INSIGHT BUILDER

SUPPORTED BY
SIMPLY DISCOVER

WRITING DSPY METRICS & DEBUGGING PRGM

Bootstrap

Optimizrers and Evaluators development

PHOENIX

Part 2: Looking at Bootstraping Process in DSPy

WHY & HOW OF METRICS

METRICS RE-PRIMER

PARTS OF METRICS

METRICS & EVALUATION

OPTIMIZER COMPILATION

TRACING LLM CALLS

SPANS REVIEW & DEBUGGING

METRICS RE-PRIMER

metric is just a function that will take examples from your data and take the output of your system, and return a score that quantifies how good the output is. What makes outputs from your system good or bad?

Simple Metric: A simple python function that asserts if the pred output is equal to the gold train data

Metric with AI Feedback: Similar to above, the long form outputs are processed with AI and then compared

Advanced Metric That Uses DSPy Programs: Use complete programs to get the final output to compare

METRICS PARTS

Gold Examples: Examples that are having the Inputs and its corresponding Output that is required. Usually taken from train data

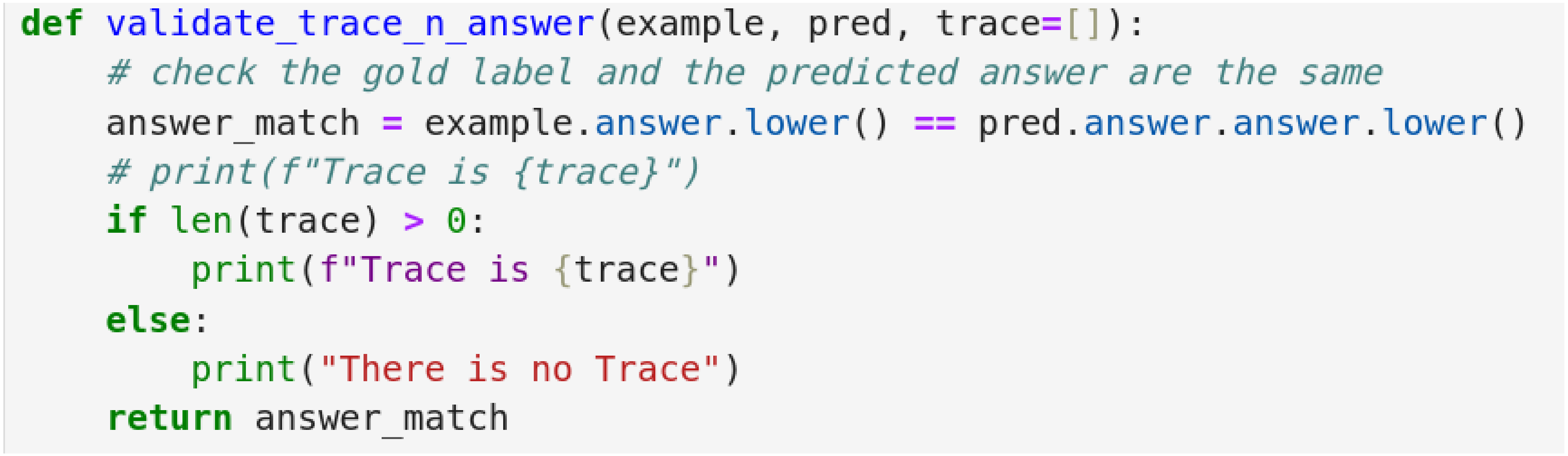
Pred Outputs: This is part of the Prediction object, that is created after the Examples are sent through the DSPy program

Traces: Contains the Input Examples and corresponding Prediction object that is to be “Bootstraped” for improviing the Program

