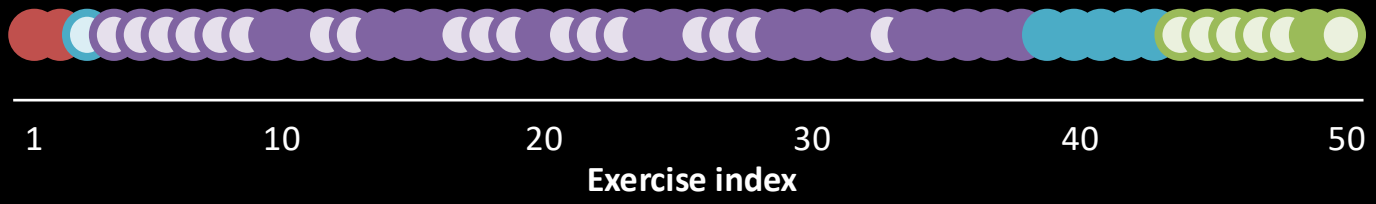
CS398: Computational Education

Insight Tracing

Story of Riley



Exercise Type:

- Solving for x-intercept
- Solving for y-intercept
- Graphing linear equations
- Square roots
- Slope of a line

Answer:

- Correct
- Incorrect

KHAN ACADEMY

Story of Riley

Exercise Type:

- Solving for x-intercept
- Solving for y-intercept
- Graphing linear equations
- Square roots
- Slope of a line

Answer:

- Correct
- Incorrect

KHAN ACADEMY

Story of Riley

Exercise Type:

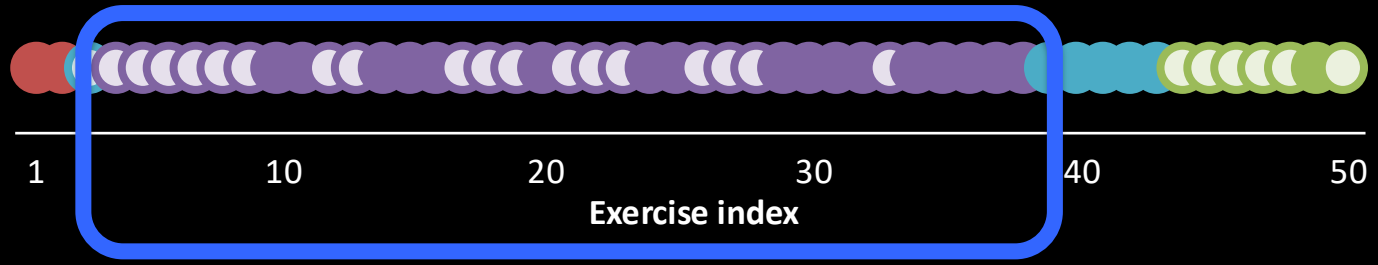
- Solving for x-intercept
- Solving for y-intercept
- Graphing linear equations
- Square roots
- Slope of a line

Answer:

- Correct
- Incorrect

KHAN ACADEMY

Story of Riley



Exercise Type:

- Solving for x-intercept
- Solving for y-intercept
- Graphing linear equations
- Square roots
- Slope of a line

Answer:

- Correct
- Incorrect

KHAN ACADEMY

Story of Riley

Exercise Type:

- Solving for x-intercept
- Solving for y-intercept
- Graphing linear equations
- Square roots
- Slope of a line

Answer:

- Correct
- Incorrect

KHAN ACADEMY

Story of Riley

Exercise Type:

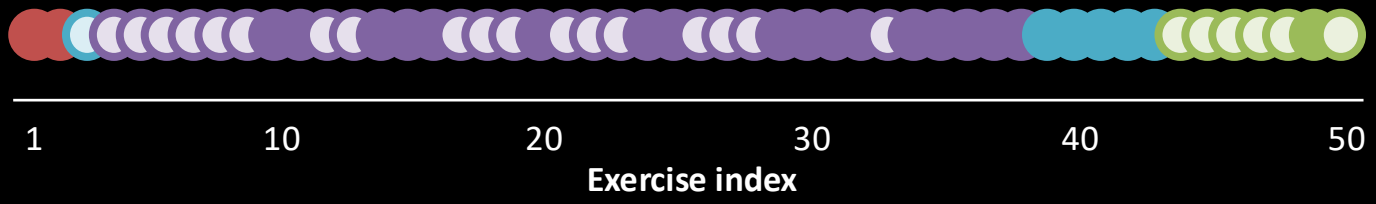
- Solving for x-intercept
- Solving for y-intercept
- Graphing linear equations
- Square roots
- Slope of a line

Answer:

- Correct
- Incorrect

KHAN ACADEMY

Story of Riley



What does Riley know?

Exercise Type:

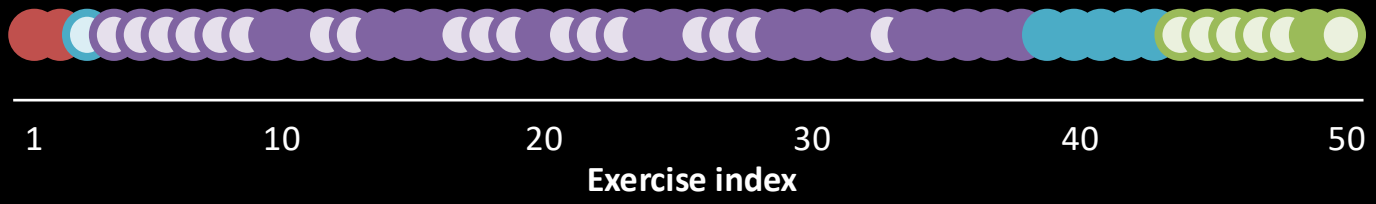
- Solving for x-intercept
- Solving for y-intercept
- Graphing linear equations
- Square roots
- Slope of a line

Answer:

- Correct
- Incorrect

KHAN ACADEMY

Story of Riley



What should Riley do next?

