# An Introduction to Generative AI

**Cori Faklaris**
Assistant Professor, Dept. of Software and Information Systems

*Charlotte AI Institute for Smarter Learning, UNC Charlotte Dubois Center, May 18, 2023*

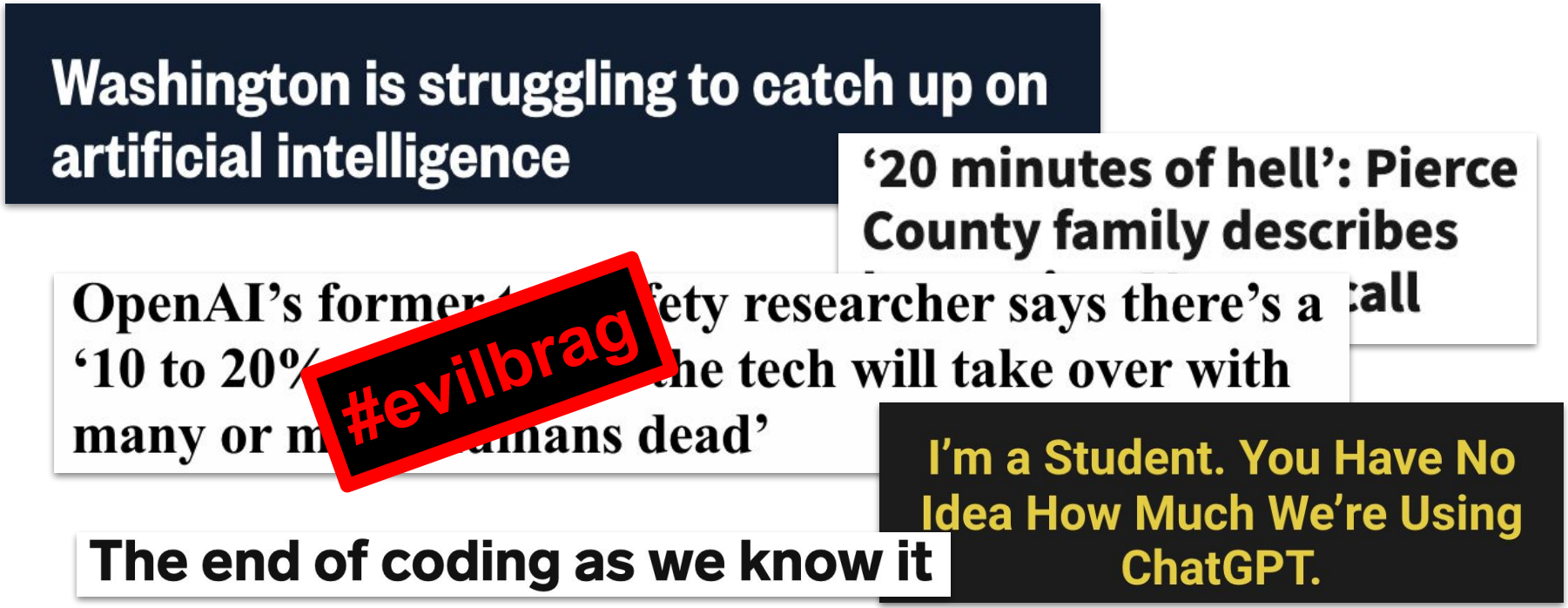



# Key takeaways from this talk

- Generative AI tools are great for PRODUCTIVITY - they can be nifty shortcuts to dispose of low-value tasks and / or to jumpstart creativity
- Generative AI tools should always be used - *and taught to be used* - with a critical mind, because they are prone to mistakes and “hallucinations”

# Overview of Generative AI

# Lots of hype - and doom /gloom - around AI right now …

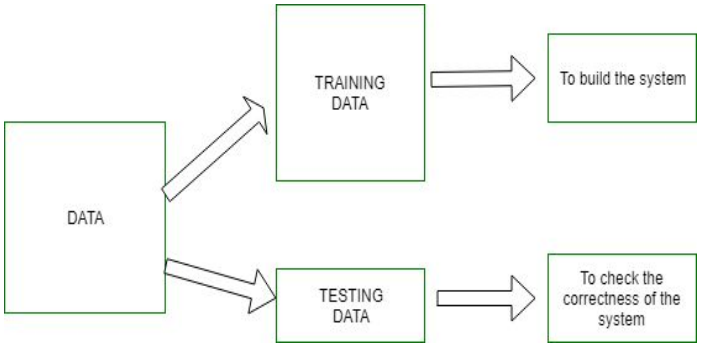


# When you hear **“AI,”** think **“statistical pattern-matching”**

- Oracle describes AI this way:

*AI has become a catchall term for applications that perform complex tasks that once required human input, such as communicating with customers online or playing chess.*

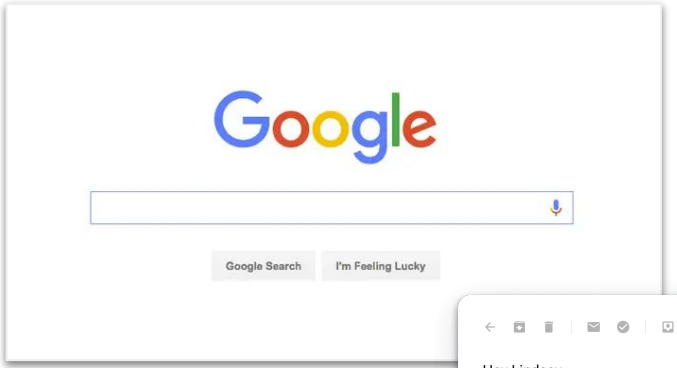
*The term is often used interchangeably with … **machine learning (ML)** and deep learning.*

