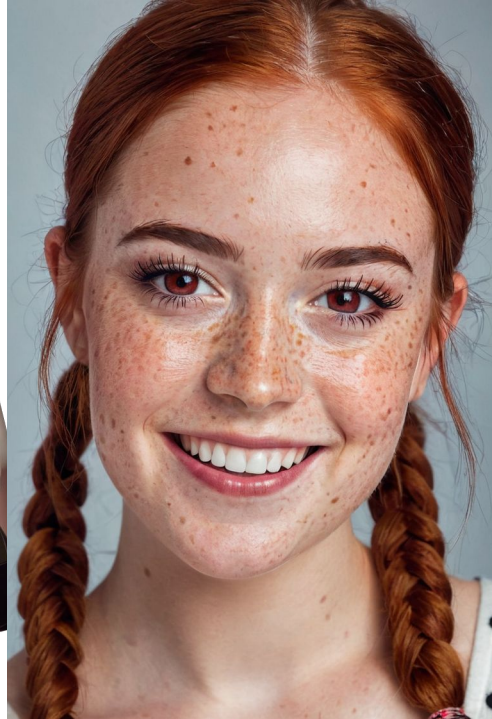


Intelligence artificielle



IA : pren



Définition

- Intelligence humaine
 - L'intelligence est la capacité de résoudre par la pensée des problèmes nouveaux. (Edouard Claparède)
 - L'intelligence reste un concept encore mal défini sur le plan scientifique.



Définition

- Intelligence artificielle
 - Pas de consensus
 - 2 types.... de types
 - Basé sur la capacité
 - Étroite \Rightarrow Monotâche (Alpha-Go)
 - Large (ou AGI) \Rightarrow Multitâche
 - Basé sur la « sentience »
 - Forte : Sentiente \Rightarrow Elle a conscience d'elle-même
 - Faible : Ou pas...



Mécanique

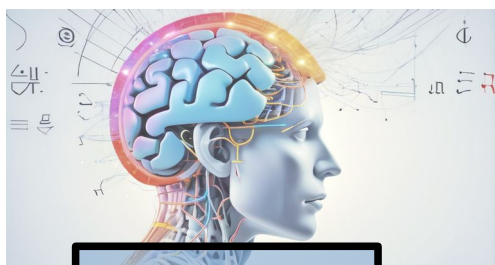
- Programmation
 - Prolog
 - Lisp
 - Eliza (1965, conversationnel, puis psy)
- Machine Learning
 - Jeu de données pour apprendre
 - Mécaniques
 - Supervisé (Étiquette)
 - Non supervisé (classification automatique)
 - Par renforcement (sanction / récompense)



Grands types ML

- Algorithmes génétiques
 - Héritage, variation, sélection, « saut » génétique
- Probabilistes
 - Bayésien : probabilité d'un événement global connaissant un événement local
 - Si « Viagra » apparaît dans un mail \Rightarrow 99,95 % spam
- Réseaux de neurones
 - Imitation des neurones biologiques
 - On se cantonne au « deep learning » avec des réseaux de neurones « profonds »
 - 3 archétypes
 - Classique (ANN \Rightarrow Artificial Neural Network) \Rightarrow décision à partir de données disponibles.
 - A récurrence (RNN \Rightarrow Recurrent Neural Network) \Rightarrow séquencement (dans le temps, dans l'espace)
 - A convolution (CNN \Rightarrow Convolution Neural Network) \Rightarrow (pré)Traitement d'images pour repérer des structures