Exercise Type:

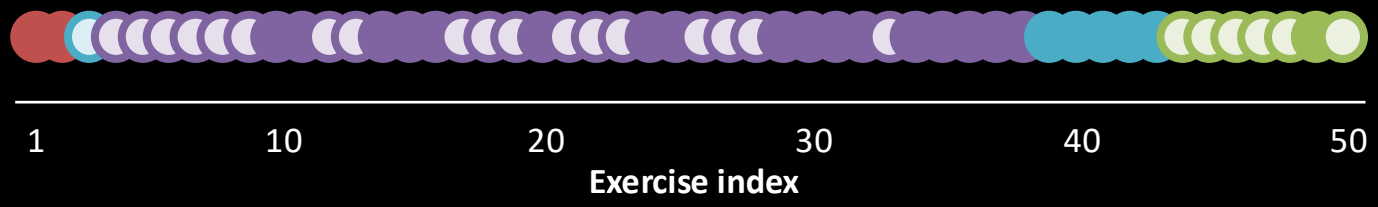
- Solving for x-intercept
- Solving for y-intercept
- Graphing linear equations
- Square roots
- Slope of a line

Answer:

- Correct
- Incorrect

KHAN ACADEMY

Story of Riley



Exercise Type:

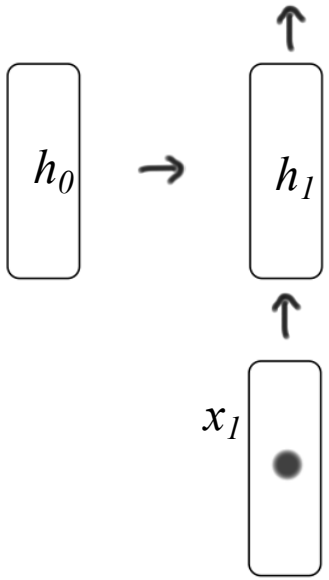
- Solving for x-intercept
- Solving for y-intercept
- Graphing linear equations
- Square roots
- Slope of a line

Answer:

- Correct
- Incorrect

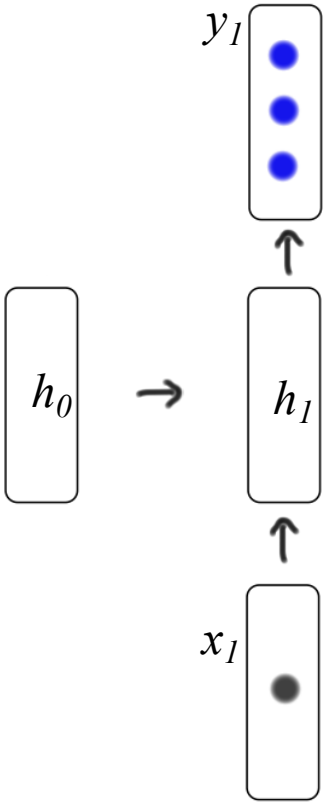
KHAN ACADEMY

Deep Knowledge Tracing



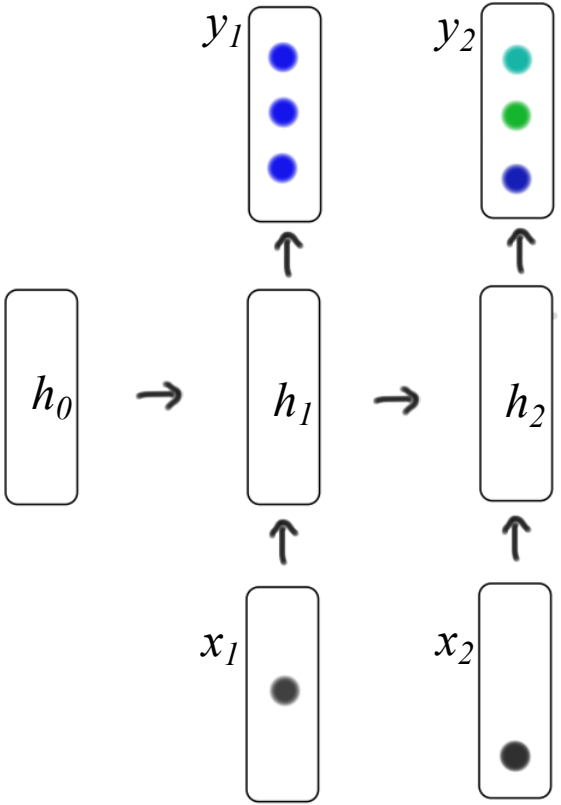
Piech

Deep Knowledge Tracing



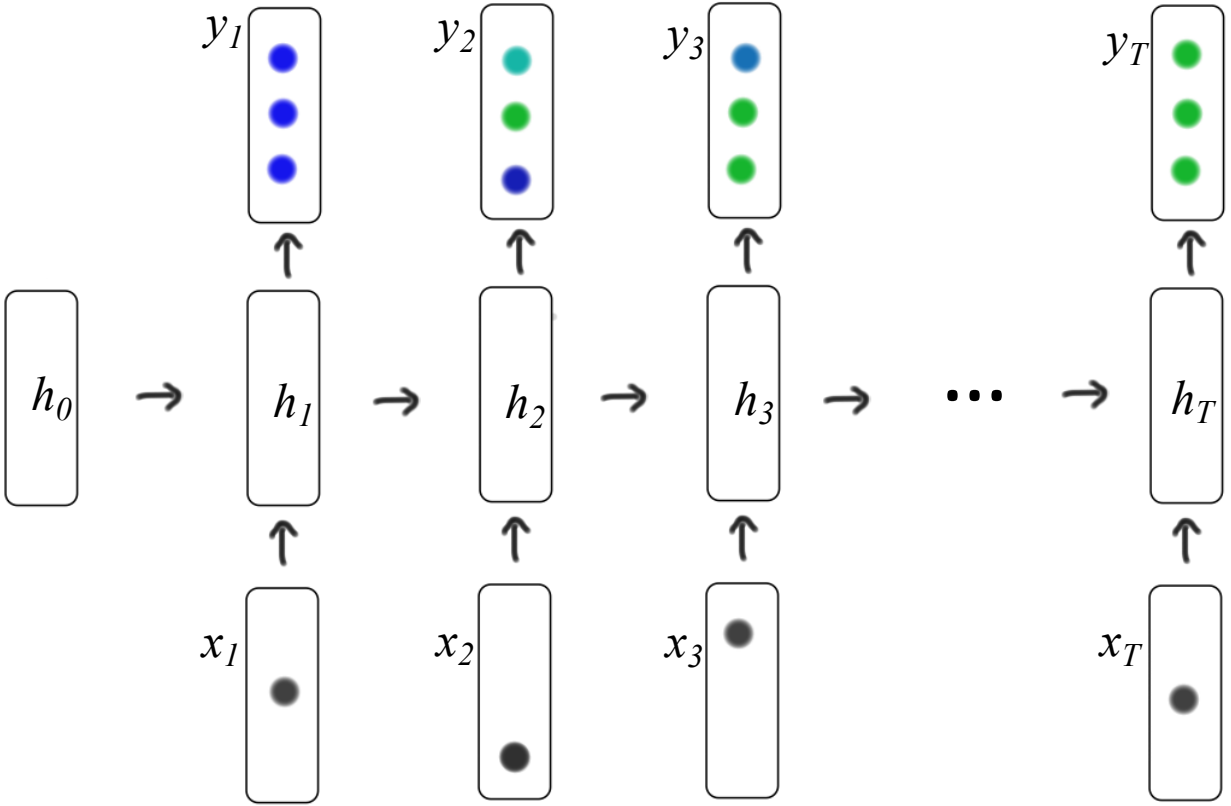
Piech

