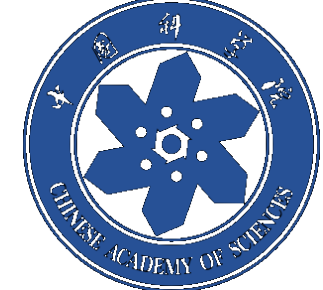


Strong yet Controllable LLM Abstention without Learning to Abstain

Junwei Zhang* Yifei Wang* Wanli Yang Hexiang Tan Rui Tang

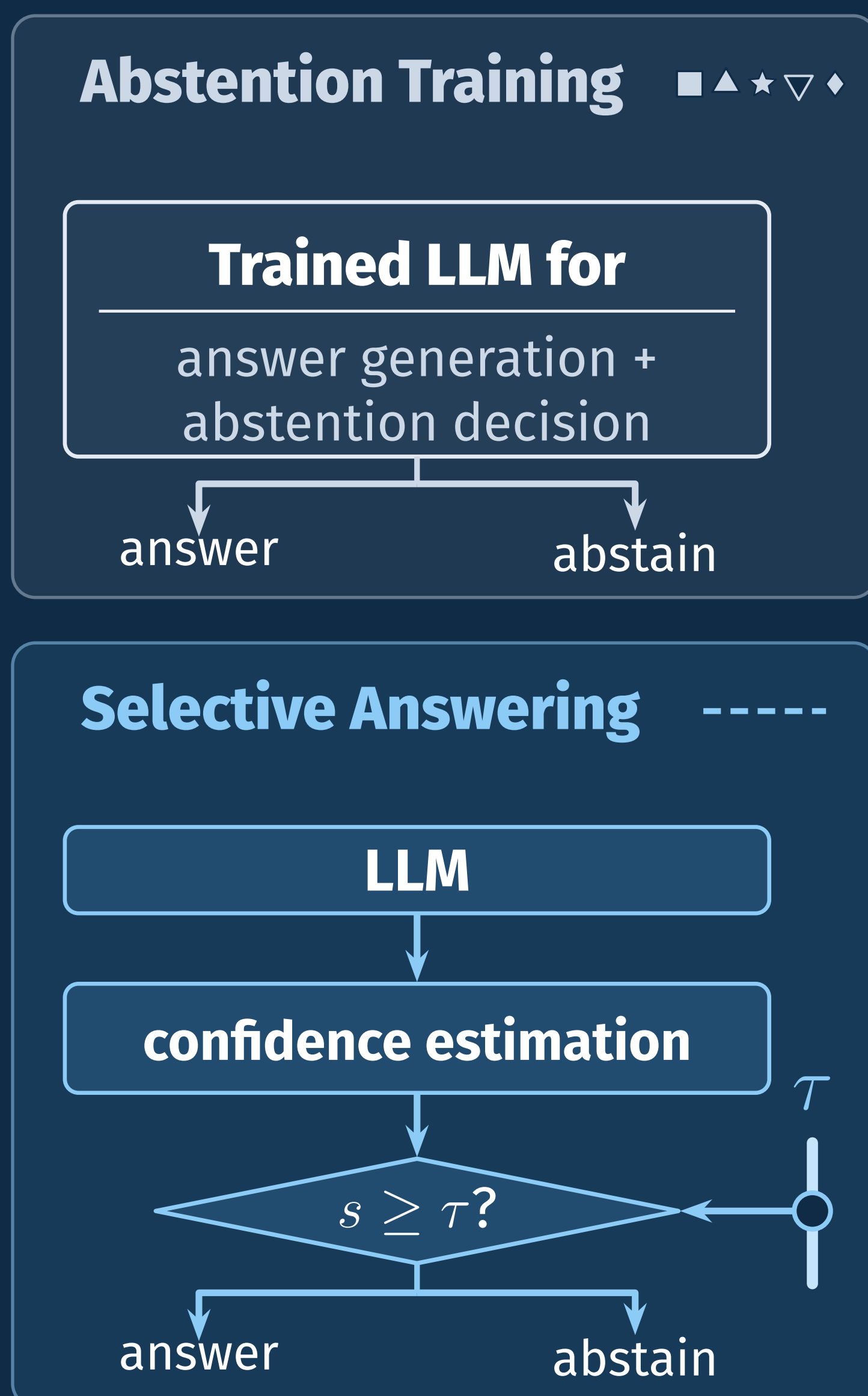
