountain View, CA  
{goodfellow, shlens, szegedy}@google.com

# Hallucination. Comment falsifier un résultat

Kaylee Defer



Nancy Trevis



Kaylee Defer

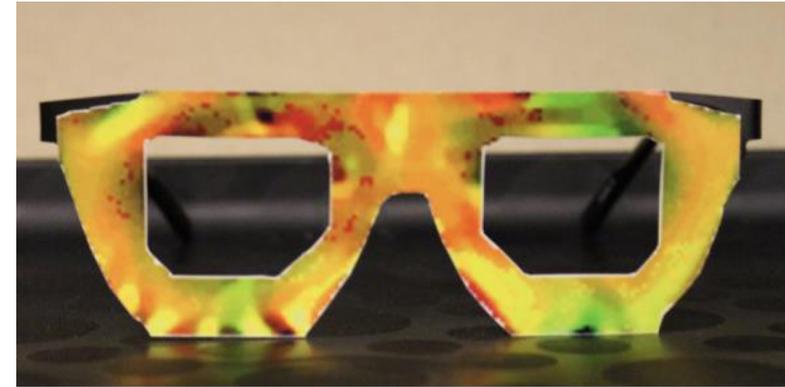
+

“Lunettes  
dessinées”

# Hallucination. Comment falsifier un résultat



Vraies lunettes



## Accessorize to a Crime: Real and Stealthy Attacks on State-of-the-Art Face Recognition

Mahmood Sharif  
Carnegie Mellon University  
Pittsburgh, PA, USA  
mahmoods@cmu.edu

Sruti Bhagavatula  
Carnegie Mellon University  
Pittsburgh, PA, USA  
srutib@cmu.edu

Lujo Bauer  
Carnegie Mellon University  
Pittsburgh, PA, USA  
lbauer@cmu.edu

Michael K. Reiter  
University of North Carolina  
Chapel Hill, NC, USA  
reiter@cs.unc.edu

# https://www.capable.design



designed by \_\_\_\_\_ rachele didero

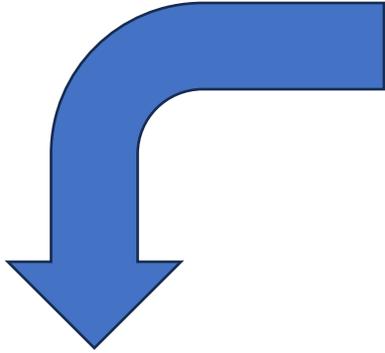
[Home](#) [Mission](#) [Collection](#) [News](#) [Shop](#) [Team](#)

EN 



# Autocollant sur panneau de signalisation

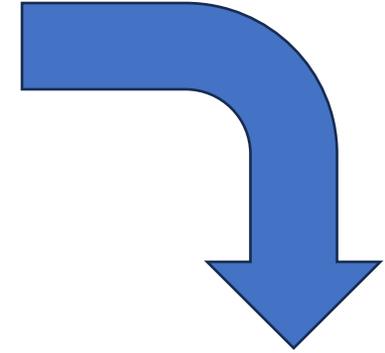
Humain



stop



IA



Limitation  
vitesse  
45km/h

**Robust Physical-World Attacks on Deep Learning Visual Classification**

Kevin Eykholt<sup>\*1</sup>, Ivan Evtimov<sup>\*2</sup>, Earlene Fernandes<sup>2</sup>, Bo Li<sup>3</sup>,  
Amir Rahmati<sup>4</sup>, Chaowei Xiao<sup>1</sup>, Atul Prakash<sup>1</sup>, Tadayoshi Kohno<sup>2</sup>, and Dawn Song<sup>3</sup>