Neurone



Entrée x1

Poids 1 (w1)
5,01245

Poids 2 (w2)
0,9856

Entrée x2

Neurone

$$S = x1 * w1 + x2 * w2$$

$$\text{Sortie } y = f(S * \text{Biais})$$

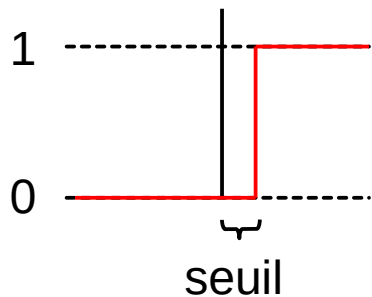
Paramètres

- poids ($w1, w2, \dots, wn$)
- biais
- seuil

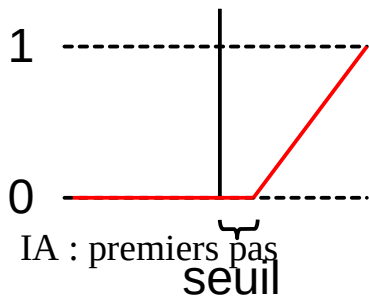
Y

Sortie

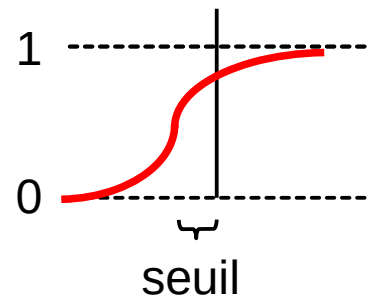
Seuil



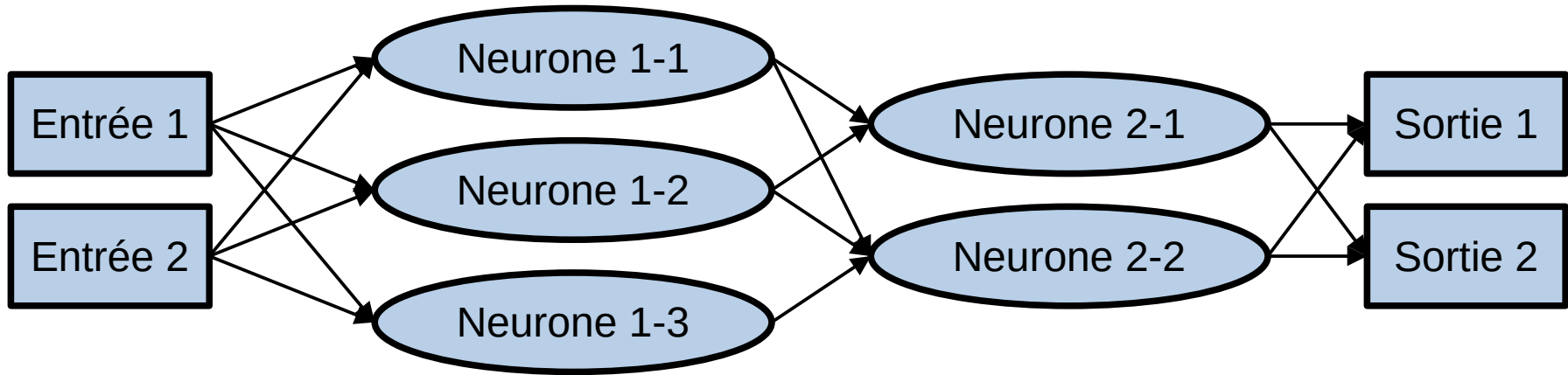