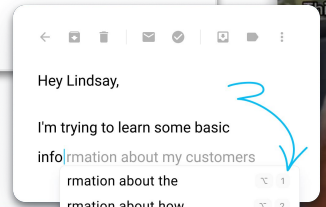


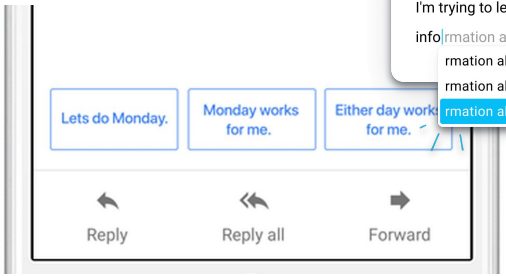

The data is **“tokenized”** *(= made into “chunks” of words, punctuation marks, pixels, etc.)* during this process - remember this for later

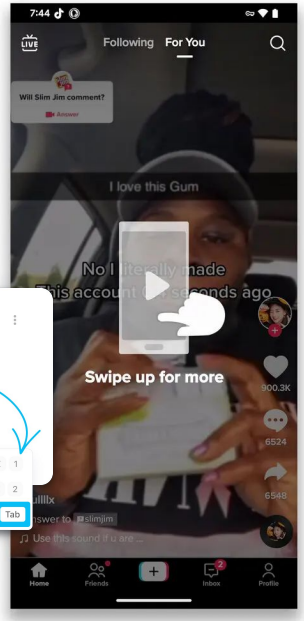
Text from What is Artificial Intelligence (AI)? Oracle, n.d. Retrieved May 16, 2023 from https://www.oracle.com/artificial-intelligence/what-is-ai/
Image from Pattern Recognition. GeeksforGeeks. Retrieved May 16, 2023 from https://www.geeksforgeeks.org/pattern-recognition-introduction/

# AI has been with us for years, whether “generative” or not

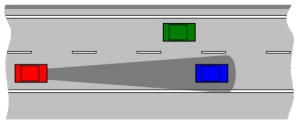




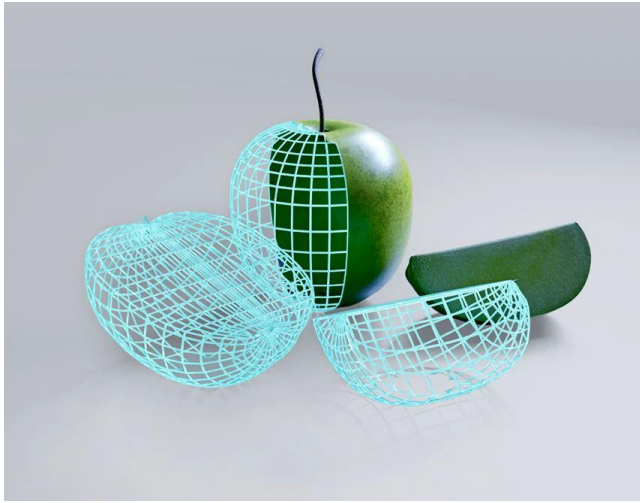
Tiktok screenshots from J. D. Biersdorfer. 2022. The Latecomer’s Guide to TikTok. *The New York Times*. Retrieved May 16, 2023 from https://www.nytimes.com/2022/10/26/technology/personaltech/tiktok-guide-latecomers.html
ADAS images from Wikipedia contributors. 2023. Advanced driver-assistance system. Wikipedia, The Free Encyclopedia. Retrieved from https://en.wikipedia.org/w/index.php?title=Advanced_driver-assistance_system&oldid=1150142876

# Now, AI can synthesize *part* or *all* of a creative work

- McKinsey defines generative AI as:

  *… Algorithms (such as ChatGPT) that can be used to create new content, including audio, code, images, text, simulations, and videos.*

  *Recent breakthroughs in the field have the potential to drastically change the way we approach content creation.*



Text and image from What is generative AI? McKinsey. Retrieved May 16, 2023 from https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-generative-ai

# How Generative AI works (admittedly oversimplified)

The system generates text or images using its previously built model of the statistical distributions of **tokens** *(= "chunks" of words, punctuation marks, pixels, etc.)* created from its *very large* training dataset.

