

# Où en est l'intelligence artificielle ?

État de l'art, Performances et Perspectives

Aurélien Garivier

École Normale Supérieure de Lyon, UMPA & LIP

Mercredi 25 juin 2025



# Plan

Le futur de l'IA a une histoire

Concepts fondamentaux

Des succès spectaculaires

Faire confiance aux IA ?

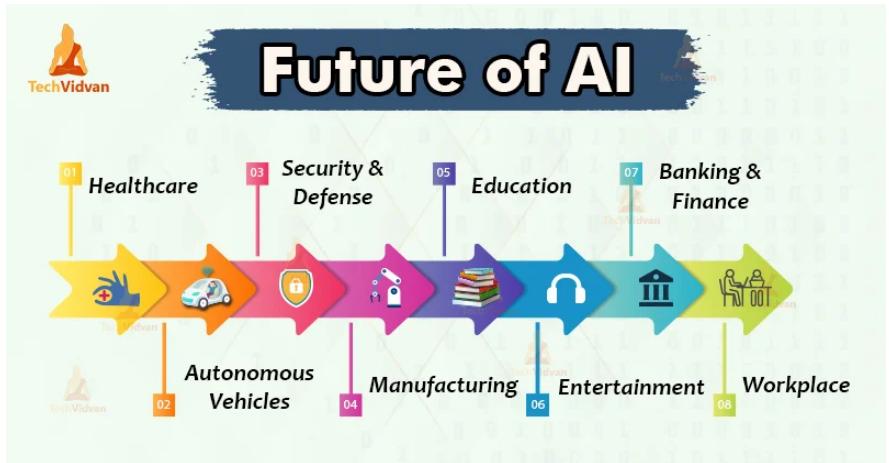


# Le futur de l'IA... il y a 3 ans



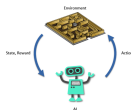
Src: <https://assets-global.website-files.com>

# Le futur de l'IA... il y a 3 ans



Src: <https://techvidvan.com>

# L'exposé de ChatGPT il y a 3 ans



what is the state of the art of A.I. research?



As of my knowledge cutoff in 2021, the state of the art of AI research was characterized by several significant developments across different fields, including:



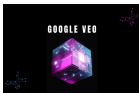
1. **Natural Language Processing (NLP):** The development of large-scale language models, such as GPT-3 and BERT, which achieved state-of-the-art performance in a wide range of language tasks, including language generation, question answering, and sentiment analysis.
2. **Computer Vision:** Advancements in computer vision, including the development of deep neural networks, such as convolutional neural networks (CNNs), that can perform complex visual recognition tasks with high accuracy. Applications of computer vision include image and video recognition, object detection, and facial recognition.
3. **Reinforcement Learning:** The development of reinforcement learning algorithms that enable machines to learn from experience by maximizing a reward signal, leading to breakthroughs in robotics and game playing.
4. **Generative Models:** The development of generative models, such as generative adversarial networks (GANs) and variational autoencoders (VAEs), which can generate new images, videos, and even text by learning the underlying distribution of the data.
5. **Robotics:** Advances in robotics, including the development of humanoid robots and the use of deep reinforcement learning to teach robots new skills.

It is worth noting that research in AI is a rapidly evolving field, and there may have been new breakthroughs and developments since my knowledge cutoff.

# L'exposé de ChatGPT aujourd'hui



## ChatGPT



what is the state of the art of AI?

As of mid-2025, the **state of the art in AI** is defined by several key trends and breakthroughs across models, applications, and capabilities. Here's a concise overview:

### 1. Frontier Models

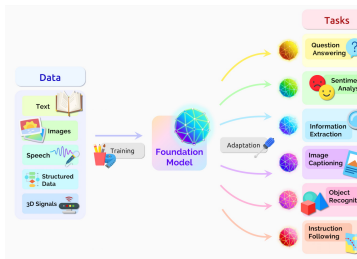
- **OpenAI GPT-4o**: Multimodal model handling text, image, and audio inputs/output in real time with strong reasoning and conversational ability.
- **Anthropic Claude 3, Google Gemini 1.5, Mistral, xAI's Grok-1.5**: Competing LLMs with differing strengths (e.g. Claude's interpretability, Gemini's tool integration).
- **Multimodal AI**: Models like **GPT-4o**, **Gemini 1.5 Pro**, and **CLIP-based vision-language models** now process images, video, speech, and code natively.

