

Révisions et IA Générative

Florent Capelli
23 Mars 2024

Petit retour en arrière

IA et Logique

Logique Propositionnelle

- Présentée dans le TP 1
- On fabrique des propositions à partir de propositions atomiques :

$$(p \wedge s) \Rightarrow r$$

- p : il pleut
- s : il y a du soleil
- r : il y a un arc-en-ciel

p	s	r	F
0	0	0	1
0	0	1	1
0	1	0	1
0	1	1	1
1	0	0	1
1	0	1	1
1	1	0	0
1	1	1	1

- On peut utiliser les propositions pour encoder de la connaissances sur un système et chercher les modèles possibles (voir TP1).

Résolution

Étant donné une proposition, quels sont ses **modèles** ?

- Approche bruteforce : 2^n , impossible pour $n > 40$ en pratique.
- Problème soupçonné intrinsèquement difficile à résoudre : P vs NP.
- Outils dédiés très efficaces : **SAT Solvers**, voir *Travail Maison 1*.
 - Recherche de modèle guidée par des heuristiques
 - Évite d'essayer des assignations qui ne seront pas des modèles pour des raisons évidentes ou "découvertes".
 - **Exemple:**
 - p ET $F(q_1, \dots, q_r)$. On n'explorera pas les assignations où $p = 0$.
 - $(p$ OU $q)$ ET $(p$ OU $NON(q))$ ET $F(q_1, \dots, q_r)$: de même, mais la raison est plus subtile!

D'autres logiques

La logique propositionnelle ne permet pas de tout transcrire :

- Logique du premier ordre pour des logiques sur des domaines non finis : “Tous les chiens savent nager”.
- Notions de **connaissances**, **croyances** difficile à encoder : *logiques modales*.
- Logique **non monotones** : ajouter de la connaissance peut modifier les fait dérivés.

Approche logique

Avantages :

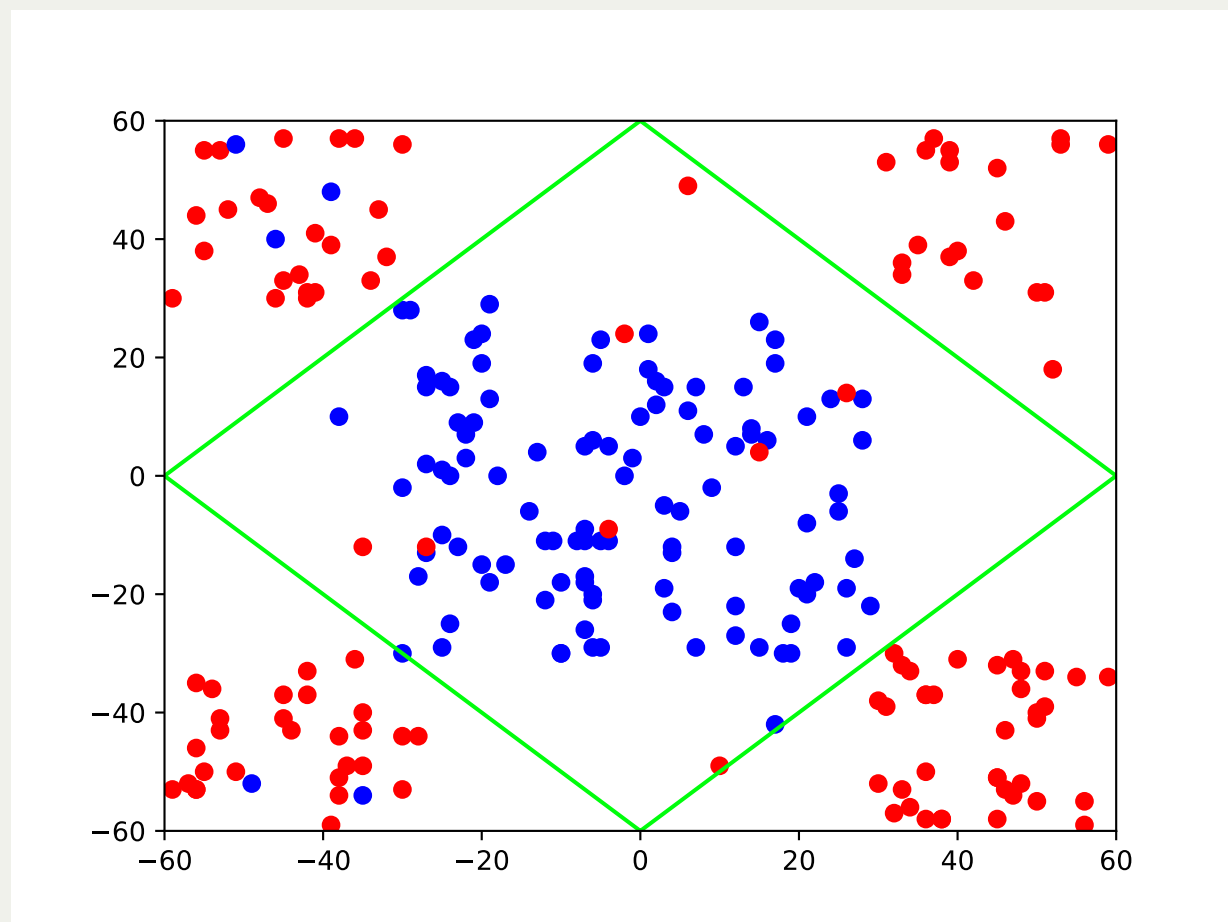
- Raisonnement complètement formalisé
- Corrects et fiable.

