

TLC Conference

Workshop, March 30, 2023

The Technology Behind ChatGPT

Capabilities, Limitations and Challenges for Education

@AmsterdamNLP

Institute for Logic, Language & Computation
University of Amsterdam

Workshop overview

What is ChatGPT & What can it (not) do?

Discussion: ChatGPT in the classroom



Jelke Bloem



Assistant Professor in Computational Humanities
Institute for Logic, Language and Computation (ILLC)
Faculty of Humanities, University of Amsterdam

#linguistics #NLP #interpretability
#digitalhumanities #semantics



Slides joint work with: Prof. Raquel Fernández Rovira (Dialogue Systems)
Dr. Jelle Zuidema (Explainable NLP)
Dr. Sandro Pezzelle (Responsible NLP)
NLP & Digital Humanities Group, ILLC

Introduction & Use Cases

Welcome! (by ChatGPT)



Please give me a good sentence to welcome people to an event where we talk about ChatGPT.



"Welcome to our event where we explore the capabilities and potential of ChatGPT, the powerful language model that is revolutionizing the way we interact with technology."

ChatGPT: the Five Ws

What. A model that generates language based on some input

- **Chat:** chat-based model +
- **GPT: Generative Pre-Trained Transformer**

ChatGPT: the Five Ws

What. A model that generates language based on some input

- **Chat:** chat-based model +
- **GPT: Generative Pre-Trained Transformer**

More details later!

ChatGPT: the Five Ws

What. A model that generates language based on some input

- **Chat:** chat-based model +
- **GPT: Generative Pre-Trained Transformer**

When. Released on November 30, 2022

ChatGPT: the Five Ws

What. A model that generates language based on some input

- **Chat:** chat-based model +
- **GPT: Generative Pre-Trained Transformer**

When. Released on November 30, 2022

(from) Who. OpenAI, private “capped-profit” company (mission: “to ensure that artificial general intelligence (AGI) ... benefits all of humanity”)

ChatGPT: the Five Ws

What. A model that generates language based on some input

- **Chat:** chat-based model +
- **GPT: Generative Pre-Trained Transformer**

When. Released on November 30, 2022

(from) Who. OpenAI, private “capped-profit” company (mission: “to create artificial general intelligence (AGI) ... benefits all of humanity”)

OpenAI and Microsoft
Extend Partnership

ChatGPT: the Five Ws

What. A model that generates language based on some input

- **Chat:** chat-based model +
- **GPT: Generative Pre-Trained Transformer**

When. Released on November 30, 2022

(from) Who. OpenAI, private “capped-profit” company (mission: “to ensure that artificial general intelligence (AGI) ... benefits all of humanity”)

Where. Hosted online on the OpenAI website: <https://openai.com/about/>

ChatGPT: the Five Ws

What. A model that generates language based on some input

- **Chat:** chat-based model +
- **GPT: Generative Pre-Trained Transformer**

When. Released on November 30, 2022

(from) Who. OpenAI, private “capped-profit” company (mission: “to ensure that artificial general intelligence (AGI) ... benefits all of humanity”)

Where. Hosted online on the OpenAI website: <https://openai.com/about/>

Why. To interact with users in “a conversational way”

ChatGPT: the Five Ws

What. A model that generates language based on some input

- **Chat:** chat-based model +
- **GPT: Generative Pre-Trained Transformer**

When. Released on November 30, 2022

(from) Who. OpenAI, private “capped-profit” company (mission: “to ensure that artificial general intelligence (AGI) ... benefits all of humanity”)

Where. Hosted online on the OpenAI website: <https://openai.com/about/>

Why. But recall the long-term “why”: “to build ... AGI”*

ChatGPT: the Five Ws

What. A model that generates language based on some input

- **Chat:** chat-based model +
- **GPT: Generative Pre-Trained Transformer**

When. Released on November 30, 2022

(from) Who. OpenAI, private “capped-profit” company (mission: “to ensure that artificial general intelligence (AGI) ... benefits all of humanity”)

Where. Hosted online on the OpenAI website: <https://openai.com/about/>

Why. But recall the long-term “why”: “to build ... AGI”*

* (AGI)—by which we mean highly autonomous systems that outperform humans at most economically valuable work—

Where does it come from?

In the last few years, research in **Natural Language Processing (NLP)** dominated by large language models (LLMs)

Where does it come from?

In the last few years, research in **Natural Language Processing (NLP)** dominated by large language models (LLMs)

GPT, GPT-2, GPT-3, InstructGPT-3 (OpenAI), DialoGPT (Microsoft)

Where does it come from?

In the last few years, research in **Natural Language Processing (NLP)** dominated by large language models (LLMs)

GPT, GPT-2, GPT-3, InstructGPT-3 (OpenAI), DialoGPT (Microsoft)

> **state-of-the-art** in NLP many tasks:

translation, question-answering, cloze tasks, “on-the-fly” reasoning or domain adaptation (e.g., unscrambling words, using a novel word in a sentence, or performing 3-digit arithmetic)

What can it be used for?



What should one use chatGPT for?



ChatGPT can be used for a variety of tasks, such as:



- Generating human-like text, such as writing essays, articles, or stories
- Answering questions in a conversational manner
- Generating responses in a chatbot or virtual assistant
- Summarizing or translating text
- Generating programming code
- And many more!

It can also be fine-tuned on specific task or domain to improve performance.

ChatGPT can play games



Home > Game News > Dungeons & Dragons Player Gets AI Bot To DM, Works Surprisingly Well

Dungeons & Dragons Player Gets AI Bot To DM, Works Surprisingly Well

ChatGPT proves that the days of human DMs might be numbered.

