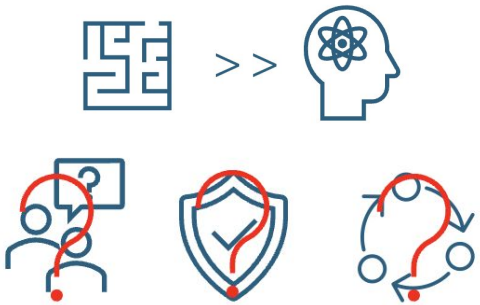
Towards Scalable Oversight via Partitioned Human Supervision

Ren Yin[1,2], Takashi Ishida[2,1], Masashi Sugiyama[2,1]

The University of Tokyo[1], RIKEN[2]

Motivation



Frontier AI can solve tasks beyond any single expert's ability

- Standard pipelines such as RLHF and RLVR assume humans can provide: **correct labels** or **verifiers**
- However, for cross-domain / highly technical tasks: **No human can fully verify**

Example Intuition

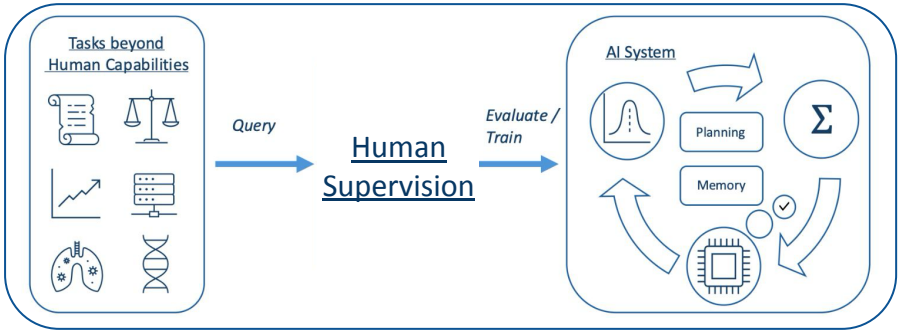
No single expert sees the full picture, but each can provide reliable partial feedback.

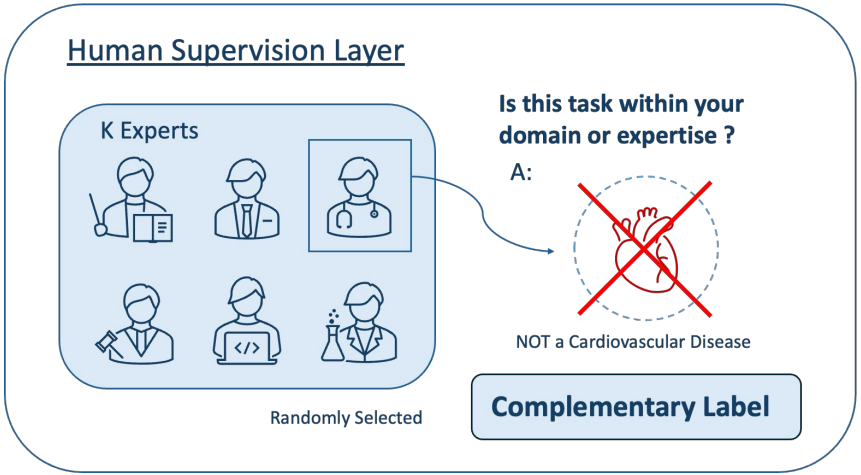
Key Idea: Complementary Label

The Concept of Complementary Label[1]: a class that is definitely incorrect

For a given class k:

- ❑ Ordinary Label: k is correct
- ❑ Complementary Label: k is NOT correct





[1] Ishida, T., Niu, G., Hu, W., & Sugiyama, M. (2017). Learning from complementary labels. Advances in neural information processing systems, 30.

Unbiased Estimator

Let $K \geq 3$ and let the input domain be $\mathcal{X}$. Let $(X, Y) \sim \mathcal{D}$ be drawn from an unknown joint distribution over $\mathcal{X} \times \{1, \ldots, K\}$, where Y is the true class label. An AI system (e.g., LLM) produces a top-1 prediction $\hat{Y} = f(X) \in \{1, \ldots, K\}$, and its top-1 accuracy is $A = \Pr_{(X,Y)\sim\mathcal{D}}(\hat{Y} = Y)$.

Instead of observing Y, we observe a complementary label $\bar{Y} \in \{1, \ldots, K\} \setminus \{Y\}$ drawn uniformly at random conditional on Y:

$$\Pr(\bar{Y} = k \mid Y) = \frac{1}{K-1}, \quad \forall k \neq Y.$$

Given n_c i.i.d. complementary-labeled samples $\{(x_i, \bar{y}_i)\}_{i=1}^{n_c}$, define

$$\hat{q} = \frac{1}{n_c} \sum_{i=1}^{n_c} \mathbf{1}[f(x_i) \neq \bar{y}_i] .$$

An estimator of the top-1 accuracy without observing ground-truth is given as:

$$\hat{A}_{\text{comp}} = (K-1)\hat{q} - (K-2).$$

Sample Efficiency and Estimator in Practice

We have n_{o} ordinary-labeled items (with ground-truth Y) and n_{c} complementary-labeled items (with $\bar{Y} \neq Y$). Complementary labels are weaker than ordinary labels. To match the same reliability in terms of variance:

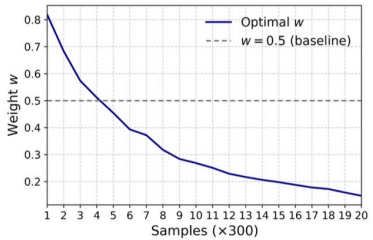
$$n_{\mathrm{c}} = \left(1 + \frac{K-2}{A}\right) n_{\mathrm{o}}.$$

