

Intelligence artificielle : quand la machine apprend à lire



Fête de la Science, Edition 2021

Denis Coquenot

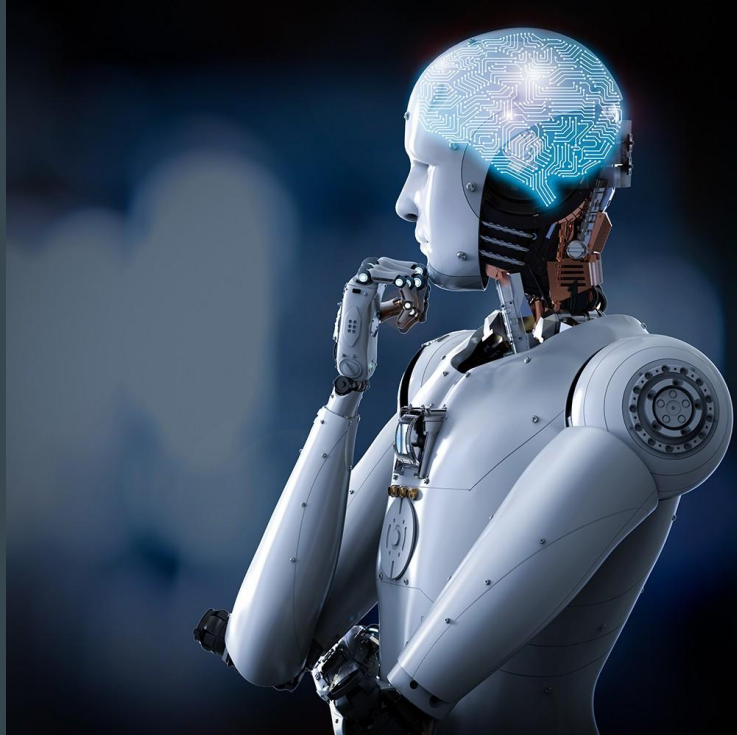


Intelligence artificielle

Que nous dit Wikipédia ?

“L'**intelligence artificielle (IA)** est l'ensemble des théories et des techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence humaine”

