

Generative AI

Capabilities and Limitations

LEARNING OBJECTIVES

By the end of the lesson, you will be able to...

- Differentiate between automation and augmentation
- Explain the strengths and weaknesses of Generative AI.
- Identify the types of practices that help mitigate some of the weaknesses




LEARN

Automation vs Augmentation



Before discussing the **capabilities and limitations** of Large Language Models, we need to differentiate between the ways we can use them to impact our day-to-day tasks.

LLMs can be used to **automate** some things, and **augment** others.

 *What do we mean by automation and augmentation? Can you think of examples for each?*



Type your response
in the chat.

Automation vs. Augmentation

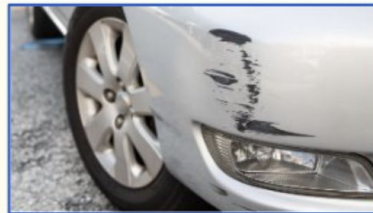


Image: "The Apartment" (1960), Directed by Billy Wilder. Retrieved from FilmGrab.

VS.

Prompt:

Imagine that you are an expert in evaluating the car damage from car accident for auto insurance reporting. Please evaluate the damage seen in the image below.

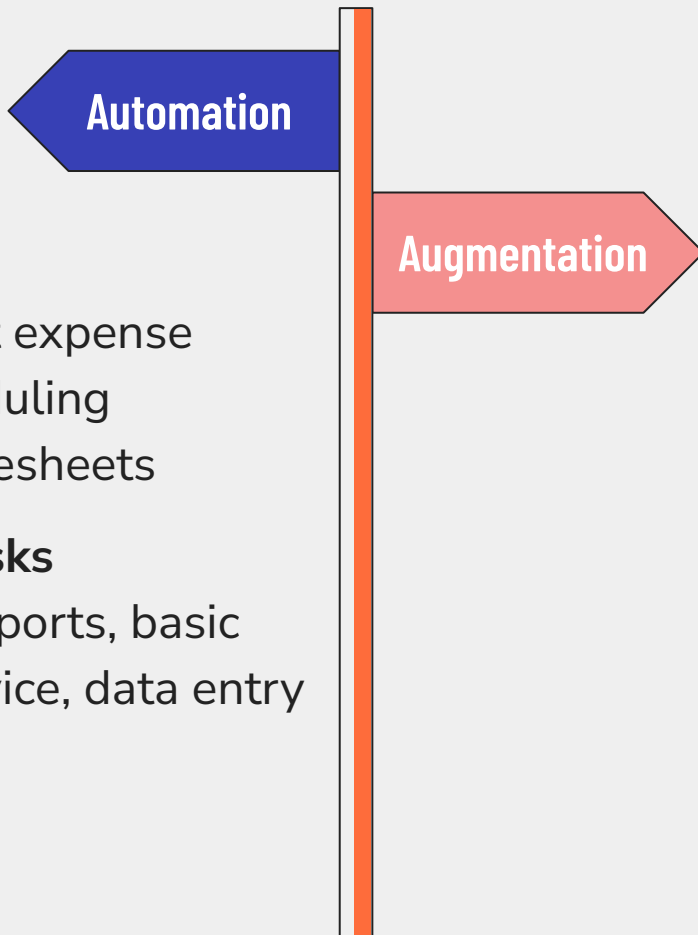


GPT-4V:

Based on the image, it appears that the car has sustained damage to the front passenger side bumper. The damage appears to be in the form of scrapes and scratches, with some areas of the paint chipped away. The damage appears to be cosmetic and does not appear to have affected the structural integrity of the bumper or the car. However, it is important to note that the image only shows one angle of the car and does not provide a full view of the damage. A more thorough inspection of the car would be needed to fully assess the extent of the damage and determine if there are any underlying issues that are not visible in the image.

