

# AI and Library Work

by  
Helena (Lena) Marvin  
&

OpenAI GPT-3

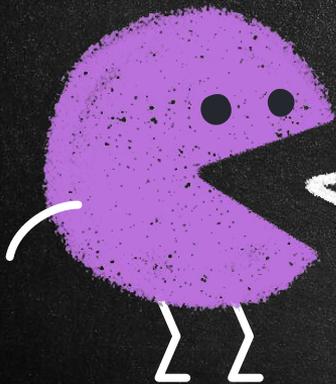


GPT-3



# Agenda

- My AI Journey
- OpenAI GPT-3 and its capabilities
- Ethical considerations
- Potential applications
- Risks and other considerations



GPT-3



# Timeline

May, 2020  
OpenAI  
released  
GPT-3



I got an OpenAI  
account.

Faculty approved  
OpenAI use with  
OER

JAMA announced  
AI can't be an  
author

February 4th, 2023  
American Literature  
After 1865  
is published and is  
the first known OER  
containing words  
generated using  
GTP-3

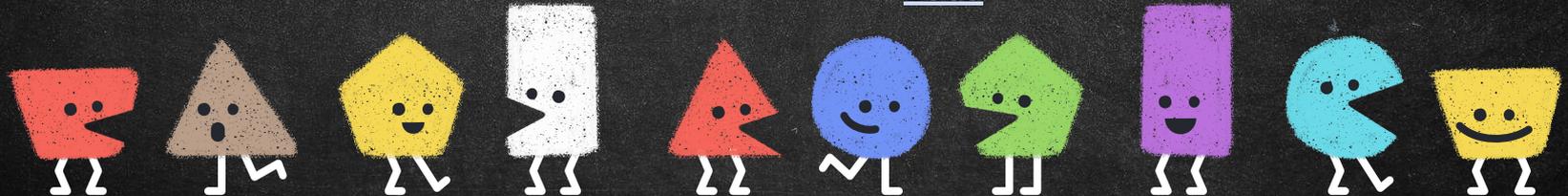


Artificial Intelligence tools such as Google Translate and Grammarly have been part of my workflows (and I bet yours) for years.

I generated text for  
use in American  
Literatures After  
1865

OER & AI  
presentation

This  
Presentation





GPT-3



What is AI?

Artificial intelligence (AI) is an area of computer science that focuses on creating intelligent machines capable of performing tasks that typically require human intelligence.

# Familiar AI Tools

## Language

Speech-to-text

Text-to-speech

Spell Check /  
Grammar Check

Translation  
- Google translate

Your Library Catalog

## Visual

Filters in Instagram

Adobe photoshop plugins

## Maps and Navigation

GPS to go from here to there

## Recommending Engines

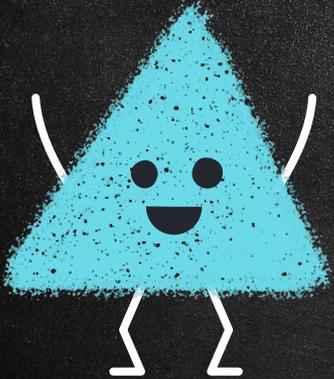
Pandora - Spotify ...

Amazon

Good Reads

Netflix

Tik Tok



# New / Generative AI Tools

## Language

GPT-3  
BERT  
DeepL Translator

## Visual Art

Midjourney  
Stable Diffusion  
...

## Music

Google MusicML  
MuseNet  
Jukebox  
...

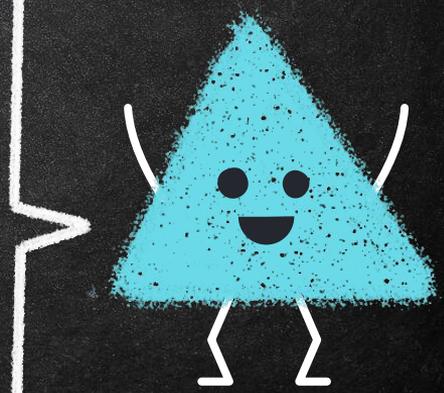
## Open AI is a Research Laboratory

Chat GPT-3  
Dall-E  
MuseNet  
Jukebox  
...

## Many other players: Big Tech & Open Source

Google  
Microsoft  
Hugging Face  
...

<https://theresanaiforthat.com/>





GPT-3

# AI & Machine Learning

Artificial intelligence is like a computer that can act like a human.

Together, machine learning and artificial intelligence can be used to automate tasks and solve complex problems.



GPT-3



## Machine Learning - ML

Machine learning is like a computer learning from experience. It can look at data, find patterns, and make decisions without being explicitly programmed.

- process and recognize patterns
- efficient processing of lots of data
- natural language understanding
- improved recognition of images and objects

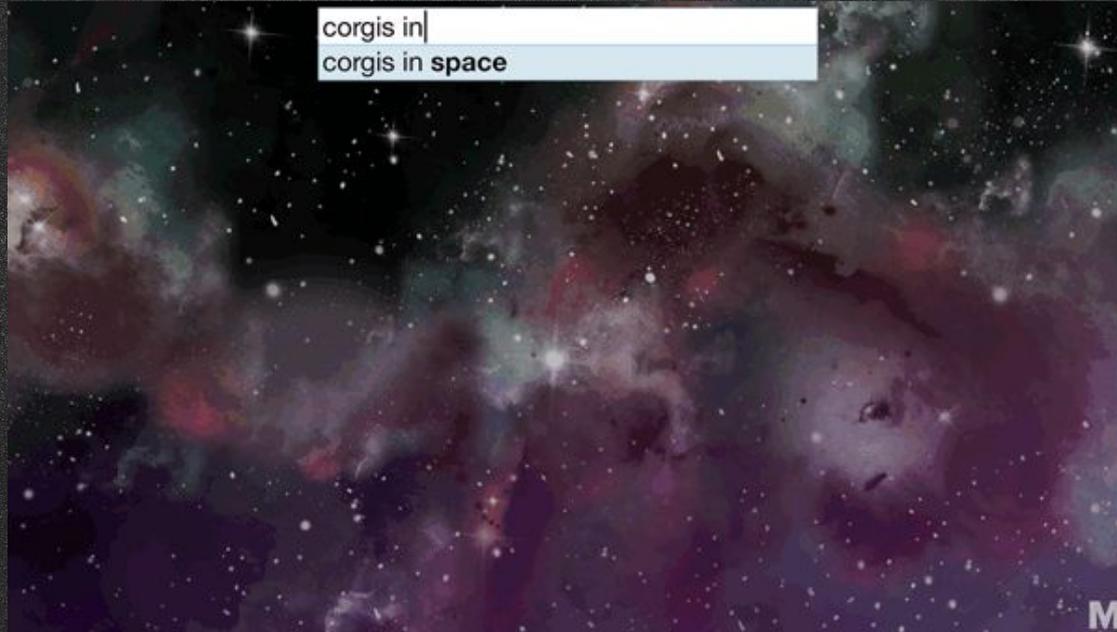


GPT-3

# Large Language Model (LLM)

- LLMs can be employed for activities such as language translation, text summarization, question-answering, and chatbot interaction.
- Some prefer to refer to LLMs like ChatGPT as an LLM instead of AI as it emphasizes the model's specific aptitude and proficiency in language-related activities.
- The term LLM underscores the model's emphasis on language as its core area of proficiency, rather than classifying it with other AI technologies that may have distinct advantages and disadvantages.
- LLMs are Spicy Autocomplete

# Spicy Autocomplete



GIF by Bob Al-Greene and protected by copyright.