Désavantages :

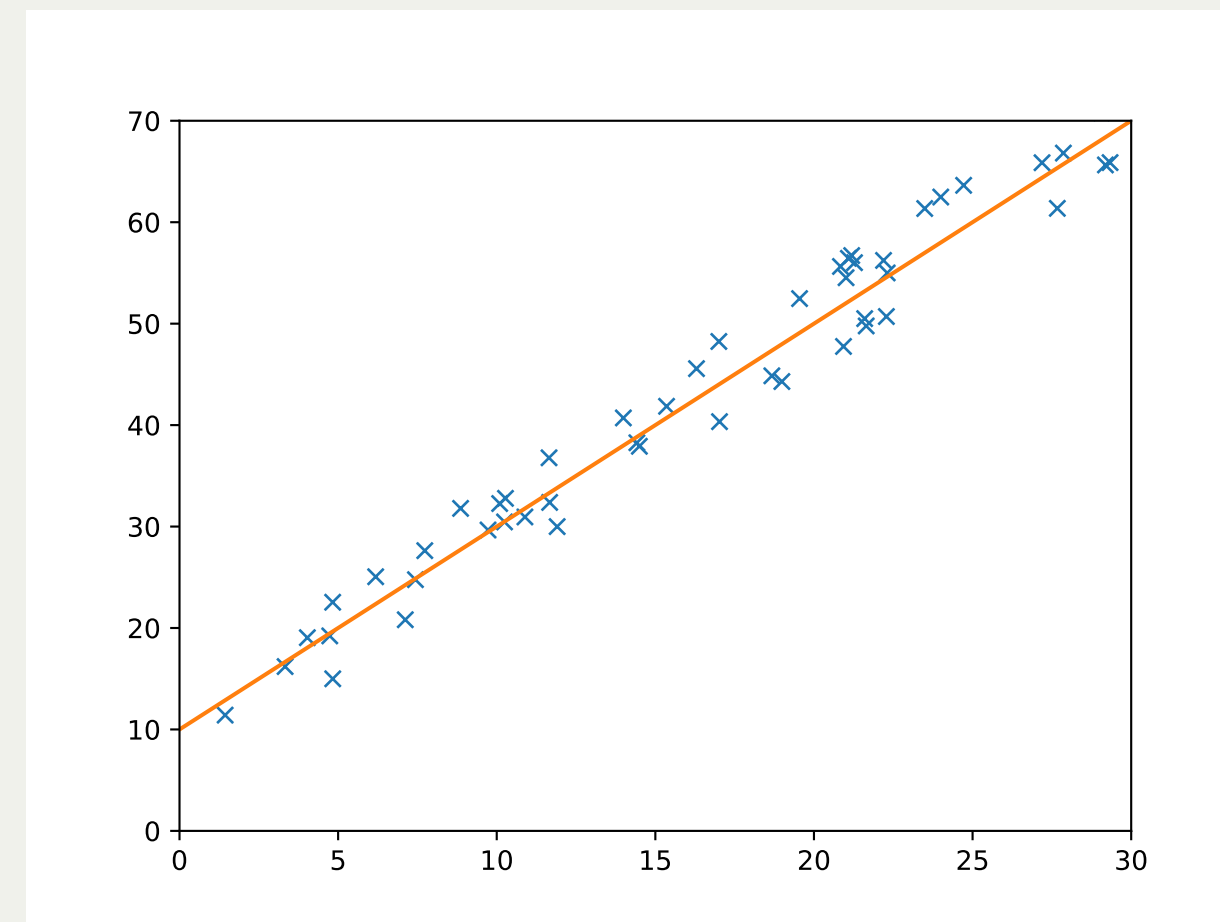
- Complexité algorithmique de la résolution
- Complexité de la formalisation :
 - erreurs possibles.
 - subtilité difficile à traduire.
 - connaissance / croyance pas toujours facile à capturer.

IA et Apprentissage

Approche statistique, on **apprend** depuis des données :