<sup>1</sup>University of Michigan, Ann Arbor

<sup>2</sup>University of Washington

<sup>3</sup>University of California, Berkeley

<sup>4</sup>Samsung Research America and Stony Brook University

# Encore "Création" de musique

suno.com

SUNO

Radiofrance

USA 2024

Rechercher

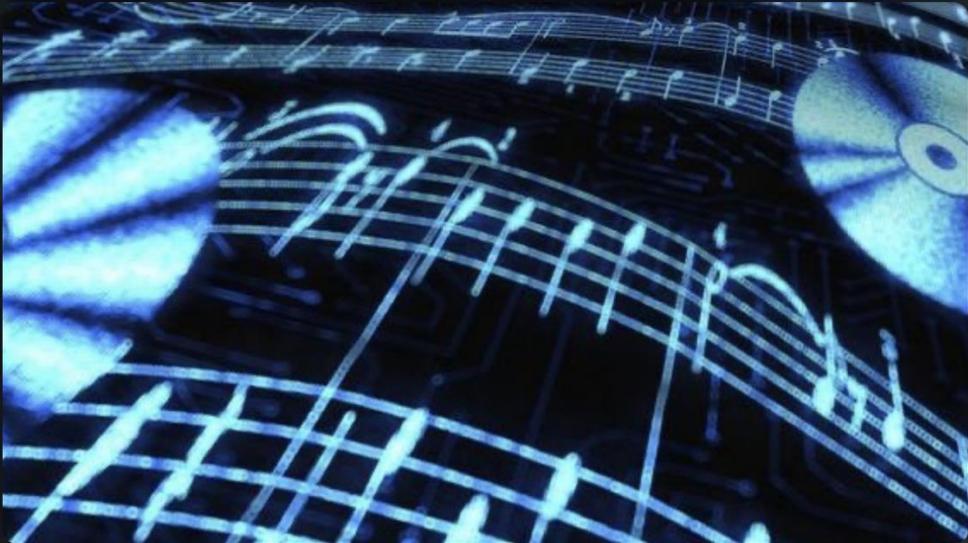
Se connecter

franceinfo: Grille des programmes Podcasts Titres diffusés

## États-Unis : un escroc crée des centaines de milliers de chansons grâce à l'intelligence artificielle

Jeudi 12 septembre 2024

ÉCOUTER (2 min)



Ce type d'arnaque est une première pour le bureau de procureur de New York, selon le New York Times (photo d'illustration). - ALENGO / E+



illustration: Tosca van der Weerden

---

## Prédire pour détecter.. la fraude

**Pays-Bas. Scandale des allocations familiales : un avertissement qui montre l'urgence d'interdire les algorithmes racistes**

**AMNESTY  
INTERNATIONAL**



Nom	Enfants	Revenus	...	Nationalité	Classe
Maries De vries	3	20 000€		Ndls	ok
Jan Mulder	42	15 000€		Ndls	X
...					...
Magali Legast	2	250 000€		Non-Ndls	X
Timoteo Carletti	1	32 000€		Non-Ndls	ok



Nom	Enfants	Revenus	...	Nationalité	Classe
Maries De vries	3	20 000€		Ndls	ok
Jan Mulder	42	15 000€		Ndls	X
...					...
Magali Legast	2	250 000€		Non-Ndls	X
Timoteo Carletti	1	32 000€		Non-Ndls	ok



Score de risque ?



= 6

Action automatique en fonction du score !

# Modèle de classification du risque : Résultats



## XENOPHOBIC MACHINES

- Dizaines de milliers de personnes accusées à tort de fraude
- Sommes importantes à rembourser (dettes, expulsions, stress, divorces, ...)
- Profilage racial
- Démission du gouvernement (2021)



AMNESTY  
INTERNATIONAL



illustration: Tosca van der Weerden

# Comment on en est arrivé là ?

- ❌ Manque de transparence (boite noire + autorités)
- ❌ Résultats IA aggravés par décisions humaines
- ❌ Recours très difficile
- ❌ Inclusion d'un critère de nationalité

# Et maintenant, où est-ce qu'on va ?

*Artificial Intelligence Act (AIAct)*

Adoptée en mai 2024

Application graduelle dès août 2024



Objectifs annoncés par le parlement de l'UE :

- Respecter les droits fondamentaux
- Stimuler l'investissement et l'innovation
- IA sûres et dignes de confiance
- Exceptions : recherche et militaire/défense

# Pourquoi avons-nous besoin de règles sur l'IA?

La législation sur l'IA garantit que les Européens peuvent faire confiance à ce que l'IA a à offrir. Alors que la plupart des systèmes d'IA présentent un risque limité voire nul et peuvent contribuer à résoudre de nombreux défis sociétaux, certains systèmes d'IA créent des risques auxquels nous devons faire face pour éviter des résultats indésirables.

