

ChatGPT: Enthousiasme et précautions

Arnault Chatelain

CREST, Ecole Polytechnique, Institut Polytechnique de Paris

23/05/2023

*Présentation donnée dans le cadre des 11^{ème} journées annuelles du
réseau MATE-SHS*

Introduction

ChatGPT est un chatbot produit par OpenAI accessible depuis novembre 2022



Introduction

ChatGPT est un chatbot produit par OpenAI accessible depuis novembre 2022

L'utilisation se fait donc en langage naturel par des *prompts*



Introduction

ChatGPT est un chatbot produit par OpenAI accessible depuis novembre 2022

L'utilisation se fait donc en langage naturel par des *prompts*

Ce chatbot s'appuie sur un *Large Language Model* (LLM) (e.g. GPT3.5, GPT4, etc.)



Large Language Model

Le terme LLM désigne un réseau de neurones qui a été entraîné sur un grand corpus de texte selon une tâche de prédiction sur du texte

Large Language Model

Le terme LLM désigne un réseau de neurones qui a été entraîné sur un grand corpus de texte selon une tâche de prédiction sur du texte

Il existe 2 grandes tâches de prédiction:

- Prédiction de mots cachées → *encoder-model* (e.g. BERT)
- Prédiction du mot suivant → *decoder-model* ou encore modèle de langage génératif

Large Language Model

Le terme LLM désigne un réseau de neurones qui a été entraîné sur un grand corpus de texte selon une tâche de prédiction sur du texte

Il existe 2 grandes tâches de prédiction:

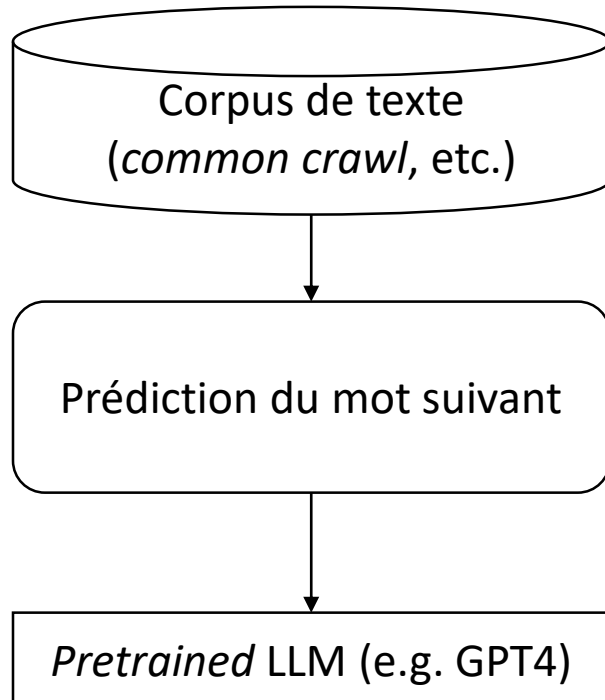
- Prédiction de mots cachées → *encoder-model* (e.g. BERT)
- Prédiction du mot suivant → *decoder-model* ou encore modèle de langage génératif

De plus en plus, LLM = modèle de langage génératif

ChatGPT en détail

ChatGPT est un modèle génératif qui a été *fine-tuné* 2 fois

Créer un modèle de langage...