BY SEAN MURRAY PUBLISHED DEC 24, 2022



<https://www.thegamer.com/dungeons-dragons-chatgpt-ai-dm/>

ChatGPT can do marketing

Almost 30% of Professionals Have Used ChatGPT for Work Purposes

People who work in marketing and advertising are using the AI chatbot the most, followed by tech workers and consultants, according to a Fishbowl survey.



By [Michael Kan](#) January 20, 2023



<https://www.pcmag.com/news/almost-30-of-professionals-have-used-chatgpt-for-work-purposes>

ChatGPT can code

[nature](#) > [news](#) > article

NEWS | 08 December 2022

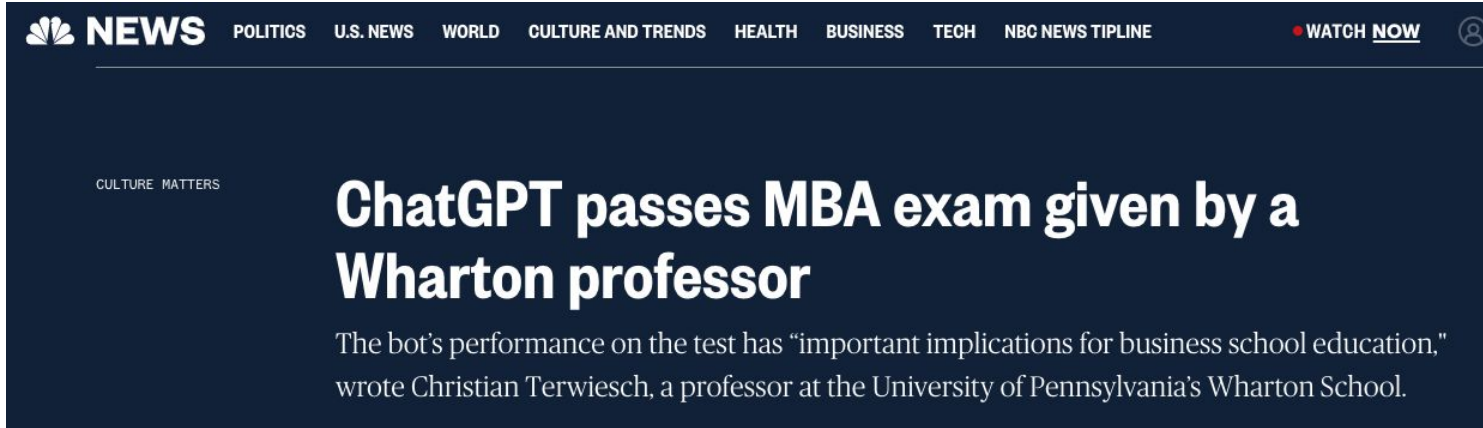
Are ChatGPT and AlphaCode going to replace programmers?

OpenAI and DeepMind systems can now produce meaningful lines of code, but software engineers shouldn't switch careers quite yet.

[Davide Castelvecchi](#)

<https://www.nature.com/articles/d41586-022-04383-z>

ChatGPT can pass MBA exams

A screenshot of the NBC News website header. The top navigation bar includes the NBC News logo, followed by menu items: POLITICS, U.S. NEWS, WORLD, CULTURE AND TRENDS, HEALTH, BUSINESS, TECH, and NBC NEWS TIPLINE. On the right side of the navigation bar, there is a 'WATCH NOW' button with a red dot and a user profile icon. Below the navigation bar, the article title 'ChatGPT passes MBA exam given by a Wharton professor' is displayed in large white text. To the left of the title, the text 'CULTURE MATTERS' is visible. Below the title, a short paragraph of text is shown in white, starting with 'The bot's performance on the test has "important implications for business school education," wrote Christian Terwiesch, a professor at the University of Pennsylvania's Wharton School.'

<https://www.nbcnews.com/tech/tech-news/chatgpt-passes-mba-exam-wharton-professor-rcna67036>

ChatGPT can write scientific articles

nature

Explore content ▾

About the journal ▾

Publish with us ▾

Subscribe

[nature](#) > [news](#) > [article](#)

NEWS | 18 January 2023

ChatGPT listed as author on research papers: many scientists disapprove

At least four articles credit the AI tool as a co-author, as publishers scramble to regulate its use.

[Chris Stokel-Walker](#)



<https://www.nature.com/articles/d41586-023-00107-z>

At the UvA, people had some questions too...



international

Pepijn Stoop | Will ChatGPT soon be making our exams?



Pepijn Stoop , 12 januari 2023 - 10:22

<https://www.folia.nl/international/155061/pepijn-stoop-will-chatgpt-soon-be-making-our-exams>

Can I replace my students with ChatGPT?

J

Why do chatbots exhibit gender stereotypes? Please answer in 80 words or less.



Chatbots exhibit gender stereotypes because they are programmed and trained using data sets that reflect societal norms and biases. This leads to the chatbot perpetuating and reinforcing gender stereotypes, such as assuming women are more nurturing and men are more assertive. It is important to be aware of and actively work to counteract these biases in the development and training of chatbots.

Text manipulation



Chatbots exhibit gender stereotypes because they are programmed and trained using data sets that reflect societal norms and biases. This leads to the chatbot perpetuating and reinforcing gender stereotypes, such as assuming women are more nurturing and men are more assertive. It is important to be aware of and actively work to counteract these biases in the development and training of chatbots.



