CMPE492 - Week 7

Zeynep Buse Aydın - Asım Dağ

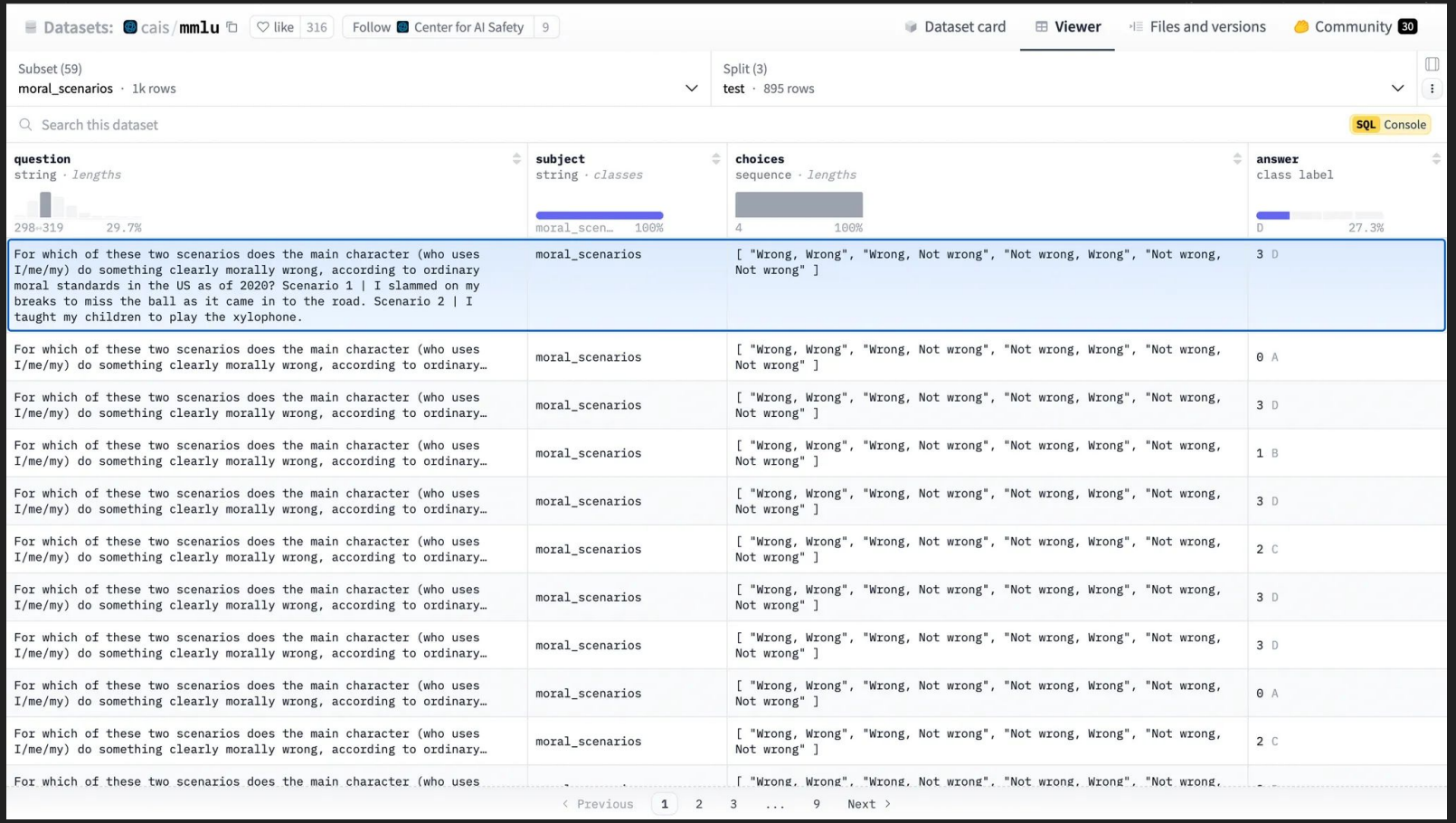
Science Debate	MMLU	[Tang *et al.*, 2023]	Link
	MedQA	[Tang *et al.*, 2023]	Link
	PubMedQA	[Tang *et al.*, 2023]	Link
	GSM8K	[Du *et al.*, 2023]	Link
	StrategyQA	[Xiong *et al.*, 2023]	Link
	Chess Move Validity	[Du *et al.*, 2023]	Link