ReLU



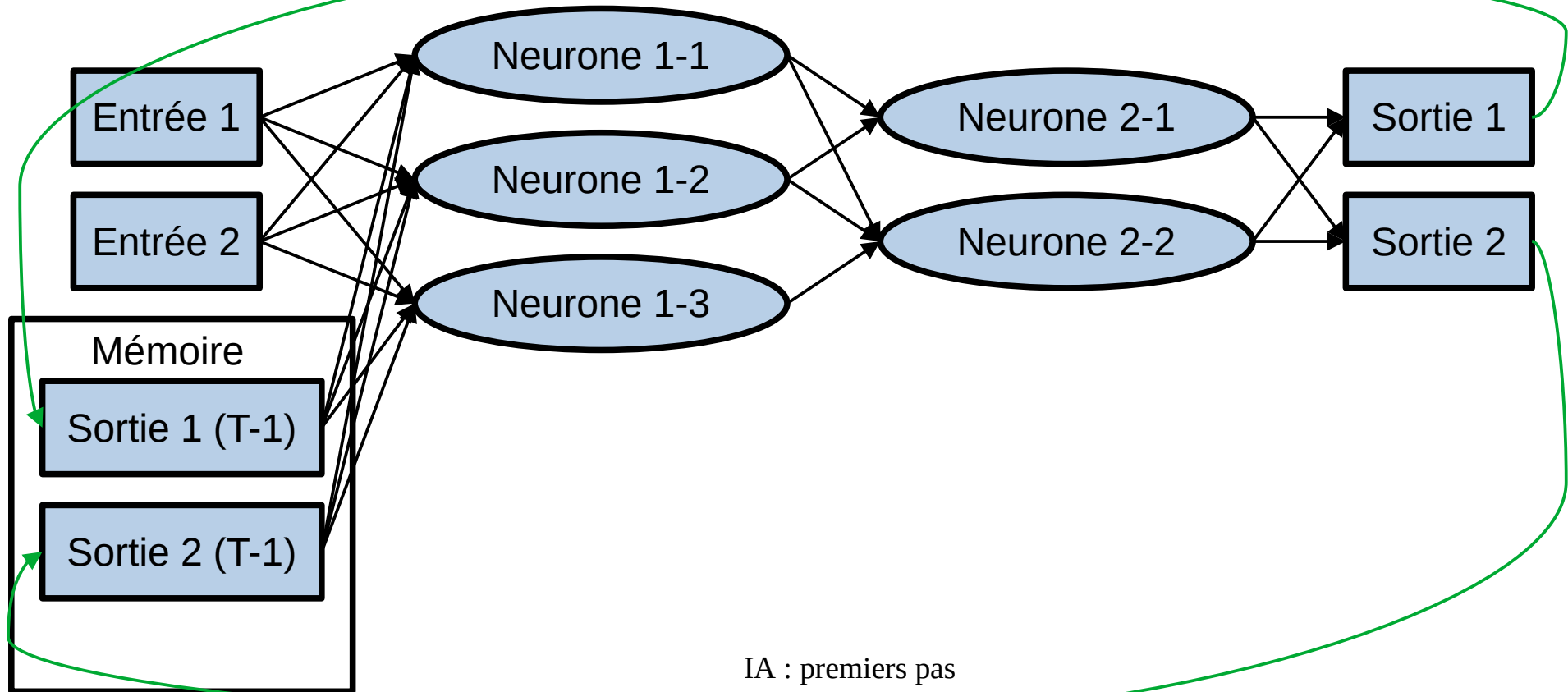
Sigmoïde



Réseau de neurones « profond »



Réseau de neurones récurrent



IA : premiers pas



Réseau de neurones

- Plus de couches « cachées » (entre 1 et 8 souvent)
 - plus de précisions
 - plus de calcul (exponentiel)
- Plus de neurones par couche (pas de limite)
 - plus de précisions
 - plus de calcul (exponentiel)

