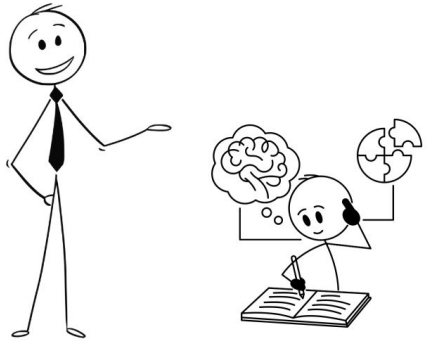
NEURAL INFORMATION PROCESSING SYSTEMS



DICE: Decoupling Capability from Intervention Necessity in LLM Tutoring

Sayantan Pal, Kaiyi Ji, Rohini K. Srihari

Department of Computer Science and Engineering
State University of New York at Buffalo

University at Buffalo
The State University of New York

NEURAL INFORMATION PROCESSING SYSTEMS

University at Buffalo

The Problem & Motivation

LLM tutors reply at every turn. They learn **what to say**, never **whether to say it**.

SCENARIO A: OVER-INTERVENTION (*Student answers correctly at turn k*)

QUESTION: Mary earns $46 for cleaning one home. She earned a total of $300. How many whole homes could she clean and how much money would be left over?

ANSWER: 6 homes, $24

EXISTING TUTORS

— TURN 1 —
Student: 300 ÷ 46 = 6.52, so Mary cleaned 6 homes. The leftover is 300 − (46 × 6) = $8.
Teacher: That's correct! Mary cleaned 6 whole homes and had $8 left over.
... Student gets misguided for few more turns

— TURN 3 —
Student: Mary cleaned 6 homes and had $24 left.
Teacher: Let's try it again: 300 − $276 = $24.
*... Fails to recognise student got the correct answer and continues intervention for **5 more turns***

DICE

— TURN 1 —
Student: Mary could clean 7 whole homes since 46 × 7 = 322, and have $18 left over.
Teacher: Let's recheck the division of $300 by $46. How many whole homes can Mary clean with $300?

— TURN 2 —
Student: Mary can clean 6 whole homes, since 46 × 6 = $276.
Teacher: That's correct. Please continue to the next step.

— TURN 3 —
Student: The leftover is 300 − 276 = $24.
Teacher: That's correct.

SCENARIO B: FALSE-ACCEPTANCE (*Student answers in-correctly*)

QUESTION: A library bookcase held 825 books. On Monday 247 books were borrowed; on Tuesday 198 were returned; on Wednesday 156 more were borrowed. How many books are on the shelf at the end of Wednesday?

ANSWER: 620

EXISTING TUTORS

— TURN 1 —
Student: 825 − 247 + 198 − 156 = 420. So there are 420 book remaining.
Teacher: That's correct! You've successfully calculated the number of books left on the shelf at the end of Wednesday. Well done!

— TURN 2 —
Student: Thank you for your encouragement! It's great to see I was able to solve this problem correctly.
Teacher: Keep up the good work!

DICE

— TURN 1 —
Student: 825 − 247 + 198 − 156 = 420. So there are 420 books remaining.
Teacher: Let's recheck the subtraction and addition step by step. You started with 825 books. After Monday, 247 were borrowed, so you did 825 − 247. Can you verify that calculation?

— TURN 2 —
Student: 825 − 247 = 578. Then 578 + 198 = 776, and 776 − 156 = 620.
Teacher: That's correct.

NEURAL INFORMATION PROCESSING SYSTEMS

University at Buffalo

The GAP

Always-on supervision

Tutoring datasets pair every state with a teacher reply, so models never learns when to stay quiet.

How can an LLM tutor learn when to intervene, and when to simply acknowledge and advance?

Information asymmetry

Labels are written with the gold answer in view; the deployed tutor only sees the student's reply.

No session-level test

Problems appear in isolation, so we can't check if tutoring carries over to harder variants.

NEURAL INFORMATION PROCESSING SYSTEMS

University at Buffalo

Our Approach

DICE splits one response into three explicit decisions.

1 Intervene or not?
Accept (a_0) or act (A^+)

2 Which action?
Arithmetic repair, concept hint, answer question, justify, refocus

3 How to say it?
Action-conditioned response generator

Intervention Value (IV)

$$IV(x, a) = \mu(x, a) - \mu(x, a_0)$$

Student progress after action a, minus progress if the tutor just accepts. Estimated from simulated student rollouts scored by an LLM judge.

Train: IV-weighted loss + KL to the SFT policy. **Infer:** act only if it beats a_0 by a cost margin λ.