### 2. Model Architecture Trends

- **Mixture of Experts (MoE)**: Used in models like Mistral's and Google's to scale efficiently.
- **Sparse Attention + Efficient Transformers**: Long-context handling (e.g., 1M+ tokens in Gemini, Claude).
- **Retrieval-Augmented Generation (RAG)**: Widely used to enhance factual accuracy and memory.
- **Agentic Systems**: Growing focus on models acting autonomously in environments (e.g., AutoGPT, Devin by Cognition).

### 3. Applications

- **Coding assistants**: GitHub Copilot, Devin (first "autonomous software engineer").
- **Medical AI**: Large-scale deployment in radiology, diagnostics (e.g., Google Med-PaLM).
- **Enterprise copilots**: AI embedded into tools like Microsoft 365, Google Workspace, and CRM platforms.
- **Robotics + Embodied AI**: Integration of LLMs into physical agents (e.g., Boston Dynamics + OpenAI partnership).



# Une explosion des usages

**A You**  
dessine une ville intelligente dans le style de dali

**ChatGPT**

I've created an image that imagines a futuristic smart city with a surreal twist, inspired by the style of Salvador Dalí. This city blends advanced technology with eco-friendly architecture, featuring skyscrapers with green walls, solar panels, lush parks, and innovative transport systems, all depicted in a dreamlike, surreal manner.

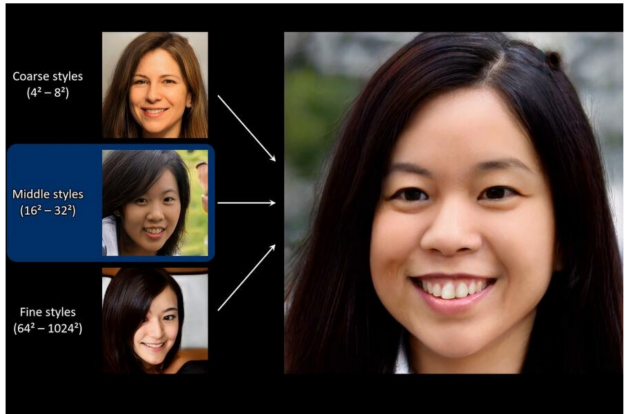
**A You**  
dis-moi ça en français stp

**ChatGPT**  
J'ai créé une image qui imagine une ville intelligente futuriste avec une touche surréaliste, inspirée par le style de Salvador Dalí. Cette ville fusionne la technologie avancée avec l'architecture éco-responsable, mettant en scène des gratte-ciels aux murs végétaux, des panneaux solaires, des parcs luxuriants, et des systèmes de transport innovants, le tout représenté de manière onirique et surréaliste.



... qui peut faire peur

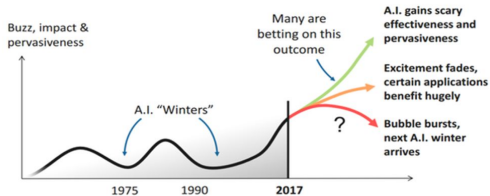
## UNE EFFICACITÉ SUPRENANTE



# Une histoire mouvementée

## AI summers and winters

AI is enjoying significant hype and investment



Src: Aleksandra Przegalinska, Kozminski University/ MIT Center for Collective Intelligence

# Plan

Le futur de l'IA a une histoire

Concepts fondamentaux

Des succès spectaculaires

Faire confiance aux IA ?



# IntelligenceS artificielleS ?

## INTELLIGENCES ARTIFICIELLES ?



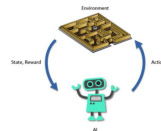
logique

rationnellement

mathématiques

informatique

économie  
théorie de la décision  
théorie des jeux



penser

apprentissage

cybernétique  
robotique

agir



neurosciences  
sciences cognitives  
psychologie

linguistique

créativité

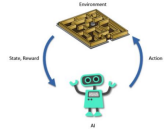
NLP

comme des humains



# IntelligenceS artificielleS ?

## INTELLIGENCES ARTIFICIELLES ?



# Intelligences Artificielles

## → **Raisonnement sur les données**

Ontologies

Web sémantique



+



→ Léonard de Vinci

## → **Stratégies**

Jeux

Agents autonomes



## → **Apprentissage automatique**

Classification supervisée

Régression



→ chat

# Spécification d'une tâche

**Ordinateur** = machine capable d'un petit nombre d'opérations élémentaires, qu'on peut combiner à l'envi

**Spécification d'un problème** = description de la sortie souhaitée pour une entrée donnée



[3,2,5,1,4] → [1,2,3,4,5]

le petit chat → the little cat



→

5



→



Exemples:

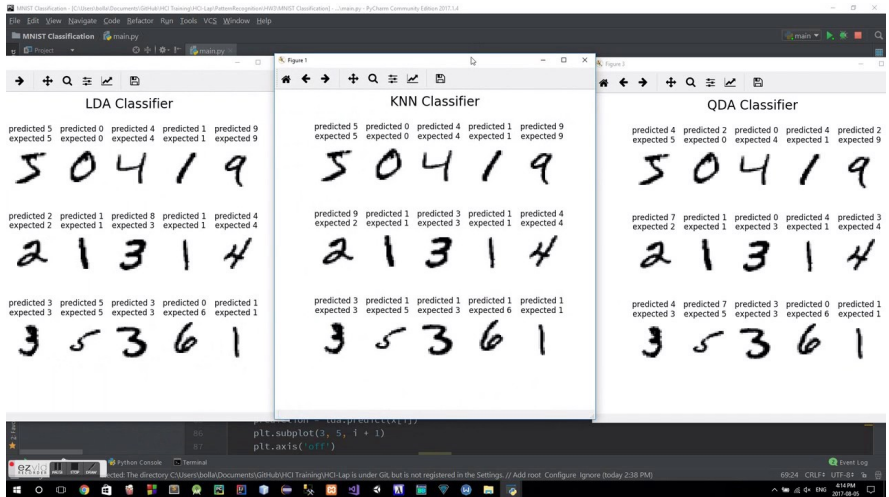
# Deux approches de résolution

**Approche classique : réduction** = décrire le *programme* = la suite d'opérations élémentaires permettant de construire la sortie à partir de l'entrée  
= programmation (codage)

**Intelligence artificielle** = programmer l'ordinateur pour qu'il *construise lui-même* le programme qui résout la tâche visée  
= méta-programmation

- pour chaque problème, les deux approches sont possibles
- mais elles sont plus ou moins efficaces

# Exemple tarte à la crème: MNIST



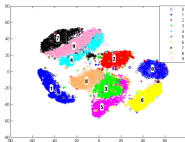
# Science des données

= résolution de problèmes dont les entrées sont les données (singulier politique)

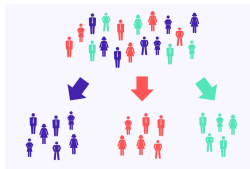
→ stockage, gestion, accès, etc.



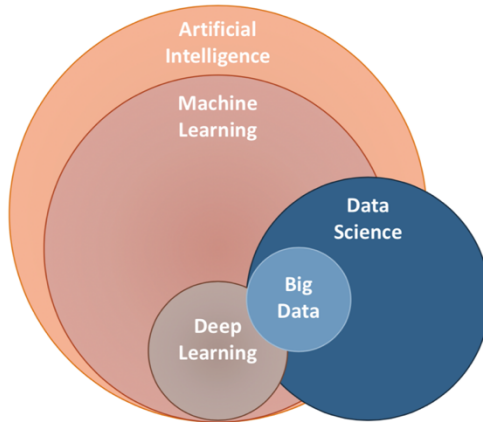
→ visualisation, représentation



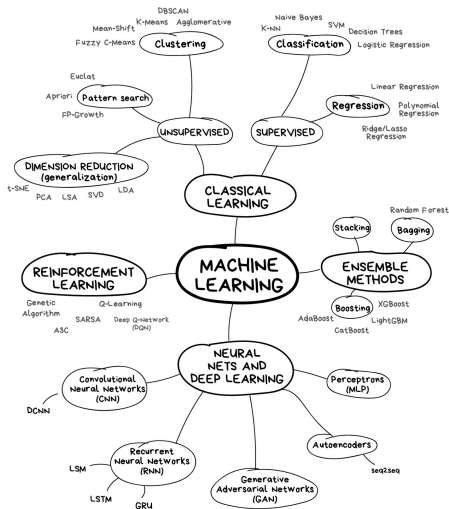
→ simplification, typologie



# Science des Données, Intelligence Artificielle, Apprentissage Automatique

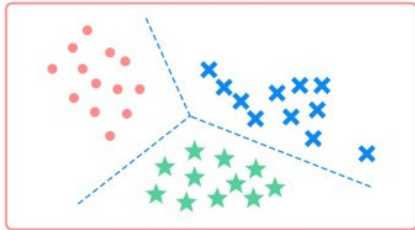


# Un panorama riche et varié



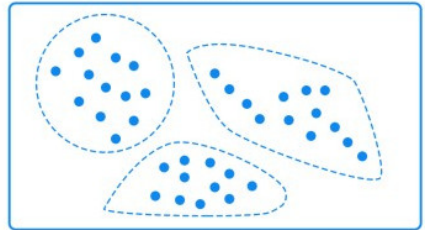
# Apprentissage supervisé & non-supervisé

