Contents

GPT – 3.5 – turbo: IB Math Question and Answer Generator

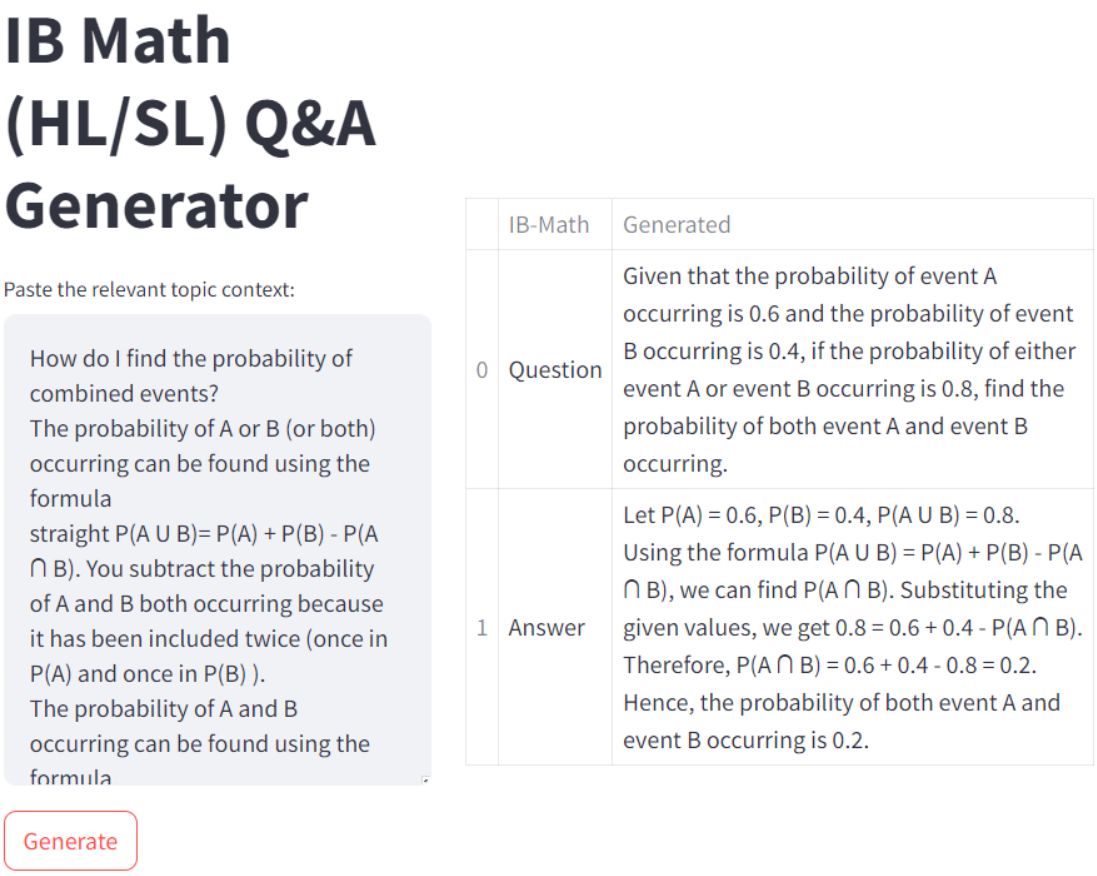
I have written some code using Open AI API to generate IB math question and answer.

For the model to provide accurate question and answer, it is best to provide relevant context: summary of topic + an example question at best.

Attached below is a simple interface showcasing the context provided to the model, and math questions and answers generated as a result.

