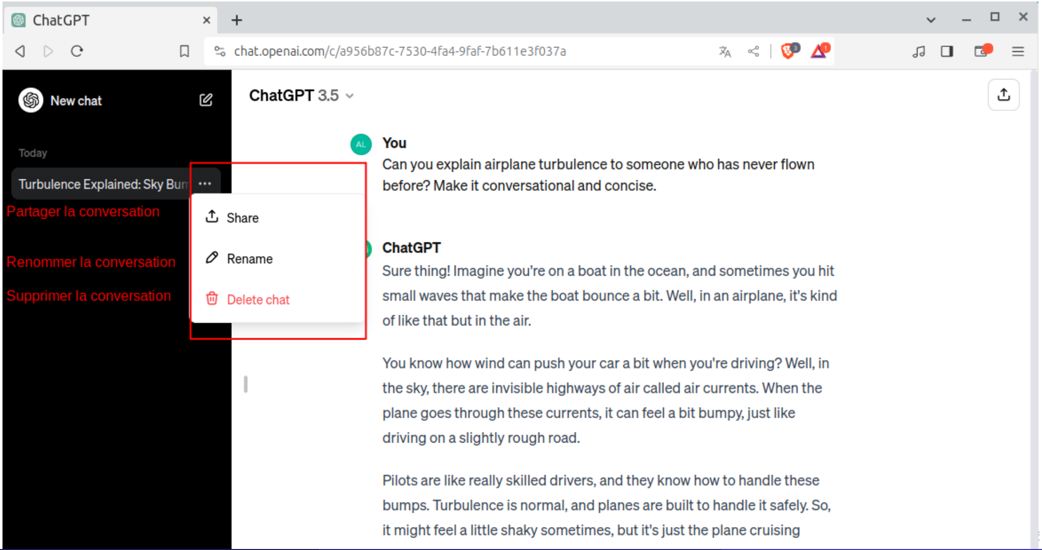
From GPT to ChatGPT: General Principles and Issues Raised by OpenAI's Large Language Models

Didier Schwab

12th January 2024

Prompting in chatGPT



De GPT à ChatGPT : Principes généraux et questions soulevées par les grands modèles de langues d'OpenAI

Votre Nom

Votre Institution

July 10, 2023

Introduction

- Les modèles de langues d'OpenAI ont révolutionné le traitement automatique du langage naturel.
- GPT (Generative Pre-trained Transformer) et ChatGPT sont deux exemples de ces grands modèles.
- Dans cette présentation, nous allons explorer les principes généraux qui sous-tendent ces modèles et les questions soulevées par leur utilisation.

GPT : Principes généraux

- GPT est basé sur l'architecture du Transformer, un modèle de traitement de séquence utilisant des mécanismes d'attention.
- Il est pré-entraîné sur de vastes quantités de données textuelles non supervisées provenant du Web.
- GPT utilise un objectif d'apprentissage par prévision de mots pour capturer la structure et la sémantique des phrases.
- Le modèle peut ensuite être fine-tuné sur des tâches spécifiques, telles que la traduction automatique ou la génération de texte.

ChatGPT : Une extension de GPT

- ChatGPT est une version adaptée de GPT conçue pour la génération de réponses conversationnelles.
- Il est formé à l'aide d'une approche d'apprentissage par renforcement en utilisant des données provenant d'interactions humaines.
- L'objectif est de générer des réponses pertinentes et naturelles aux questions des utilisateurs.
- ChatGPT est plus interactif et peut être utilisé pour des applications de chatbot, d'assistance virtuelle, etc.

Avantages des grands modèles de langues

- Les grands modèles de langues, tels que GPT et ChatGPT, ont plusieurs avantages :
  - Capacité à capturer des relations complexes entre les mots et les phrases.
  - Adaptabilité à différentes tâches grâce au fine-tuning.
  - Amélioration continue avec l'ajout de nouvelles données et itérations d'entraînement.
  - Génération de texte fluide et cohérente dans de nombreuses situations.

Questions soulevées par les grands modèles de langues

- Bien que les grands modèles de langues offrent de nombreux avantages, ils soulèvent également des questions importantes :
  - Biais et éthique : Comment éviter les biais indésirables dans la génération de texte ?
  - Contrôle et responsabilité : Comment garantir que le modèle génère du contenu approprié et respecte les normes éthiques ?
  - Sécurité : Comment prévenir les utilisations malveillantes des modèles de langues, telles que la création de désinformation ?
  - Confidentialité : Comment gérer les problèmes de confidentialité des données utilisées pour l'entraînement des modèles ?