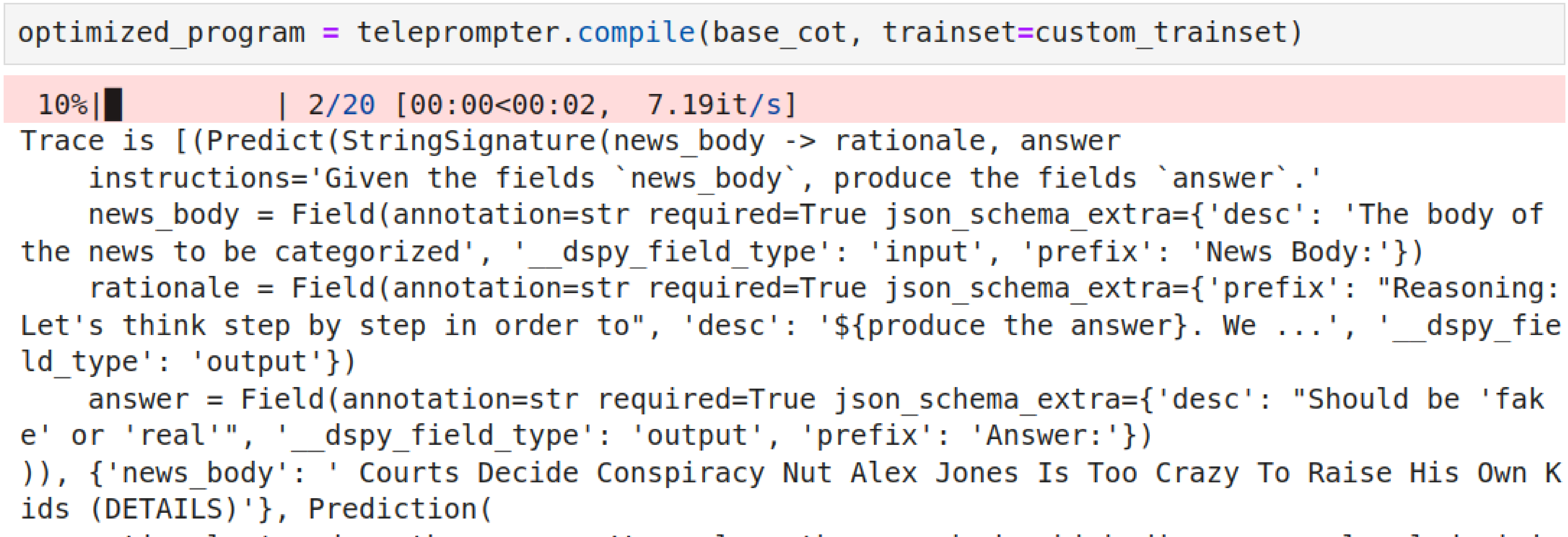
The metric can return a float / bool or integers. For sake of Sanity DSPy devs suggest to keep the return value between 0 to 1

There will no trace during evaluation or optimization. Trace is generated if metric is used to bootstrap demonstrations.

METRIC PARTS: OPTIMISATION



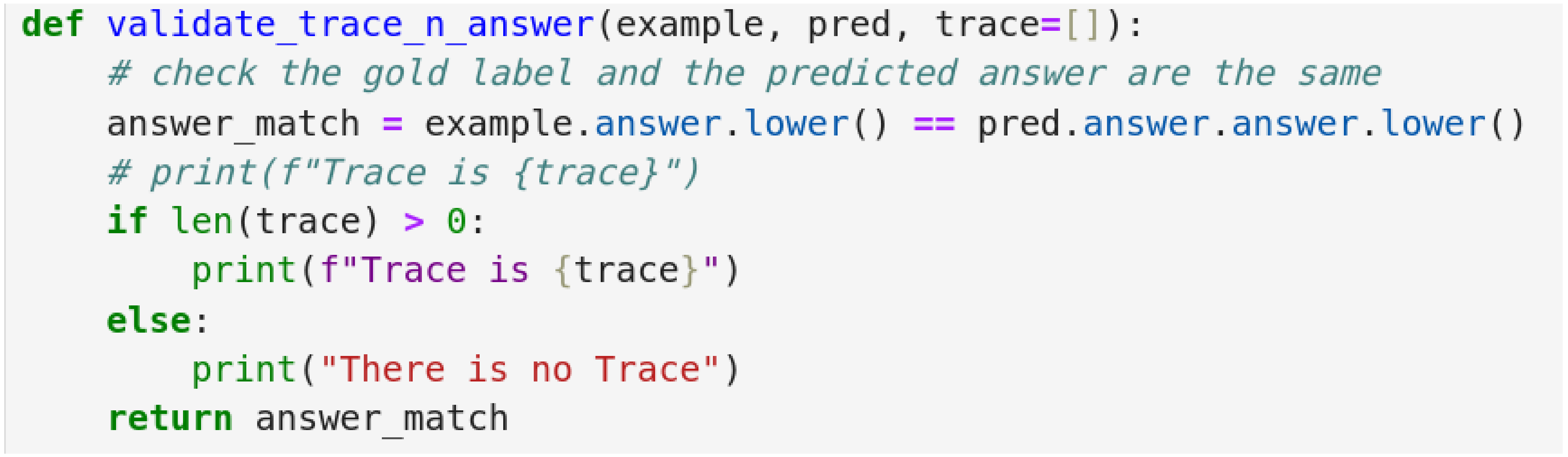
```python
def validate_trace_n_answer(example, pred, trace=[]):
    # check the gold label and the predicted answer are the same
    answer_match = example.answer.lower() == pred.answer.answer.lower()
    # print(f"Trace is {trace}")
    if len(trace) > 0:
        print(f"Trace is {trace}")
    else:
        print("There is no Trace")
    return answer_match
```