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

Example 1 – Optimal case



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

straight P(A U B)= P(A) + P(B) - P(A ∩ 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.
P(A ∩ B) = P(A) P(B|A)

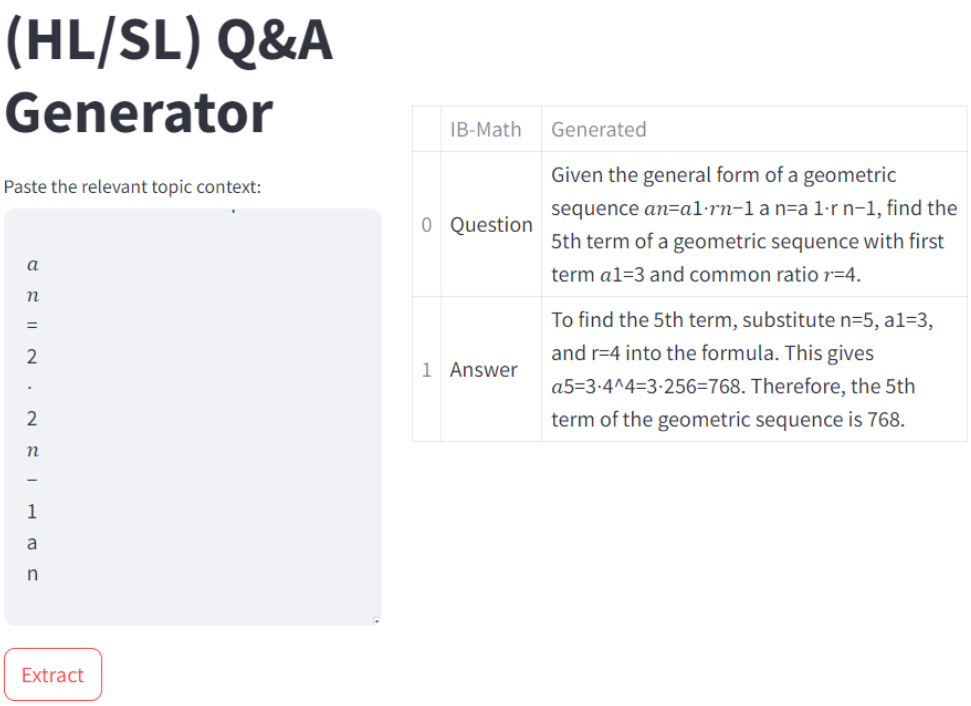
Question

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.

Answer

Let P(A) = 0.6, P(B) = 0.4, P(A U B) = 0.8. Using the formula P(A U B) = P(A) + P(B) - P(A ∩ B), we can find P(A ∩ B). Substituting the given values, we get 0.8 = 0.6 + 0.4 - P(A ∩ B). Therefore, P(A ∩ B) = 0.6 + 0.4 - 0.8 = 0.2. Hence, the probability of both event A and event B occurring is 0.2.

Example 2 – Providing topic background and example question



Context

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:

$an=a1\cdot rn-1an=a1\cdot rn-1$

Where:

- $anan$ is the nth term of the sequence
- $a1a1$ is the first term of the sequence
- rr is the common ratio between each term of the sequence

For example, consider the geometric sequence 2, 4, 8, 16, 32, ... with the first term $a1=2a1=2$ and the common ratio $r=2r=2$. Using the formula, we can find the nth term of the sequence:

$an=2\cdot 2n-1an=2\cdot 2n-1$

Thus, the 6th term of the sequence is $a6=2\cdot 26-1=64a6=2\cdot 26-1=64$.

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:

$$a_n = a_1 \cdot r^{n-1}$$

Where:

- a_n is the nth term of the sequence
- a_1 is the first term of the sequence
- r is the common ratio between each term of the sequence

For example, consider the geometric sequence 2, 4, 8, 16, 32, ... with the first term $a_1 = 2$ and the common ratio $r = 2$. Using the formula, we can find the nth term of the sequence:

$$a_n = 2 \cdot 2^{n-1}$$

Thus, the 6th term of the sequence is $a_6 = 2 \cdot 2^{6-1} = 64$.