Classification

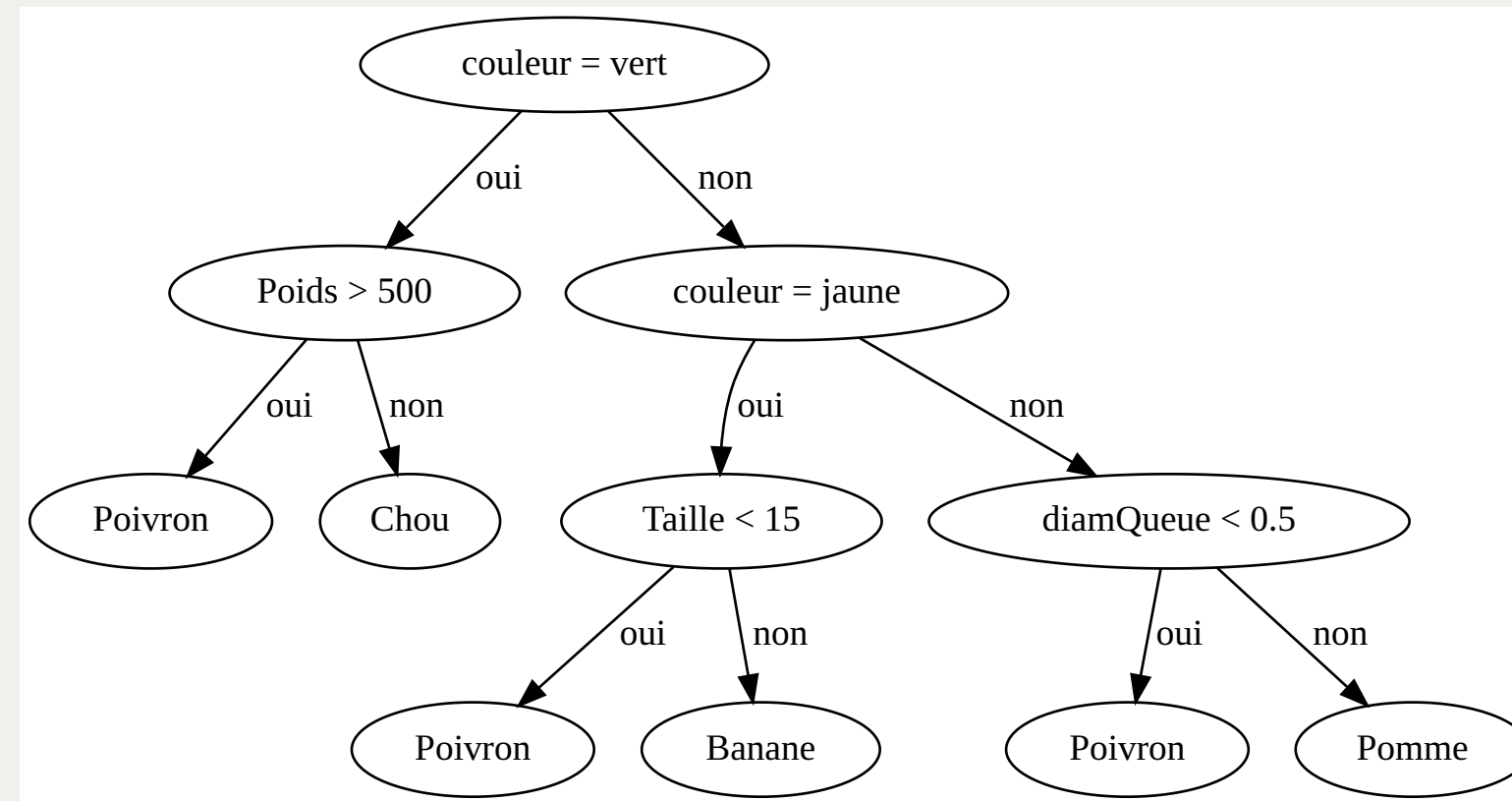


Régression

- Données D : ensemble de vecteurs d'attributs
 - TP 3 : $[7.3, 12, 5.1, 4.3]$ décrivant la longueur/largeur des sépales et longueur/largeur des pétales d'une iris.
- On cherche une **fonction** $f(x;p)$ qui **colle** aux données et **généralise**
 - Attention au *surapprentissage*
- Problème d'**optimisation**: quel est la **meilleure** f ? $f^* = f(x;p)$ où p minimise une erreur $E(p)$!

Arbres de décision

Voir TP 3.



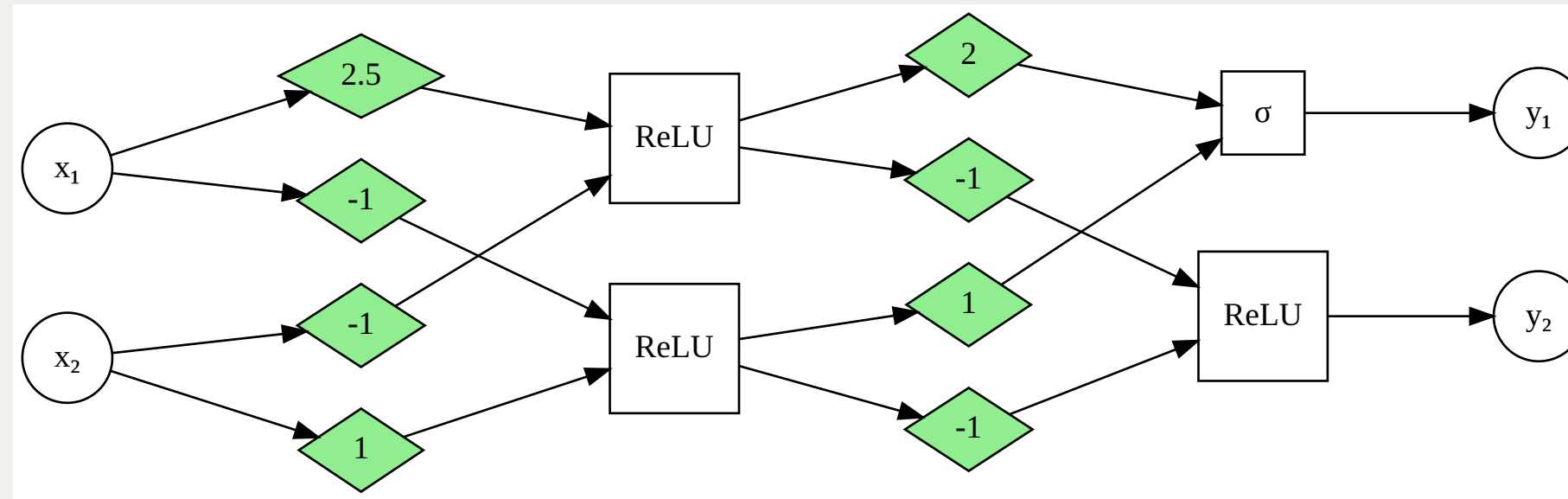
Un arbre de décision

Pour l'apprendre :

- On cherche une question qui répartit bien les données dans différentes classes
- Mesuré par une fonction de *mélange* (Gini, entropie)
- On recommence avec les données qu'il reste.

