RLxF: Reinforcement Learning from World Feedback · ICML 2026 Workshop

An Adoption-Aware Crop Recommender from Farmer World Feedback

Vairaaj Bindal
Lynbrook High School, San Jose, California

Oral presentation · Seoul · July 10, 2026

The problem

Extension services recommend the highest-yield plan.

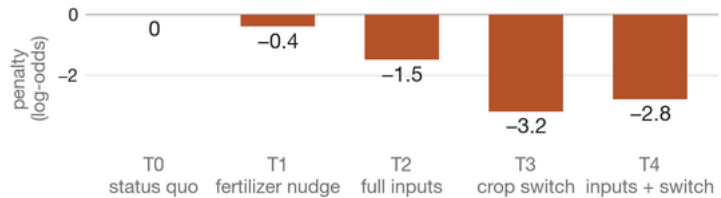
For a smallholder, that plan is often the most expensive, the riskiest, and the biggest change to how they farm.

So it goes unadopted, and unadopted advice adds nothing to the harvest.

The idea

Recommend what will actually happen:

$$\mathrm{ERB}(a \mid x) = \underbrace{\mathbb{P}(\text{adopt } a \mid x)}_{\text{will they do it?}} \cdot \underbrace{\left(\mathbb{E}[Y \mid a, x] - \mathbb{E}[Y \mid a_0, x]\right)}_{\text{what is it worth?}}$$



The tier penalties come from Ruzzante et al.'s 367-study meta-analysis, not from our data.

This talk

1 The data it learns from.

2 The result, and what it is worth to a farm.

3 Trying to break it: four ablations, two statistical checks, a placebo, and a field experiment in Kenya.

4 What carries beyond crops.

Where the feedback comes from

8,744 plots across 23 districts of Ethiopia: inputs, adoption decisions, and realized yields, from the World Bank's LSMS-ISA survey.

Nothing was deployed; the loop is learned offline.

+50 **kg/ha**

over the accuracy-only policy, on 800 held-out plots

That's a 3–4% bump on typical yields of about 1.5 t/ha:
roughly $27 per hectare each season,
or two weeks of a family's calories.

95% CI [+24, +74]; the policies differ on 46% of plots.

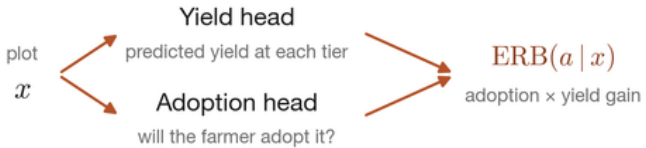
The catch

This is not a randomized experiment. Farmers were not assigned their choices; they chose, based on risk tolerance, knowledge of their own plot, and family constraints, none of which is recorded.

The standard off-policy toolkit needs the odds behind each logged choice, and here those cannot be known.

Every number in this talk compares two policies through the same model. None of it is an identified causal effect.

How it’s built: two heads

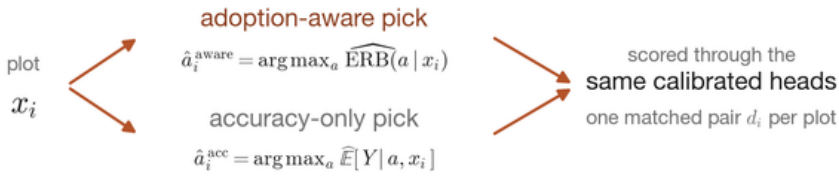


Yield: gradient boosting, plus fixed response coefficients from the agronomy literature.

Adoption: logistic propensity, plus the tier penalties above, recalibrated by education.

Without the fixed coefficients, counterfactual predictions collapse to zero.

