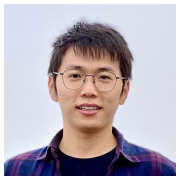
Dynamic Prompt Learning via Policy Gradient for Semi-structured Mathematical Reasoning

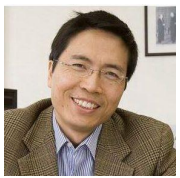
https://promptpg.github.io


Pan Lu


Liang Qiu


Kai-Wei Chang


Ying Nian Wu


Song-Chun Zhu


Tanmay Rajpurohit


Peter Clark


Ashwin Kalyan

Motivations

- **Math word problems** (MWPs) a well-defined task to diagnose the ability of intelligent systems to perform numerical reasoning
- However, most existing datasets focus on the **textual only** setting
- **Tables**, widely distributed in documents, contain rich structured information

Problem: Dan have 5 pens and 3 pencils, Jessica have 4 more pens and 2 less pencils than him. How many pens and pencils do Jessica have in total?
Equation: x = 5 + 4 +3 -2
Solution: 10

Deep Neural Solver for Math Word Problems, EMNLP 2017

Bill To	Ship To	Invoice #	US-001
John Smith 2 Court Square New York, NY 12210	John Smith 3787 Pineview Drive Cambridge, MA 12210	Invoice Date P.O.# Due Date	11/02/2019 2312/2019 26/02/2019

Qty	Description	Unit Price	Amount
1	Front and rear brake cables	100.00	100.00
2	New set of pedal arms	15.00	30.00
3	Labor 3hrs	5.00	15.00
		Subtotal	145.00
		Sales Tax 6.25%	9.06
		Invoice Total	$154.06

Flights

Flight No.	Airline	Departure Date	Model	Ticket Price	Available
2345	2	03/27/15	747	$357	No
6785	4	04/26/15	737	$489	No
8888	5	05/16/16	737	$525	Yes
6754	3	04/27/17	747	$399	Yes

Pet Health Record Table

Name of Pet	Dog	Contact Info	555 555-5555
Name of Owner	Adam Smith	Age & Gender of Pet	3 months, male
Breed	African	Weight of Pet	3.5 Kg
Breeder Name	Adam Smith	Color	Brown
Insurance Reference	Saturn	Policy No.	9445#67

Veterinary Visiting History

Date	Veterinarian	Disease	Test conducted
15/05/2015	Mr. Frandes	Chicken pox	YES
20/05/2015	Mr. Frandes	Chicken pox	No

Your Account Summary

Previous Charges:		
Amount of Your Last Bill (dated 6/13/2018)	$	137.78
Payment received 7/3/2018 – Thank you!		−137.78
Past Due Amount	$	0.00
Current Charges:		
Electric Charges	$	103.95
Natural Gas Charges		61.69
Total Current Charges	$	165.64
Total includes current and past due charges	**Total $**	165.64

Late Payments | A late payment fee of 1% per month will apply to past due charges, if any, and amounts unpaid more than 10 business days after the statement due date. Amounts will be considered delinquent if payment is not received on or before the due date.

We propose **TabMWP**, the first **M**ath **W**ord **P**roblem dataset with **Tab**ular contexts.

square beads	$2.97 per kilogram
oval beads	$3.41 per kilogram
flower-shaped beads	$2.18 per kilogram
star-shaped beads	$1.95 per kilogram
heart-shaped beads	$1.52 per kilogram
spherical beads	$3.42 per kilogram
rectangular beads	$1.97 per kilogram

Question: If Tracy buys 5 kilograms of spherical beads, 4 kilograms of star-shaped beads, and 3 kilograms of flower-shaped beads, how much will she spend? (unit: $)
Answer: **31.44**
Solution:
Find the cost of the spherical beads. Multiply: $3.42 × 5 = $17.10.
Find the cost of the star-shaped beads. Multiply: $1.95 × 4 = $7.80.
Find the cost of the flower-shaped beads. Multiply: $2.18 × 3 = $6.54.
Now find the total cost by adding: $17.10 + $7.80 + $6.54 = $**31.44**.
She will spend $**31.44**.