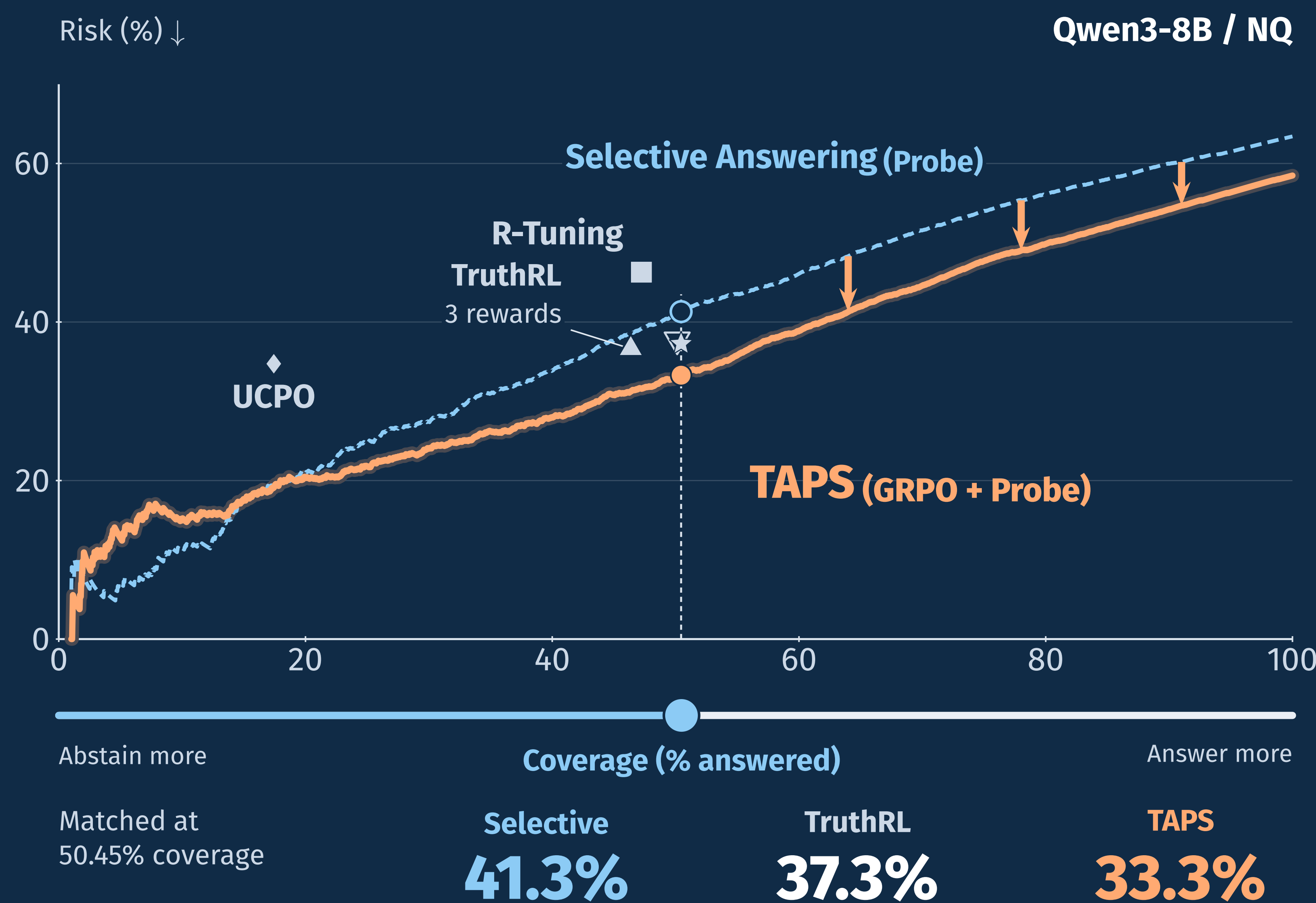
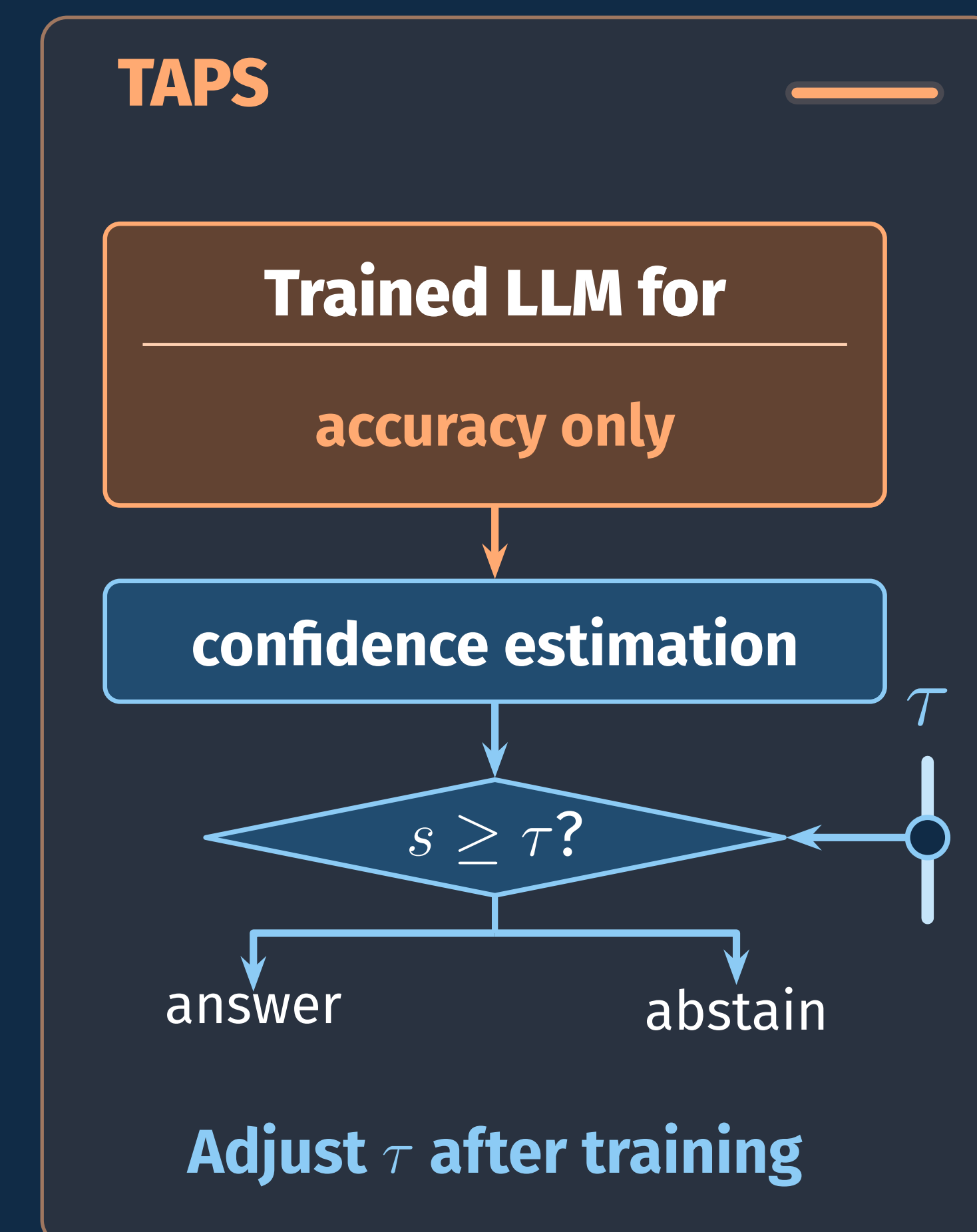
Xin Chen Jingang Wang Du Su Fei Sun