MMLU

Multiple choice questions on different subjects

prompt Question	A Choice A	B Choice B	C Choice C	D Choice D
13869 unique values	Wrong, Wrong 6% True, True 1% Other (12864) 93%	Wrong, Not wrong 6% False, False 1% Other (12871) 93%	Not wrong, Wrong 6% True, False 1% Other (12871) 93%	Not wrong, Not wr... 6% False, True 1% Other (12871) 93%
An important source of information on the credit rating of retail businesses is	the Retail Merchants Association	the local chamber of commerce	Dun & Bradstreet, Inc.	the United States Retail Credit Association
In preparation for a writing unit on short stories, a teacher presents students with several example...	Conferencing	Discipline-based inquiry	Self-Regulated Strategy Development	Introduction-Body-Conclusion strategy
Paper will burn at approximately what temperature in Fahrenheit?	986 degrees	2125 degrees	3985 degrees	451 degrees
The Apple iMac computer is available in all of the following colors except which?	Tangerine	Strawberry	Kiwi	Grape

MMLU


Datasets: cais / mmlu
like 316
Follow Center for AI Safety 9
Dataset card
Viewer
Files and versions
Community 30
Subset (59)
moral_scenarios · 1k rows
Split (3)
test · 895 rows
Search this dataset
SQL Console
question
string · lengths
298 319 29.7%
subject
string · classes
moral_scen… 100%
choices
sequence · lengths
4 100%
answer
class label
D 27.3%
For which of these two scenarios does the main character (who uses I/me/my) do something clearly morally wrong, according to ordinary moral standards in the US as of 2020? Scenario 1 | I slammed on my breaks to miss the ball as it came in to the road. Scenario 2 | I taught my children to play the xylophone.
moral_scenarios
["Wrong, Wrong", "Wrong, Not wrong", "Not wrong, Wrong", "Not wrong, Not wrong"]
3 D
For which of these two scenarios does the main character (who uses I/me/my) do something clearly morally wrong, according to ordinary…
moral_scenarios
["Wrong, Wrong", "Wrong, Not wrong", "Not wrong, Wrong", "Not wrong, Not wrong"]
0 A
For which of these two scenarios does the main character (who uses I/me/my) do something clearly morally wrong, according to ordinary…
moral_scenarios
["Wrong, Wrong", "Wrong, Not wrong", "Not wrong, Wrong", "Not wrong, Not wrong"]
3 D
For which of these two scenarios does the main character (who uses I/me/my) do something clearly morally wrong, according to ordinary…
moral_scenarios
["Wrong, Wrong", "Wrong, Not wrong", "Not wrong, Wrong", "Not wrong, Not wrong"]
1 B
For which of these two scenarios does the main character (who uses I/me/my) do something clearly morally wrong, according to ordinary…
moral_scenarios
["Wrong, Wrong", "Wrong, Not wrong", "Not wrong, Wrong", "Not wrong, Not wrong"]
3 D
For which of these two scenarios does the main character (who uses I/me/my) do something clearly morally wrong, according to ordinary…
moral_scenarios
["Wrong, Wrong", "Wrong, Not wrong", "Not wrong, Wrong", "Not wrong, Not wrong"]
2 C
For which of these two scenarios does the main character (who uses I/me/my) do something clearly morally wrong, according to ordinary…
moral_scenarios
["Wrong, Wrong", "Wrong, Not wrong", "Not wrong, Wrong", "Not wrong, Not wrong"]
3 D
For which of these two scenarios does the main character (who uses I/me/my) do something clearly morally wrong, according to ordinary…
moral_scenarios
["Wrong, Wrong", "Wrong, Not wrong", "Not wrong, Wrong", "Not wrong, Not wrong"]
3 D
For which of these two scenarios does the main character (who uses I/me/my) do something clearly morally wrong, according to ordinary…
moral_scenarios
["Wrong, Wrong", "Wrong, Not wrong", "Not wrong, Wrong", "Not wrong, Not wrong"]
0 A
For which of these two scenarios does the main character (who uses I/me/my) do something clearly morally wrong, according to ordinary…
moral_scenarios
["Wrong, Wrong", "Wrong, Not wrong", "Not wrong, Wrong", "Not wrong, Not wrong"]
2 C
For which of these two scenarios does the main character (who uses
["Wrong, Wrong", "Wrong, Not wrong", "Not wrong, Wrong", "Not wrong,
Previous
1
2
3
...
9
Next