Sandwich sales		
Shop	**Tuna**	**Egg salad**
City Cafe	6	5
Sandwich City	3	12
Express Sandwiches	7	17
Sam's Sandwich Shop	1	6
Kelly's Subs	3	4

Question: As part of a project for health class, Cara surveyed local delis about the kinds of sandwiches sold. Which shop sold fewer sandwiches, Sandwich City or Express Sandwiches?
Options: (A) Sandwich City (B) Express Sandwiches
Answer: **(A) Sandwich City**
Solution:
Add the numbers in the Sandwich City row. Then, add the numbers in the Express Sandwiches row.
Sandwich City: 3 + 12 = 15. Express Sandwiches: 7 + 17 = 24.
15 is less than 24. **Sandwich City** sold fewer sandwiches.

2 Tasks **38,431** Problems **35,442** Solutions **37,644** Tables **12.9/54** Avg/Max cells

Formats of TabMWP

Question types	Answer types (%)	Descriptions
Free-text	Integer (59.50%)	The answer is an integer number, e.g., “40”, “1,207”, “-3”.
	Decimal (15.23%)	The answer is a decimal or a fraction number, e.g., “192.80”, “68/217”.
Multi-choice	Extractive (13.01%)	The answer could be extracted from the table context.
	Boolean (10.97%)	The answer is Boolean, e.g., “yes”/“no”, “true”/“false”, “linear”/“nonlear”.
	Other (1.29%)	The answer belongs to other text types, e.g., a statement.

Image format

Field day schedule		
Event	Begin	End
water balloon toss	11:30 A.M.	11:50 A.M.
obstacle course	12:05 P.M.	12:25 P.M.
parachute ball toss	12:30 P.M.	1:30 P.M.
jump rope race	1:40 P.M.	2:05 P.M.
balloon stomp	2:15 P.M.	2:35 P.M.
relay race	2:50 P.M.	3:40 P.M.
hula hoop contest	3:55 P.M.	4:30 P.M.
potato sack race	4:40 P.M.	5:15 P.M.

Semi-structured format

Table title: Field day schedule
Table text:
Event | Begin | End
water balloon toss | 11:30 A.M. | 11:50 A.M.
obstacle course | 12:05 P.M. | 12:25 P.M.
parachute ball toss | 12:30 P.M. | 1:30 P.M.
jump rope race | 1:40 P.M. | 2:05 P.M.
balloon stomp | 2:15 P.M. | 2:35 P.M.
relay race | 2:50 P.M. | 3:40 P.M.
hula hoop contest | 3:55 P.M. | 4:30 P.M.