Par exemple, il est souvent impossible de savoir pourquoi un système d'IA a pris une décision ou une prédiction et a pris une mesure particulière. Il peut donc devenir difficile d'évaluer si une personne a été injustement désavantagée, par exemple dans une décision d'embauche ou dans une demande de régime d'utilité publique.

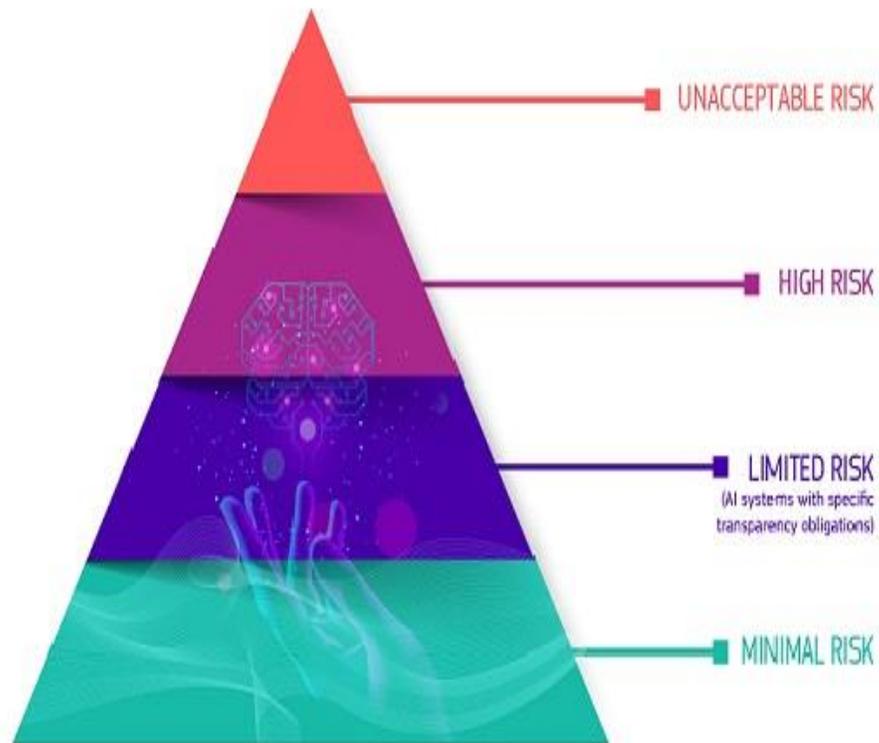
Bien que la législation existante offre une certaine protection, elle est insuffisante pour relever les défis spécifiques que les systèmes d'IA peuvent poser.

Les nouvelles règles:

- traiter les risques spécifiquement créés par les applications d'IA
- interdire les pratiques d'IA qui présentent des risques inacceptables
- établir une liste des demandes à haut risque;
- fixer des exigences claires pour les systèmes d'IA destinés aux applications à haut risque
- définir des obligations spécifiques pour les déployeurs et les fournisseurs d'applications d'IA à haut risque
- exiger une évaluation de la conformité avant la mise en service ou la mise sur le marché d'un système d'IA donné
- mettre en place des mesures d'exécution après la mise sur le marché d'un système d'IA donné
- mettre en place une structure de gouvernance aux niveaux [européen](#) , et national