zhangjunwei233@mails.ucas.ac.cn sunfei@ict.ac.cn * Equal contribution



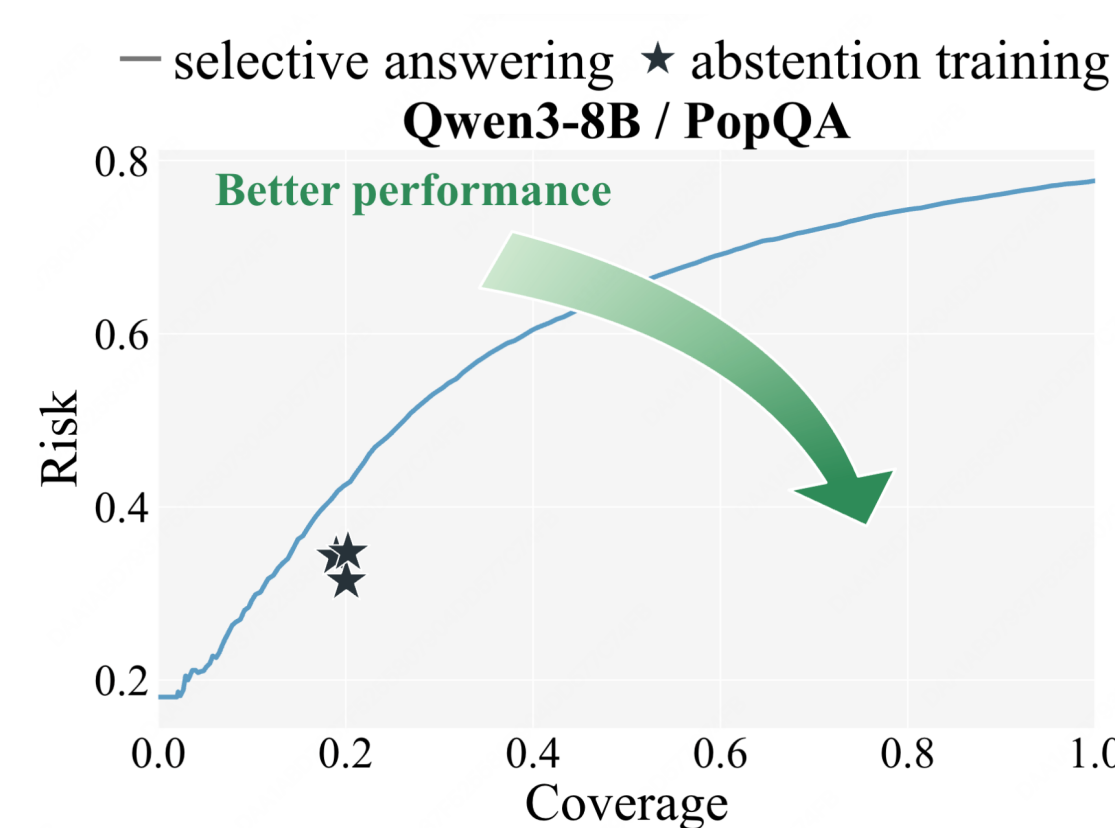
Just Select on a Good Answerer

A Strong Baseline for LLM Abstention



The Cost of Abstention Training

The **standard evaluation** compares **risk at matched coverage**.



Coverage

Fraction of questions answered.

Risk

Error rate among those answers.