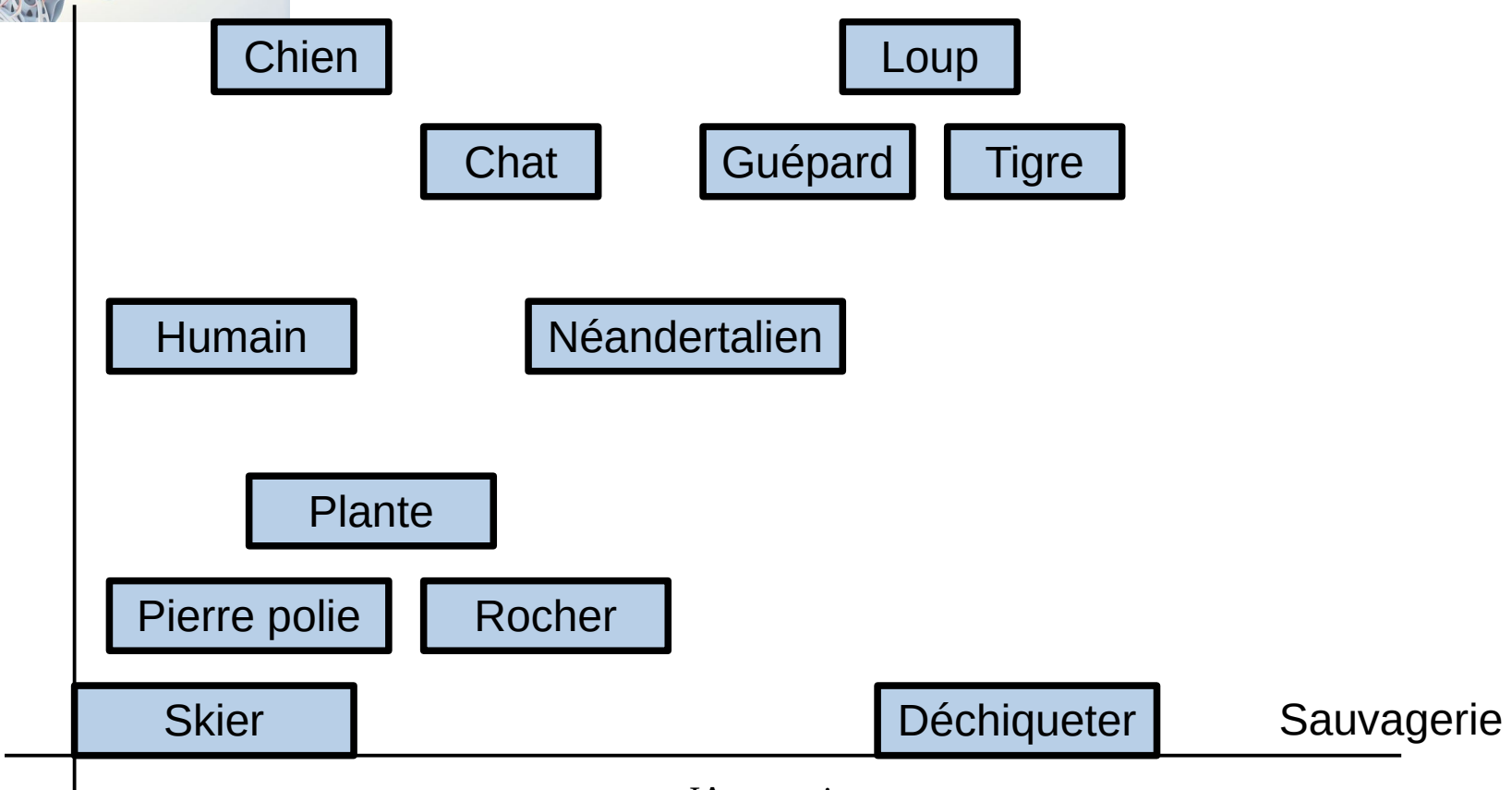


Quelles entrées pour un RNN ?

- Entrées numériques
 - Données scientifiques \Rightarrow facile
 - Image, son \Rightarrow Facile...ou pas (convolution)
- Transformation des mots en numérique ?
 - Code ASCII ? Non !
 - Vectorisation sémantique (word2vec)