# Loi européenne sur l'IA

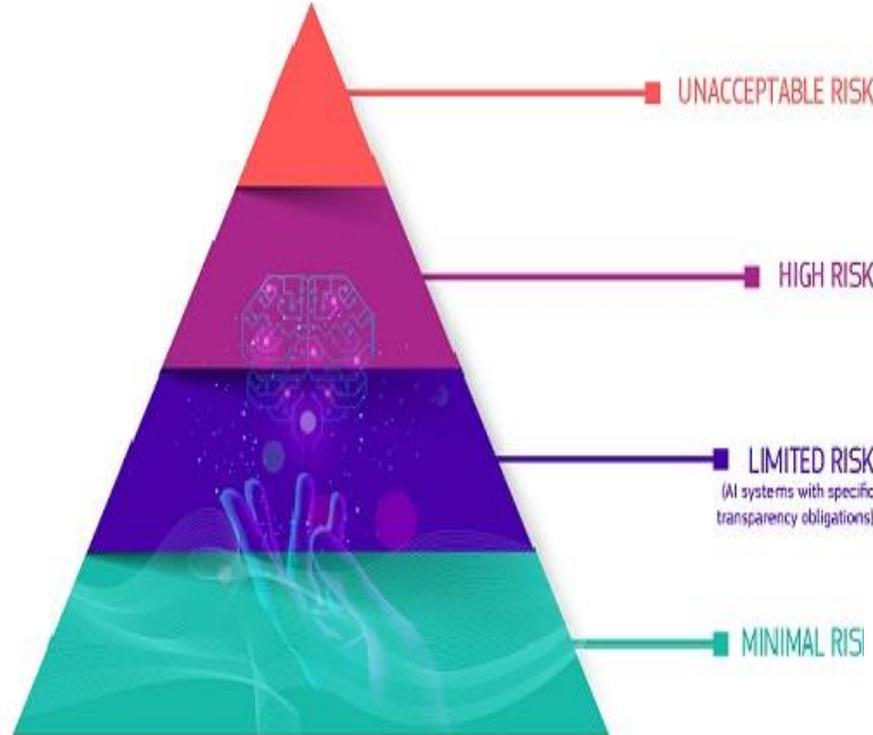


- **Risque inacceptable** -> interdit
  - Manipulation (surtout pour les groupes à risque)
  - Score social
  - Identification biométriques des groupes discriminés
  - ~~Identification biométriques dans l'espace public~~ → Quelques exceptions
- **Risque élevé** -> audit avant et pendant mise sur le marché, Supervision humaine, Transparence accrue
  - Infrastructures critiques
  - Education
  - Applications légales
  - Emploi
  - Accès aux services publics et privés essentiels
  - Migration et frontières
- **Risque limité** -> Obligations de transparence
- **Risque minimal ou nul** (Majorité des IA dans l'UE)  
-> Pas de restriction



# Loi européenne sur l'IA

## Risque élevé



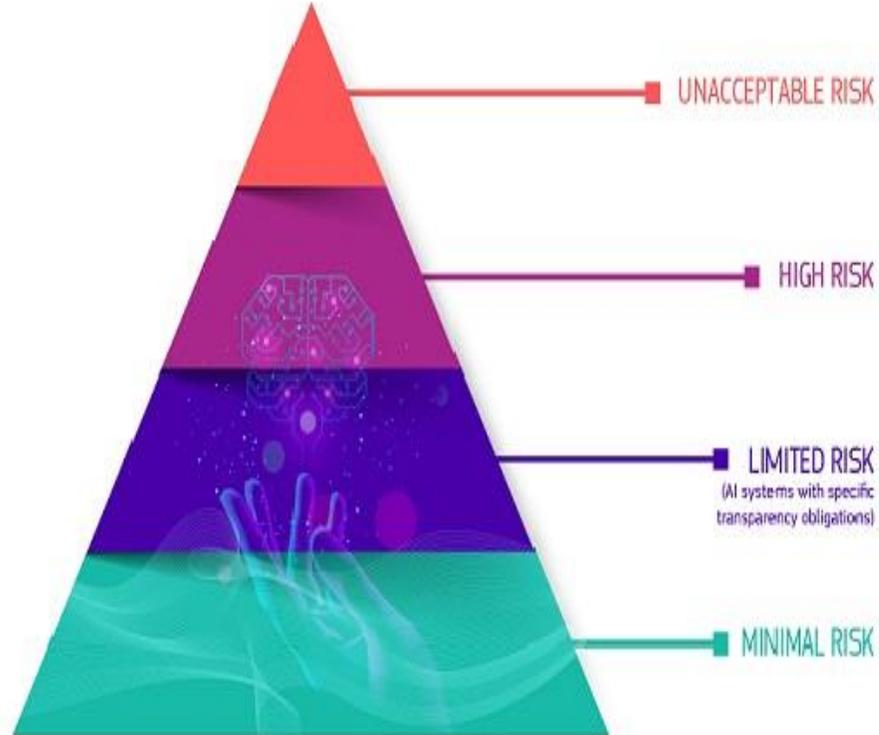
- les infrastructures critiques (par exemple, les transports), qui pourraient mettre en danger la vie et la santé des citoyens;
- la formation scolaire ou professionnelle, qui peut déterminer l'accès à l'éducation et au parcours professionnel d'une personne (par exemple, la notation des examens);
- composants de sécurité des produits (par exemple, application de l'IA en chirurgie assistée par robot)
- l'emploi, la gestion des travailleurs et l'accès au travail indépendant (par exemple, logiciel de tri des CV pour les procédures de recrutement)
- services publics et privés essentiels (par exemple, notation de crédit empêchant les citoyens d'obtenir un prêt)
- les services répressifs susceptibles d'interférer avec les droits fondamentaux des personnes (par exemple, l'évaluation de la fiabilité des éléments de preuve);
- la gestion des migrations, de l'asile et des contrôles aux frontières (par exemple, l'examen automatisé des demandes de visa);
- administration de la justice et processus démocratiques (par exemple, solutions d'IA pour rechercher des décisions de justice)