Deep Knowledge Tracing



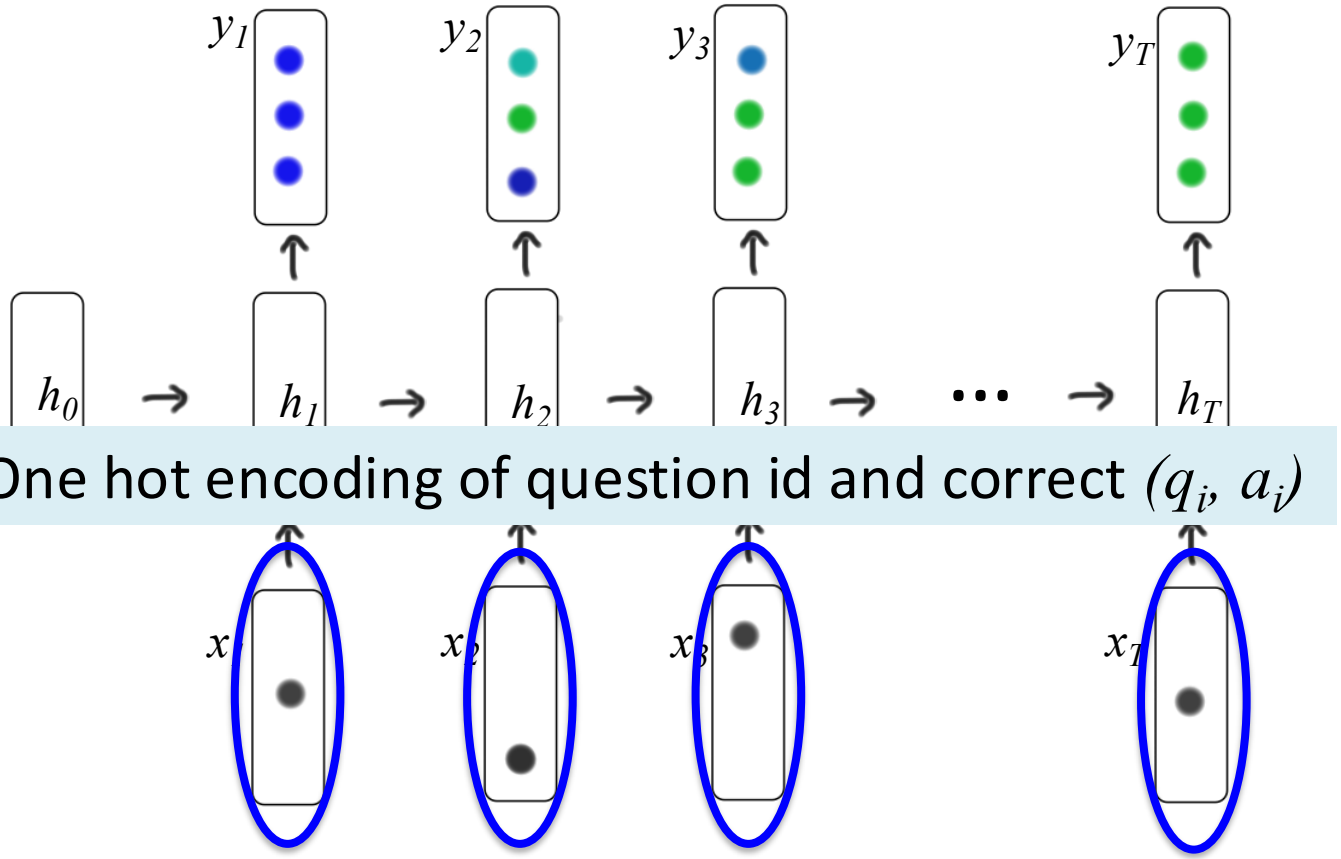
Piech

Deep Knowledge Tracing



Piech

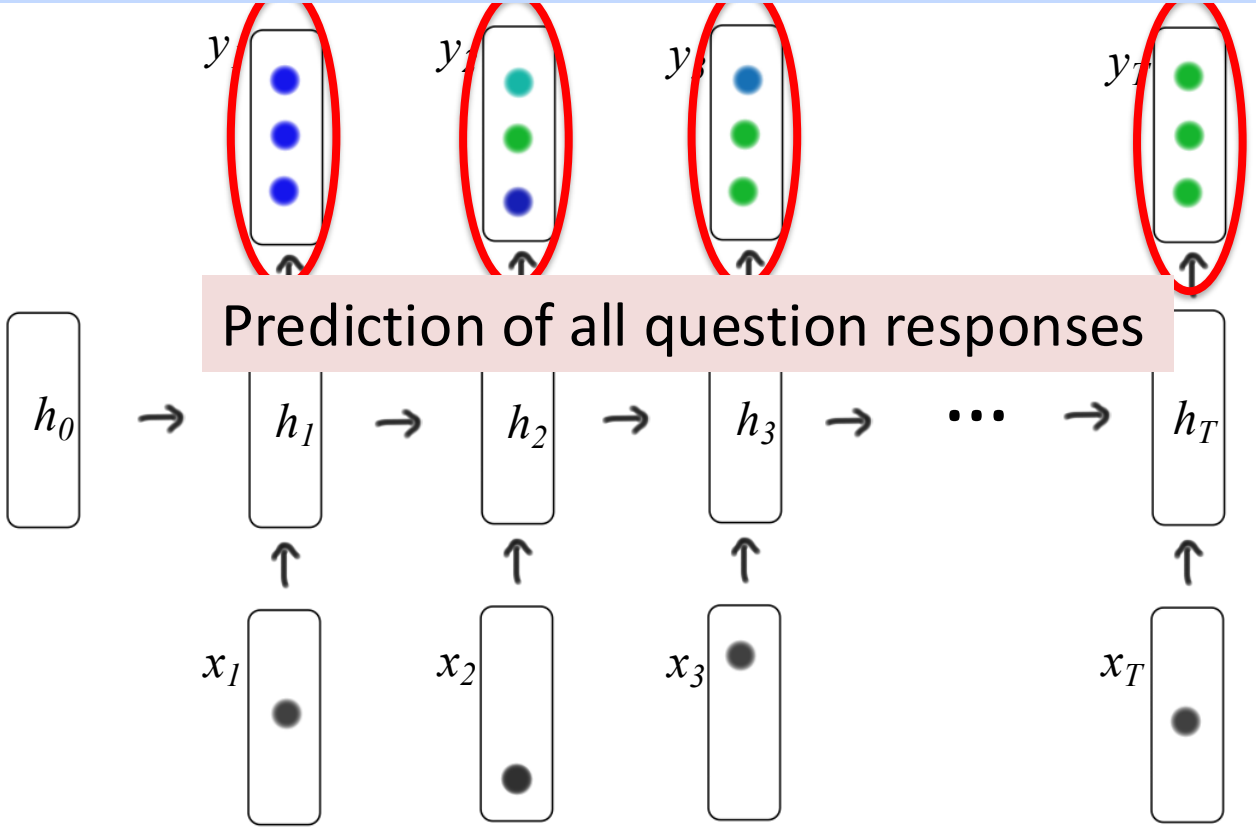
Deep Knowledge Tracing



One hot encoding of question id and correct (q_i, a_i)