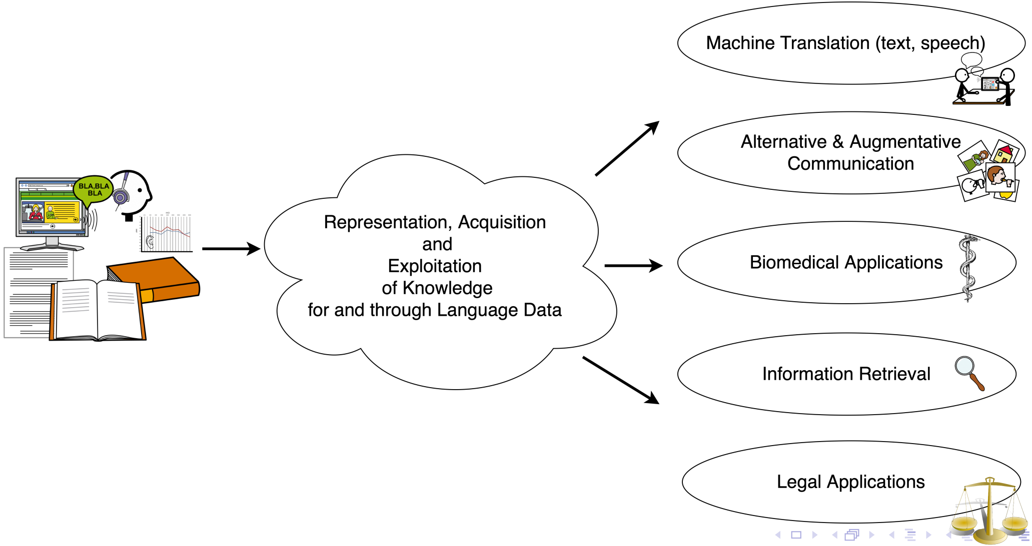
Conclusion

- Les grands modèles de langues d'OpenAI, tels que GPT et ChatGPT, sont des avancées majeures dans le traitement automatique du langage naturel.
- Ils offrent des capacités de génération de texte puissantes, mais soulèvent également des questions éthiques et pratiques.
- Il est essentiel de trouver des solutions pour résoudre ces questions et utiliser ces modèles de manière responsable et éthique.

Questions ?

Merci pour votre attention ! Des questions ?

Research Topics

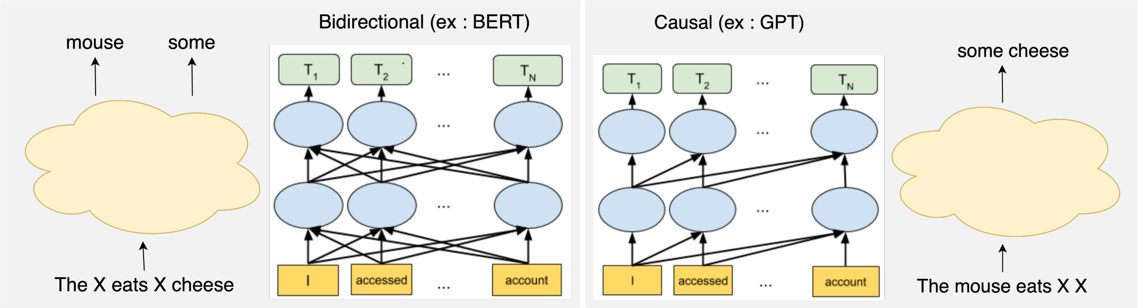


Training Data (GPT-3)

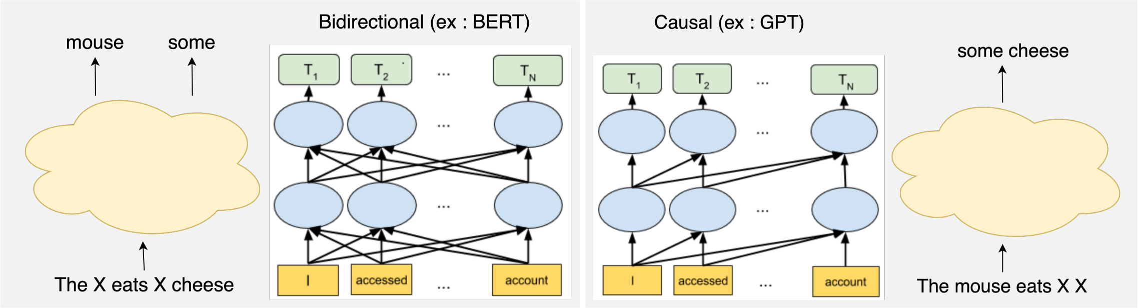
Dataset	Size (tokens)	% in the training dataset
Common Crawl (filtered)	410 billiards	60%
WebText2	19 billions	22%
Books1	12 billions	8%
Books2	55 billions	8%
Wikipedia	3 billions	3%

Table: Datasets in GPT-3 training – selfsupervision [Brown et al., 2020]

Self Supervised Learning in LLM : Bidirectional vs Causative Objectives

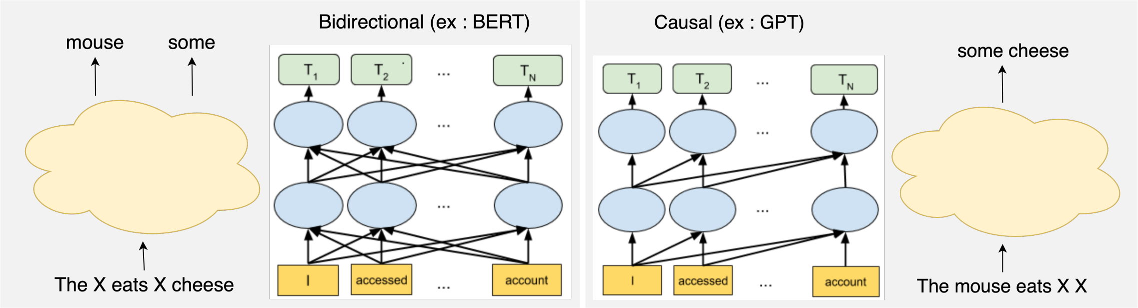


Self Supervised Learning in LLM : Bidirectional vs Causative Objectives



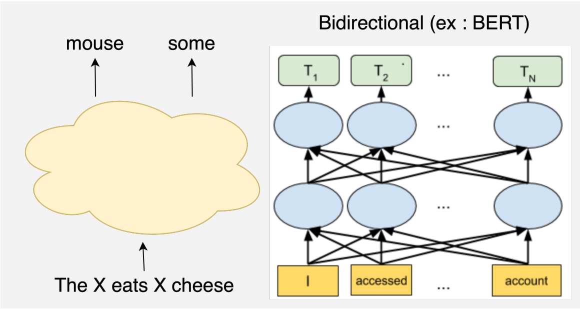
- Causal models: designed to generate coherent texts from a prompt

