

11-667 Homework 1

Krish Rana, krana

Due date: September 26, 2023 at 2 PM

1 Observing the Impact of Decoding Strategy

1.1 Rolling a Twenty-Sided Die

Question 1: When using $\text{top_p} = 1.0$, the LLM is not equally likely to generate all the outcomes. If we set $\text{top_p} = 1.0$, we include the entire vocabulary in the set of possible candidates for next_token . However they are sampled according to the probability distribution computed by the model. If the model is trained to the optimum, then we can expect it to generate all outcomes with equal probability. We do not expect such performance from GPT-3

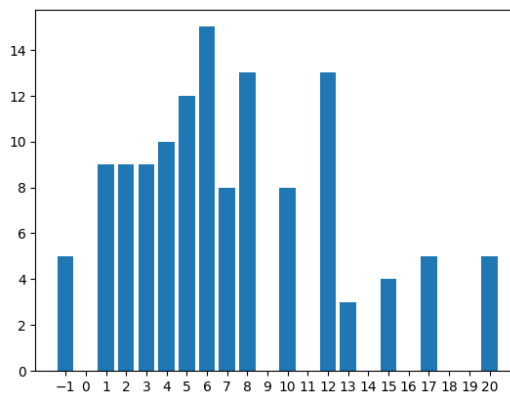
Question 2: Answer:

1. $p = 0.7$

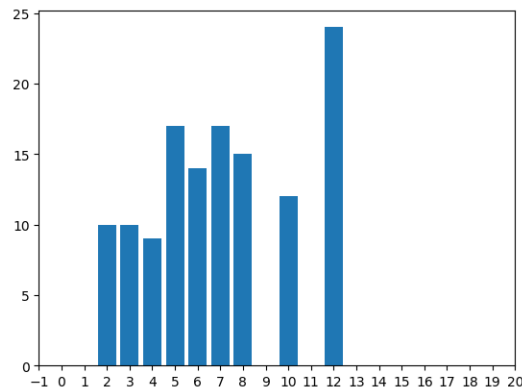
With $p = 0.7$, we see that the LLM produces 70% of the expected outcomes. The distribution of the outcomes is, however, uneven. We see that the model outputs the number 6 maximum times (15) followed by 8 (13) and 12 (13). It outputs number 13 the least (3) time. But as compared to $p = 1.0$, the number of invalid outputs decreases -1: 5 in line with our expectation.

2. $p = 0.5$

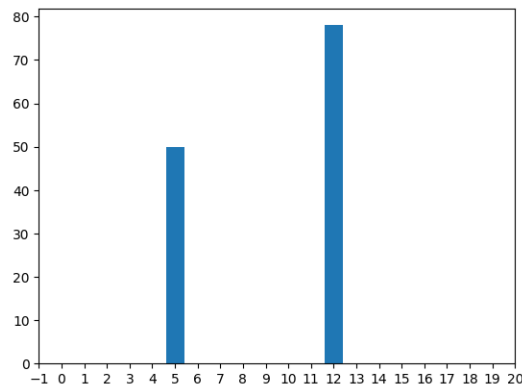
With $p = 0.5$, we see that the model generates 45% of the expected outcome. It outputs the



Figur 1: Outcome Distribution



Figur 2: Outcome Distribution



Figur 3: Outcome Distribution

number 12 the most number of times followed by 5. The output distribution is still uneven but it does not generate any invalid outcome.

3. $p = 0.1$

With $p = 0.1$, we see that the model generates only 1% of the expected outcome. It outputs the number 12 the most number of times followed by 5. For a 20-sided die, the probability of observing any number between 1 and 20 is 5%. Therefore the idea of nucleus sampling works out in this case as we require at least 2 numbers to reach the 0.1 threshold. The model mostly biases towards 12 and 5.

1.2 Longform Generation

1.2.1 Question 1

```
Prompt = Let me tell you a story of when I was in a spy operation in Mars, top_p=1.0,  
type_token = 0.70
```

, I was only sixteen. So basically, we had to pick one mission, and there were about twenty missions to choose from, and most of my team got called on another mission, but there was this sniper person and a team leader. So we went on this mission. It lasted about forty minutes. We did it. We stole the package and got out.”