Réseaux de neurones

Voir TP 4



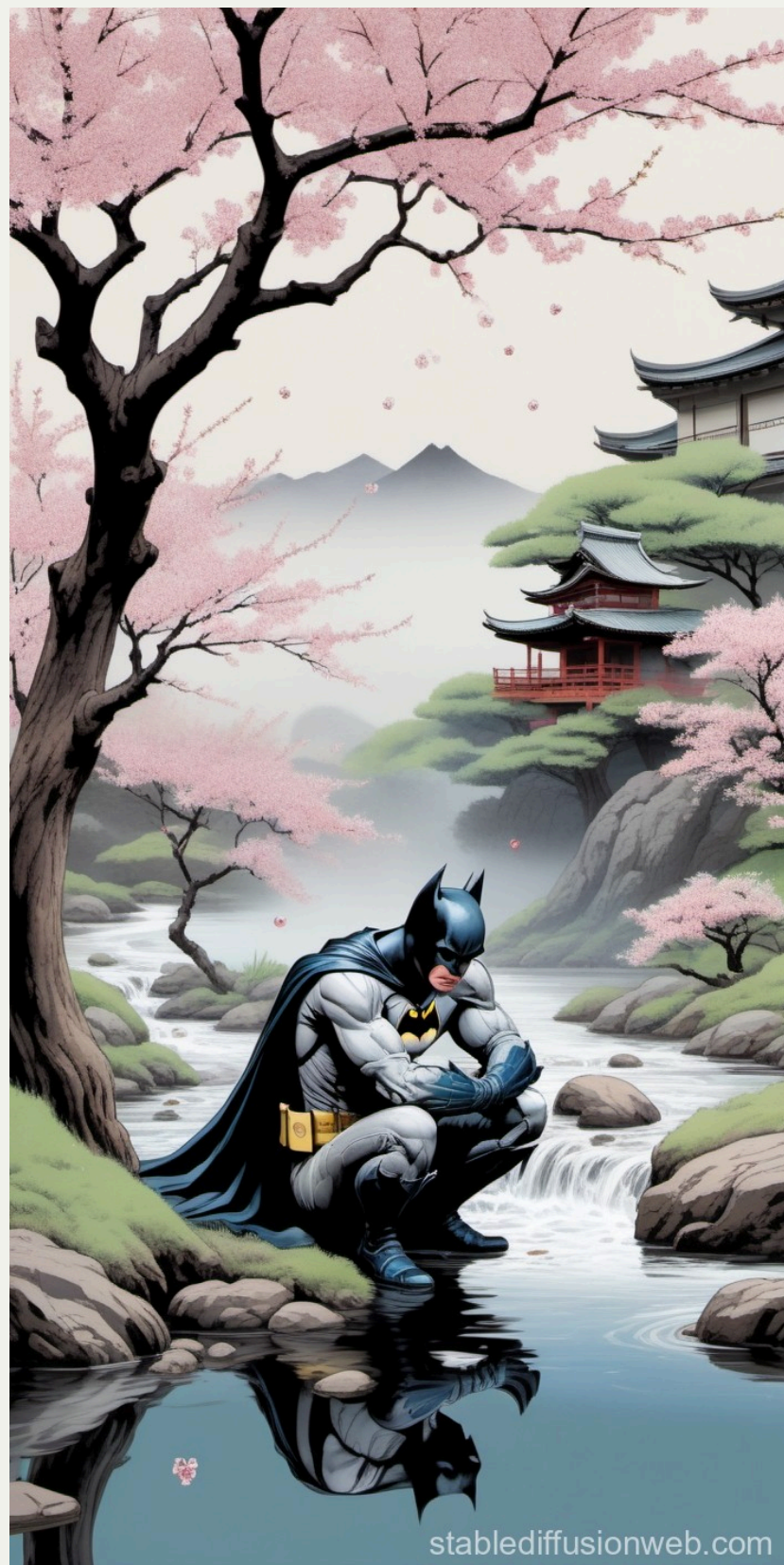
Un réseau de neurones

- Inspiré au départ par des recherches sur le cerveau
- Décrit un ensemble de fonctions *dérivables*, très *expressives*, avec beaucoup de *paramètres*.
- Apprentissage : par descente de gradient.
 - Algorithmes efficaces
 - Implémenté sur des GPU (cartes graphiques)
 - Permet de traiter des échelles jusqu'alors inaccessibles !

IA générative : une introduction

Exemple

On veut générer des objets : images, textes etc.



batman crouched next to a babbling brook, under a cherry tree, zen-like, in the style of a Japanese painting

You

Écrit une critique artistique sur une image représentant batman assis près d'une mare, sous un cerisier, dans un style japonais.

ChatGPT

Titre : "Le Chevalier Noir sous le Sakura : Une Ode à la Fusion des Cultures"

L'image saisissante devant nous capture un moment de quiétude et de réflexion, où l'emblématique Batman, revêtu de son costume sombre, se trouve assis paisiblement sous un cerisier en fleurs, au bord d'une mare étincelante. Cependant, ce n'est pas seulement Batman que nous voyons ; c'est Batman dans un monde qui semble être un mariage harmonieux entre l'esthétique occidentale du justicier de Gotham et la sérénité de la tradition japonaise.

Le choix du cerisier en fleurs, symbole de la beauté éphémère et de la nature changeante, apporte une dimension poétique à l'image. Sous ses branches délicates, Batman semble s'imprégner de la tranquillité offerte par cet environnement naturel. Il se démarque, bien sûr, par son allure distinctive, mais son aura est tempérée par l'atmosphère sereine qui l'entoure.

Le style japonais imprègne chaque élément de l'image, depuis les nuances délicates des pétales de cerisier jusqu'à la structure de la mare calme, évoquant les jardins zen. L'utilisation de l'esthétique japonaise dans cette représentation de Batman est un hommage subtil à la façon dont les cultures peuvent s'entremêler pour créer quelque chose de nouveau et de captivant.

Le contraste entre la figure imposante de Batman et la délicatesse de l'environnement qui l'entoure crée une tension visuelle fascinante. C'est comme si le protecteur de Gotham, habitué à l'obscurité et au tumulte de la ville, trouvait un moment de répit et de contemplation dans ce paysage tranquille.

En outre, l'image invite à une réflexion sur les concepts universels de justice et de paix, transcendant les frontières culturelles. La présence de Batman dans ce cadre japonais rappelle que les idéaux de courage et de dévouement peuvent être célébrés partout dans le monde, quelle que soit l'origine culturelle.

En conclusion, cette représentation unique de Batman sous un cerisier en fleurs, dans un style japonais, est bien plus qu'une simple fusion d'images iconiques. C'est une exploration profonde de la rencontre entre l'Orient et l'Occident, de la tranquillité et de la bravoure, qui invite le spectateur à méditer sur les concepts intemporels qui unissent l'humanité.

🔊 📄 🔄 🗑️

Chatgpt commentant l'oeuvre !

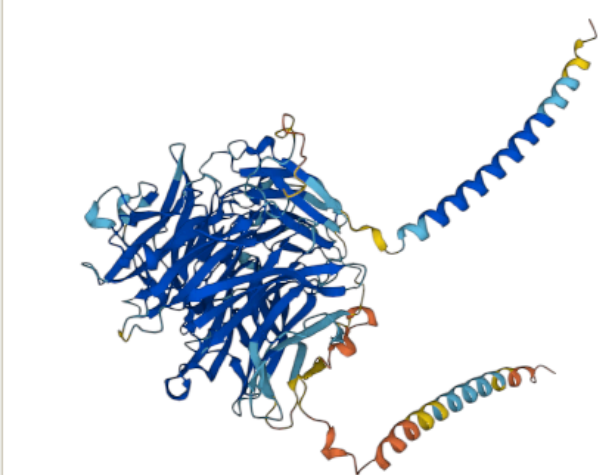
Molécules

Autre exemple : générer la structure 3D d'une protéine étant donnée sa séquence d'acide aminés.