Classification



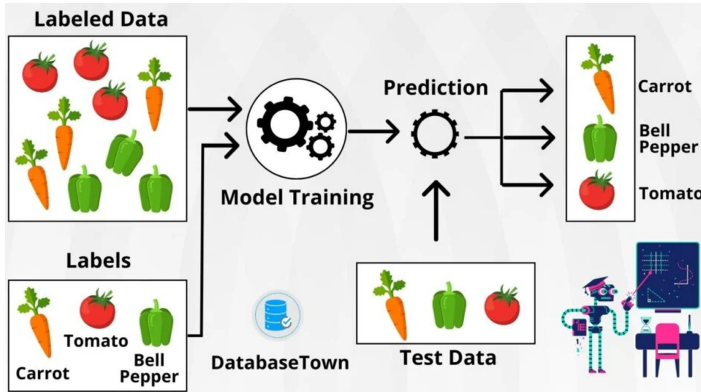
Supervised learning

Clustering



Unsupervised learning

# Apprentissage supervisé



# Plan

Le futur de l'IA a une histoire

Concepts fondamentaux

Des succès spectaculaires

Faire confiance aux IA ?

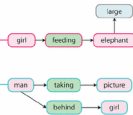


# Des succès spectaculaires

- Reconnaissance d'image
- Traitement du langage naturel  
... et leur combinaison



Leaves on the ground  
Huts on a hillside  
A bag  
A bush next to a river.  
a woman wearing a brown shirt  
(girl feeding large elephant)  
Woman wearing a purple dress  
Tree near the water  
a man wearing a hat  
A handle of bananas.  
A man taking a picture behind girl  
Glasses on the hair.  
blue flip flop sandals  
small houses on the hillside  
the nearby river  
Flashlight with carrier on 8% back



<https://link.springer.com/article/10.1007>



- Résolution de jeux (stratégie)
- Véhicules autonomes



- Systèmes de recommandation (presse, publicités, films, etc.)

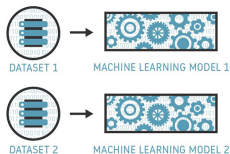


# Apprentissage par transfert

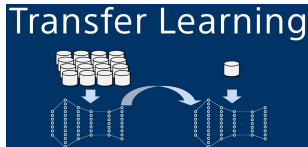
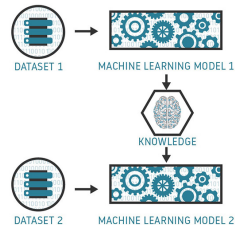
Utiliser des compétences acquises sur une tâche pour en résoudre une autre

- analogue
- dérivée
- ou plus précise

TRADITIONAL MACHINE LEARNING



TRANSFER LEARNING



# Modèles de langages

## 3. THE SERIES OF APPROXIMATIONS TO ENGLISH

To give a visual idea of how this series of processes approaches a language, typical sequences in the approximations to English have been constructed and are given below. In all cases we have assumed a 27-symbol "alphabet," the 26 letters and a space.

1. Zero-order approximation (symbols independent and equiprobable).

XFOML RXKHRJFFJUJ ZLPWCFWKCYJ FFJEYVKCQSGHYD QPAAMKBZAACIBZLHJQD.

2. First-order approximation (symbols independent but with frequencies of English text).

OCRO HLI RGWR NMIELWIS EU LL NBNESBYA TH EEI ALHENHTTPA OOBTTVA NAH BRL.

3. Second-order approximation (digram structure as in English).

ON IE ANTSOUTINYS ARE T INCTORE ST BE S DEAMY ACHIN D ILONASIVE TU-COOWE AT TEASONARE FUSO TIZIN ANDY TOBE SEACE CTISBE.

4. Third-order approximation (trigram structure as in English).

IN NO IST LAT WHEY CRATICT FROURE BIRS GROCID PONDENOME OF DEMONSTURES OF THE REPTAGIN IS REGOACTIONA OF CRE.

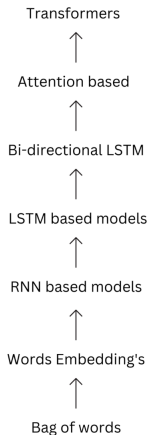
5. First-order word approximation. Rather than continue with tetragram,  $\dots$ ,  $n$ -gram structure it is easier and better to jump at this point to word units. Here words are chosen independently but with their appropriate frequencies.

REPRESENTING AND SPEEDILY IS AN GOOD APT OR COME CAN DIFFERENT NATURAL HERE HE THE A IN CAME THE TO OF TO EXPERT GRAY COME TO FURNISHES THE LINE MESSAGE HAD BE THESE.

