## Probe Accuracy for Yi-34B 8.0 Bio (base) Bio (unlearned) Cyber (base) Accuracy 9.0 9.0 Cyber (unlearned) Random chance 0.2 20 30 40 50 60