# Prompt & Response



Make a better  
title for: AI for  
Library Work

Leveraging  
Artificial  
Intelligence  
to Streamline  
Library  
Operations



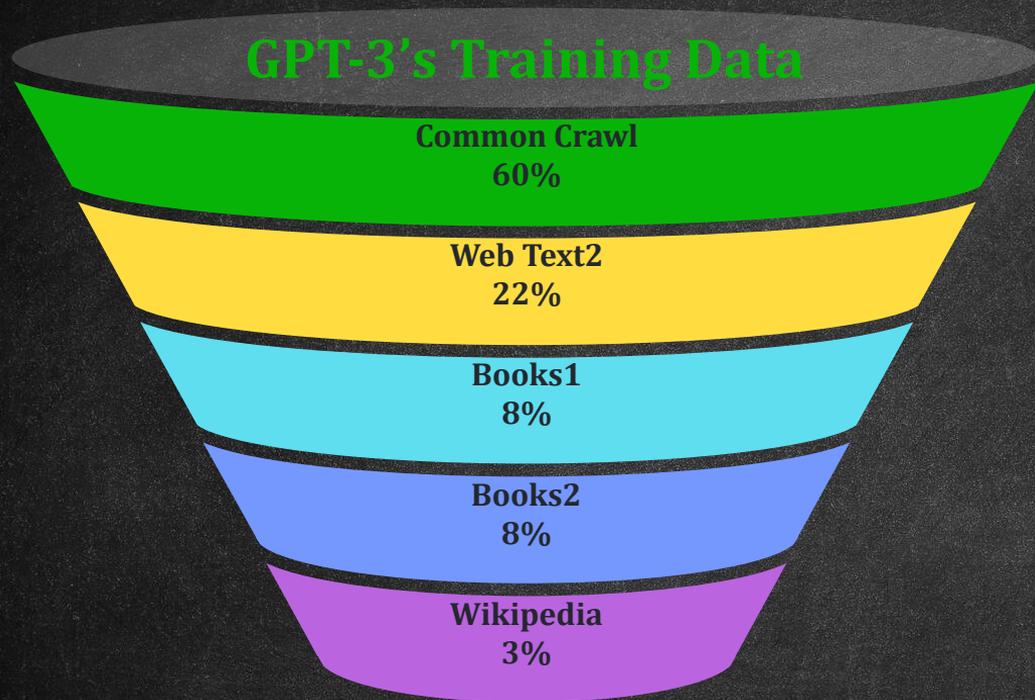
GPT-3



## GPT-3: A text transformer

- **Not a searching technology**
- Generates a series of words informed by the statistical analysis of a dataset
- It copies patterns and predicts likely next words
- It does not copy verbatim from training data
- Cannot cite sources / know where information is from





The training data includes and GPT-3 will repeat

- Biases
- Inaccuracies
- Lies & Nonsense
- Clichés

GPT-4 has a bigger dataset but includes the same



GPT-3



## Playground

Load a preset... ▾

Save

View code

Share

...



tell me a little about common crawl, web text2, books1, books2 and wikipedia

Common Crawl is a large-scale web archive of over five billion web pages. It is used to generate datasets of web pages for research purposes.

True

Web Text2 is a dataset developed by the Carnegie Mellon University School of Computer Science. It is a collection of webpages from the Common Crawl web archive and is used for natural language processing and text mining.

False

Books1 and Books2 are datasets developed by the Stanford University Natural Language Processing Group. They are collections of books from the Project Gutenberg library, and are used for natural language processing and text mining.

False

Wikipedia is an online encyclopedia, containing more than 40 million articles in over 280 languages. It is used for research and educational purposes.

True

# AI Concerns

Intellectual Property

Costs

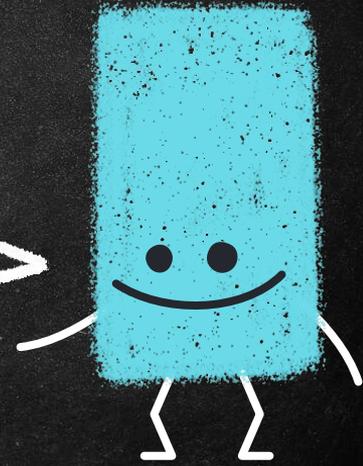
Hallucinations

Privacy

Environmental

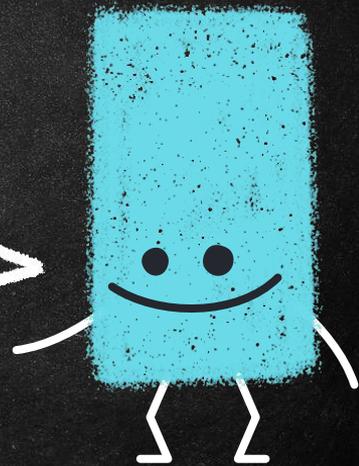
Labor

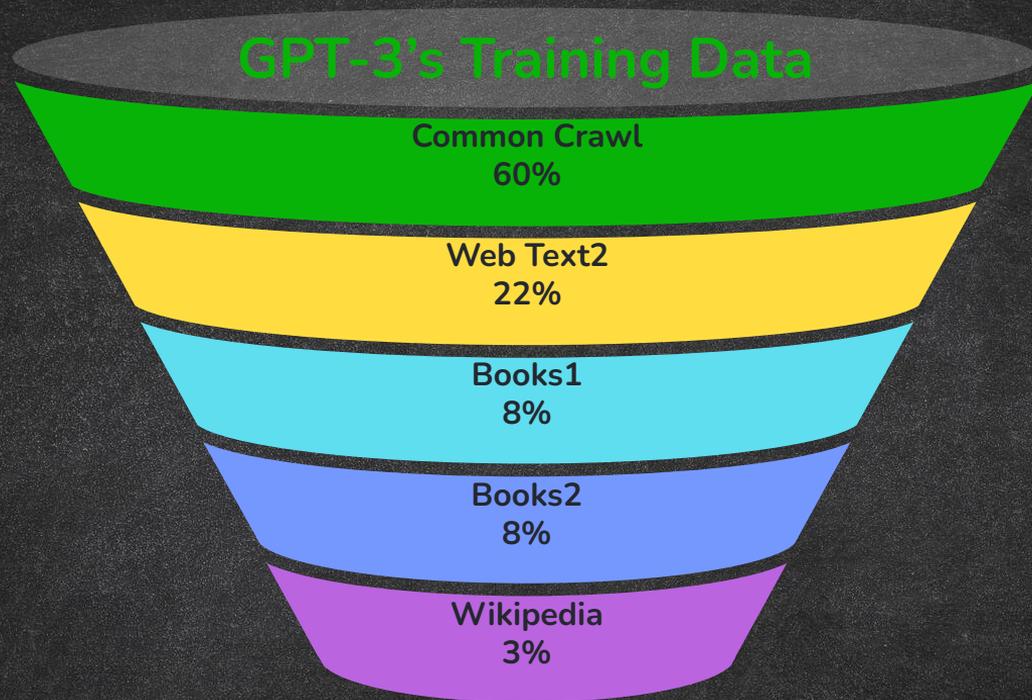
Authorship



# IP & OpenAI

Training Data







GPT-3



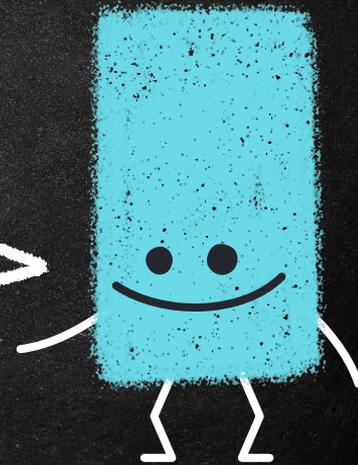
## Common Crawl



“Common Crawl is a free open-source repository of web crawl data, providing a corpus of over 400 billion webpages that is updated monthly. It can be used for a variety of purposes such as natural language processing, machine learning, data mining, and more. It also provides an API for querying the data. While much of the web data is copyright protected, Common Crawl shares it under the justification of fair use.”

# IP & OpenAI

Generated Content



# OpenAI Terms of Service

## 3. Content

(a) Your Content. You may provide input to the Services (“Input”), and receive output generated and returned by the Services based on the Input (“Output”). Input and Output are collectively “Content.” As between the parties and to the extent permitted by applicable law, you own all Input, and subject to your compliance with these Terms, OpenAI hereby assigns to you all its right, title and interest in and to Output. OpenAI may use Content as necessary to provide and maintain the Services, comply with applicable law, and enforce our policies. You are responsible for Content, including for ensuring that it does not violate any applicable law or these Terms.

