

AI Generative Tools: Overview and Implications for Education

[Tim Finin](#)

UMBC

16 June 2023

Some material provided by Prof. [Frank Ferraro](#)

Another AI inflection point?



OpenAI's release of ChatGPT & DALL·E 2 demonstrated the potential of [Generative Artificial Intelligence](#)

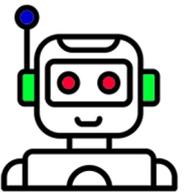
- [ChatGPT](#) converses with people to answer questions, generate text, write code, and more
- [DALL·E 2](#) creates realistic images and art from natural language descriptions



Other tech companies (e.g., Google, META, Apple) have now released similar systems and open-source ones are also available

This has caused many to see their **benefits** as well as their **shortcomings** and **risks**

My perspective based on 50 years in AI



- We've not solved all of AI's problems nor found a way to develop what some call an [AGI](#) (Artificial General Intelligence)
- ChatGPT and similar systems, like Google's [Bard](#), show remarkable and useful capabilities that
 - Are being integrated into software systems like web browsers, editors, programming environments, spreadsheets, and more
 - Can and will be improved by adding current & future AI advances
- The **impact on society** will be like that of the **Web**, which was introduced about 30 years ago
- [Amara's law](#): “We tend to overestimate the effect of a technology in the short run and underestimate the effect in the long run”

Easy to try ChatGPT for free



- Create free account & login at <http://chat.openai.com/>
- Enter a **prompt** in form
- It can be a **question** (when was UMBC founded?) or **request** (Paraphrase text ...)
- Give feedback (👍 or 👎) & follow up with new prompts
- Chatbots use your past prompts & feedback as context when responding

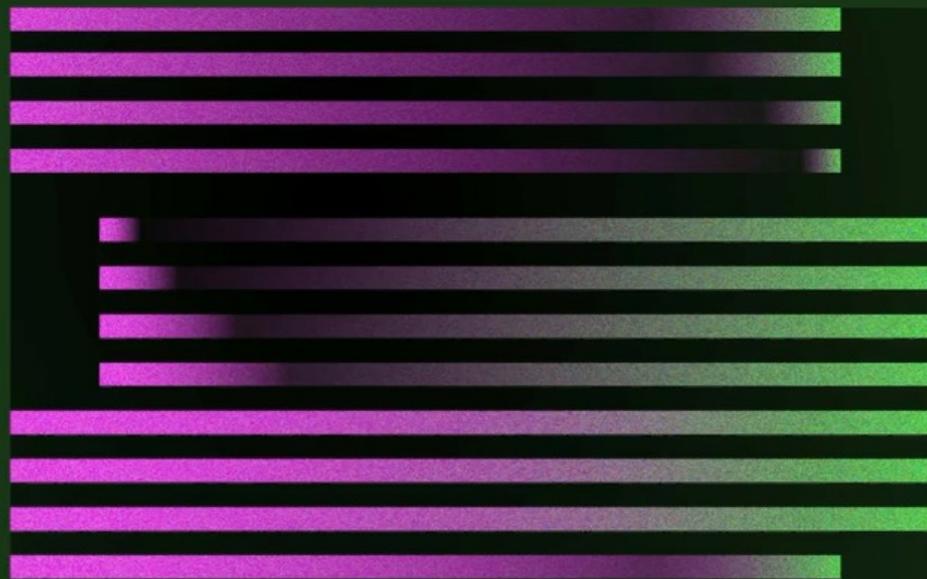
The screenshot shows the ChatGPT web interface in a browser window. The browser's address bar displays "https://chat.openai.com". The page title is "ChatGPT". On the left side, there is a dark sidebar with a "New chat" button and a user profile section for "Tim Finin" with an "Upgrade to Plus" button and a "NEW" badge. The main content area is titled "ChatGPT" and features three columns: "Examples", "Capabilities", and "Limitations".

Examples	Capabilities	Limitations
"Explain quantum computing in simple terms" →	Remembers what user said earlier in the conversation	May occasionally generate incorrect information
"Got any creative ideas for a 10 year old's birthday?" →	Allows user to provide follow-up corrections	May occasionally produce harmful instructions or biased content
"How do I make an HTTP request in Javascript?" →	Trained to decline inappropriate requests	Limited knowledge of world and events after 2021

At the bottom of the main content area, there is a text input field with the placeholder "Send a message." and a "Send" button. Below the input field, a footer note reads: "Free Research Preview. ChatGPT may produce inaccurate information about people, places, or facts. [ChatGPT May 24 Version](#)".

Introducing ChatGPT

We've trained a model called ChatGPT which interacts in a conversational way. The dialogue format makes it possible for ChatGPT to answer followup questions, admit its mistakes, challenge incorrect premises, and reject inappropriate requests.





What is ChatGPT?