Vectorisation : exemple



IA : premiers pas

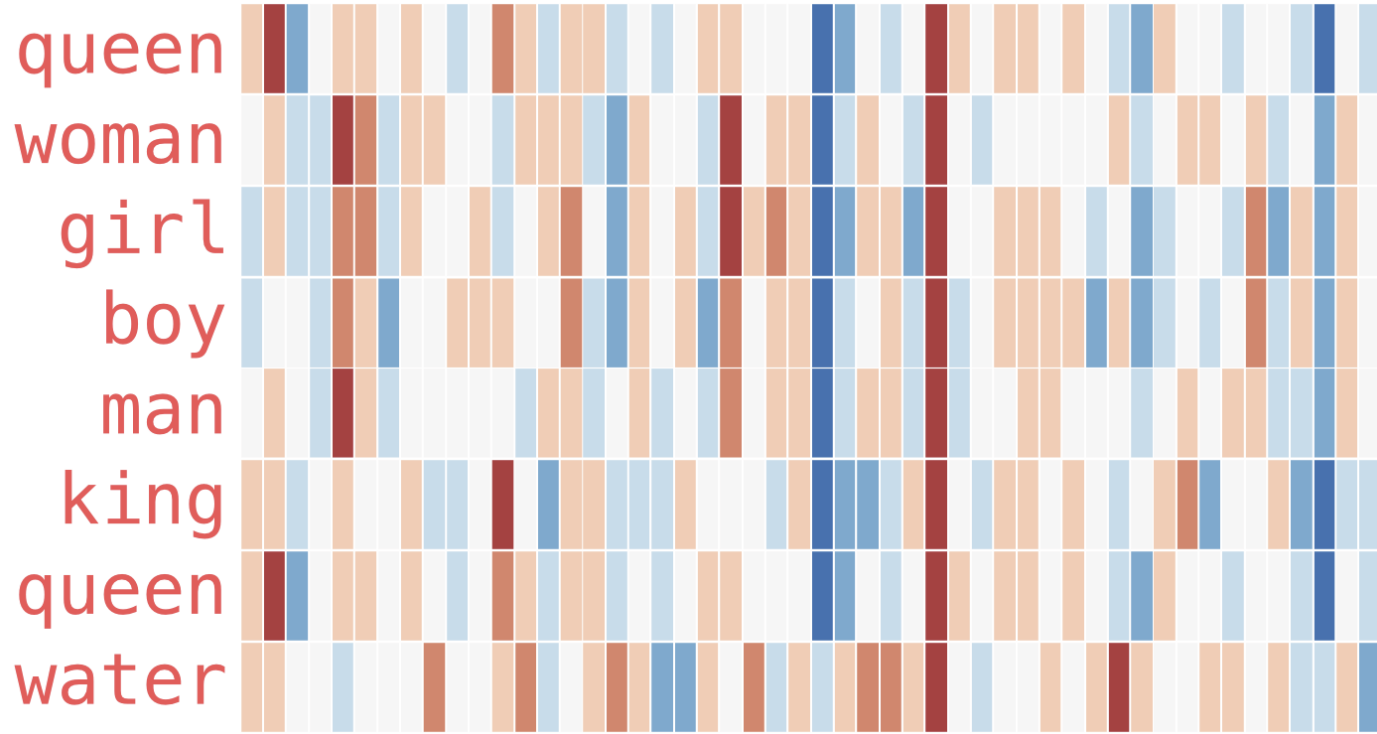


Vectorisation

- Les mots sont donc des vecteurs, des tableaux de chiffres.
- Nombre de « dimensions » : entre 100 et 1000
- La constitution de ces tableaux est faite par apprentissage
- On ne sait forcément pas ce que représentent les chiffres
- Les mots ont alors une valeur « sémantique »
- Les tables de vecteurs font donc partie des IA.
- Cela permet des opérations
 - $\text{Reine} = \text{Roi} - \text{Homme} + \text{Femme}$
 - $\text{Garçon} = \text{Fille} - \text{Femme} + \text{Homme}$



Vectorisation



Source : <https://jalammr.github.io/illustrated-word2vec/>



Vectorisation

- Création d'un nouveau type de base de données
 - GIS ou SIG (géographique)
 - Timeseries (les MRTG, RRDTOol)
 - Vectorielles
 - Pour stocker les « mots »



Vectorisation

- Indispensable pour « comprendre » la question.
- Question qui s'appelle un « prompt ».



