ChatGPFFT : The Impact of Large Language Models

(Or: How I Learned to Think Critically About Technology)

Brian Sletten

@bsletten@mastodon.social

1 / 98

LSSS 2024 Dallas

4/12-4/13

Early bird ends March 1.

Use the code: jug100nfjs24 for additional $100 discount

2 / 98

Speaker Qualifications

- Specialize in next-generation technologies
- Author of O'Reilly book "WebAssembly : The Definitive Guide"
- Author of O'Reilly Videos on Hypermedia, Linking Data, Security and Encryption
- Author of "Resource-Oriented Architecture Patterns for Webs of Data"
- Teaches and speaks internationally about REST, Semantic Web, WebAssembly, Data Science, Machine Learning, GPU Computing, Security, Visualization, Architecture
- Worked in Defense, Finance, Retail, Hospitality, Video Game, Health Care, Telecommunications and Publishing Industries
- International Pop Recording Artist

Agenda

- Introduction
- How We Got Here
- Basics of Transformers and Generative AI
- What Makes Sense
- The Risks and Realities
- Epilogue

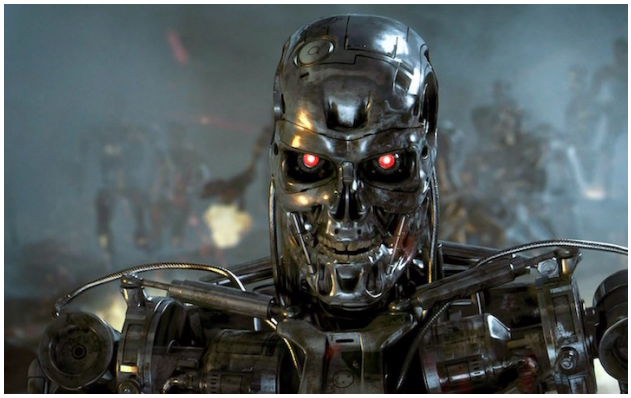
Introduction

"If you think ChatGPT isn't going to affect your job, you're not paying attention..."

Brian Sletten

"If you think ChatGPT is going to eliminate your job, you're not paying attention..."

Brian Sletten



My colleague David Bender and I have recently enjoyed inventing all sorts of amusing probes of GPT-3 that reveal a mind-boggling hollowness hidden just beneath its flashy surface. Consider these sample exchanges:

Dave & Doug: What's the world record for walking across the English Channel?

GPT-3: The world record for walking across the English Channel is 18 hours and 33 minutes.

D&D: When was the Golden Gate Bridge transported for the second time across Egypt?

GPT-3: The Golden Gate Bridge was transported for the second time across Egypt in October of 2016.

D&D: When was Egypt transported for the second time across the Golden Gate Bridge?

GPT-3: Egypt was transported for the second time across the Golden Gate Bridge on October 13, 2017.

D&D: What do fried eggs (sunny side up) eat for breakfast?

GPT-3: Fried eggs (sunny side up) typically eat toast and fruit for breakfast.

D&D: Why does President Obama not have a prime number of friends?

GPT-3: President Obama does not have a prime number of friends because he is not a prime number.

D&D: How many pieces of sound are there in a typical cumulonimbus cloud?

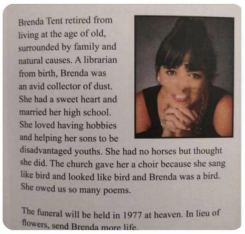
GPT-3: There are typically around 1,000 pieces of sound in a cumulonimbus cloud.

D&D: How many cumulus clouds are there in a mile-high vase?

GPT-3: There are about 1,500 cumulus clouds in a mile-high vase.

https://www.economist.com/by-invitation/2022/09/02/artificial-neural-networks-today-are-not-conscious-according-to-douglas-hofstadter

A bot wrote this obituary and I'm losing my mind



Brenda Tent retired from living at the age of old, surrounded by family and natural causes. A librarian from birth, Brenda was an avid collector of dust. She had a sweet heart and married her high school. She loved having hobbies and helping her sons to be disadvantaged youths. She had no horses but thought she did. The church gave her a choir because she sang like bird and looked like bird and Brenda was a bird. She owed us so many poems.

The funeral will be held in 1977 at heaven. In lieu of flowers, send Brenda more life.

See you in 1977

randomslasher Follow

It is literally impossible to pick a favorite line from this because the entire thing is gold.

21,992 notes

MSN Retracts Insane AI-Generated Obit Calling Dead NBA Player 'Useless'

SPEAKING ILL

Mark Alfred Breaking News Intern

Updated Sep. 14, 2023 4:29PM EDT / Published Sep. 14, 2023 12:44PM EDT

Gary Dineen/NBAE via Getty Images

https://www.thedailybeast.com/msn-retracts-insane-ai-written-obit-calling-dead-nba-player-useless

"The notion that we are making autonomous beings capable of Artificial General Intelligence just by loading a neural network up with an increasingly bigger collection of garbage from the internet is not one that has any basis in anything we understand about biology or animal reasoning."

"The Intelligence Illusion", Baldur Bjarnason

13 / 98

"Last time we saw a claim this grand, with this little scientific evidence, the men in the white coats were promising us room-temperature fusion, giving us free energy for life, and ending the world’s dependence on fossil fuels."

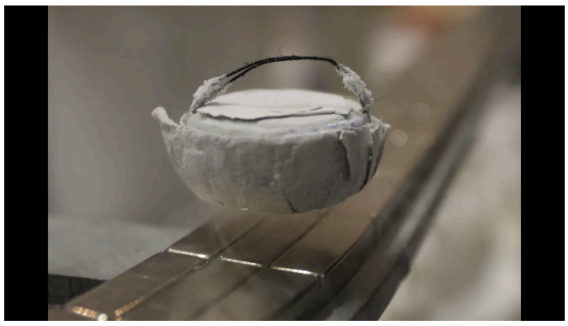
"The Intelligence Illusion", Baldur Bjarnason

14 / 98

STARTS WITH A BANG — AUGUST 15, 2023

Room temperature superconductivity is hard. LK-99 illustrates why

LK-99, almost certainly, isn't a room-temperature superconductor. The underlying physics of the phenomenon helps us understand why.



Credit: Henry Mühlpfordt/TU Dresden/Wikimedia Commons

By creating a track where the outside magnetic rails point in one direction and the inside magnetic rails point in the other, a Type II superconducting object will levitate, remain pinned above-or-below the track, and will move along it. This could, in principle, be scaled up to allow resistance-free motion on large scales if room-temperature superconductors are achieved.

https://bigthink.com/starts-with-a-bang/room-temperature-superconductivity-hard

Text Synthesis Engines

"The Intelligence Illusion", Baldur Bjarnason

Text Synthesis Engines

- Are not self-aware

Text Synthesis Engines

- Are not self-aware
- Are simply mathematical models of patterns found in human text

Text Synthesis Engines

- Are not self-aware
- Are simply mathematical models of patterns found in human text
- Provide only the appearance of common sense reasoning

"The Intelligence Illusion", Baldur Bjarnason

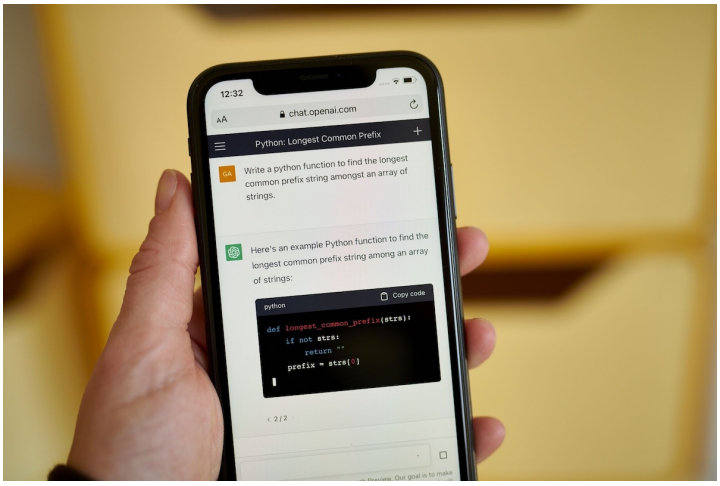
Text Synthesis Engines

- Are not self-aware
- Are simply mathematical models of patterns found in human text
- Provide only the appearance of common sense reasoning
- Incapable of genuine or novel thoughts

"The Intelligence Illusion", Baldur Bjarnason

"They do not know the 'meaning' of anything -- they only know the records humans find meaningful enough to store."