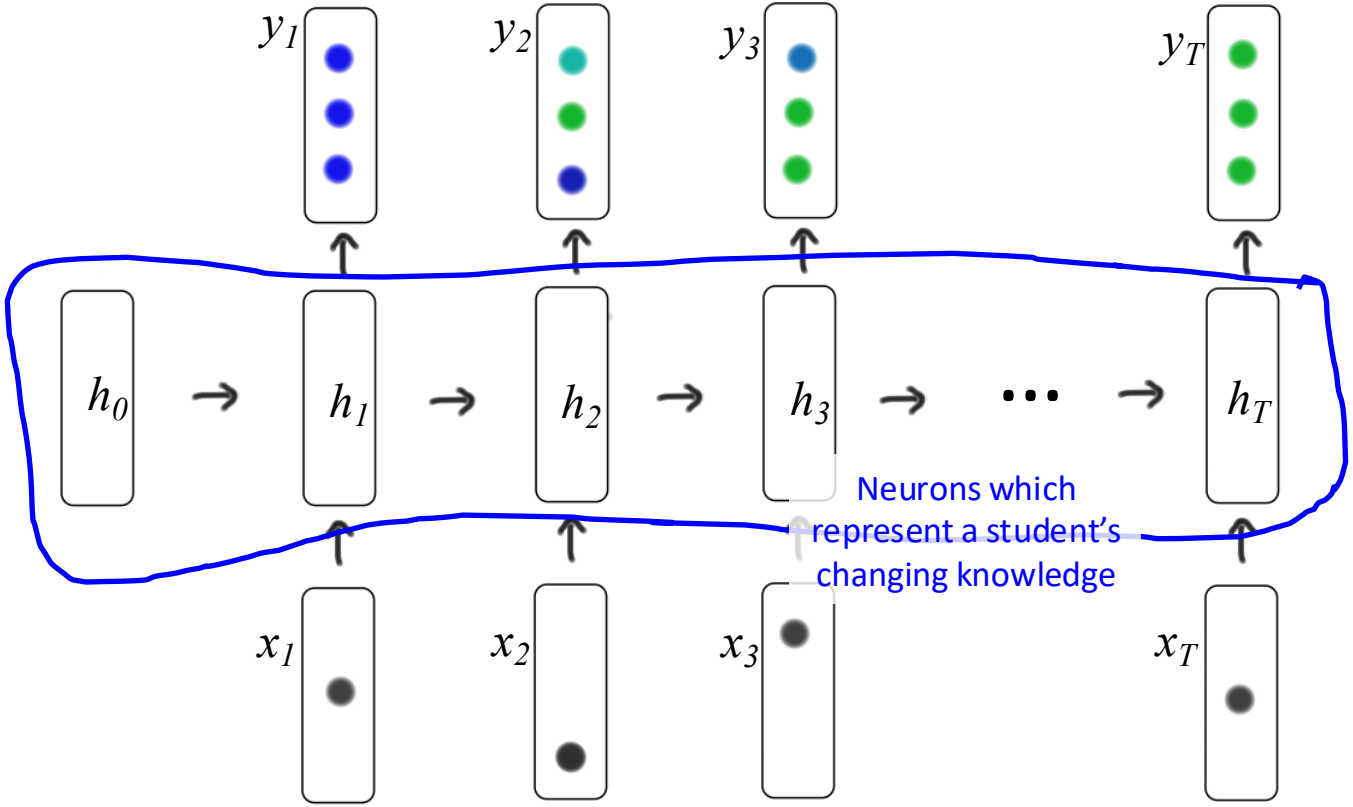
Piech

Deep Knowledge Tracing



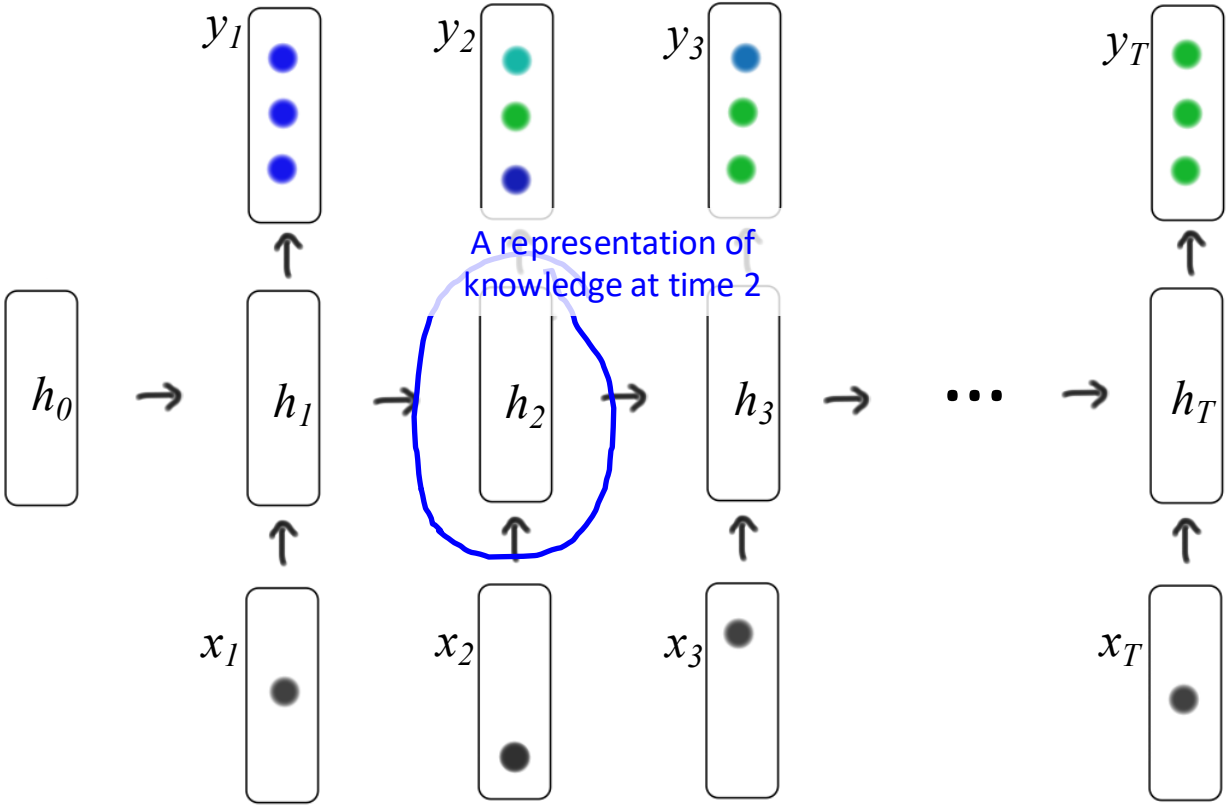
Piech

Deep Knowledge Tracing



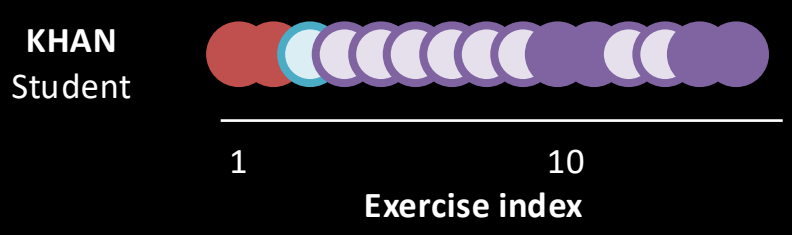
Piech

Deep Knowledge Tracing



Piech

Deep Knowledge Tracing

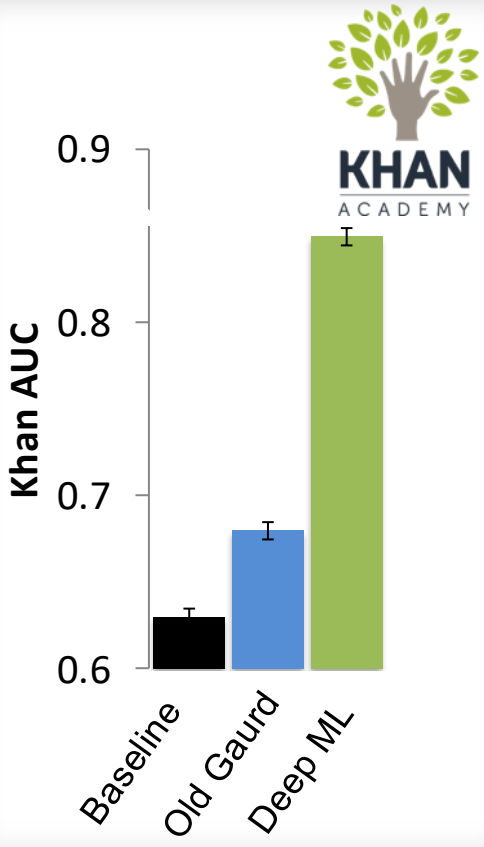


Exercise Type:

- Solving for x-intercept
- Solving for y-intercept
- Graphing linear equations
- Square roots
- Slope of a line