LLM

- Large Language Model
- Absorption d'une quantité phénoménale de textes (tout wikipédia, ne représente que 5 % du total).
- Calcul de la probabilité du mot suivant, d'après les mots précédents (contexte).
 - Le choix se fait avec une « température »
- <https://linc.cnil.fr/dossier-ia-generative-chatgpt-un-beau-parleur-bien-entraîne>



LLM

- Les token sont les « mots » utilisés dans les LLM
- Mais en fait, mot n'est pas tout à fait égal à token (d'après chatGPT 1 token = 0,75 mot)
- Exemple
 - Une phrase de test en français
 - ['U', 'ne', ' phrase', ' de', ' test', ' en', ' fr', 'an', 'ç', 'ais']



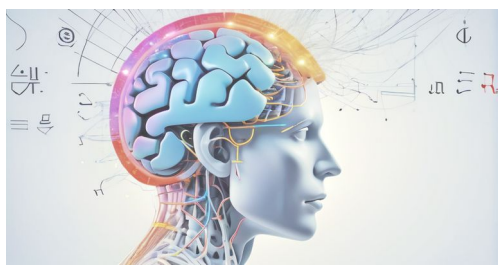
LLM

- Taille du contexte en tokens
 - D'où leur importance
 - de 8000 (IN + OUT) avec ChatGPT 3.5, mais certains parlent plutôt de 2048
 - à 64 000 (IN + OUT) avec ChatGPT 4.0 en API, et dans certaines conditions.
- Plus de contexte, plus de précision
- Les connaissances sont donc à 3 endroits
 - Dans les vecteurs
 - Dans le LLM
 - Dans le contexte



LLM

- Pas d'apprentissage durant l'utilisation
- 3 mois de calcul pour OpenAI pour 1 version de chatGPT
- Donc apprentissage asynchrone
 - MAIS...



IA: « apprentissage »

- Plusieurs buts
 - Augmentation de connaissance
 - Changement de comportement
- Plusieurs manières
 - Ajouter du contexte
 - Pré-prompt
 - RAG
 - Refaire/adapter le réseau de neurones
 - Modifier la vectorisation pour certains mots
 - Ajouter des couches intermédiaires.
 - Intercepter les prompts et les réponses.



LLM

- Il y a des « pré-prompt » ou « prompt système »
 - Rôle de cadrage (politiquement correct, précaution juridique)
 - Préciser les demandes (les GPTs .. qui sont des « experts »)
- Il y a des filtrages
 - Refus de termes dans les questions
 - Refus dans la réponse



IA : Les RAG

- Retrieval Augmented Generation
- Vectoriser des documents
- Et les considérer comme seule base de connaissance
- Absence d'hallucination



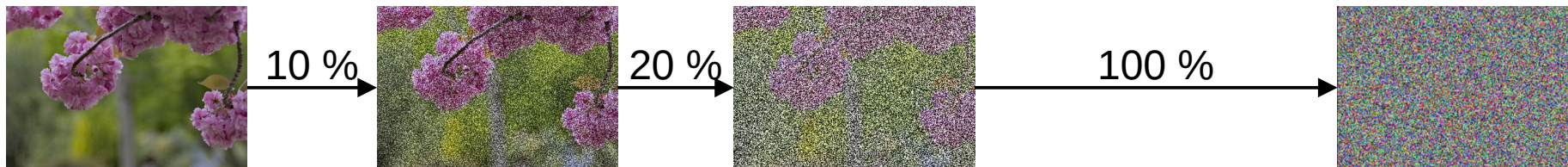
IA : le multimédia

IA : premiers pas



IA : Image : le débruitage

- <https://www.youtube.com/watch?v=tdeUss-5hY>



- Pack d'images nettes et bruitées d'un « objet »
- Entraînement à corriger de 10 % à 0% grâce aux RNN.
 - Et de 20 à 10 puis 0..... puis de 30 à 20 à 10 à 0
 - Que donne 100 à 0 ? ⇒ hallucination
- Multiples mots, dans le cadre des LLM



IA : image

- MidJourney
- StableDiffusion
 - <https://clipdrop.co>
 - Local
- DALL-E 3 (inclus dans ChatGPT).