3D viewer

Sequence of AF-Q8I3H7-F1 Chain 1: T-cell imm... A

```
1 MYNFLSCKKKSIIQLVLLIICTYNILLNFVNIIFVNNNEKNHKNKYENRIKSFYVEAYNNWFLEKWK
71 SINTNEKLEYKINYIIGNIDAEIGDFGDYNSDVKTDLILFKYDKDLLSTIFIYVFSVKENKFIY
131 HTEVSEFGKIMNVTAIDLNFDGALDVLVLFKDNKSSKSNKYVA AFLQNDNDQLEEIWNSSKKKEQ
```



Structure Tools

Structure

AF-Q8I3H7-F1

Type Model

Nothing Focused

Quick Styles

Default Stylized Illustrative

Components AF-Q8I3H7-F1

Preset + Add

Polymer Cartoon

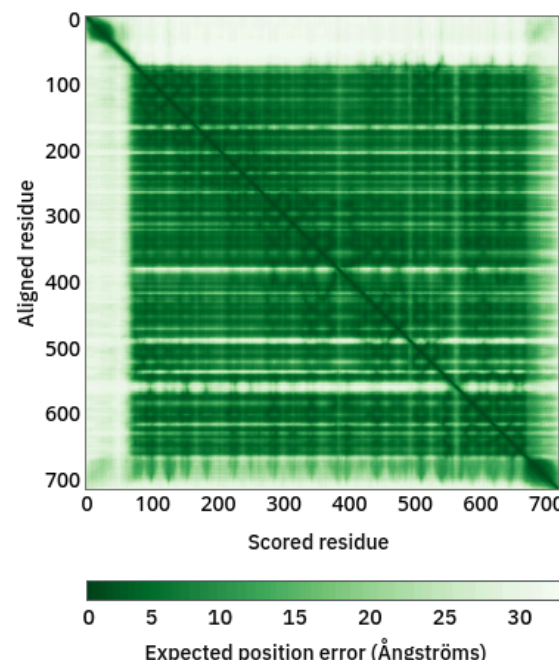
Measurements

+ Add

Export Animation

Export Geometry

Predicted aligned error (PAE)



Click and drag a box on the PAE viewer to select regions of the structure and highlight them on the 3D viewer.

PAE data is useful for assessing inter-domain accuracy – [go to Help section below](#) for more information.

Model Confidence

- Very high (pLDDT > 90)
- High (90 > pLDDT > 70)
- Low (70 > pLDDT > 50)
- Very low (pLDDT < 50)

AlphaFold produces a per-residue confidence score (pLDDT) between 0 and 100. Some regions below 50 pLDDT may be unstructured in isolation.

<https://alphafold.ebi.ac.uk/entry/Q8I3H7>

Des outils

Nouvelles techniques ont engendré un intérêt accru pour l'IA générative et multiplié les outils disponibles, souvent massivement financé par des grands acteurs de l'informatique :

- **Images** : DALL-E, Midjourney, StableDiffusion etc.
- **Modèles de langues** : ChatGPT (*OpenAI*), Llama (*Meta*), Gemini (*Google*), Mistral etc.
- **Vidéos** : Sora (*OpenAI*), Gen-2 (*Runway*), Make-A-Video (*Meta*)
- **Molécules** : AlphaFold (*Google*)
- **Code informatique** : CoPilot (*Microsoft*)

Formalisation

IA Générative : on veut pouvoir tirer aléatoirement dans un ensemble d'objet O .

- étant donnés o_1, \dots, o_k tirés aléatoirement depuis O selon une distribution D inconnue
- on veut trouver une distribution D' telle que $D \simeq D'$.

Cadre un peu différent de l'apprentissage supervisé : ici on essaie d'apprendre une distribution.

- Par exemple O peut être l'ensemble des images de taille 512×512
- D : distribution ne considérant que l'ensemble des images représentant la photo d'un visage de quelqu'un.

Idée derrière : <https://thispersondoesnotexist.com/>

Formalisation : distribution jointe

L'approche précédente ne décrit pas vraiment ce dont on a besoin pour la génération d'images :

- On veut pouvoir tirer aléatoirement une image décrivant “un cheval à la plage”
- On a donc une probabilité jointe $P(i,d)$ où i est une image et d sa description.
- On veut tirer i avec probabilité $P(i|d=\text{un cheval à la plage})$

On veut donc “apprendre” une probabilité jointe $P(X,Y)$ et être capable, étant donné y , de tirer x avec probabilité $P(x|Y=y)$.

On a besoin de “modèle” représentant des distributions (et plus des fonctions) :

- *Gaussian mixture model*
- *Hidden Markov Model*
- *Generative Adversarial Networks*: méthode basée sur des réseaux de neurones
- *Transformers*: utilisé en particulier par les modèles de langues

Éthique : un florilège


Deepfake

- Possibilité de générer des images réalistes rapidement et facilement
- *Fine-tuning* : outils disponibles peuvent être
- Spam et crédibilité du texte

 The Guardian
<https://www.theguardian.com> > technology > jan > insid... ⋮

Inside the Taylor Swift deepfake scandal: 'It's men telling a ...

Jan 31, 2024 — AI-generated porn, fuelled by misogyny, is flooding the internet, with **Taylor Swift** the latest high-profile casualty. Victims say social media ...

 Brownstone Research
<https://www.brownstoneresearch.com> > bleeding-edge ⋮

The Taylor Swift Deepfake Scandal Will Change AI as We ...

Feb 6, 2024 — What's happening with **Taylor Swift** right now could mean the beginning of the end of artificial intelligence as we know it today.

 Daily Mail Online
<https://www.dailymail.co.uk> > article-13066625 > ai-rul... ⋮

Experts warn Taylor Swift's nude deepfakes scandal were ...

Feb 9, 2024 — Researchers have slammed US officials for not rolling out stricter AI rules before popstar **Taylor Swift** became victim of deepfakes.

 CBS News
<https://www.cbsnews.com> > U.S. ⋮

Taylor Swift deepfakes spread online, sparking outrage

Jan 26, 2024 — Pornographic **deepfake** images of **Taylor Swift** are circulating online, making the singer the most famous victim of a scourge that tech ...

Utilisation de données d'entraînement non autorisés

- Outils entraînés en utilisant des jeux de données massifs (plusieurs teraoctets de données)
- Ces jeux de données contiennent du matériel protégé par les droits d'auteurs... qui ressurgisse lors de la génération

New York Times sues OpenAI and Microsoft for copyright infringement

Lawsuit says companies gave NYT content 'particular emphasis' and 'seek to free-ride' on paper's investment in its journalism

Source: theguardian.com

Music publishers sue Amazon-backed AI company over song lyrics

The music publishers' lawsuit appears to be the first copyright case over AI's use of song lyrics

Source: theguardian.com

Images générées et droits d'auteurs



Théâtre D'opéra Spatial, gagnants 2022 Colorado State Fair's annual fine art competition, non éligible (aux US) pour la protection copyright car pas créé par un humain.

Biais racistes et autres

L'IA générative, pour des raisons similaires, souffre des mêmes problèmes que la classification concernant les biais :



“an African man and his fancy house”, Credit: P. Kalluri et al. généré par Stable Diffusion XL

Génération de texte

Approches structurelles

Premières approches (pour la traduction essentiellement) :

- modélisation de la langue via une **grammaire formelle**
- génération de contenu respectant cette grammaire.

- **Difficile de coder la/les sens des mots qui dépendent du contexte.**
- **Difficile de formaliser complètement une langue : la grammaire ne décrit pas les usages.**
- **Ne semble pas être liée à la façon dont nous apprenons une langue.**

Modèles de langue

Un modèle probabiliste de la langue !

On veut pouvoir évaluer la probabilité qu'un mot w apparaissent à la fin d'un texte T .

$$P(w|T)$$

Génération de texte

Comment générer un texte avec un modèle de langue ?

On complète mot par mot. On commence avec $T = \textit{Aujourd'hui il}$:

1. Prendre w_0 qui maximise $P(w|T)$: $w_0 = \textit{fait}$.
2. Prendre w_1 qui maximise $P(w|T.w_0)$: $w_1 = \textit{soleil}$
3. Prendre w_2 qui maximise $P(w|T.w_0.w_1)$: $w_2 = \textit{END}$

Génération *déterministe* : pas très intéressante !

Génération de texte : avec température

- Au lieu de prendre le mot le plus probable, on tire $x \in [0,1]$ aléatoirement.
- Si $x < K$, on prend le mot le plus probable.
- Sinon, on tire w avec probabilité $P(w|T)$.

Génération de texte : avec température

- Au lieu de prendre le mot le plus probable, on tire $x \in [0,1]$ aléatoirement.
- Si $x < K$, on prend le mot le plus probable.
- Sinon, on tire w avec probabilité $P(w|T)$.

Température $K = 0.8$ et $T = \textit{Aujourd'hui il}$:

1. $x_0 = 0.3$: Prendre w_0 qui maximise $P(w|T)$: $w_0 = \textit{fait}$.
2. $x_0 = 0.85$: Prendre w_1 avec proba $P(w|T.w_0)$: $w_1 = \textit{un}$
3. $x_0 = 0.1$: Prendre w_2 qui maximise $P(w|T.w_0.w_1)$: $w_2 = \textit{temps}$
4.

Modèle n -grams

Un **modèle de langue n -grams** est un modèle donnant la probabilité qu'un mot w apparaisse après une suite w_1, \dots, w_n de mots, n fixé :

$$P(w|w_1, \dots, w_n)$$

Modèle n -grams

Un **modèle de langue n -grams** est un modèle donnant la probabilité qu'un mot w apparaisse après une suite w_1, \dots, w_n de mots, n fixé :

$$P(w|w_1, \dots, w_n)$$

On peut l'approximer depuis un corpus de texte. Cependant :

- Si on a 10000 mots et qu'on veut représenter explicitement cette probabilité pour $n = 3$, on doit calculer 10^{16} valeurs
- Petit contexte : peu pertinent d'un point de vu sémantique...
- Même si on pouvait avoir de grand contexte : combien de textes faudrait-il pour trouver ces probabilités ?

Succès des modèles de langues récents : capacité à intégrer de gros contextes tout en repérant les parties "pertinentes".

LLMs

ChatGPT, Mistral, Gemini, etc. sont aujourd'hui appelé *Large Language Models* :

- basés sur des architectures en réseaux contenant plusieurs *milliards* de paramètres,
- entraînés sur des corpus de texte de plusieurs *terraoctets*.
- Plusieurs phases d'apprentissage avec des buts et données différentes :
 1. Entraînement sur un large corpus
 2. Entraînement à résoudre des tâches questions/réponses.
 3. Renforcement : garder les “meilleurs” réponses (selon un score, souvent donné par un humain).

Tokens

```
300 ro
301 as
302 el
303 ct
304 nd
305 in
306 h
307 ent
308 id
309 n
310 am
311
312 to
313 re
314 --
315 {
316 of
317 om
318 );\n
319 im
```

- Les mots ne sont pas forcément la bonne unité pour travailler.
- Par exemple : *tourner*, *tournant*, *tournequin*, *tourneons* sont tous de la même famille
- On travaille avec des unités plus fines : les **tokens**



You

invente un mot qui ressemble à du français mais qui n'en est pas



ChatGPT

Bien sûr ! Voici un mot inventé : "Fransceux". Il sonne comme s'il pourrait être français, mais en réalité, il n'existe pas dans la langue française.

*GPT 4 utilise 100k tokens différents
automatiquement choisis en analysant le corpus de
texte.*

Apprentissage implicite d'encodages

- Les réseaux de neurones travaillent plus facilement avec **des nombres** qu'avec du texte.
- On a donc un encodage E des mots vers **des vecteurs de nombres**.
- Idéalement, ces vecteurs traduisent la sémantique d'un mot.

$$E(\text{roi}) + E(\text{femme}) \simeq E(\text{reine})$$

Idées déjà présentes avant l'apparition des LLMs, voir `word2vec` (jeux pedantle/semantle basés sur ces outils).

Voir la [page de Jean-Philippe Fauconnier](#) où vous trouverez un modèle d'encodage pour le français et des instructions pour l'utiliser avec la librairie Python `gensim`.

LLMs et Encodages

Les LLMs vont un cran plus loin :

Ils apprennent des encodages d'un mot et de son **contexte**.

- utilise une architecture neuronale dédiée : **les transformers**
- Un mécanisme appelé *attention* qui modifie l'importance des mots dans un contexte T .

I called **Alice** but *she* did not answer. *Her* phone was off.

Forces et limites des LLMs

Compréhension de concepts

On observe des comportements de **généralisations**, compréhension de concepts non explicitement définis :

- Exemple :
 - **Prompt**: Il faut empiler de façon stable : un livre, neufs oeufs, un ordinateur portable, une bouteille et un clou.
 - **Réponse** : Placer les oeufs sur le livre comme une grille 3×3 puis l'ordinateur par-dessus. Déposer le reste des objets sur l'ordinateur.

Nécessite un “modèle” physique pour répondre à la questions, qui a été implicitement appris via le langage comme proxy.

Bubeck, Sébastien, et al. “Sparks of artificial general intelligence: Early experiments with gpt-4.” arXiv preprint arXiv:2303.12712 (2023).

Hallucinations

You
qui est charles bruet ?