In practice, we often have a small n_{o} and a large n_{c}. We can combine them to get a more stable estimate:

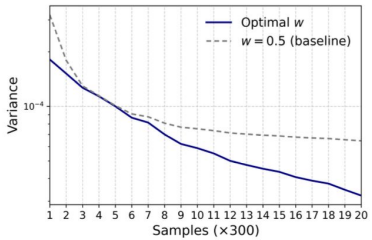
- **Inverse Variance Weighted Estimator (IVW)**
- **Closed-Form Maximum-Likelihood Estimator (ML)**

Main Experiment

Estimator	MMLU-Pro	MedQA-USMLE	GPQA	MATH†	MATH(CoT)‡	Average
Ord	78.33 ± 1.73 (±2.38)	92.89 ± 1.35 (±1.48)	64.17 ± 1.67 (±4.39)	47.56 ± 3.91 (±2.88)	84.89 ± 0.77 (±2.07)	73.57
Comp-n_o	77.00 ± 12.49 (±7.95)	**92.67 ± 1.53 (±2.67)**	**59.17 ± 3.82 (±9.42)**	48.44 ± 10.78 (±7.72)	80.44 ± 2.78 (±4.98)	71.54
Comp-Var	75.67 ± 2.15 (±2.51)	90.61 ± 1.43 (±1.69)	63.67 ± 5.01 (±4.28)	41.10 ± 3.17 (±2.93)	81.35 ± 0.29 (±2.28)	70.48
IVW-0.5	77.89 ± 1.58 (±1.80)	91.61 ± 1.11 (±1.15)	65.28 ± 1.34 (±3.33)	43.44 ± 3.95 (±2.52)	83.56 ± 1.17 (±1.58)	72.36
IVW	**77.97 ± 1.58 (±1.79)**	91.86 ± 1.11 (±1.13)	65.14 ± 1.38 (±3.30)	44.87 ± 3.82 (±2.37)	**83.86 ± 0.83 (±1.56)**	72.74
ML	77.94 ± 1.58 (±1.79)	91.65 ± 1.08 (±1.18)	65.11 ± 1.38 (±3.28)	**44.75 ± 3.79 (±2.36)**	83.65 ± 1.04 (±1.59)	72.62
Ord-Eval	77.97	92.66	59.52	44.21	83.89	–



(a) Optimal w vs. number of samples.

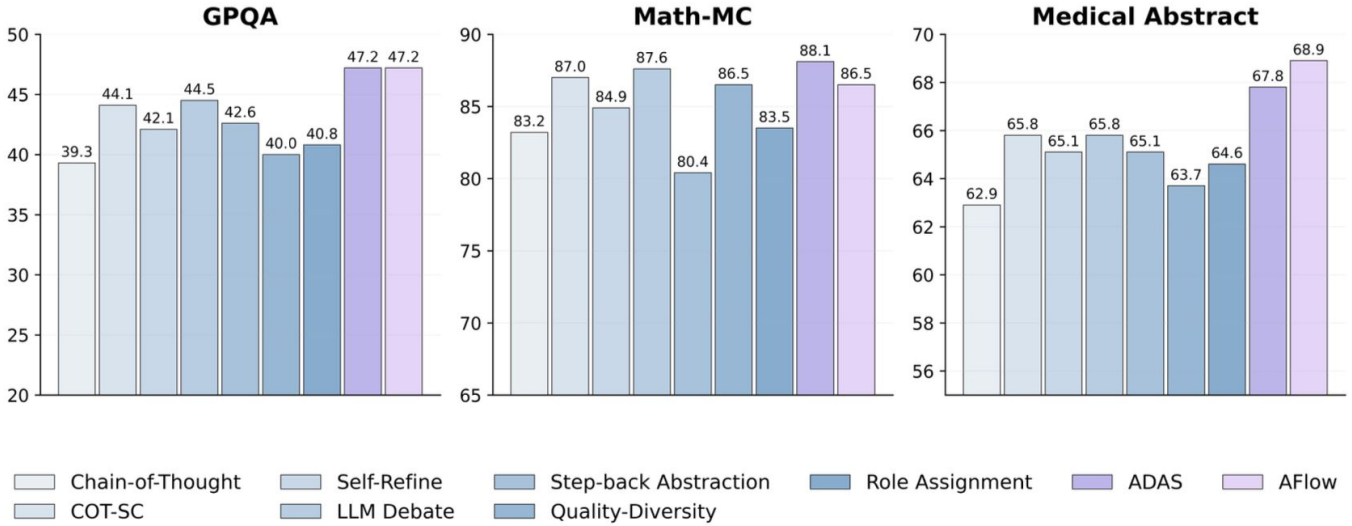


(b) Corresponding variance on a log scale.

Optimal weighting consistently achieves lower variance compared to the equal-weighted baseline, demonstrating more reliable estimation with limited samples.

Agentic Training with Partial Feedback

Accuracy estimated from complementary labels can be used as a training signal. It improves agentic systems such as ADAS[1] and AFlow[2].



[1] Hu, S., Lu, C., & Clune, J. (2025). Automated design of agentic systems. International Conference on Learning Representations.

[2] Zhang, J., Xiang, J., Yu, Z., Teng, F., Chen, X.-H., Chen, J., Zhuge, M., Cheng, X., Hong, S., Wang, J., Zheng, B., Liu, B., Luo, Y., & Wu, C. (2025). AFlow: Automating agentic workflow generation. International Conference on Learning Representations.

Towards Scalable Oversight via Partitioned Human Supervision

Thanks for Listening