



AS OF SEPTEMBER 23, 2024

# DEMYSTIFYING GENERATIVE AI FOR CPAs

**MAILYN FERNANDEZ**  
PHD CANDIDATE — ACCOUNTING

# Agenda

1. **Brief overview of how these models work**
2. **Important Considerations**
3. **Example use-case(s)**

# Your job for the next 45 minutes

## 1. Sit back and listen

## 2. Ask questions

- You probably aren't the only one with the same question
- I learn just as much from your questions as you do from my responses to them

## After this session...

Take a photo of this slide  
with your phone

### 1. Download slides:

**<https://cpa.chat-lab.ai>**

### 2. Review additional resources on site

### 3. Check site often for updates

### 4. Reach out to me

**In-person**

**[mailynf@vt.edu](mailto:mailynf@vt.edu)**

**305-898-1455**

# How does ChatGPT work?

## Word Prediction:

- ChatGPT predicts the next word based on context

## Large Language Model (LLM):

- Trained on vast amounts of data (free internet & licensed)
- Estimates probabilities for word combinations

## Continuous Improvement:

- Human feedback reinforcement learning (HFRL)
- User interactions for fine-tuning

# Key Features and Limitations

## Context Windows:

- Allows for long conversations (ChatGPT <128K tokens)
- 1 token  $\approx$  3/4 of a word

## Temperature:

- Controls randomness in responses
- Higher temperature = more creative, varied responses

## Limitations:

- Knowledge cut-offs
- Can generate seemingly plausible, but false & incorrect text ("*hallucinations*")
- Unable to admit lack of knowledge
- Potential for inconsistent performance over time

# How can we improve LLMs?

## 1. Prompt Engineering

- Iterate on what you tell the LLM to generate the best response for your purpose