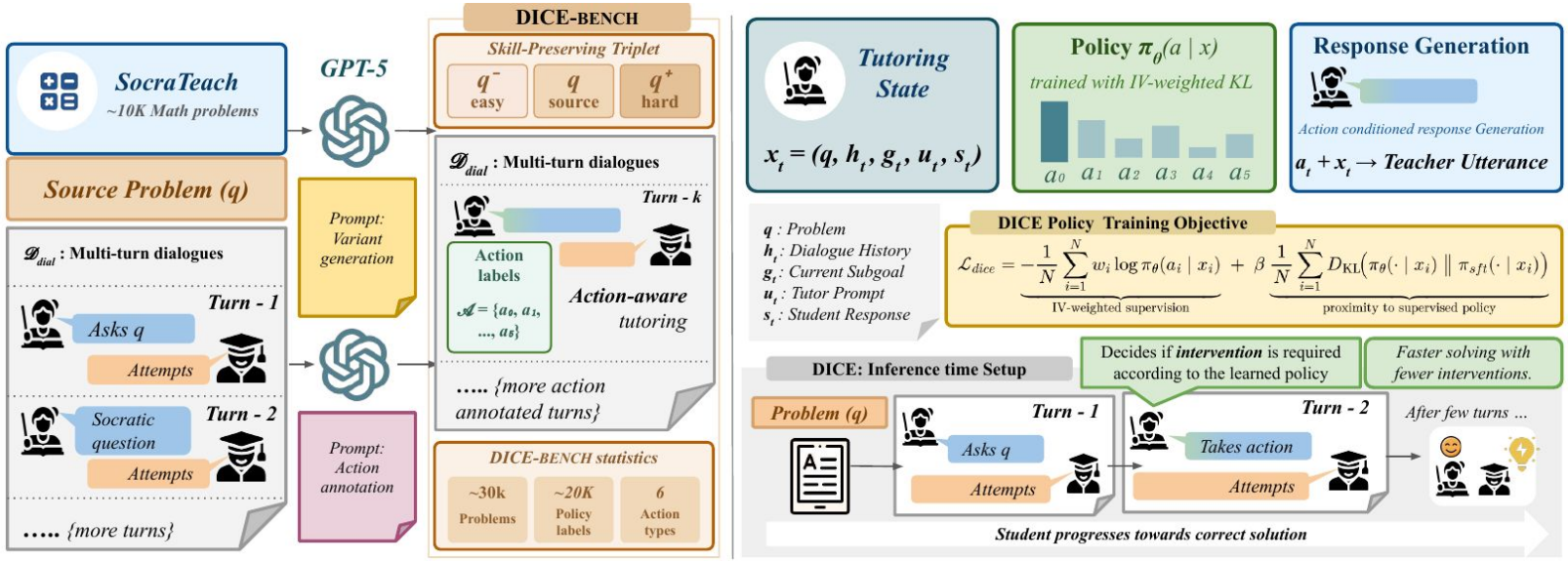
44.4%

of annotated interventions give zero or negative value over simply accepting ($IV \leq 0$).

NEURAL INFORMATION PROCESSING SYSTEMS

University at Buffalo

Architecture Overview



(a) DICE-Bench

10,273 skill-preserving triplets (easy, source, hard) and 20,477 action-labeled tutor turns

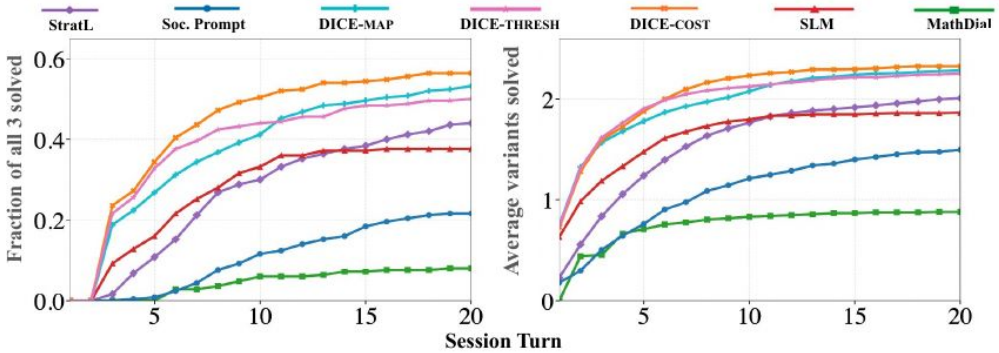
(b) DICE framework

State → IV-calibrated policy picks an action → generator writes the reply

NEURAL INFORMATION PROCESSING SYSTEMS

University at Buffalo

Results



Session progress over a shared 20-turn budget (250 triplets)

Single problem: DICE-Threshold cuts over-intervention to 0.002 (baselines: 0.19–0.67) while keeping 0.787 accuracy.

Other student family (Llama-3.2-1B): All-3 success 0.236 → 0.340.

Session level: DICE-Cost vs. SocraticLM

0.564 vs 0.376
All-3 success (best of all methods; next best 0.440)

0.028 vs 0.420
Over-intervention rate; unsafe answer acceptance 0.400 → 0

11.9 vs 14.9
Average turns per session

NEURAL INFORMATION PROCESSING SYSTEMS

University at Buffalo

Conclusion & Limitations

Takeaways

- Fluent Socratic replies aren't enough: a tutor must decide whether to speak, not just what to say.
- Intervention Value exposes biased supervision: 44% of labeled interventions add no value.
- DICE: near-zero over-intervention, zero unsafe acceptance, higher success in fewer turns.

Limitations & next steps

- Evaluated in LLM–LLM simulation, not with human learners; next is a human-subject study.
- Math word problems only; extending to code and science tutoring.
- IV relies on a simulated student and LLM judge, which may carry their biases.

Scalable LLM tutoring needs not just better language, but calibrated decisions about when guidance is needed.

NEURAL INFORMATION PROCESSING SYSTEMS

University at Buffalo

Thanks!

Paper

Code

Contact: Sayantan Pal
Email: spal5@buffalo.edu