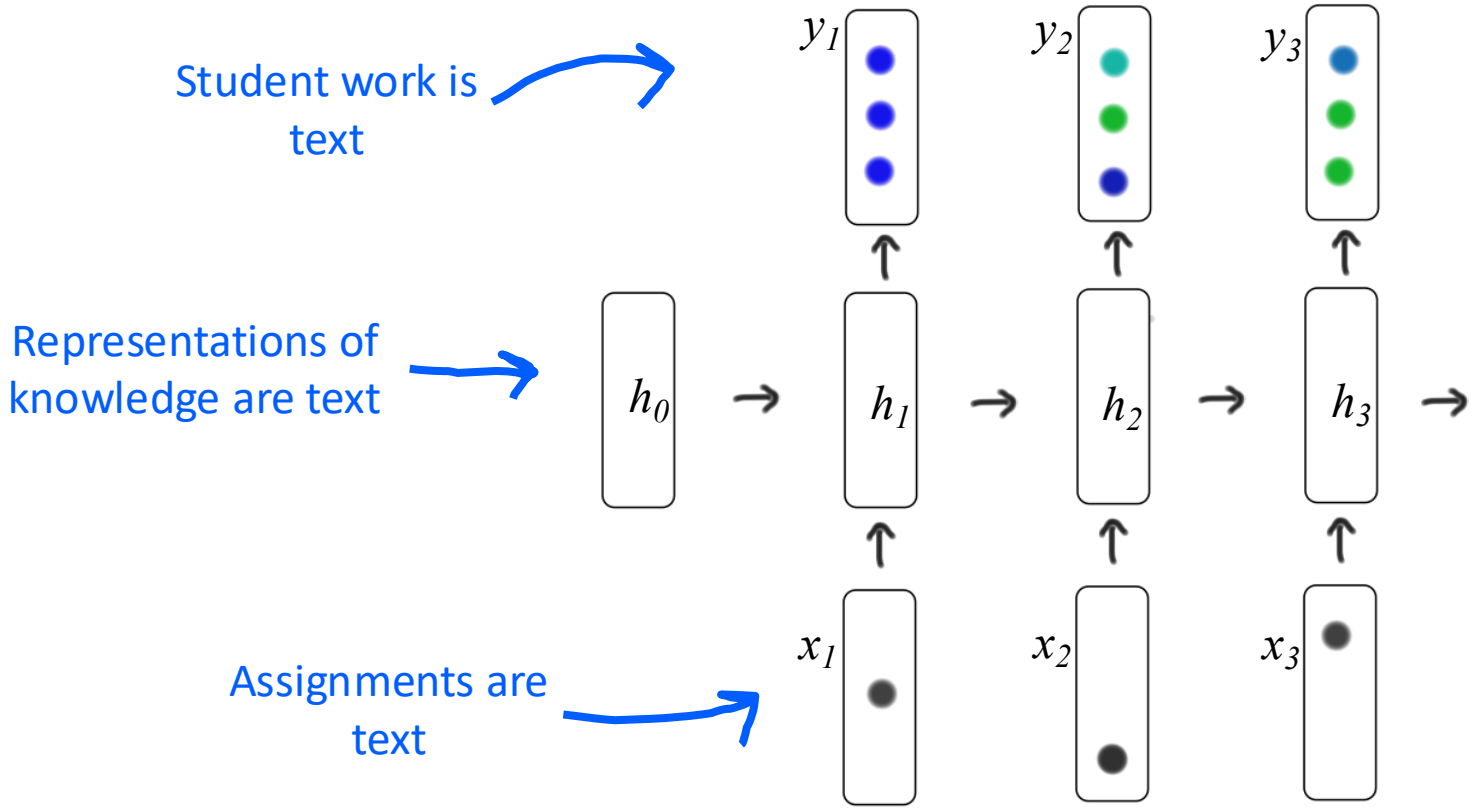
Answer:

- Correct
- Incorrect

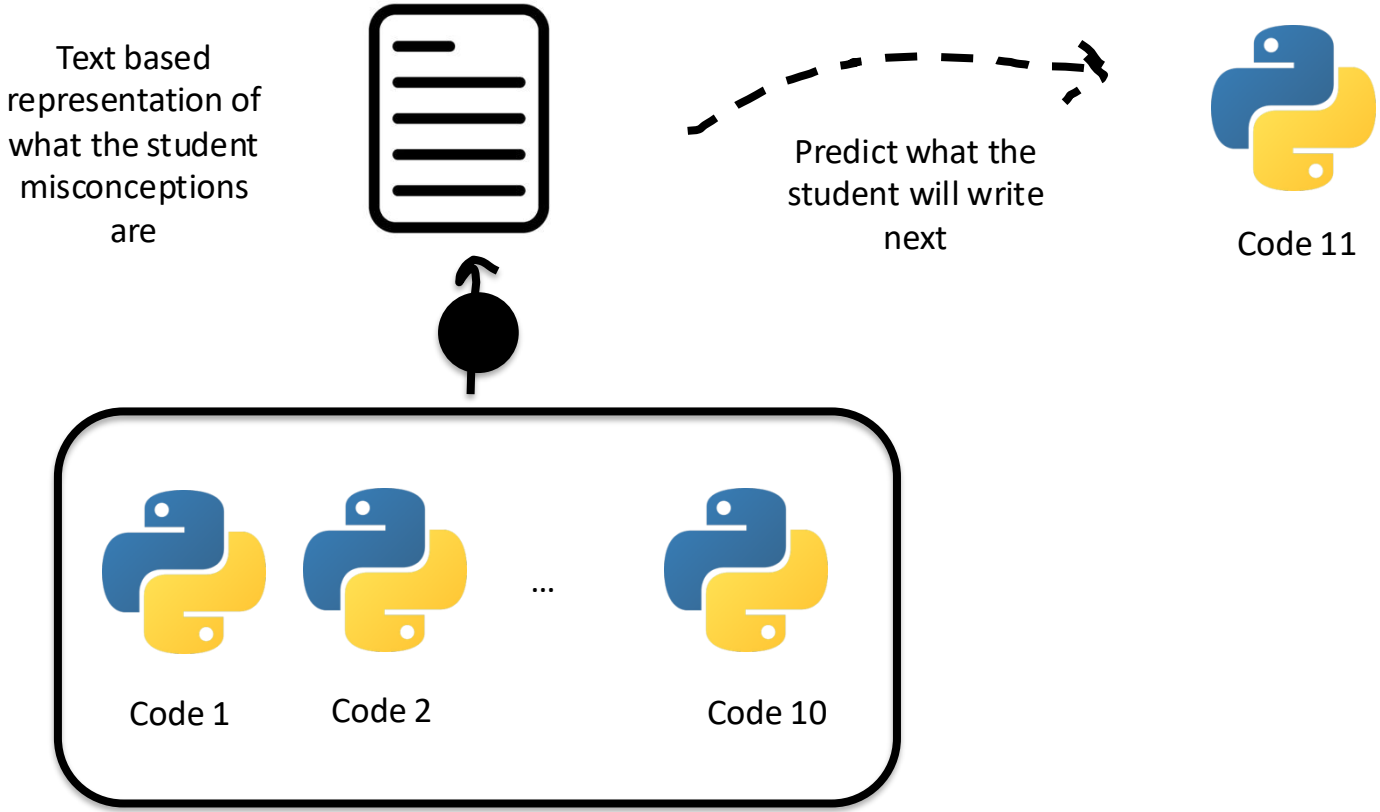


Powers DuoLingo Bird Brain (as of 2023)

Insight Tracing



Insight Tracing



Example of Insight Tracing

Student **before** TeachNow session

```
3   def main():
7       num2 = random.randint(10, 99)
8       print(f"What is {num1} + {num2}?")
9       answer = int(input("Your answer: "))
10      if answer = num1 + num2:
11          print("Correct!")
12      else:
13          print("Incorrect.")
14  if __name__ == '__main__':
15      main()
```