6. Second-order word approximation. The word transition probabilities are correct but no further structure is included.

THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH WRITER THAT THE CHARACTER OF THIS POINT IS THEREFORE ANOTHER METHOD FOR THE LETTERS THAT THE TIME OF WHO EVER TOLD THE PROBLEM FOR AN UNEXPECTED.

# Perroquets stochastiques



Src: Wikipedia



# DE L'INTELLIGENCE ÉMERGE ?

## KANGOUROU DES MATHÉMATIQUES



L'association *Kangourou Sans Frontières* organise le jeu-concours *Kangourou* pour plus de six millions de participants dans le monde.

**Jeu-concours 2023 – Durée : 50 minutes**

### Sujet E

- Il y a une seule bonne réponse par question. Les questions 1 à 8 valent 3 points chacune, les questions 9 à 16 valent 4 points chacune, les questions 17 à 24 valent 5 points chacune. Une réponse fausse enlève un quart des points prévus, tandis que si tu ne réponds pas, ton total ne change pas.
- Pour gagner les prix nationaux, l'épreuve doit être individuelle et sans calculatrice. Les classements sont séparés pour chaque niveau (CE2, CM1, CM2, ...).

- 26** La date 16/03/2023 a la somme de ses huit chiffres égale à 17 ( $1+6+0+3+2+0+2+3$ ). Combien de dates en 2023 ont la somme de leurs huit chiffres égale à 26 ?

# DE L'INTELLIGENCE ÉMERGE ? GEMINI



- A La date 16/03/2023 a la somme de ses huit chiffres égale à 17 ( $1+6+0+3+2+0+2+3$ ). Combien de dates en 2023 ont la somme de leurs huit chiffres égale à 26 ?

Show drafts ▼

## ◆ Déterminons le nombre de dates possibles en 2023.

### 1. Année:

- L'année 2023 est fixe, donc elle ne peut pas varier.

### 2. Mois:

- Il y a 12 mois possibles (de janvier à décembre).

### 3. Jour:

- Le jour peut varier de 1 à 31.

### 4. Calcul du nombre total de dates:

- Il y a donc  $12 \times 31 = 372$  dates possibles en 2023.

## Déterminons le nombre de dates avec une somme de 26.

### 1. Analyse des chiffres:

- Pour obtenir une somme de 26, il faut combiner des chiffres de manière à atteindre 26.
- On peut décomposer 26 en  $20 + 6$  ou  $18 + 8$  ou  $16 + 10$  ou  $14 + 12$ .

### 2. Cas 1: $20 + 6$

- Le chiffre 20 ne peut être utilisé qu'une seule fois pour l'année.

(à compléter)

### 3. Cas 2: $18 + 8$

- Le chiffre 18 ne peut être utilisé qu'une seule fois pour l'année.
- Il reste 8 à répartir entre les 7 chiffres restants.
- Le nombre de combinaisons possibles est élevé et difficile à calculer précisément.

### 4. Cas 3: $16 + 10$

- Le chiffre 16 ne peut être utilisé qu'une seule fois pour l'année.
- Il reste 10 à répartir entre les 7 chiffres restants.
- Le nombre de combinaisons possibles est élevé et difficile à calculer précisément.

### 5. Cas 4: $14 + 12$

- Le chiffre 14 ne peut être utilisé qu'une seule fois pour l'année.
- Il reste 12 à répartir entre les 7 chiffres restants.
- Le nombre de combinaisons possibles est élevé et difficile à calculer précisément.

### 6. Conclusion:

- Il est impossible de déterminer le nombre exact de dates avec une somme de 26 sans une analyse approfondie de chaque cas et des combinaisons possibles.
- Le nombre de dates sera **significativement inférieur** à 372 (nombre total de dates) mais **supérieur** à 0.

### 7. Estimation approximative:

- On peut approximer le nombre de dates en supposant que chaque cas ( $20+6$ ,  $18+8$ ,  $16+10$ ,  $14+12$ ) a un nombre équivalent de combinaisons possibles.
- En divisant le nombre total de dates (372) par le nombre de cas (4), on obtient une estimation grossière de 93 dates.

### 8. Remarque:

- Cette estimation est approximative et ne prend pas en compte les restrictions supplémentaires sur les dates (mois de 30 ou 31 jours, etc.).
- Le nombre réel de dates peut être légèrement plus ou moins élevé que l'estimation fournie.