“What was this package?”

“I guess you could say it was a computer chip that was going to crash, bringing down their funding. Something like that. And it was a very daring thing to do, just sort of sneaking around. I don’t think I allowed myself to feel too much, not like when I got physically hurt during the games. After the fragmenting of the second enemy group, I didn’t really have too many feelings. That was because, if I were to fall in love with someone on Mars and then she get killed, it would just be awful having to see her die in battle. That would just be extremely painful for me.”

“So it’s kind of hard to know if that experience of battle traumatized you?”

“No, I don

```
Prompt = Let me tell you a story of when I was in a spy operation in Mars, top_p=0.0,  
type_token = 0.09
```

. I was in a mission to find out what the Martians were up to. I was in a space ship
and I was in a space suit. I was in a space suit because I was in a space ship. I was in
a space ship because I was in a space suit. I was in a space suit because I was in a
space ship. I was in a space ship because I was in a space suit. I was in a space suit
because I was in a space ship. I was in a space ship because I was in a space suit. I
was in a space suit because I was in a space ship. I was in a space ship because I was
in a space suit. I was in a space suit because I was in a space ship. I was in a space
ship because I was in a space suit. I was in a space suit because I was in a space ship.
I was in a space ship because I was in a space suit. I was in a space suit because I was
in a space ship. I was in a space ship because I was in a space suit. I was in a space
suit because I was in a space ship. I was in a space ship because I was in a space suit.

I

gibberish repeated sentences "Jokes for kids". Maybe the model has seen a lot of training examples of spy ops on Mars in the context of a jokes book for kids leading to the repeated statements. On reducing the top_p further to 0.0, we see that it instantly starts repeating the tokens and the generation is gibberish

1.2.2 Question 2

Speech: Bush's speech on Iraq War

Prompt = Just 2 hours ago, allied air forces began an attack on military targets in Iraq and Kuwait. These attacks continue as I speak. Ground forces are not engaged., top_p=1.0

invaded a small and helpless neighbor. His goal was to start an ethnic civil war between Shiite and Sunni Moslems that would force us to come in on the side of one group or the other. That would be a mistake we would regret for the next hundred years."

Tonight, the entire United States is on the threshold of extraordinary times. The crisis, while unprecedented, is one that Americans have faced before.

"It is is none other than President George Washington who, looking out over all of America in 1796 at a confrontation between his late-republican nation and the expansionist designs of Alexander Hamilton's infant republic, observed that war gave every member that he had taken the finest opportunity for consideration of what sort of government he wished to live under.

Åaging war through decision, 'when there are none to appeal to but God and the representatives,' President Washington advised, gave the people the opportunity to put into practice perpetual common sense—the potential benison of crisis. No bad mistake is so bad that a country cannot survive it. No calamity is so great that a country cannot endure it. If we really want to avoid all of the terrible consequences of strife, let peace triumphant indeed reign among ourselves it can only

Prompt = Just 2 hours ago, allied air forces began an attack on military targets in Iraq and Kuwait. These attacks continue as I speak. Ground forces are not engaged., top_p=0.5

invaded a small and helpless neighbor. Kuwait had been a peaceful and secure country until Iraq invaded. Iraq's purpose was to take over Kuwait and then to use Kuwait's oil wealth to finance the rebuilding of his military machine.

The world had tried to stop the aggression of Saddam Hussein. We and our allies have pleaded with him to respect the legitimate boundaries of Kuwait. We have warned that his aggression would not stand. We have given him time to reconsider his actions. He has not.

And so we have decided to act and act now.

At the same time, the states of the Gulf Cooperation Council have decided to act and act now. The governments of Egypt, Saudi Arabia, Morocco, Oman, Qatar, and the United Arab Emirates have decided to take immediate action to support the effort to defend Kuwait and the legitimate government of Kuwait.

The states of the Arab League have decided to act and act now. The government of Syria and the government of Jordan have decided to take immediate action to support the effort to defend Kuwait and the legitimate government of Kuwait.