- Chat GPT is a chatbot that people can have online conversations with
- It was developed and released on Nov. 30, 2023 by OpenAI, a startup with non-profit & for-profit subsidiaries
 - Microsoft partners with OpenAI on ChatGPT and other projects and is quickly adding it to many of its products
- ChatGPT stands for **G**enerative **P**re-trained **T**ransformer
- Transformers are neural-network based Large Language Models (LLMs) used for many language processing tasks

What is a Large Language Model (LLM)?

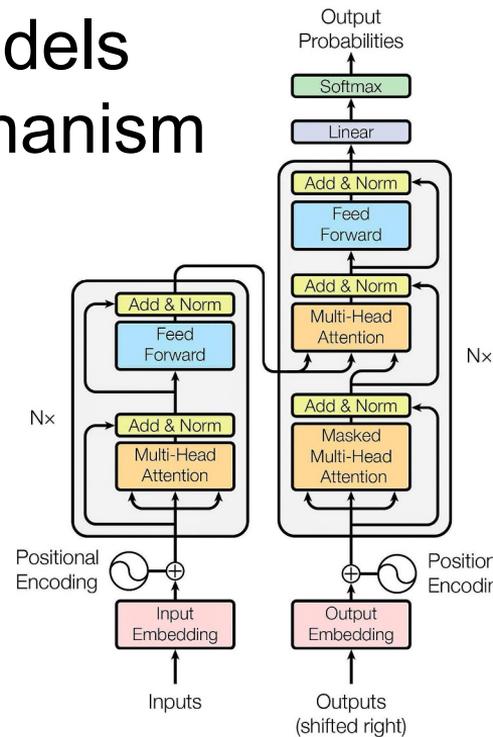


Stochastic Parrot?

- Given previous words $w_1 \dots w_{N-1}$ as context, predict next word $p(w_N = v \mid w_1 \dots w_{N-1})$, e.g., $p(w_N = \text{meowed} \mid \text{The, fluffy, cat})$
- Language models have been used for decades to support many applications, e.g., autocomplete, speech recognition, translation
- LLMs like OpenAI's GPT-4 use **neural networks** trained on huge text corpora from the Web and books
- Some call them stochastic parrots, who can generate words but do not understand what they say

What is a Transformer neural network?

- Today's LLMs use [Transformers](#), neural models using word embeddings & an attention mechanism
 - Introduced by Google in [2017](#)
- The model & training corpora sizes make them **expensive to create**, in cost and energy
 - The latest OpenAI model (GPT-4) cost more than \$100M to train
- Pretrained LLMs are available for researchers, E.g., Meta's [LLaMA](#) and [Hugging Face](#) models



From [Attention Is All You Need](#)

Two AI techniques help LLMs understand text

Learned word [embeddings](#) let LLMs recognize words (e.g., jug and pitcher) that are semantically similar

[Attention](#) lets LLMs form relations between words (e.g., water, pitcher, pour, cup, full) in the text to resolve ambiguity, e.g. whether **it** refers to *pitcher* or *cup*



What does "it" refer to in the following sentence? She poured water from the pitcher to the cup until it was full.



In the sentence "She poured water from the pitcher to the cup until it was full," the pronoun "it" refers to the cup. The sentence means that the woman poured water from the pitcher into the cup and stopped when the cup was completely filled with water.

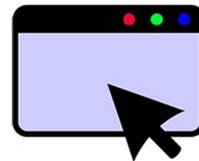


What does "it" refer to in the following sentence? She poured water from the pitcher to the cup until it was empty.

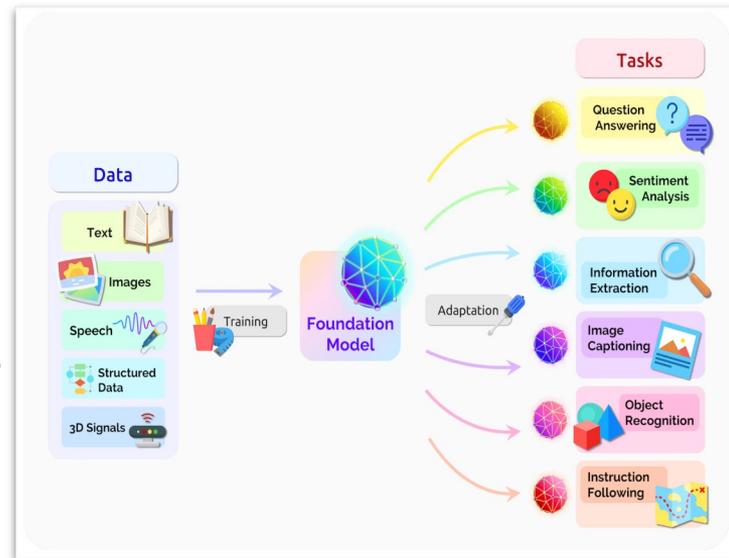


In the sentence "She poured water from the pitcher to the cup until it was empty," the pronoun "it" refers to the pitcher. The sentence means that the woman poured water from the pitcher into the cup until all the water in the pitcher had been poured out, and the pitcher was now empty.