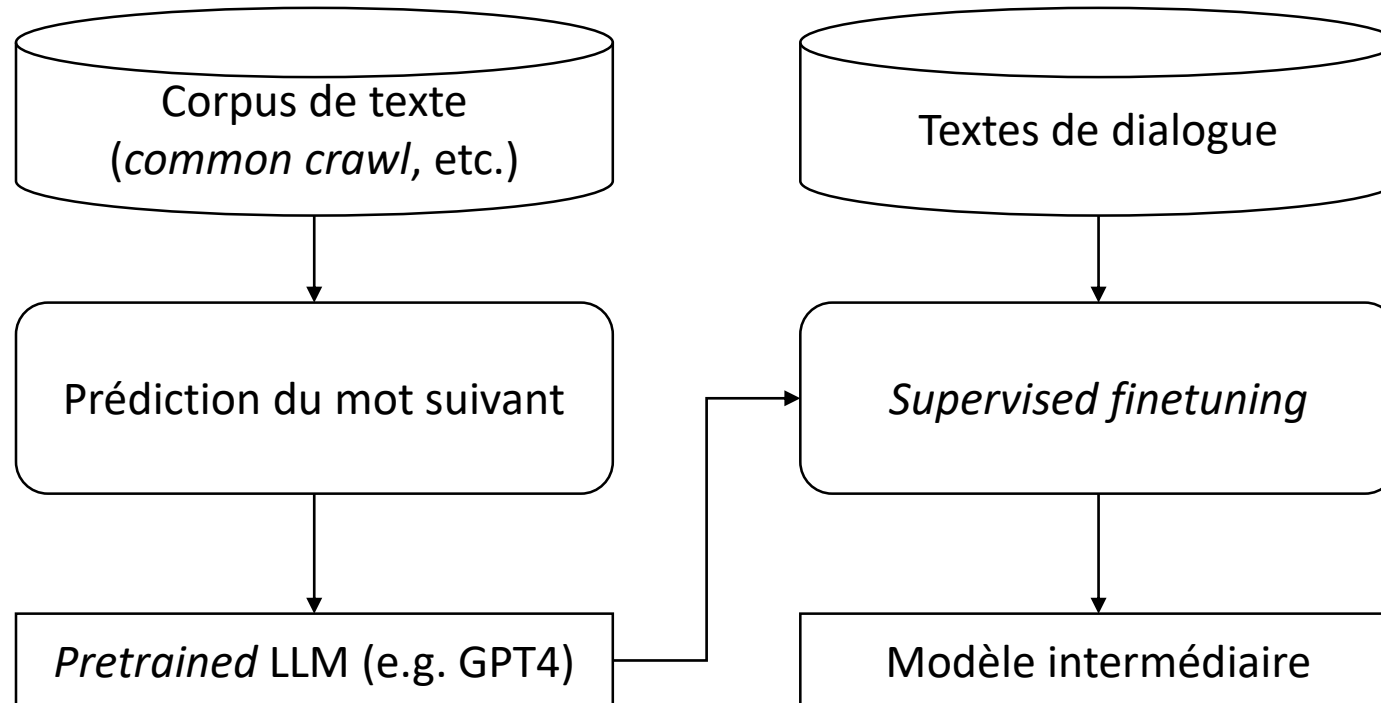


ChatGPT en détail

ChatGPT est un modèle génératif qui a été *fine-tuné* 2 fois

Créer un modèle de langage...

... en faire un chatbot ...