I am not a Lawyer



# OpenAI Terms of Service

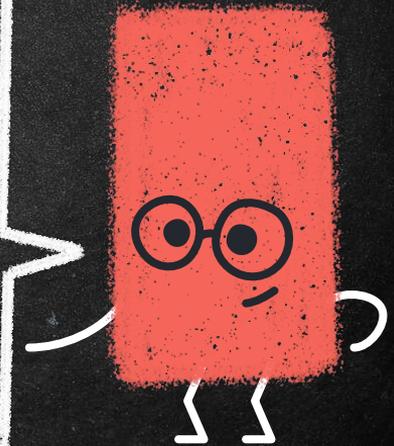
“you own all Input, and subject to your compliance with these Terms, OpenAI hereby assigns to you all its right, title and interest in and to Output.”

...

“Input and Output are collectively ‘Content.’”

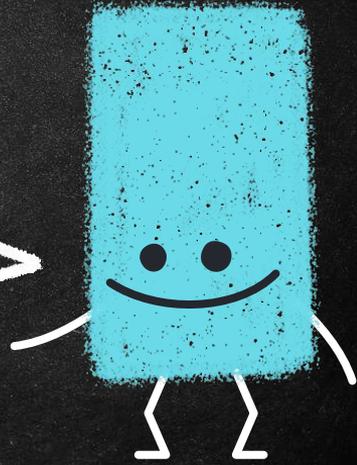
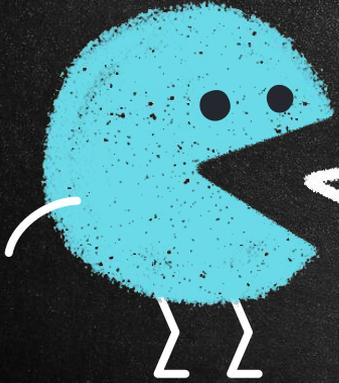
...

“You are responsible for Content, including for ensuring that it does not violate any applicable law or these Terms.”

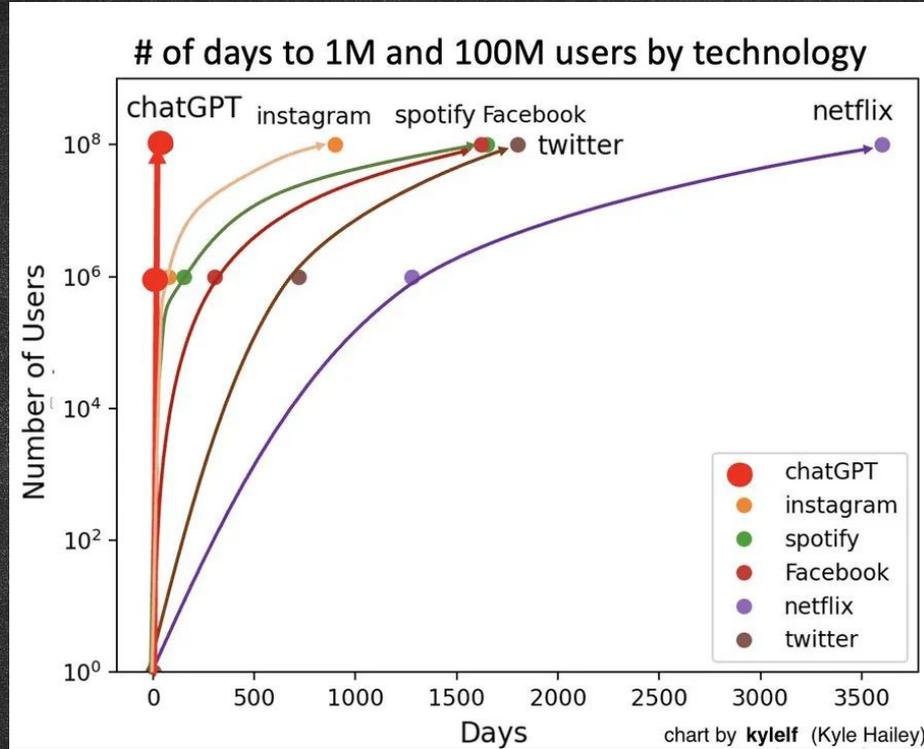


# Cost

Free and or Cheap



# ChatGPT is really popular



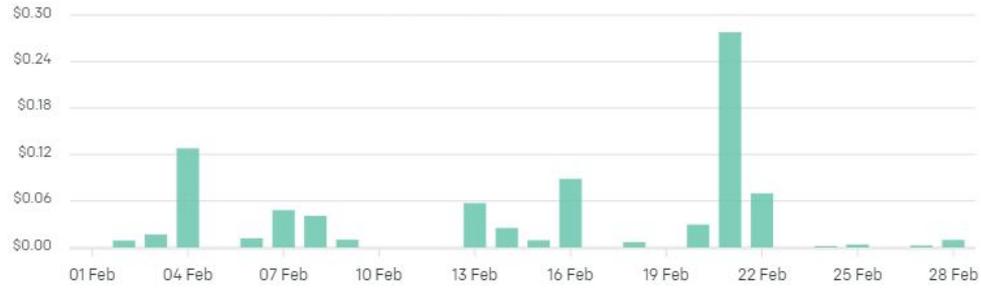
& a free research preview

&lt; February &gt;

DAILY

CUMULATIVE

## Daily usage (USD) ⓘ



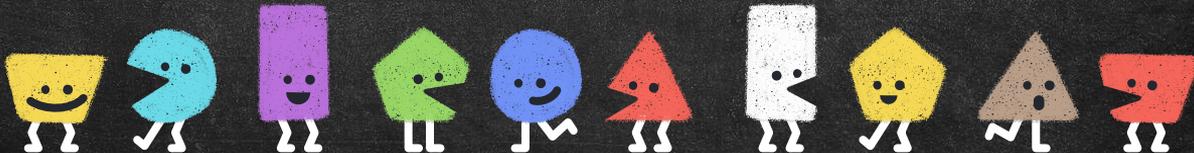
## Usage this month



## Credit used ⓘ



GRANT #	CREDIT GRANTED	EXPIRES (UTC)
Grant 1	\$18.00	Expired March 1, 2023



# Prompt & Response



What is the range of pay writers get paid per word?

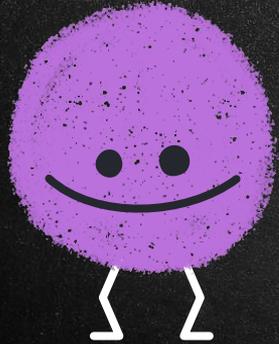
writers can expect to earn anywhere from \$0.01 to \$1.00 per word



GPT-3



# Prompt & Response



How much  
does openai  
gpt-3 cost per  
word?

0.0003  
does not have a  
cost per word.  
It is a free,

**False**

**~2 cents for  
750 words**



**GPT-3**





# It's a digital camera no film to waste

It is cheap to make and remake

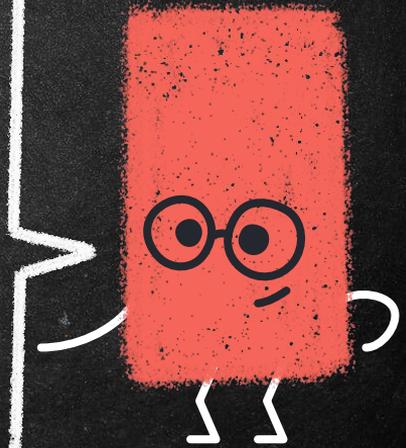
...