How it’s scored: matched pairs

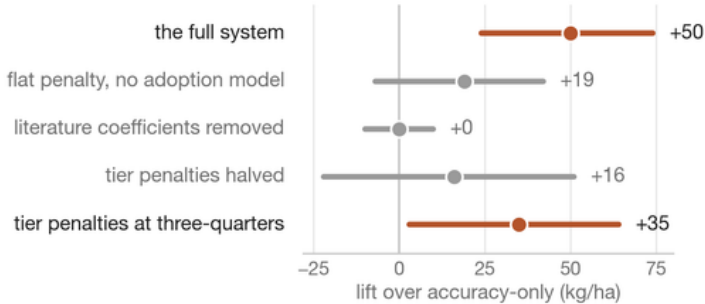


$$d_i = \Delta\text{ERB}(\hat{a}_i^{\text{aware}} \mid x_i) - \Delta\text{ERB}(\hat{a}_i^{\text{acc}} \mid x_i)$$

The same model scores both policies,
so its errors largely cancel in the difference.
800 held-out plots in 44 villages the model never saw.

Remove a piece, rerun everything

Gray means the lift is gone.



Every piece earns its keep,
and the borrowed penalties must be right to within a factor of two.

Statistical checks

Hidden confounding. An unmeasured confounder would need to shift the odds behind farmers' choices eight-fold to flip the sign (Rosenbaum $\Gamma^{*} \geq 8$; agricultural economics calls 1.5 strong).

Chance. Permutation testing: $p < 0.0005$, at both the plot level and the village level.

The placebo

The lift is biggest in the no-education group. Targeting?

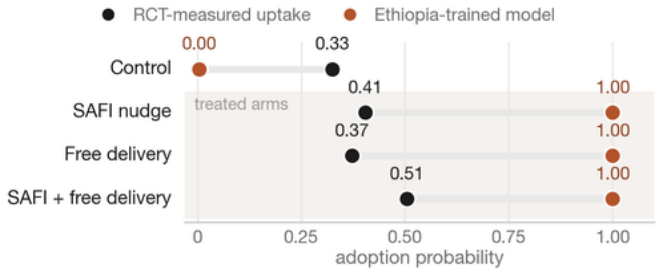
Test: shuffle the labels 2,000 times; does the real pattern stand out?

Stratum	n^{dis}	Placebo p
No education	257	0.14
Early primary	17	0.86
Late primary	47	0.87
Secondar y	25	0.50

It does not stand out. The no-education group is simply 68% of the test set; there is no targeting signal.

An independent experiment

Everything so far is internal to the model. The Kenya SAFI trial measured real adoption; we score our Ethiopia-trained model on its four arms.

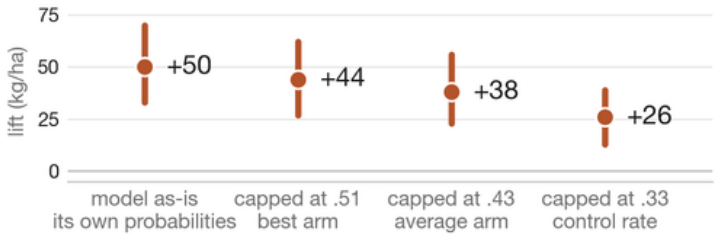


It gets the ordering right: control lowest, every treated arm above.

But it calls treated-arm adoption cer tain (1.00); measured: 37–51%.

How wrong could the +50 be?

The overconfidence caps how far to trust the result. So replace the model's certainty with the experiment's measured rates, generous to harsh, and rerun.



Even the harshest replacement leaves +26 kg/ha: wherever the truth is, the lift sits between 26 and 44.

Takeaways

1 Numbers you cannot estimate can be borrowed: the tier penalties and the Kenya cap both came from other people's experiments.

2 When your model and an experiment disagree, the disagreement is usable: the saturation failure became the 26–44 bound.

3 Trying to break your own result buys more trust than adding to it.

Limits: magnitudes are bounded rather than identified, and the subgroup pattern does not replicate in Tanzania.

An adoption-aware crop recommender
from farmer world feedback:

+50 kg/ha over the accuracy-only policy,
bracketed at 26–44 kg/ha by an independent RCT.

Thank you.

vairaajb@gmail.com

full paper: