

## Contents

Introduction .....	2
Comparing 2 models as Math Q&A Generators .....	3
Finetuning GPT – 3.5 – turbo.....	12
Finetuning existing pre-trained / fine-tuned model e.g. T5 model .....	13
Challenges.....	16

# Introduction

To demonstrate some basic understanding of the VeNTUre project, I have done some research and come out with this little demo project.

This project showcases 2 text generation models using 2 different APIs. The models and corresponding APIs used are:

Model 1: GPT-3.5-turbo (OpenAI API)

Model 2: Llama 3 70B instruct (Groq API)

The section below will compare some math questions and answers generated by 2 different models given a relevant topic context.

# Comparing 2 models as Math Q&A Generators

The two models compared against each other are:

Model 1: GPT-3.5-turbo-0125 (OpenAI API)

Model 2: Llama 3 70B instruct (Groq API)

I will call them as GPT 3.5 and llama3 in the comparison table below. For every example below, there is a snapshot of a simple user interface of the generator.

I also attach a short video on how the generator runs down below. Kindly access the link.

[ib\\_math\\_qa\\_generator.mp4](#)

I briefly categorize the type of prompts into 4 examples:

1. Only providing context
2. Providing topic context and example question
3. Symbols annotated in text format
4. Providing an example question only

Prompt for the examples below include keywords “**SL math question**”, “**medium difficulty level**”, “**non-calculator type**”

## Conclusion

We could observe that sometimes one model would generate an application / scenario problem as a question, while sometimes the question would be in the form of algebra (**e.g. Example 1**)

For both models to provide accurate question and answer, it is best to provide relevant context: summary of topic + an example question at best. (**e.g. Example 2**)

Also both models are able to understand symbols annotated in text format (**e.g. Example 3** – “open parentheses a plus b close parentheses to the power of n equals a to the power of n plus blank scriptbase straight C subscript 1 space end scriptbase presubscript blank presuperscript”), but this kind of input is of course not encouraged due to its low readability.

Lastly, this model could just act as a math question answering model! We just have to paste a math question as the context. (**e.g. Example 4**)

Comparison: The performances of both models seem to be very close to each other. From the aspect of pricing,

For gpt 3.5:

Model	Input	Output
gpt-3.5-turbo-0125	\$0.0005 / 1K tokens	\$0.0015 / 1K tokens

For llama 3: It is free as long as we don't exceed the rate limit below.

ID	REQUESTS PER MINUTE	REQUESTS PER DAY	TOKENS PER MINUTE
llama3-70b-8192	30	14,400	6,000

## Example 1 – Only providing topic context

# IB Math (HL/SL) Q&A Generator

Paste the relevant topic context:

How do I find the probability of combined events?  
The probability of A or B (or both) occurring can be found using the formula  
 $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ . You subtract the probability of A and B both occurring because it has been included twice (once in P(A) and once in P(B)).  
The probability of A and B occurring can be found using the formula

Generate

	IB-Math	Generated
0	Question	Given that the probability of event A occurring is 0.6 and the probability of event B occurring is 0.4, if the probability of either event A or event B occurring is 0.8, find the probability of both event A and event B occurring.
1	Answer	Let $P(A) = 0.6$ , $P(B) = 0.4$ , $P(A \cup B) = 0.8$ . Using the formula $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ , we can find $P(A \cap B)$ . Substituting the given values, we get $0.8 = 0.6 + 0.4 - P(A \cap B)$ . Therefore, $P(A \cap B) = 0.6 + 0.4 - 0.8 = 0.2$ . Hence, the probability of both event A and event B occurring is 0.2.

	GPT 3.5	Llama 3
<b>Context 1</b>	<p>How do I find the probability of combined events? The probability of A or B (or both) occurring can be found using the formula straight <math>P(A \cup B) = P(A) + P(B) - P(A \cap B)</math>. You subtract the probability of A and B both occurring because it has been included twice (once in P(A) and once in P(B)). The probability of A and B occurring can be found using the formula. <math>P(A \cap B) = P(A) P(B A)</math></p>	
<b>Question 1</b>	Given that the probability of event A occurring is 0.6 and the probability of event B occurring is 0.4, if the probability of either event A or event B occurring is 0.8, find the probability of both event A and event B occurring.	A fair six-sided die is rolled twice. Let A be the event that the first roll is a 2, and let B be the event that the sum of the two rolls is 7. Find the probability that A or B (or both) occurs.
<b>Answer 1</b>	Let $P(A) = 0.6$ , $P(B) = 0.4$ , $P(A \cup B) = 0.8$ . Using the formula	Let P(A) be the probability that the first roll is a 2, so $P(A) = 1/6$ . Let P(B) be the

	<p> <math>P(A \cup B) = P(A) + P(B) - P(A \cap B)</math>, we can find <math>P(A \cap B)</math>. Substituting the given values, we get <math>0.8 = 0.6 + 0.4 - P(A \cap B)</math>. Therefore, <math>P(A \cap B) = 0.6 + 0.4 - 0.8 = 0.2</math>. Hence, the probability of both event A and event B occurring is 0.2. </p>	<p> probability that the sum of the two rolls is 7. There are 6 outcomes that sum to 7: (1,6), (2,5), (3,4), (4,3), (5,2), (6,1), so <math>P(B) = 6/36 = 1/6</math>. Since the events are independent, <math>P(A \cap B) = P(A)P(B A) = P(A)P(B) = (1/6)(1/6) = 1/36</math>. Using the formula, <math>P(A \cup B) = P(A) + P(B) - P(A \cap B) = (1/6) + (1/6) - (1/36) = 11/36</math>. </p>
--	--	---

Example 2 – Providing topic context and example question

(HL/SL) Q&A  
Generator

Paste the relevant topic context:

a

n

=

2

.

2

n

-

1

a

n

Extract

	IB-Math	Generated
0	Question	Given the general form of a geometric sequence $a_n=a_1 \cdot r^{n-1}$ a $n=a_1 \cdot r^{n-1}$ , find the 5th term of a geometric sequence with first term $a_1=3$ and common ratio $r=4$ .
1	Answer	To find the 5th term, substitute $n=5$ , $a_1=3$ , and $r=4$ into the formula. This gives $a_5=3 \cdot 4^4=3 \cdot 256=768$ . Therefore, the 5th term of the geometric sequence is 768.

	GPT 3.5	Llama 3
Context 2	<p>A geometric sequence is a sequence of numbers that follow a particular pattern of multiplication by a constant ratio. The sequence is formed by multiplying each term of the sequence by a constant ratio to obtain the next term. A geometric sequence can be written in the general form as:</p> $a_n = a_1 \cdot r^{n-1}$ <p>Where:</p> <ul style="list-style-type: none"><li><math>a_n</math> is the <math>n</math>th term of the sequence</li><li><math>a_1</math> is the first term of the sequence</li><li><math>r</math> is the common ratio between each term of the sequence</li></ul> <p>For example, consider the geometric sequence 2, 4, 8, 16, 32, ... with the first term <math>a_1 = 2</math> and the common ratio <math>r = 2</math>. Using the formula, we can find the <math>n</math>th term of the sequence:</p> $a_n = 2 \cdot 2^{n-1}$ <p>Thus, the 6th term of the sequence is <math>a_6 = 2 \cdot 2^{6-1} = 64</math>.</p>	
Question 2	Given the general form of a geometric sequence $a_n=a_1 \cdot r^{n-1}$ a $n=a_1 \cdot r^{n-1}$ , find the 5th term of a geometric sequence with first term $a_1=3$ and common ratio $r=4$ .	In a geometric sequence, the first term is 3 and the common ratio is 2. Find the 5th term of the sequence.
Answer 2	To find the 5th term, substitute $n=5$ , $a_1=3$ , and $r=4$ into the formula. This gives	Using the formula $a_n = a_1 \cdot r^{n-1}$ , we can find the 5th term of the sequence: $a_5 =$

	$a_5 = 3 \cdot 4^4 = 3 \cdot 256 = 768$ . Therefore, the 5th term of the geometric sequence is 768.	$3 \cdot 2^{5-1} = 3 \cdot 2^4 = 3 \cdot 16 = 48$ . Hence, the 5th term of the sequence is 48.
--	--	--

**Example 3 – symbols annotated in the form of text**

**IB Math  
(HL/SL) Q&A  
Generator**

Paste the relevant topic context:

What is the Binomial Theorem?

The binomial theorem (sometimes known as the binomial expansion) gives a method for expanding a two-term expression in a bracket raised to a power

A binomial expression is in fact any two terms inside the bracket, however in IB the expression will usually be linear

To expand a bracket with a two-term expression in:

First choose the most appropriate

Generate

	IB-Math	Generated
0	Question	Expand the expression $(2x - 3)^5$ using the Binomial Theorem.
1	Answer	Expanding $(2x - 3)^5$ using the Binomial Theorem: $(2x)^5 + 5 \cdot (2x)^4 \cdot (-3) + 10 \cdot (2x)^3 \cdot (-3)^2 + 10 \cdot (2x)^2 \cdot (-3)^3 + 5 \cdot (2x) \cdot (-3)^4 + (-3)^5 = 32x^5 - 240x^4 + 720x^3 - 1080x^2 + 810x - 243$

	GPT 3.5	Llama 3
Context 3	What is the Binomial Theorem? The binomial theorem (sometimes known as the binomial expansion) gives a method for expanding a two-term expression in a bracket raised to a power A binomial expression is in fact any two terms inside the bracket, however in IB the expression will usually be linear To expand a bracket with a two-term expression in:	



	<p>First choose the most appropriate parts of the expression to assign to a and b</p> <p>Then use the formula for the binomial theorem:</p> <p>open parentheses a plus b close parentheses to the power of n equals a to the power of n plus blank scriptbase straight C subscript 1 space end scriptbase presubscript blank presuperscript space n end presuperscript a to the power of n minus 1 space end exponent b space plus space horizontal ellipsis space plus space scriptbase straight C subscript r space end scriptbase presubscript blank presuperscript n a blank to the power of n minus r end exponent space b blank to the power of r space end exponent plus space horizontal ellipsis space plus space b to the power of n</p> <p>where scriptbase straight C subscript r end scriptbase presubscript blank presuperscript n equals fraction numerator n factorial over denominator r factorial open parentheses n minus r close parentheses factorial end fraction</p> <p>See below for more information on scriptbase straight C subscript r end scriptbase presubscript blank presuperscript n</p> <p>You may also see scriptbase straight C subscript r end scriptbase presubscript blank presuperscript n written as open parentheses n r close parentheses</p> <p>or straight C presubscript space n end presubscript subscript r</p> <p>You will usually be asked to find the first three or four terms of an expansion</p> <p>Look out for whether you should give your answer in ascending or descending powers of x</p> <p>For ascending powers start with the constant term, an</p> <p>For descending powers start with the term with x in</p> <p>You may wish to swap a and b over so that you can follow the general formula given in the formula book</p>	
<b>Question 3</b>	Expand the expression $(2x - 3)^5$ using the Binomial Theorem.	Expand $(2x + 3)^4$ up to and including the term in $x^2$ , and hence find the value of k if the coefficient of $x^2$ is 72.
<b>Answer 3</b>	Expanding $(2x - 3)^5$ using the Binomial Theorem: $(2x)^5 + 5(2x)^4 * (-3) + 10(2x)^3 * (-3)^2 + 10(2x)^2 * (-3)^3 + 5(2x) * (-3)^4 + (-3)^5 = 32x^5 - 240x^4 + 720x^3 - 1080x^2 + 810x - 243$	Using the binomial theorem, we have $(2x + 3)^4 = 2^4x^4 + 4C1(2x)^3(3) + 4C2(2x)^2(3)^2 + 4C3(2x)(3)^3 + 3^4$ . Expanding, we get $16x^4 + 32x^3(3) + 24x^2(9) + 32x(27) + 81$ . Hence, the coefficient of $x^2$ is $24(9) = 216$ . Since the coefficient of $x^2$ is 72, we have $216 = 72$ , which gives $k = 3$ .

## Example 4 – Straightaway providing an example question

# IB Math (HL/SL) Q&A Generator

Paste the relevant topic context:

For the function  $f(x)=x^3+1, 2 \leq x \leq 10$ :  
a) write down the value of  $f(7)$ .  
b) find the range of  $f(x)$

Generate

	IB-Math	Generated
0	Question	For the function $f(x)=x^3+1, 2 \leq x \leq 10$ : a) write down the value of $f(7)$ . b) find the range of $f(x)$
1	Answer	a) $f(7) = 7^3 + 1 = 343 + 1 = 344$ . b) To find the range of $f(x)$ , we need to look at the possible values that $f(x)$ can take. Since $x^3$ is always positive or zero for real $x$ , and adding 1 doesn't change that, the minimum value of $f(x)$ occurs at $x=2$ where $f(2) = 2^3 + 1 = 8 + 1 = 9$ . The function $f(x)$ increases with increasing $x$ without bound. Therefore, the range of $f(x)$ is $[9, \infty)$ .