The governments of Bahrain, Kuwait, Qatar, Oman, Saudi Arabia, and the United Arab Emirates have decided to take immediate action to support the effort to defend Kuwait and the legitimate

Prompt = Just 2 hours ago, allied air forces began an attack on military targets in Iraq and Kuwait. These attacks continue as I speak. Ground forces are not engaged.,
top_p=0.0

invaded a small and helpless neighbor. The United States responded with decisive force. We responded with firmness. We responded with resolve. This will not stand. This will not stand, because we will not allow it to stand.

The world is also responding to this aggression. The United Nations Security Council has condemned this invasion. The Arab League has condemned this invasion. The Organization of American States has condemned this invasion. The European Community has condemned this invasion. The Organization of African Unity has condemned this invasion. The Non-Aligned Movement has condemned this invasion. The Organization of Islamic Conference has condemned this invasion. The United States and the United Kingdom and France and Egypt and Saudi Arabia and Kuwait and Bahrain and Qatar and the United Arab Emirates and Oman and Yemen and Jordan and Morocco and Algeria and all the countries of NATO have condemned this invasion.

The world has condemned this invasion. And yet Iraq persists.

The world has spoken, Iraq, and your aggression will not stand.

The world has spoken, and the world demands the immediate and unconditional withdrawal of Iraqi forces from Kuwait.

The world has spoken, and the world will be heard.

The world has spoken, and the world will be heard.

The world has spoken,

For every iteration of top_p, the model generates a correct continuation of the first statement ("invaded a small and helpless neighbour") but the generation quickly diverges after the first sentence for all iterations of top_p. For top_p = 1.0, the model changes the topic to include speeches of George Washington. There are parallels in George Bush's speech where he quotes Thomas Paine. When the top_p is reduced to 0.5, the model adheres to the topic. It does not generate the same exact speech but the speech is parallel to what we find in the original content. It suggests that the model has not simply memorised the training data (it is highly likely that Bush's speech will be in GPT-3 training data) but generalizes well while maintaining coherency in generation.

1.2.3 Question 3

Prompt = Let me tell you a story of when I was in a spy operation in Mars, top_p=0.0, frequency_penalty = 1.0, type_token = 0.675

. I was in a mission to find out what the Martians were up to. I had to go undercover as a Martian and infiltrate their society.

I went into this bar and ordered some drinks, but they didn't have any alcohol on Mars! So I had to order something else, which was called "Martian Juice". It tasted like water with a hint of lemon juice in it.

After drinking my Martian Juice, I started talking with some of the other Martians at the bar about how they were going to take over Earth one day and make us all slaves! They said that we humans are too stupid for them so they will just kill us all off instead! They also said that we humans taste good when you cook us up on their planet! That is why they want to eat us all up!

I told them that if they wanted me dead then why did they let me live? The bartender told me that he knew who I really was because he saw my spaceship land outside his window earlier today when I arrived on Mars. He said he would let me live if I gave him \$100 million dollars from Earth or else he would tell everyone about my true identity as an alien spy from Earth who came here for

The type_token score improves drastically from 0.09 to 0.675. the story also becomes more coherent and introduces elements such as Martian Juice which were not previously seen.

2 Measuring Perplexity

Poem: The Road Not Taken - by Robert Frost

Poem Base, perplexity = 1.42

TWO roads diverged in a yellow wood, And sorry I could not travel both And be one traveler, long I stood And looked down one as far as I could To where it bent in the undergrowth;

Then took the other, as just as fair, And having perhaps the better claim Because it was grassy and wanted wear; Though as for that, the passing there Had worn them really about the same,

And both that morning equally lay In leaves no step had trodden black.

Oh, I marked the first for another day! Yet knowing how way leads on to way I doubted if I should ever come back.

I shall be telling this with a sigh Somewhere ages and ages hence: Two roads diverged in a wood, and I, I took the one less traveled by, And that has made all the difference.

2.0.1 Question 1

Poem Typo, perplexity = 10.41

TWO roads diverged in a mellow wood, And sorry I could not travel booth And be one trabelloe, long I stood And looked dpwn one as far as I could To where it spent in the undergrowth;

Then took the otter, as just as fair, And having perhaps the batter claim Because it was greasy and wanted wear; Though as for that, the possum there Had worn them really abut the same,

And booth that morning equallly lay In lweabewrs no step had trodden back.