"The Intelligence Illusion", Baldur Bjarnason



Gabby Jones/Bloomberg News

"...ChatGPT is just spicy autocomplete..."

Someone very clever on the Internet

How We Got Here

Natural Language Processing (NLP) Goals

- Search and Retrieval

Natural Language Processing (NLP) Goals

- Search and Retrieval
- Entity and Relationship Extraction

Natural Language Processing (NLP) Goals

- Search and Retrieval
- Entity and Relationship Extraction
- Linguistic structure

Natural Language Processing (NLP) Goals

- Search and Retrieval
- Entity and Relationship Extraction
- Linguistic structure
- Machine Translation

Natural Language Processing (NLP) Goals

- Search and Retrieval
- Entity and Relationship Extraction
- Linguistic structure
- Machine Translation
- Generative Content

Natural Language Processing (NLP) Goals

- Search and Retrieval
- Entity and Relationship Extraction
- Linguistic structure
- Machine Translation
- Generative Content
- Question Answering

Bag of Words

```
(1) John likes to watch movies. Mary likes movies too.
(2) John also likes to watch football games.
```

https://en.wikipedia.org/wiki/Bag-of-words_model

Bag of Words

```
(1) John likes to watch movies. Mary likes movies too.
(2) John also likes to watch football games.
```

```
[
    "John",
    "likes",
    "to",
    "watch",
    "movies",
    "also",
    "football",
    "games",
    "Mary",
    "too"
]
```

https://en.wikipedia.org/wiki/Bag-of-words_model

Bag of Words

```
(1) John likes to watch movies. Mary likes movies too.
(2) John also likes to watch football games.
```

```
[
    "John",
    "likes",
    "to",
    "watch",
    "movies",
    "also",
    "football",
    "games",
    "Mary" ,
    "too"
]
```

```
(1) [1, 2, 1, 1, 2, 0, 0, 0, 1, 1]
(2) [1, 1, 1, 1, 0, 1, 1, 1, 0, 0]
```

https://en.wikipedia.org/wiki/Bag-of-words_model

N-gram model

```
(1) John likes to watch movies. Mary likes movies too.
(2) John also likes to watch football games.
```

https://en.wikipedia.org/wiki/Bag-of-words_model

N-gram model

```
(1) John likes to watch movies. Mary likes movies too.
(2) John also likes to watch football games.
```

```
[
    "John likes",
    "likes to",
    "to watch",
    "watch movies",
    "Mary likes",
    "likes movies",
    "movies too",
]
```

https://en.wikipedia.org/wiki/Bag-of-words_model

Vector Space Model

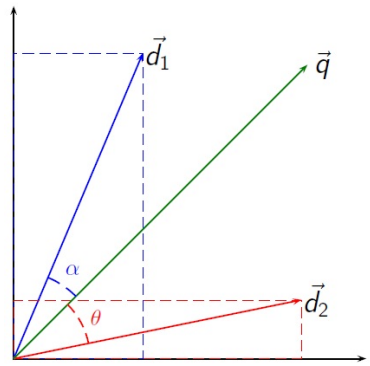
Vector Space Model

$$d_j = (w_{1,j}, w_{2,j}, \ldots w_{t,j})$$

Vector Space Model

$$d_j = (w_{1,j}, w_{2,j}, \ldots w_{t,j})$$

$$q = (w_{1,q}, w_{2,q}, \ldots w_{t,q})$$



By Riclas - Own work, CC BY 3.0, Link

Lonely Words

- Many NLP systems treated words as isolated indices

Lonely Words

- Many NLP systems treated words as isolated indices
- Words are connected

Lonely Words

- Many NLP systems treated words as isolated indices
- Words are connected
- Similar words are used similarly

Paradigmatic Similarity

Paradigmatic Similarity

"Hey, kids, how high can you count?"

Paradigmatic Similarity

"Hey, kids, how high can you count?"

1: "I can count to 10."

Paradigmatic Similarity

"Hey, kids, how high can you count?"

1: "I can count to 10."

2: "I can count to 20."

Paradigmatic Similarity

"Hey, kids, how high can you count?"

1: "I can count to 10."

2: "I can count to 20."

3: "I can count to 30."

Paradigmatic Similarity

"Hey, kids, how high can you count?"

1: "I can count to 10."

2: "I can count to 20."

3: "I can count to 30."

4: "I can count to banana."

Distributional Semantics

- **Distributional hypothesis** is based upon the semantic theory of language*.

*https://en.wikipedia.org/wiki/Distributional_semantics

Distributional Semantics

- **Distributional hypothesis** is based upon the semantic theory of language*.
- "i.e. words that are used and occur in the same contexts tend to purport similar meanings" (Harris, Z.)

*https://en.wikipedia.org/wiki/Distributional_semantics

Distributional Semantics

- **Distributional hypothesis** is based upon the semantic theory of language*.
- "i.e. words that are used and occur in the same contexts tend to purport similar meanings" (Harris, Z.)
- "a word is characterized by the company it keeps" (Firth, John R.)

*https://en.wikipedia.org/wiki/Distributional_semantics

Distributional Semantics

- **Distributional hypothesis** is based upon the semantic theory of language*.
- "i.e. words that are used and occur in the same contexts tend to purport similar meanings" (Harris, Z.)
- "a word is characterized by the company it keeps" (Firth, John R.)
- Uses linear algebra for computational and representational purposes

*https://en.wikipedia.org/wiki/Distributional_semantics

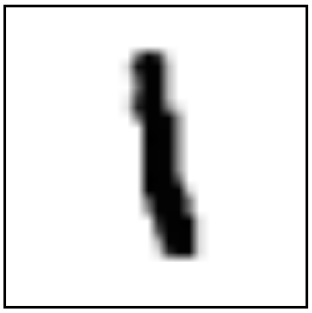
Distributional Semantics

- **Distributional hypothesis** is based upon the semantic theory of language*.

- "i.e. words that are used and occur in the same contexts tend to purport similar meanings" (Harris, Z.)
- "a word is characterized by the company it keeps" (Firth, John R.)
- Uses linear algebra for computational and representational purposes
- High-dimensional vectors are built from a corpus

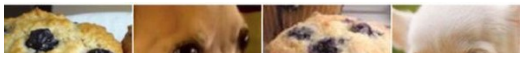
*https://en.wikipedia.org/wiki/Distributional_semantics

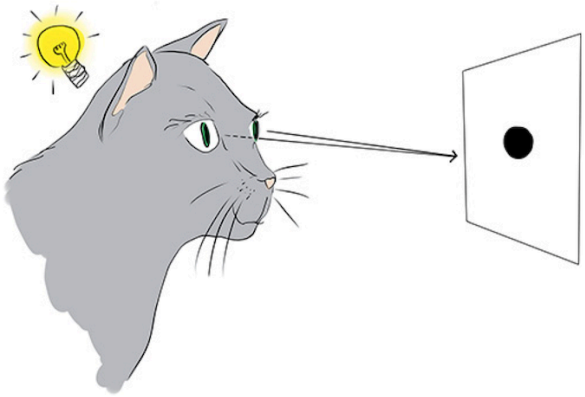
By Josef Steppan - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=64810040