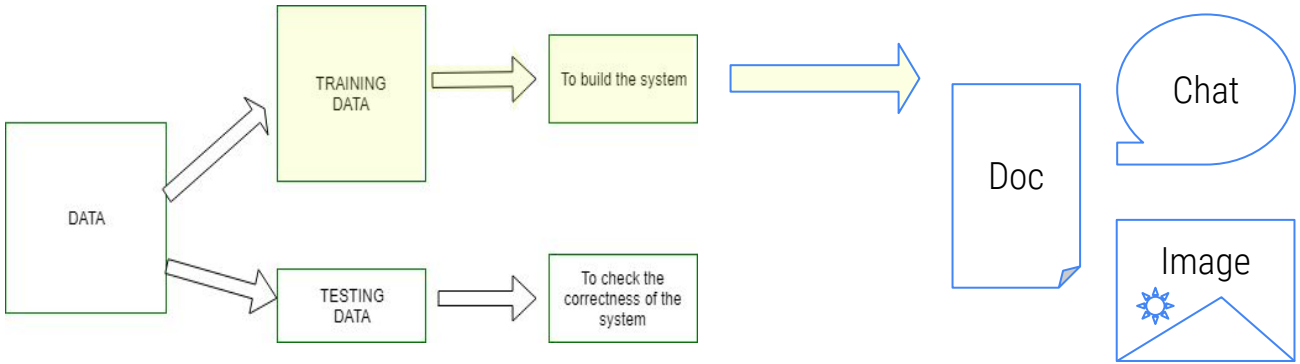
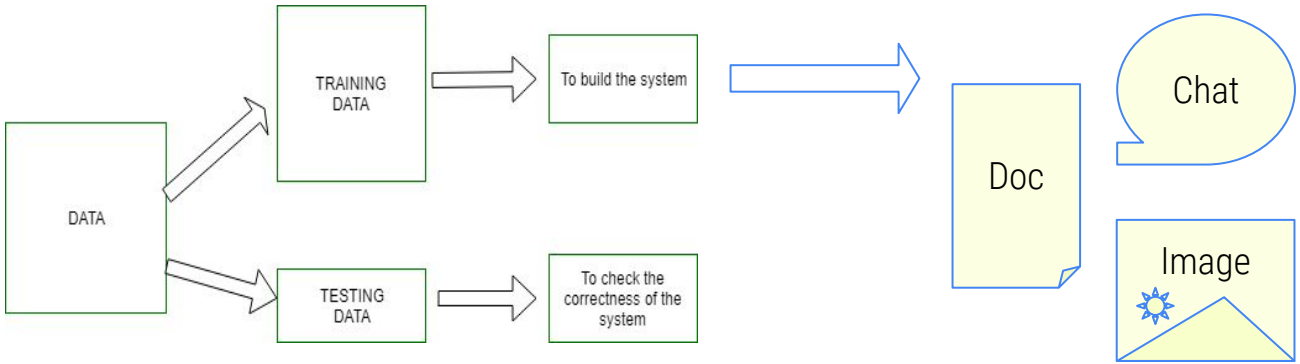



Image from Pattern Recognition. GeeksforGeeks. Retrieved May 16, 2023 from https://www.geeksforgeeks.org/pattern-recognition-introduction/
Murray Shanahan. 2022. Talking About Large Language Models. arXiv [cs.CL]. Retrieved from http://arxiv.org/abs/2212.03551
Bea Stollnitz. How generative language models work. Retrieved May 10, 2023 from https://bea.stollnitz.com/blog/how-gpt-works/

# How Generative AI works (admittedly oversimplified)

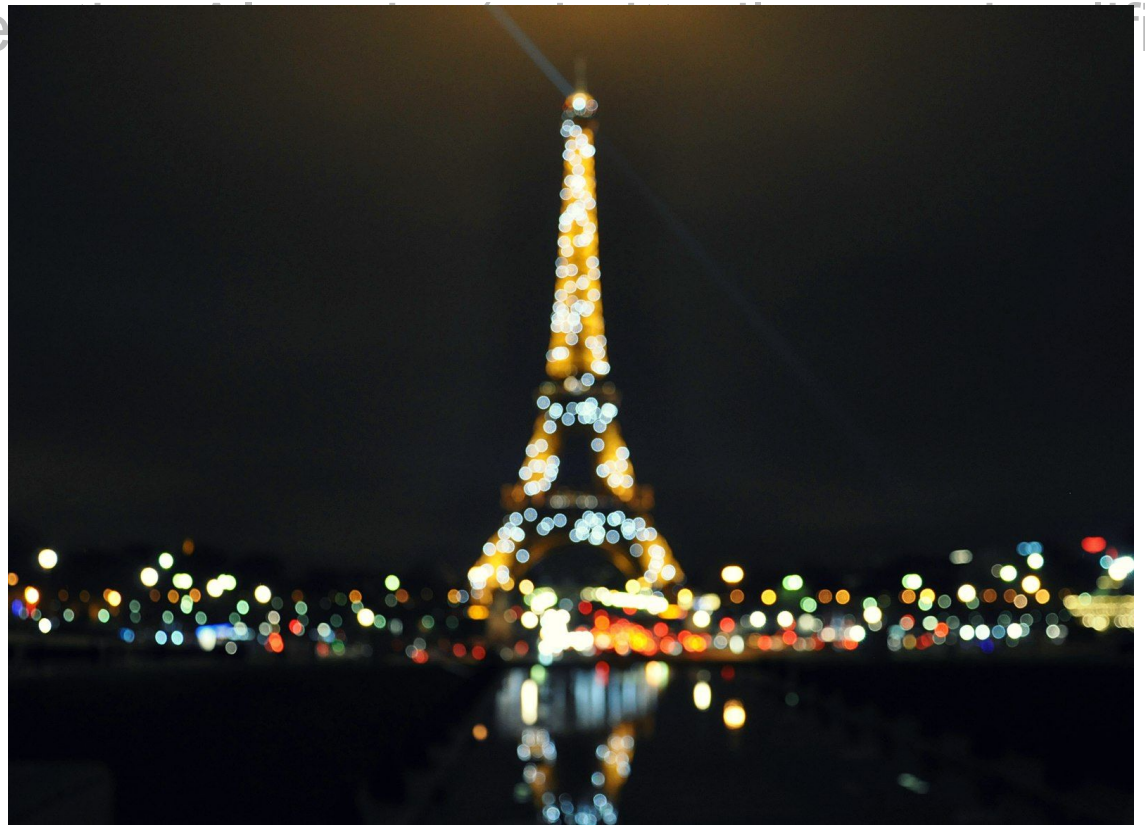
**It might make mistakes or "hallucinate" based on the limitations of its process, but the output still might look like what you wanted.**