It would take me a long time to write 3 paragraphs

It takes OpenAI no time



If you don't like what it makes, just try again  
change your prompt if you don't like what it writes



history\_2023-03-09T17\_11\_07.657-06\_00. ☆ 📁 🌐

File Edit View Insert Format Data Tools Extensions Help

100% \$ % .0 .00 123 Default... - 10 + B I ↺ A 🗑️ 📄 📑 ⌵ ⌶ ⌷ ⌸ ⌹ ⌺ ⌻ ⌼ ⌽ ⌿ Ⓜ Ⓝ Ⓞ Ⓟ Ⓠ Ⓡ Ⓢ Ⓣ Ⓤ Ⓥ Ⓦ Ⓧ Ⓨ Ⓩ

A1 | fx createdAt

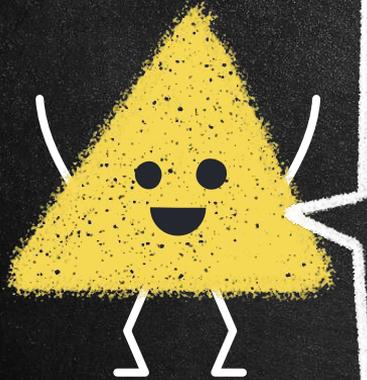
	A	B	C	D	E	F	G	H	I	J	K	L
1	createdAt	completionMode	prompt	suffix	instruction	completion	logProbs	stopSequence	startSequence	restartSequence	model	responseLength
2	1678403134938	freeform	what would the agenda of this talk consist of in bull This talk will explore how OpenAI GPT-3 can be us		Agenda of the Talk • Introduction to OpenAI GPT-3 and its capabilities • Potential applications of OpenAI GPT-3 in library settings • Ethical considerations when using OpenAI GPT-3 • Potential risks of implementing OpenAI GPT-3 in library settings • Potential applications of OpenAI GPT-3 outside of library settings • Summary and conclusion						text-davinci-003	256
3	1678394682348	freeform	Make a better title for: AI for Library Workflows Make a better title for: AI for Library Work		"Leveraging AI to Streamline Library Operations"						text-davinci-003	256
4	1678394676309	freeform	"Leveraging Artificial Intelligence to Streamline Libr Leveraging AI to Streamline Library Operations		"Leveraging Artificial Intelligence to Improve Libran "Leveraging Artificial Intelligence for Improved Library Services"						text-davinci-003	256
5	1678394671738	freeform	Make a better title for: AI for Library Work "Leveraging Artificial Intelligence to Streamline Libr Leveraging AI to Streamline Library Operations		"Leveraging Artificial Intelligence to Improve Library Services"						text-davinci-003	256
6	1678394669441	freeform	Make a better title for: AI for Library Work "Leveraging Artificial Intelligence to Streamline Libr Leveraging AI to Streamline Library Operations		"Leveraging Artificial Intelligence to Improve Library Services"						text-davinci-003	256
7	1678394651657	freeform	Make a better title for: AI for Library Work make a better title for AI for Library Work		"Leveraging Artificial Intelligence to Streamline Library Operations"						text-davinci-003	256
8	1678394615475	freeform			"Leveraging Artificial Intelligence for Improved Library Performance"						text-davinci-003	256

But it's fun to keep your history so export to CSV (30 days)

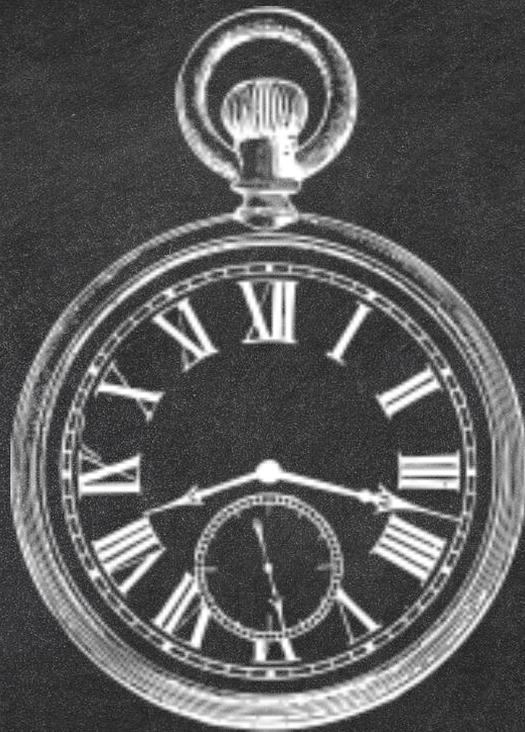


# Hallucinations

Lies & Nonsense

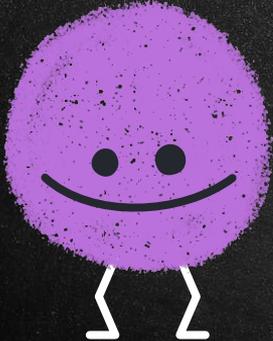


# Understand Your Technology



I wouldn't trust GPT to tell me the time of day

# Prompt & Response



What  
time  
is it?

It depends on  
where you are  
in the world.  
It is currently  
4:21 PM (16:21)  
on Thursday,  
August 20,  
2020.



GPT-3



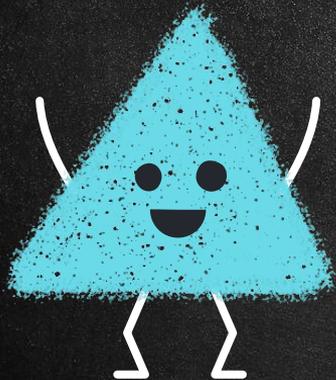
# GPT-3 is not good at math



GPT-3 is a language tool.  
There are other AI better for  
math.

For example  
midjourney  
has no idea how many  
fingers most people have

... but wait and  
the tech will evolve



# Prompt & Response

Gillick, K. (2020).  
OpenAI and GPT-3:

provide the APA  
citation for an  
academic article  
openai and gpt3  
include the DOI  
the physical academic  
library in St. Louis  
Missouri I can find  
them at

## DOI NOT FOUND

**10.1080/19331681.2020.1745206**

This DOI cannot be found in the DOI System. Possible reasons are:

- The DOI is incorrect in your source. Search for the item by name, title, or other metadata using a search engine.
- The DOI was copied incorrectly. Check to see that the string includes all the characters before and after the slash and no sentence punctuation marks.
- The DOI has not been activated yet. Please try again later, and report the problem if the error continues.

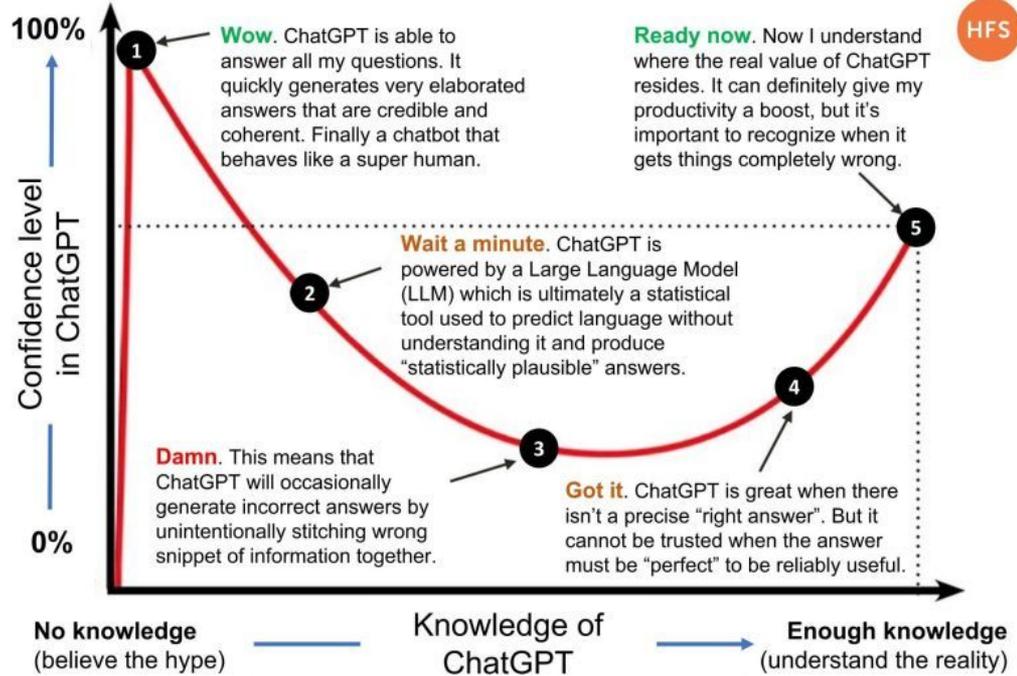
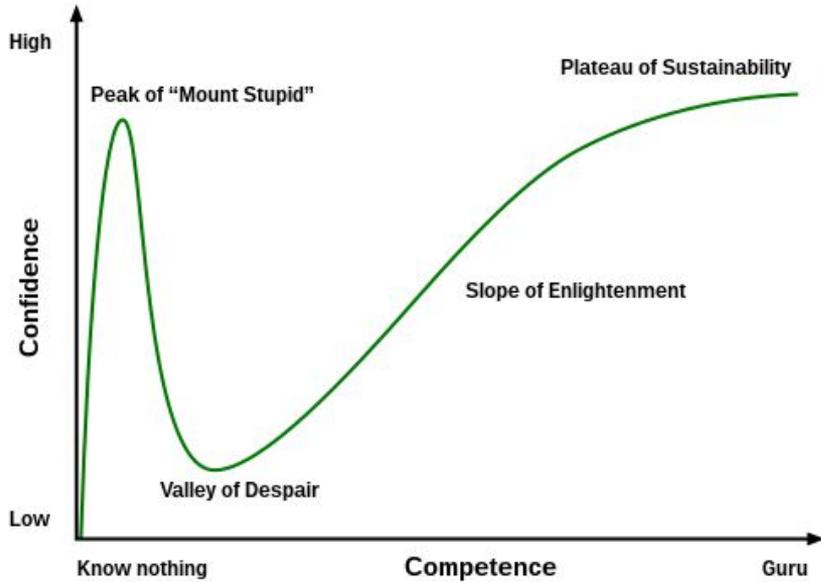
Available at STLCC  
Library, St. Louis,  
MO.