## 2. RAG

- Feeding LLM additional context from your files

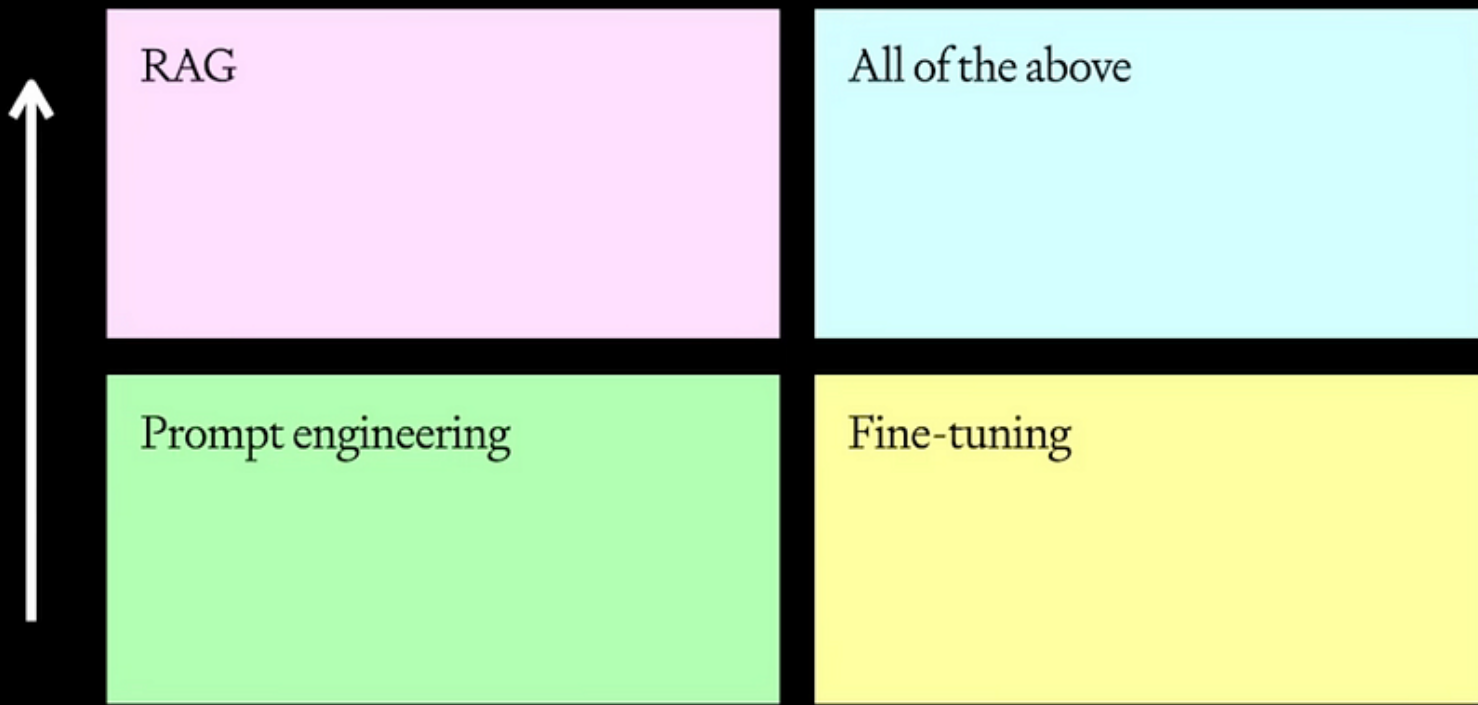
## 3. Fine-tuning

- Feeding it enough examples of the desired response from a given prompt that it changes the behavior of the model

## The optimization flow

Context  
optimization

What the model  
needs to know



LLM optimization

How the model needs to act



# Anecdotal Advice

- 1. Clear the chat context often**
  - Can feel like longer conversations are less useful
    - Start a new chat
    - Give it just the context that you want it to have
- 2. Natural language is usually good enough**
  - Use clear, step-by-step instructions
  - Ask it to explain its reasoning step-by-step
- 3. Be polite to the AI**
  - Human interactions with technology can change the way Humans interact with other Humans