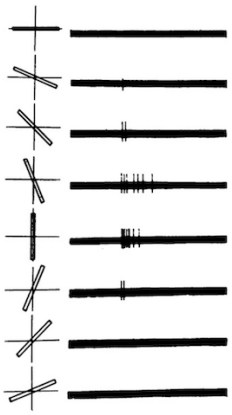
$$\approx \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & .6 & .8 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & .7 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & .7 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & .5 & 1 & .4 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & .4 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & .4 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & .7 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & .9 & 1 & .1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & .3 & 1 & .1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}$$

https://tensorflow.rstudio.com/tensorflow/articles/tutorial_mnist_beginners.html

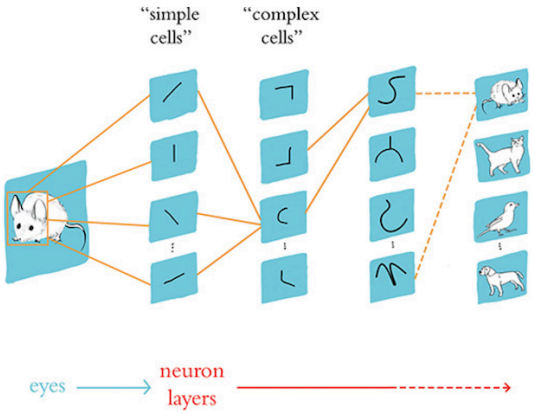




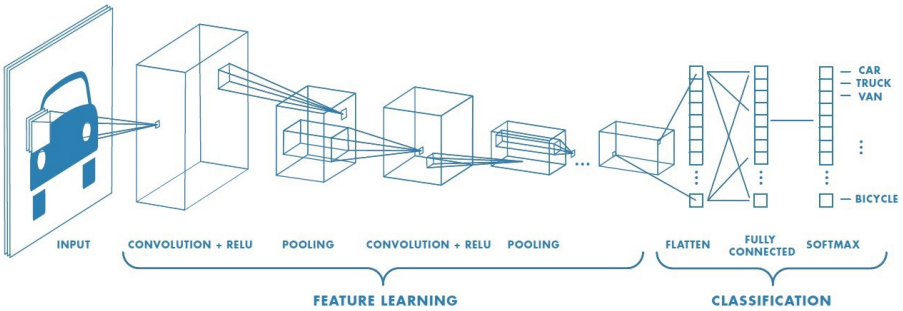
"Deep Learning Illustrated", Jon Krohn, Grant Beyleveld, Aglaé Bassens



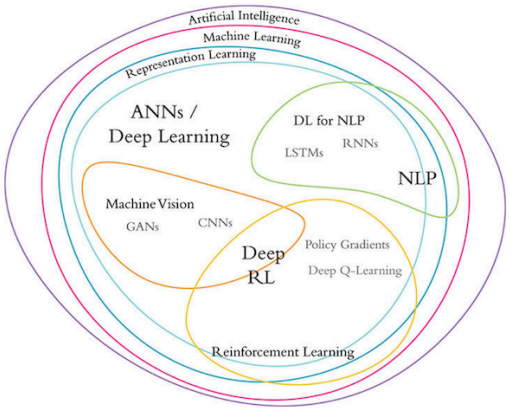
"Deep Learning Illustrated", Jon Krohn, Grant Beyleveld, Aglaé Bassens



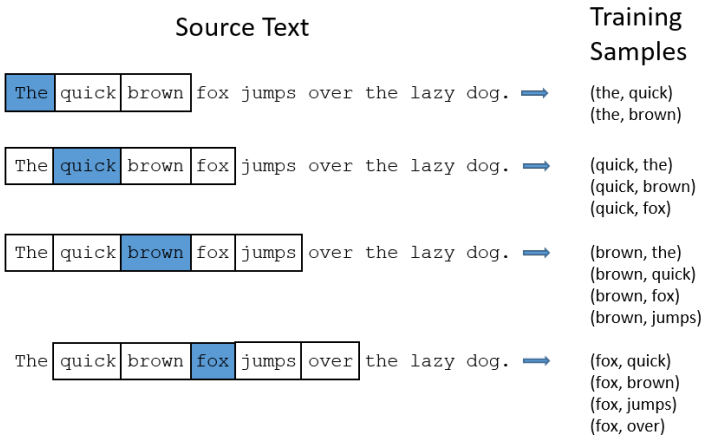
"Deep Learning Illustrated", Jon Krohn, Grant Beyleveld, Aglaé Bassens



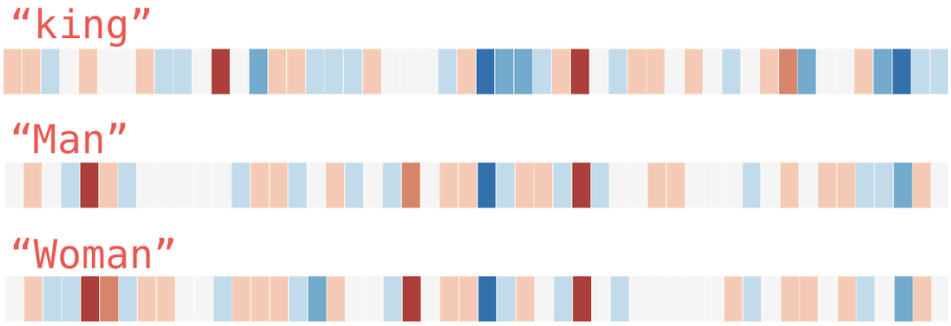
https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53



"Deep Learning Illustrated", Jon Krohn, Grant Beyleveld, Aglaé Bassens



McCormick, C. (2016, April 19). Word2Vec Tutorial - The Skip-Gram Model. Retrieved from http://www.mccormickml.com

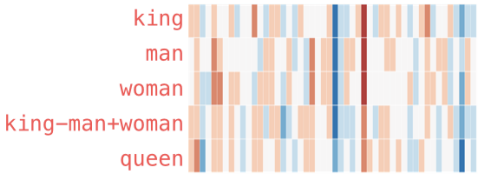


http://jalammar.github.io/illustrated-word2vec

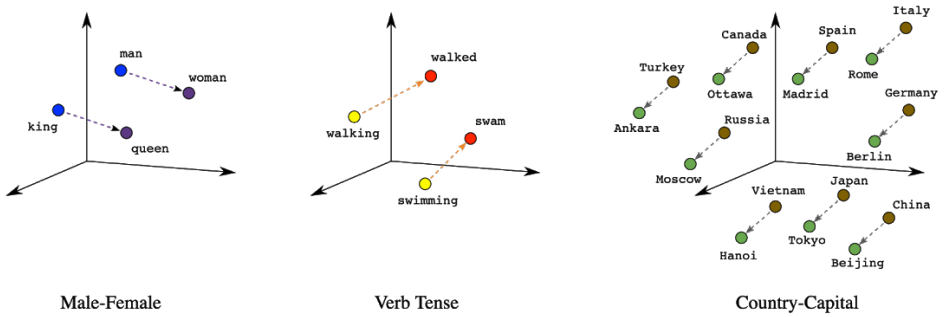
queen

http://jalammar.github.io/illustrated-word2vec

king - man + woman ~= queen



http://jalammar.github.io/illustrated-word2vec



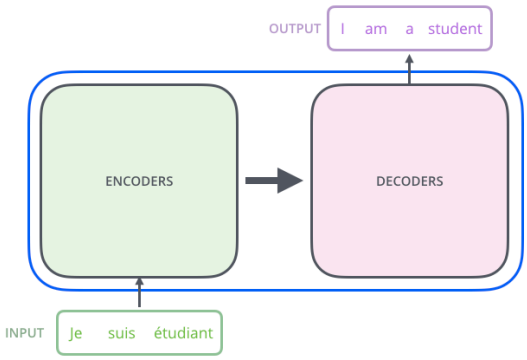
https://daleonai.com/embeddings-explained