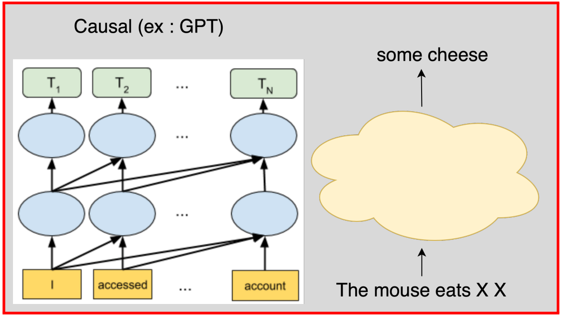
Self Supervised Learning in LLM : Bidirectional vs Causative Objectives



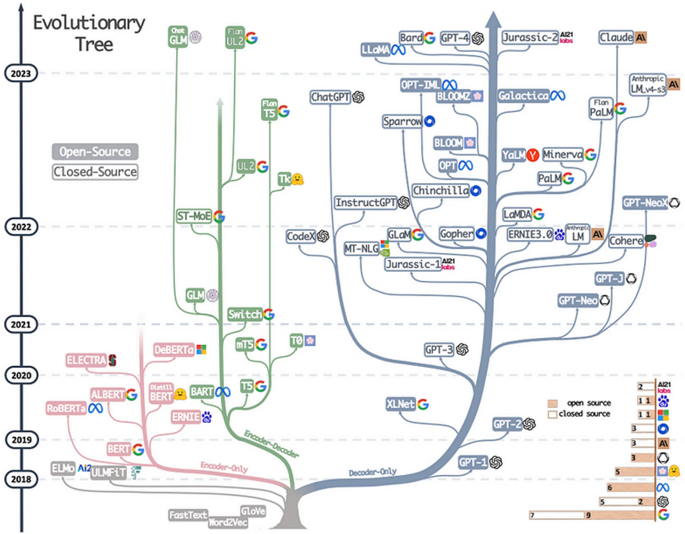
- Causal models: designed to generate coherent texts from a prompt
- Bidirectional models: designed to build good representations of words, sentences or documents