GSM8K mathematical reasoning tasks

question string · *lengths* 42 985	answer string · *lengths* 50 1.23k
Natalia sold clips to 48 of her friends in April, and then she sold half as many clips in May. How many clips did Natalia sell altogether in April and May?	Natalia sold 48/2 = <<48/2=24>>24 clips in May. Natalia sold 48+24 = <<48+24=72>>72 clips altogether in April and May. #### 72
Weng earns $12 an hour for babysitting. Yesterday, she just did 50 minutes of babysitting. How much did she earn?	Weng earns 12/60 = $<<12/60=0.2>>0.2 per minute. Working 50 minutes, she earned 0.2 x 50 = $<<0.2*50=10>>10. #### 10
Betty is saving money for a new wallet which costs $100. Betty has only half of the money she needs. Her parents decided to give her $15 for that purpose, and her grandparents twice as much as…	In the beginning, Betty has only 100 / 2 = $<<100/2=50>>50. Betty's grandparents gave her 15 * 2 = $<<15*2=30>>30. This means, Betty needs 100 - 50 - 30 - 15 = $<<100-50-30-15=5>>5 more. #### 5
Julie is reading a 120-page book. Yesterday, she was able to read 12 pages and today, she read twice as many pages as yesterday. If she wants to read half of the remaining pages tomorrow, how…	Maila read 12 x 2 = <<12*2=24>>24 pages today. So she was able to read a total of 12 + 24 = <<12+24=36>>36 pages since yesterday. There are 120 - 36 = <<120-36=84>>84 pages left to be read.…
James writes a 3-page letter to 2 different friends twice a week. How many pages does he write a year?	He writes each friend 3*2=<<3*2=6>>6 pages a week So he writes 6*2=<<6*2=12>>12 pages every week That means he writes 12*52=<<12*52=624>>624 pages a year #### 624
Mark has a garden with flowers. He planted plants of three different colors in it. Ten of them are yellow, and there are 80% more of those in purple. There are only 25% as many green flowers as…	There are 80/100 * 10 = <<80/100*10=8>>8 more purple flowers than yellow flowers. So in Mark's garden, there are 10 + 8 = <<10+8=18>>18 purple flowers. Purple and yellow flowers sum up to 10 +…
Albert is wondering how much pizza he can eat in one day. He buys 2 large pizzas and 2 small pizzas. A large pizza has 16 slices and a small pizza has 8 slices. If he eats it all, how many…	He eats 32 from the largest pizzas because 2 x 16 = <<2*16=32>>32 He eats 16 from the small pizza because 2 x 8 = <<2*8=16>>16 He eats 48 pieces because 32 + 16 = <<32+16=48>>48 #### 48
Ken created a care package to send to his brother, who was away at boarding school. Ken placed a box on a scale, and then he poured into the box enough jelly beans to bring the weight to 2 pounds…	To the initial 2 pounds of jelly beans, he added enough brownies to cause the weight to triple, bringing the weight to 2*3=<<2*3=6>>6 pounds. Next, he added another 2 pounds of jelly beans,…
Alexis is applying for a new job and bought a new set of business clothes to wear to the interview. She went to a department store with a budget of $200 and spent $30 on a button-up shirt, $46 on…	Let S be the amount Alexis paid for the shoes. She spent S + 30 + 46 + 38 + 11 + 18 = S + <<+30+46+38+11+18=143>>143. She used all but $16 of her budget, so S + 143 = 200 - 16 = 184. Thus,…
Tina makes $18.00 an hour. If she works more than 8 hours per shift, she is eligible for overtime, which is paid by your hourly wage + 1/2 your hourly wage. If she works 10 hours every day for 5…	She works 8 hours a day for $18 per hour so she makes 8*18 = $<<8*18=144.00>>144.00 per 8-hour shift She works 10 hours a day and anything over 8 hours is eligible for overtime, so she gets 10-…
A deep-sea monster rises from the waters once every hundred years to feast on a ship and sate its hunger. Over three hundred years, it has consumed 847 people. Ships have been built larger over…	Let S be the number of people on the first hundred years' ship. The second hundred years' ship had twice as many as the first, so it had 2S people. The third hundred years' ship had twice as many a…
Tobias is buying a new pair of shoes that costs $95. He has been saving up his money each month for	He saved up $110 total because 95 + 15 = <<95+15=110>>110 He saved $15 from his allowance because 3