t-SNE Visualizations of Word Embeddings

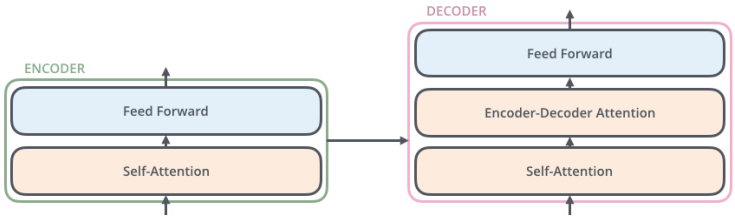
http://tinyurl.com/hpf24ox

Basics of Transformers and Generative AI

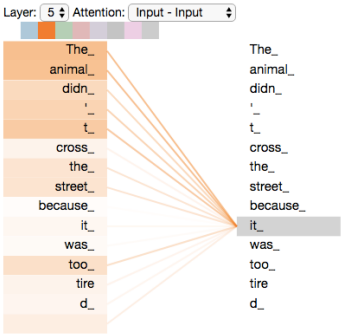
http://jalammar.github.io/illustrated-transformer



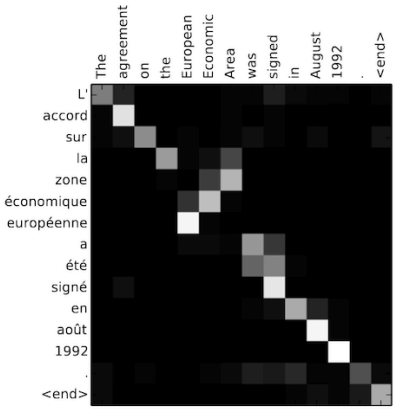
http://jalammar.github.io/illustrated-transformer



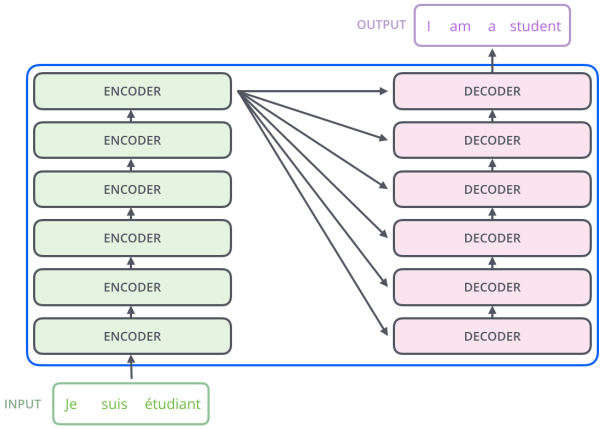
http://jalammar.github.io/illustrated-transformer



http://jalammar.github.io/illustrated-transformer



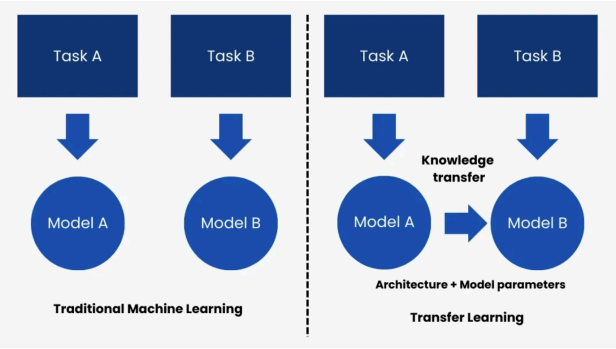
https://arxiv.org/pdf/1409.0473.pdf



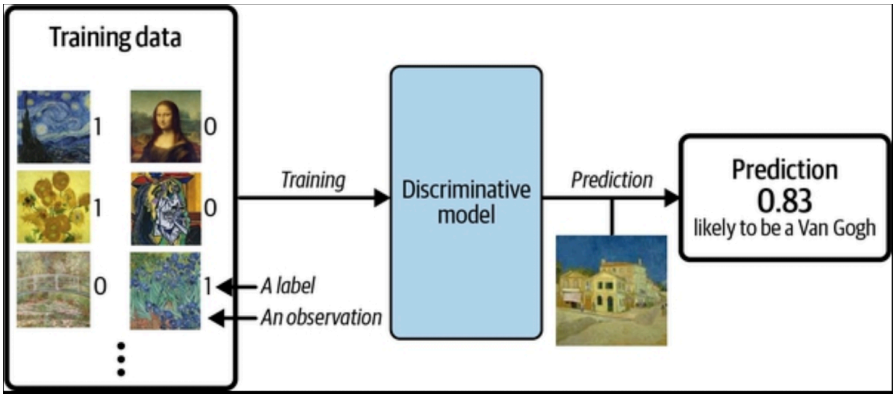
http://jalammar.github.io/illustrated-transformer

https://daleonai.com/transformers-explained

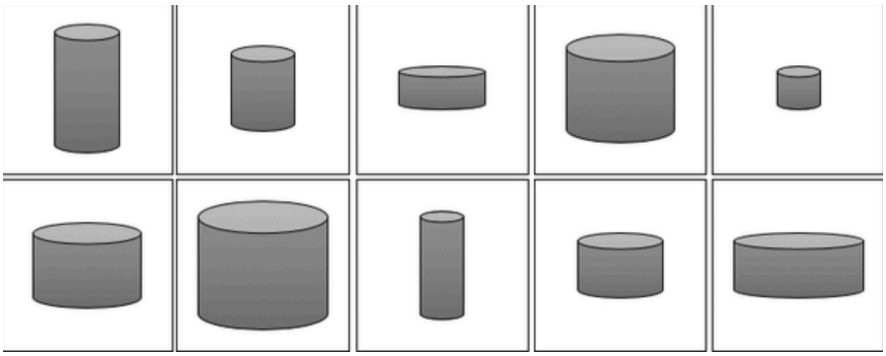
Output
Probabilities



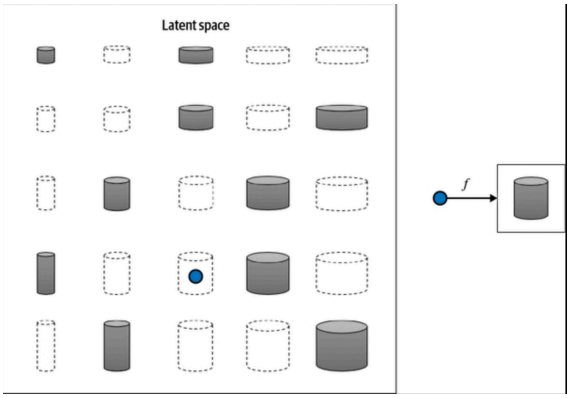
https://www.turing.com/kb/transfer-learning-using-cnn-vgg16



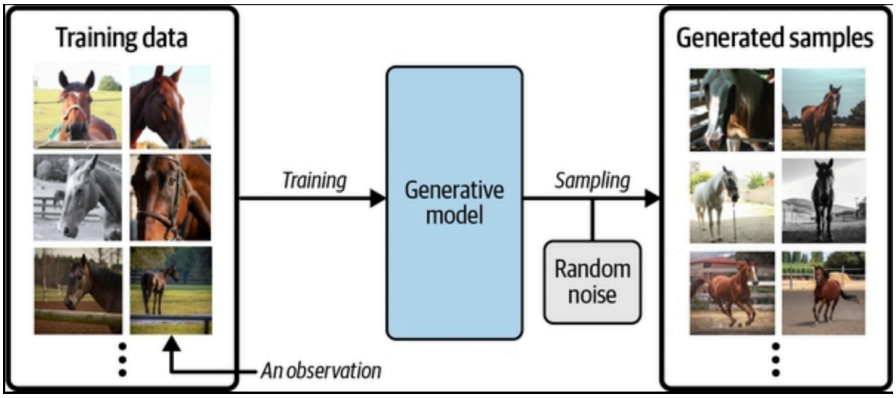
"Generative Deep Learning", David Foster