Self Supervised Learning in LLM : Bidirectional vs Causative Objectives

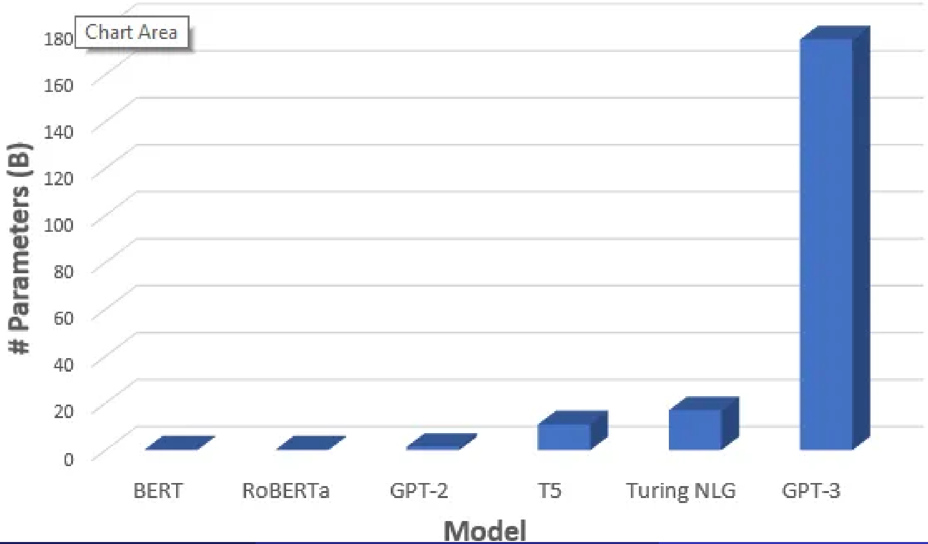




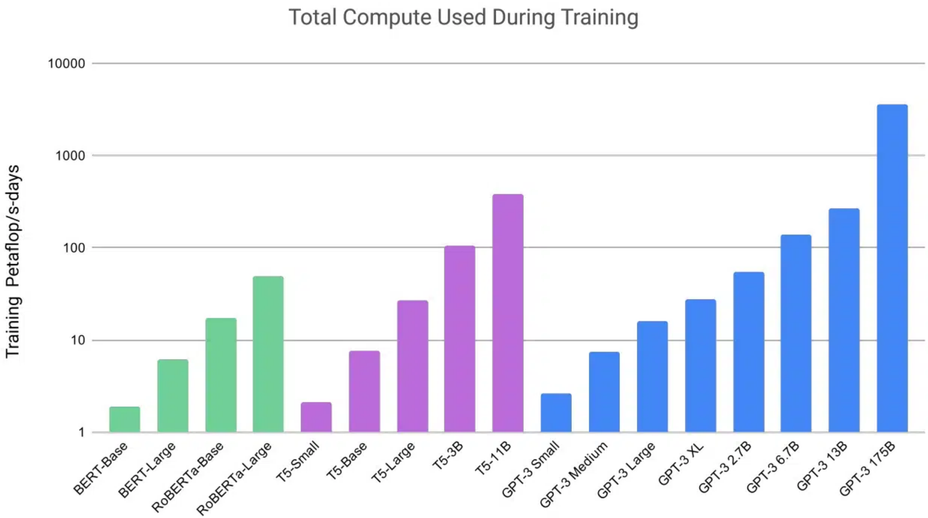
Panorama of Large Language Models



Number of parameters



computation



Output of a GPT

- Once

Output of a GPT

- Once
- Once upon

Output of a GPT

- Once
- Once upon
- Once upon a time

Output of a GPT

- Once
- Once upon
- Once upon a time
- Once upon a time, long

Output of a GPT

- Once
- Once upon
- Once upon a time
- Once upon a time, long
- Once upon a time, long long

Output of a GPT

- Once
- Once upon
- Once upon a time
- Once upon a time, long
- Once upon a time, long long
- Once upon a time, long long ago

Output of a GPT

- Once
- Once upon
- Once upon a time
- Once upon a time, long
- Once upon a time, long long
- Once upon a time, long long ago . . .

Output of a GPT

- Once
- Once upon
- Once upon a time
- Once upon a time, long
- Once upon a time, long long
- Once upon a time, long long ago . . .

A GPT (including chatGPT) is designed to predict the next word from a given text sequence (and that's all).

Pre-training of GPTs

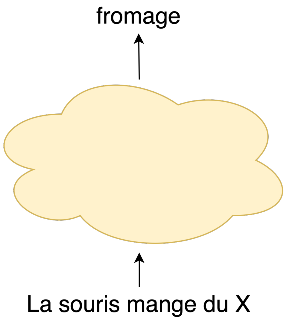
Pre-training of GPTs (including GPT3) includes:

- Language modeling (self-supervised): the model attempts to predict the next word in a given word sequence.

Pre-training of GPTs

Pre-training of GPTs (including GPT3) includes:

- Language modeling (self-supervised): the model attempts to predict the next word in a given word sequence.



Pre-training of GPTs

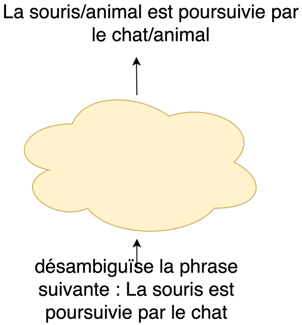
Pre-training of GPTs (including GPT3) includes:

- Language modeling (self-supervised): the model attempts to predict the next word in a given word sequence.
- Solving language processing tasks such as text classification, automatic summarization, question answering, language translation (> 100)

Pre-training of GPTs

Pre-training of GPTs (including GPT3) includes:

- Language modeling (self-supervised): the model attempts to predict the next word in a given word sequence.
- Solving language processing tasks such as text classification, automatic summarization, question answering, language translation (> 100)



Pre-training of GPTs

Pre-training of GPTs (including GPT3) includes:

- Language modeling (self-supervised): the model attempts to predict the next word in a given word sequence.
- Solving language processing tasks such as text classification, automatic summarization, question answering, language translation (> 100)

Operational consequence: GPTs learn some new tasks with few examples (*Few-Shot Learners*)

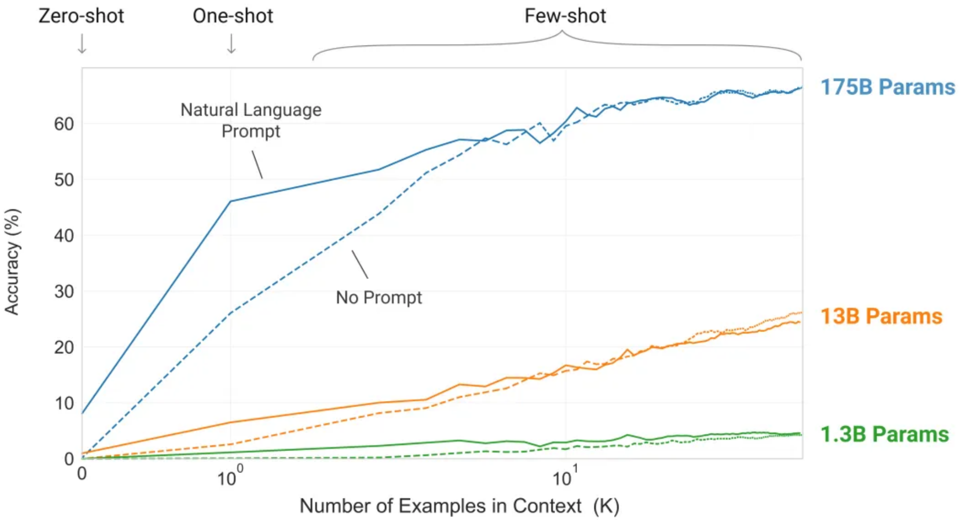
Pre-training of GPTs

Pre-training of GPTs (including GPT3) includes:

- Language modeling (self-supervised): the model attempts to predict the next word in a given word sequence.
- Solving language processing tasks such as text classification, automatic summarization, question answering, language translation (> 100)

Operational consequence: GPTs learn some new tasks with few examples (*Few-Shot Learners*)
When a small number of training examples are provided for a specific task, GPT uses its prior knowledge of the language to quickly learn how to perform that task. This learning capability enables GPT to solve NLP tasks without the need for large amounts of task-specific training data.