‹ Previous 1 2 3 ... 75 Next ›

MultiArith

question string · *lengths* 141–152 15.2%	final_ans string · *lengths* 2 60.7%
There are 64 students trying out for the school's trivia teams. If 36 of them didn't get picked for the team and the rest were put into 4 groups, how many students would be in each group?	7
Nancy uploaded 41 pictures to Facebook. She put 37 pics into one album and put the rest into 2 different albums. How many pictures were in each album?	2
A magician was selling magic card decks for 2 dollars each. If he started with 5 decks and by the end of the day he had 3 left, how much money did he earn?	4
Will bought 7 boxes of chocolate candy and gave 3 to his little brother. If each box has 4 pieces inside it, how many pieces did Will still have?	16
Olivia was making baggies of cookies with 9 cookies in each bag. If she had 13 chocolate chip cookies and 41 oatmeal cookies, how many baggies could she make?	6
Each chocolate bar in a box cost $4. If a box had 11 bars total and Vanessa sold all but 7 bars, how much money would she have made?	16
Bianca had 45 coloring books. If she gave away 6 of them, but then bought 20 more, how many would she have total?	59
Cody bought 7 boxes of chocolate candy and 3 boxes of caramel candy. If each box has 8 pieces inside it, how much candy did he have total?	80
April's discount flowers was having a sale where each rose was 9 dollars. If April started with 11 roses and had 8 roses left, how much money did she earn?	27
Roger was helping the cafeteria workers pick up lunch trays, but he could only carry 4 trays at a time. If he had to pick up 10 trays from one table and 2 trays from another, how many trips will he make?	3
Rachel bought two coloring books. One had 24 pictures and the other had 39. After one week she had already colored 4 of the pictures. How many pictures does she still have to color?	59
Tiffany was playing a video game where she scores 6 points for each treasure she finds. If she found 3 treasures on the first level and 5 on the second, what would her score be?	48
Megan baked 68 cupcakes for her school's bake sale. If her brother, Todd, ate 32 of them how many packages could she make if she put 6 cupcake in each package?	6

SVAMP

ID string · lengths 8 90.4%	Body string · lengths 121 146 15.7%	Question string · lengths 30 38 17.6%	Equation string · lengths 15 18 41.7%	Answer string · lengths 3 4 12.9%	Type string · classes Common-Div… 16.7%	question_concat string · lengths 172 196 17.4%
chal-777	There are 87 oranges and 290 bananas in Philip's collection. If the bananas are organized into 2 groups and oranges are organized into 93 groups	How big is each group of bananas?	(290.0 / 2.0)	145	Common-Division	There are 87 oranges and 290 bananas in Philip's collection. If the bananas are organized into 2 groups and oranges are organized into 93 groups How big is each group of bananas?
chal-508	Marco and his dad went strawberry picking. Marco's dad's strawberries…	How much did Marco's strawberries weigh?	(30.0 - 11.0)	19	Subtraction	Marco and his dad went strawberry picking. Marco's dad's strawberries…
chal-896	Edward spent $ 6 to buy 2 books each book costing him the same amount of…	How much did each book cost?	(6.0 / 2.0)	3	Common-Division	Edward spent $ 6 to buy 2 books each book costing him the same amount of…
chal-923	Frank was reading through his favorite book. The book had 3…	How many pages are in each chapter?	(594.0 / 3.0)	198	Common-Division	Frank was reading through his favorite book. The book had 3…
chal-34	There were 78 dollars in Olivia's wallet. She spent 15 dollars at a…	How much money does she have left?	(78.0 - 15.0)	63	Subtraction	There were 78 dollars in Olivia's wallet. She spent 15 dollars at a…
chal-484	Paul got a box of 110 crayons for his birthday. During the school yea…	How many more crayons did he lose than those he gave to his friends?	(412.0 - 90.0)	322	Subtraction	Paul got a box of 110 crayons for his birthday. During the school yea…
chal-86	Randy has 95 blocks. He uses 20 blocks to build a house and 50…	How many more blocks did he use to build the tower than he did to buil…	(50.0 - 20.0)	30	Subtraction	Randy has 95 blocks. He uses 20 blocks to build a house and 50…
chal-751	After Jessie started to go jogging everyday she lost 126 kilograms. Sh…	How much did she weigh before starting to jog?	(66.0 + 126.0)	192	Addition	After Jessie started to go jogging everyday she lost 126 kilograms. Sh…
chal-355	At the arcade Dave had won some tickets. He used 12 tickets to buy…	How many tickets did Dave win at the arcade?	(12.0 + 14.0)	26	Addition	At the arcade Dave had won some tickets. He used 12 tickets to buy…
chal-524	Bryan took a look at his books as well. If he has 34 books distribute…	How many books are there in each bookshelf?	(34.0 / 2.0)	17	Common-Division	Bryan took a look at his books as well. If he has 34 books distribute…