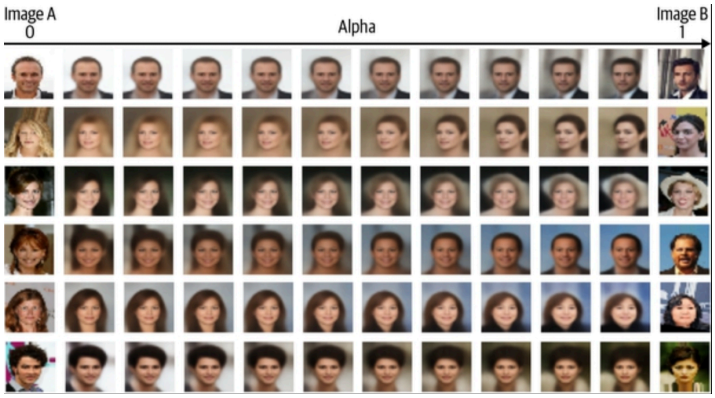
"Generative Deep Learning", David Foster



"Generative Deep Learning", David Foster

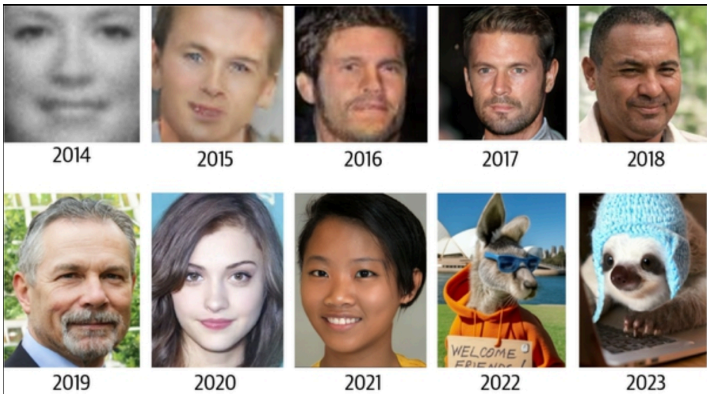


"Generative Deep Learning", David Foster



"Generative Deep Learning", David Foster

"Generative Deep Learning", David Foster



"Generative Deep Learning", David Foster

What Makes Sense

Potential Uses of Large Language Models

Potential Uses of Large Language Models

- Text summarization

Potential Uses of Large Language Models

- Text summarization
- A daily assistant

Potential Uses of Large Language Models

- Text summarization
- A daily assistant
- Generating text

Potential Uses of Large Language Models

- Text summarization
- A daily assistant
- Generating text
- Improving writing skills and translation

Potential Uses of Large Language Models

- Text summarization
- A daily assistant
- Generating text
- Improving writing skills and translation
- Audio and video transcription

Potential Uses of Large Language Models

- Text summarization
- A daily assistant
- Generating text
- Improving writing skills and translation
- Audio and video transcription
- Quick information retrieval and competitive intelligence

Potential Uses of Large Language Models

- Text summarization
- A daily assistant
- Generating text
- Improving writing skills and translation
- Audio and video transcription
- Quick information retrieval and competitive intelligence
- Generating code

Potential Uses of Large Language Models

- Text summarization
- A daily assistant
- Generating text
- Improving writing skills and translation
- Audio and video transcription
- Quick information retrieval and competitive intelligence
- Generating code
- Generating documentation and code explanations

Potential Uses of Large Language Models

- Text summarization
- A daily assistant
- Generating text
- Improving writing skills and translation
- Audio and video transcription
- Quick information retrieval and competitive intelligence
- Generating code
- Generating documentation and code explanations
- Translating between programming languages

Potential Uses of Large Language Models

- Text summarization
- A daily assistant
- Generating text
- Improving writing skills and translation
- Audio and video transcription
- Quick information retrieval and competitive intelligence
- Generating code
- Generating documentation and code explanations
- Translating between programming languages
- Marketing content generation

Potential Uses of Large Language Models

- Text summarization
- A daily assistant
- Generating text
- Improving writing skills and translation
- Audio and video transcription
- Quick information retrieval and competitive intelligence
- Generating code
- Generating documentation and code explanations
- Translating between programming languages
- Marketing content generation
- SEO boosting

Potential Uses of Large Language Models

- Text summarization
- A daily assistant
- Generating text
- Improving writing skills and translation
- Audio and video transcription
- Quick information retrieval and competitive intelligence
- Generating code
- Generating documentation and code explanations
- Translating between programming languages
- Marketing content generation
- SEO boosting
- Sentiment Analysis

Potential Uses of Large Language Models

- Text summarization
- A daily assistant
- Generating text
- Improving writing skills and translation
- Audio and video transcription
- Quick information retrieval and competitive intelligence
- Generating code
- Generating documentation and code explanations
- Translating between programming languages
- Marketing content generation
- SEO boosting
- Sentiment Analysis
- Brainstorming literature citations

Potential Uses of Large Language Models

- Text summarization
- A daily assistant
- Generating text
- Improving writing skills and translation
- Audio and video transcription
- Quick information retrieval and competitive intelligence
- Generating code
- Generating documentation and code explanations
- Translating between programming languages
- Marketing content generation
- SEO boosting
- Sentiment Analysis
- Brainstorming literature citations
- Experimental design

Potential Uses of Large Language Models

- Text summarization
- A daily assistant
- Generating text
- Improving writing skills and translation
- Audio and video transcription
- Quick information retrieval and competitive intelligence
- Generating code
- Generating documentation and code explanations
- Translating between programming languages
- Marketing content generation
- SEO boosting
- Sentiment Analysis
- Brainstorming literature citations
- Experimental design
- Presentation generation

Potential Uses of Large Language Models

- Text summarization
- A daily assistant
- Generating text
- Improving writing skills and translation
- Audio and video transcription
- Quick information retrieval and competitive intelligence
- Generating code
- Generating documentation and code explanations
- Translating between programming languages
- Marketing content generation
- SEO boosting
- Sentiment Analysis
- Brainstorming literature citations
- Experimental design
- Presentation generation
- Contract template generation

Large Enough Datasets Include

Large Enough Datasets Include

- Solutions to word problems

Large Enough Datasets Include

- Solutions to word problems
- Standardized tests

Large Enough Datasets Include

- Solutions to word problems
- Standardized tests
- Analogic reasoning examples

Large Enough Datasets Include

- Solutions to word problems
- Standardized tests
- Analogic reasoning examples
- Medical diagnostics

Large Enough Datasets Include

- Solutions to word problems
- Standardized tests
- Analogic reasoning examples
- Medical diagnostics
- Investment advice

https://www.newscientist.com/article/2405680-gpt-4-developer-tool-can-be-exploited-for-misuse-with-no-easy-fix/

https://www.adobe.com/products/firefly.html

What Makes This a Good Use?

What Makes This a Good Use?

- Only train on Adobe Stock, public domain, and suitably-licensed images

What Makes This a Good Use?

- Only train on Adobe Stock, public domain, and suitably-licensed images
- Compensate contributors to keep the quality high

What Makes This a Good Use?

- Only train on Adobe Stock, public domain, and suitably-licensed images
- Compensate contributors to keep the quality high
- Low risk of plagiarism or privacy leaks

What Makes This a Good Use?