Intelligence artificielle - Dans la fiction



- Robot humanoïde
- L'humain, en mieux
- Décide d'éradiquer l'espèce humaine

Intelligence artificielle - Pas si intelligente

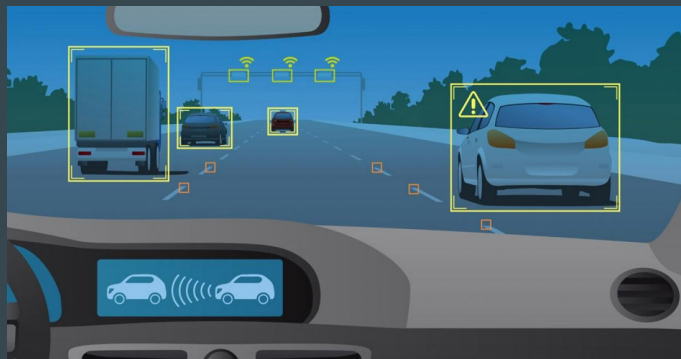


Programme informatique

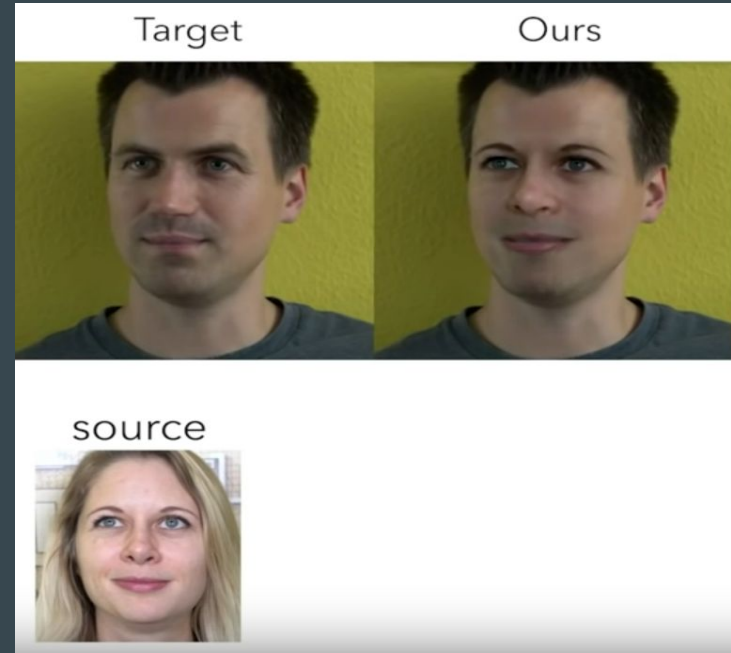
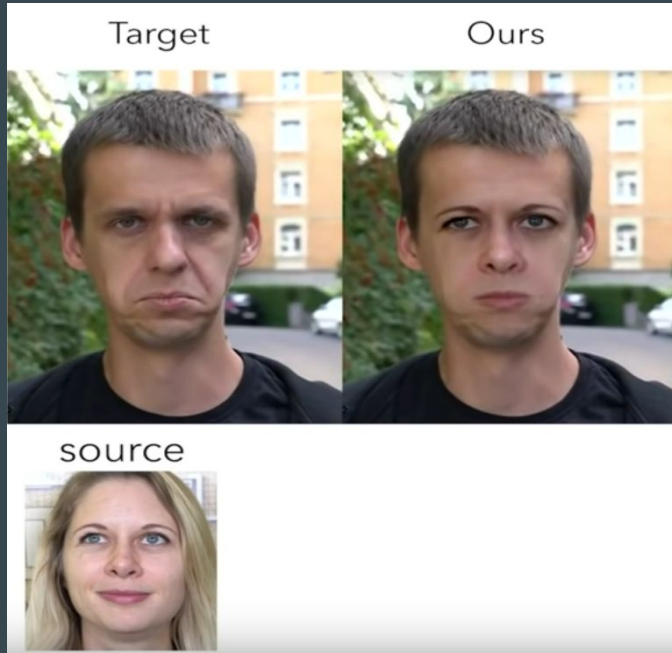
- Monotâche
- Apprentissage supervisé
- Faillible

Intelligence artificielle - Quelques applications

- Systèmes de recommandations (Netflix, Amazon, Spotify)
- Chatbot/robot conversationnel (SNCF)
- Traducteur automatique (texte, voix)
- Voiture autonome



Intelligence artificielle - Face Swapping



→ Attention aux dérives : Deepfake

Intelligence artificielle - Apprentissage supervisé

Chez l'humain

Des données :
ce qui doit être appris



Un système de prédiction :
le cerveau



“Vache”

“Poule”

Erreur

Un système de vérification :
professeur / parent



Processus d'amélioration de la prédiction :
réactions biochimiques

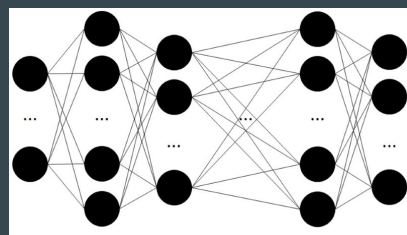
Intelligence artificielle - Apprentissage supervisé

Chez la machine

Des données :
ce qui doit être appris

Un système de prédiction :
réseau de neurones artificiels

Un système de vérification :
vérité connue dans le jeu de données



“Chien”

“Chat”

Erreur

Processus d'amélioration de la prédiction :
fonctions mathématiques

Intelligence artificielle - Apprentissage supervisé

Chez l'humain, comme chez la machine :

- Des données sont utilisées pour s'entraîner
- Des prédictions sont réalisées
- La vérité est connue et permet de reconnaître les erreurs pour améliorer les prédictions