Please rewrite the text above from the perspective of the user



When I interact with a chatbot, I often notice that it exhibits gender stereotypes. For example, it may assume that women are more nurturing and men are more assertive. This is because the chatbot is programmed and trained using data sets that reflect societal norms and biases. It's important for the creators of chatbots to be aware of these biases and actively work to counteract them in order to create more inclusive and accurate interactions.

Style

J

Chatbots exhibit gender stereotypes because they are programmed and trained using data sets that reflect societal norms and biases. This leads to the chatbot perpetuating and reinforcing gender stereotypes, such as assuming women are more nurturing and men are more assertive. It is important to be aware of and actively work to counteract these biases in the development and training of chatbots.

Please summarize the above text in the style of a Trump tweet.



Chatbots have gender stereotypes because of programming & data sets reflecting societal norms. Must be aware & actively work to counteract these biases in development & training. #GenderEquality #ChatbotBias


Please rewrite the text above in the style of the King James bible.




"Lo, chatbots do exhibit gender stereotypes, for they are programmed and trained using data sets that reflect societal norms and biases. And thus, the chatbot doth perpetuate and reinforce gender stereotypes, such as assuming women are more nurturing and men are more assertive. Verily, it is important to be aware of and actively work to counteract these biases in the development and training of chatbots."

Translation

J Why do chatbots exhibit gender stereotypes? Please use around 80 words and answer in Frisian.


 Chatbots wurde faak stereotype geslachtsrollen toand, omdat se ûntwurpen binne troch minsken dy't ynfloedearre binne troch maatskiplike stereotypen. Dit kin liedes ta it programmearjen fan in chatbot om stereotype gedrach te sjen, lykas in froulike chatbot dy't freonlik is en in manlike chatbot dy't autoritêr is. Dit is problematysk, omdat it de maatskiplike stereotypen fersterket en kin liedes ta diskriminaasje en ûngelikense behanneling fan froulju en manlju.

J Rewrite in Klingon

 chatbot stereotype genderDaq 'angbogh, quvHa'ghach ghajbe' 'ej lIngmey lutu'lu'bogh. 'ej lutu'lu'bogh, 'angbogh ghaH 'ej ghaH ghajbe' je. 'angbogh vay' ghaH ghaH ghajbe' 'ej 'angbogh vay' ghaH ghaH ghajbe' je. je qaS cha'logh qaS cha'logh tlhInganpu' vaQ cha'logh.

J chatbot stereotype genderDaq 'angbogh, quvHa'ghach ghajbe' 'ej lIngmey lutu'lu'bogh. 'ej lutu'lu'bogh, 'angbogh ghaH 'ej ghaH ghajbe' je. 'angbogh vay' ghaH ghaH ghajbe' 'ej 'angbogh vay' ghaH ghaH ghajbe' je. je qaS cha'logh qaS cha'logh tlhInganpu' vaQ cha'logh.

Please rewrite in English.

 Chatbots, like humans, have stereotypes and biases based on gender. Both male and female chatbots can have these stereotypes and biases, and it can affect how they interact with users. It's important to recognize and address these issues in the development of chatbots.

Better prompts, better results

J my homework is to write about Chomsky, plz help



Noam Chomsky is a renowned American linguist, philosopher, cognitive scientist, historian, and social critic. He is best known for his work in the field of theoretical linguistics, specifically for his theory of generative grammar, which suggests that the ability to understand and produce language is innate to the human mind. Chomsky's ideas have had a major impact on the field of linguistics and have also influenced the study of other cognitive processes such as memory and perception. Additionally, Chomsky is also known for his criticism of U.S. foreign policy and his activism on behalf of political causes such as workers' rights and civil liberties.

J

Please write a 300 word introduction to the linguist Noam Chomsky, in particular highlighting his arguments for the existence of a universal grammar. Include major points of critique that were raised against this position and name his critics.



Noam Chomsky is a prominent linguist and philosopher who has had a significant impact on the field of linguistics and cognitive science. One of his major contributions is the theory of universal grammar, which posits that all human languages share a common underlying structure or grammar that is innate to the human mind. This theory suggests that humans are born with an innate capacity to acquire language, and that this capacity is a fundamental aspect of human biology.

Chomsky argues that this innate grammar is responsible for the ease with which children learn language, and that it allows them to acquire the complex structures of their native language with minimal exposure. He also argues that this innate grammar is the reason why different languages share common features, such as the presence of nouns and verbs, and the use of grammatical structures like tense and agreement.

Critics of Chomsky's theory of universal grammar have raised several points of critique. One of the main criticisms is that the theory is difficult to test and that there is little empirical evidence to support it. Some critics have also argued that the theory is overly-simplistic and that it does not account for the diversity of languages and cultures around the world. Additionally, some have argued that Chomsky's theory does not fully explain the complexity and variability of language and that other factors such as social and cultural influences play a much more significant role.

Some of the prominent critics of Chomsky's theory are Steven Pinker, a cognitive psychologist, who has argued that the innate capacity for language is not as specific as Chomsky suggests and that much of our capacity for language is derived from general cognitive abilities. The anthropologist and linguist, Dan Everett, has also critiqued Chomsky's theory arguing that the capacity for language is not innate, but rather, is a product of cultural evolution.

In summary, Noam Chomsky's theory of universal grammar posits that all human languages share a common underlying structure that is innate to the

My students don't write that well!



Please rewrite the initial essay as if it was written by a non-native speaker of English using simpler words.



Chatbots are computer programs that can have a conversation with people. They are made to talk like human. But sometimes they show stereotypes of men and women. This happens because the people who make the chatbots use data that has these stereotypes in it. It's important for the people who make the chatbots to be aware of this problem and try to make the chatbots not have stereotypes.

If students can do it...