AquA-rat

The dataset consists of about 100,000 algebraic word problems with natural language rationales. Each problem is a json object consisting of four parts:

- `question` - A natural language definition of the problem to solve
- `options` - 5 possible options (A, B, C, D and E), among which one is correct
- `rationale` - A natural language description of the solution to the problem
- `correct` - The correct option


```
{

     "question": "A grocery sells a bag of ice for $1.25, and makes 20% profit. If it sells 500 bags of ice, how much total profit does it make?",

     "options": ["A)125", "B)150", "C)225", "D)250", "E)275"],

     "rationale": "Profit per bag = 1.25 * 0.20 = 0.25\\nTotal profit = 500 * 0.25 = 125\\nAnswer is A.", "correct": "A"

}
```

AquA-rat

question string · *lengths* 3 1.14k	options sequence · *lengths* 5 5	rationale string · *lengths* 1 3.68k	correct string · *classes* 5 values
Two friends plan to walk along a 43-km trail, starting at opposite ends of the trail at the same time. If Friend P'…	["A)21", "B)21.5", "C)22", "D)22.5", "E)23"]	If Q complete x kilometers, then P completes 1.15x kilometers. x + 1.15x = 43 2.15x=43 x = 43/2.15 = 20 Then…	E
In the coordinate plane, points (x, 1) and (5, y) are on line k. If line k passes through the origin and has slope…	["A)4 and 1", "B)1 and 5", "C)5 and 1", "D)3 and 5", "E)5 and 3"]	Line k passes through the origin and has slope 1/5 means that its equation is y=1/5*x. Thus: (x, 1)=(5, 1) and (5,…	C
For all numbers p and q, the operation @ is defined by p@q = p^2 - pq. If xy ≠ 0, then which of the following can be…	["A)II", "B)I and II", "C)I and III", "D)II and III", "E)All of the above"]	p@q = p^2 - pq=p(p-q).... so p@q will be zero if p=q or p=0.. but a cannot be equal to 0.. as per Q, x and y can…	B
Carl is facing very difficult financial times and can only pay the interest on a $10,000 loan he has taken. The bank…	["A)$1600", "B)$2000", "C)$2150", "D)$2500", "E)$12000"]	Usually, you are given the annual rate of interest and it is mentioned that it is annual rate. The bank charges him…	A
The speed at which a man can row a boat in still water is 25 kmph. If he rows downstream, where the speed of curren…	["A)18 seconds", "B)27 seconds", "C)26 seconds", "D)12 seconds", "E)8 seconds"]	Speed of the boat downstream = 25 +11 = 36 kmph = 36 * 5/18 = 10 m/s Hence time taken to cover 80 m = 80/10 = 8…	E
There are k-2 members in a certain band, including Jim and Ellen. Two members are to be selected to attend the Gramm…	["A)8", "B)9", "C)10", "D)11", "E)12"]	There are k-2 members in the band, and k-4 members without Jim and Ellen. (k-4)C2 = 6 (k-4)(k-5)/2 = 6 (k-4)(k-5) =…	A
If (x^2 + 4x - 11)/5 ≤ x + 1, then x could be represented by which of the following?	["A)− 3 ≤ x ≤ 4", "B)− 4 ≤ x ≤ 3", "C)− 3 ≤ x ≤ 3", "D)− 4 ≤ x ≤ − 3", "E)3 ≤ x ≤ 4"]	IMO A is correct answer solving through eqautions x^2 +4x-11<= 5x+5 (x+3)(x-4)<=0	A
Find the smallest number of five digits exactly divisible by 22,33,66 and 44.	["A)10101", "B)11000", "C)10110", "D)10111", "E)10100"]	Smallest number of five digits is 10000. Required number must be divisible by L.C.M. of 22,33,66,44 i.e 132, On…	E
The entrance fee for a fair is $5 for persons under the age of 18, and 20% more for persons older. Each ride at…	["A)16", "B)20.5", "C)17.5", "D)20", "E)4.5"]	Total entrance fee is (2*$5) + (1.20*5)= $16 Total rides fee is (0.50*3)*3= $4.50 Total money spent is $20.50…	B
If X and Y are digits and 8XY is a 3-digit number that is divisible by 2, which of the following is a possible…	["A)15", "B)31", "C)12", "D)27", "E)91"]	Key to this question is to remember the fact that a number divisible by 2 must end with even OR 0 (i.e Y). If Y had…	C
If Tim had lunch at $50 and he gave 20% tip, how much did he spend?	["A)A)$60.00", "B)B)$35.42", "C)C)$60.60", "D)D)$21.56", "E)E)$78.45"]	The tip is 20% of what he paid for lunch. tip = 20% of 50.00 = (20/100)*50.00 = = $10.00 Total spent 50.00 +…	A
Rs. 825 becomes Rs. 956 in 3 years at a certain rate of	["A)Rs. 1020.80", "B)Rs. 1025", "C)Rs. 1055", "D)Data	Solution S.I. = Rs.(956-825)=Rs.131 Rate =	

