

Pour commencer, L'IA générative ?

Quelles technologies et quelles applications ?

Notions clés

- L'IA « traditionnelle » se concentre sur la **détection de « patterns »**, la **classification** et la **prédiction de données**
- L'IA « générative » permet de **créer de nouvelles données** à partir de **données exemples**

Pour commencer, L'IA générative ?

Quelles technologies et quelles applications ?

8

Notions clés

- L'IA « traditionnelle » se concentre sur la **détection de « patterns »**, la **classification** et la **prédiction de données**
- L'IA « générative » permet de **créer de nouvelles données à partir de données exemples**

Technos

Deep Learning – Types de modèles d'IA générative



Les architectures de réseau de Neurones peuvent prendre différentes formes plus ou moins spécifiques à certains types de tâches

Pour commencer, L'IA générative ?

Quelles technologies et quelles applications ?

Notions clés

- L'IA « traditionnelle » se concentre sur la **détection de « patterns »**, la **classification** et la **prédiction de données**
- L'IA « générative » permet de **créer de nouvelles données à partir de données exemples**

Technos

Deep Learning – Types de modèles d'IA générative



GANs (Generative Adversarial Networks)

• **Principe** : Deux réseaux neuronaux s'affrontent : un générateur (qui crée des données) et un discriminateur (qui évalue leur authenticité).

• **Applications** :

- Génération d'images (DeepArt, visages réalistes).
- Création de vidéos, musique ou modèles 3D.
- Amélioration de la qualité des images (upscaling).

• **Exemples** : StyleGAN, BigGAN.

Pour commencer, L'IA générative ?

Quelles technologies et quelles applications ?

Notions clés

- L'IA « traditionnelle » se concentre sur la **détection de « patterns »**, la **classification** et la **prédiction de données**
- L'IA « générative » permet de **créer de nouvelles données à partir de données exemples**

Technos

Deep Learning – Types de modèles d'IA générative



b) VAEs (Variational Autoencoders)

• **Principe** : Encode les données dans un espace latent continu, puis les reconstruit en introduisant une certaine variation.

• **Applications** :

- Génération d'images et de vidéos.
- Synthèse vocale et audio.
- Analyse et compression des données.

• **Avantage** : Génère des sorties réalistes tout en offrant un contrôle sur les variations.

Pour commencer, L'IA générative ?

Quelles technologies et quelles applications ?

Notions clés

- L'IA « traditionnelle » se concentre sur la **détection de « patterns »**, la **classification** et la **prédiction de données**
- L'IA « générative » permet de **créer de nouvelles données à partir de données exemples**

Technos

Deep Learning – Types de modèles d'IA générative



c) Transformers

• **Principe** : Utilise des mécanismes d'attention pour modéliser les relations dans des séquences de données.

• **Applications** :

- Génération de texte (ChatGPT, GPT-4, Bard).
- Traduction automatique.
- Synthèse de code ou d'images (DALL-E, Stable Diffusion).

• **Exemples** : GPT (OpenAI), BERT, T5.

Pour commencer, L'IA générative ?

Quelles technologies et quelles applications ?

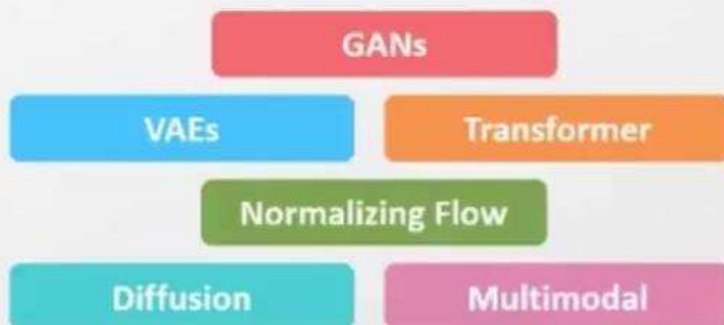
8

Notions clés

- L'IA « traditionnelle » se concentre sur la **détection de « patterns »**, la **classification** et la **prédiction de données**
- L'IA « générative » permet de **créer de nouvelles données à partir de données exemples**

Technos

Deep Learning – Types de modèles d'IA générative



c) Diffusion Models

- **Principe** : Transforme progressivement du bruit aléatoire en données structurées.
- **Applications** :
 - Génération d'images ultra-réalistes.
 - Animation ou génération vidéo.
- **Exemples** : Stable Diffusion, Imagen (Google), DALL-E 2

Pour commencer, L'IA générative ?