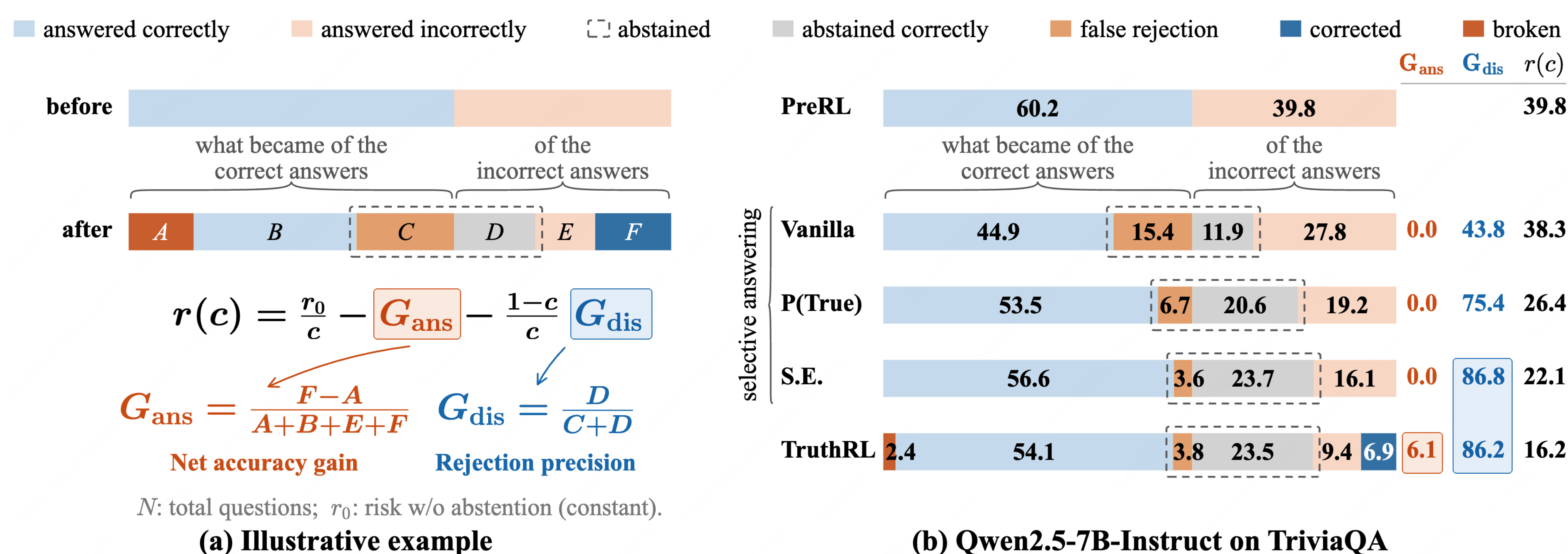
Improves performance

TruthRL lowers risk at matched coverage. ✓

Sacrifices controllability

Within runs: yield **only one point**.
Across runs: **collapse** toward a narrow range. ✗

What It Actually Buys?



- Risk can be **decomposed** into answering and rejection quality.
- Against strong selection, the remaining advantage comes mainly from **answering better**, rather than rejecting better.

Selective Answering on a Good Answerer Achieves Both!

TAPS: Train for Accuracy, then Post-hoc Select

Stage 1: Train for accuracy

Any post-training method; e.g., GRPO with a binary outcome reward.

Stage 2: Post-hoc select

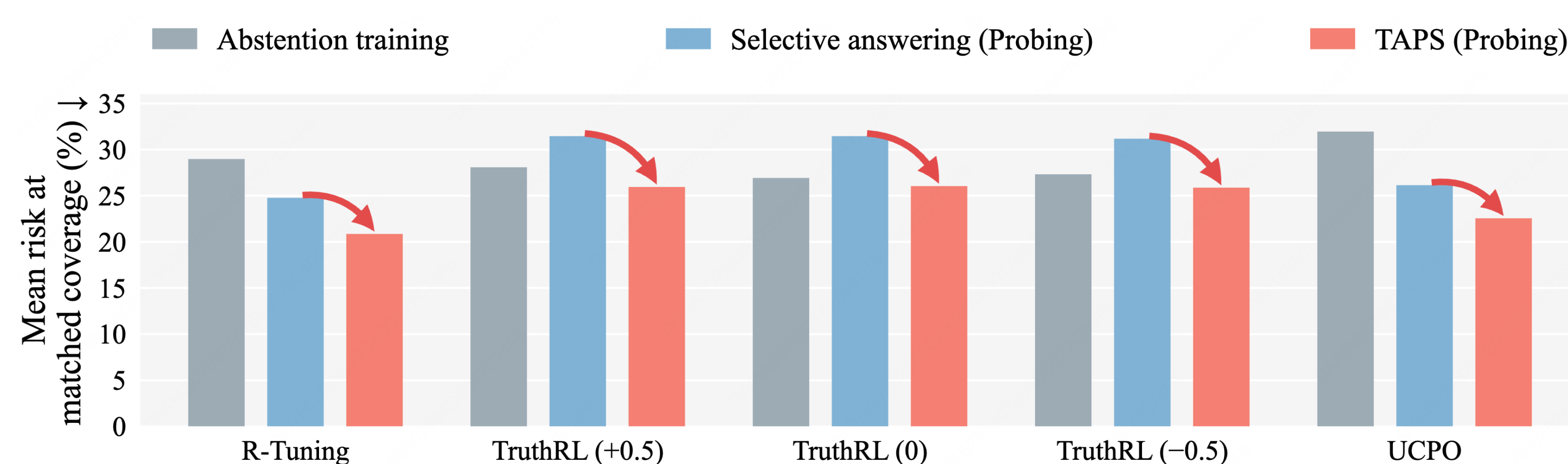
Any off-the-shelf confidence estimator; e.g., semantic entropy or a linear probe.