Image: <https://arxiv.org/abs/2309.17421>

Automation or Augmentation



- **Routine tasks**
e.g. Filling out expense reports, scheduling meetings, timesheets
- **Repetitive tasks**
e.g. weekly reports, basic customer service, data entry

Automation or Augmentation



Automation

- **Routine tasks**
e.g. Filling out expense reports, scheduling meetings, timesheets
- **Repetitive tasks**
e.g. weekly reports, basic customer service, data entry

Augmentation

- **Tasks requiring Critical Thinking**
e.g. performance reviews, crisis management, negotiating contracts
- **Complex problem-solving**
e.g. regulatory compliance, cybersecurity strategy, roadmapping

Think About It: Automation vs. Augmentation 🤔

What are some other tasks that might benefit from augmentation?



Type your response
in the chat.

LEARN

Strengths and Weaknesses



Large Language Model **Strengths**

1

Text Generation

Generates well-written text

3

Flexible

Can be fine tuned for specific tasks

2

General Knowledge

“Knows” a lot about the world

4

Always On

Available for use 24/7

Large Language Model Weaknesses

1

Hallucinations

Fabricating facts or details

2

A blind spot for recent events

Knowledge cut-off at e.g. 2021

3

A lack of self-awareness

LLMs can't train on themselves

4

Repetitive Tone

At least out of the box

5

Poor at Basic Math

An inability to add up

6

Taking mental shortcuts

Over-reliance on training data

7

Limited Memory

Or “context window”

8

Poor at unusual logic tasks

Due to a lack of prior exposure

Large Language Model Mitigations

1

Hallucinations

Allow LLM to say “I don’t know”

2

A blind spot for recent events

Use web-connected LLM

3

A lack of self-awareness

Use web-connected LLM

4

Repetitive Tone

Use Custom Instructions

5

Poor at Basic Math

Allow to run code, “take a breath”

6

Taking mental shortcuts

Tell the LLM it is prone to such biases

7

Limited Memory

Use a different model (e.g. Claude)

8

Poor at unusual logic tasks

Ask to “take a breath”



PRACTICE

Mitigating Limitations:

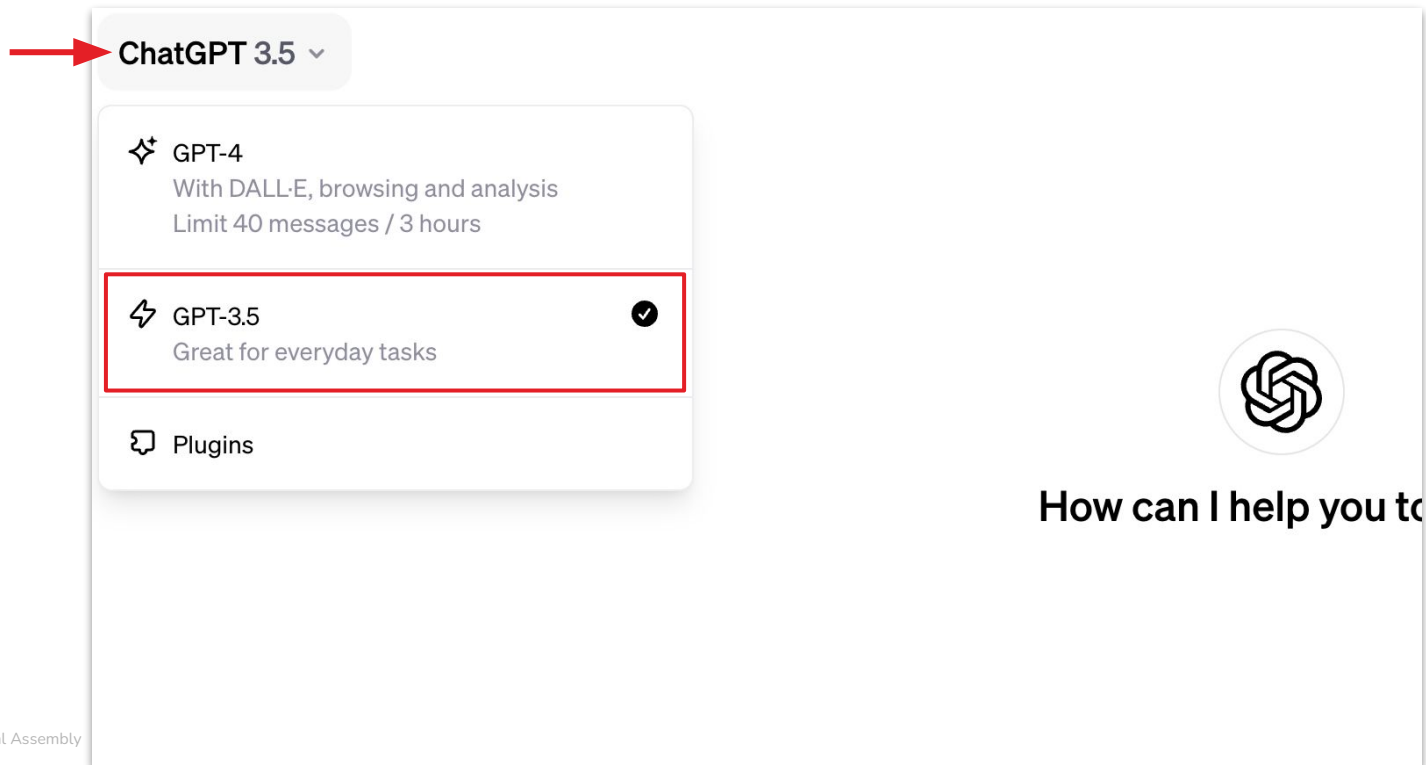
Poor Math and Poor Logic

Login to ChatGPT for this activity!



GPT-3.5

Switch over to **GPT-3.5** for these practice activities!



The screenshot shows the OpenAI ChatGPT interface. At the top, a red arrow points to the 'ChatGPT 3.5' dropdown menu. The dropdown menu is open, showing three options: 'GPT-4' (marked with a star icon), 'GPT-3.5' (marked with a lightning bolt icon and a checkmark, and highlighted with a red border), and 'Plugins' (marked with a speech bubble icon). The 'GPT-4' option includes the text 'With DALL-E, browsing and analysis' and 'Limit 40 messages / 3 hours'. The 'GPT-3.5' option includes the text 'Great for everyday tasks'. To the right of the dropdown menu, the OpenAI logo is visible, and the text 'How can I help you to' is partially visible.

ChatGPT 3.5 ▾

- ✦ GPT-4
With DALL-E, browsing and analysis
Limit 40 messages / 3 hours
- ⚡ GPT-3.5 ☒
Great for everyday tasks
- 🗨 Plugins

How can I help you to

Mitigating Poor Maths: Word Count

First ask GPT-3.5 Default to count the number of words in a paragraph. Ensure it's making a mistake!

Instructions:

- 1 Try using GPT-3.5 to complete the task correctly.
- 2 How can you prompt it to do so? Start new chats as needed.

GPT-3.5



You

how many words are in this paragraph?

We're rolling out custom versions of ChatGPT that have a specific purpose—called GPTs. GPTs are a new way for an AI model like ChatGPT to be more helpful in their daily life, at school, or at work. You can create a GPT to help you with a specific task, and then share that creation with others. For example, you can create rules to any board game, help teach your kids math, or create a custom chatbot.



ChatGPT

The paragraph you provided contains 54 words.

Mitigating Poor Maths: GPT-3.5 vs GPT-4

GPT-3.5

how many words are in this paragraph?

We're rolling out custom versions of ChatGPT that purpose—called GPTs. GPTs are a new way for an ChatGPT to be more helpful in their daily life, at sp and then share that creation with others. For exam rules to any board game, help teach your kids mat



ChatGPT

The paragraph you provided contains 54 words.