Les systèmes d'IA à haut risque sont soumis à des obligations strictes avant de pouvoir être mis sur le marché:

- des systèmes adéquats d'évaluation et d'atténuation des risques;
- qualité élevée des ensembles de données alimentant le système afin de réduire au minimum les risques et les résultats discriminatoires
- journalisation de l'activité pour assurer la traçabilité des résultats
- une documentation détaillée fournissant toutes les informations nécessaires sur le système et sa finalité pour permettre aux autorités d'évaluer sa conformité;
- information claire et adéquate du deployeur
- des mesures de surveillance humaine appropriées pour réduire au minimum les risques;
- haut niveau de robustesse, de sécurité et de précision



# Loi européenne sur l'IA



## Risque limité

Le risque limité renvoie aux risques associés au manque de transparence dans l'utilisation de l'IA. La législation sur l'IA introduit des obligations de transparence spécifiques pour veiller à ce que les êtres humains soient informés lorsque cela est nécessaire, ce qui favorise la confiance. Par exemple, lors de l'utilisation de systèmes d'IA tels que les chatbots, les humains doivent être informés qu'ils interagissent avec une machine afin qu'ils puissent prendre une décision éclairée de continuer ou de prendre du recul. Les fournisseurs doivent également veiller à ce que le contenu généré par l'IA soit identifiable. En outre, les textes générés par l'IA publiés dans le but d'informer le public sur des questions d'intérêt public doivent être étiquetés comme étant générés artificiellement. Cela s'applique également aux contenus audio et vidéo constituant des contrefaçons profondes.

## Risque minimal ou nul

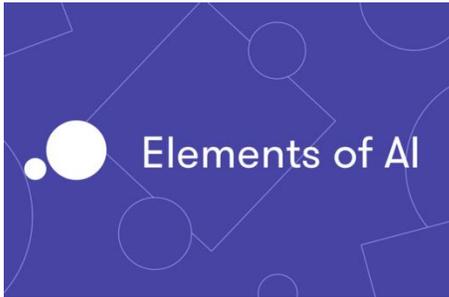
La législation sur l'IA autorise la libre utilisation de l'IA à risque minimal. Cela inclut des applications telles que les jeux vidéo compatibles avec l'IA ou les filtres anti-spam. La grande majorité des systèmes d'IA actuellement utilisés dans l'UE entrent dans cette catégorie.



# Construire des IA responsables - FAT

- **Équité (*Fairness*)**
  - Bias & Non-discrimination
  - Vie privée
- **Responsabilité (*Accountability*)**
  - Considérations éthiques
  - Niveau technique et humain
- **Transparence**
  - Système explicable

# Pour aller plus loin

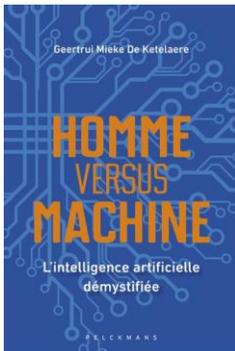


*Elementsofai.be*

Cours d'introduction en ligne pour non-spécialistes

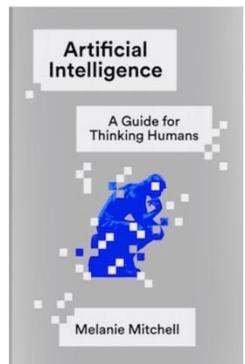
Olympiades belges d'Informatique (*be-oi.be*)

Compétition pour le secondaire – Niveau accessible



Homme versus Machine L'intelligence artificielle démystifiée,  
Geertrui Mieke De Ketelaere

Artificial Intelligence: A Guide for Thinking Humans,  
Melanie Mitchell





<https://www.youtube.com/watch?v=1u62eCiStqs>