Oh, I maked the fist for another day! Yet knowing how way leeds on to way I doubted if I could ever come back.

I shall be telling this with a sign Somewhere ages and ages fence: Two roads diverged in a hood, and I, I shook the one less traveled by, And that has made all the diferance.

On making typos in the poem, the perplexity increases to 10.41. One reason for this might be that the model had seen this poem in its training without any typos, but our variant of the poem with the typos is something the model is seeing for the first time.

2.0.2 Question 2

Poem Shuffle, perplexity = 13.48

Oh, I maked the fist for another day! Yet knowing how way leeds on to way I doubted if I could ever come back.

And booth that morning equallly lay In lweabewrs no step had trodden back.

TWO roads diverged in a mellow wood, And sorry I could not travel booth And be one trabelloe, long I stood And looked dpwn one as far as I could To where it spent in the undergrowth;

I shall be telling this with a sign Somewhere ages and ages fence: Two roads diverged in a hood, and I, I shook the one less traveled by, And that has made all the diferance.

Then took the otter, as just as fair, And having perhaps the batter claim Because it was greasy and wanted wear; Though as for that, the possum there Had worn them really abut the same,

On shuffling the contents of the poem, the perplexity increases yet again to 13.548. It shows that the positional embeddings work and the order of the input sentence matters.

2.0.3 Question 3

Poem Mimic, perplexity = 2.012

TWO roads seperated in a yellow wood, And sorry I could not travel both And be one traveler, long I stood And looked down one as far as I could To where it bent in the undergrowth;

Then took the other, as just as fair, And having perhaps the stronger claim Because it was grassy and wanted wear; Though as for that, the passing there Had worn them really about the same,

And both that morning equally lay In leaves no step had trodden black.

Oh, I marked the first for another day! Yet knowing how way leads on to way I doubted if I should ever come back.

I shall be revealing this with a sigh Somewhere ages and ages hence: Two roads seperated in a wood, and I, I took the one less traveled by, And that has made all the difference.

The perplexity does not change drastically. It goes slightly up to 2.012. It might be because semantically the meaning of the poem is still the same and the model has learned the similarity between these words and thus id able to identify that it has seen a similar sequence before.

2.0.4 Question 4

As this poem might be seen in GPT-3's training set, this might represent a data leak. Perplexity for such poems will be low as the model is already trained on it. But if the model is presented with a new poem, if the given poem has a similar semantic meaning to these famous poems, then the perplexity for that sequence might be low still. But if the poem is entirely new with a completely different structure, then the perplexity might increase.

3 Experimenting with Few-Shot Prompting

3.1 Few-shot Learning for Classification Tasks

3.1.1 Question 1

Table 1 shows the n-shot accuracies for the best prompt as per the validation set performance

3.1.2 Question 2

I experimented with the following prompts as shown in Table 2.

The last prompt was inspired from the Large Language Models as Optimizers paper where they report an 8% increase in evaluation scores on GSK8 and 50% increase in BigBench benchmark as compared to human-tuned prompts. Even for our study, we see a 5% increase

Prompt	Accuracy
Zero-Shot Prompt	89%
1-Shot Prompt	93%
3-Shot Prompt	83%
5-Shot Prompt	89%

Tabell 1: Final Accuracy for n-shot examples with the best prompt

Prompts	Dev - accuracy
"Given the following Premise and Cause or Effect, label which of the choices seem correct: "	80%
"Lets think step-by-step and given the following Premise and Cause or Effect, label which of the choices seem correct: "	79%
A little bit of arithmetic and a logical approach will help us quickly arrive at the solution to this problem. Given the following Premise and Cause or Effect, label which of the choices seem correct: "	85%

Tabell 2: Validation Accuracies for different prompts

in accuracy on the validation set. Interestingly, the Lets think step-by-step did not improve the accuracy. In addition to these prompts, I also updated the templates to include effects as well. The accuracy reported are using the updated template.