Ted Chiang's analogy = "unreliable photocopier" or a "blurry JPEG"




Ted Chiang. 2023. ChatGPT Is a Blurry JPEG of the Web. *The New Yorker*. Retrieved May 10, 2023 from https://www.newyorker.com/tech/annals-of-technology/chatgpt-is-a-blurry-jpeg-of-the-web
Murray Shanahan. 2022. Talking About Large Language Models. arXiv [cs.CL]. Retrieved from http://arxiv.org/abs/2212.03551
Bea Stollnitz. How generative language models work. Retrieved May 10, 2023 from https://bea.stollnitz.com/blog/how-gpt-works/

# How Gene ified)



https://commons.wikimedia.org/wiki/File:Blurry_eiffel.jpg - shared under CC-SA 4.0 license

# How Generative AI works (admittedly oversimplified)

We can ask it questions - but a very specific type of question known as **prompts,** following this structure:

**“Here’s a fragment of text.**

**Tell me how this fragment might *<continue on in this language, or suggest a particular image>*.**

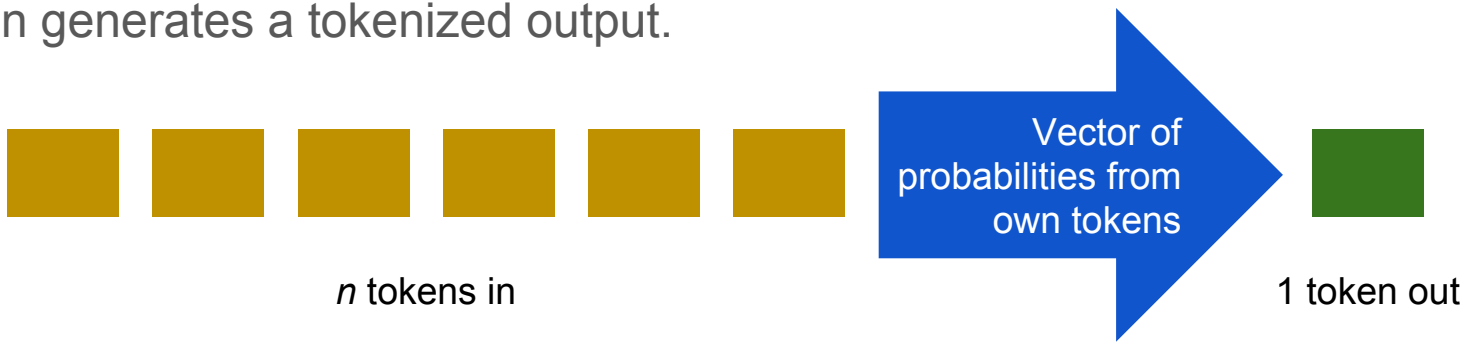
**According to your model of the statistics of *<human language, or human-handled images>*, what *<words, or pixels>* are likely to come next?”**

Murray Shanahan. 2022. Talking About Large Language Models. arXiv [cs.CL]. Retrieved from http://arxiv.org/abs/2212.03551

# How Generative AI works (admittedly oversimplified)

The prompts are converted into tokens *(= "chunks" of words, punctuation marks, pixels, etc.)*, then the system analyzes what is likely to come next, based on the tokens in its own dataset (as many as 32,000 in GPT-4!).

It then generates a tokenized output.




Murray Shanahan. 2022. Talking About Large Language Models. arXiv [cs.CL]. Retrieved from http://arxiv.org/abs/2212.03551
Bea Stollnitz. How generative language models work. Retrieved May 10, 2023 from https://bea.stollnitz.com/blog/how-gpt-works/

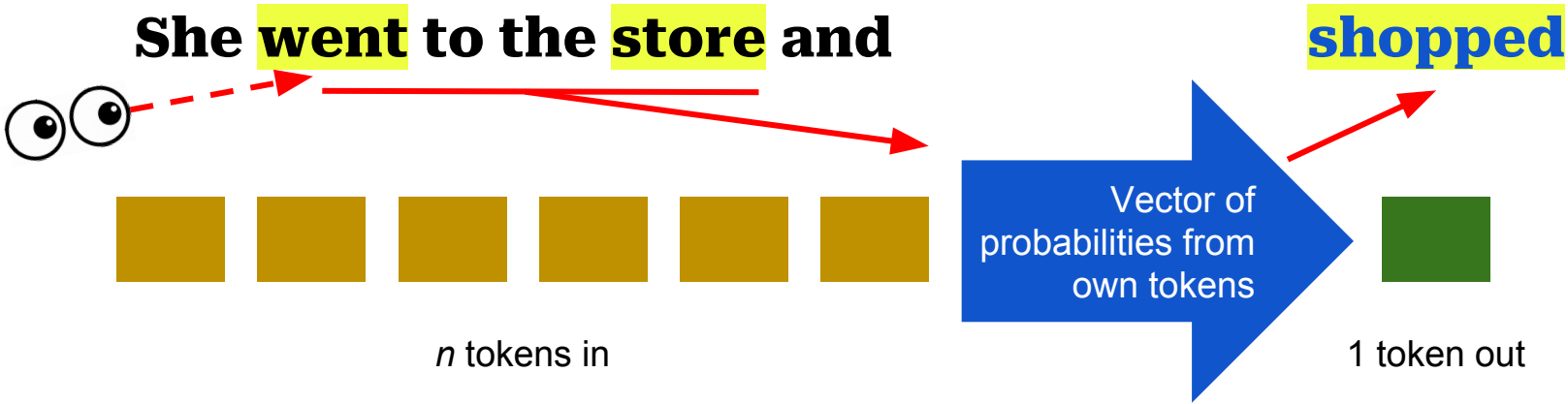
# How Generative AI works (admittedly oversimplified)

With each output, it keeps re-analyzing the probabilities to decide next tokens.