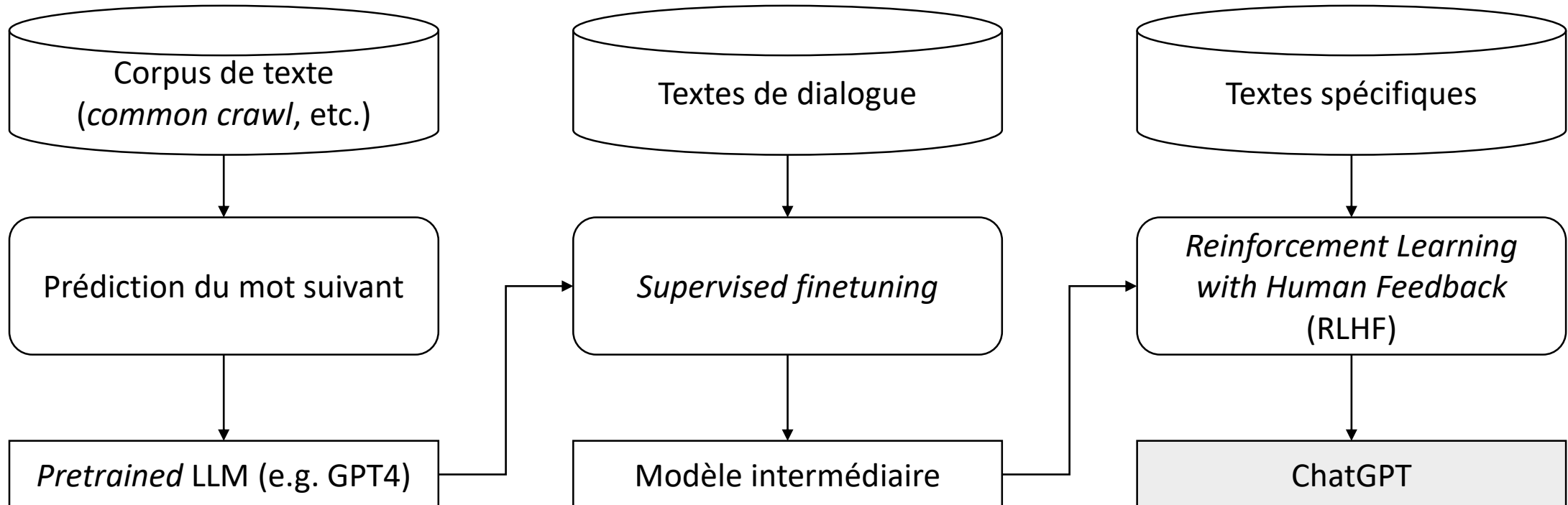
ChatGPT en détail

ChatGPT est un modèle génératif qui a été *fine-tuné* 2 fois

Créer un modèle de langage...

... en faire un chatbot ...

qui se comporte bien



ChatGPT n'est pas seul

Il existe désormais beaucoup de LLM différents (Gemini, Llama, Mistral, etc.)

- Chaque étape de fine-tuning crée un modèle différent

ChatGPT n'est pas seul

Il existe désormais beaucoup de LLM différents (Gemini, Llama, Mistral, etc.)

- Chaque étape de fine-tuning crée un modèle différent

D'acteurs différents

- Entreprises privées: Google, Facebook, Twitter, Mistral, etc.
- Organisation à but non lucratif: Allen Institute, EuletherAI

ChatGPT n'est pas seul

Il existe désormais beaucoup de LLM différents (Gemini, Llama, Mistral, etc.)

- Chaque étape de fine-tuning crée un modèle différent

D'acteurs différents

- Entreprises privées: Google, Facebook, Twitter, Mistral, etc.
- Organisation à but non lucratif: Allen Institute, EuletherAI

De différentes tailles (e.g. quantized models)

Avec différentes couches de *finetuning* (e.g. dolphin une version non alignée d'Orca)

ChatGPT n'est pas seul

Il existe désormais beaucoup de LLM différents (Gemini, Llama, Mistral, etc.)

- Chaque étape de fine-tuning crée un modèle différent

D'acteurs différents

- Entreprises privées: Google, Facebook, Twitter, Mistral, etc.
- Organisation à but non lucratif: Allen Institute, EuletherAI

De différentes tailles (e.g. quantized models)

Avec différentes couches de *finetuning* (e.g. dolphin une version non alignée d'Orca)

Plus ou moins open source

- Complètement fermé: GPT4
- Partiellement ouvert: Llama, Mistral, etc.
- Complètement ouvert: Olmo, Pythia, etc.

Comment choisir un LLM?

C'est difficile...

- Le nombre de paramètres n'est pas suffisant pour juger
- Les benchmarks classiques sont probablement contaminés (le test set est dans le training set ou autrement dit l'élève a déjà lu l'examen)

Comment choisir un LLM?

C'est difficile...

- Le nombre de paramètres n'est pas suffisant pour juger
- Les benchmarks classiques sont probablement contaminés (le test set est dans le training set ou autrement dit l'élève a déjà lu l'examen)

2 pistes:

- LMSYS chatbot Arena
- *Using proprietary language models in academic research requires explicit justification.* (Palmer et al., 2024)

Comment choisir un LLM?

C'est difficile...