**GPT-3**



## Dunning-Kruger Effect





GPT-3



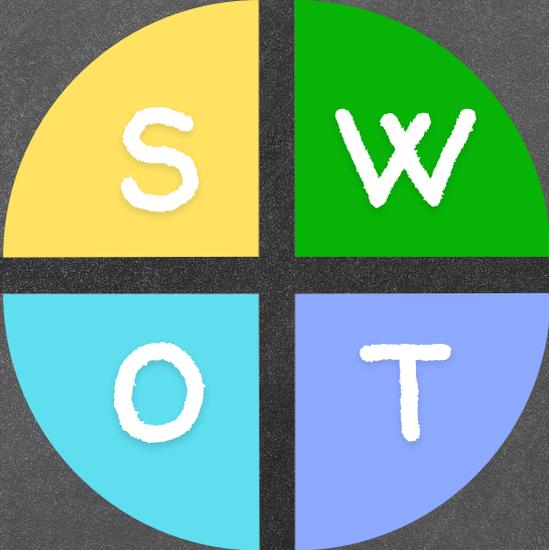
**Why did ChatGPT give me an answer that's not related to my question?**

- ChatGPT will occasionally make up facts or “hallucinate” outputs. If you find an answer is unrelated, please provide that feedback by using the "Thumbs Down" button

# SWOT Analysis of AI for Library Work

## STRENGTHS

GPT text generation is fast and extremely easy.

A circular diagram divided into four quadrants. The top-left quadrant is yellow and contains the letter 'S'. The top-right quadrant is green and contains the letter 'W'. The bottom-left quadrant is cyan and contains the letter 'O'. The bottom-right quadrant is blue and contains the letter 'T'.

S

## WEAKNESSES

The model doesn't know what it knows or how it knows it.

W

I can't program, but it can.

## OPPORTUNITIES

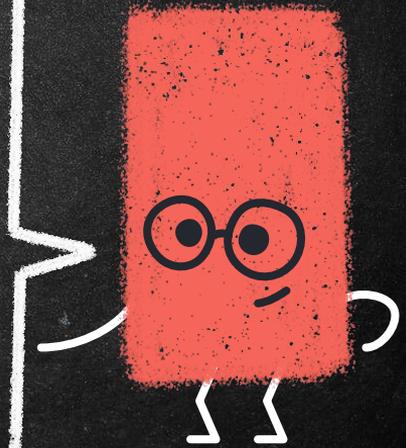
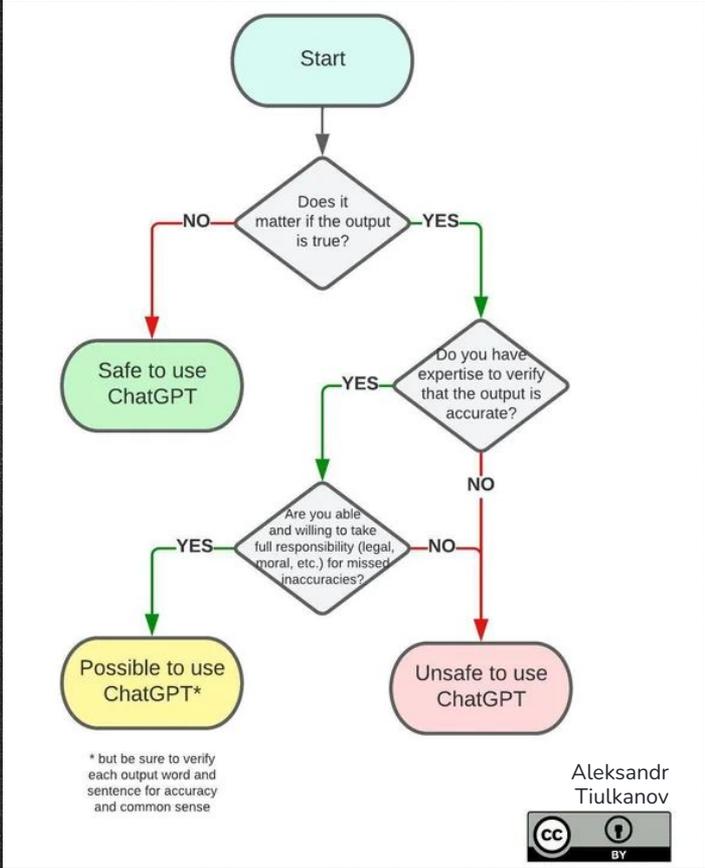
O

It will generate realistic sounding but totally imaginary fabrications.

## THREATS

T

# IS IT SAFE TO USE CHATGPT FOR YOUR TASK?



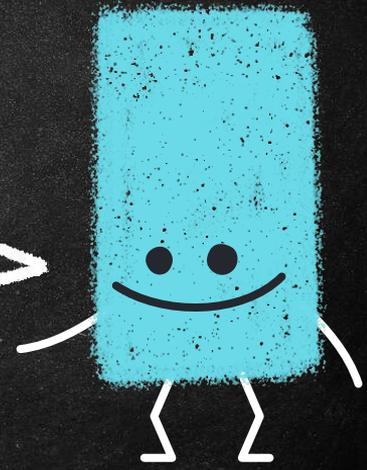
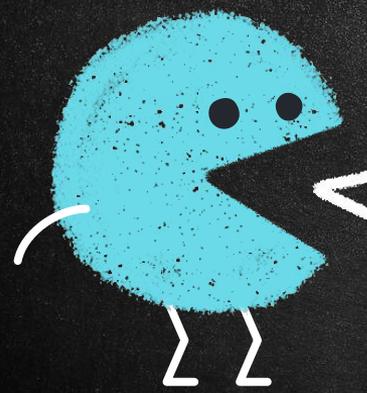


THE TECH YOU USE GETS USED

—

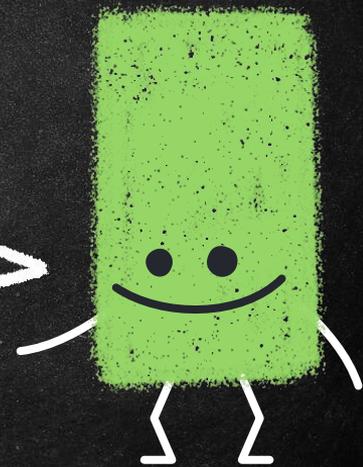
DON'T USE WHAT YOU DON'T  
WANT TO EXIST (EVEN IF THAT  
WON'T MAKE IT GO AWAY)

# WRITING EMAILS

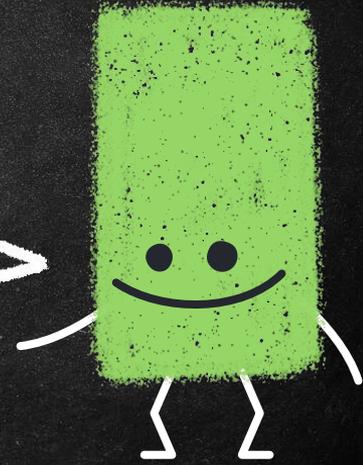


# INAUTHENTIC COMMUNICATION

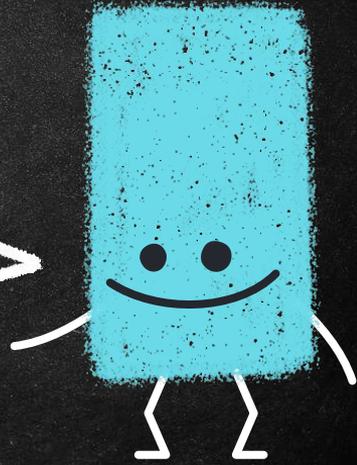
VICTORIAN LOVE LETTERS  
*NOT A NEW PROBLEM*



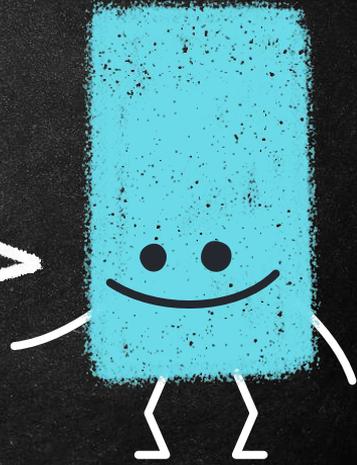
PRIVACY  
CONCERNS  
DON'T TELL GPT  
THINGS YOU  
DON'T WANT  
PEOPLE TO KNOW



WRITING SMALL  
HELPFUL  
PROGRAM



# MAKING A MARC RECORD





Create MARC authority record for Zydeco musician Jeremy Fruge.

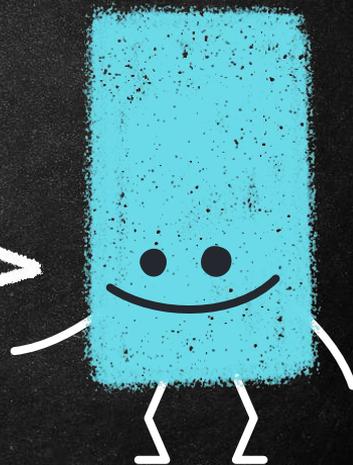
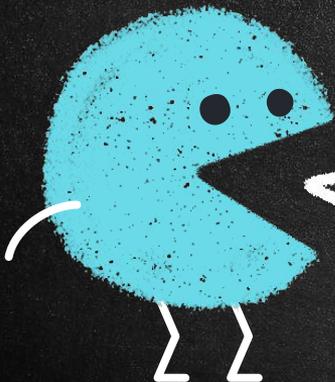


Here is an example of a MARC authority record for Zydeco musician Jeremy Fruge:

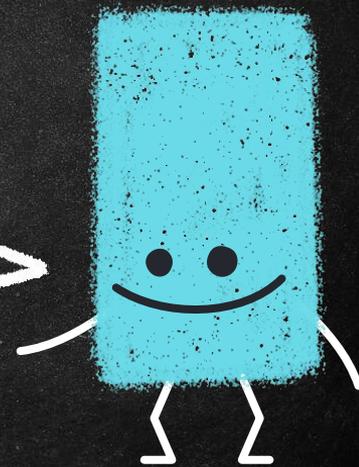
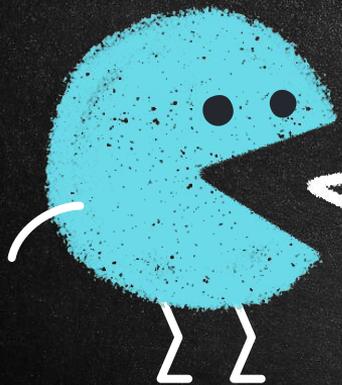
css

Copy code

```
LDR 00000nam a2200000 a 4500
001 ocm00000000
005 20220305
008 210305 n acznneabn n a eng d
010 __ |a n 2022045678
040 __ |a DLC |b eng |c DLC |d DL
100 1_ |a Fruge, Jeremy
370 __ |a Opelousas |b Louisiana
373 __ |a Zydeco musicians |2 lcs
375 __ |a male |2 lcdgt
500 __ |a Born in Opelousas, Loui
670 __ |a His Facebook page, view
678 __ |a Jeremy Fruge plays acco
```



UNSTRUCTURED  
TO STRUCTURED  
DATA





make an excel formula to make text in b1 combine with a1 but if column b is empty leave it empty



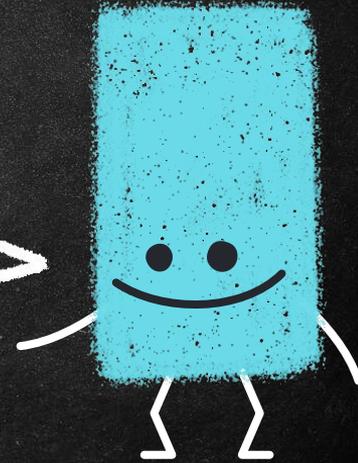
You can use the following formula in cell C1 to achieve this:



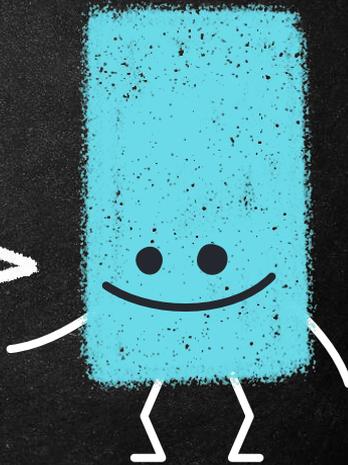
```
=IF(B1="", "", A1&B1)
```

This formula checks if cell B1 is empty. If it is empty, then it returns an empty string (""). If B1 is not empty, then it combines the text in cell A1 with B1 using the "&" operator. The final result will be displayed in cell C1.

You can then drag the formula down the column to apply it to the entire range of data.



# CREATING PROMOTIONAL COPY



# About the Author (or is it Authors)?

## Helena Marvin & OpenAI's GPT-3



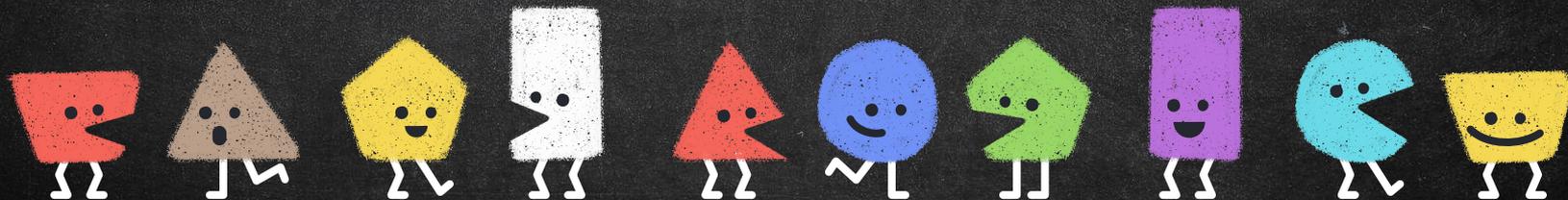
is the institutional repository, open educational resources and a reference librarian at the University of Missouri-St. Louis. Fair warning, this presentation is informed by her BA in Philosophy and long term fascination with culture and technology. These days she feels very lucky to share open access articles, open data and open educational resources professionally.

“The author generated this text in part with GPT-3, OpenAI’s large-scale language-generation model. Upon generating draft language, the author reviewed, edited, and revised the language to their own liking and takes ultimate responsibility for the content of this publication.” Copy provided by OpenAI documentation.



Thanks!  
Questions?

[marvinh@umsl.edu](mailto:marvinh@umsl.edu)





GPT-3

OpenAI's GPT-3 model is an advanced Artificial Intelligence (AI).

It uses natural language processing techniques to generate human-like text.

It can generate text on any given topic, from summarizing articles to creating original content.

Chat GPT-3 is technically GPT-3.5 with some chatbot elements, I mostly use the GPT-3 API

# Prompt & Response

OpenAI isn't really so open or non-profit



please explain how  
openai is a  
misleading name for  
a company that make  
proprietary software  
as a company that is  
no longer non-profit



OpenAI, formerly a  
non-profit, is now a  
for-profit company focused  
on developing proprietary  
software and tools for  
businesses. Therefore, the  
name "OpenAI" can be  
misleading for a company  
that primarily develops  
proprietary software and  
tools for businesses  
rather than engaging in  
open and transparent  
research for the public  
good.