Murray Shanahan. 2022. Talking About Large Language Models. arXiv [cs.CL]. Retrieved from http://arxiv.org/abs/2212.03551
Bea Stollnitz. How generative language models work. Retrieved May 10, 2023 from https://bea.stollnitz.com/blog/how-gpt-works/

# HERE'S THE REALLY COOL PART!!!

Transformers (the "T in "GPT") know how to **direct attention to specific parts of the input** to guide their selection of the output - such as verb tenses, objects.




Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention Is All You Need. arXiv [cs.CL]. Retrieved from http://arxiv.org/abs/1706.03762
Bea Stollnitz. How generative language models work. Retrieved May 10, 2023 from https://bea.stollnitz.com/blog/how-gpt-works/

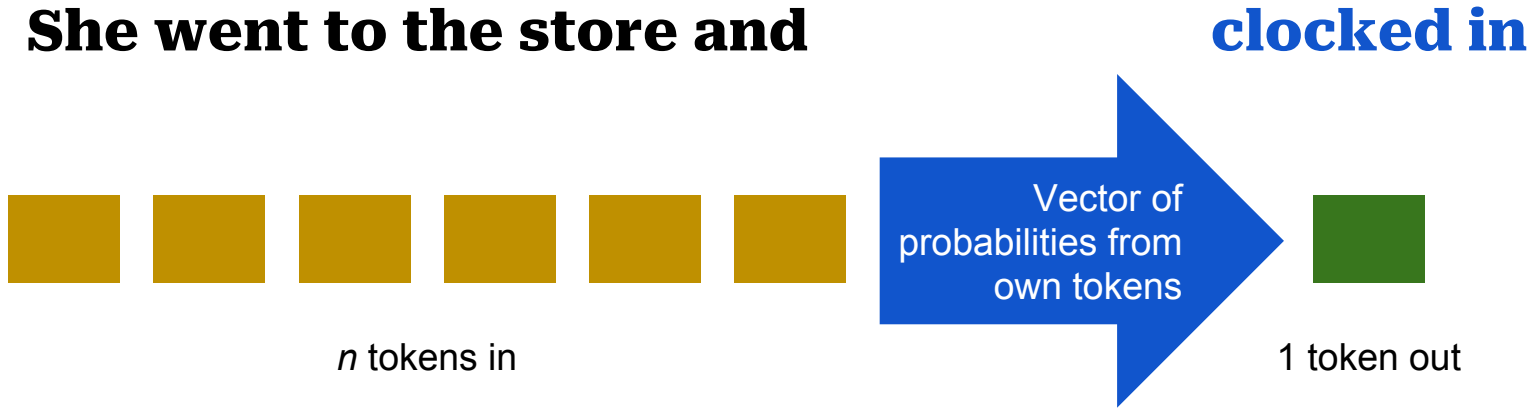
# How Generative AI works (admittedly oversimplified)

The system can give you different answers to the same inputs:


Murray Shanahan. 2022. Talking About Large Language Models. arXiv [cs.CL]. Retrieved from http://arxiv.org/abs/2212.03551
Bea Stollnitz. How generative language models work. Retrieved May 10, 2023 from https://bea.stollnitz.com/blog/how-gpt-works/

# How Generative AI works (admittedly oversimplified)

The system can give you different answers to the same inputs:




Murray Shanahan. 2022. Talking About Large Language Models. arXiv [cs.CL]. Retrieved from http://arxiv.org/abs/2212.03551
Bea Stollnitz. How generative language models work. Retrieved May 10, 2023 from https://bea.stollnitz.com/blog/how-gpt-works/

# How Generative AI works (admittedly oversimplified)

The system can give you different answers to the same inputs:


Murray Shanahan. 2022. Talking About Large Language Models. arXiv [cs.CL]. Retrieved from http://arxiv.org/abs/2212.03551
Bea Stollnitz. How generative language models work. Retrieved May 10, 2023 from https://bea.stollnitz.com/blog/how-gpt-works/

# How Generative AI works (admittedly oversimplified)

"Hallucinations" - when the output doesn't seem to make sense - are why it is important not to accept everything it outputs at face value.

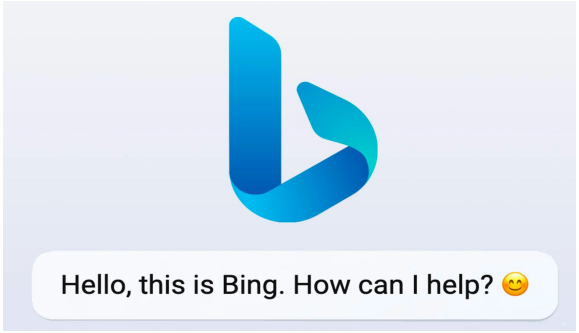

Murray Shanahan. 2022. Talking About Large Language Models. arXiv [cs.CL]. Retrieved from http://arxiv.org/abs/2212.03551
Bea Stollnitz. How generative language models work. Retrieved May 10, 2023 from https://bea.stollnitz.com/blog/how-gpt-works/

