

ZALO AI 2023

Zalo  Challenge / 2023

[Home](#)

[Challenges](#)

[Advisors](#)

[Timeline](#)

[Portal](#)

[News](#)

[FAQ](#)

[Regulations](#)



Advertisement Banner
Generation



GENAI

TA Hùng An

Problem - NLP

Challenge focuses on developing a language model/system capable of answering elementary-level math questions in alignment with the Vietnamese Education Program

Input: A multiple choices math question which have 4 possible choices which only one is the correct answer.

Output: The correct choice for the given math question.

Problem

	question	explanation	A	B	C	D	answer
0	Một người bán hàng bỏ ra 80,000 đồng tiền vốn ...	Theo đề bài, số tiền lỗ bằng 6% của 80 000 đôn...	A. 80,000 : 6	B. 80,000 x 6	C. 80,000 : (6 x 100)	D. (80,000 x 6) : 100	D
1	8 dm ² 24 cm ² = dm ² . Số thích hợp điền v...	Ta có 24 cm ² = 0,24 dm ² \n Vậy 8 dm ² 24 cm ² = ...	A. 824	B. 82,4	C. 8,24	D. 0,824	C
2	10% của 11,5m ² là:	10% của 11,5m ² là: 11,5 \times 10 : 100 = ...	A. 10,15dm ²	B. 1,5m ²	C. 15,5m ²	D. 1,15m ²	D
3	Một tam giác có đáy bằng 15 cm và gấp 2,5 l...	Chiều cao của tam giác đó là: 15 : 2,5 = 6 (c...	A. 90 cm ²	B. 37,5 cm ²	C. 45 cm ²	D. 18,75 cm ²	C
4	Một tấm thiệp hình chữ nhật có chiều rộ...	Chiều dài hình chữ nhật là: 9 \times 2 ...	A. 162 cm ²	B. 81 cm ²	C. 99 cm ²	D. 63 cm ²	A
...
1195	Một ô tô đi được 60 km với vận tốc 60 km/giờ, ...	vi đoạn đường thứ nhất ô tô đi hết 1 giờ, đoạn...	A. 1,5 giờ	B. 2 giờ	C. 3 giờ	D. 4 giờ	C
1196	Một ô tô đi từ A đến B với vận tốc 38,5k...	Thời gian ô tô đã đi là: 100,1 : 38,5 = 2,6 (...)	A. 2 giờ 1 phút	B. 2 giờ 6 phút	C. 2 giờ 36 phút	D. 2 giờ 30 phút	C
1197	Một ô tô đi từ Hà Nội lúc 11 giờ trưa và đến V...	5 giờ 30 phút chiều = 17 giờ 30 phút\n Thời gi...	A. 4 giờ 30 phút	B. 6 giờ 30 phút	C. 6 giờ 15 phút	D. 6 giờ	D
1198	Một sân vận động hình chữ nhật có chiều ...	Chiều dài sân vận động là: 95 \times 3 ...	A.380 m	B.285 m	C.670 m	D.760 m	D
1199	Chữ số 7 trong số thập phân 53,675 có giá trị là:	None	A: 7/10	B: 7/100	C: 70	D: 700	B

1200 rows \times 7 columns

Problem

Input	Output
<p>Số "bảy triệu hai trăm nghìn" có:</p> <p>A. Ba chữ số 0</p> <p>B. Bốn chữ số 0</p> <p>C. Năm chữ số 0</p> <p>D. Sáu chữ số 0</p>	<p>C. Năm chữ số 0</p>
<p>Mẫu số của một phân số thập phân có thể là những số nào?</p> <p>A. Các số chẵn</p> <p>B. Các số 10; 100; 1000; ...</p> <p>C. Các số lẻ</p> <p>D. Mọi số tự nhiên khác 0</p>	<p>B. Các số 10; 100; 1000; ...</p>

Problem

Allowances:

- You are encouraged to utilize pre-trained Large Language Models (LLMs) that are freely available online, such as Bloom, Llama, and others. However, it is essential to ensure that the pre-trained models you use are both free and publicly accessible on the internet. If you decide to train your own pre-trained models, you must provide us with the original textual data and the complete script outlining the entire process of creating these models.
- We have set a limit on the size of pre-trained models to 14 billion parameters, which allows you to make use of some open-source pre-trained models.

Restrictions:

- Using APIs that call any external services (e.g. Open AI, Google, ...) during training and inference.
- Training & inference with Internet access.

Problem

Additional finetuning data:

- You have the option to leverage the capabilities of other large language models to enhance your models or collect data from the internet for fine-tuning. However, it is mandatory to declare and submit all data sources and the data cleaning process to ensure that the results can be replicated.

Constraints:

- The inference time for each test case must not exceed 30 seconds on a single A30 or RTX 3090 GPU.

Data Augmentation

Text Augmentation

const stop_words

func get_only_chars

func synonym_replacement

func get_synonyms

func random_deletion

func random_swap

func swap_word

func random_insertion

func add_word

func eda

▼ **class** RandomErasing

func __init__

func __call__

▼ **class** RandomGrayscaleErasing

func __init__

func __call__

Data Augmentation

Augmentation with LLM