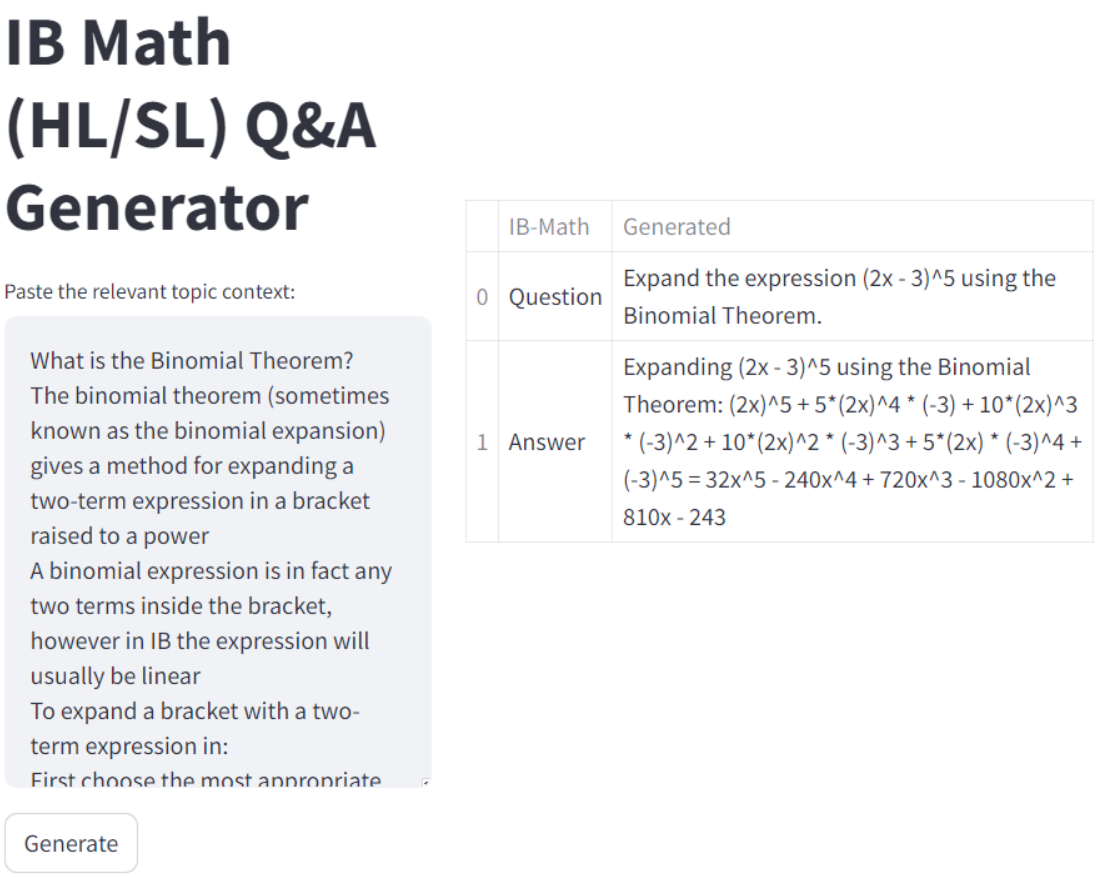
Question

Given the general form of a geometric sequence $an=a1\cdot rn-1$ a n=a 1⋅r n−1, find the 5th term of a geometric sequence with first term $a1=3$ and common ratio $r=4$.

Answer

To find the 5th term, substitute n=5, a1=3, and r=4 into the formula. This gives $a5=3\cdot4\text{^}4=3\cdot256=768$. Therefore, the 5th term of the geometric sequence is 768.

Example 3 – symbols annotated in the form of text



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 parts of the expression to assign to a and b

Then use the formula for the binomial theorem:

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

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

See below for more information on scriptbase straight C subscript r end scriptbase presubscript blank presuperscript n

You may also see scriptbase straight C subscript r end scriptbase presubscript blank presuperscript n written as open parentheses n

r close parentheses

 or straight C presubscript space n end presubscript subscript r

You will usually be asked to find the first three or four terms of an expansion

Look out for whether you should give your answer in ascending or descending powers of x

For ascending powers start with the constant term, an

For descending powers start with the term with x in

You may wish to swap a and b over so that you can follow the general formula given in the formula book

Question

Expand the expression (2x - 3)^5 using the Binomial Theorem.

Answer

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

Example 4 – Straightaway providing an example question

IB Math (HL/SL) Q&A Generator

Paste the relevant topic context:

For the function f(x)=x3+1,2≤x≤10:
a) write down the value of f(7).
b) find the range of f(x)

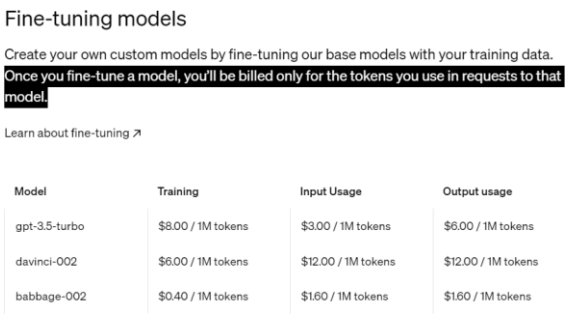
Generate

	IB-Math	Generated
0	Question	For the function f(x)=x^3+1, 2≤x≤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, ∞).

Finetuning OGPT – 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/calcformer-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 S
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:

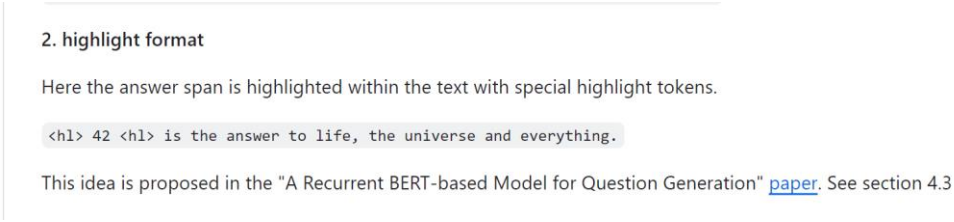
It is still unable to perform more complex operations that is outside the scope of (+-*/). Do not match the difficulty level of IB curriculum.

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

   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)

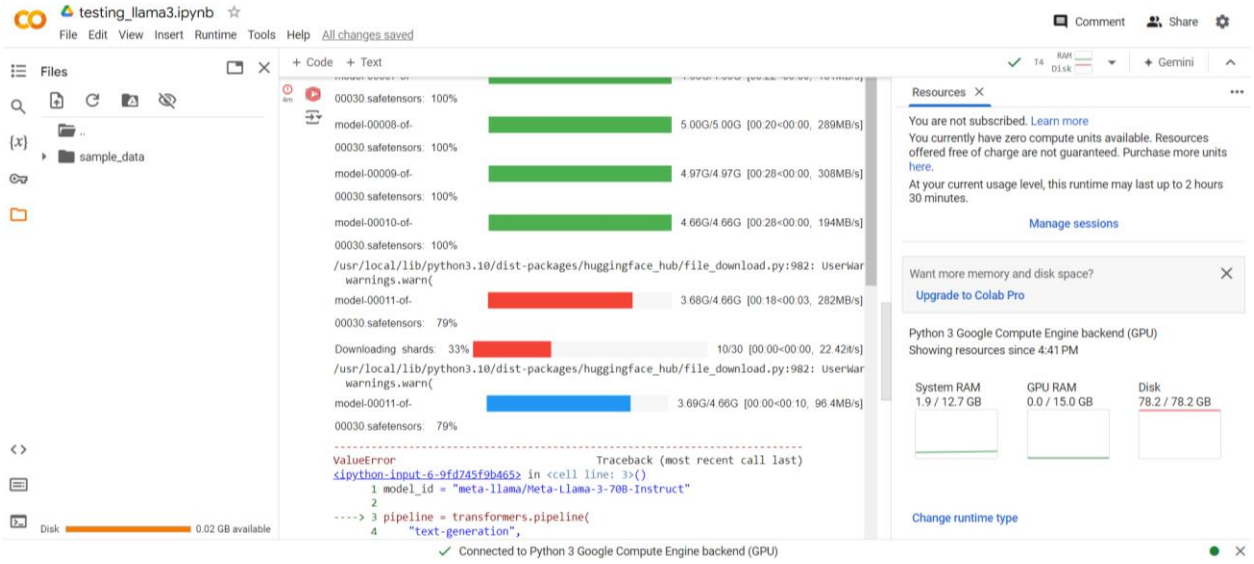
   https://huggingface.co/JDBN/t5-base-fr-qg-fquad

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

   https://github.com/patil-suraj/question_generation

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

Running llama 3 on Google Colab



Ran out of disk space

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/