Auto Generated Insight:

"The learner misunderstands equality vs assignment with variables"

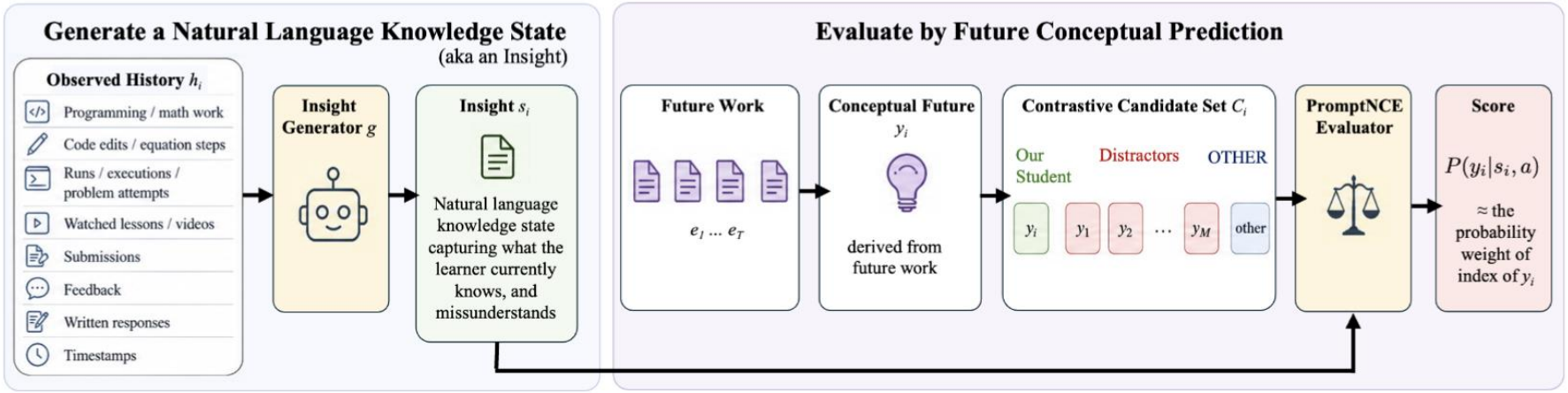
Student **after** TeachNow session

```
3   def main():
        {num2}?")
9       answer = int(input("Your
        answer: "))
10      if answer == num1 + num2:
11          print("Correct!")
12      else:
13          print("Incorrect.")
14  if __name__ == '__main__':
15      main()
```

Auto Generated Insight:

"The learner correctly fixed an = to be an == but may have a lingering misconception with assignment"

Overview


Generate a Natural Language Knowledge State
(aka an Insight)
Observed History h_i
Programming / math work
Code edits / equation steps
Runs / executions / problem attempts
Watched lessons / videos
Submissions
Feedback
Written responses
Timestamps
Insight Generator g
Insight s_i
Natural language knowledge state capturing what the learner currently knows, and missunderstands
Evaluate by Future Conceptual Prediction
Future Work
$e_1 \dots e_T$
Conceptual Future
y_i
derived from future work
Contrastive Candidate Set C_i
Our Student
Distractors
OTHER
y_i
y_1
y_2
$\cdots$
y_M
other
PromptNCE Evaluator
Score
$P(y_i|s_i, a)$
$\approx$ the probability weight of index of y_i

PromptNCE: Conditional Probabilities and PMI Using Only LLMs and Contrastive Estimation Prompts

Juliette Woodrow
Department of Computer Science
Stanford University
jwoodrow@cs.stanford.edu

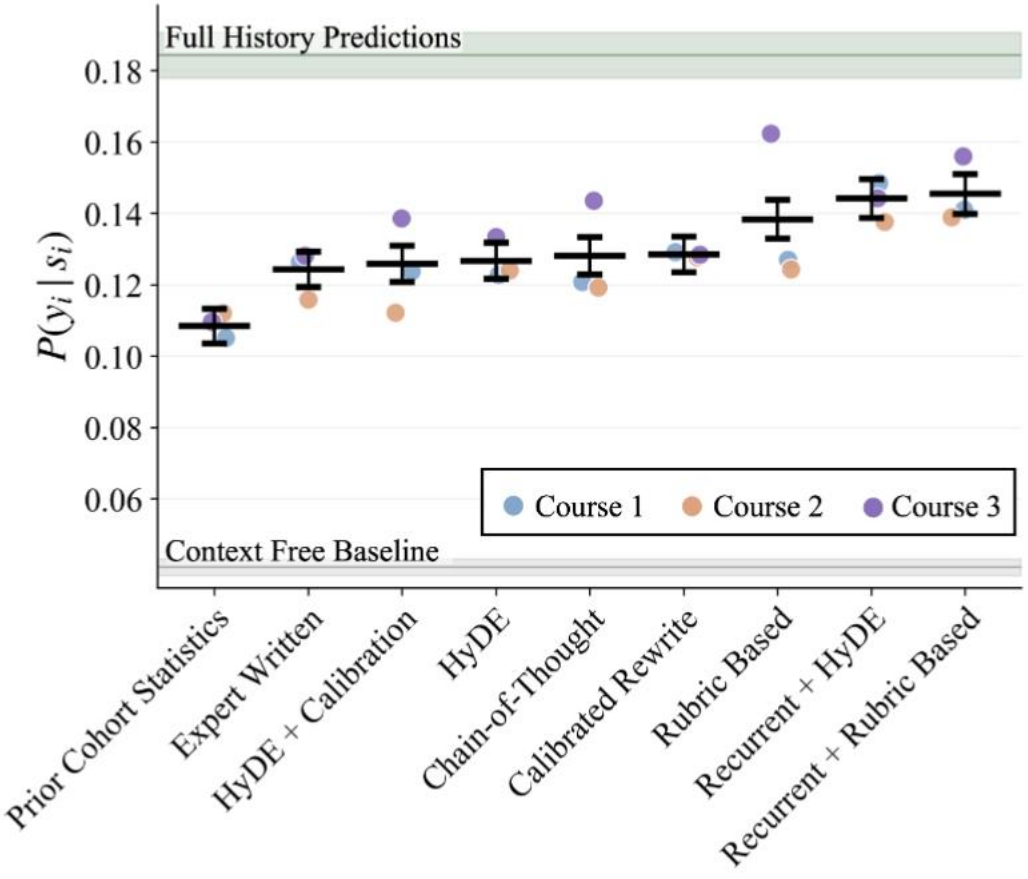
Chris Piech
Department of Computer Science
Stanford University
piech@cs.stanford.edu