	GPT 3.5	Llama 3
<b>Context 4</b>	For the function $f(x) = x^3+1, 2 \leq x \leq 10$ : a) write down the value of $f(7)$ . b) find the range of $f(x)$	
<b>Question 4</b>	For the function $f(x) = x^3 + 1, 2 \leq x \leq 10$ : a) Write down the value of $f(7)$ . b) Find the range of $f(x)$ .	For the function $f(x) = x^3 + 1, 2 \leq x \leq 10$ , find the value of $f(7)$ and the range of $f(x)$ .
<b>Answer 4</b>	a) To find $f(7)$ , simply substitute $x=7$ into the function: $f(7) = 7^3 + 1 = 343 + 1 = 344$ . Therefore, the value of $f(7)$ is 344. b) The range of $f(x)$ can be found by analyzing the behavior of the function. Since the function is a cubic function with a positive leading coefficient, it	a) To find the value of $f(7)$ , substitute $x = 7$ into $f(x) = x^3 + 1$ . So, $f(7) = 7^3 + 1 = 343 + 1 = 344$ . b) To find the range of $f(x)$ , find the minimum and maximum values of $f(x)$ in the given domain. Since $x^3$ is an increasing function, the minimum value of $f(x)$ occurs at $x = 2$ , and the maximum value occurs at $x = 10$ . So, the

	continuously increases as x increases. The minimum value of the function occurs at $x=2$ and the maximum occurs at $x=10$ . Hence, the range of $f(x)$ is $f(2)$ to $f(10)$ , which is 9 to 1001.	range of $f(x)$ is $[f(2), f(10)] = [2^3 + 1, 10^3 + 1] = [9, 1001]$ .
--	---	--

# Finetuning GPT – 3.5 – turbo

Purpose: to suit more domain-specific task – suit IB math syllabus HL and SL standard, and also customize style / tone of questions and answers generated.

Advantages: higher quality, save tokens and costs, lower latency request

## Fine-tuning models

Create your own custom models by fine-tuning our base models with your training data.

Once you fine-tune a model, you'll be billed only for the tokens you use in requests to that model.

[Learn about fine-tuning](#)

Model	Training	Input Usage	Output usage
gpt-3.5-turbo	\$8.00 / 1M tokens	\$3.00 / 1M tokens	\$6.00 / 1M tokens
davinci-002	\$6.00 / 1M tokens	\$12.00 / 1M tokens	\$12.00 / 1M tokens
babbage-002	\$0.40 / 1M tokens	\$1.60 / 1M tokens	\$1.60 / 1M tokens

Prepare dataset in the format below:

1. System prompt: (i) HL/SL (ii) Calculator/Non-calculator (iii) Topic
2. User prompt: the topic given
3. Assistant prompt: the response returned

Json file → Jsonl file → load into OpenAI

<https://platform.openai.com/docs/guides/fine-tuning/create-a-fine-tuned-model>

# Finetuning existing pre-trained / fine-tuned model e.g. T5 model

Incorporating mathematical knowledge into model:

1. Tokenization

For e.g. GPT2 tokenizer:

Given an expression  $3x^2 + 4x + 5 = 0$ , individual components: numbers, variables, operators and exponentiation are identified and tokenized into ['3', 'x', '^', '2', '+', '4', 'x', '+', '5', '=', '0'].

2. Embedding representation

If a language model doesn't handle mathematical problems well, it means the embeddings for mathematical symbols and notation haven't been learned properly. Potential issues: insufficient training data, improper tokenization, lack of contextual training, and failure to learn mathematical patterns.

3. Libraries to deal with symbolic processing

SymPy, LaTeX and MathJax

Can translate mathematical expressions into formats that LLMs can understand and process.

Examples of fine-tuned models on math dataset

1. [MU-NLPC/calcfomer-t5-large](#)

T5 model fine-tuned on CalcX – math problems dataset

In the training dataset, in “answers” section, the steps that are needed to calculate for giving a right answer is converted into Chain-of-Thought, which is in “calculator terms”. The model interacts with an external system: a SYMPY calculator to calculate the mathematical operations in the training examples.

