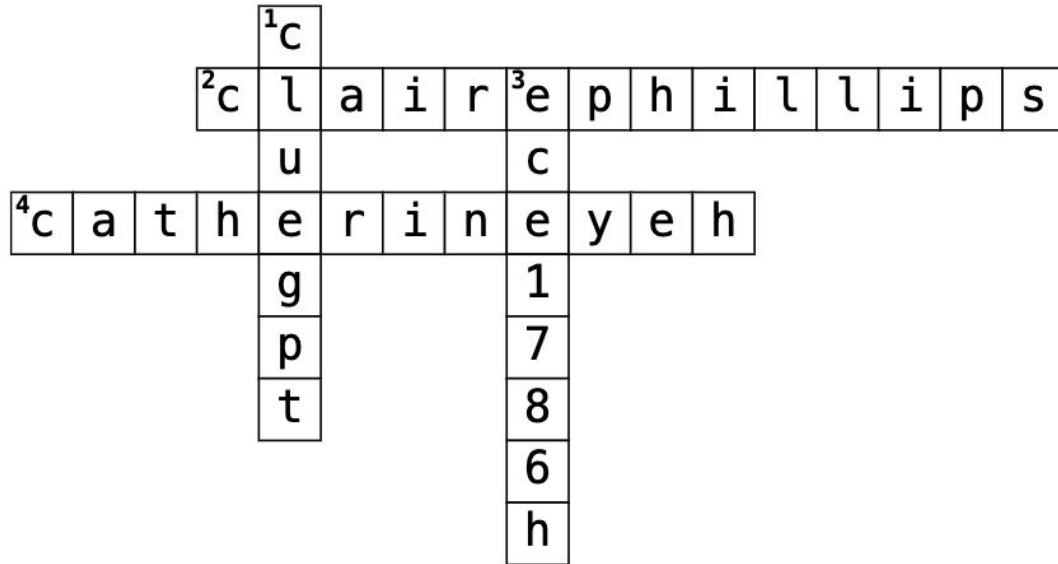


# ClueGPT



## Across

2. Team Mate

4. Team Mate

## Down

1. Project Name

3. Course Code

# Goal & Motivation

- Goal of ClueGPT: supply people solving crosswords with a second clue generated by AI.
- Motivation:
  - Crossword solvers often encounter difficulties when solving puzzles.
  - Limited options available when stuck, often leading to frustration.
  - Many may resort to searching for answers online.
  - This situation creates a feeling of choosing between cheating and giving up.
- ClueGPT is an alternative to this and is a tool which can be used to get help without being given the answer.

# Data

- **Dataset:** New York Times Crossword Repository by The Text Machine Lab at the University of Massachusetts.

Clue	Answer
Line at an airport	limos
Kind of omelet	egg white
Susa was its capital	elam

# Data Cleaning

Several errors and clues which referenced other clues needed to be removed from the dataset.

Clue	Answer	Type of outlier	Removed
in 23-Down	\nfreshened\nplateau\nalo ha\ntreaty result\nad...	Entry error	Yes, through preprocessing
Slot-car track section	b&orailroad	Entry error	Yes,through preprocessing
Ohio city whose name means "hospitality" in Greek	ho hoho hoho hoho hoho hoho hoho hoho hoho hoh...	Entry error	Yes, through preprocessing

# Data Processing

## Answer Processing:

- Remove special characters
- Remove spaces
- Make all letters lowercase

## Clue Processing:

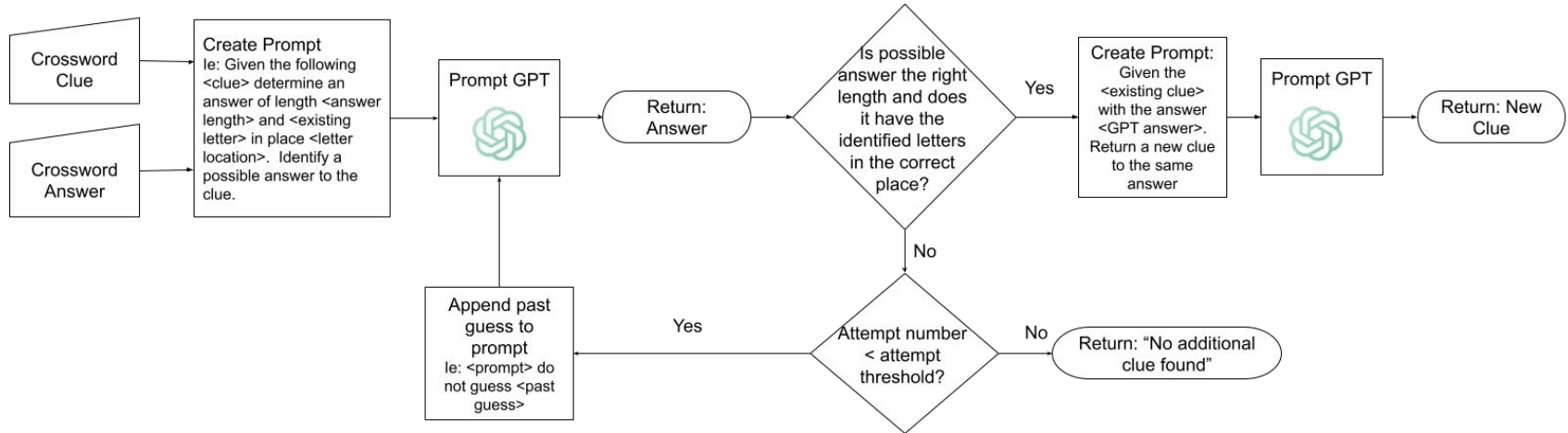
- Get length of answer
- Get included letters (randomly selected from answer with probably n)