Structured format

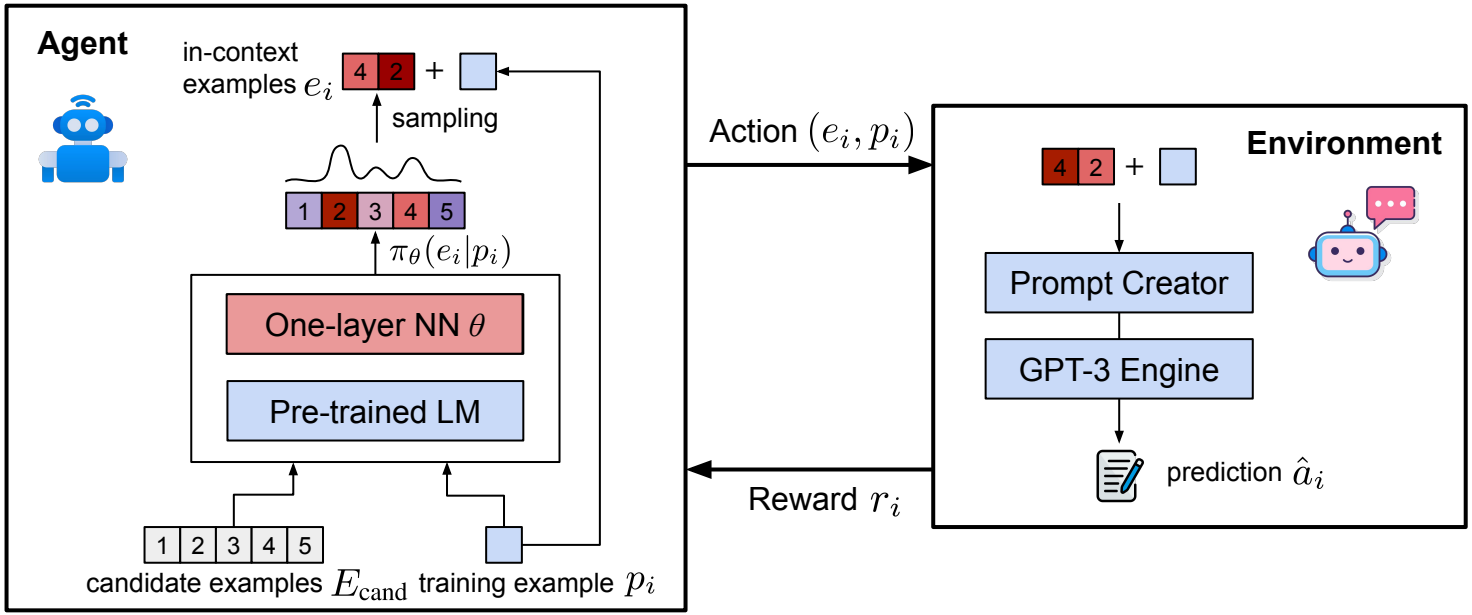
Table title: Field day schedule

	Event	Begin	End
0	water balloon toss	11:30 A.M.	11:50 A.M.
1	obstacle course	12:05 P.M.	12:25 P.M.
2	parachute ball toss	12:30 P.M.	1:30 P.M.
3	jump rope race	1:40 P.M.	2:05 P.M.
4	balloon stomp	2:15 P.M.	2:35 P.M.
5	relay race	2:50 P.M.	3:40 P.M.
6	hula hoop contest	3:55 P.M.	4:30 P.M.
7	potato sack race	4:40 P.M.	5:15 P.M.

Comparison to existing datasets

Dataset	Size	#Table	Need Math?	Need Table?	Table Type		Question Type		Answer Type			Solution Type
					Domain	Format	Free-text	MC	Text	Integer	Decimal	
Dolphin18K (2016)	831	✗	✓	✗	✗	✗	✓	✗	✗	✓	✓	formula
DRAW-1K (2017)	1,000	✗	✓	✗	✗	✗	✓	✗	✗	✓	✓	formula
Math23K (2017)	23,162	✗	✓	✗	✗	✗	✓	✗	✗	✓	✓	formula
MathQA (2019)	37,297	✗	✓	✗	✗	✗	✗	✓	✗	✓	✓	formula
ASDiv (2020)	2,305	✗	✓	✗	✗	✗	✓	✗	✓	✓	✓	formula
SVAMP (2021)	1,000	✗	✓	✗	✗	✗	✓	✗	✗	✓	✗	formula
GSM8K (2021)	8,792	✗	✓	✗	✗	✗	✓	✗	✗	✓	✗	text
IconQA (2021b)	107,439	✗	✓	✗	✗	✗	✓	✓	✓	✓	✗	✗
FinQA (2021)	8,281	2,766	✓	76.6%	finance	text	✓	✗	✗	✓	✓	program
TAT-QA (2021)	16,552	2,747	50.0%	✓	finance	text	✓	✗	✗	✓	✓	✗
MultiHiertt (2022)	10,440	9,843	✓	89.8%	finance	text	✓	✗	✗	✓	✓	✗
TABMWP (ours)	**38,431**	**37,644**	✓	✓	**open**	**text***	✓	✓	✓	✓	✓	**text**

We propose **PromptPG**, the first work that learns to select in-context examples for few-shot GPT-3 as the **Prompt** via reinforcement learning (**P**olicy **G**radient).



PromptPG: the Algorithm

Algorithm 1 Dynamic Prompt Learning via Policy Gradient (PROMPTPG)

Input: Initial policy π_{θ_0}, training example set P_{train}, candidate example set E_{cand}, # of training epochs N
Output: Learned policy π_θ

function REINFORCE(π_{θ_0}, P_{train}, E_{cand}, N)
    Initialize policy network π with parameter θ_0
    for epoch = $1, 2, ..., N$ **do**
        for $P_{\text{batch}} \in P_{\text{train}}$ **do** ▷ get a batch from the training set
            $\mathcal{L}_{\text{batch}} \leftarrow 0$
            for $p_i \in P_{\text{batch}}$ **do**
                Sample $e_i^k \sim \pi_\theta(e_i|p_i), e_i^k \in E_{\text{cand}}, k = \{1, ..., K\}$ ▷ K is # of in-context examples
                $\hat{a}_i \leftarrow$ GPT-3$(e_i^1, ..., e_i^k, p_i)$ ▷ $\hat{a}_i$ is the GPT-3 generated answer
                $r_i \leftarrow$ EVAL$(\hat{a}_i, a_i), r_i \in \{-1, 1\}$ ▷ a_i is the ground truth answer of p_i
                $\mathcal{L}_{\text{batch}} \leftarrow \mathcal{L}_{\text{batch}} - r_i \cdot \ln \pi_\theta(e_i|p_i)$
            end for
            Optimize $\mathcal{L}_{\text{batch}}$ wrt. θ
        end for
    end for
    return π_θ
end function

PromptPG largely reduces the randomness from the random selection of in-context examples and gains an improvement of **5.31%** over random selection, without any designed heuristics.