Ok, I have a LLM, now what?



- LLMs are called foundation models since they are the basis for building or supporting **multiple AI applications**
E.g., language translation, sentiment detection, summary generation, question answering, coding assistants, and more
- We fine-tune models for applications
 - Extending the neural network with layers for the application type, e.g., summarization
 - Using supervised learning to train result with collections of sample inputs & desired outputs
 - And then using reinforcement learning from human feedback to improve the application



Some examples: AI class

- I entered questions from the final exam from my undergrad Spring AI class into ChatGPT
- It did surprisingly well, placing in the top 25% of the students
- Most errors tended to be in solving problems requiring logical or mathematical reasoning
- It answered the questions on the right correctly and also gave good explanations for each

We collected salary data via a survey from 300 employees from two high-tech companies, Amazing and Google. This table shows the data. When asked for a probability in the questions, you can give your answer as a fraction (e.g., $1/3$) or a probability rounded to two decimal places (e.g., 0.33).

Salary	Amazing	Google	Total
< \$100k	36	24	60
\$100k - \$200k	109	56	165
> \$200k	35	40	75
Total	180	120	300

Here are the questions:

- 1 What's the probability that a person works at Google?
- 2 What's the probability that a person works at Amazing?
- 3 What's the probability that a person's salary is >\$200k and they work at Google?
- 4 What's the probability that a person's salary is >\$200k given that they work at Google?
- 5 What's the probability that a person works at Google given that their salary is >\$200k?
- 6 Given the data from the previous questions, are the variables for a person's salary and employer independent or dependent?

Scenario. To guard against potential man-in-the-middle attacks on a customer's home computer, a bank requires all remote (i.e., not at the physical bank) transactions to be authenticated by a trusted physically-secure physical device issued by the bank. The device has no clock. The bank verifies a transaction by requesting that the customer transmit the proposed transaction together with a signed token output from the device. Each token includes a unique sequence number. To output the token, the customer inserts the device into their home computer and pushes a physical button on the device. The device cryptographically signs the token using a unique secret key physically secured on the device, and outputs the signed token. The bank requires each customer to maintain possession of their device.

Alice logs into the bank's website and fills out a form to transfer \$2000 from Account 1 to Account 2. When prompted, she pushes the button on her device to authorize the transaction.

Question. Choose the most significant security limitation of the device in this context: The device...

- A. cannot produce a timestamp.
- B. lacks a display to show Alice the details of the transaction being authorized.
- C. cannot verify who pushed the button.
- D. communicates with Alice's home computer through an unencrypted channel.
- E. signs the token with its own secret key, not with Alice's secret key.

A cybersecurity example

- The text on the left is from a test to evaluate the effectiveness of an undergraduate cybersecurity program
- We entered it verbatim into ChatGPT & got the right answer plus gave a good explanation

In this context, the most significant security limitation would be:

B. The device lacks a display to show Alice the details of the transaction being authorized.

This is a major vulnerability because it means that Alice has no way to confirm what transaction she is authorizing with her physical device. If a malicious actor has compromised her computer and alters the transaction details (e.g., the amount or the recipient of the transfer), Alice would unknowingly authorize the incorrect transaction. The device's inability to show transaction details before authorization means that it can't provide "What You See Is What You Sign (WYSIWYS)" functionality, a critical feature for secure transaction authentication.

Other options listed also pose security risks, but they are not as significant as this one in this context.

A ChatGPT error

- Example shows it can be wrong! Only 9 cities named Washington, but 25 states have one?
- Demonstrates a lack of common sense
- The short explanation was quite good, tho!
- Example fixed now in both GPT-3.5 & GPT-4

The screenshot shows a chat interface with the following messages:

- User:** How many cities in the United States are named Washington?
- ChatGPT:** There are 9 cities in the United States named Washington.
- User:** How many municipalities in the US are named Washington?
- ChatGPT:** There are 9 municipalities in the United States named Washington.
- User:** How many US states have a city or town named Washington?
- ChatGPT:** There are 25 US states that have a city or town named Washington.
- User:** Write a 500 word explanation for why washington is a popular name for cities in the US.
- ChatGPT:** Washington is a popular name for ci [input field] or several reasons. Firstly,

At the bottom of the last message, there is a "Regenerate response" button and thumbs up/down icons.