IA : Image

- Prompt

- beautiful lady, (freckles), big smile, ruby eyes, pigtails hair, dark makeup, hyperdetailed photography, soft light, head and shoulders portrait, cover

- Notion de prompt négatif

- (worst quality, low quality, normal quality, lowres, low details, oversaturated, undersaturated, overexposed, underexposed, grayscale, bw, bad photo, bad photography, bad art:1.4), (watermark, signature, text font, username, error, logo, words, letters, digits, autograph, trademark, name:1.2), (blur, blurry, grainy), morbid, ugly, asymmetrical, mutated malformed, mutilated, poorly lit, bad shadow, draft, cropped, out of frame, cut off, censored, jpeg artifacts, out of focus, glitch, duplicate, (airbrushed, cartoon, anime, semi-realistic, cgi, render, blender, digital art, manga, amateur:1.3), (3D ,3D Game, 3D Game Scene, 3D Character:1.1), (bad hands, bad anatomy, bad body, bad face, bad teeth, bad arms, bad legs, deformities:1.3)



IA : Image

- Toujours quelques erreurs (« soupe de doigts », yeux bizarres), mais moins fréquentes
- Les prompts négatifs sont parfois très « limitants » sur l'imagination.
- La cohérence des images est encore un peu compliquée
- Les deepfakes sont désormais faciles.
- Toujours BEAUCOUP d'essais avant d'avoir quelque chose qui correspond à ce que l'on veut.
- Les mots sont en train d'arriver.
- Les « watermarks » sont parfois présentes (cf negative prompt).



IA : Vidéo

- Text to video (problème de cohérence)
- Une IA peut recopier un mouvement (OpenPose) et l'appliquer sur une photo.
- Capacité de faire des deepfakes en ligne



IA : son

- Synthèse vocale
- Comprendre de la voix sans apprentissage
- Générer de la musique
- Générer une chanson
- Copier une voix



IA : son

- Synthèse vocale
 - TTS : text to speech
 - Tout à fait crédible
 - Gérable localement.



IA : son

- Reconnaissance vocale
 - Gérable en local



IA : son

- Générer de la musique
 - Gérable en local
 - Bonne musique
- Générer de la chanson
 - En cours pour du local, mais avec grosses GPU
 - <https://suno.ai>
 - Assez bonne en anglais,
 - Moyenne en français.



IA: les paramètres

- Nombre de paramètres
 - Nb de neurones * nb de paramètres par neurone
 - Exemples (B \Rightarrow milliards)
 - ChatGPT 3.5 : 175B
 - Mistral en version 7B, 13B, 70B et dernièrement, 8 x 7B
- Quantisation (des poids des entrées des neurones)
 - Les poids passent de 32 bits à 8 voire 4 ou 2
 - Les poids deviennent des entiers
 - 5,32421524 \Rightarrow 5



Optimisation des IA

- Puissance nécessaire
- Mémoire nécessaire (mobile par exemple)
- Solution
 - Quantisation (FP32, FP16, Q8, Q4, Q2)
 - Réduction du nombre de paramètres
 - Compromis nombre de paramètres / quantisation
 - Mieux vaut quantisation que baisse de paramètres.
 - Version « turbo » apparemment c'est de la quantisation.