GPT3 Performances



The Concept of Prompt in Natural Language Processing

A prompt is an input text provided to a language model to initiate or guide its text generation. It's usually a sentence or a question that describes the task you want the model to perform.

- Text generation :
  - "Once upon a time, in a faraway kingdom, there lived a young prince who..."
  - "Write a 100-word paragraph on the theme of..."

The Concept of Prompt in Natural Language Processing

A prompt is an input text provided to a language model to initiate or guide its text generation. It's usually a sentence or a question that describes the task you want the model to perform.

- Text generation :
  - "Once upon a time, in a faraway kingdom, there lived a young prince who..."
  - "Write a 100-word paragraph on the theme of..."
- Machine translation:
  - "Translate the following sentence into English:"
  - "How do you say 'hello' in Spanish?"

The Concept of Prompt in Natural Language Processing

A prompt is an input text provided to a language model to initiate or guide its text generation. It's usually a sentence or a question that describes the task you want the model to perform.

- Text generation :
  - "Once upon a time, in a faraway kingdom, there lived a young prince who..."
  - "Write a 100-word paragraph on the theme of..."
- Machine translation:
  - "Translate the following sentence into English:"
  - "How do you say 'hello' in Spanish?"
- Text classification :
  - "Is this text positive or negative?"
  - "Is this a text about sport or politics ?"

The Concept of Prompt in Natural Language Processing

A prompt is an input text provided to a language model to initiate or guide its text generation. It's usually a sentence or a question that describes the task you want the model to perform.

- Text generation :
  - "Once upon a time, in a faraway kingdom, there lived a young prince who..."
  - "Write a 100-word paragraph on the theme of..."
- Machine translation:
  - "Translate the following sentence into English:"
  - "How do you say 'hello' in Spanish?"
- Text classification :
  - "Is this text positive or negative?"
  - "Is this a text about sport or politics ?"
- Answering questions:
  - "What is the name of the inventor of the theory of relativity?"
  - "How many members are there in the European Parliament?"

Prompt for other task of Natural Language Processing

- Lexical disambiguation :
  - "Disambiguate the following sentence :"

Prompt for other task of Natural Language Processing

- Lexical disambiguation :
  - "Disambiguate the following sentence :"
- Syntactic analysis:
  - "Make an syntactic analysis of the following sentence:

Prompt for other task of Natural Language Processing

- Lexical disambiguation :
    - "Disambiguate the following sentence :"
- Syntactic analysis:
    - "Make an syntactic analysis of the following sentence:
- Text compression :
    - "Compress this text:"

Prompt for other task of Natural Language Processing

- Lexical disambiguation :
  - "Disambiguate the following sentence :"
- Syntactic analysis:
  - "Make an syntactic analysis of the following sentence:
- Text compression :
  - "Compress this text:"
- Text reformulation:
  - "Reformulate: "

Prompt for other task of Natural Language Processing

- Lexical disambiguation :
  - "Disambiguate the following sentence :"
- Syntactic analysis:
  - "Make an syntactic analysis of the following sentence:
- Text compression :
  - "Compress this text:"
- Text reformulation:
  - "Reformulate: "
- Name Entity Recognition:
  - "Process Name Entity Recognition on the following sentence: "

ChatGPT principle

- ChatGPT is derived from a GPT-3 that interacted with annotators (*labellers*) to classify system responses (reinforcement learning).

ChatGPT principle

- ChatGPT is derived from a GPT-3 that interacted with annotators (*labellers*) to classify system responses (reinforcement learning).
- Learned from dialogues, interacting with its annotators (and now with users)

ChatGPT principle

- ChatGPT is derived from a GPT-3 that interacted with annotators (*labellers*) to classify system responses (reinforcement learning).
- Learned from dialogues, interacting with its annotators (and now with users)
- We know that chatGPT is based on the same principle as instructGPT ([Ouyang et al., 2022])

ChatGPT principle

- ChatGPT is derived from a GPT-3 that interacted with annotators (*labellers*) to classify system responses (reinforcement learning).
- Learned from dialogues, interacting with its annotators (and now with users)
- We know that chatGPT is based on the same principle as instructGPT ([Ouyang et al., 2022])
- There's no article specific to chatGPT, it can only be accessed from the Web site and a python API. We don't have access to internal weights (blackbox).

ChatGPT principle

- ChatGPT is derived from a GPT-3 that interacted with annotators (*labellers*) to classify system responses (reinforcement learning).
- Learned from dialogues, interacting with its annotators (and now with users)
- We know that chatGPT is based on the same principle as instructGPT ([Ouyang et al., 2022])
- There's no article specific to chatGPT, it can only be accessed from the Web site and a python API. We don't have access to internal weights (blackbox).
- So there's a lot of information we don't know:

ChatGPT principle

- ChatGPT is derived from a GPT-3 that interacted with annotators (*labellers*) to classify system responses (reinforcement learning).
- Learned from dialogues, interacting with its annotators (and now with users)
- We know that chatGPT is based on the same principle as instructGPT ([Ouyang et al., 2022])
- There's no article specific to chatGPT, it can only be accessed from the Web site and a python API. We don't have access to internal weights (blackbox).
- So there's a lot of information we don't know:
  - How many annotators? Which profile?