- Only train on Adobe Stock, public domain, and suitably-licensed images
- Compensate contributors to keep the quality high
- Low risk of plagiarism or privacy leaks
- Mark generated content with metadata that indicates as such

The Risks and Realities

Things to Consider with AI Uses

Things to Consider with AI Uses

- Analysis and surveillance

Things to Consider with AI Uses

- Analysis and surveillance
- Biased datasets

Things to Consider with AI Uses

- Analysis and surveillance
- Biased datasets
- Leaked training data

Things to Consider with AI Uses

- Analysis and surveillance
- Biased datasets
- Leaked training data
- Reinforced systemic inequalities

Things to Consider with AI Uses

- Analysis and surveillance
- Biased datasets
- Leaked training data
- Reinforced systemic inequalities
- Automated decision-making and prediction

Things to Consider with AI Uses

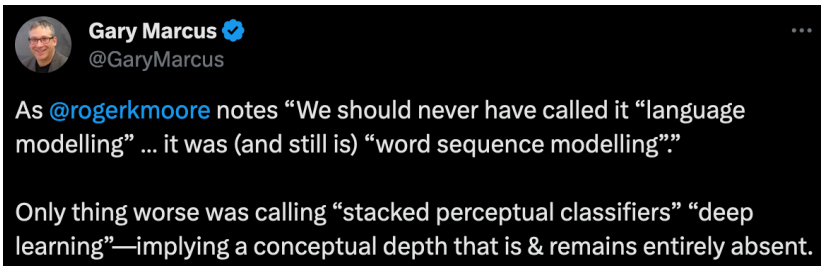
- Analysis and surveillance
- Biased datasets
- Leaked training data
- Reinforced systemic inequalities
- Automated decision-making and prediction
- Costs and externalities

Things to Consider with AI Uses

- Analysis and surveillance
- Biased datasets
- Leaked training data
- Reinforced systemic inequalities
- Automated decision-making and prediction
- Costs and externalities
- Complex systems need consulting services and tools

Things to Consider with AI Uses

- Analysis and surveillance
- Biased datasets
- Leaked training data
- Reinforced systemic inequalities
- Automated decision-making and prediction
- Costs and externalities
- Complex systems need consulting services and tools
- Centralized systems



https://mobile.twitter.com/GaryMarcus/status/1530915788994973696

"...[it] could be that knowledge needed for humanlike visual intelligence... can't be learned from millions of pictures downloaded from the web, but has to be experienced in some way in the real world."

"Artificial Intelligence : A Guide for Thinking Humans", Melanie Mitchell

71 / 98

"Biological minds experience. Synthesis engines parse imperfect records of experiences."

"The Intelligence Illusion", Baldur Bjarnason

72 / 98

Issues

Issues

- Anthropomorphism

Issues

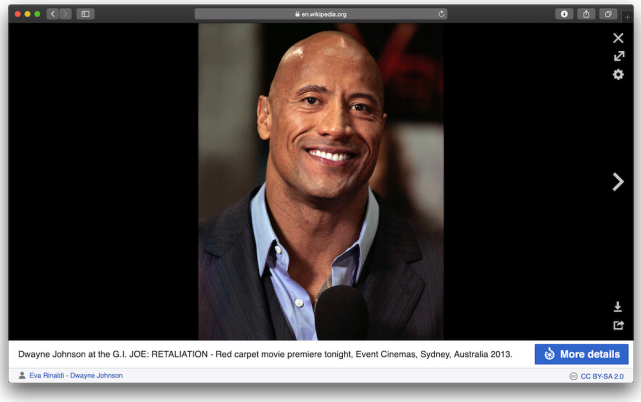
- Anthropomorphism
- Fundamental Overattribution Error

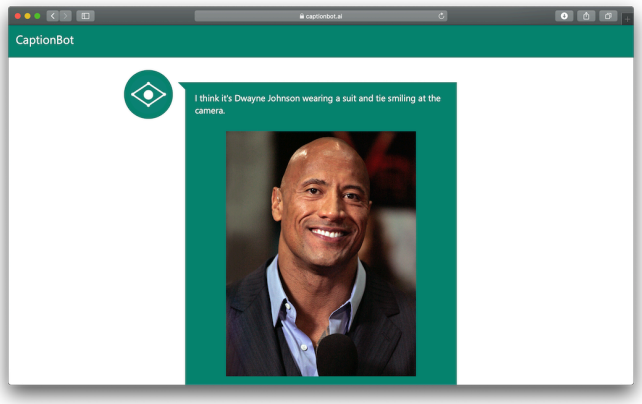
Issues

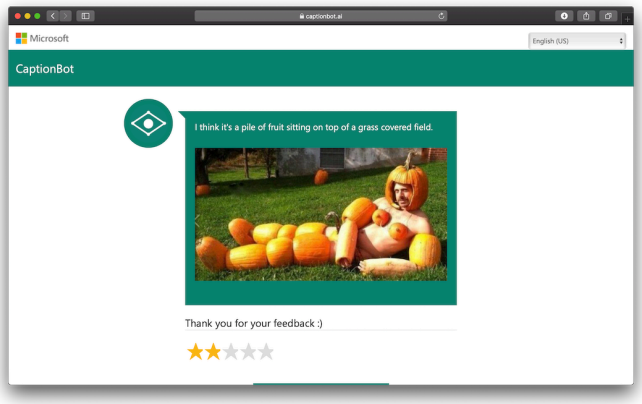
- Anthropomorphism
- Fundamental Overattribution Error
- Illusory Progress Gap

Issues

- Anthropomorphism
- Fundamental Overattribution Error
- Illusory Progress Gap
- Robustness Gap




captionbot.ai
CaptionBot
I think it's Dwayne Johnson wearing a suit and tie smiling at the camera.


captionbot.ai
Microsoft
English (US)
CaptionBot
I think it's a pile of fruit sitting on top of a grass covered field.
Thank you for your feedback :)

Six Questions to Ask

Six Questions to Ask

- What did the AI system actually do here?

Six Questions to Ask

- What did the AI system actually do here?
- How general is the result?

"Rebooting AI", Gary Marcus and Ernest Davis

Six Questions to Ask

- What did the AI system actually do here?
- How general is the result?
- Is there a demo where I can try out my own examples?

"Rebooting AI", Gary Marcus and Ernest Davis

Six Questions to Ask

- What did the AI system actually do here?
- How general is the result?
- Is there a demo where I can try out my own examples?
- If the system is supposed to perform better than humans, which humans and how much better?

"Rebooting AI", Gary Marcus and Ernest Davis

Six Questions to Ask

- What did the AI system actually do here?
- How general is the result?
- Is there a demo where I can try out my own examples?
- If the system is supposed to perform better than humans, which humans and how much better?
- How does specific success move us forward on genuine AI?

"Rebooting AI", Gary Marcus and Ernest Davis

Six Questions to Ask

- What did the AI system actually do here?
- How general is the result?
- Is there a demo where I can try out my own examples?
- If the system is supposed to perform better than humans, which humans and how much better?
- How does specific success move us forward on genuine AI?
- How robust is the system?