# AI BIAS

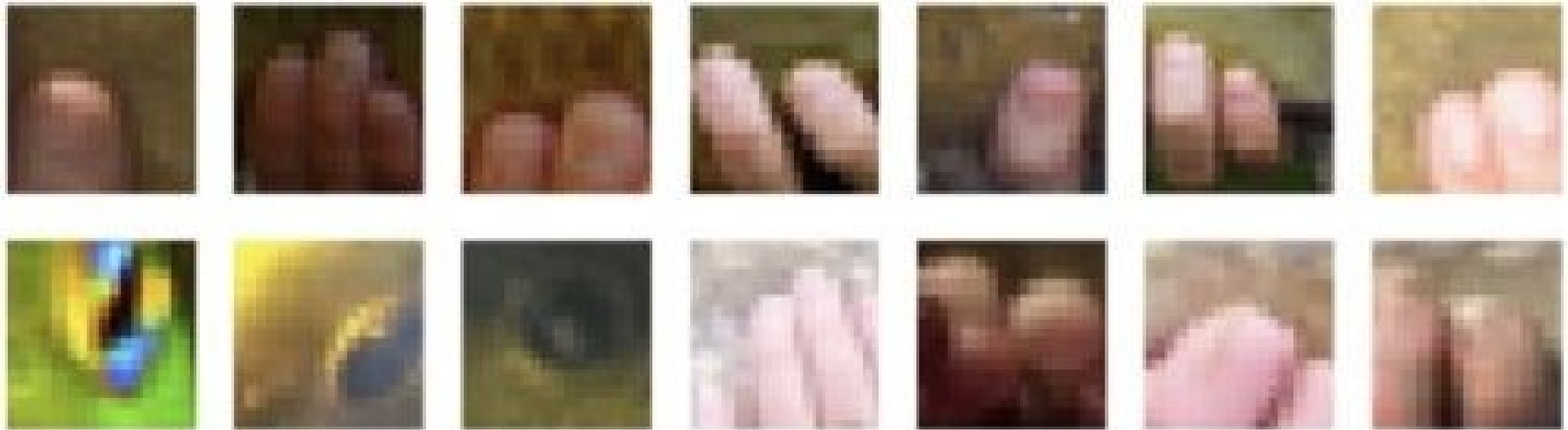
## Bias isn't inherently good or bad

- In statistical context, it's just a feature of data
- We talked about improving LLMs by fine-tuning with context-specific data
  - purposely biasing the model

**Need to understand the biases that exist in the underlying training data to assess whether a model's outputs are reliable**

# Thinking about underlying training data

Researchers at the University of Tuebingen trained a neural net to recognize images, and then had it point out which parts of the images were the most important for its decision. When they asked it to highlight the most important pixels for the category “Tench” (a kind of fish), this is what it highlighted:



Human fingers against a green background!

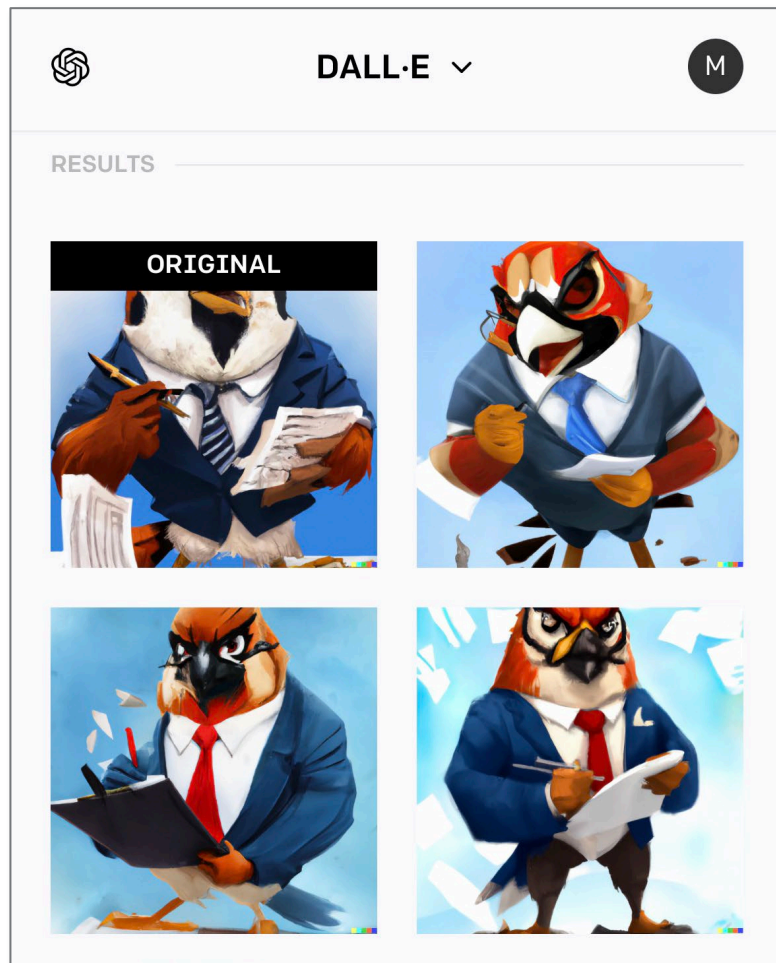
# Thinking about underlying training data

The image-generating neural net in ArtBreeder (called BigGAN) was also trained on the same dataset, called ImageNet, and when you ask it to generate Tenches, this is what it does:



# Thinking about underlying training data

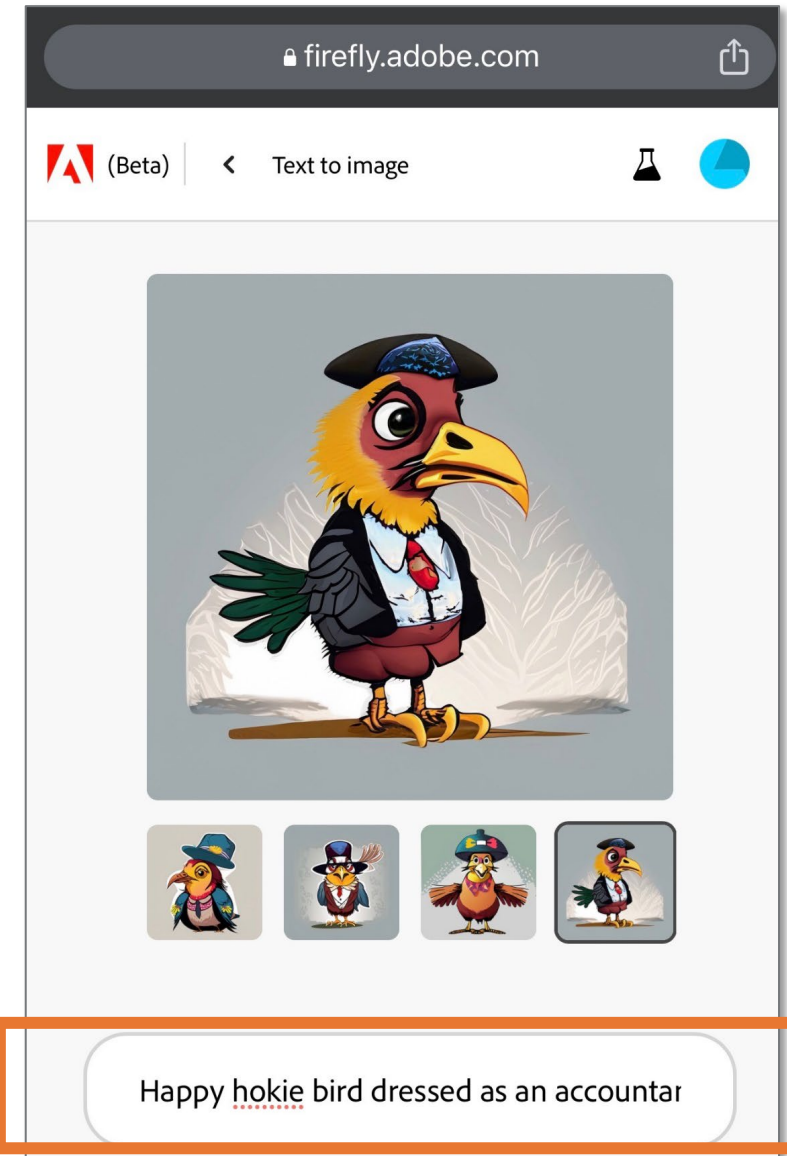
Prompt: A Hokie bird accountant, digital art



Why do they all seem so angry?

**Training data reflects biases present in real world data.**

Perhaps we should reflect on outsider perceptions of accountants as we consider solutions for addressing the recent declines in accounting major enrollments.



# Thinking about underlying training data

1. Can we predict when we'll receive biased responses?
2. How will we know when we receive biased responses?

## Recently:

- OpenAI bot will crawl the web. Newsrooms and content creators opting out
- Others signing content licensing deals
- Reddit users mutiny over Reddit new API pricing (i.e., charging OpenAI to crawl site content)
- “As an AI language model...” phrase growing in prevalence online

**Going forward...**How will training data differ for future LLM versions?

3. Will the heuristics we develop today apply in the future?



# Tool for thinking about underlying data

[chatGPT4google.com](https://chatGPT4google.com)\*

Browser extension that shows ChatGPT results right next to google search results

## Options:

- Query ChatGPT on every search
- Query ChatGPT using “?” at end of search

\* Do as I say, not as I do—I have not reviewed the privacy policy of this 3<sup>rd</sup> party extension. Use with caution.