Crossword clue	num_of_char	included_letters	Answer
Line at an airport	5	{4: 's'}	limos
Kind of omelet	8	{0: 'e', 1: 'g', 2: 'g', 3: 'w', 6: 't'}	eggwhite
Susa was its capital	4	{1: 'l', 2: 'a'}	lima
Absolutely out of the question!	7	{0: 'n', 2: 's', 3: 'i', 4: 'r', 6: 'e'}	nosiree
Inky	4	{0: 'e', 2: 'o'}	ebon
Green wheels	6	{0: 'e', 1: 'c', 2: 'o'}	ecocar

# Software/Platforms Used

- GPT 4 Model
- Python
- Google Colab
- Google Sheets

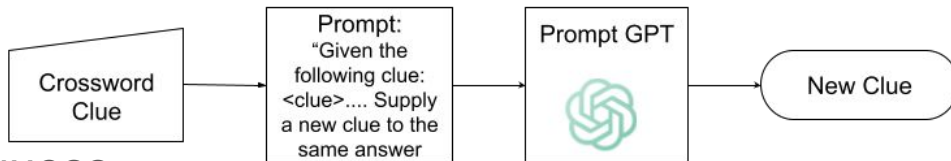
# ClueGPT Model Architecture



# Comparing Performance

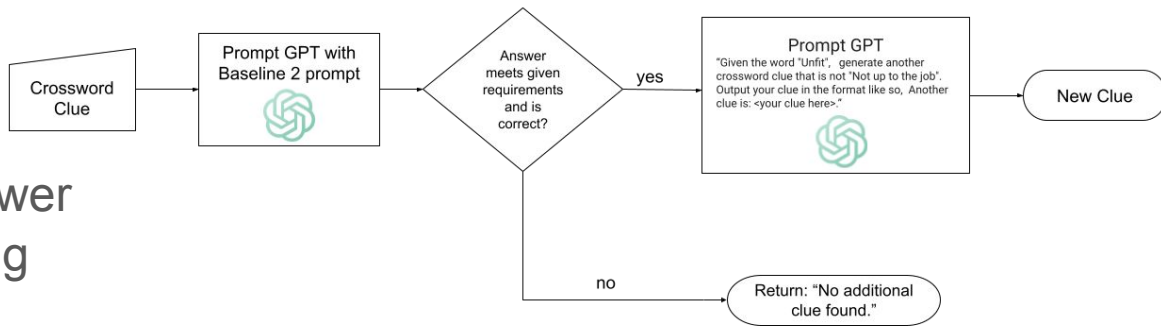
## Version 0

- Zero shot prompt
- No validation of the models guess



## Version 1

- No iteration to finding answer
- Chain of thought prompting





# Measuring Success

The output for each clue was measured qualitatively via the following definition of success: A clue that leads to correct answers, but cannot give away the answer explicitly (eg. answer: ‘ous’; clue: Common ending for words like "dangerous" or "hilarious".).

We compare the success rate of ClueGPT with the two other versions described earlier.

	Version 0	Version 1	ClueGPT
Percent of generated clues which meet the definition of success	65.7%	100%	100%
Percent of clues the model did not generate a clue.	0%	40%	33%
Percent of clues generated which meet the definition of success and are not a “No clue generated” response	65.7%	60%	67%

# Example of Output

ClueGPT V1 had no incorrect clues created given our best prompts.

Original Clue	Clue GPT Guessed Answer	ClueGPT Output	Answer
Au naturel	bare	Another clue is: "Stripped of covering"	bare
___ said ...	sooner	No clue available.	easier
Place of rest	haven	Another clue is: "Sanctuary or safe place"	haven

# Discussion

Our best results came from ClueGPT.

Although version 0 had a very close rate of correct clues it produced incorrect clues while, thanks to its architecture, version 1 supplied user with a “no clue available” response.

This response is better than an incorrect guess and incorrect clues would confuse users and hurt their experience with the tool.

	Input Clue	Output Clue	Actual Answer
Version 0	Cash in the music business	Symbol of revenue in the entertainment industry	rosanne
Version 1	Minstrel, often	No clue available	lutist

Incorrect clue

# Learning

- We didn't consider that GPT models will output letters with accents.
- System prompt messages are very useful in creating output of the correct form.
- Tuning the temperature and top\_p values are ineffective at producing a wide range of coherent one word responses. These parameters are likely more useful when longer form answers are desired.
- Model behaves differently depending on the number of letters passed in.