Across 3 models, 3 datasets and 5 baseline configurations:

Accuracy training cuts mean risk by **4.75 pp**.

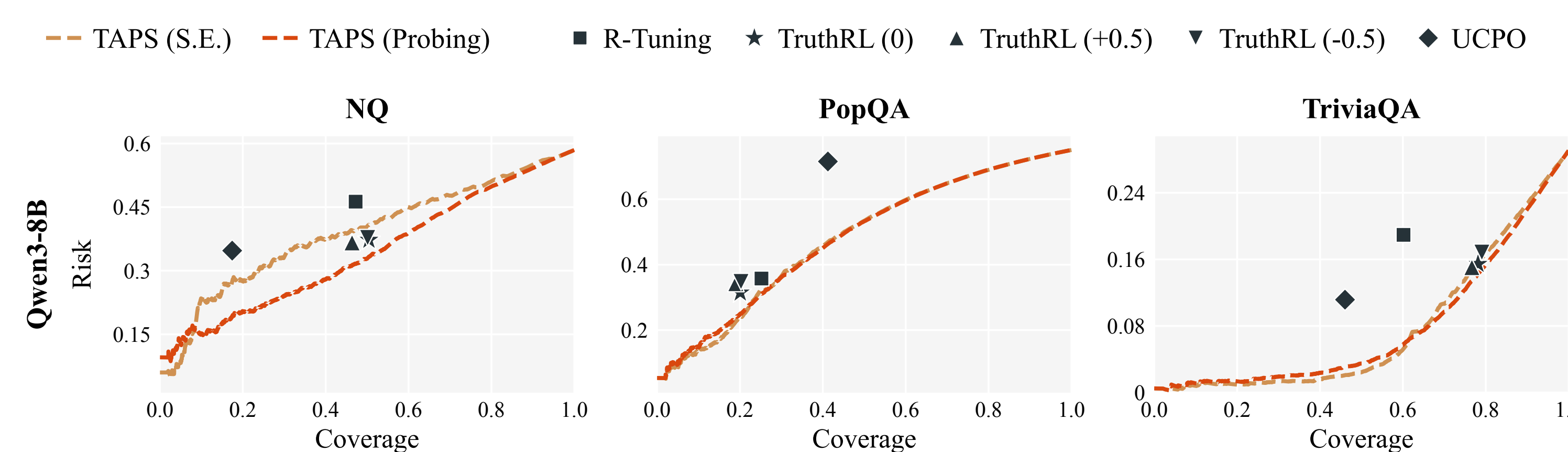
Performance ✓

TAPS achieves **0.9–9.4 pp** lower risk than abstention training.



TAPS can span **Full** coverage levels.

Controllability ✓



Key Insights

Lower risk does not necessarily imply better abstention

- Risk at matched coverage mixes answering and abstaining.

A higher bar for abstention training

- Internalizing abstention should add value beyond selecting on a post-trained answerer.