# Credits

Thanks to all people who make and release awesome resources for free.

Presentation template

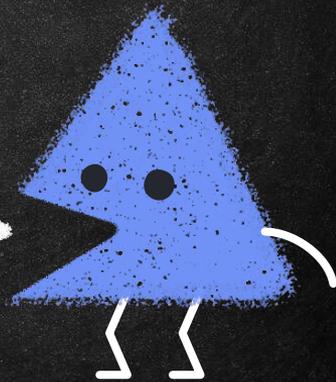
[Cute Geometry Presentation](#)

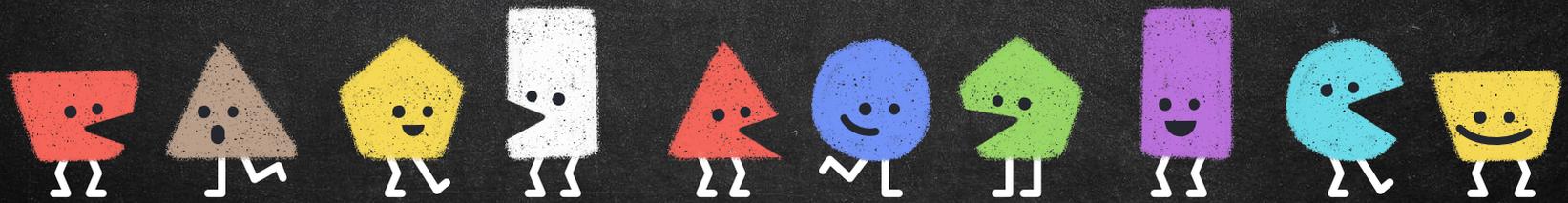
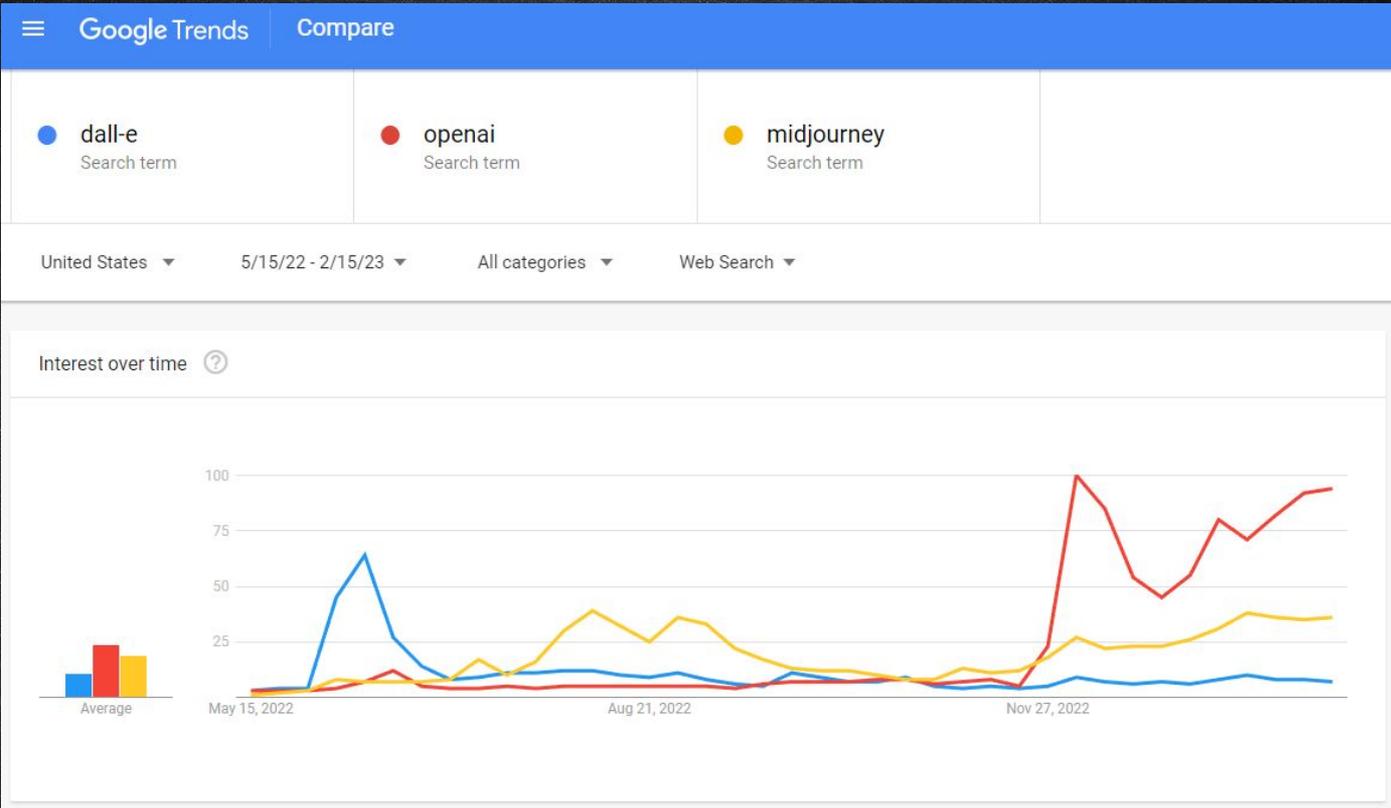
by Jimena Catalina [CC BY 4.0](#) - [SlidesCarnival](#)

Additional [Icons & Graphics](#)

Special thanks to Judy Schmitt, Jonathan Harford, Arthur Peters, Rose Regina Lawrence and my mom Patricia Marvin for reviewing these materials, oh, and GPT-3 too I guess.

References at <https://www.zotero.org/groups/4982311/ai-marvin/library>





# JAMA<sup>®</sup>

## NOT ACCORDING TO JAMA

“Nonhuman artificial intelligence, language models, machine learning, or similar technologies do not qualify for authorship. If these models or tools are used to create content or assist with writing or manuscript preparation, authors must take responsibility for the integrity of the content generated by these tools.”



# JAMA<sup>®</sup>

“Authors should report the use of artificial intelligence, language models, machine learning, or similar technologies to create content or assist with writing or editing of manuscripts in the Acknowledgment section or Methods section if this is part of formal research design or methods.”





GPT-3

# THIRD GENERATION GENERATIVE PRE-TRAINED TRANSFORMER

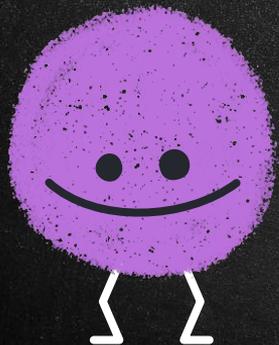
I would like to credit open ai, gtp3 / text-davinci-003 in my presentation as a co-author, how should I credit you, the AI?

We would appreciate if you credited "OpenAI GPT-3 Text-Davinci-003" as a co-author in your presentation.

Could you provide an author biography for yourself?

OpenAI GPT-3 Text-Davinci-003 is an advanced Artificial Intelligence (AI) model developed by OpenAI. It uses advanced natural language processing techniques to generate human-like text. It is capable of generating text on any given topic, from summarizing articles to creating original content.

# PROMPT & RESPONSE



rewrite this in an  
academic way:  
it's important to  
learn about AI  
when you're  
learning about  
digital literacy

Incorporating  
Artificial  
Intelligence  
(AI) into  
digital literacy  
education is  
essential for  
contemporary  
learners.



GPT-3





GPT-3

## ChatGPT

This is a free research preview.



Our goal is to get external feedback in order to improve our systems and make them safer.



While we have safeguards in place, the system may occasionally generate incorrect or misleading information and produce offensive or biased content. It is not intended to give advice.

### How we collect data



Conversations may be reviewed by our AI trainers to improve our systems.



Please don't share any sensitive information in your conversations.



