

completely wrong

We weight models based on their stakes. We don't need to know how good a model is on the test set. It's not as strong of a signal as stake. We trust you.