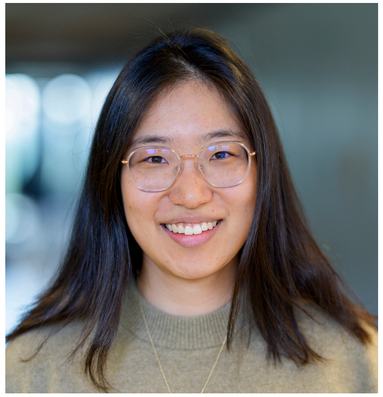
NEURAL INFORMATION PROCESSING SYSTEMS

Strategic Hypothesis Testing

NeurIPS 2025 · Spotlight



Yatong Chen*

Safwan Hossain*

Yiling Chen

*equal contribution

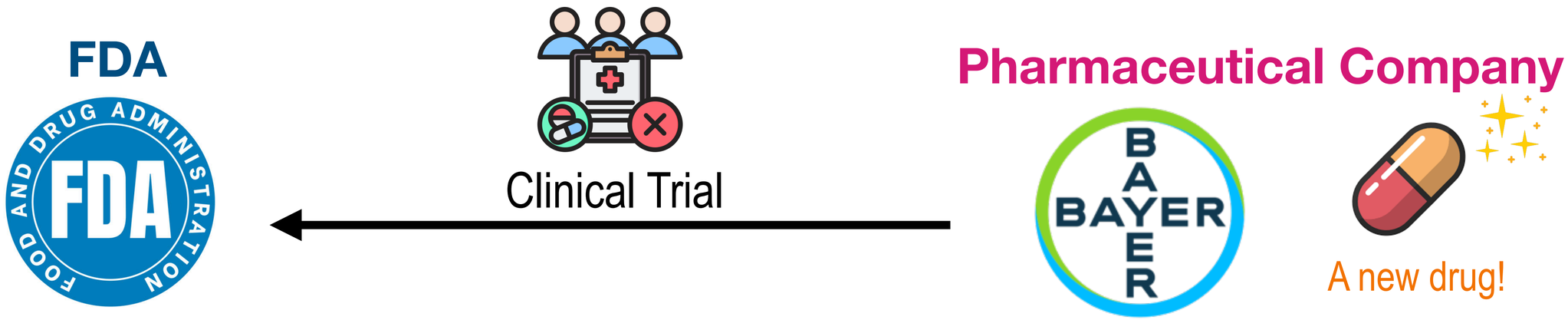
My Research

- My research lies in the intersections of **Game Theory, Machine Learning and Decision-Making**
  - how **strategic** human agents interact with ML systems?
  - how **incentives from the decision rule** shape the outcomes of algorithmic decisions

- This paper fits into this theme!

Decision-Making Based on Statistical Evidence

- Many real-life decisions have to rely on (noisy) evidence from a statistical test
- Example: the FDA's drug approval process



- **FDA's decision** (Approve ✅ or Reject ❌) on the drug is entirely based on clinical trial results

Decision-Making Based on Statistical Evidence

FDA

Clinical Trial

Pharmaceutical Company

A new drug!

FDA's Testing Protocol

FDA chooses a p-value threshold α (e.g., 0.05) $\Rightarrow$ **Company** runs a trial with n patients $\Rightarrow$ **FDA** computes the p-value p $\Rightarrow$ $p \leq \alpha$?

- *Yes* → ***Approve*** 💰
- *No* → ***Reject*** 😞

Drug Company's Decision & Utility

FDA's Testing Protocol

⇒ **Company** runs a trial with n patients ⇒ ⇒ $p \leq \alpha$?

- **Company** (with effectiveness μ) chooses the size of the clinical trial n^* to maximize the expected utility

$$u(n;\alpha) = R \cdot \Pr[\text{approve} \mid n, \alpha] - \text{Cost}(n)$$

Revenue once approved 📈💰📊

Cost of the trial (with n samples) 💵💵💵

For any fixed rule α, sample size n ↑
⇒ *Pr[approve| n, α] ↑ but Cost(n)↑*

Company will do **strategic self-selection**:
If $u(n^*;\alpha) < 0$, agent will not participate

FDA's Decision & Utility

FDA's Testing Protocol

FDA chooses a p-value threshold α (e.g., 0.05) $\Rightarrow$ $\Rightarrow$ $\Rightarrow$ $p \leq \alpha?$

$$u(n;\alpha) = R \cdot \Pr[\text{approve} \mid n, \alpha] - \text{Cost}(n)$$

- **The FDA,** knowing agents will do a ***strategic self-selection,*** chooses α^* to minimize total errors

$$\ell(\alpha) = \mathbb{E}_Q[\lambda_1 \text{FP} + \lambda_2 \text{FN}]$$

Approve inefficient drug

Disapprove effective drug/ effective drug not participate

Results & Implications

- We provide characterizations of the **optimal behavior** of both the principal (selection criteria α^*) and the agent (optimal size of the trial n^*)
  - We characterize how different loss components (FP, FN) change as we vary α
  - They are still monotonic in α (same as in classic hypothesis testing!)
- Our results offer insights for policymakers/regulators
  - Agent's strategic self-selection works in favor of the principal!

Proposition $\forall \alpha > 0$, there exists a participation threshold $\mu_\tau(\alpha)$, s.t. only drug with effectiveness $\mu \geq \mu_\tau(\alpha)$ will participate. $\mu_\tau(\alpha)$ is monotonic in α.

  - Principal can choose a stricter α so only good agents participate

Final Remarks

- A statistical decision-rule on strategic agents **not only makes decisions** — it also acts as a **gatekeeper**, shaping who chooses to participate.

- Our model can be used to study much broader settings:
  - A/B testing for a new feature
  - Scientific discovery
  - Etc…

Thank you! Question?