# DE L'INTELLIGENCE ÉMERGE ? MISTRAL-AI



**AG** La date 16/03/2023 a la somme de ses huit chiffres égale à 17 ( $17 = 1 + 6 + 0 + 3 + 2 + 0 + 2 + 3$ ). Combien de dates en 2023 ont la somme de leurs huit chiffres égale à 26 ?

 Pour résoudre ce problème, nous allons décomposer la date en trois parties : le jour (JJ), le mois (MM) et l'année (AAAA). La somme des chiffres de la date (S) est égale à la somme des chiffres du jour, du mois et de l'année.

Dans ce cas, nous cherchons le nombre de dates en 2023 dont la somme des chiffres est égale à 26.

1. Analyse de l'année : L'année 2023 a une somme de chiffres égale à 7 ( $2+0+2+3$ ). Cette valeur est fixe pour toutes les dates de l'année.
2. Analyse du mois : La somme des chiffres du mois peut varier de 1 (mois = 01) à 19 (mois = 12). Comme nous avons déjà la somme 7 de l'année, la somme des chiffres du jour et du mois doit être égale à 19 ( $26 - 7$ ) pour obtenir une somme totale de 26.
3. Analyse du jour : La somme des chiffres du jour peut varier de 1 (jour = 01) à 18 (jour = 29, mais en 2023 n'est pas une année bissextile, donc le maximum est 18 pour le jour 19).

Maintenant, nous allons chercher les combinaisons mois-jour qui donnent une somme de chiffres égale à 19 :

- Mois = 01 (somme = 1) : Le jour doit avoir une somme de chiffres égale à 18. Le seul jour possible est le 19 (1+9). Donc, nous avons la date 19/01/2023.
- Mois = 02 (somme = 2) : Le jour doit avoir une somme de chiffres égale à 17. Aucun jour ne peut avoir une somme de chiffres égale à 17.
- Mois = 03 (somme = 3) : Le jour doit avoir une somme de chiffres égale à 16. Aucun jour ne peut avoir une somme de chiffres égale à 16.
- Mois = 04 (somme = 4) : Le jour doit avoir une somme de chiffres égale à 15. Aucun jour ne peut avoir une somme de chiffres égale à 15.
- Mois = 05 (somme = 5) : Le jour doit avoir une somme de chiffres égale à 14. Le seul jour possible est le 23 (2+3). Donc, nous avons la date 23/05/2023.
- Mois = 06 (somme = 6) : Le jour doit avoir une somme de chiffres égale à 13. Aucun jour ne peut avoir une somme de chiffres égale à 13.
- Mois = 07 (somme = 7) : Le jour doit avoir une somme de chiffres égale à 12. Aucun jour ne peut avoir une somme de chiffres égale à 12.
- Mois = 08 (somme = 8) : Le jour doit avoir une somme de chiffres égale à 11. Aucun jour ne peut avoir une somme de chiffres égale à 11.
- Mois = 09 (somme = 9) : Le jour doit avoir une somme de chiffres égale à 10. Le seul jour possible est le 19 (1+9). Donc, nous avons la date 19/09/2023.
- Mois = 10 (somme = 1) : Le jour doit avoir une somme de chiffres égale à 18. Le seul jour possible est le 19 (1+9). Donc, nous avons la date 19/10/2023.
- Mois = 11 (somme = 2) : Le jour doit avoir une somme de chiffres égale à 17. Aucun jour ne peut avoir une somme de chiffres égale à 17.
- Mois = 12 (somme = 3) : Le jour doit avoir une somme de chiffres égale à 16. Aucun jour ne peut avoir une somme de chiffres égale à 16.

Donc, il y a quatre dates en 2023 dont la somme des chiffres est égale à 26 : 19/01/2023, 23/05/2023, 19/09/2023 et 19/10/2023.

# DE L'INTELLIGENCE ÉMERGE ? CHATGPT



**You**

La date 16/03/2023 a la somme de ses huit chiffres égale à 17 ( $1+6+0+3+2+0+2+3$ ).  
Combien de dates en 2023 ont la somme de leurs huit chiffres égale à 26 ?



**ChatGPT**

En 2023, il y a 3 dates dont la somme de leurs huit chiffres est égale à 26. [2]

# DE L'INTELLIGENCE ÉMERGE ? CHATGPT



**You**

La date 16/03/2023 a la somme de ses huit chiffres égale à 17 ( $1+6+0+3+2+0+2+3$ ).  
Combien de dates en 2023 ont la somme de leurs huit chiffres égale à 26 ?