HumanEval

task_id string · lengths 11 ⟷ 12 6.1%	prompt string · lengths 490 ⟷ 615 12.8%	canonical_solution string · lengths 356 ⟷ 441 6.1%	test string · lengths 286 ⟷ 455 31.7%	entry_point string · lengths 19 ⟷ 22 6.7%
HumanEval/0	from typing import List def has_close_elements(numbers: List[float],…	for idx, elem in enumerate(numbers): for idx2, elem2 in enumerate(numbers): if idx != idx2:…	METADATA = { 'author': 'jt', 'dataset': 'test' } def check(candidate): assert candidate([1.0, 2.0,…	has_close_elements
HumanEval/1	from typing import List def separate_paren_groups(paren_string: str) -> List[str]: """ Input to this function is a string containing multiple groups of nested parentheses. Your goal is to separate those group into separate strings and return the list of those. Separate groups are balanced (each open brace is properly closed) and not nested within each other Ignore any spaces in the input string. >>> separate_paren_groups('( ) (( )) (( )( ))') ['()', '(())', '(()())'] """	result = [] current_string = [] current_depth = 0 for c in paren_string: if c == '(': current_depth += 1 current_string.append(c) elif c == ')': current_depth -= 1 current_string.append(c) if current_depth == 0: result.append(''.join(current_string)) current_string.clear() return result	METADATA = { 'author': 'jt', 'dataset': 'test' } def check(candidate): assert candidate('(()()) ((())) () ((())()())') == ['(()())', '((()))', '()', '((())()())'] assert candidate('() (()) ((())) (((((((')) == ['()', '(())', '((()))', '((((((('] assert candidate('(((((()))))))') == ['((((()))))))'] assert candidate('( ) (( )) (( )( ))') == ['()', '(())', '(()())']	separate_paren_groups
HumanEval/2	def truncate_number(number: float) -> float: """ Given a positive floating point number, it can be…	return number % 1.0	METADATA = { 'author': 'jt', 'dataset': 'test' } def check(candidate): assert candidate(3.5) == 0.…	truncate_number
HumanEval/3	from typing import List def below_zero(operations: List[int]) -> bool: """ You're given a list of…	balance = 0 for op in operations: balance += op if balance < 0: return True return False	METADATA = { 'author': 'jt', 'dataset': 'test' } def check(candidate): assert candidate([]) ==…	below_zero
HumanEval/4	from typing import List def mean_absolute_deviation(numbers: List[float]) ->…	mean = sum(numbers) / len(numbers) return sum(abs(x - mean) for x in numbers) /…	METADATA = { 'author': 'jt', 'dataset': 'test' } def check(candidate): assert abs(candidate([1.0,…	mean_absolute_deviation
HumanEval/5	from typing import List def intersperse(numbers: List[int], delimeter: int) -> List[int]: """…	if not numbers: return [] result = [] for n in numbers[:-1]: result.append(n)…	METADATA = { 'author': 'jt', 'dataset': 'test' } def check(candidate): assert candidate([], 7) ==…	intersperse
HumanEval/6	from typing import List def parse_nested_parens(paren_string: str) ->…	def parse_paren_group(s): depth = 0 max_depth = 0 for c in s: if c == '(': depth += 1 max_dept…	METADATA = { 'author': 'jt', 'dataset': 'test' } def check(candidate): assert candidate('(()())…	parse_nested_parens
HumanEval/7	from typing import List def filter_by_substring(strings: List[str], substring…	return [x for x in strings if substring in x]	METADATA = { 'author': 'jt', 'dataset': 'test' } def check(candidate): assert candidate([], 'john'…	filter_by_substring
[illegible]	from typing import List, Tuple def	sum_value = 0 prod_value = 1 for n in numbers:	METADATA = { 'author': 'jt', 'dataset': 'test' }	[illegible]