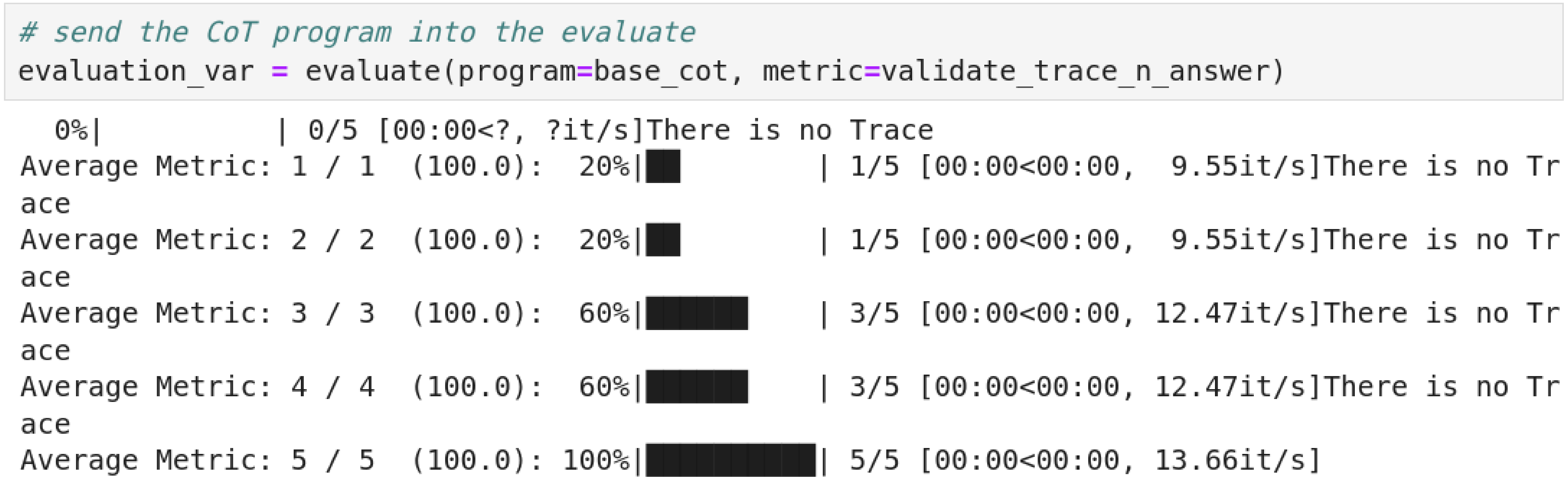
```python
optimized_program = teleprompter.compile(base_cot, trainset=custom_trainset)
```

```
 10%|█          | 2/20 [00:00<00:02,  7.19it/s]
Trace is [(Predict(StringSignature(news_body -> rationale, answer
    instructions='Given the fields `news_body`, produce the fields `answer`.'
    news_body = Field(annotation=str required=True json_schema_extra={'desc': 'The body of
the news to be categorized', '__dspy_field_type': 'input', 'prefix': 'News Body:'})
    rationale = Field(annotation=str required=True json_schema_extra={'prefix': "Reasoning:
Let's think step by step in order to", 'desc': '${produce the answer}. We ...', '__dspy_fie
ld_type': 'output'})
    answer = Field(annotation=str required=True json_schema_extra={'desc': "Should be 'fak
e' or 'real'", '__dspy_field_type': 'output', 'prefix': 'Answer:'})
)), {'news_body': ' Courts Decide Conspiracy Nut Alex Jones Is Too Crazy To Raise His Own K
ids (DETAILS)'}, Prediction(
```