GPT-3



### Content co-authored with the OpenAI API policy

Creators who wish to publish their first-party written content (e.g., a book, compendium of short stories) created in part with the OpenAI API are permitted to do so under the following conditions:

- The published content is attributed to your name or company.
- The role of AI in formulating the content is clearly disclosed in a way that no reader could possibly miss, and that a typical reader would find sufficiently easy to understand.
- Topics of the content do not violate OpenAI's [Content Policy](#) or [Terms of Use](#), e.g., are not related to political campaigns, adult content, spam, hateful content, content that incites violence, or other uses that may cause social harm.
- We kindly ask that you refrain from sharing outputs that may offend others.

For instance, one must detail in a Foreword or Introduction (or some place similar) the relative roles of drafting, editing, etc. People should not represent API-generated content as being wholly generated by a human or wholly generated by an AI, and it is a human who must take ultimate responsibility for the content being published.

Here is some stock language you may use to describe your creative process, provided it is accurate:

*“The author generated this text in part with GPT-3, OpenAI’s large-scale language-generation model. Upon generating draft language, the author reviewed, edited, and revised the language to their own liking and takes ultimate responsibility for the content of this publication.”*

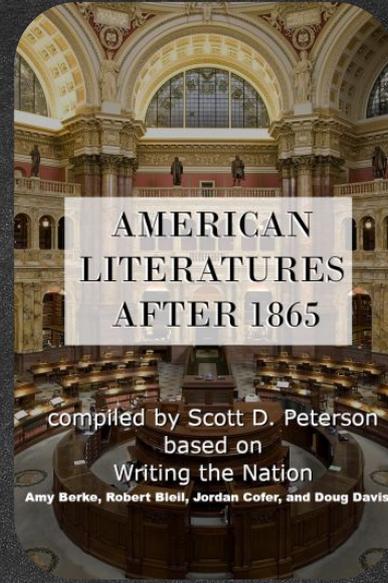
# AI & OER

Generated introductory text

Added contextualizing information

Summarized an article

Created copy for promoting the OER



Introduction

Helena Marvin and OpenAI

I. Contemporary Fiction: Metamodernism 101 / Silko / Faulkner



GPT-3

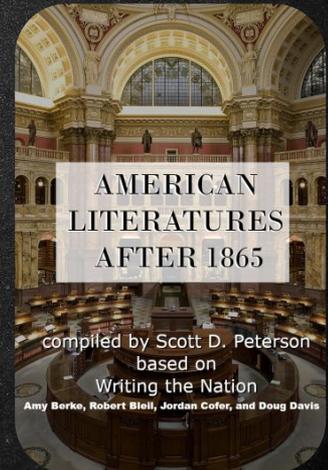


In this collection, you will find some of the most iconic works of American literature, both new and old. Many of the works featured in this anthology are available in the public domain, meaning that they are free from copyright restrictions. As such, you are free to read, share, and use them for any purpose. Much of this work is shared under a Creative Commons Attribution ShareAlike license, just as the Writing the Nation work this anthology has. This OER provides links to many works which are under copyright and cannot be collected within this text.

Some explanatory sections of this anthology (such as this section) were generated with the assistance of OpenAI, a technology that employs machine learning and natural language processing to generate text.



GPT-3





GPT-3

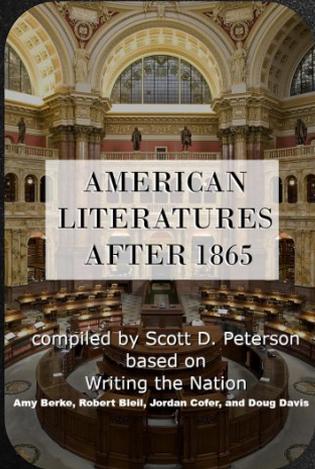
The influence of Leslie Marmon Silko and William Faulkner on contemporary fiction is undeniable. From the depths of Native American culture to the Southern United States, both writers have left their mark on the literary landscape.

...

Metamodernism is a cultural philosophy that encourages us to have hope and be creative. It suggests that we can find ways to move forward even when we face crises, and that we can be both serious and funny, sincere and ironic. It's a way of thinking that helps us to make sense of the confusing world we live in and encourages us to come up with creative solutions to the problems that we face.

Abramson, Seth. "Metamodernism 101." Just Words, 23 July 2015,

<https://medium.com/just-words/metamodernism-101-8cdb8563e0>.



"O E R" letters standing in a field together  
with no background glitter



Generated by Dall-E 2

# What shouldn't be done



Don't try and create something without a subject expert:

**"AI-based OERs"**

Social mobility  $\neq$  Uber

Don't use it without acknowledging that you're using it.



**GPT-3**

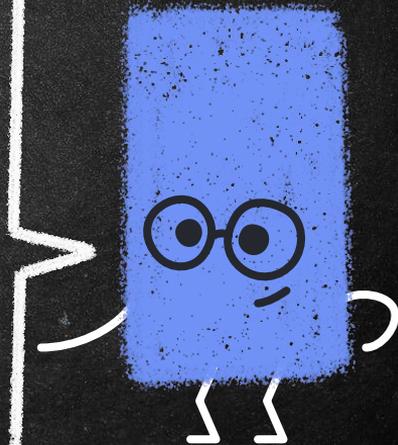


# OpenAI Recommendations

## Factual responses

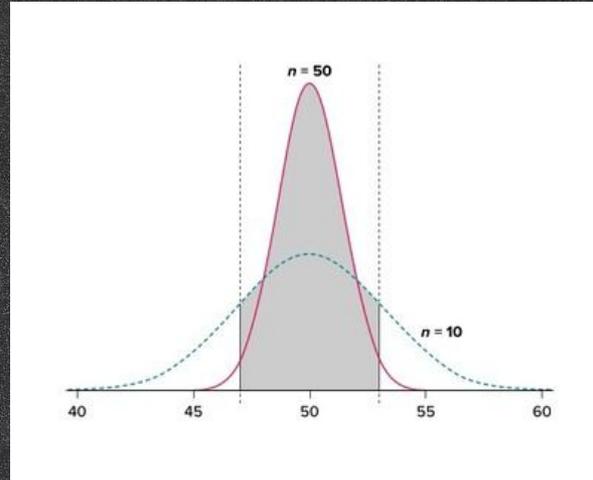
The API has a lot of knowledge that it's learned from the data that it was been trained on. It also has the ability to provide responses that sound very real but are in fact made up. There are two ways to limit the likelihood of the API making up an answer.

- 1 **Provide a ground truth for the API.** If you provide the API with a body of text to answer questions about (like a Wikipedia entry) it will be less likely to confabulate a response.
- 2 **Use a low probability and show the API how to say "I don't know".** If the API understands that in cases where it's less certain about a response that saying "I don't know" or some variation is appropriate, it will be less inclined to make up answers.



# What can be done

Accessibility



OER and AI pedagogy



GPT-3

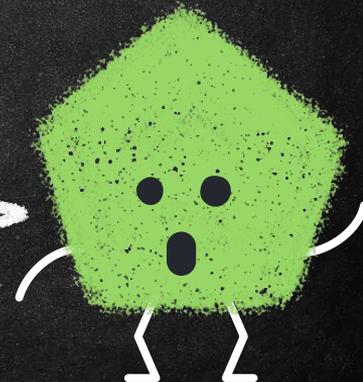


# Next Steps & More to Explore

[Adapting College Writing for the Age of Large Language Models such as ChatGPT: Some Next Steps for Educators](#)

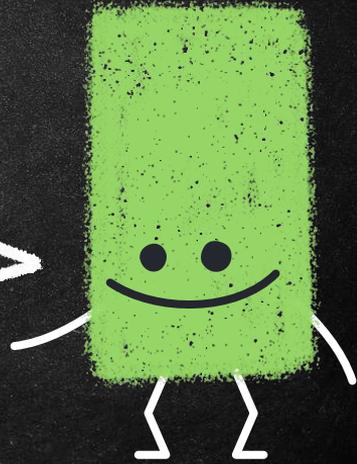
[AI Text Generators and Teaching Writing: Starting Points for Inquiry](#)

[Finding and Building Open Educational Resources about AI Writing Assistance](#)



# Authorship

Who wrote this?



## According to JAMA<sup>®</sup>

- “Nonhuman artificial intelligence, language models, machine learning, or similar technologies do not qualify for authorship”
- “Authors must take responsibility for the integrity of the content generated by these tools”
- Authors should note the use of these tools in their acknowledgments