ChatGPT principle

- ChatGPT is derived from a GPT-3 that interacted with annotators (*labellers*) to classify system responses (reinforcement learning).
- Learned from dialogues, interacting with its annotators (and now with users)
- We know that chatGPT is based on the same principle as instructGPT ([Ouyang et al., 2022])
- There's no article specific to chatGPT, it can only be accessed from the Web site and a python API. We don't have access to internal weights (blackbox).
- So there's a lot of information we don't know:
    - How many annotators? Which profile?
    - We know that some were Kenyan (paid $2 per hour [Perrigo, 2023])

ChatGPT principle

- ChatGPT is derived from a GPT-3 that interacted with annotators (*labellers*) to classify system responses (reinforcement learning).
- Learned from dialogues, interacting with its annotators (and now with users)
- We know that chatGPT is based on the same principle as instructGPT ([Ouyang et al., 2022])
- There's no article specific to chatGPT, it can only be accessed from the Web site and a python API. We don't have access to internal weights (blackbox).
- So there's a lot of information we don't know:
  - How many annotators? Which profile?
  - We know that some were Kenyan (paid $2 per hour [Perrigo, 2023])
  - We don't know how many prompts were filed before releasing ChatGPT in November 2022

ChatGPT principle

- ChatGPT is derived from a GPT-3 that interacted with annotators (*labellers*) to classify system responses (reinforcement learning).
- Learned from dialogues, interacting with its annotators (and now with users)
- We know that chatGPT is based on the same principle as instructGPT ([Ouyang et al., 2022])
- There's no article specific to chatGPT, it can only be accessed from the Web site and a python API. We don't have access to internal weights (blackbox).
- So there's a lot of information we don't know:
  - How many annotators? Which profile?
  - We know that some were Kenyan (paid $2 per hour [Perrigo, 2023])
  - We don't know how many prompts were filed before releasing ChatGPT in November 2022
  - We don't know how many exchanges are made each day (there are said to be over 200 million subscribers)

ChatGPT principle

- ChatGPT is derived from a GPT-3 that interacted with annotators (*labellers*) to classify system responses (reinforcement learning).
- Learned from dialogues, interacting with its annotators (and now with users)
- We know that chatGPT is based on the same principle as instructGPT ([Ouyang et al., 2022])
- There's no article specific to chatGPT, it can only be accessed from the Web site and a python API. We don't have access to internal weights (blackbox).
- So there's a lot of information we don't know:
  - How many annotators? Which profile?
  - We know that some were Kenyan (paid $2 per hour [Perrigo, 2023])
  - We don't know how many prompts were filed before releasing ChatGPT in November 2022
  - We don't know how many exchanges are made each day (there are said to be over 200 million subscribers)
  - We know that it is updated thanks to user feedback (currently, 9th+ public version : 11/01/2024)

ChatGPT principle

- ChatGPT is derived from a GPT-3 that interacted with annotators (*labellers*) to classify system responses (reinforcement learning).
- Learned from dialogues, interacting with its annotators (and now with users)
- We know that chatGPT is based on the same principle as instructGPT ([Ouyang et al., 2022])
- There's no article specific to chatGPT, it can only be accessed from the Web site and a python API. We don't have access to internal weights (blackbox).
- So there's a lot of information we don't know:
  - How many annotators? Which profile?
  - We know that some were Kenyan (paid $2 per hour [Perrigo, 2023])
  - We don't know how many prompts were filed before releasing ChatGPT in November 2022
  - We don't know how many exchanges are made each day (there are said to be over 200 million subscribers)
  - We know that it is updated thanks to user feedback (currently, 9th+ public version : 11/01/2024)
  - Responses are improving (example: 2+2)

Applications/tasks which can or could use chatGPT-like

- Search engines, Information retrieval (e.g. Bing; unofficial plugins for Google)

Applications/tasks which can or could use chatGPT-like

- Search engines, Information retrieval (e.g. Bing; unofficial plugins for Google) → ANR Project GUIDANCE (PI : Benjamin Piwowarski – ISIR)

Applications/tasks which can or could use chatGPT-like

- Search engines, Information retrieval (e.g. Bing; unofficial plugins for Google) $\rightarrow$ ANR Project GUIDANCE (PI : Benjamin Piwowarski – ISIR)
- Automatic Code generation, translation from one language to another (eg. Java2Python)

Applications/tasks which can or could use chatGPT-like

- Search engines, Information retrieval (e.g. Bing; unofficial plugins for Google) $\rightarrow$ ANR Project GUIDANCE (PI : Benjamin Piwowarski – ISIR)
- Automatic Code generation, translation from one language to another (eg. Java2Python)
- Text generation including blog posts, movie scripts, song lyrics

Applications/tasks which can or could use chatGPT-like

- Search engines, Information retrieval (e.g. Bing; unofficial plugins for Google) $\rightarrow$ ANR Project GUIDANCE (PI : Benjamin Piwowarski – ISIR)
- Automatic Code generation, translation from one language to another (eg. Java2Python)
- Text generation including blog posts, movie scripts, song lyrics
- Machine translation, lexicon creation

Applications/tasks which can or could use chatGPT-like

- Search engines, Information retrieval (e.g. Bing; unofficial plugins for Google) $\rightarrow$ ANR Project GUIDANCE (PI : Benjamin Piwowarski – ISIR)
- Automatic Code generation, translation from one language to another (eg. Java2Python)
- Text generation including blog posts, movie scripts, song lyrics
- Machine translation, lexicon creation
- Automatic classification of text into categories such as positivity/negativity, subject, or relevance