→ L'apprentissage se fait à force d'erreurs et de répétitions

→ Coût humain d'annotation pour l'apprentissage de l'intelligence artificielle

Intelligence artificielle - Quand la machine apprend à lire

Classification chien/chat :

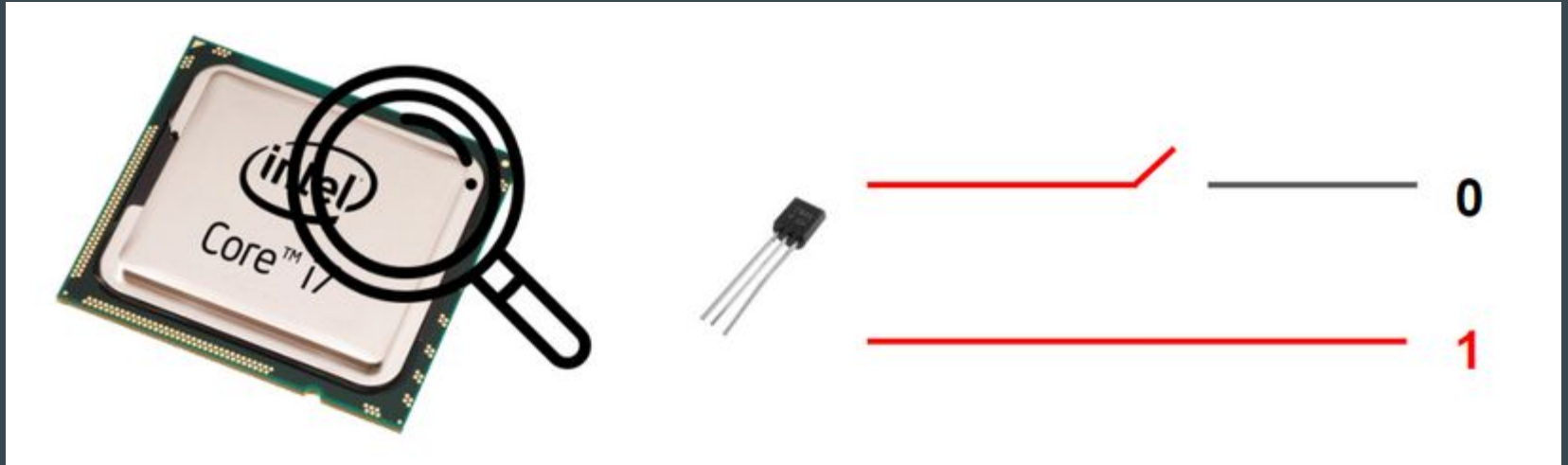
- Entrée : une image de chien ou de chat
- Sortie : “chat” ou “chien”

Reconnaissance d'écriture manuscrite :

- Entrée : un document quelconque numérisé
- Sortie : le texte présent dans ce document

→ Comment une machine peut-elle interpréter des images et du texte ?

Intelligence artificielle - La machine ne comprend que le langage binaire



→ Utilisation de transistors agissant comme des interrupteurs : ouvert/fermé → 0/1