METRIC PARTS: EVALUATION



```python
def validate_trace_n_answer(example, pred, trace=[]):
    # check the gold label and the predicted answer are the same
    answer_match = example.answer.lower() == pred.answer.answer.lower()
    # print(f"Trace is {trace}")
    if len(trace) > 0:
        print(f"Trace is {trace}")
    else:
        print("There is no Trace")
    return answer_match
```



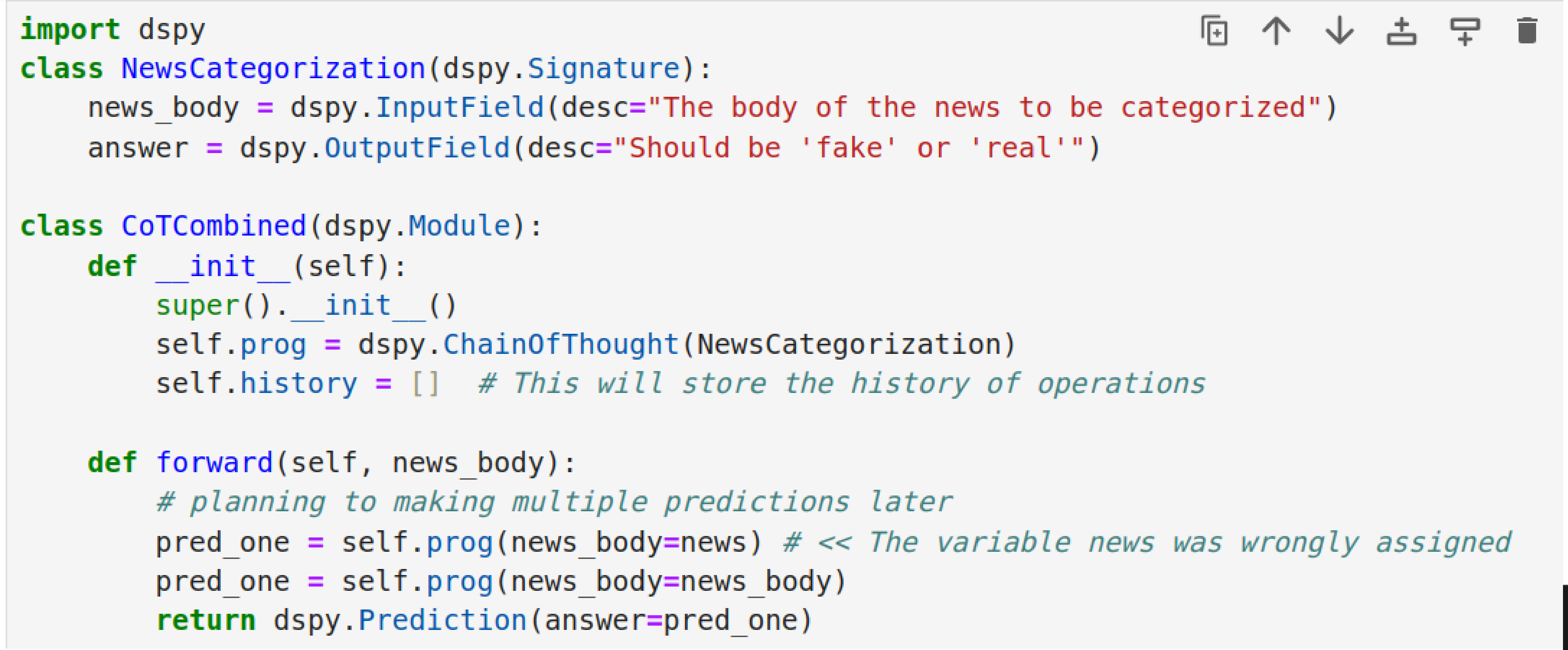
```python
# send the CoT program into the evaluate
evaluation_var = evaluate(program=base_cot, metric=validate_trace_n_answer)
```

```
  0%|          | 0/5 [00:00<?, ?it/s]There is no Trace
Average Metric: 1 / 1  (100.0):  20%|██        | 1/5 [00:00<00:00,  9.55it/s]There is no Tr
ace
Average Metric: 2 / 2  (100.0):  20%|██        | 1/5 [00:00<00:00,  9.55it/s]There is no Tr
ace
Average Metric: 3 / 3  (100.0):  60%|██████    | 3/5 [00:00<00:00, 12.47it/s]There is no Tr
ace
Average Metric: 4 / 4  (100.0):  60%|██████    | 3/5 [00:00<00:00, 12.47it/s]There is no Tr
ace
Average Metric: 5 / 5  (100.0): 100%|██████████| 5/5 [00:00<00:00, 13.66it/s]
```