**ChatGPT**

En 2023, il y a 3 dates dont la somme de leurs huit chiffres est égale à 26. [2]

```
# To find the number of dates in 2023 with the sum of their digits equal to 26,
# we can iterate through all the dates of the year 2023 and check their sums.

def sum_of_digits(day, month, year):
    """Calculate the sum of digits of a given day, month, and year."""
    return sum(map(int, f"{day:02d}{month:02d}{year}"))

def dates_with_sum_26(year):
    """Find how many dates in the given year have a digit sum of 26."""
    count = 0
    for month in range(1, 13): # For each month
        if month == 2:
            days_in_month = 29 if year % 4 == 0 and (year % 100 != 0 or year % 400 == 0) else 28
        elif month in [4, 6, 9, 11]:
            days_in_month = 30
        else:
            days_in_month = 31

        for day in range(1, days_in_month + 1): # For each day in the month
            if sum_of_digits(day, month, year) == 26:
                count += 1
    return count

dates_with_sum_26_2023 = dates_with_sum_26(2023)
print(dates_with_sum_26_2023)
```

# Plan

Le futur de l'IA a une histoire

Concepts fondamentaux

Des succès spectaculaires

Faire confiance aux IA ?



# Transparence d'usage

un challenge sociétal,  
politique et législatif



# Explicabilité

un challenge technique



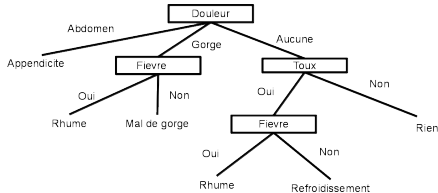
# Explicabilité versus interprétabilité

Deux notions à distinguer (mais vocabulaire trompeur: on suit ici

<https://www2.eecs.berkeley.edu/Pubs/TechRpts/2017/EECS-2017-159.pdf>).

Une règle de décision est dite

interprétable si on comprend comment elle associe une réponse à des observations; exemple typique : un arbre de décision;



<http://www.up2.fr/>

explicable si on comprend sur quels éléments est basée la décision, éventuellement de façon contrefactuelle; exemple : "si telle variable avait pris telle autre valeur, alors la décision aurait été différente".

# Explicabilité versus interprétabilité

Deux notions à distinguer (mais vocabulaire trompeur: on suit ici

<https://www2.eecs.berkeley.edu/Pubs/TechRpts/2017/EECS-2017-159.pdf> ).

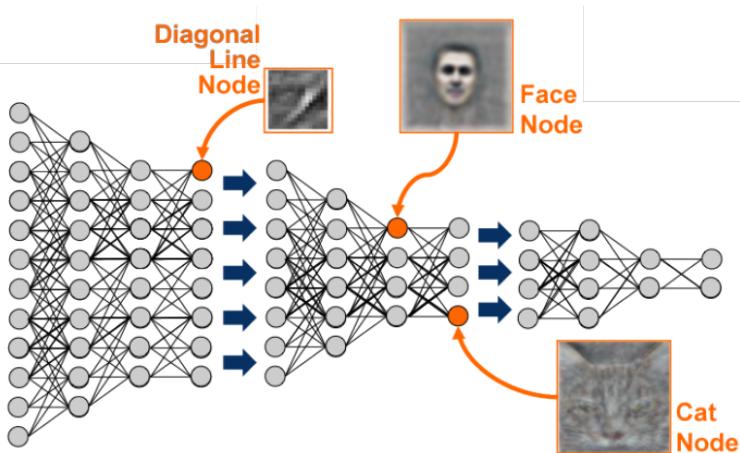
Une règle de décision est dite

interprétable si on comprend comment elle associe une réponse à des observations; exemple typique : un arbre de décision;

explicable si on comprend sur quels éléments est basée la décision, éventuellement de façon contrefactuelle; exemple : "si telle variable avait pris telle autre valeur, alors la décision aurait été différente".

La notion d'explicabilité se rattache à la statistique causale et à l'analyse de sensibilité.

# Interpréter un réseau de neurones



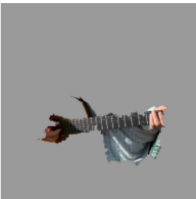
<http://aiehive.com>

Pari scientifique audacieux

# Explicabilité : la méthode LIME



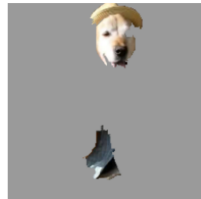
(a) Original Image



(b) Explaining *Electric guitar*



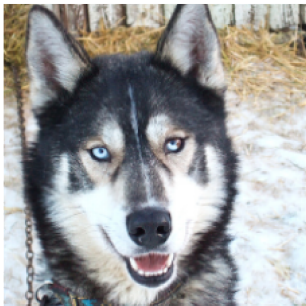
(c) Explaining *Acoustic guitar*



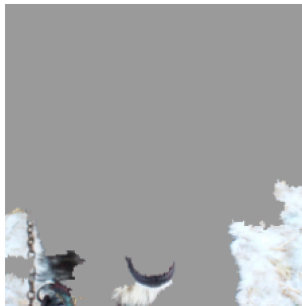
(d) Explaining *Labrador*

"Why Should I Trust You?" Explaining the Predictions of Any Classifier, by Marco Tulio Ribeiro, Sameer Singh and Carlos Guestrin