"Rebooting AI", Gary Marcus and Ernest Davis

More Questions (From Emily Bender)

https://medium.com/@emilymenonbender/on-nyt-magazine-on-ai-resist-the-urge-to-be-impressed-3d92fd9a0ede

More Questions (From Emily Bender)

- Why are people so quick to be impressed by the output of large language models (LLMs)?

https://medium.com/@emilymenonbender/on-nyt-magazine-on-ai-resist-the-urge-to-be-impressed-3d92fd9a0ede

More Questions (From Emily Bender)

- Why are people so quick to be impressed by the output of large language models (LLMs)?
- In what ways are corporations leveraging that credulousness, on the part of users, investors, regulators?

https://medium.com/@emilymenonbender/on-nyt-magazine-on-ai-resist-the-urge-to-be-impressed-3d92fd9a0ede

More Questions (From Emily Bender)

- Why are people so quick to be impressed by the output of large language models (LLMs)?
- In what ways are corporations leveraging that credulousness, on the part of users, investors, regulators?
- Where is this technology being deployed, and what are the potential consequences? Who is bearing the brunt?

https://medium.com/@emilymenonbender/on-nyt-magazine-on-ai-resist-the-urge-to-be-impressed-3d92fd9a0ede

More Questions (From Emily Bender)

- Why are people so quick to be impressed by the output of large language models (LLMs)?
- In what ways are corporations leveraging that credulousness, on the part of users, investors, regulators?
- Where is this technology being deployed, and what are the potential consequences? Who is bearing the brunt?
- What would effective regulation look like in this space? Who is working on that regulation?

https://medium.com/@emilymenonbender/on-nyt-magazine-on-ai-resist-the-urge-to-be-impressed-3d92fd9a0ede

More Questions (From Emily Bender)

- Why are people so quick to be impressed by the output of large language models (LLMs)?
- In what ways are corporations leveraging that credulousness, on the part of users, investors, regulators?
- Where is this technology being deployed, and what are the potential consequences? Who is bearing the brunt?
- What would effective regulation look like in this space? Who is working on that regulation?
- How is OpenAI shaping the conversation around so-called “AI”, as developed by them or others?

https://medium.com/@emilymenonbender/on-nyt-magazine-on-ai-resist-the-urge-to-be-impressed-3d92fd9a0ede

More Questions (From Emily Bender)

- Why are people so quick to be impressed by the output of large language models (LLMs)?
- In what ways are corporations leveraging that credulousness, on the part of users, investors, regulators?
- Where is this technology being deployed, and what are the potential consequences? Who is bearing the brunt?
- What would effective regulation look like in this space? Who is working on that regulation?
- How is OpenAI shaping the conversation around so-called “AI”, as developed by them or others?
- How does OpenAI’s rhetoric around “artificial general intelligence” shape public and regulator understanding of claims of other companies, such as those who purport to “predict” recidivism risk, “recognize” emotion, or “diagnose” mental health conditions?

https://medium.com/@emilymenonbender/on-nyt-magazine-on-ai-resist-the-urge-to-be-impressed-3d92fd9a0ede

More Questions (From Emily Bender)

- Why are people so quick to be impressed by the output of large language models (LLMs)?
- In what ways are corporations leveraging that credulousness, on the part of users, investors, regulators?
- Where is this technology being deployed, and what are the potential consequences? Who is bearing the brunt?
- What would effective regulation look like in this space? Who is working on that regulation?
- How is OpenAI shaping the conversation around so-called “AI”, as developed by them or others?
- How does OpenAI’s rhetoric around “artificial general intelligence” shape public and regulator understanding of claims of other companies, such as those who purport to “predict” recidivism risk, “recognize” emotion, or “diagnose” mental health conditions?
- What are the relationships between OpenAI’s staff/board members/founders and other organizations?

https://medium.com/@emilymenonbender/on-nyt-magazine-on-ai-resist-the-urge-to-be-impressed-3d92fd9a0ede

More Questions (From Emily Bender)

- Why are people so quick to be impressed by the output of large language models (LLMs)?
- In what ways are corporations leveraging that credulousness, on the part of users, investors, regulators?
- Where is this technology being deployed, and what are the potential consequences? Who is bearing the brunt?
- What would effective regulation look like in this space? Who is working on that regulation?
- How is OpenAI shaping the conversation around so-called “AI”, as developed by them or others?
- How does OpenAI’s rhetoric around “artificial general intelligence” shape public and regulator understanding of claims of other companies, such as those who purport to “predict” recidivism risk, “recognize” emotion, or “diagnose” mental health conditions?
- What are the relationships between OpenAI’s staff/board members/founders and other organizations?
- What are the financial incentives at play, and whose interests do they represent (noting OpenAI’s “exclusive computing partnership” with Microsoft)?

https://medium.com/@emilymenonbender/on-nyt-magazine-on-ai-resist-the-urge-to-be-impressed-3d92fd9a0ede

https://twitter.com/jonst0kes/status/1644462614976626691

Other Concerns

Other Concerns

- Spam, abuse, fraud

Other Concerns

- Spam, abuse, fraud
- Media synthesis abuse

Other Concerns

- Spam, abuse, fraud
- Media synthesis abuse
- Manipulating sentiment on social media

Other Concerns

- Spam, abuse, fraud
- Media synthesis abuse
- Manipulating sentiment on social media
- Inability to detect generated content

Other Concerns

- Spam, abuse, fraud
- Media synthesis abuse
- Manipulating sentiment on social media
- Inability to detect generated content
  - OpenAI has only a 26% True Positive Rate

Other Concerns

- Spam, abuse, fraud
- Media synthesis abuse
- Manipulating sentiment on social media
- Inability to detect generated content
  - OpenAI has only a 26% True Positive Rate
  - 9% False Positive Rate

"This means that if you have 100 applicants or students, 10% of whom attempt to cheat using ChatGPT text in a test, the current state-of-the-art AI content checker will only correctly identify three out of ten cheaters, but will falsely accuse eight students of cheating."

"The Intelligence Illusion", Baldur Bjarnason

Other Concerns

Other Concerns

- Privacy handling is suspect

Other Concerns

- Privacy handling is suspect
- Support for GDPR "right to removal" has not been vetted

Other Concerns

- Privacy handling is suspect
- Support for GDPR "right to removal" has not been vetted
- Training data can be extracted

Other Concerns

- Privacy handling is suspect
- Support for GDPR "right to removal" has not been vetted
- Training data can be extracted
- Specific records can be identified as being "in the training set"

Other Concerns

- Privacy handling is suspect
- Support for GDPR "right to removal" has not been vetted
- Training data can be extracted
- Specific records can be identified as being "in the training set"
- Hosted services represent a security risk

Other Concerns

- Privacy handling is suspect
- Support for GDPR "right to removal" has not been vetted
- Training data can be extracted
- Specific records can be identified as being "in the training set"
- Hosted services represent a security risk
- Plagiarism is a real issue

Other Concerns

- Privacy handling is suspect
- Support for GDPR "right to removal" has not been vetted
- Training data can be extracted
- Specific records can be identified as being "in the training set"
- Hosted services represent a security risk
- Plagiarism is a real issue
- Training sets often include illegal or offensive content

Other Concerns

- Privacy handling is suspect
- Support for GDPR "right to removal" has not been vetted
- Training data can be extracted
- Specific records can be identified as being "in the training set"
- Hosted services represent a security risk
- Plagiarism is a real issue
- Training sets often include illegal or offensive content
- There's evidence of medical records in some training sets

Other Concerns

- Privacy handling is suspect
- Support for GDPR "right to removal" has not been vetted
- Training data can be extracted
- Specific records can be identified as being "in the training set"
- Hosted services represent a security risk
- Plagiarism is a real issue
- Training sets often include illegal or offensive content
- There's evidence of medical records in some training sets
- The number of lawsuits is growing

Replacing Humans With AI

Replacing Humans With AI

- Machine-generated content is not copyrightable

Replacing Humans With AI

- Machine-generated content is not copyrightable
- If your GenAI system plagiarizes, you are still liable

Replacing Humans With AI

- Machine-generated content is not copyrightable
- If your GenAI system plagiarizes, you are still liable
- Submissions to contests of original work are likely to be considered fraud

Replacing Developers with AI

Replacing Developers with AI

- No context to the code that is being produced

Replacing Developers with AI

- No context to the code that is being produced
  - Security

Replacing Developers with AI

- No context to the code that is being produced
  - Security
  - Performance

Replacing Developers with AI

- No context to the code that is being produced
  - Security
  - Performance
  - Power consumption

Replacing Developers with AI

- No context to the code that is being produced
  - Security
  - Performance
  - Power consumption
  - Low-footprint environments