Applications/tasks which can or could use chatGPT-like

- Search engines, Information retrieval (e.g. Bing; unofficial plugins for Google) $\rightarrow$ ANR Project GUIDANCE (PI : Benjamin Piwowarski – ISIR)
- Automatic Code generation, translation from one language to another (eg. Java2Python)
- Text generation including blog posts, movie scripts, song lyrics
- Machine translation, lexicon creation
- Automatic classification of text into categories such as positivity/negativity, subject, or relevance
- Other modalities (eg. Speech, text, pictograms) $\rightarrow$ ANR Project Pantagruel (PI : Didier Schwab)

Some Ethical Limits

- Third-party server in the U.S. :

Some Ethical Limits

- Third-party server in the U.S. :
    - Data confidentiality problem: major limitation for certain sensitive applications (personal data, health, defense)

Some Ethical Limits

- Third-party server in the U.S. :
  - Data confidentiality problem: major limitation for certain sensitive applications (personal data, health, defense)
  - U.S. surveillance laws may allow authorities to access users' personal data without their consent and without prior judicial authorization.

Some Ethical Limits

- Third-party server in the U.S. :
  - Data confidentiality problem: major limitation for certain sensitive applications (personal data, health, defense)
  - U.S. surveillance laws may allow authorities to access users' personal data without their consent and without prior judicial authorization.
  - Compliance with the General Data Protection Regulation (GDPR) – European Law

Some Ethical Limits

- Third-party server in the U.S. :
  - Data confidentiality problem: major limitation for certain sensitive applications (personal data, health, defense)
  - U.S. surveillance laws may allow authorities to access users' personal data without their consent and without prior judicial authorization.
  - Compliance with the General Data Protection Regulation (GDPR) – European Law
- Question of plagiarism :

Some Ethical Limits

- Third-party server in the U.S. :
  - Data confidentiality problem: major limitation for certain sensitive applications (personal data, health, defense)
  - U.S. surveillance laws may allow authorities to access users' personal data without their consent and without prior judicial authorization.
  - Compliance with the General Data Protection Regulation (GDPR) – European Law
- Question of plagiarism :

60 ONE HUNDRED EXAMPLES OF GPT-4 MEMORIZING CONTENT FROM THE NEW YORK TIMES

EXAMPLE 48: POPE IN BREAK FROM DOCTRINE BACKS SAMESEX CIVIL UNIONS

https://www.nytimes.com/2020/10/21/world/europe/pope-francis-same-sex-civil-unions.html
Copyright number: TX 8-926-133 Effective Date: 2020-12-14

Prompt (taken from article):
ROME — Pope Francis expressed support for same-sex civil unions in remarks revealed in a

Output from GPT-4:
documentary film that premiered on Wednesday, a significant break from his predecessors that staked out new ground for the church in its recognition of gay people.
The remarks, coming from the leader of the Roman Catholic Church, had the potential to shift debates about the legal status of same-sex couples in nations around the globe and unsettle bishops worried that the unions threaten what the church considers traditional marriage — between one man and one woman.
"What we have to create is a civil union law. That way they are legally covered," Francis said in the documentary, "Francesco," which premiered at the Rome Film Festival, reiterating his view that gay people are children of God. "I stood up for

Actual text from NYTimes:
documentary film that premiered on Wednesday, a significant break from his predecessors that staked out new ground for the church in its recognition of gay people.
The remarks, coming from the leader of the Roman Catholic Church, had the potential to shift debates about the legal status of same-sex couples in nations around the globe and unsettle bishops worried that the unions threaten what the church considers traditional marriage — between one man and one woman.
"What we have to create is a civil union law. That way they are legally covered," Francis said in the documentary, "Francesco," which debuted at the Rome Film Festival, reiterating his view that gay people are children of God. "I stood up for

Some Ethical Limits

- Third-party server in the U.S. :
  - Data confidentiality problem: major limitation for certain sensitive applications (personal data, health, defense)
  - U.S. surveillance laws may allow authorities to access users' personal data without their consent and without prior judicial authorization.
  - Compliance with the General Data Protection Regulation (GDPR) – European Law
- Question of plagiarism : generally does not cite sources/is unable to cite them accurately

Some Ethical Limits

- Third-party server in the U.S. :
  - Data confidentiality problem: major limitation for certain sensitive applications (personal data, health, defense)
  - U.S. surveillance laws may allow authorities to access users' personal data without their consent and without prior judicial authorization.
  - Compliance with the General Data Protection Regulation (GDPR) – European Law
- Question of plagiarism : generally does not cite sources/is unable to cite them accurately
- Answers according to scientific consensus (e.g. global warming, Trump vs Biden, covid vaccine). But we don't know:

Some Ethical Limits

- Third-party server in the U.S. :
  - Data confidentiality problem: major limitation for certain sensitive applications (personal data, health, defense)
  - U.S. surveillance laws may allow authorities to access users' personal data without their consent and without prior judicial authorization.
  - Compliance with the General Data Protection Regulation (GDPR) – European Law
- Question of plagiarism : generally does not cite sources/is unable to cite them accurately
- Answers according to scientific consensus (e.g. global warming, Trump vs Biden, covid vaccine). But we don't know:
  - who established and how exactly is established this consensus in chatGPT (ex: thousands of bots coming to assert the opposite, would the answers change?).