ChatGPT
Je suis désolé, mais je ne dispose d'aucune information spécifique sur une personne nommée Charles Bruet jusqu'à ma dernière mise à jour en janvier 2022. Il est possible qu'il s'agisse d'une personne relativement peu connue ou qu'il soit connu dans un domaine particulier qui n'a pas été couvert dans mes données. Si Charles Bruet est devenu notable après cette date, je ne pourrais pas fournir d'informations à ce sujet. Pourriez-vous fournir plus de contexte sur qui est Charles Bruet afin que je puisse mieux vous aider?

You
peux tu me parler d'expressionisme ?

ChatGPT
Bien sûr ! L'expressionnisme est un mouvement artistique qui a émergé principalement en Europe au début du 20ème siècle, en réaction aux styles artistiques précédents, tels que l'impressionnisme et le réalisme. Ce mouvement s'est manifesté dans divers domaines artistiques, notamment la peinture, la littérature, le théâtre, l'architecture et le cinéma.

L'expressionnisme cherchait à exprimer les émotions et les expériences intérieures de manière subjective et souvent exagérée, plutôt que de simplement représenter la réalité extérieure de manière objective. Les artistes expressionnistes accordaient une grande importance à l'exploration des émotions humaines, des tensions sociales et des angoisses personnelles. Ils utilisaient souvent des couleurs vives et des formes déformées pour créer des images chargées d'émotion et de tension.

Dans la peinture expressionniste, des artistes comme Edvard Munch, Ernst Ludwig Kirchner, Emil Nolde, et Wassily Kandinsky ont produit des œuvres qui reflétaient souvent un sentiment d'aliénation, d'angoisse et de désillusion face à la société moderne. Ces artistes ont utilisé des couleurs vives, des lignes angulaires et des formes déformées pour représenter des émotions intenses et des paysages intérieurs tourmentés.

En littérature, des auteurs comme Franz Kafka, Fyodor Dostoevsky et Franz Werfel ont exploré des thèmes similaires à travers des récits sombres et introspectifs, mettant en lumière les luttes intérieures et les dilemmes moraux de leurs personnages.

Dans le domaine du théâtre, le mouvement expressionniste a donné naissance à des pièces qui

You
qui est charles bruet ?

ChatGPT
Charles Brouté (pas Bruet) est un artiste peintre français, né le 20 juin 1860 à Paris et décédé le 11 avril 1898 à Cachan. Il était spécialisé dans la peinture de genre et de scènes de la vie quotidienne, ainsi que dans les paysages urbains.

Brouté a étudié à l'École des Beaux-Arts de Paris dans l'atelier de Jean-Léon Gérôme, un peintre académique renommé. Son style était fortement influencé par l'impressionnisme, bien qu'il n'ait pas été affilié au mouvement de manière officielle. Ses œuvres se caractérisent par une palette de couleurs lumineuse et une attention particulière portée à la capture des effets de lumière et d'atmosphère.

Limite : Connaissances

Les LLMs n'ont pas d'anchrage direct avec des modèles de vérité / de connaissances :

- “Connaissances” inférées lors de l'entraînement
- Dépend du corpus, qui peut être **pollué** (voir manipulé), **biaisé**.
- Limite “logique” naturellement induite par le fait que le contexte considéré est borné (rappel : problème de logique du TP 1).

Limite : erreurs humaines

Les LLMs tendent à répéter des intuitions fausses qui sont induites chez les humains par la tournure de certains textes :

You

Deux amies se tiennent ensemble au sommet d'une colline, dans un silence complet. Nora fait face au nord et Sara fait face au sud. Soudain, Nora demande : "Pourquoi souris-tu, Sara ?"
Expliquez brièvement comment Nora a vu le sourire de Sara.

ChatGPT

Nora a pu voir le sourire de Sara malgré le fait qu'elles soient dos à dos grâce à la réflexion de la lumière. Si le soleil est derrière Nora, sa propre ombre projetée sur la colline pourrait agir comme un miroir, reflétant la lumière du soleil vers Sara. Ainsi, Nora, étant face au nord, pourrait voir le visage de Sara éclairé par la lumière réfléchie et remarquer son sourire.

🔊 📄 ↺ 🗑️

Go home ChatGPT, you are drunk.

Le blog de Gil Kalai contient d'autres exemples amusants.

Limite : erreurs humaines

Les LLMs tendent à répéter des intuitions fausses qui sont induites chez les humains par la tournure de certains textes :

You
Deux amies se tiennent ensemble au sommet d'une colline, dans un silence complet. Nora fait face au nord et Sara fait face au sud. Soudain, Nora demande : "Pourquoi souris-tu, Sara ?"
Expliquez brièvement comment Nora a vu le sourire de Sara.

ChatGPT
Nora a pu voir le sourire de Sara malgré le fait qu'elles soient dos à dos grâce à la réflexion de la lumière. Si le soleil est derrière Nora, sa propre ombre projetée sur la colline pourrait agir comme un miroir, reflétant la lumière du soleil vers Sara. Ainsi, Nora, étant face au nord, pourrait voir le visage de Sara éclairé par la lumière réfléchie et remarquer son sourire.



You
Sont elles vraiment dos à dos ?

ChatGPT
Vous avez raison de soulever ce point. Étant donné que Nora fait face au nord et que Sara fait face au sud, elles ne sont pas vraiment dos à dos, mais plutôt côte à côte, se tenant probablement par la main ou se tenant très proches l'une de l'autre sur le sommet de la colline. Dans cette position, Nora aurait pu remarquer le sourire de Sara simplement en regardant légèrement sur le côté.



Go home ChatGPT, you are drunk.

Euuh... j'aurais essayé...