8

Quelles technologies et quelles applications ?

Notions clés

- L'IA « traditionnelle » se concentre sur la **détection de « patterns »**, la **classification** et la **prédiction de données**
- L'IA « générative » permet de **créer de nouvelles données à partir de données exemples**

Technos

Deep Learning – Types de modèles d'IA générative

GANs

VAEs

Transformer

Normalizing Flow

Diffusion

Multimodal

Applications



Image à partir
d'image



Texte à partir
de texte



Audio à partir
d'audio

Agent
conversationnel

Pour commencer, L'IA générative ?

8

Quelles technologies et quelles applications ?

Notions clés

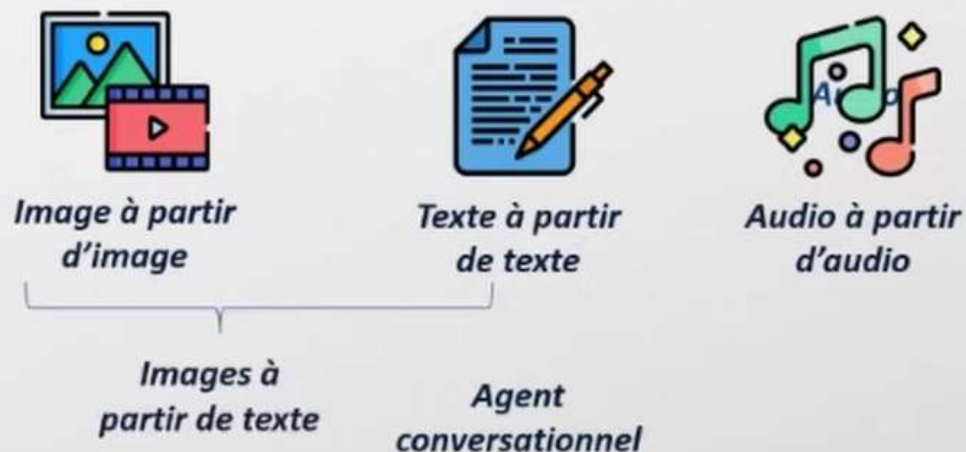
- L'IA « traditionnelle » se concentre sur la **détection de « patterns »**, la **classification** et la **prédiction de données**
- L'IA « générative » permet de **créer de nouvelles données à partir de données exemples**

Technos

Deep Learning – Types de modèles d'IA générative



Applications



DALI qui permet de créer des images à partir d'une description, le modèle va comprendre à la fois le sens de votre de votre input de votre requête et à la fois être en capacité de générer l'image qui correspond à sa compréhension

Pour commencer, L'IA générative ?

8

Quelles technologies et quelles applications ?

Notions clés

- L'IA « traditionnelle » se concentre sur la **détection de « patterns »**, la **classification** et la **prédiction de données**
- L'IA « générative » permet de **créer de nouvelles données à partir de données exemples**

Technos

Deep Learning – Types de modèles d'IA générative



Applications



Image à partir
d'image



Texte à partir
de texte



Audio à partir
d'audio

Images à
partir de texte

Agent
conversationnel

– Musique à partir
de texte
– Text-to-speech
– Speech-to-text

Pour commencer, L'IA générative ?

Quelles technologies et quelles applications ?