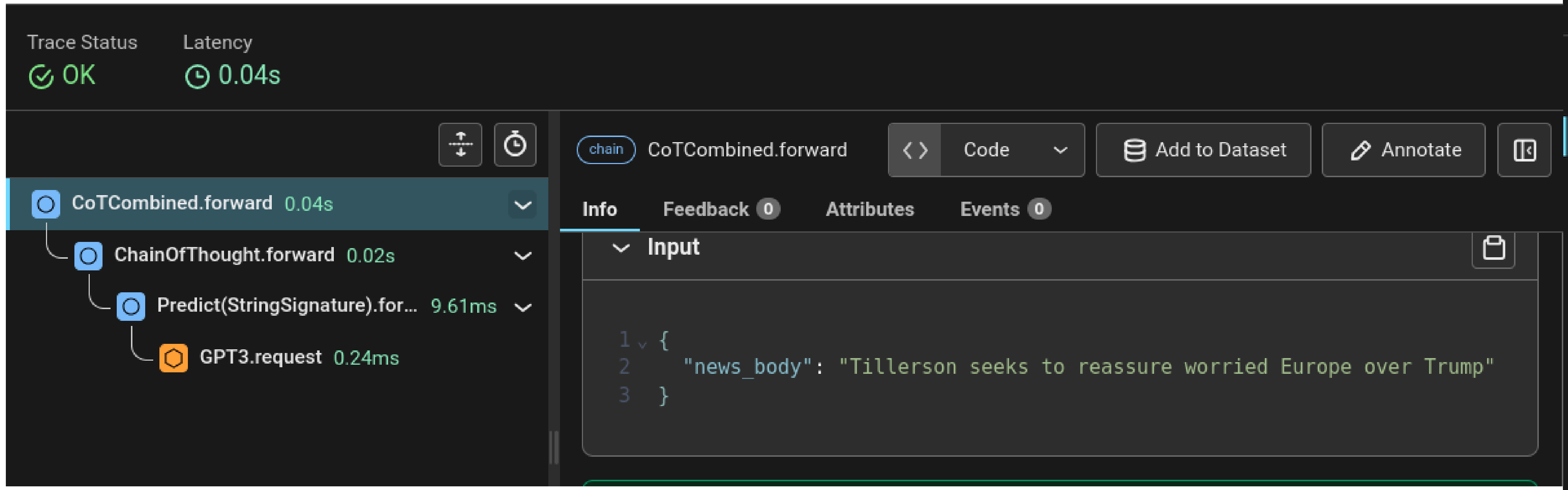
DSPY TRACES ARE NOT PRESENT ALL THE TIME