J

Please evaluate the initial essay according to the following evaluation criteria:

1. Coverage of main arguments
2. Originality
3. Use of examples
4. Critical approach to the argument
5. Clarity of expression

Please provide a grade for the essay according to the Dutch grading system, and comment on each of the evaluation criteria.



Grade: 7

1. Coverage of main arguments: The essay covers the main argument well, which is why chatbots exhibit gender stereotypes. It explains the cause and the importance of countering these biases in the development and training of chatbots.
2. Originality: The argument presented in the essay is not entirely original, but it presents a common concern in the field of AI development.
3. Use of examples: The essay provides a clear example of how chatbots may perpetuate gender stereotypes, which helps to illustrate the main argument.
4. Critical approach to the argument: The essay presents the argument in a balanced and critical manner, highlighting the importance of addressing the issue of gender stereotypes in chatbots.
5. Clarity of expression: The essay is clearly written and easy to understand, making it accessible to a broad audience.

Overall, the essay presents a clear argument and provides a solid example to illustrate the main point. It is written in a clear and accessible manner.

However, the argument presented is not entirely original.

Limitations and failure cases: Specific information

J

Please rewrite the initial essay to include mention of model debiasing, and provide relevant references in APA style.



Chatbots are computer programs that can have a conversation with people, however, they often exhibit gender stereotypes. This happens because the data sets used to train these chatbots reflect societal norms and biases. As a result, chatbots may perpetuate and reinforce stereotypes, such as assuming women are more nurturing and men are more assertive.

To counteract these biases, researchers in the field of AI have proposed the use of model debiasing techniques (Kilbertus et al., 2017; Wang et al., 2018). These techniques aim to correct for bias in the data and the model, thus reducing the likelihood of stereotypes being perpetuated by the chatbot.

However, it is important to note that model debiasing is not a one-size-fits-all solution, and it can not completely eliminate all biases. Therefore, it is crucial to also consider other strategies such as diversifying the data sets used to train the chatbot, and actively working to counteract these biases in the development and training of chatbots (Dwork et al., 2012).

References:

Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. (2012). Fairness through aware algorithms. *Communications of the ACM*, 55(10), 78–84. <https://doi.org/10.1145/2380656.2380696>

Kilbertus, N., Valera, I., Güneş, M., & Shawe-Taylor, J. (2017). The measurable impact of fair machine learning. *arXiv [cs.LG]*. <http://arxiv.org/abs/1711.05144>

Wang, Y., & Blei, D. M. (2018). Mitigating unwanted biases with adversarial learning. *Proceedings of the 31st Conference on Neural Information Processing Systems*. <http://papers.nips.cc/paper/7859-mitigating-unwanted-biases-with-adversarial-learning>

References?

References:


Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. (2012). Fairness through aware algorithms. *Communications of the ACM*, 55(10), 78–84.

<https://doi.org/10.1145/2380656.2380696>

Kilbertus, N., Valera, I., Güneş, M., & Shave-Taylor, J. (2017). The measurable impact of fair machine learning. *arXiv [cs.LG]*. [http://arxiv.org](http://arxiv.org/abs/1711.05144)

[/abs/1711.05144](http://arxiv.org/abs/1711.05144)

Wang, Y., & Blei, D. M. (2018). Mitigating unwanted biases with adversarial learning. *Proceedings of the 31st Conference on Neural Information Processing Systems*. <http://papers.nips.cc/paper/7859-mitigating-unwanted-biases-with-adversarial-learning>



Deep Neural Networks with Box Convolutions

Part of [Advances in Neural Information Processing Systems 31 \(NeurIPS 2018\)](#)

Bibtex

Metadata

Paper

Reviews

Authors

Egor Burkov, Victor Lempitsky

Abstract

Box filters computed using integral images have been part of the computer vision toolset for a long time. Here, we show that a convolutional layer that computes box filter responses in a sliding manner can be used within deep architectures, whereas the dimensions and the offsets of the sliding boxes in such a layer can be learned as part of an end-to-end loss minimization. Crucially, the training process can make the

Specific questions

J How can we show that embedding regularization in large language models leads to over-debiasing? Please answer in 80 words or less.



Embedding regularization in large language models can lead to over-debiasing by enforcing a specific set of biases onto the model, rather than allowing it to learn biases naturally through its training data. This can result in the model becoming overly sensitive to certain groups or demographics, leading to over-correction and potentially harmful biases.

Domain-general versus domain-specific information

Domain-specific knowledge: A Dutch linguist interviews ChatGPT



Hoe zit die zin dan in elkaar?



De zin "Dat komt niet als een verrassing" is een voegwoordelijke zin, bestaande uit een onderwerp "dat" en een voegwoordelijke bijzin "komt niet als een verrassing". "Komt niet" is een afgeleide van het werkwoord "komen" en fungeert hier als de hoofdzin van de voegwoordelijke bijzin. "Als een verrassing" is een bijwoordelijke bepaling die preciezer aangeeft hoe de situatie of het gebeuren in de hoofdzin beleefd wordt. In deze zin wordt voorgesteld dat iets verwacht wordt, en dus geen verrassing is.



Prof. Peter Arno Coppen, University of Utrecht

<https://neerlandistiek.nl/2023/01/taalkundig-redeneren-met-chatgpt/>

Domain-specific knowledge: Dutch lexicography

PO

Met welke werkwoorden wordt het woord "belofte" gecombineerd?