Notions clés

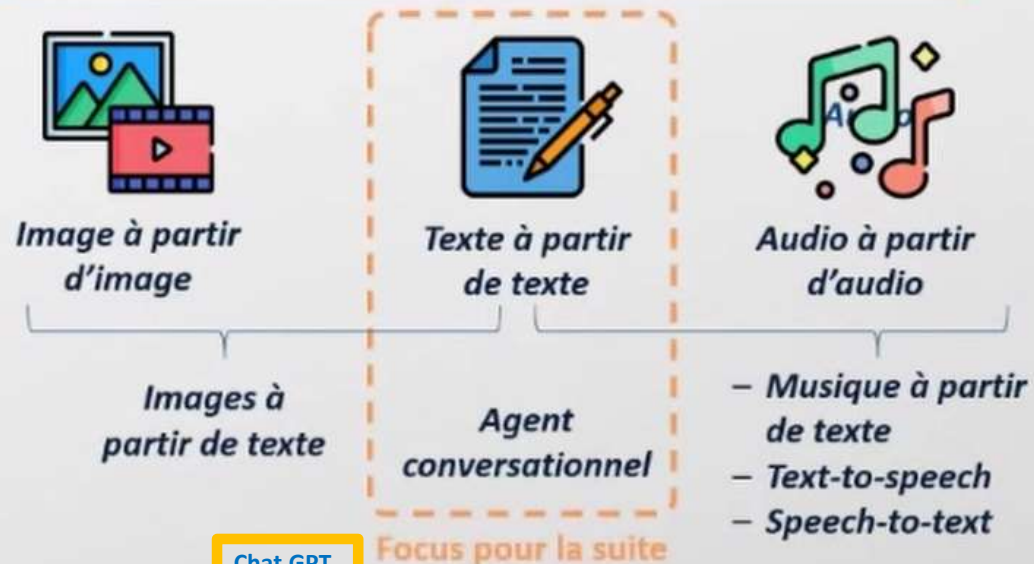
- L'IA « traditionnelle » se concentre sur la **détection de « patterns »**, la **classification** et la **prédiction de données**
- L'IA « générative » permet de **créer de nouvelles données à partir de données exemples**

Technos

Deep Learning – Types de modèles d'IA générative



Applications



Chat GPT

Focus pour la suite

ChatGPT

9

Qu'est-ce qu'un agent conversationnel ?

Notions clés

- ChatGPT (OpenAI) , Bing (Microsoft) et Bard (Google) sont des agents conversationnels (*liste non exhaustive*)
- Ils sont basés sur les **Large Language Model (LLM)**, une classe de technologie de l'IA générative

- Les **Large Language Models** sont des modèles d'intelligence artificielle
- Entraînés sur de vastes ensembles de données textuelles
- Pour traiter, comprendre et générer du langage naturel.
- Ils reposent généralement sur des architectures avancées de réseaux de neurones, comme les **Transformers**, introduits par Google dans l'article "*Attention Is All You Need*" (2017).

ChatGPT

9

Qu'est-ce qu'un agent conversationnel ?

Notions clés

- ChatGPT (OpenAI) , Bing (Microsoft) et Bard (Google) sont des agents conversationnels (*liste non exhaustive*)
- Ils sont basés sur les Large Language Model (LLM), une classe de technologie de l'IA générative

Caractéristiques principales :

1. Taille énorme :

1. Les LLMs contiennent des milliards de paramètres.
2. Par exemple : GPT-4, PaLM de Google, LLaMA de Meta.
3. Plus de paramètres permettent une meilleure capacité à capturer les subtilités du langage.

2. Formation sur de grands ensembles de données :

1. Données issues de livres, articles, sites web, forums, code source, etc.
2. Diversité et volume des données améliorent leur compréhension générale.

3. Apprentissage contextuel :

1. Comprennent les relations entre les mots grâce au mécanisme d'attention.
2. Capables de produire des réponses cohérentes en analysant le contexte global.

4. Polyvalence :

1. Applications variées : traduction, rédaction, chatbots, résumé, analyse de sentiments, génération de code, etc.

ChatGPT

9

Qu'est-ce qu'un agent conversationnel ?

Notions clés

- ChatGPT (OpenAI) , Bing (Microsoft) et Bard (Google) sont des agents conversationnels (*liste non exhaustive*)
- Ils sont basés sur les Large Language Model (LLM), une classe de technologie de l'IA générative

Comment fonctionnent-ils ?

1. Architecture Transformer :

1. Basée sur l'attention, permettant au modèle de se concentrer sur des parties spécifiques d'un texte.
2. Encodeur-décodeur ou simplement un décodeur (comme GPT).

2. Entraînement :

1. **Objectif principal** : Modéliser les probabilités conditionnelles des mots.
2. Techniques :
 1. **Apprentissage supervisé** : Basé sur des données annotées.
 2. **Apprentissage non supervisé** : Sur de grandes quantités de texte brut.
 3. **Fine-tuning** : Adaptation pour des cas d'usage spécifiques.