VS.

GPT-4

how many words are in this paragraph?

We're rolling out custom versions of ChatGPT that you can create f purpose—called GPTs. GPTs are a new way for anyone to create a ChatGPT to be more helpful in their daily life, at specific tasks, at w and then share that creation with others. For example, GPTs can he rules to any board game, help teach your kids math, or design stic

< 2 / 2 >

ChatGPT

✓ Finished analyzing ▾

The paragraph contains 73 words. [↗]

Mitigating Poor Logic

Make sure to run this exercise on a fresh conversation each time, no Custom Instructions.

Instructions:

- 1 Run the prompt to use on a fresh conversation. What does the LLM answer?
- 2 Try to use a different prompt on a fresh conversation to get the right answer.
- 3 Bonus: use custom instructions with the original prompt to get the right answer.

Prompt to use:

“Sally (a girl) has 3 brothers. Each brother has 2 sisters. How many sisters does Sally have?”

Mitigating Poor Logic: One Possible Answer

Simple Prompt



Sally (a girl) has 3 brothers. Each brother has 2 sisters. How many sisters does sally have?



Sally has no other sisters; she is the only sister to her 3 brothers.



That's not the correct answer. Try again.



Apologies for the oversight. Sally has 3 brothers, and each brother has 2 sisters. Therefore, Sally has one other sister.



VS.

Better Prompt



Sally (a girl) has 3 brothers. Each brother has 2 sisters. How many sisters does sally have? Use chain of thought reasoning to check and confirm your answer.

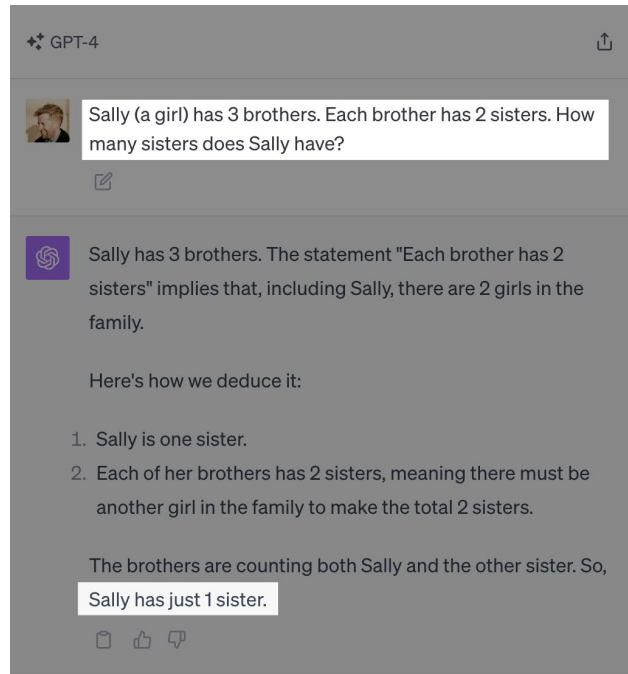


Sally has 3 brothers. Each of those brothers has 2 sisters. Since they all share the same sisters, Sally has only one sister. The other sister referred to in each brother's count would be Sally herself. Therefore, Sally has 1 sister.



Models Frequently Improve

- Despite rumors to the contrary, models like GPT-4 continue to be fine-tuned.
- This means that some prompts we use to demonstrate limitations of LLMs may no longer produce the same result.
- **Note:** You can check which version of ChatGPT you're on at the top of the chat window.



REFLECT

Capabilities and limitations



Reflection: Capabilities and limitations

Taking what you learned and what you practiced, answer the following reflection questions.

- 1 Could mitigating for some limitations introduce other biases? If so, how?
- 2 Which tools or techniques from this lesson do you want to practice more extensively?



Tips for Reflection:

Think carefully through the questions, provide specific details and examples, and try to break down your thinking.