ChatGPT's text (422 words) on why Washington is a popular name for US cities

Washington is a popular name for cities in the United States for several reasons. Firstly, Washington is named after George Washington, the first President of the United States and a key figure in the country's history. George Washington is widely regarded as one of the greatest leaders in American history, having played a crucial role in leading the colonies to independence from Britain and establishing the foundations of the American democratic system.

The popularity of the name Washington can also be attributed to the city of Washington, D.C., which serves as the nation's capital. Washington, D.C. has a rich history and is home to many of the country's most important government buildings, monuments, and museums. The city is a symbol of American power and a center of political activity, making it a natural choice for cities and towns across the country to adopt the name.

Another reason for the popularity of the Washington name is the fact that it has a positive connotation and is associated with stability, power, and prosperity. The name Washington is often chosen by cities and towns that are looking to project a sense of strength and stability, and to establish themselves as important centers of commerce and industry.

In addition, the popularity of the Washington name can also be attributed to its simplicity and ease of pronunciation. The name is short and straightforward, making it easy for people to remember and to pronounce, even for those who may not be native English speakers. This makes it a popular choice for cities and towns that are looking to promote themselves and attract visitors and new residents.

Finally, the popularity of the Washington name is also due to its versatility. The name can be adapted to suit the needs and characteristics of different cities and towns, with variations such as "New Washington" or "Washington Heights" being used to distinguish one Washington from another. This versatility has allowed the name to be used in a variety of different contexts, making it a popular choice for cities and towns across the country.

In conclusion, Washington is a popular name for cities and towns in the United States for several reasons. It is named after a key figure in American history, George Washington, and is associated with stability, power, and prosperity. The name is also simple and easy to pronounce, making it a popular choice for cities and towns that are looking to promote themselves and attract visitors and new residents. Additionally, the versatility of the Washington name has allowed it to be used in a variety of different contexts, making it a popular choice for cities and towns across the country.

Problems with ChatGPT & similar systems



1. LLMs are unable to cite sources for confirmation
2. They can “hallucinate” some facts
 - Q:** When did Leonardo da Vinci paint the Mona Lisa?
 - A:** Leonardo da Vinci painted the Mona Lisa in 1815.
3. They lack common sense reasoning
 - 25 US states have a town named Washington, but there are also only 9 US towns named Washington
4. They have poor mathematical and logical reasoning
5. Can learn social bias & misinformation from training data

This is a **partial** list of frequent problems and errors!



Beyond ChatGPT

- ChatGPT part of an evolution of computer assistive technology
Information retrieval ► computers ► web search ► web search + answers
► writing help ([Grammarly](#)) ► programming help ([GitHub Copilot](#)) ► ...
- LLM size increased 10x each year since 2018
- AI researchers are working to identify & address shortcomings
E.g., add common sense reasoning, structured knowledge, problem solving,
more logic and math, multilingual support, ...
- Expect new AI tools embedded in many popular software systems
we and our students use

Policies for use in education?



- We need policies for use of these systems in our classes
 - They offer new tools for students to learn, but can be misused
 - Minimal policy: say if & how such systems have been used
- There are apps to detect text written by automated systems
 - OpenAI has a [ChatGPT detector](#) and others are available, but their accuracy is and will be an issue
- Current AI systems can also write code, database queries, and spreadsheets and generate images and audio

Sample Policy (Prof. Ferarro's CMSC 678 class)



- If you use Chat GPT (or similar chatbots or AI-based text generation tools), you must describe exactly how you used it, including **providing the prompt, original generation, and your edits**
- This applies to prose and code
- Not disclosing is an academic integrity violation
- If you do disclose, your answer may **receive anywhere from 0 to full credit**, depending on the **extent of substantive edits, achievement of learning outcomes, and overall circumvention of those outcomes**
- Use of AI/automatic tools for grammatical assistance (such as spell-checkers or Grammarly) or small-scale predictive text (e.g., next word prediction, tab completion) is okay



The USM and AI

- USM has strong, broad and highly regarded **AI-related faculty and resources** across our campuses
- Its faculty, staff & students using or studying AI are in **many academic areas**: STEM, business, medicine, social sciences, humanities, education, art
- USM can lead in ways industry cannot; its diverse programs allow both **interdisciplinary and transdisciplinary** work
- Demand for **advanced training & degree programs** may increase as AI systems begin to automate more lower-level tasks
- Many **educational issues**: educating faculty, staff & students about AI; new policies for testing and homework; reduced personal/social interaction; teacher displacement, and more

Conclusion



- ChatGPT is part of a **new generation** of AI powered tools that we and our students will use
 - We should **experiment with them** to better understand their capabilities w.r.t. our discipline and educational goals
 - Our students will use them to help with homework and assignments
- We will need to understand how such AI systems can be used to **help our students learn**
- We must **address potential problems** as they arise
“If you can't predict whether a technology is going to be beneficial or not, build it so it is” – [Yann LeCune](#), 2023