The screenshot shows a Google search interface. The search bar contains the text "rules for rock, paper, scissors, lizard, spock?". Below the search bar, the results show "About 78,100 results (0.40 seconds)". The first result is from wikiHow, titled "How to Play Rock Paper Scissors Lizard Spock - wikiHow". The second result is from Instructables, titled "How to Play Rock, Paper, Scissors, Lizard, Spock : 6 Steps".

Overlaid on the right side of the search results is a ChatGPT extension window. The window has a header with icons for ChatGPT, GPT-4, Claude2, and Bard. The main content area is titled "Response" and contains the following text:

Certainly! "Rock, Paper, Scissors, Lizard, Spock" is an extended version of the classic game "Rock, Paper, Scissors." It was popularized by the TV show "The Big Bang Theory." In this version, there are additional hand gestures and rules. Here are the basic rules:

1. Rock crushes Scissors.
2. Scissors cuts Paper.
3. Paper covers Rock.
4. Rock crushes Lizard.
5. Lizard poisons Spock.
6. Spock smashes Scissors.
7. Scissors decapitates Lizard.
8. Lizard eats Paper.
9. Paper disproves Spock.
10. Spock vaporizes Rock.

Each item can defeat two other items, be defeated by two other items, and is neutral against one item. Players choose one of the five items and reveal their choices simultaneously. The winner of each round is

# Open-Source v. Closed-Source Models

## Closed-Source

### Examples:

- OpenAI GPT models
- Anthropic Claude
- Google Gemini
- Cohere R

### Pros:

- Largest models, best performance
- Minimal setup relative to local development

### Cons:

- Cannot run locally
- Your data can be used to train the current model and future models

## Open-Source Models

### Examples:

- Mistral Large 2
- Meta Llama 3.1
- Google Gemma2

### Pros:

- Can run locally
- No concern over your data being used in model training
- Models are “static” – don’t have to worry about performance degradation with more human interaction

### Cons:

- Lots of DIY



# Example use-case: Information Retrieval

## RAG Agents – information retrieval from

- Firm policies
- HR handbook
- Rules and regulations
- CPA Exam review materials
- Client information

## Enhance CRM

- Built-in features in practice management tools
- Local files and folders

# Example use-case: Information Retrieval

## CRM Example

- **Keep ALL client info in one folder**
- **Separate folders for storing documents thematically**
  - Sub-folders and filenames named descriptively
- **One text file with descriptive notes about client**
  - Information to keep handy about client
  - Document interactions (including emails) with the client
  - Document internal interactions re: the client
  - Personal notes about the person—birthdays, wife/kids names, business plans, observations about the person's character, temperament, and personality, etc.

# Example use-case: Information Retrieval

## CRM Example

### Client information retrieval

- When was the last time I met with Joe from Joe's Plumbing?
- When is X due for Joe's Plumbing?
- How many times did I ask Joe for <document>?

### Aggregate Information Retrieval

- Which clients have not been contacted over the past 3 months?
- Which clients have X due next month?
- Which clients require more than 2 follow-up emails to request <document>?

# Example use-case: Information Retrieval

**Vendors might have these features, but...**

- Information is in proprietary environments
- Scattered across multiple products

**Automate frequent data exports across vendors and products**

- Consolidate your information locally to use for RAG
- Take on some cyber risk, but also mitigate others

# Example use-case: Employee Training

## Role-playing

1. Have ChatGPT take on the role of a difficult client
  - Young employees → practice difficult client conversations
  - Experienced employees → practice sales calls
  - Audio recording of real calls to ask ChatGPT for feedback
    - Get appropriate consent for this exercise
2. Practice public speaking
  - PowerPoint built-in feature
    - Slide Show → Rehearse with coach
  - Feedback from ChatGPT