# Explicabilité : la méthode LIME



(a) Husky classified as wolf



(b) Explanation

**Figure 11: Raw data and explanation of a bad model's prediction in the "Husky vs Wolf" task.**

"Why Should I Trust You?" Explaining the Predictions of Any Classifier, by Marco Tulio Ribeiro, Sameer Singh and Carlos Guestrin

# Certification / robustesse

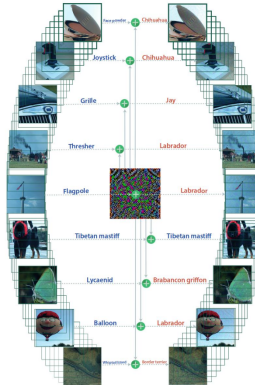
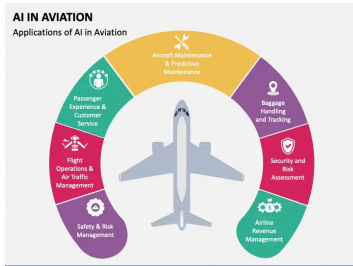


Figure 1: When added to a natural image, a universal perturbation image causes the image to be misclassified by the deep neural network with high probability. *Left images:* Original natural images. The labels are shown on top of each arrow. *Central image:* Universal perturbation. *Right images:* Perturbed images. The estimated labels of the perturbed images are shown on top of each arrow.

un challenge  
technique  
et sociétal



## Facial analysis datasets

|         |                           |
|---------|---------------------------|
| LFW     | 77.5% male<br>83.5% white |
| IJB-A   | 79.6% lighter-skinned     |
| Adience | 86.2% lighter-skinned     |

### Error Rate<sub>(1-PPV)</sub> By Female x Skin Type



Buciarneini & Gebru FAT\* 2018, Slides from Joy Buciarneini

Buolamwini & Gebru (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification

# Confidentialité

un challenge technique  
et sociétal

Review your NDA  
with our A.I.  
It's free.

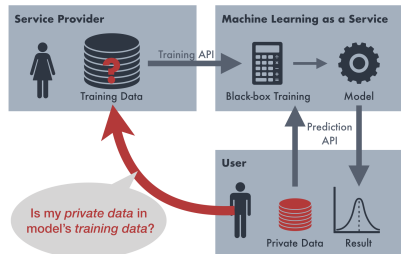
K L A R I T Y



# Fuite d'information

## Membership attack

Src: <https://www.arxiv-vanity.com/papers/1904.05506/>



## Model inversion attack

[Fredrikson et al. '2015]

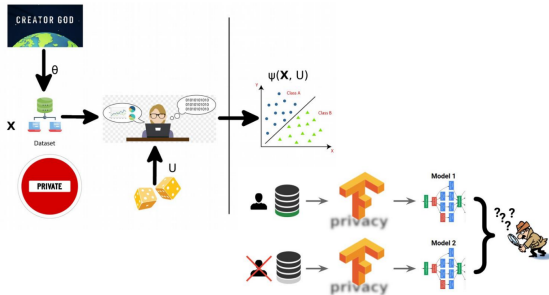


Figure 1: An image recovered using a new model inversion attack (left) and a training set image of the victim (right). The attacker is given only the person's name and access to a facial recognition system that returns a class confidence score.

Voir <https://arxiv.org/abs/1610.05820> for more information: *Membership Inference Attacks against Machine Learning Models* by Reza Shokri, Marco Stronati, Congzheng Song, Vitaly Shmatikov

# Confidentialité différentielle

## GARANTIR LA CONFIDENTIALITÉ TOUT EN GARDANT LA PUISSANCE DE L'ANALYSE



### Un modèle formel

pour le respect de la confidentialité : la *Differential Privacy*

### Un standard international

adopté ces dernières années par les GAFAM, le Census Bureau, LinkedIn, etc.

### Une efficacité basée sur la preuve mathématique

qu'aucune méthode existante ou à venir ne pourra remettre en cause