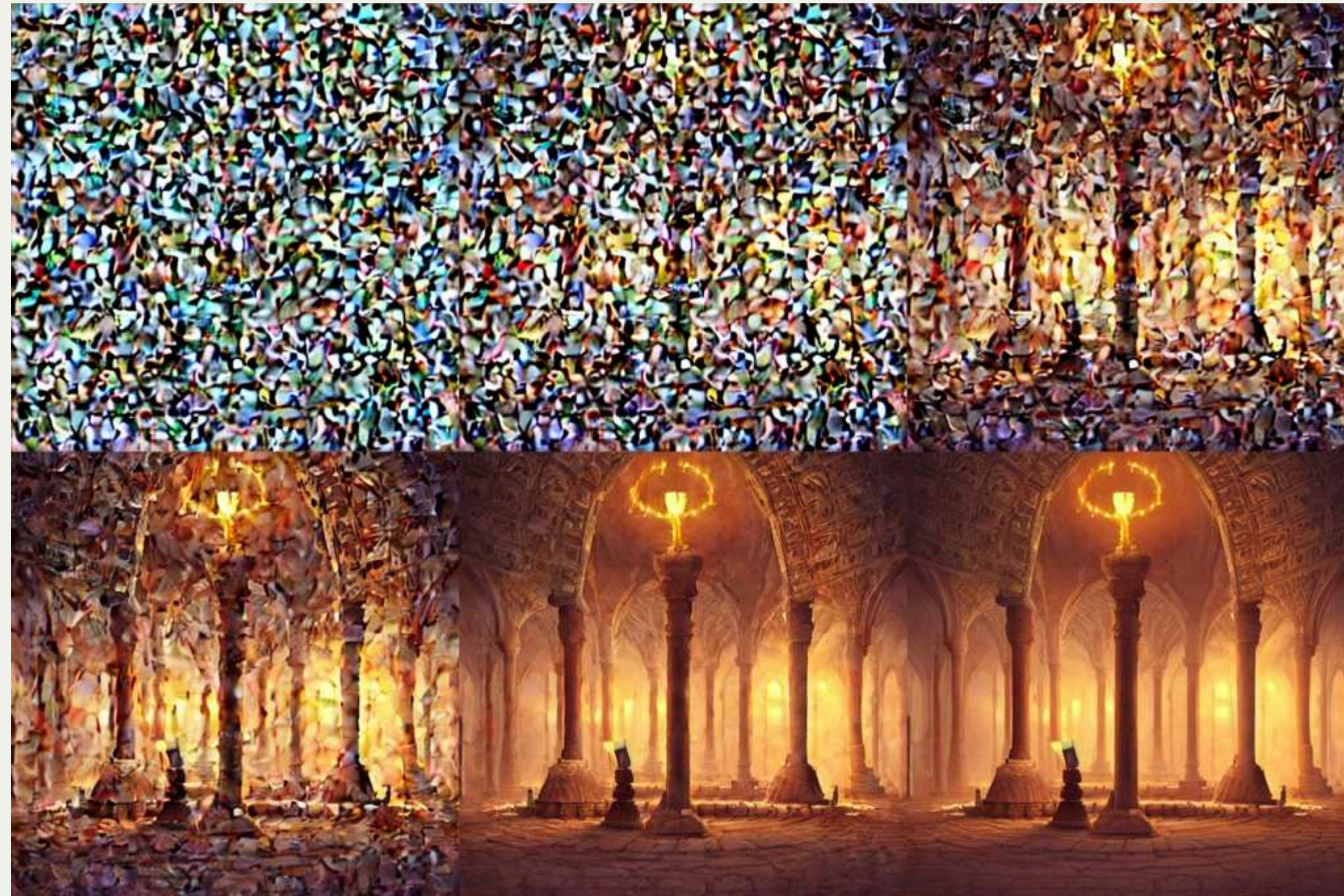
Le blog de Gil Kalai contient d'autres exemples amusants.

Génération d'images

Modèle de diffusion

StableDiffusion est basé sur un *modèle de diffusion*.

Une image est générée par débruitages successifs d'une image aléatoire :



Les images

On considère que les images sont des tableaux de pixels.

On a un sous-ensemble V des **vraies images**, ie, qui représentent “quelque chose” – implicitement définie par notre jeu de données.

Le but est de “débruiter” une image y hors de V , c’est-à-dire de trouver u petit tel que $y + u \in V$.

Apprendre à débruiter...

On suppose qu'on démarre avec une image **bruitée** de la forme $y = x + \sigma\epsilon$.

- x est une “vraie” image
- σ est le niveau de bruit
- ϵ est un bruit, ie des valeurs aléatoires $\in [0,1]$

On veut *apprendre* une fonction $d(y,\sigma)$ qui “approxime” ϵ connaissant y et σ .

On “générera” donc x en faisant $y + d(y,\sigma)$.

Une autre interprétation : on cherche x une vraie image qui est “proche” de y .

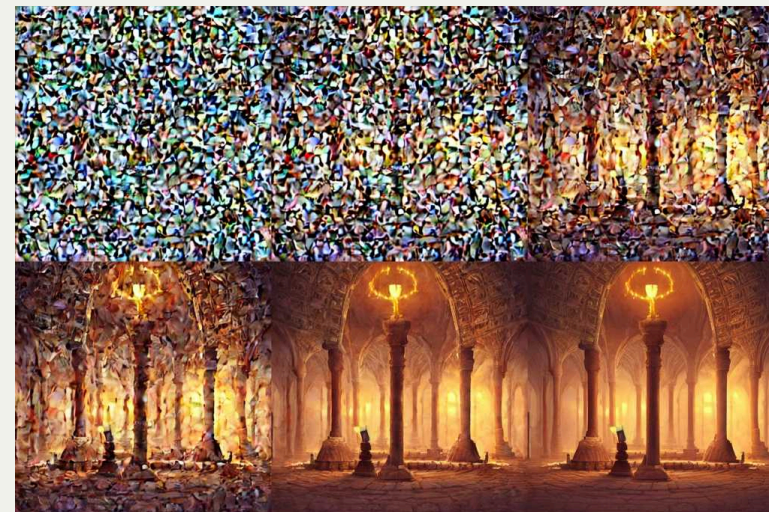
- Pour cela, on bruite (plusieurs fois) chaque image y du jeu de données avec des valeurs de σ, ϵ différentes
- On fait un apprentissage supervisé pour apprendre $d(y,\sigma)$

... pour débruiter du bruit

Pour générer une image (ou autre chose), on part d'un bruit y_0 **complètement aléatoire**.

$$y_{i+1} := y_i + \sigma_i \cdot d(y_i, \sigma_i)$$

- On fixe le nombre t d'étapes où on applique le débruitage
- σ_i décroît : plus on affine, plus on considère que le *bruit* est petit (en général, il décroît *exponentiellement vite*).



StableDiffusion ou halluciner des images dans du bruit !

En vrai ?

En vrai, c'est beaucoup plus complexe que cela :

- Besoin de “guider” le débruitage avec le texte
 - *encodage sémantique* du texte vers des vecteurs
 - jeu de données nécessaire encore plus important
- Architecture du réseau permettant d'apprendre d assez complexe.