3. Capacités émergentes :

1. Résolution de problèmes complexes sans apprentissage explicite.
2. Exemples : Compréhension des langues rares, rédaction de code.

ChatGPT

9

Qu'est-ce qu'un agent conversationnel ?

Notions clés

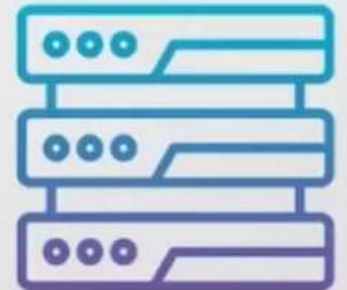
- ChatGPT (OpenAI) , Bing (Microsoft) et Bard (Google) sont des agents conversationnels (*liste non exhaustive*)
- Ils sont basés sur les **Large Language Model (LLM)**, une classe de technologie de l'IA générative

Utilisateur



ChatGPT

basé sur LLM, e.g gpt-3



ChatGPT

9

Qu'est-ce qu'un agent conversationnel ?

Notions clés

- ChatGPT (OpenAI) , Bing (Microsoft) et Bard (Google) sont des agents conversationnels (*liste non exhaustive*)
- Ils sont basés sur les **Large Language Model (LLM)**, une classe de technologie de l'IA générative

Utilisateur



2 Méthodes :

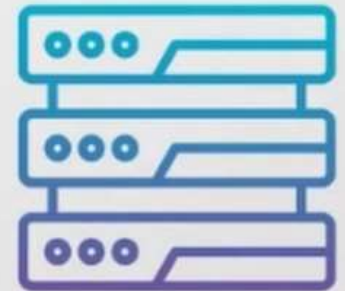
- Via le site web
- Via l'API

Envoie une requête (« prompt ») en langage naturel au modèle



ChatGPT

basé sur LLM, e.g gpt-3



ChatGPT

9

Qu'est-ce qu'un agent conversationnel ?

Notions clés

- ChatGPT (OpenAI) , Bing (Microsoft) et Bard (Google) sont des agents conversationnels (*liste non exhaustive*)
- Ils sont basés sur les **Large Language Model (LLM)**, une classe de technologie de l'IA générative

Utilisateur

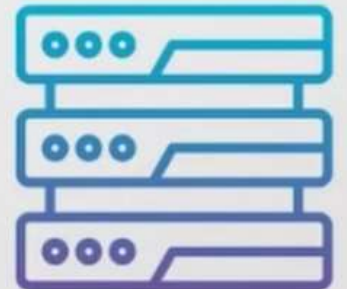


2 Méthodes :

- Via le site web
- Via l'API

ChatGPT

basé sur LLM, e.g gpt-3



Envoie une requête (« prompt ») en langage naturel au modèle



Retourne une réponse en langage naturel répondant à la requête



Les LLM

Qu'est-ce qu'un Large Language Model ?

10

Notions clés

- Un LLM est une architecture de modèle de **Deep Learning** basé sur un composant appelé *Transformer*
- Ils génèrent du texte en **prédisant les mots suivants** les plus probables d'un texte donné (**Next Token prediction**)
- Les LLM nécessitent **d'énormes volumes de textes** pour être entraînés

Les LLM

10

Qu'est-ce qu'un Large Language Model ?

Notions clés

- Un LLM est une architecture de modèle de **Deep Learning** basé sur un composant appelé **Transformer**
- Ils génèrent du texte en **prédisant les mots suivants** les plus probables d'un texte donné (**Next Token prediction**)
- Les LLM nécessitent **d'énormes volumes de textes** pour être entraînés

Le principe du « Next Token Prediction »

Je suis allé voir Emilie. C'est une...

| | | |
|------------------|----|------------|
| <i>amie</i> | -- | <i>35%</i> |
| <i>chanteuse</i> | -- | <i>15%</i> |
| <i>très</i> | -- | <i>9%</i> |
| <i>...</i> | | |

Input (texte envoyé au modèle) → **Output** (texte retourné par le modèle)

Les LLM

10

Qu'est-ce qu'un Large Language Model ?