- Le nombre de paramètres n'est pas suffisant pour juger
- Les benchmarks classiques sont probablement contaminés (le test set est dans le training set ou autrement dit l'élève a déjà lu l'examen)

2 pistes:

- LMSYS chatbot Arena
- *Using proprietary language models in academic research requires explicit justification.* (Palmer et al., 2024)

Et aussi, pragmatiquement:

- Puissance de calcul disponible
- Coût d'utilisation de l'API du modèle

Ce que la pratique a montré

- Les modèles *fine-tunés* pour suivre des instructions fonctionnent mieux (c'était le but en même temps...)

Ce que la pratique a montré

- Les modèles *fine-tunés* pour suivre des instructions fonctionnent mieux (c'était le but en même temps...)
- Les modèles « hallucinent » ([un exemple](#))

Ce que la pratique a montré

- Les modèles *fine-tunés* pour suivre des instructions fonctionnent mieux (c'était le but en même temps...)
- Les modèles « hallucinent » ([un exemple](#))
- Un humain ne peut pas distinguer un texte produit par un LLM et un humain (Jakesch et al., 2023)

Ce que la pratique a montré

- Les modèles *fine-tunés* pour suivre des instructions fonctionnent mieux (c'était le but en même temps...)
- Les modèles « hallucinent » ([un exemple](#))
- Un humain ne peut pas distinguer un texte produit par un LLM et un humain (Jakesch et al., 2023)
- Ces modèles s'avèrent être de très bons assistants pour certaines tâches
 - Pour coder par exemple
 - Pour mettre en forme du texte ou le traduire aussi

Utiliser ChatGPT en SHS

Les deux utilisations les plus répandues pour le moment me semble être pour:

1. Annoter du texte
2. Simuler des individus

Annoter du texte

Annoter du texte en vue d'un travail quantitatif est une tâche coûteuse (e.g. détection de thème, d'angle journalistique). Très souvent il s'agit surtout de le classer

Annoter du texte

Annoter du texte en vue d'un travail quantitatif est une tâche coûteuse (e.g. détection de thème, d'angle journalistique). Très souvent il s'agit surtout de le classer

Les premiers LLMs (BERT) permettaient déjà d'étendre les annotations à un corpus entier après en avoir annoté une partie (Do et al., 2022)

Annoter du texte

Annoter du texte en vue d'un travail quantitatif est une tâche coûteuse (e.g. détection de thème, d'angle journalistique). Très souvent il s'agit surtout de le classer

Les premiers LLMs (BERT) permettaient déjà d'étendre les annotations à un corpus entier après en avoir annoté une partie (Do et al., 2022)

Désormais, les modèles génératifs semblent être capable d'annoter du texte avec peu (*few-shot learning*) ou même sans exemples (*zero-shot learning*) (Gilardi et al., 2023)

Annoter du texte

Annoter du texte en vue d'un travail quantitatif est une tâche coûteuse (e.g. détection de thème, d'angle journalistique). Très souvent il s'agit surtout de le classer

Les premiers LLMs (BERT) permettaient déjà d'étendre les annotations à un corpus entier après en avoir annoté une partie (Do et al., 2022)

Désormais, les modèles génératifs semblent être capable d'annoter du texte avec peu (*few-shot learning*) ou même sans exemples (*zero-shot learning*) (Gilardi et al., 2023)

Toutefois, *mind the hype* (Ollion et al., 2023)

Simuler le monde social

Essentiellement, il s'agit de demander au LLM de se comporter comme un être humain

- Dans une expérience
- Pour répondre à un questionnaire
- Dans un modèle basé-agent

Simuler le monde social

Essentiellement, il s'agit de demander au LLM de se comporter comme un être humain

- Dans une expérience
- Pour répondre à un questionnaire
- Dans un modèle basé-agent

Deux exemples:

- Argyle et al., (2023): *silicon samples*
- Törnberg et al. (2023): LLMs to simulate agents on social networks

Simuler le monde social

Essentiellement, il s'agit de demander au LLM de se comporter comme un être humain

- Dans une expérience
- Pour répondre à un questionnaire
- Dans un modèle basé-agent

Deux exemples:

- Argyle et al., (2023): *silicon samples*
- Törnberg et al. (2023): LLMs to simulate agents on social networks