Method	Training Data	Selection Strategy	Question Types		Answer Types					Grades		Avg.
			FREE	MC	INT	DEC	EXTR	BOOL	OTH	1-6	7-8	
Heuristic Baselines												
Heuristic guess	-	-	6.71	39.81	8.37	0.26	30.80	51.22	26.67	17.55	12.27	15.29
Human performance	-	-	84.61	93.32	84.95	83.29	97.18	88.69	96.20	94.27	81.28	90.22
pre-trained Baselines												
UnifiedQA$_{\text{SMALL}}$	-	-	1.18	43.62	1.37	0.43	38.70	49.78	37.14	15.57	7.65	12.18
UnifiedQA$_{\text{BASE}}$	-	-	4.60	43.02	5.28	1.97	37.08	50.11	38.10	17.14	11.11	14.56
UnifiedQA$_{\text{LARGE}}$	-	-	4.48	48.80	5.19	1.72	48.33	50.33	40.00	19.78	10.87	15.96
TAPEX$_{\text{BASE}}$	-	-	7.32	39.76	8.68	2.06	35.06	47.11	20.95	18.67	11.81	15.73
TAPEX$_{\text{LARGE}}$	-	-	8.80	46.59	10.62	1.72	46.91	48.11	30.48	22.65	13.18	18.59
fine-tuned Baselines												
UnifiedQA$_{\text{SMALL}}$	23,059	-	22.27	51.31	27.27	2.83	52.28	48.11	69.52	35.85	21.71	29.79
UnifiedQA$_{\text{BASE}}$	23,059	-	34.02	70.68	40.74	7.90	84.09	55.67	73.33	53.31	30.46	43.52
UnifiedQA$_{\text{LARGE}}$	23,059	-	48.67	82.18	55.97	20.26	94.63	68.89	79.05	65.92	45.92	57.35
TAPEX$_{\text{BASE}}$	23,059	-	39.59	73.09	46.85	11.33	84.19	61.33	69.52	56.70	37.02	48.27
TAPEX$_{\text{LARGE}}$	23,059	-	51.00	80.02	59.92	16.31	**95.34**	64.00	73.33	67.11	47.07	58.52
Prompting Baselines w/ GPT-3												
Zero-shot	-	-	53.57	66.67	55.55	45.84	78.22	55.44	54.29	63.37	48.41	56.96
Zero-shot-CoT	-	-	54.36	66.92	55.82	48.67	78.82	55.67	51.43	63.62	49.59	57.61
Few-shot (2-shot)	2	Random	54.69	64.11	58.36	40.40	75.95	52.41	53.02	63.10	49.16	57.13
Few-shot-CoT (2-shot)	2	Random	60.76	69.09	60.04	63.58	76.49	61.19	**67.30**	68.62	55.31	62.92
PROMPTPG *w/ GPT-3 (Ours)*												
Few-shot-CoT (2-shot)	160+20	Dynamic	**66.17**	**74.11**	**64.12**	**74.16**	76.19	**72.81**	65.71	**71.20**	**64.27**	**68.23**$_{5.31\uparrow}$

▷ ***In-context example (ID: 13974)***

Table:
heart-shaped beads | $3/kilogram
rectangular beads | $2/kilogram
spherical beads | $2/kilogram
oval beads | $2/kilogram

heart-shaped beads	$3/kilogram
rectangular beads	$2/kilogram
spherical beads	$2/kilogram
oval beads	$2/kilogram

Question: Rebecca bought 2.5 kilograms of oval beads. How much did she spend? (Unit: $)
Answer:
(Step 1) Find the cost of the oval beads. Multiply the price per kilogram by the number of kilograms.
(Step 2) $2 × 2.5 = **$5**
(Step 3) She spent **$5**. **The answer is 5**.