Notions clés

- Un LLM est une architecture de modèle de **Deep Learning** basé sur un composant appelé **Transformer**
- Ils génèrent du texte en **prédisant les mots suivants** les plus probables d'un texte donné (**Next Token prediction**)
- Les LLM nécessitent **d'énormes volumes de textes** pour être entraînés

Le principe du « Next Token Prediction »

Je suis allé voir Emilie. C'est une...

| | | |
|------------------|----|-----|
| <i>amie</i> | -- | 35% |
| <i>chanteuse</i> | -- | 15% |
| <i>très</i> | -- | 9% |
| ... | | |

Input (texte envoyé au modèle) → **Output** (texte retourné par le modèle)

Construction d'un LLM et de ChatGPT

Comment obtient-on un agent conversationnel ?

Construction d'un LLM et de ChatGPT

11

Comment obtient-on un agent conversationnel ?

Larges corpus textuels (570 GB)



**Common
Crawl (filtré)**
410B de tokens



Webtext2
19B de tokens



Livres numérisés
67B de tokens



Wikipedia
3B de tokens

Construction d'un LLM et de ChatGPT

11

Comment obtient-on un agent conversationnel ?

Entraînement d'un
LLM « Next Token
Prediction »



Larges corpus textuels (570 GB)



Common
Crawl (filtré)
410B de tokens



Webtext2
19B de tokens



Livres numérisés
67B de tokens

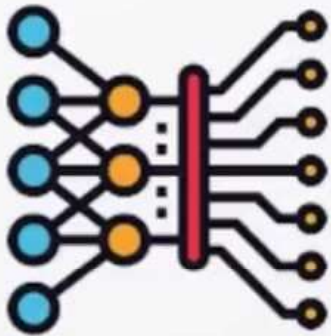


Wikipedia
3B de tokens

Construction d'un LLM et de ChatGPT

Comment obtient-on un agent conversationnel ?

LLM GPT-3 (par exemple)



Deep Neural Network
(175B de paramètres)

Basé sur les Transformers

Entraînement d'un LLM « Next Token Prediction »



Larges corpus textuels (570 GB)



Common Crawl (filtré)
410B de tokens



Webtext2
19B de tokens



Livres numérisés
67B de tokens

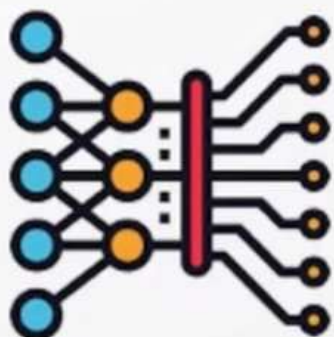


Wikipedia
3B de tokens

Construction d'un LLM et de ChatGPT

Comment obtient-on un agent conversationnel ?

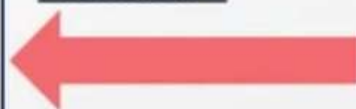
LLM GPT-3 (par exemple)



Deep Neural Network
(175B de paramètres)

Basé sur les Transformers

Entraînement d'un LLM « Next Token Prediction »



Larges corpus textuels (570 GB)



Common Crawl (filtré)
410B de tokens



Webtext2
19B de tokens



Livres numérisés
67B de tokens



Wikipedia
3B de tokens

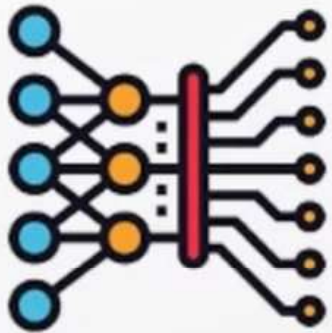


Entraînement spécifique à la conversation

Construction d'un LLM et de ChatGPT

Comment obtient-on un agent conversationnel ?

LLM GPT-3 (par exemple)



Deep Neural Network
(175B de paramètres)

Basé sur les Transformers

Entraînement d'un LLM « Next Token Prediction »

Larges corpus textuels (570 GB)



Common Crawl (filtré)
410B de tokens



Webtext2
19B de tokens



Livres numérisés
67B de tokens



Wikipedia
3B de tokens

3 étapes d'entraînements

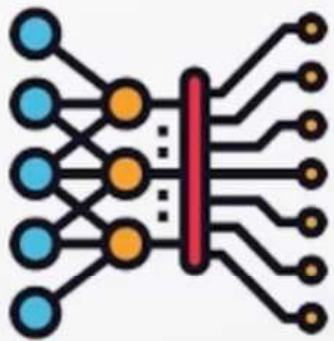
Entraînement spécifique à la conversation



Construction d'un LLM et de ChatGPT

Comment obtient-on un agent conversationnel ?