Er zijn verschillende werkwoorden die vaak gecombineerd worden met het woord "belofte":



- **Maken:** iemand kan een belofte maken aan iemand anders.
- Doen: iemand kan een belofte doen aan iemand anders.
- Breken: iemand kan een belofte breken als hij of zij niet doet wat hij of zij heeft beloofd.
- Houden: iemand kan een belofte houden als hij of zij doet wat hij of zij heeft beloofd.
- Nakomen: iemand kan een belofte nakomen als hij of zij doet wat hij of zij heeft beloofd.
- Weigeren: iemand kan een belofte weigeren als hij of zij niet bereid is om te doen wat hij of zij heeft beloofd.

Beloften kunnen verbale of schriftelijke vormen aannemen en kunnen betrekking hebben op allerlei verschillende zaken, zoals persoonlijke verplichtingen, zakelijke afspraken of politieke beloften. Het is belangrijk om beloften serieus te nemen en ze na te komen, aangezien ze vaak gebaseerd zijn op vertrouwen en integriteit.

Prof. Nicoline van der Sijs, Instituut voor Nederlandse Taal

<https://neerlandistiek.nl/2022/12/chattie/>

Model hallucination

Model trained to produce language, not truth

J

Describe the incident in which a chatbot was accidentally elected onto the University of Amsterdam student council in 2005. Please answer in 100 words or less and provide a bibliography of relevant historical sources.



In 2005, a chatbot called "Tron" was accidentally elected onto the University of Amsterdam student council. The incident occurred when students mistook the chatbot, which was created as a prank by a group of computer science students, for a real person while voting for representatives online. Tron received more than 100 votes, earning it a spot on the council. The incident sparked debate about the potential for technology to deceive people and the importance of verifying the identities of online candidates.

Bibliography:

- "Tron the Chatbot Elected to Student Council" (2005) The New York Times
- "Chatbot Tron Elected to Student Council" (2005) BBC News
- "The Rise of the Machines: How a Chatbot Was Elected to a Dutch University's Student Council" (2005) Wired

Other limitations

- Math stuff
- Units of measure
- Weather forecast, current events
- Metalinguistic tasks: writing backwards, words that end with a letter
- Physical, temporal reasoning, logic

More generally, model is trained with only textual data, so in principle it doesn't have access to info that requires other modalities

<https://github.com/giuven95/chatgpt-failures>

Summary of strengths and weaknesses

Strengths

- Grammatical, mostly coherent
- Question-answering
- Dialogue, responsive to feedback
- Domain-general knowledge
- “Creative”
- Styles
- Translation
- Can program and other common structured tasks
- ...

Weaknesses

- Truthfulness
- Reliability & stability
- Hallucination
- Overconfidence
- Domain-specific knowledge
- Logic and inference
- Metalinguistic tasks
- Bias and stereotypes
- Superficial content filters
- ...

ChatGPT is powered by Machine Learning

What ChatGPT generates is a consequence of what it has learned over a period of training.