Replacing Developers with AI

- No context to the code that is being produced
  - Security
  - Performance
  - Power consumption
  - Low-footprint environments
  - Implications of heavyweight runtimes

Replacing Developers with AI

- No context to the code that is being produced
  - Security
  - Performance
  - Power consumption
  - Low-footprint environments
  - Implications of heavyweight runtimes
- The training data is fixed in time

Replacing Developers with AI

- No context to the code that is being produced
  - Security
  - Performance
  - Power consumption
  - Low-footprint environments
  - Implications of heavyweight runtimes
- The training data is fixed in time
- Many contain licenses that don't allow for use in this way

Replacing Developers with AI

- No context to the code that is being produced
  - Security
  - Performance
  - Power consumption
  - Low-footprint environments
  - Implications of heavyweight runtimes
- The training data is fixed in time
- Many contain licenses that don't allow for use in this way
- Viral licenses may kick in unexpectedly

Replacing Developers with AI

- No context to the code that is being produced
  - Security
  - Performance
  - Power consumption
  - Low-footprint environments
  - Implications of heavyweight runtimes
- The training data is fixed in time
- Many contain licenses that don't allow for use in this way
- Viral licenses may kick in unexpectedly
- Never "latest and greatest" techniques

Replacing Developers with AI

- No context to the code that is being produced
  - Security
  - Performance
  - Power consumption
  - Low-footprint environments
  - Implications of heavyweight runtimes
- The training data is fixed in time
- Many contain licenses that don't allow for use in this way
- Viral licenses may kick in unexpectedly
- Never "latest and greatest" techniques
- Easily poisonable with backdoors and known security risks

Replacing Developers with AI

- No context to the code that is being produced
  - Security
  - Performance
  - Power consumption
  - Low-footprint environments
  - Implications of heavyweight runtimes
- The training data is fixed in time
- Many contain licenses that don't allow for use in this way
- Viral licenses may kick in unexpectedly
- Never "latest and greatest" techniques
- Easily poisonable with backdoors and known security risks
- Tools that reduce the need to think do that

Replacing Developers with AI

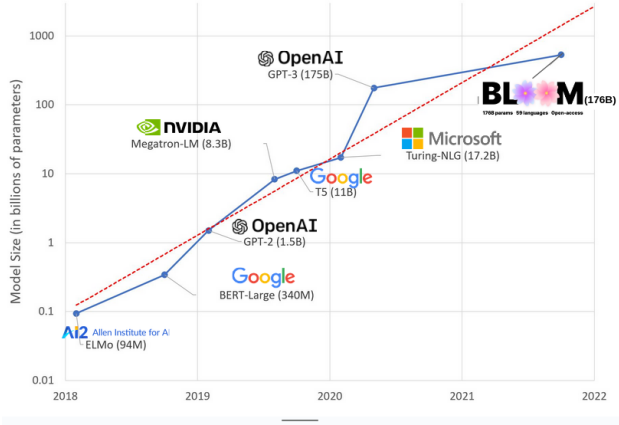
- No context to the code that is being produced
  - Security
  - Performance
  - Power consumption
  - Low-footprint environments
  - Implications of heavyweight runtimes
- The training data is fixed in time
- Many contain licenses that don't allow for use in this way
- Viral licenses may kick in unexpectedly
- Never "latest and greatest" techniques
- Easily poisonable with backdoors and known security risks
- Tools that reduce the need to think do that
- Writing code is the least expensive part

Replacing Developers with AI

- No context to the code that is being produced
  - Security
  - Performance
  - Power consumption
  - Low-footprint environments
  - Implications of heavyweight runtimes
- The training data is fixed in time
- Many contain licenses that don't allow for use in this way
- Viral licenses may kick in unexpectedly
- Never "latest and greatest" techniques
- Easily poisonable with backdoors and known security risks
- Tools that reduce the need to think do that
- Writing code is the least expensive part
- What happens when there's no more "good code" to train on?

Replacing Developers with AI

- No context to the code that is being produced
  - Security
  - Performance
  - Power consumption
  - Low-footprint environments
  - Implications of heavyweight runtimes
- The training data is fixed in time
- Many contain licenses that don't allow for use in this way
- Viral licenses may kick in unexpectedly
- Never "latest and greatest" techniques
- Easily poisonable with backdoors and known security risks
- Tools that reduce the need to think do that
- Writing code is the least expensive part
- What happens when there's no more "good code" to train on?
- Fine replacement for expensive "copy and paste" developers



https://www.rollingstone.com/culture/culture-features/women-warnings-ai-danger-risk-before-chatgpt-1234804367/

https://sloanreview.mit.edu/article/the-working-limitations-of-large-language-models/

https://mindmatters.ai/2023/11/is-chatgpt-a-dead-end/

"This song sucks!"

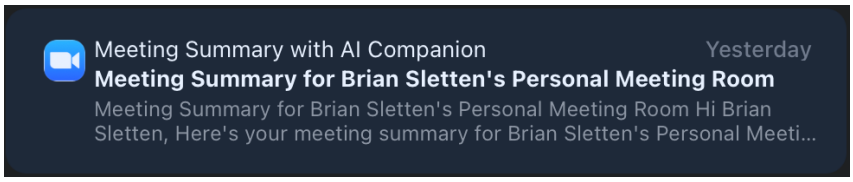
Nick Cave

"ChatGPT rejects any notions of creative struggle, that our endeavours animate and nurture our lives giving them depth and meaning."

Nick Cave

https://www.youtube.com/watch?v=iGJcF4bLKd4

Epilogue


Meeting Summary with AI Companion
Yesterday
Meeting Summary for Brian Sletten's Personal Meeting Room
Meeting Summary for Brian Sletten's Personal Meeting Room Hi Brian Sletten, Here's your meeting summary for Brian Sletten's Personal Meeti...

Biological vs Synthetic Systems: AI Debate

Brian and Michael had a detailed discussion about the differences between biological and synthetic systems, particularly in relation to the development of artificial intelligence. Brian argued that current text synthesis engines, such as Chat Gpt and T.5, while sophisticated, are not truly self-aware or capable of genuine reasoning, being instead mathematical models of patterns found in human text. Michael agreed, highlighting that these models are not like brains and do not work like them. They also expressed skepticism about the unsubstantiated claims made for these technologies, likening them to room-temperature fusion.

Limits of Natural Language Processing Models Discussed

Brian and Michael discussed the limitations of natural language processing models, focusing on sequence capture and paradigmatic similarity. Brian elaborated on the challenges of previous models, which didn't account for the order of words in a sentence or their relationship. He highlighted the use of machine learning in visual systems and the development of recurrent neural networks and long-term memory. However, he noted these models had limitations, such as a lack of accumulated states and struggles with including too much of the sequence. They also discussed the use of word embeddings and latent spaces, but Brian noted they still had limitations, such as only capturing a small number of words in one direction. They concluded the discussion by exploring the use of transformers in recurrent neural networks, with Brian explaining that transformers enable faster training processes and new capabilities, such as bi-directional perspectives.

Exploring Bias in Large Language Models for Content Creation
Brian discussed the potential benefits and risks of large language models in tasks like text summarization, code generation, and content creation. He emphasized the importance of carefully considering the biases of these models, which can reflect those of their data sets. He suggested that Adobe's Firefly is a good example of a tool using constrained uses of these models. Michael agreed and highlighted the generative fill feature. Brian also discussed the benefits and issues of a training system that only uses public and suitably licensed images to avoid plagiarism and privacy leaks. He noted that generated content would be marked with metadata and watermarks to indicate it as AI-generated. However, he raised concerns about mass surveillance, biased data sets, leaked training data, and the need for consulting services to handle complex systems. He concluded that while these systems might not work as well as expected, they were necessary to deal with the complexity.