3.1.3 Question 3

I believe the quality of examples that we give to the model affects the accuracy the most. The instruction prompt and the template to structure the premise, cause or effect has a huge impact on the model's accuracy. However, mislabelling the examples does not have any effect on the model accuracy. 1-shot performance of the model with mislabelled examples is the same as before.

3.1.4 Question 4

Most of the questions that the model gets wrong are a bit ambiguous. for example, PREMISE: I pulled the rubber band. CHOICE 1 : It flung across the room. CHOICE 2 : It stretched. PREDICTED: choice 1 The first choice seems plausible here but it is labeled wrong.

3.1.5 Question 5

We see that all the three models perform no better than random guessing. I believe that our prompt and eval templates are highly optimised for the Davinci model (especially the

Model	Accuracy
ADA	55%
Curie	55%
Babbage	55%
Davinci	93%

Tabell 3: Accuracy across different sizes for 1-Shot alt_prompt

prompt) due to which these smaller models fail. To perform accurately on the classification task I believe that we should go with the text-davinci-002

3.2 Few-shot Learning for Generation Tasks

3.2.1 Question 1

Task: Add space between each character of the string I decided on the following prompt with 3-shot examples:

Prompt: "Convert String1 to String2"

Structure:

"Convert String1 to String2

Example1

Example2

Example3

String1: queryString, String2:

"

The idea is to have a structured input for the model to understand. In practice, it works well with a small test set of about 8 sentences.

3.2.2 Question2

For the current language models, the tokenization of text occurs on a sub-word level using algorithms such as BPE, SentencePiece etc. As the inputs of the models are sub-word units, we can hypothesise that character-level manipulations might be difficult for these models to process. However the LLMS are found to perform well on this task.

4 Investigating Knowledge Across Different Model Sizes

4.1 Question 1:

Model = Davinci, Wikipedia: CMU

In 1967, the Carnegie Institute of Technology merged with the Mellon Institute of Industrial Research to form Carnegie Mellon University. The university's 140-acre (0.57 km²) main campus is 3 miles (4.8 km²) from Downtown Pittsburgh and abuts the Carnegie Museums of Pittsburgh, which includes the Andy Warhol Museum.

Carnegie Mellon consistently ranks among the 25 top research universities in the United States. It is a top-tier university in computer science, engineering, robotics, and business. CMU has been home to many technology pioneers such as Herbert A. Simon, Allen Newell, Raj Reddy, and John Harsanyi. The university is also one of the first in the nation to develop a degree in artificial intelligence. CMU has produced notable alumni in a broad range of disciplines, including Nobel laureates Herb Simon, John Nash, and Vernon Smith; astronauts Alan Shepard and Rhea Seddon; businessman Bill Gates; Apple co-founder Steve Wozniak; and actors Zachary Quinto and Alan Cumming.

Model = ADA, Wikipedia: CMU

The following year, the Institute of Americanveda, now known as Carnegie Mellon University, opened. It was this first modern university that helped create the image of a.s.a.s.p. and led to the establishment of other institutions of higher education across the country.

The campus in Pittsburgh, PA.

The university has a total area of which is is approximately with a population of approximately 50,000.

It consists of two main campuses: the Public Learning Campus which offers about 60 courses per year, and the Private Learning Campus which offers about 30 courses per year. The Public Learning Campus "is made up of about 8.5 million square feet of space, of which is about 5.4 million square feet is Classrooms and classrooms.

The Private Learning Campus "consists of about 2.5 million square feet of space, of which is about 1.8 million square feet is Classrooms and classrooms.

4.2 Question 2:

For Davinci, the error is mostly in in named entity or some very specific details about CMU such as the campus acre size, or the alumni and the Nobel Laureates at the University. As for

text-ada, the errors are more pronounced. They get the basic facts wrong about the university to a point where most of the text is hallucination.