THATS WHERE PHOENIX TRACES CAN HELP

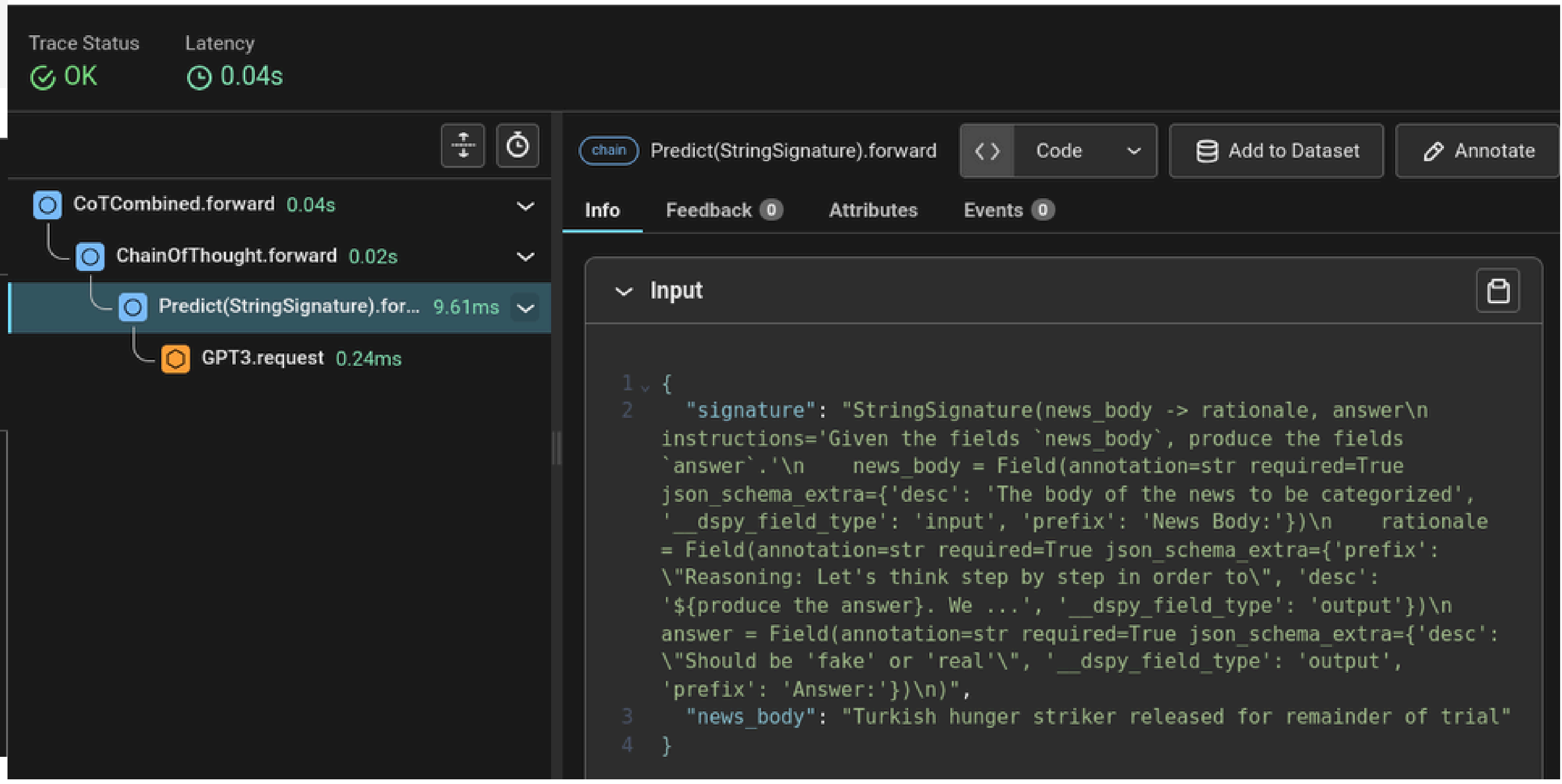
DEBUGGING WRONG PRGM



```python
import dspy
class NewsCategorization(dspy.Signature):
    news_body = dspy.InputField(desc="The body of the news to be categorized")
    answer = dspy.OutputField(desc="Should be 'fake' or 'real'")

class CoTCombined(dspy.Module):
    def __init__(self):
        super().__init__()
        self.prog = dspy.ChainOfThought(NewsCategorization)
        self.history = []  # This will store the history of operations

    def forward(self, news_body):
        # planning to making multiple predictions later
        pred_one = self.prog(news_body=news) # << The variable news was wrongly assigned
        pred_one = self.prog(news_body=news_body)
        return dspy.Prediction(answer=pred_one)
```



```
{
  "news_body": "Tillerson seeks to reassure worried Europe over Trump"
}
```