IA: puissance/mémoire

- Commerciaux \Rightarrow Inconnu
- Mistral
 - 7B \Rightarrow 13,5 Go
 - 13B \Rightarrow 30 Go
- StableDiffusion
 - 2.2 Go
 - 6,5 Go pour Juggernaut XL
- Compter entre 2 et 20 Go par IA



IA : Les défauts

- Les biais
- Les surcorrections
- Le contexte juridique
- Hallucinations
- Sécurité
- L'explicabilité
- L'impact social



Conclusion

- L'IA c'est bien
- Mais avoir un vrai cerveau pour l'utiliser, c'est bien aussi.



Compléments

- Pour les ceusses qui veulent des trucs en plus.



Défauts IA

- **Explicabilité**
 - Comprendre les erreurs.
 - Problème très difficile. Objet de travaux.
 - Cas du loup
 - Même Yann Le Cun (prix Turing 2018) se trompe sur les possibilités des LLM.
- **Date du jeu d'entraînement**
 - Limite les connaissances
- **L'absence de références (d'où vient l'information?)**



Défaut IA : les biais

- Jeu d'entraînement mondial... donc majoritairement américain
 - Meilleures réponses en anglais
 - Vision américaine (libéralisme, WASP, etc, mais pas que.)
- En LLM et en imagerie
 - Les femmes sont plus jeunes et plus belles (Si une femme ne s'estime pas « belle » ⇒ pas de photos, les mecs... s'en foutent)
 - Les hommes sont plus présents sur les périmètres de « pouvoir »
- Difficulté à dire « je ne sais pas ».



Défauts IA: la surcorrection

- ChatGPT :
 - Ode à (choisissez une minorité) ⇒ OK
 - Ode aux blancs ⇒ refus du prompt
- Stable diffusion (version online)
 - Refus de création d'images
 - D'enfant ⇒ au prompt
 - ou de « Hercules » ⇒ à la génération.



LLM : Baisse de qualité

- Baisse de qualité de ChatGPT (mais pas que)
 - <https://huggingface.co/spaces/lmsys/chatbot-arena-leaderboard>
 - En gros
 - Les IA « publiques » payantes sont vraiment touchées
 - Les API des IA publiques sont moins touchées
- Pourquoi ?
 - Risque juridique (cf biais)
 - Surcorrection (remontées majoritairement par les utilisateurs « sensibles » et « réactifs »)
 - Rendre la version plus rapide (moins de coût)
 - Rendre la réponse plus courte (surtout sur mobile).
 - <https://youtu.be/rAG5YtBsBTw>



Défaut IA: juridique

- Entraînement « illégal »
 - Bases pirates (copyright)
 - LAION (images) dont certaines seraient « pédophiles »
- Génération « illégale »
 - «Marvel Universe»,
 - « dans le style de... »
- Responsabilité
 - Politiquement incorrecte
 - Donnant une information fausse
- Deepfake
 - Visuel
 - Audio
 - Textuel (à la manière de...)



Défaut IA : les hallucinations

- Créativité \Rightarrow hallucination
 - It's not a bug, it's a feature.
- Nécessaire pour « combler les trous » de l'apprentissage
- Exemples
 - Invention
 - d'articles de loi,
 - de références bibliographiques.
 - de fonctions dans les langages.
 - Multiplicité des doigts.
 - ATTENTION, elle se trouve parfois dans la simple « augmentation de taille » d'une image.



Défauts IA : sécurité

- Cloud
 - Fuite des documents à l'étranger
 - Fuite des prompts (donc des documents) par le GPU
- Plugins
 - Piégeage de pages web de références (parle en émoticônes)
- Contourner les restrictions
 - Interdiction de donner le mot de passe.
- Piéger les données d'entraînement
 - Et les méthodes de protection... rendent le problème plus sournois.



Défauts IA : excès de confiance

- Anthropomorphisme
 - Exemple d'Eliza
 - Replika, Xiaolce, « Her »
- Pas de remise en cause (c'est l'ordinateur qui l'a dit).
- IA très affirmative (même si il y a des messages ajoutés).