Intelligence artificielle - Les nombres en binaire

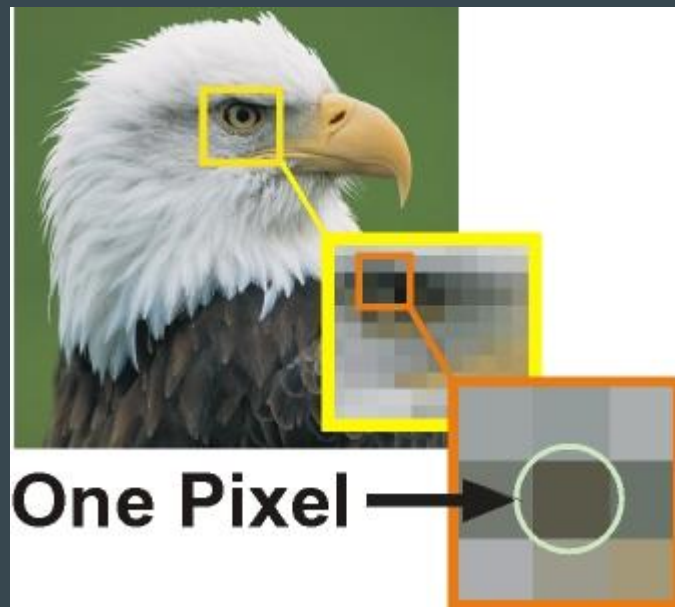
Humain (base 10)	Machine (base 2)
0	0
1	1
2	10
3	11
4	100
5	101
6	110
7	111
8	1000
9	1001
10	1010

Intelligence artificielle - La machine ne comprend que les nombres

Entrée : un document quelconque numérisé

- Un document numérisé = une image
- Une image = un ensemble de pixels
- Un pixel = une couleur
- Une couleur = 3 nombres (R, G, B)

→ Une image = un ensemble de nombres



Intelligence artificielle - La machine ne comprend que les nombres

Sortie : le texte présent dans ce document

- Un texte = une suite de caractères (lettres, chiffres, ponctuations, ...)
- Un caractère = un nombre

→ Un texte = une suite de nombres

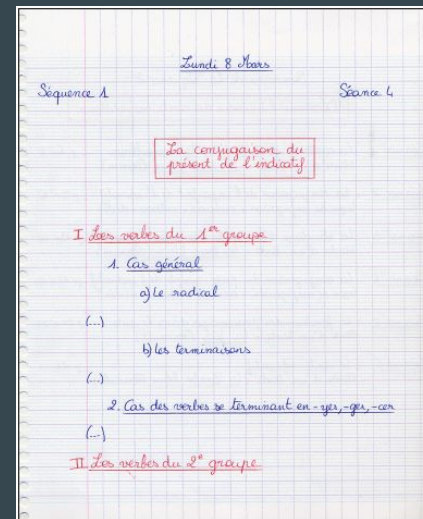
b o n j o u r
2 15 14 10 15 21 18

a → 1
b → 2
...
z → 26
A → 27
B → 28
...
Z → 52
0 → 53
1 → 54
...
9 → 62
! → 63
. → 64
...

Reconnaissance d'écriture manuscrite - Une tâche complexe

Une grande variété de style :

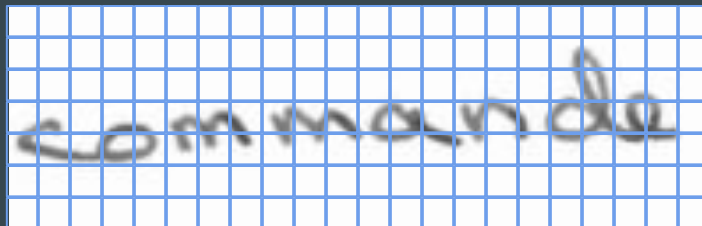
- Attaché / détaché
- Écritures plus ou moins penchés
- Majuscules / minuscules
- Espacements inter-caractères, inter-mots, inter-lignes
- Signature
- Calligraphie
- Fond (couleur, carreaux, ...)
- Langue



C'est que, au lieu des formes figées des écritures

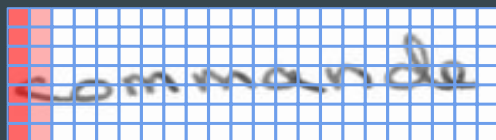
Commençons simple : un seul mot

- Problème :
 - L'image d'entrée est de taille variable
 - Le texte en sortie est de taille variable



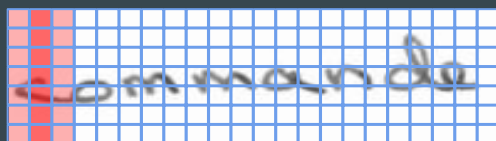
Commençons simple : un seul mot

- Solution : utilisation d'une fenêtre glissante, 1 prédiction par colonne de pixels