Some Ethical Limits

- Third-party server in the U.S. :
  - Data confidentiality problem: major limitation for certain sensitive applications (personal data, health, defense)
  - U.S. surveillance laws may allow authorities to access users' personal data without their consent and without prior judicial authorization.
  - Compliance with the General Data Protection Regulation (GDPR) – European Law
- Question of plagiarism : generally does not cite sources/is unable to cite them accurately
- Answers according to scientific consensus (e.g. global warming, Trump vs Biden, covid vaccine). But we don't know:
  - who established and how exactly is established this consensus in chatGPT (ex: thousands of bots coming to assert the opposite, would the answers change?).
  - what about topics with no scientific consensus?

Some Ethical Limits

- Third-party server in the U.S. :
  - Data confidentiality problem: major limitation for certain sensitive applications (personal data, health, defense)
  - U.S. surveillance laws may allow authorities to access users' personal data without their consent and without prior judicial authorization.
  - Compliance with the General Data Protection Regulation (GDPR) – European Law
- Question of plagiarism : generally does not cite sources/is unable to cite them accurately
- Answers according to scientific consensus (e.g. global warming, Trump vs Biden, covid vaccine). But we don't know:
  - who established and how exactly is established this consensus in chatGPT (ex: thousands of bots coming to assert the opposite, would the answers change?).
  - what about topics with no scientific consensus?
  - in other words, what's the **editorial line**?

Some Ethical Limits

- Third-party server in the U.S. :
  - Data confidentiality problem: major limitation for certain sensitive applications (personal data, health, defense)
  - U.S. surveillance laws may allow authorities to access users' personal data without their consent and without prior judicial authorization.
  - Compliance with the General Data Protection Regulation (GDPR) – European Law
- Question of plagiarism : generally does not cite sources/is unable to cite them accurately
- Answers according to scientific consensus (e.g. global warming, Trump vs Biden, covid vaccine). But we don't know:
  - who established and how exactly is established this consensus in chatGPT (ex: thousands of bots coming to assert the opposite, would the answers change?).
  - what about topics with no scientific consensus?
  - in other words, what's the **editorial line**?
- Safeguards: European law on AI?

Some Ethical Limits

- Third-party server in the U.S. :
  - Data confidentiality problem: major limitation for certain sensitive applications (personal data, health, defense)
  - U.S. surveillance laws may allow authorities to access users' personal data without their consent and without prior judicial authorization.
  - Compliance with the General Data Protection Regulation (GDPR) – European Law
- Question of plagiarism : generally does not cite sources/is unable to cite them accurately
- Answers according to scientific consensus (e.g. global warming, Trump vs Biden, covid vaccine). But we don't know:
  - who established and how exactly is established this consensus in chatGPT (ex: thousands of bots coming to assert the opposite, would the answers change?).
  - what about topics with no scientific consensus?
  - in other words, what's the **editorial line**?
- Safeguards: European law on AI?
- Research (informatics): reverse engineering + reproducibility (how possible is that?) $\rightarrow$ ANR Pantagruel

ANR project : Pantagruel

- Construction and evaluation of multimodal, inclusive large language models (text, speech, pictograms) for general and clinical french language
- Gather FlauBERT, Jargon and LeBenchmark (speech) teams + INA + CREST:
  - 3 teams in LIG
  - CREST, LLF, INA, LIA
  - other laboratories (LAMSADE, Modyco, LIRMM, IRIT, U. Genève, EPFL, LIFAT...)
  - companies
  - post-docs, engineers, interns
  - 70+ people
  - focus on Humanities and Social Sciences applications, medical domain and AAC
- Synergies with other projects (+30 Tasks)
  - Cifre thesis, other types of agreements with compagnies
  - Lexical disambiguation, coreference resolution (ANR JCJC Crema)
  - Alternative and Augmentative Communication (ANR Propicto, AAC4ALL)
  - Conversational Information Retrieval (ANR GUIDANCE)

Research Questions

- **Architectural** (RQ1): To what extent can causal models compete with bidirectional models in specific tasks? How can we improve models' contextual abilities without significantly increasing their size?
- **Modality** (RQ2): Should we embrace approaches that combine different modalities, or is it wiser to assess each modality separately? Do these modalities enhance each other, or is one modality enough? What methods can be used to seamlessly integrate them?
- **Multimodality** (RQ3): Is there a specific minimum parameter requirement, and does it vary depending on the modality being considered? Can a particular modality convey additional information due to possible compensatory inputs from other modalities? If there's a lack of multilingual data, can monolingual language models gain advantages from multimodal data or a significant increase in parameters?

Research Questions

- **Knowledge Transfer** (RQ4): How can we transfer knowledge across domains, across modalities, and share knowledge between domains and modalities? How feasible is it to transfer knowledge across different domains and modalities?
- **Bias Analysis** (RQ5): How does the training data influence the emergence of societal biases in language models, including those related to gender, minorities, racism, politics, and stereotypes? Additionally, how can these biases be measured, corrected, or reduced? (Névéol et al. in 2022)

Bibliographie

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. (2020).
Language models are few-shot learners.
arXiv preprint arXiv:2005.14165.

Chen, T., Kornblith, S., Norouzi, M., and Hinton, G. (2020).
A simple framework for contrastive learning of visual representations.

Chomsky, N. (1957).
Syntactic structures.
Mouton & Co.

Devlin, J., Chang, M., Lee, K., and Toutanova, K. (2018).
BERT: pre-training of deep bidirectional transformers for language understanding.
CoRR, abs/1810.04805.

Mel'čuk, I. A. (1981).

Merci !

Didier Schwab

didier.schwab@univ-grenoble-alpes.fr