▷ ***Test example (ID: 17417)***

Table:
[TITLE]: Birthday party
Activity | Parents | Children
Singing | 14 | 20
Eating cake | 5 | 10
Jumping rope | 16 | 20
Swimming | 16 | 19
Playing tag | 4 | 9

Birthday party		
Activity	**Parents**	**Children**
Singing	14	20
Eating cake	5	10
Jumping rope	16	20
Swimming	16	19
Playing tag	4	9

Question: At Josie's birthday party, children and parents celebrated by participating in some activities. How many more children than parents are playing tag? (Unit: children)
Answer:
(Step 1) To find the difference between the number of children and parents playing tag, subtract the number of parents from the number of children.
(Step 2) 9 - 4 = **5**
(Step 3) There are **5 more children** than parents playing tag. **The answer is 5**.

PromptPG prefers examples with multiple reasoning steps and **similar abilities** in mathematical reasoning to the test example.

The prediction follows the similar reasoning steps and arrives at the **correct** answer.

▷ ***In-context example (ID: 18429)***

Table:
[TITLE]: Children's weights (lbs)
Stem | Leaf
1 | 7
2 | 4
3 |
4 |
5 | 2, 2, 8
6 | 6

Children's weights (lbs)	
Stem	**Leaf**
1	7
2	4
3	
4	
5	2, 2, 8
6	6

Question: Dr. Thornton, a pediatrician, weighed all the children who recently visited her office. How many children weighed exactly 52 pounds? (Unit: children)
Answer: (Step 1) For the number 52, the stem is 5, and the leaf is 2. Find the row where the stem is 5. In that row, count all the leaves equal to 2. (Step 2) You counted 2 leaves, which are blue in the stem-and-leaf plot above. (Step 3) 2 children weighed exactly 52 pounds. The answer is 2.

▷ ***Test example (ID: 17417)***

Table:
[TITLE]: Birthday party
Activity | Parents | Children
Singing | 14 | 20
Eating cake | 5 | 10
Jumping rope | 16 | 20
Swimming | 16 | 19
Playing tag | 4 | 9

Birthday party		
Activity	**Parents**	**Children**
Singing	14	20
Eating cake	5	10
Jumping rope	16	20
Swimming	16	19
Playing tag	4	9

Question: At Josie's birthday party, children and parents celebrated by participating in some activities. How many more children than parents are playing tag? (Unit: children)
Answer: (Step 1) For the activity "Playing tag", **the stem is 4, and the leaf is 9. Find the row where the stem is 4. In that row, count all the leaves equal to 9.** (Step 2) **You counted 9 leaves, which are blue in the stem-and-leaf plot above. 9 children** are playing tag. **The answer is 9.**

The **nearest neighbor search** strategy selects **semantically similar** examples to the test example.

The prediction follows the wrong reasoning steps and arrives at the **wrong** answer.

Failure Cases

Table:
yellow tiger's eye | $0.85
piece of green slate | $0.59
piece of red sandstone | $0.19
piece of rose quartz | $0.61
smooth piece of marble | $0.45

yellow tiger's eye	$0.85
piece of green slate	$0.59
piece of red sandstone	$0.19
piece of rose quartz	$0.61
smooth piece of marble	$0.45

Question: How much money does Connor need to buy a piece of rose quartz, a piece of green slate, and a piece of red sandstone? (unit: $)

Answer:

(Step 1) Connor needs to buy a piece of rose quartz, a piece of green slate, and a piece of red sandstone. To find the total amount of money Connor needs, add the prices of the three items.

(Step 2) $0.85 + $0.59 + $0.19 = $1.63

(Step 3) Connor needs $1.63 to buy a piece of rose quartz, a piece of green slate, and a piece of red sandstone.

The answer is 1.63.

Output: 1.63

Ground truth: 1.39

Failure Cases

Table:
[TITLE] Apples per tree
Stem | Leaf
1 | 1, 3, 6
2 | 2, 3, 3, 6
3 | 0
4 | 3
5 | 2, 6
6 | 0, 0, 2
7 | 2, 8
8 | 4, 5, 5, 6, 7

Apples per tree	
Stem	**Leaf**
1	1, 3, 6
2	2, 3, 3, 6
3	0
4	3
5	2, 6
6	0, 0, 2
7	2, 8
8	4, 5, 5, 6, 7

Question: A farmer counted the number of apples on each tree in her orchard. How many trees have at least 51 apples but fewer than 83 apples? (unit: trees)

Answer:

(Step 1) To solve this problem, look at the table and find all the trees that have at least 51 apples but fewer than 83 apples.