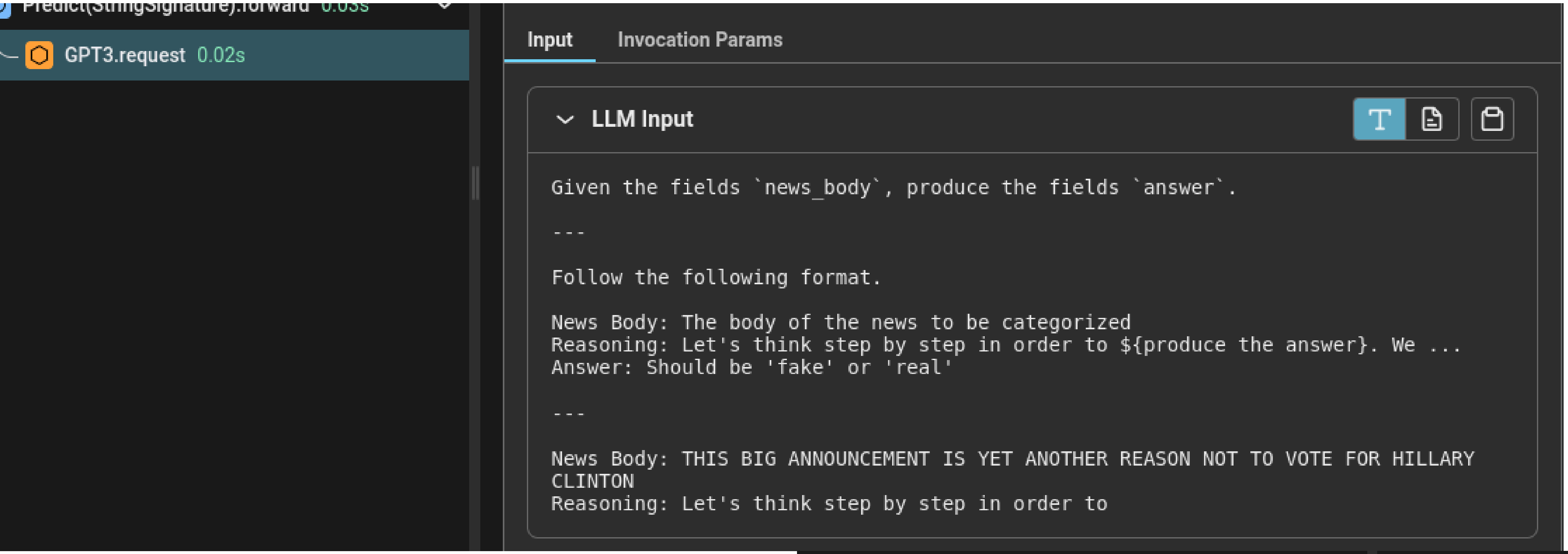


Trace Status
OK
Latency
0.04s
CoTCombined.forward 0.04s
ChainOfThought.forward 0.02s
Predict(StringSignature).for... 9.61ms
GPT3.request 0.24ms
chain
Predict(StringSignature).forward
Code
Add to Dataset
Annotate
Info
Feedback 0
Attributes
Events 0
Input

```
{
  "signature": "StringSignature(news_body -> rationale, answer\n
instructions='Given the fields `news_body`, produce the fields
`answer`.'\n    news_body = Field(annotation=str required=True
json_schema_extra={'desc': 'The body of the news to be categorized',
'__dspy_field_type': 'input', 'prefix': 'News Body:'})\n    rationale
= Field(annotation=str required=True json_schema_extra={'prefix':
\"Reasoning: Let's think step by step in order to\", 'desc':
'${produce the answer}. We ...', '__dspy_field_type': 'output'})\n
answer = Field(annotation=str required=True json_schema_extra={'desc':
\"Should be 'fake' or 'real'\", '__dspy_field_type': 'output',
'prefix': 'Answer:'})\n)",
  "news_body": "Turkish hunger striker released for remainder of trial"
}
```