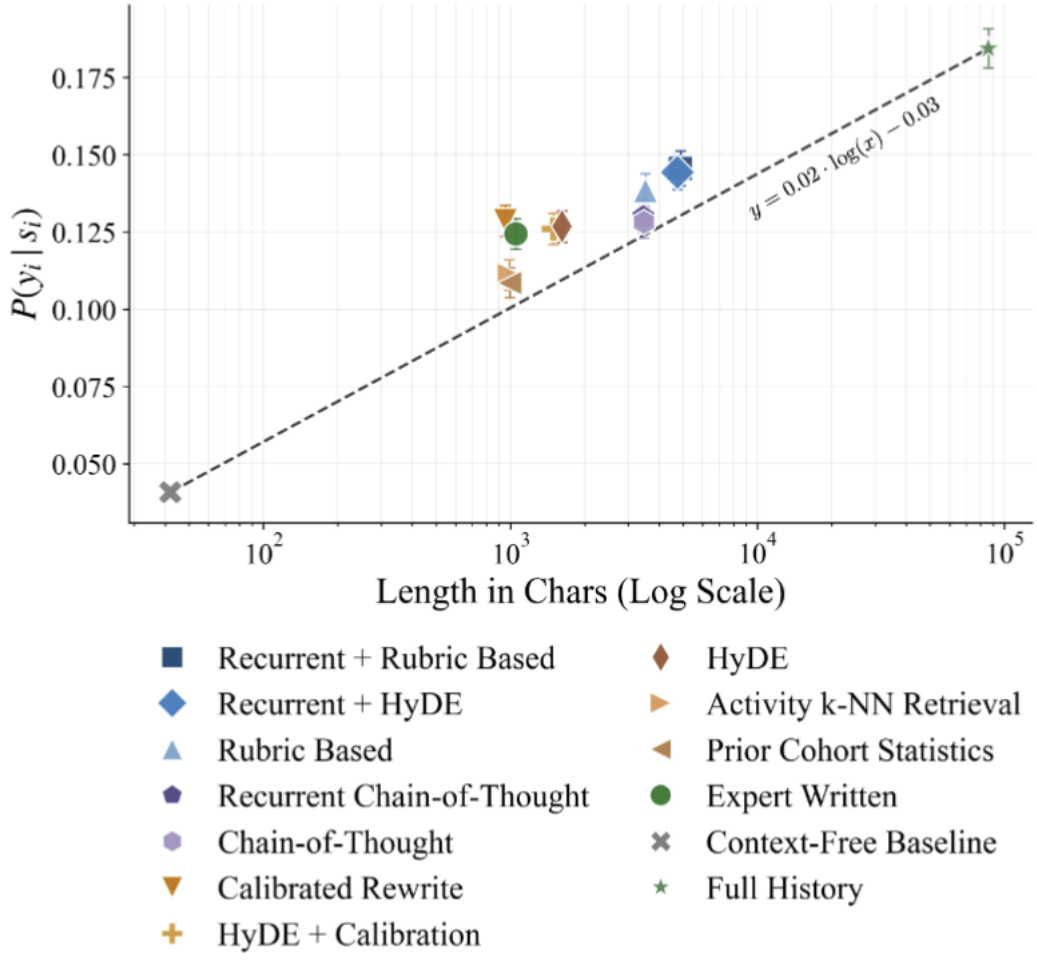
Abstract

Estimating mutual information from text usually requires training a task-specific critic, which limits its use in low-data settings. We ask whether large language models can instead estimate pointwise mutual information zero-shot, using only prompts and elicited probabilities. We construct a benchmark from three publicly available human-annotated datasets with ground-truth PMI, and evaluate five information-theoretic prompting-based estimators. Our main method, PROMPTNCE, frames conditional probability estimation as a contrastive task and augments the candidate set with an explicit OTHER category. The OTHER category allows the model to assign probability mass outside the candidate set, avoiding the closed-set normalization of standard contrastive prompts. PROMPTNCE gives the best conditional probability estimates on all three datasets. For full PMI, we find that estimating label base rates is the primary bottleneck on two of the three datasets, with the best methods reaching Spearman correlation up to 0.78. We also present a case study in computer science education showing how these estimators can be used to score student knowledge summaries in a low-data setting. We release our code and prompts.

Insight Tracing: Now We Can…

- ✓ Trace when **misconceptions** started and stopped for students
- ✓ Give **summaries to teachers** before they interact with students.
- ✓ Ask **arbitrary questions** about the student state
- ✓ Measure **impact of teaching** experiences (TeachNow, Section, Lectures)
- ✓ Novel **teacher puzzles** (with principled scoring)
- ✓ **Transfer insights** from one platform to another (including offline)


Full History Predictions
Context Free Baseline
$P(y_i \mid s_i)$
0.18
0.16
0.14
0.12
0.10
0.08
0.06
Course 1
Course 2
Course 3
Prior Cohort Statistics
Expert Written
HyDE + Calibration
HyDE
Chain-of-Thought
Calibrated Rewrite
Rubric Based
Recurrent + HyDE
Recurrent + Rubric Based


0.175
0.150
0.125
0.100
0.075
0.050
$P(y_i \mid s_i)$
$y = 0.02 \cdot \log(x) - 0.03$
10^2
10^3
10^4
10^5
Length in Chars (Log Scale)
Recurrent + Rubric Based
Recurrent + HyDE
Rubric Based
Recurrent Chain-of-Thought
Chain-of-Thought
Calibrated Rewrite
HyDE + Calibration
HyDE
Activity k-NN Retrieval
Prior Cohort Statistics
Expert Written
Context-Free Baseline
Full History

Leaderboard

Rank	Team	Best model	Best accuracy	Submissions
1	MKN	gradient-boosted tree model	66.7%	3
2	AMF	DKT + IRT head (GRU-64)	65.6%	3
3	Big E	DKT	65.2%	1
4	rube_goldberg	irt	64.5%	1
5	peach	Kalman filter	63.5%	3
6	Ehm	IRT	62.6%	1