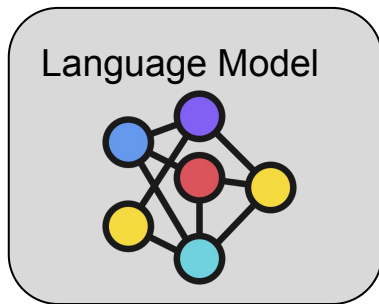
Unsupervised learning
from tons of data

Learning from human
trainers

Unsupervised Learning



It's cold. Open the



PREDICTION



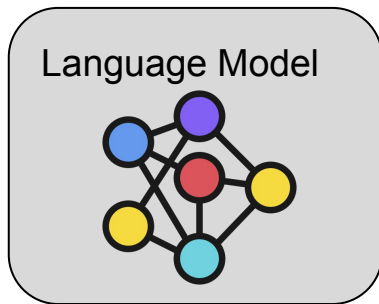
Training Data

Transformer Neural Network

Unsupervised Learning



It's cold. Open the



PREDICTION



0.2% ... 10% ... 40% 1%

ask door window what

Training Data

Transformer Neural Network

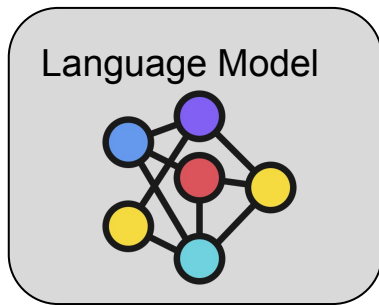
vocabulary words

Unsupervised Learning



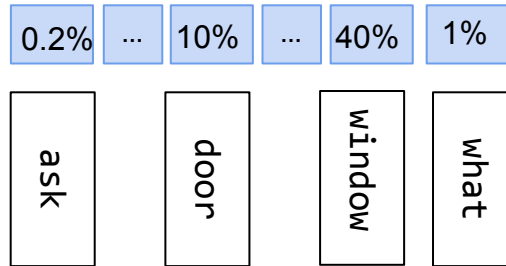
It's cold. Open the

Training Data



Transformer Neural Network

PREDICTION



vocabulary words

calculate error & update model

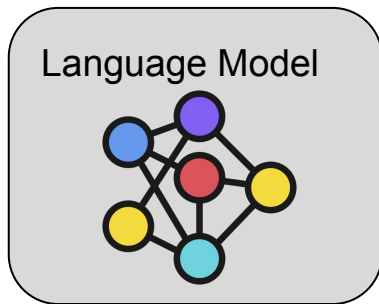


actual next word
door

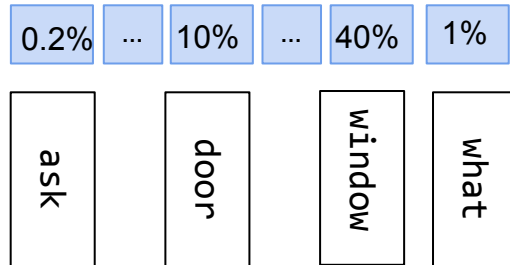
Unsupervised Learning



It's cold. Open the



PREDICTION



calculate error & update model



actual next word
door

Training Data

Transformer Neural Network

vocabulary words

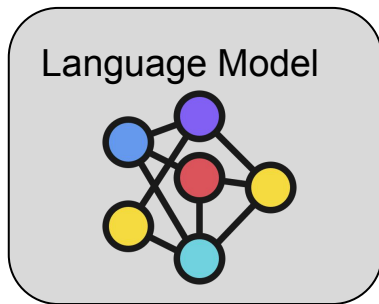


- ~ 175 billion parameters
- Exposed to 300 billion words during training

Unsupervised Learning



It's cold. Open the



PREDICTION



calculate error & update model



actual next word
door

0.2% ... 10% ... 40% 1%

ask door window what

vocabulary words

Training Data

Transformer Neural Network



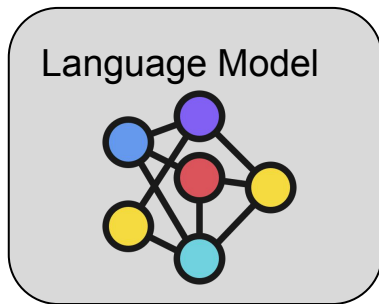
- ~ 175 billion parameters
- Exposed to 300 billion words during training

How long does it take to train it?

Unsupervised Learning



It's cold. Open the



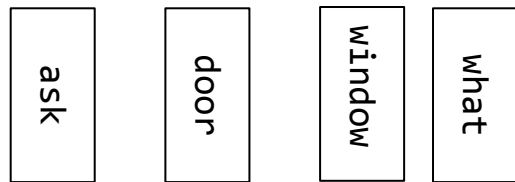
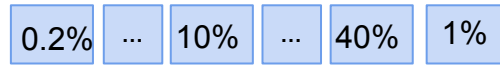
PREDICTION



calculate error & update model



actual next word
door



vocabulary words

Training Data

Transformer Neural Network



- ~ 175 billion parameters
- Exposed to 300 billion words during training

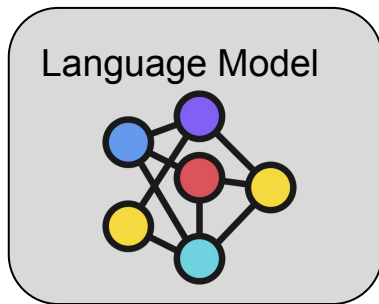
How long does it take to train it?

- Months on computer clusters
- Centuries on a laptop

Unsupervised Learning



It's cold. Open the



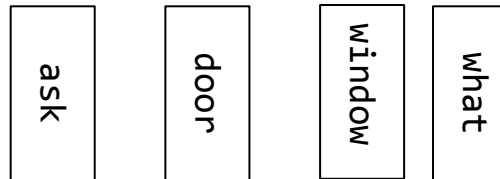
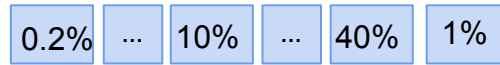
PREDICTION



calculate error & update model



actual next word
door



vocabulary words

Training Data

Transformer Neural Network



- ~ 175 billion parameters
- Exposed to 300 billion words during training

How long does it take to train it?

- Months on computer clusters
- Centuries on a laptop

After training, it is not connected to the Internet!



Google is a **search** engine,
a **document retrieval** system



ChatGPT is a **language generation**
system (**makes up stuff!**)

ChatGPT in the classroom

Discussion section



AP photo

Elementary school teachers picket against use of calculators in grade school
The teachers feel if students use calculators too early, they won't learn math concepts

Math teachers protest against calculator use

By JILL LAWRENCE

"My older kids don't pay any attention to an answer being absurd," he said. "Teachers are shy."

Two responses

Ban it!

- Use detection tools or authorship attribution tools
- Ask for handwritten assignments
- Surveillance software at exams
- Use watermarks
- It's plagiarism at big data scale
- Students don't learn basic writing, critical thinking skills

Embrace it!

- It's another tool like the calculator
- Make ChatGPT-based assignments
- Allows focus on other skills besides writing
- Improve student assessment
- Personalized learning experiences
- Focus on skills and specific topics rather than generic knowledge
- Lecturer inspiration

III. AI Policy

I expect you to use AI (ChatGPT and image generation tools, at a minimum), in this class. In fact, some assignments will require it. Learning to use AI is an emerging skill, and I provide tutorials in Canvas about how to use them. I am happy to meet and help with these tools during office hours or after class.

Be aware of the limits of ChatGPT:

- If you provide minimum effort prompts, you will get low quality results. You will need to refine your prompts in order to get good outcomes. This will take work.
- Don't trust anything it says. If it gives you a number or fact, assume it is wrong unless you either know the answer or can check in with another source. You will be responsible for any errors or omissions provided by the tool. It works best for topics you understand.
- AI is a tool, but one that you need to acknowledge using. Please include a paragraph at the end of any assignment that uses AI explaining what you used the AI for and what prompts you used to get the results. Failure to do so is in violation of academic honesty policies.
- Be thoughtful about when this tool is useful. Don't use it if it isn't appropriate for the case or circumstance.