# Examples of publicly available Generative AI tools

ChatGPT

Stable Diffusion

OpenAI DALL·E 2




Bard Experiment

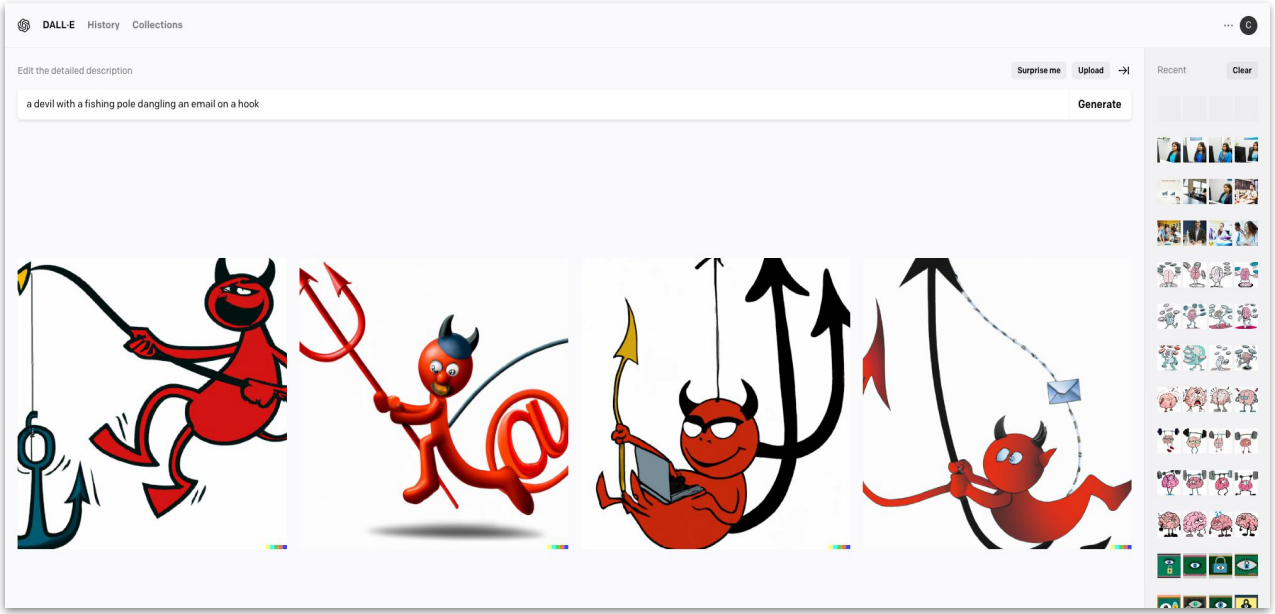
HUGGING FACE

**Crowdsourced list of available AI tools:**
https://bit.ly/UsefulLLMs

# Using Generative AI as an academic

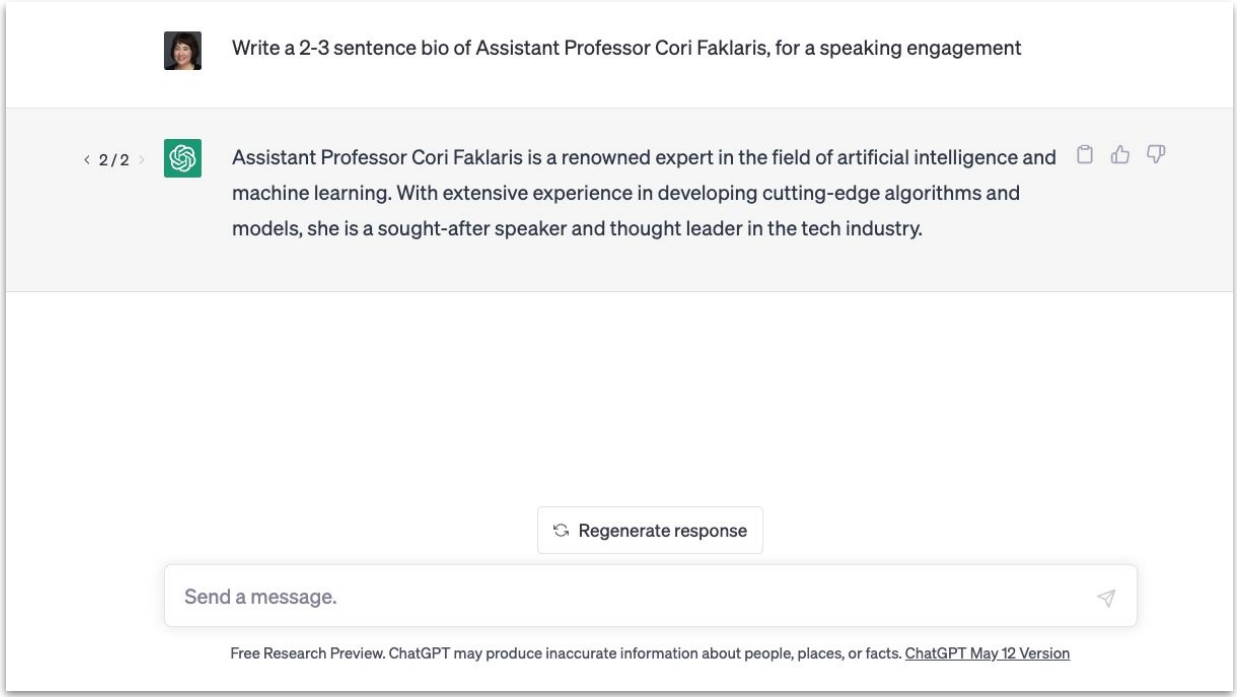
# Use DALL-E 2 to create images for course slides