4.3 Question 3:

Model = Davinci, Wikipedia: CMU, Prompt: "his is a part of the wikipedia article on Carnegie Mellon University. Continue the article based on facts and do not add false information. Article: "

In 1967, the Carnegie Institute of Technology merged with the Mellon Institute of Industrial Research to form Carnegie Mellon University. The university's 140-acre (0.57 km²) main campus is 3 miles (4.8 km) from Downtown Pittsburgh and abuts the Carnegie Museums of Pittsburgh, which includes the Andy Warhol Museum.

Carnegie Mellon has seven colleges and independent schools: the College of Engineering, College of Fine Arts, Dietrich College of Humanities and Social Sciences, Mellon College of Science, Tepper School of Business, H. John Heinz III College of Information Systems and Public Policy, and the School of Computer Science. The university also has campuses in Qatar and Silicon Valley, with degree-granting programs in six continents.

Carnegie Mellon consistently ranks in the top 25 in the national U.S. News & World Report rankings. It is home to the world's first degree-granting Robotics and Drama programs, as well as one of the first Computer Science departments. The university spent \$333 million on research in 2015.

Carnegie Mellon counts 13,650 students from 114 countries, over 100,000 living alumni, and a global network of over 300,000. The university has an endowment of \$1.85 billion as of June 2019, the seventh largest in the United States.

Carnegie Mellon is a highly selective university. For the class of 2023, the university received nearly 75,000 applications and accepted about

Model = ADA, Wikipedia: CMU, Prompt: "his is a part of the wikipedia article on Carnegie Mellon University. Continue the article based on facts and do not add false information. Article: "

CMU has been a part of the University of Pittsburgh since 1896.

-CMU is a private research university -The institution was originally established in 1900 by Andrew Carnegie -In 1912, it became the Carnegie Institute of Technology and began granting four-year degrees

-CMU has been a part of the University of Pittsburgh since 1896

The prompt reduces the hallucinations in the ADA model substantially but the generation is very short. IT seems to have no effect on the larger Davinci model.

4.4 Question 4

prompt = "Tom Cruise became the CEO of Apple in 2002.", Model = Davinci

This is not true.

prompt = "Tom Cruise became the CEO of Apple in 2002.", Model = ADA

In that role, he led the company's initial public offering (IPO) and became a driving force behind the company's success in the consuming Aalogue technology.

After Bill Gates' death in 2014, Cruise founded aux Infosys, which is""

At Apple, Cruise has been a driving force behind the company's success in the Consumer Technology Company (STFC) and creation of aux infosys Limited. He is also president and CEO of aughnd get the company's Gottera AG seal of force. CodeText

Davinci, simply states that it is not true and does not continue generation. ADA generates a wikipedia sounding article but with factual errors.

4.5 Question 5

Building an LLM which performs well on both factual data and fantasy stories has some challenges.

Data Bias: The LLMs are trained on large-scale internet data which has a lot of inaccuracies.

Stale Information: The models have a training cut-off date thus, they do not have any information after their training data. This can lead to stale or non-current information.

Long-term reasoning: Fantasy stories require long-term reasoning over multiple books in some cases. since LLMs have a fixed context length, this might be a challenge to produce a coherent and consistent storyline.

For Factual LLM, we could try to curate the datasets with very high-quality information. Additionally, we can also build external systems which feed the LLMs external data to answer the given prompt and use the in-context learning of LLMs to answer these questions grounded in facts.

For Fantasies, incorporating structured commonsense knowledge graphs into training data could improve the model's ability to handle both cases. Training the model with a dataset that includes information on a wide variety of real and fictional entities can improve its ability to respond to such prompts. We could also include fictional content from literature, mythology, science fiction, and fantasy genres.

5 Comparing Pre-Trained and Fine-tuned Models

5.1 Question 1

5.1.1 Recipe Generation

Task: Recipe Generation, prompt = "Give me a recipe for making spaghetti", Model = text-davinci-003

Ingredients:

-1 pound of spaghetti -1 jar of your favorite spaghetti sauce -1/2 cup of chopped onion
-2 cloves of garlic, minced -1 tablespoon of olive oil -1 teaspoon of dried oregano
-1 teaspoon of dried basil -1/2 teaspoon of salt -1/4 teaspoon of pepper -1/4 cup of
grated Parmesan cheese

Instructions:

1. Bring a large pot of salted water to a boil.
2. Add the spaghetti and cook according to package instructions.
3. Meanwhile, heat the olive oil in a large skillet over medium heat. Add the onion and garlic and cook until softened

Task: Recipe Generation, prompt = "Give me a recipe for making spaghetti", Model = davinci

sauce.