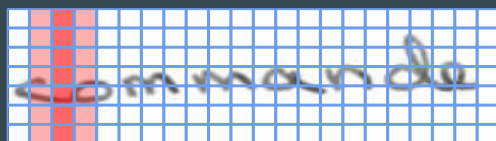
→ c

Résultat :
 $cc^*omm^*mma^*n^*ddee^*$



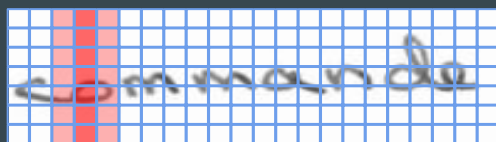
→ c

Étape 1 : on supprime les caractères identiques successifs
 $c^*om^*ma^*nde^*$



→ *

Étape 2 : on supprime les *
commande

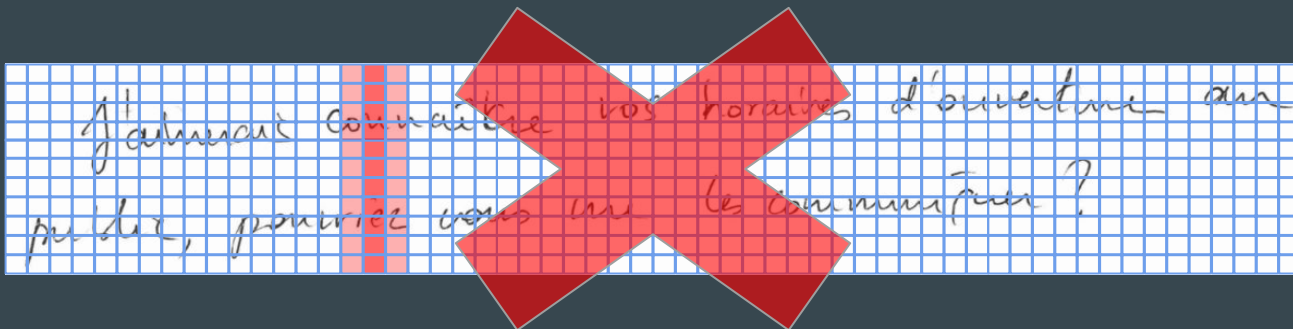


→ o

→ Même chose pour une ligne, l'espace est considéré comme un caractère à reconnaître

Reconnaissance d'un paragraphe

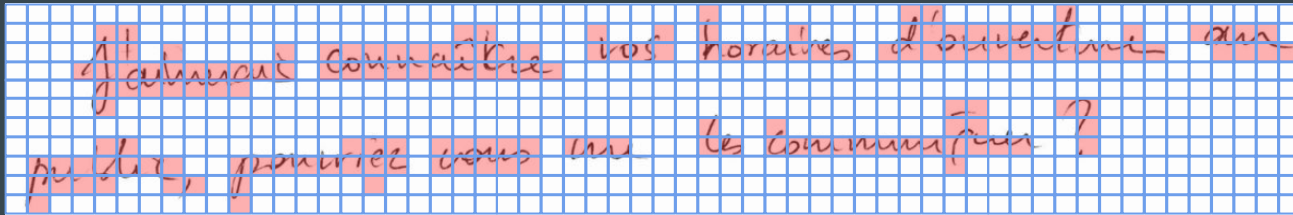
- Problème : impossible d'utiliser la même technique



- Une seule prédiction par colonne de pixels
- On ne peut reconnaître qu'une seule des deux lignes