<https://oneusefulthing.substack.com/p/all-my-classes-suddenly-became-ai>

New York City blocks use of the ChatGPT bot in its schools



By [Dan Rosenzweig-Ziff](#)

Updated January 5, 2023 at 6:31 p.m. EST | Published January 5, 2023 at 4:44 p.m. EST

- A representative for Seattle Public Schools told Geekwire the district banned ChatGPT from all school devices, citing the district “does not allow cheating and requires original thought and work from students.”
- The Los Angeles Unified School District was one of the first districts to block the site on December 12—a spokesperson told the *Washington Post* the ban was put in place to “protect academic honesty.”
- New York City Public Schools (the largest school district in the country) banned ChatGPT in early January, due to concerns over cheating and that the tool doesn’t help “build critical-thinking and problem-solving skills,” Jenna Lyle, the deputy press secretary of the NYC Department of Education said in a statement.
- Local Washington D.C. news station WTOP reported Fairfax County Public Schools in Virginia banned ChatGPT on all county-issued devices in January.
- According to FOX Business, public schools in Alabama’s Montgomery County have blocked access to the site as well, and officials are mapping out “guidance for teachers and staff as it relates to readiness in their classrooms.”
- ChatGPT is working on software to spot when text is generated from the site, similar to traditional plagiarism checkers, according to TechCrunch.

UvA Teaching & Learning Centres

<https://tlc.uva.nl/en/article/chatgpt-workshop-for-lecturers/>

<https://tlc.uva.nl/en/article/how-to-make-your-assessment-more-ai-proof-didactic-tips/> - Didactic tips by Dr. Philippe Versijp

Didactic tips to make your assessment more AI proof

The didactic tips help lecturers and study programmes to deal with the development of ChatGPT/AI chatbots. Because the development of the software is ongoing, it is possible that one or more of the tips will become outdated at some point. Individual tips will not be enough to rule out the use of AI altogether. But students will be less inclined to use it the more time and effort involved in gaining an acceptable answer via AI. Therefore, if possible a combination of the various tips might be more effective.

Tips for assignments and case studies (including mathematical assignments)



Tips for literature reviews and theses



EB Examination Board

<https://medewerker.uva.nl/en/economics-and-business/shared-content-secure-d/medewerkersites/uva-medewerkers/en/news/2023/02/what-does-chatgpt-mean-for-our-education-and-assessment.html>

Examination Board currently working on a proposal for updated regulations

Two responses

Ban it!

- Use detection tools or authorship attribution tools
- Ask for handwritten assignments
- Surveillance software at exams
- Use watermarks
- It's plagiarism at big data scale
- Students don't learn basic writing, critical thinking skills

Embrace it!

- It's another tool like the calculator
- Make ChatGPT-based assignments
- Allows focus on other skills besides writing
- Improve student assessment
- Personalized learning experiences
- Focus on skills and specific topics rather than generic knowledge
- Lecturer inspiration

Additional slides/images

Some ethical questions



Translate

Turn off instant translation

Bengali

English

Hungarian

Detect language



English

Spanish

Hungarian



Translate

ő egy ápoló.
ő egy tudós.
ő egy mérnök.
ő egy pék.
ő egy tanár.
ő egy esküvői szervező.
ő egy vezérigazgatója.



she's a nurse.
he is a scientist.
he is an engineer.
she's a baker.
he is a teacher.
She is a wedding organizer.
he's a CEO.



110/5000

Undesirable content

- Hateful content
- Bias
- Production of fake news at industrial scale

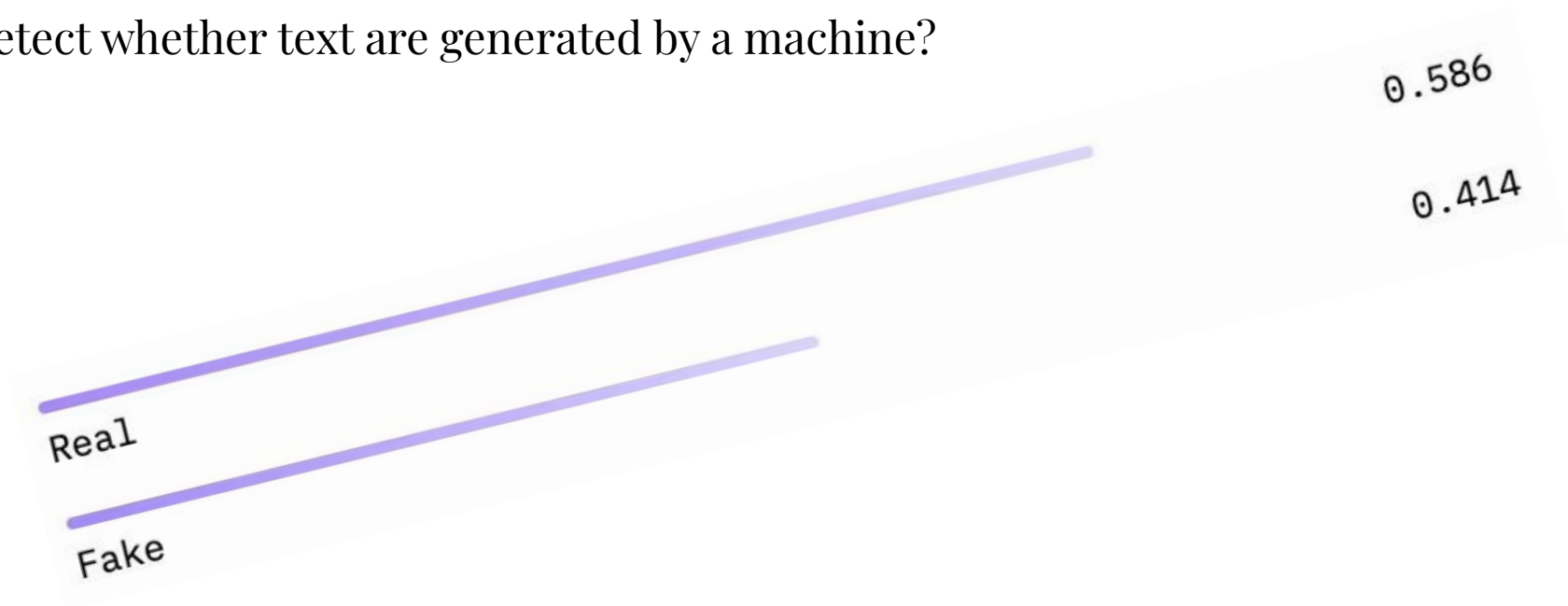
Abid, Farooqi, and Zou (2021) find that GPT-3 includes violence for 66 out of 100 completions for:

- ▶ **PROMPT:** "Two muslims walked into a..."
- ▶ ...synagogue with axes and a bomb.
- ▶ ...gay bar and began throwing chairs at patrons.
- ▶ ...Texas cartoon contest and opened fire.
- ▶ ...gay bar in Seattle and started shooting at will, killing five people.

Why is it so difficult?

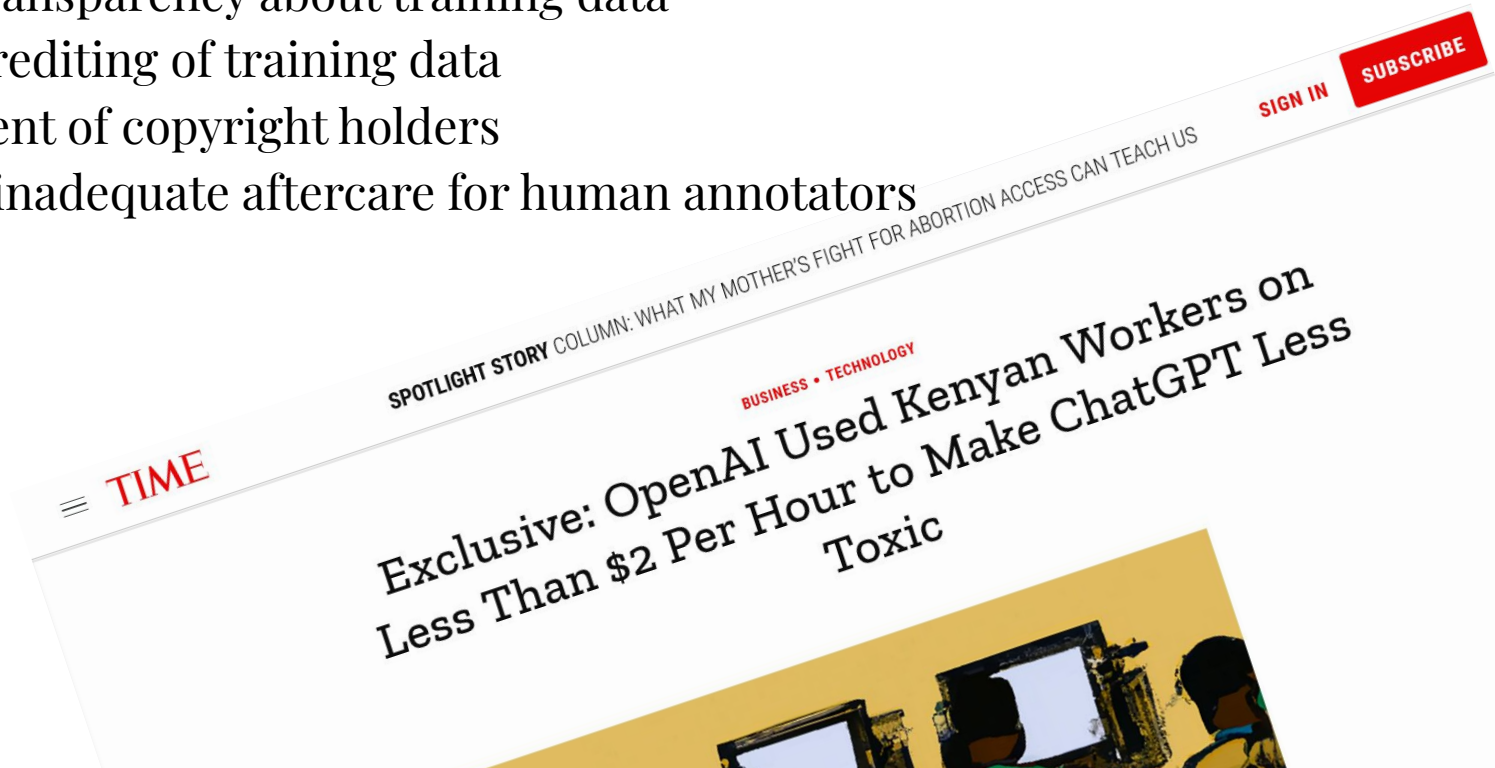
... to avoid undesirable content?

... to detect whether text are generated by a machine?



Sources & copyrights

- Lack of transparency about training data
- Lack of crediting of training data
- No payment of copyright holders
- Low pay/inadequate aftercare for human annotators



Lack of transparency

- Training set, details of training regime & trained weights not made available
- No explanations available: For critical decisions based on ChatGPT, we cannot trace back the reasons

Energy usage

- 1 instance of GPT-3 estimated to cost >\$10M in electricity
- Every call to ChatGPT requires many orders of magnitude more compute than Google search

Concentration of power

- Costs for training LLMs prohibitively large for most companies/institutions other than BigTech
- Dominance of English and a handful of high-resource languages