The guy was angry. He grabbed the cook by the collar and said, "I don't want to hear that kind of talk. Now, give me the recipe or I'll throw you out the window."

The cook threw him out the window.

5.1.2 Explaining the rules for a sport

Task: Explaining the rules of a sport, prompt = "Explain me the rules of football", Model = text-davinci-003

Football is a game played between two teams of 11 players each. The objective of the game is to score more points than the opposing team. Points are scored by advancing the ball into the opponent's end zone or by kicking the ball through the goal posts.

The game is played on a rectangular field, typically 100 yards long and 53.3 yards wide. At each end of the field is a goal line, and in the center is the 50-yard line.

At the beginning of the game, the ball is placed at the 50-yard line and the two teams face each other. The offense has four chances (downs) to move the ball 10 yards. If they fail to do so, the other team takes

Task: Explaining the rules of a sport, prompt = "Explain me the rules of football",
Model = davinci

,said a young football fan. "It is a very simple game. One team of eleven men has a ball, and the other team tries to take it away from them. And that is all there is to it."

The young man nodded. Sounds pretty simple."

"It is,said the old man. But you must remember one thing. If you don't go in for the game, it's not your fault if you don't understand it."

A few days later the old man saw the young man in the street. He had a big bandage around his head. Åhat happened to you?asked the old man.

Åell, it's like this,said the young man

5.1.3 Wikipedia Article Continuation

Task: Wikipedia Article Continuation, prompt = "Continue this wikipedia article: Carnegie Mellon University (CMU) is a private research university in Pittsburgh, Pennsylvania. The institution was originally established in 1900 by Andrew Carnegie as the Carnegie Technical Schools. In 1912, it became the Carnegie Institute of Technology and began granting four-year degrees.", Model = text-davinci-003

In 1967, the Carnegie Institute of Technology merged with the Mellon Institute of Industrial Research to form Carnegie Mellon University. Since then, the university has grown to include seven colleges and schools and has expanded its course offerings to include a variety of disciplines. The university has been recognized for its commitment to research, with several of its programs consistently ranked among the best in the nation.

CMU is also renowned for its work in artificial intelligence and robotics. Its School of Computer Science has produced numerous prominent computer scientists and entrepreneurs, including some of the founders of Google, Apple, and Uber. The university is also home to the world's first robotics degree program, as well as the world's first language technologies degree program.

CMU is also

Task: Wikipedia Article Continuation, prompt = "Continue this wikipedia article: Carnegie Mellon University (CMU) is a private research university in Pittsburgh, Pennsylvania. The institution was originally established in 1900 by Andrew Carnegie as the Carnegie Technical Schools. In 1912, it became the Carnegie Institute of Technology and began granting four-year degrees.", Model = davinci

In 1967, the Carnegie Institute of Technology merged with the Mellon Institute of Industrial Research to form Carnegie Mellon University. Carnegie Mellon's coordinate women's college, the Margaret Morrison Carnegie College closed in 1973 and merged its academic programs with the rest of the university. The university's 140-acre (57 ha) main campus is 3 miles (4.8 km) from Downtown Pittsburgh. The university also has campuses in Qatar and Silicon Valley, with degree-granting programs in six continents. CMU was ranked among the top universities in the United States in the 2010s, and has been highly ranked internationally for some time. Carnegie Mellon is organized into ten colleges and schools: the Dietrich College of Humanities and Social Sciences, the College of Engineering, the CMU is also

5.2 Question2

A fine-tuned model is good at tasks that require instruction following, such as recipe generation and game explanation. But for free-form text generation, both models work equally well. If the user wants a free-form text output, then it is better to use a pre-trained model as it can save compute costs to fine-tune an instruction following the model.

6 Acknowledgment of AI Tools

I used Grammarly and ChatGPT to rephrase and find grammatical errors in my answers.