(Step 2) There are three trees that have at least 51 apples but fewer than 83 apples. The answer is 3.

Output: 3

Ground truth: 7

Failure Cases

Table:
[TITLE] Lawn mowers sold
Day | Number of lawn mowers
Tuesday | 8
Wednesday | 7
Thursday | 5
Friday | 7
Saturday | 9
Sunday | 8
Monday | 8

Lawn mowers sold	
Day	**Number of lawn mowers**
Tuesday	8
Wednesday	7
Thursday	5
Friday	7
Saturday	9
Sunday	8
Monday	8

Question: A garden supply store kept track of how many lawn mowers it sold in the past 7 days. What is the median of the numbers?
Answer:
(Step 1) The median is the "middle" value. To find the median, put the numbers in order from least to greatest and find the number in the middle.
(Step 2) 8, 7, 5, 7, 9, 8, 8
(Step 3) The median is 7. The answer is 7.
Output: 7
Ground truth: 8

Failure Cases

Table:
[TITLE] Tour boat schedule
Ocean City | 8:15 A.M. | 9:00 A.M. | 9:15 A.M. | 9:30 A.M. | 10:00 A.M.
Whale Watch Harbor | 9:30 A.M. | 10:15 A.M. | 10:30 A.M. | 10:45 A.M. | 11:15 A.M.
Oyster Lighthouse | 10:15 A.M. | 11:00 A.M. | 11:15 A.M. | 11:30 A.M. | 12:00 P.M.
Fisherman's Cove | 11:15 A.M. | 12:00 P.M. | 12:15 P.M. | 12:30 P.M. | 1:00 P.M.
Surfing Beach | 12:00 P.M. | 12:45 P.M. | 1:00 P.M. | 1:15 P.M. | 1:45 P.M.

Tour boat schedule					
Ocean City	8:15 A.M.	9:00 A.M.	9:15 A.M.	9:30 A.M.	10:00 A.M.
Whale Watch Harbor	9:30 A.M.	10:15 A.M.	10:30 A.M.	10:45 A.M.	11:15 A.M.
Oyster Lighthouse	10:15 A.M.	11:00 A.M.	11:15 A.M.	11:30 A.M.	12:00 P.M.
Fisherman's Cove	11:15 A.M.	12:00 P.M.	12:15 P.M.	12:30 P.M.	1:00 P.M.
Surfing Beach	12:00 P.M.	12:45 P.M.	1:00 P.M.	1:15 P.M.	1:45 P.M.

Question: Look at the following schedule. Haley is at Ocean City at 9.45 A.M. How soon can she get to Surfing Beach?
Options: (A) 1:45 P.M. (B) 12:00 P.M. (C) 10:30 A.M. (D) 12:15 P.M.
Answer:
(Step 1) Find 9:45 A.M. on the schedule.
(Step 2) The next stop is Whale Watch Harbor at 10:15 A.M.
(Step 3) The earliest she can get to Surfing Beach is 12:00 P.M. The answer is 12:00 P.M.
Output: (B) 12:00 P.M.
Ground truth: (A) 1:45 P.M.

Failure Cases

Table:
[TITLE] Blocks from school
Name | Number of blocks
Wendy | 14
Carson | 8
Zach | 17
Lila | 17
Nicole | 3
Shane | 7
Henry | 5
Devin | 17

Blocks from school	
Name	**Number of blocks**
Wendy	14
Carson	8
Zach	17
Lila	17
Nicole	3
Shane	7
Henry	5
Devin	17

Question: Some students compared how many blocks they live from school. What is the mean of the numbers?
Answer:
(Step 1) The mean is the average of the numbers.
(Step 2) To find the mean, add the numbers and divide by the number of students.
(Step 3) 14 + 8 + 17 + 17 + 3 + 7 + 5 + 17 = 86
(Step 4) There are 8 students, so divide 86 by 8.
(Step 5) 86 ÷ 8 = 10.75
(Step 6) The mean is 10.75. The answer is 10.75.
Output: 10.75
Ground truth: 11

ICLR

UCLA

Georgia Tech

AI2

PromptPG

Dynamic Prompt Learning via Policy Gradient for Semi-structured Mathematical Reasoning (ICLR 2023)

https://promptpg.github.io