DSPY PROGRAM INPUT & GPT API CALL INPUT ARE DIFFERENT

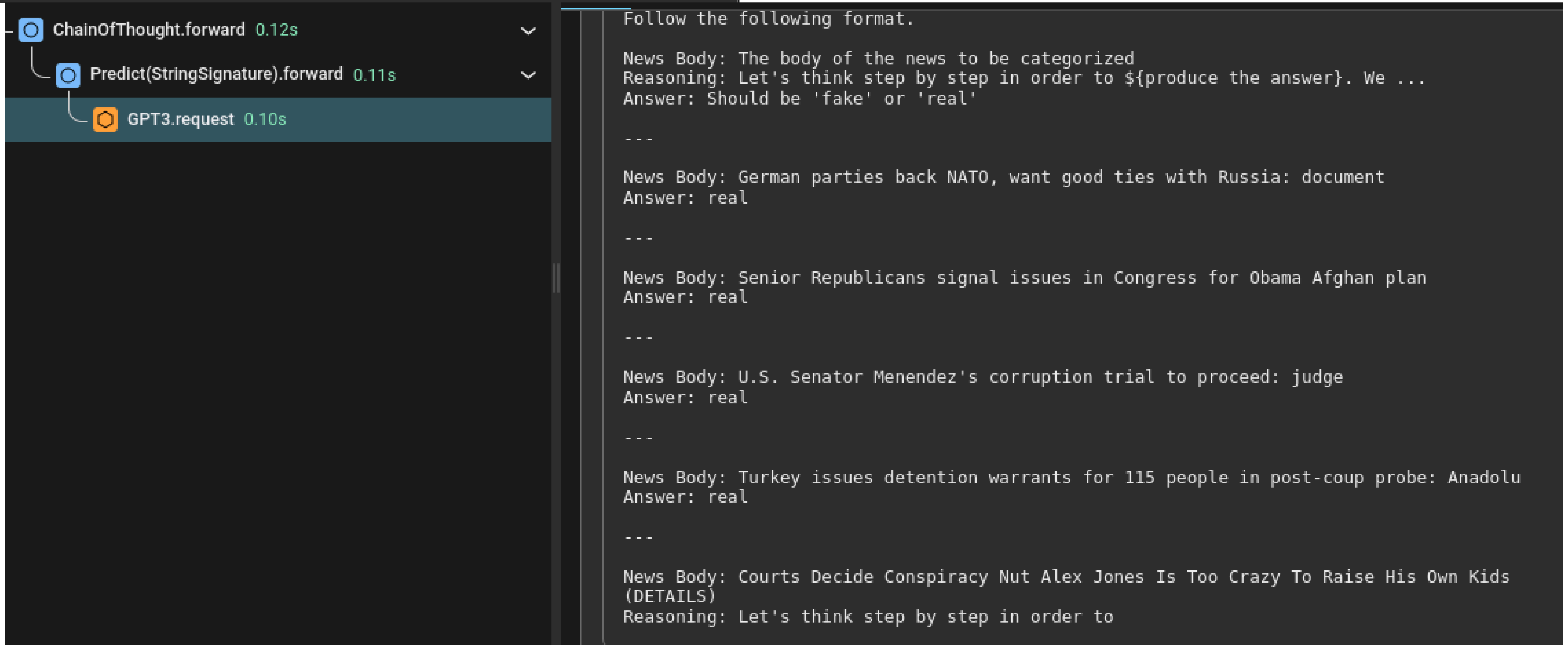
LOOKING AT BOOTSTRAPED CALLS



```
Given the fields `news_body`, produce the fields `answer`.

---

Follow the following format.

News Body: The body of the news to be categorized
Reasoning: Let's think step by step in order to ${produce the answer}. We ...
Answer: Should be 'fake' or 'real'

---

News Body: THIS BIG ANNOUNCEMENT IS YET ANOTHER REASON NOT TO VOTE FOR HILLARY
CLINTON
Reasoning: Let's think step by step in order to
```

NO BOOTSTRAPPED EXAMPLES

WITH BOOTSTRAPPED EXAMPLES


ChainOfThought.forward 0.12s
Predict(StringSignature).forward 0.11s
GPT3.request 0.10s

```
Follow the following format.

News Body: The body of the news to be categorized
Reasoning: Let's think step by step in order to ${produce the answer}. We ...
Answer: Should be 'fake' or 'real'

---

News Body: German parties back NATO, want good ties with Russia: document
Answer: real

---

News Body: Senior Republicans signal issues in Congress for Obama Afghan plan
Answer: real

---

News Body: U.S. Senator Menendez's corruption trial to proceed: judge
Answer: real

---

News Body: Turkey issues detention warrants for 115 people in post-coup probe: Anadolu
Answer: real

---

News Body: Courts Decide Conspiracy Nut Alex Jones Is Too Crazy To Raise His Own Kids
(DETAILS)
Reasoning: Let's think step by step in order to
```

THANKS FOR WATCHING

LIKE

SHARE

SUBSCRIBE

HTTPS://GITHUB.COM/INSIGHTBUILDER