A training example of CalcX dataset:

id string · lengths	question string · lengths	chain string · lengths	result string · lengths	source_ds string · classes
17~18 61.2%	6~128 23.4%	18~428 88.5%	1~5 88.4%	ape210k 61.2%
ape210k_00666205	A semicircular aquarium has a radius of 5 meters, what is the perimeter of the aquarium in meters?	<gadget id="calculator">3.14 * 5</gadget> <output>15.7</output> <gadget id="calculator">5 * 2</gadget> <output>10</output> <gadget id="calculator">15.7 + 10</gadget> <output>25.7</output> <result>25.7</result>	25.7	ape210k

T5 model is then subsequently fine-tuned on this dataset: using tags to wrap text. An example of tag:

tags	
<result>18</result>	Result tag: to display calculated answer
<gadget id="calculator">27/3</gadget>	Gadget tag: acts as input / queries to an external system. External system here refers to a calculator
<output>9</output>	Output tag: acts as response of the “calculator” to the input / query

An inference example of the model

```

default_max_tokens=512)

query = """
    The profit from a business transaction is shared among 2 business partners,
    Mike and Johnson in the ratio 2:5 respectively.
    If Johnson got $2500, how much will Mike have
    after spending some of his share on a shirt that costs $200?
    """

inputs = tokenizer(query, return_tensors="pt")
output_ids = model.generate(**inputs)
tokenizer.decode(output_ids[0], spaces_between_special_tokens=False)

```

This returns:

```

According to the ratio, for every 5 parts that Johnson gets, Mike gets 2 parts
each part is therefore $2500/5 = $<gadget id="calculator">2500/5</gadget><output
Mike will get 2*$500 = $<gadget id="calculator">2*500</gadget><output>1_000</out
After buying the shirt he will have $1000-$200 = $<gadget id="calculator">1000-2
Final result is<result>800</result></s>

```

Flaw:

It is still unable to perform more complex operations that is outside the scope of (+-\*/). Do not match the difficulty level of IB curriculum.

Might need further research on the ability of (i) tokenizing complex symbols. (ii) capturing of embedding representation of complex symbols by the model (ii) external systems: “calculator” to work on more complex operations

# Challenges

1. Pipeline: there are 3 elements in dataset – context, question and answer. What should be the input and what should be the output
  - ➔ Recommended to go through 2 pipelines: one for question generation, and for question answering
2. Question generation in the form of equations: so far, seen questions are all in the form of English texts: e.g.
  - a. Question: What structure is classified as a definite lie algebra?
  - b. Answer: A definite Lie algebra is a Lie algebra equipped with an inner product that is positive definite. Such an algebra is called a...But according to IB math syllabus, it should be in the form of equations.
  - ➔ Create new tokenizer based on the fed dataset
  - ➔ Utilize other models that are pretrained on math equations
  - ➔ Combining neural systems and symbolic systems (calculator)
3. Tagging and categorization of text: should questions in different categories be processed separately: meaning a different tag OR should it be stated explicitly in the prompt???
4. The answer is not found in context, can't be processed in this way:

2. highlight format

Here the answer span is highlighted within the text with special highlight tokens.

```
<hl> 42 </hl> is the answer to life, the universe and everything.
```

This idea is proposed in the "A Recurrent BERT-based Model for Question Generation" [paper](#). See section 4.3

[https://github.com/patil-suraj/question\\_generation?tab=readme-ov-file#question-generation-using-transformers](https://github.com/patil-suraj/question_generation?tab=readme-ov-file#question-generation-using-transformers)

Can we generate answers which can't be found in the context?

Resources:

1. A model that can run both question generation and question answering (finetuned on French dataset)

<https://huggingface.co/JDBN/t5-base-fr-qg-fquad>

the pre-processing of the dataset is pre-processed as such:

[https://github.com/patil-suraj/question\\_generation](https://github.com/patil-suraj/question_generation)



multi-task QG and QA: [https://github.com/patil-suraj/question\\_generation?tab=readme-ov-file#question-generation-using-transformers](https://github.com/patil-suraj/question_generation?tab=readme-ov-file#question-generation-using-transformers)

#### Additional Resources

1. Abstractive Question Answering – retriever, building dataset, uploading data and querying and getting CoT for GPT-3.5, get result from GPT-3.5

<https://myscale.com/docs/en/sample-applications/abstractive-qa/>