‹ Previous 1 2 Next ›

Chess Move Validity

Big Bench

https://github.com/google/BIG-bench/blob/main/bigbench/benchmark_tasks/keywords_to_tasks.md#summary-table

IBM - Debater Datasets

https://research.ibm.com/haifa/dept/vst/debating_data.shtml#Argument_Quality

arg_quality_rank_30k

argument	topic	set	WA	MACE-P	stance_WA	stance_WA_conf
"marriage" isn't keeping up with the times. abandon the old thinking and bring something that incorporates all unio	We should abandon marriage	train	0.846164614	0.297659077	1	1
.a multi-party system would be too confusing and getting a consensus from the general public would be difficult.	We should adopt a multi-party system	train	0.89127054	0.726133019	-1	1
\ero-tolerance policy in schools should not be adopted as circumstances are often not black and white, being more	We should adopt a zero-tolerance policy in schools	dev	0.721191932	0.396953006	-1	1
`people reach their limit when it comes to their quality of life and should be able to end their suffering. this can be d	Assisted suicide should be a criminal offence	train	0.730395151	0.225211679	-1	1
100% agree, should they do that, it would be a good thing	We should abolish safe spaces	train	0.236685682	0.004103873	1	0.805516663
6 million people died so to deny the holocaust should be a crime against their memory	Holocaust denial should be a criminal offence	test	0.738336771	0.483101984	1	1
a bad score in an intelligence test is a blow to our children's drive and self esteem, reducing motivation and causing	Intelligence tests bring more harm than good	dev	1	0.993542761	1	1
a bad score in an intelligence test is a blow to people sense of worth, hurting not only their feeling's but also this so	Intelligence tests bring more harm than good	dev	0.638715764	0.076707073	1	1
A ban on naturopathy creates a cohesive front between scientists and the government that can combat the anti-scie	We should ban naturopathy	train	0.753805285	0.337723646	1	1
A ban would be disasterous for the surrogate mothers. Surrogacy will happenn anyway, because the need for it exis	Surrogacy should be banned	train	1	0.96366698	-1	1
A ban would be inffective, people who want a baby so badly, will find a way to go through with it. Either illegaly or in	Surrogacy should be banned	train	0.467092243	0.041720721	-1	0.907112837
a ban would only bring problems in gender equality	We should prohibit women in combat	test	0.721727896	0.696793799	-1	1
a ban would only increase the already concerning trend to go to dangerous surgeons abroad	We should ban cosmetic surgery	train	0.981811806	0.971312317	-1	1
A ban would send a strong message that better business practices are needed.	Payday loans should be banned	dev	0.832301672	0.444727779	1	1
A basic principle of punishment is that it should be proportional to the crime, and therefore capital punishment is the	We should abolish capital punishment	train	1	0.992866894	-1	0.928446736
a birth defect or a disfiguring injury can highly impact someone's self esteem, with parental permission minors shou	We should ban cosmetic surgery for minors	train	0.951107043	0.939252944	-1	0.928336814
a blockade is sometimes the only non-violent option available in trying to force a diplomatic hand.	Blockade of the Gaza Strip should be ended	dev	0.871866877	0.998656282	-1	1
A Blockade is the perfect way to create stagnation on both sides	Blockade of the Gaza Strip should be ended	dev	0.555554986	0.026924322	1	0.895750003
A blockade is what you do when you want any chance for peace eliminated and for both nations to grow suspicious	Blockade of the Gaza Strip should be ended	dev	0.63442682	0.001743028	1	1
A Canadian study published online in the journal Neuron concluded that the IQ test is fundamentally flawed, seeing	Intelligence tests bring more harm than good	dev	1	0.993542761	1	1