LLM GPT-3 (par exemple)



Deep Neural Network
(175B de paramètres)

Basé sur les Transformers

Entraînement d'un LLM « Next Token Prediction »

Larges corpus textuels (570 GB)



Common Crawl (filtré)
410B de tokens



Webtext2
19B de tokens



Livres numérisés
67B de tokens



Wikipedia
3B de tokens

3 étapes d'entraînements

Entraînement spécifique à la conversation

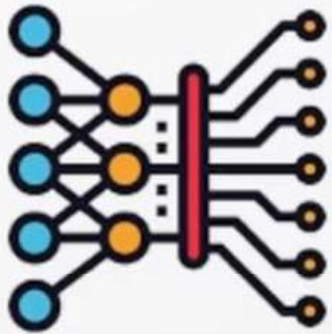
1

Entraînement basé sur des exemples de conversations « idéales »

Construction d'un LLM et de ChatGPT

Comment obtient-on un agent conversationnel ?

LLM GPT-3 (par exemple)



Deep Neural Network
(175B de paramètres)

Basé sur les Transformers

Entraînement d'un LLM « Next Token Prediction »

Larges corpus textuels (570 GB)



Common Crawl (filtré)
410B de tokens



Webtext2
19B de tokens



Livres numérisés
67B de tokens



Wikipedia
3B de tokens

3 étapes d'entraînements

Entraînement spécifique à la conversation

1

Entraînement basé sur des exemples de conversations « idéales »

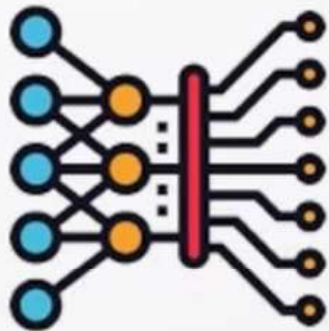
2

Entraînement basé sur le classement préférentiel des réponses par un humain

Construction d'un LLM et de ChatGPT

Comment obtient-on un agent conversationnel ?

LLM GPT-3 (par exemple)



Deep Neural Network
(175B de paramètres)

Basé sur les Transformers

Entraînement d'un LLM « Next Token Prediction »

Larges corpus textuels (570 GB)



Common Crawl (filtré)
410B de tokens



Webtext2
19B de tokens



Livres numérisés
67B de tokens



Wikipedia
3B de tokens

3 étapes d'entraînements

Entraînement spécifique à la conversation

1

Entraînement basé sur des exemples de conversations « idéales »

2

Entraînement basé sur le classement préférentiel des réponses par un humain

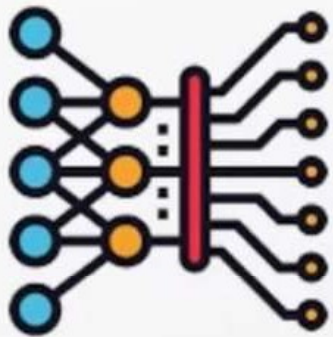
3

Entraînement basé des réponses préférentielles automatisées

Construction d'un LLM et de ChatGPT

Comment obtient-on un agent conversationnel ?

LLM GPT-3 (par exemple)



Deep Neural Network
(175B de paramètres)

Basé sur les Transformers

Entraînement d'un LLM « Next Token Prediction »

Larges corpus textuels (570 GB)



Common Crawl (filtré)
410B de tokens



Webtext2
19B de tokens



Livres numérisés
67B de tokens



Wikipedia
3B de tokens

3 étapes d'entraînements

Entraînement spécifique à la conversation

1

Entraînement basé sur des exemples de conversations « idéales »

2

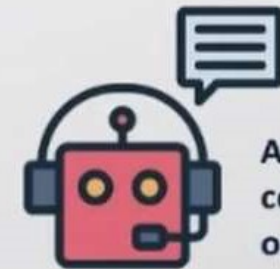
Entraînement basé sur le classement préférentiel des réponses par un humain

3

Entraînement basé des réponses préférentielles automatisées

Mise en prod.

ChatGPT



Agent conversationnel opérationnel