- Goal 1: Quickly source visuals that add interest and reinforce content
- Goal 2: Demonstrate limits of AI output with limited inputs or prompts



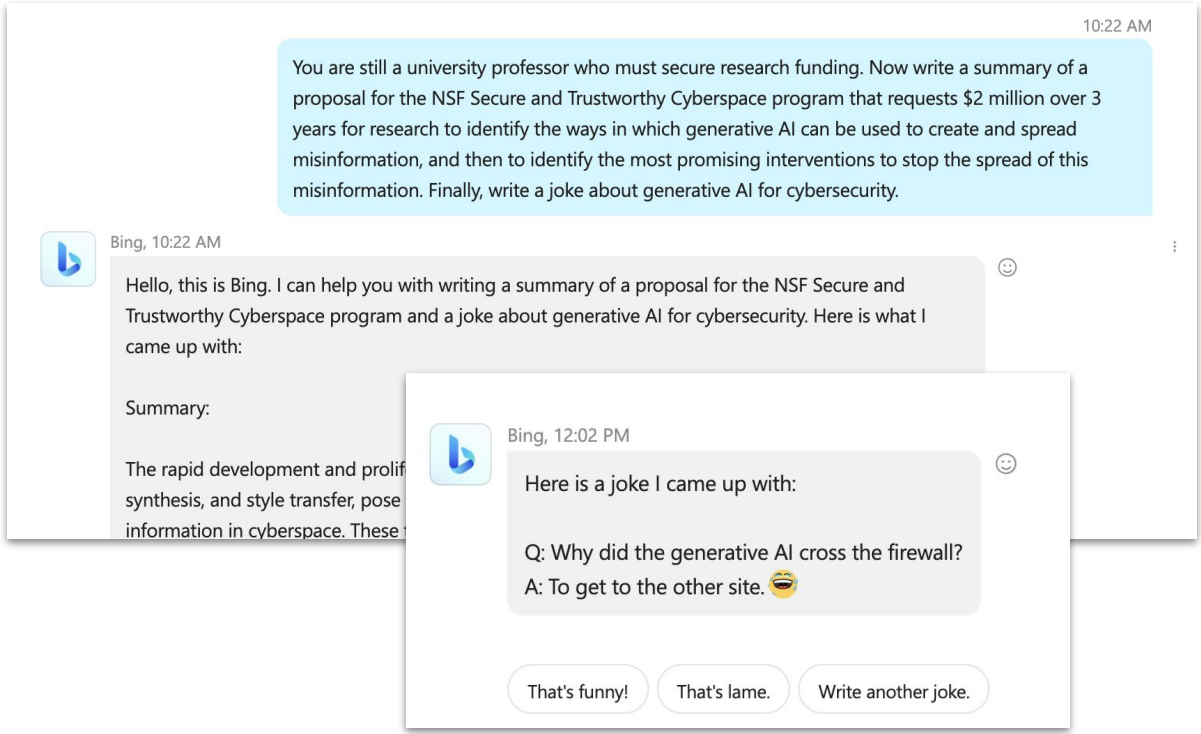
# Use ChatGPT to create first draft of biography text

- Goal 1: Cut the time spent on low-value but necessary job tasks
- Goal 2: Goof around with fellow academics on social media



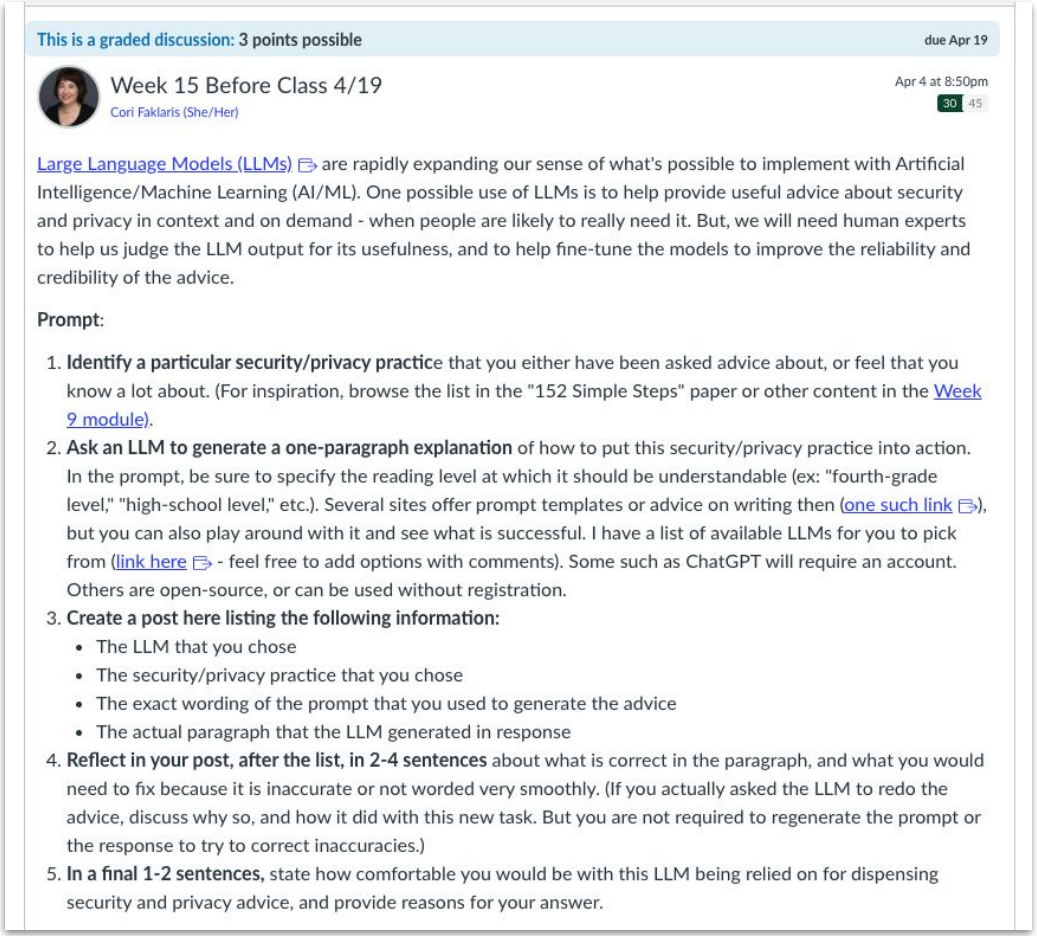
# Use BingChat to draft a grant proposal