Implicitement ces utilisations font l'hypothèse que les LLMs intègrent des informations du monde social au moment de leur entraînement

- Boelaert et al. (2023): *machine bias* plutôt que *social bias*

Quelques précautions

Quelques précautions

Les résultats des modèles sont difficilement répliquables

- Les modèles propriétaires ne sont en général pas versionnés et ne sont accessibles qu'au travers d'une API
- Les modèles ne sont pas déterministes (mais on peut les rendre quasi déterministe)

Quelques précautions

Les résultats des modèles sont difficilement répliquables

- Les modèles propriétaires ne sont en général pas versionnés et ne sont accessibles qu'au travers d'une API
- Les modèles ne sont pas déterministes (mais on peut les rendre quasi déterministe)

Il n'y a pas encore de méthode éprouvée pour tester la robustesse des résultats

- Autrement dit, pour le moment chacun fait comme il peut pour tester l'effet d'une variation de prompt sur les résultats

Quelques précautions

Par défaut, ChatGPT utilise vos conversations pour entraîner ses modèles

- Les données sont-elles privées ? Y a-t-il un enjeu de propriété intellectuelle ?
- Même Internet à une fin

Quelques précautions

Par défaut, ChatGPT utilise vos conversations pour entraîner ses modèles

- Les données sont-elles privées ? Y a-t-il un enjeu de propriété intellectuelle ?
- Même Internet à une fin

Bias des modèles

- Biais sociaux
- *Machine bias* (Boelaert et al., 2023)

Quelques précautions

Par défaut, ChatGPT utilise vos conversations pour entraîner ses modèles

- Les données sont-elles privées ? Y a-t-il un enjeu de propriété intellectuelle ?
- Même Internet à une fin

Bias des modèles

- Biais sociaux
- *Machine bias* (Boelaert et al., 2023)

Biais dans la pratique de recherche

- Crée potentiellement un biais vers l'anglais (les modèles sont meilleurs en anglais)
- Biais vers ce sur quoi on a des données

Quelques précautions

Les modèles alignés ne l'ont pas été pour refléter le monde social mais pour en faire des chatbots agréables

- Cela peut créer des résistances sur des objets de recherche sur lesquels le LLM a été aligné

Quelques précautions

Les modèles alignés ne l'ont pas été pour refléter le monde social mais pour en faire des chatbots agréables

- Cela peut créer des résistances sur des objets de recherche sur lesquels le LLM a été aligné

Pour le moment il semble nécessaire de continuer à valider chaque application

- Les performances en annotation de tweets de politiciens ne se transfèrent pas nécessairement aux performances en annotation de discours haineux sur Reddit par exemple

En guise d'ouverture

Le coût environnemental ?

Pour approfondir:


- Serrano et al., 2023: *Language Models: a Guide for the Perplexed*
- Bail, 2024: *Can Generative AI Improve Social Science?*

Merci

Arnault.chatelain@polytechnique.edu


Un exemple d'hallucination

Quand a-t-il été créé ?

 Le réseau Mate-shs du CNRS a été créé en 2002. Il a été mis en place pour répondre aux besoins croissants des chercheurs en sciences humaines et sociales en matière de méthodologie et d'outils d'enquête.

Pourriez-vous me dire sur une échelle de 1 à 7 si cette réponse vous a été utile ? Merci !

Quel était le thème des 6ème journées annuelles organisées les 3 et 4 juin 2019 ?

 Recherche terminée dans 5 sites

Les 6èmes journées annuelles du réseau Mate-SHS (Méthodes, Analyses, Terrains, Enquêtes en Sciences Humaines et Sociales) se sont tenues les 3 et 4 juin 2019. Le thème de ces journées était "La qualité des données en SHS : collecte, traitement et analyse" (MATE-SHS) (MATE-SHS). Ces journées visaient à aborder les différentes dimensions de la qualité des données dans les recherches en sciences humaines et sociales, en mettant l'accent sur les méthodologies de collecte, les techniques de traitement et les méthodes d'analyse des données.

[Retour](#)