Reconnaissance d'un paragraphe

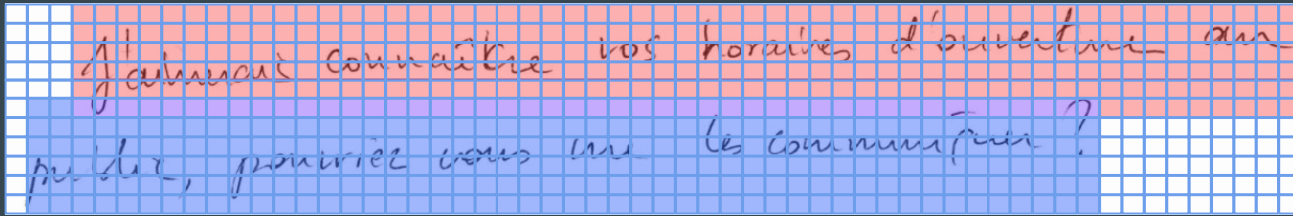
- Solution : ajouter une étape de segmentation



- Nouvelle étape = deuxième intelligence artificielle
- Nouvel apprentissage → nouvelles données nécessaires

Reconnaissance d'un paragraphe

- Annotation au niveau pixel très coûteux → boîtes englobantes



→ L' IA apprend à classer chaque pixel en deux catégories : "texte" / "fond"

Reconnaissance d'un paragraphe

- Étape 1 : une IA de segmentation en lignes

J'aimerais connaître vos horaires d'ouverture au public, pourriez vous me les communiquer ?



J'aimerais connaître vos horaires d'ouverture au public, pourriez vous me les communiquer ?

- Étape 2 : une IA de reconnaissance de texte

J'aimerais connaître vos horaires d'ouverture au public, pourriez vous me les communiquer ?



J'aimerais connaître vos horaires d'ouverture au public, pourriez vous me les communiquer ?

J'aimerais connaître vos horaires d'ouverture au public, pourriez vous me les communiquer ?



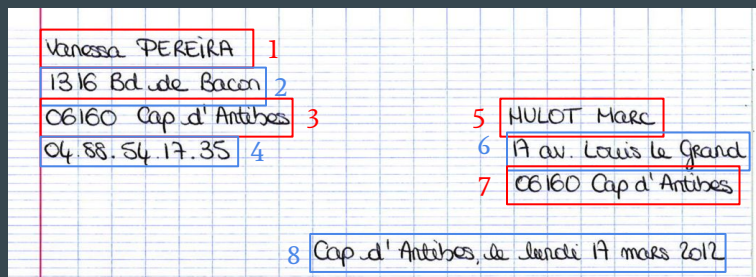
public, pourriez vous me les communiquer ?

- Étape 3 : assemblage par ordre vertical

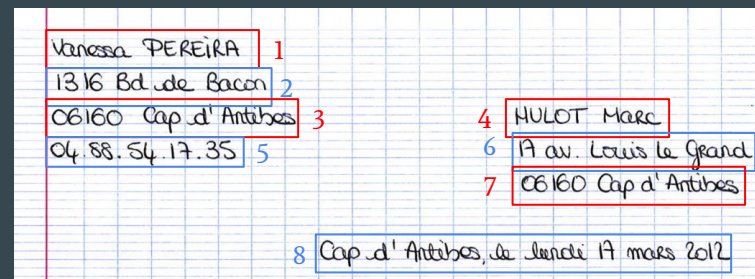
J'aimerais connaître vos horaires d'ouverture au public, pourriez vous me les communiquer ?

Reconnaissance d'un document complexe

Ordre de lecture humain



Ordre de lecture machine



- Étape de segmentation : OK
- Étape de reconnaissance : OK
- Étape d'assemblage : problème

→ Pas d'IA sur l'étape d'assemblage : ordre de lecture fixe, de haut en bas et de gauche à droite

Reconnaissance d'un document complexe

→ Solution : on apprend à l'IA à lire comme un humain, de caractère en caractère

VIDEO

Quelques chiffres

Jeu de données : 1 000 documents manuellement annotés

Temps d'apprentissage : quelques heures / jours

Temps de prédiction : quelques millisecondes

Performances : ~5% d'erreurs

Questions ?