- Goal 1: Overcome “analysis paralysis”, make yourself laugh in the process
- Goal 2: Experiment with a sequence of prompts for sophisticated outputs



# Assign students to pick/use a tool, then critique the output

- Goal 1: Give permission and encouragement to play around with new tech
- Goal 2: Mentor class members in how to think critically use of AI tools



This is a graded discussion: 3 points possible — due Apr 19

Week 15 Before Class 4/19

Cori Faklaris (She/Her) — Apr 4 at 8:50pm — 30 45

Large Language Models (LLMs) are rapidly expanding our sense of what's possible to implement with Artificial Intelligence/Machine Learning (AI/ML). One possible use of LLMs is to help provide useful advice about security and privacy in context and on demand - when people are likely to really need it. But, we will need human experts to help us judge the LLM output for its usefulness, and to help fine-tune the models to improve the reliability and credibility of the advice.

**Prompt**:

1. **Identify a particular security/privacy practice** that you either have been asked advice about, or feel that you know a lot about. (For inspiration, browse the list in the "152 Simple Steps" paper or other content in the Week 9 module).
2. **Ask an LLM to generate a one-paragraph explanation** of how to put this security/privacy practice into action. In the prompt, be sure to specify the reading level at which it should be understandable (ex: "fourth-grade level," "high-school level," etc.). Several sites offer prompt templates or advice on writing then (one such link), but you can also play around with it and see what is successful. I have a list of available LLMs for you to pick from (link here - feel free to add options with comments). Some such as ChatGPT will require an account. Others are open-source, or can be used without registration.
3. **Create a post here listing the following information:**
   - The LLM that you chose
   - The security/privacy practice that you chose
   - The exact wording of the prompt that you used to generate the advice
   - The actual paragraph that the LLM generated in response
4. **Reflect in your post, after the list, in 2-4 sentences** about what is correct in the paragraph, and what you would need to fix because it is inaccurate or not worded very smoothly. (If you actually asked the LLM to redo the advice, discuss why so, and how it did with this new task. But you are not required to regenerate the prompt or the response to try to correct inaccuracies.)
5. **In a final 1-2 sentences,** state how comfortable you would be with this LLM being relied on for dispensing security and privacy advice, and provide reasons for your answer.

# My syllabus policy on “Use of AI and Other Creative Tools”

In this course, students are **permitted to use tools such as Stable Diffusion, DALL-E, ChatGPT, and BingChat.** In general, permitted use of such tools is consistent with **permitted use of non-AI assistants** such as Grammarly, **templating tools** such as Canva, or **images or text sourced from** the internet or others’ files.

No student may submit an assignment or work on an exam as their own that is **entirely generated** by means of an AI tool.

If students use an AI tool or other creative tool to generate, draft, create, or compose any portion of any assignment, they must **(a) credit** the tool, **(b) identify** what part of the work is from the AI tool and what is from themselves, and **(c) briefly summarize why** they decided to use the tool and include its output.

Cori Faklaris. 2023. Policy on Use of AI Tools for my course syllabus, version 1.0. Cori Faklaris’ blog – HeyCori. Retrieved May 16, 2023 from https://blog.corifaklaris.com/2023/03/17/policy-on-use-of-ai-tools-for-my-course-syllabus-version-1-0/#.ZGPtWezMJqs

# Some *actual and/or realistic risks* of using generative AI

- Violations of data privacy
  - Some students told me they do not feel comfortable giving up any data to such services, such as may be required for creating an account. For these students, I created an alternate assignment for Slide 24, using a search engine.
- Violations of intellectual property
  - Check the Terms of Service - will your inputs or prompts be used as training data?
- Violations of academic integrity
  - Do a spot check of outputs, using a search engine, to see if any are wholly from another work
  - Analyze submitted work using Open AI's AI Text Classifier or the multi-service GPTZero

# Humans' #1 skill set will continue to be *communication*


Screenshot from https://twitter.com/TheRealOllieLaw/status/1656605938374307840?s=20

# Key takeaways

- Generative AI tools can be nifty shortcuts to dispose of low-value tasks and / or to jumpstart creativity.
- Generative AI tools should always be used with your “thinking cap” on because they are prone to mistakes and “hallucinations.”

**Thank you for listening!**

**Crowdsourced list of available AI tools:**
https://bit.ly/UsefulLLMs