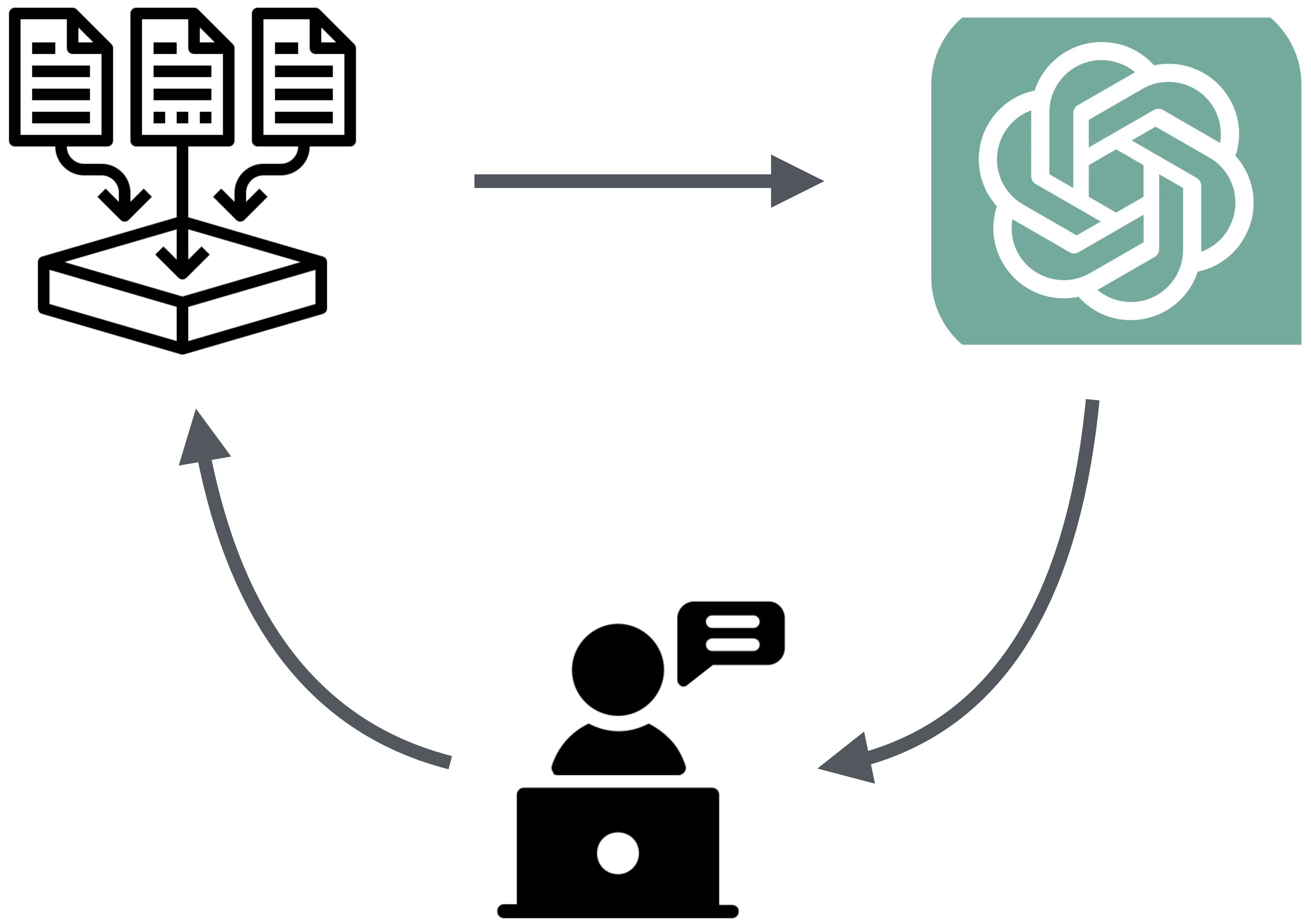
Learning with Feedback Loops

DJ Thornton — UNSW Sydney

ESAM 2024

Motivation

Many learning processes have the potential problem that realizations of the data affect the data generating process itself.



Motivation

Recommendation platforms

- An LLM provides information to users who then post that information online, which in turn is trained on by future iterations of the LLM.
- A platform promotes certain products leading more users to purchase them, which in turn reinforces their status as best-sellers.
- A search engine ranks high-traffic websites higher, resulting in increased visits to those sites, which in turn boosts their rankings further.
- A social media platform favors certain content types, causing more users to engage with them, **which causes the algorithm to prioritize similar content.**

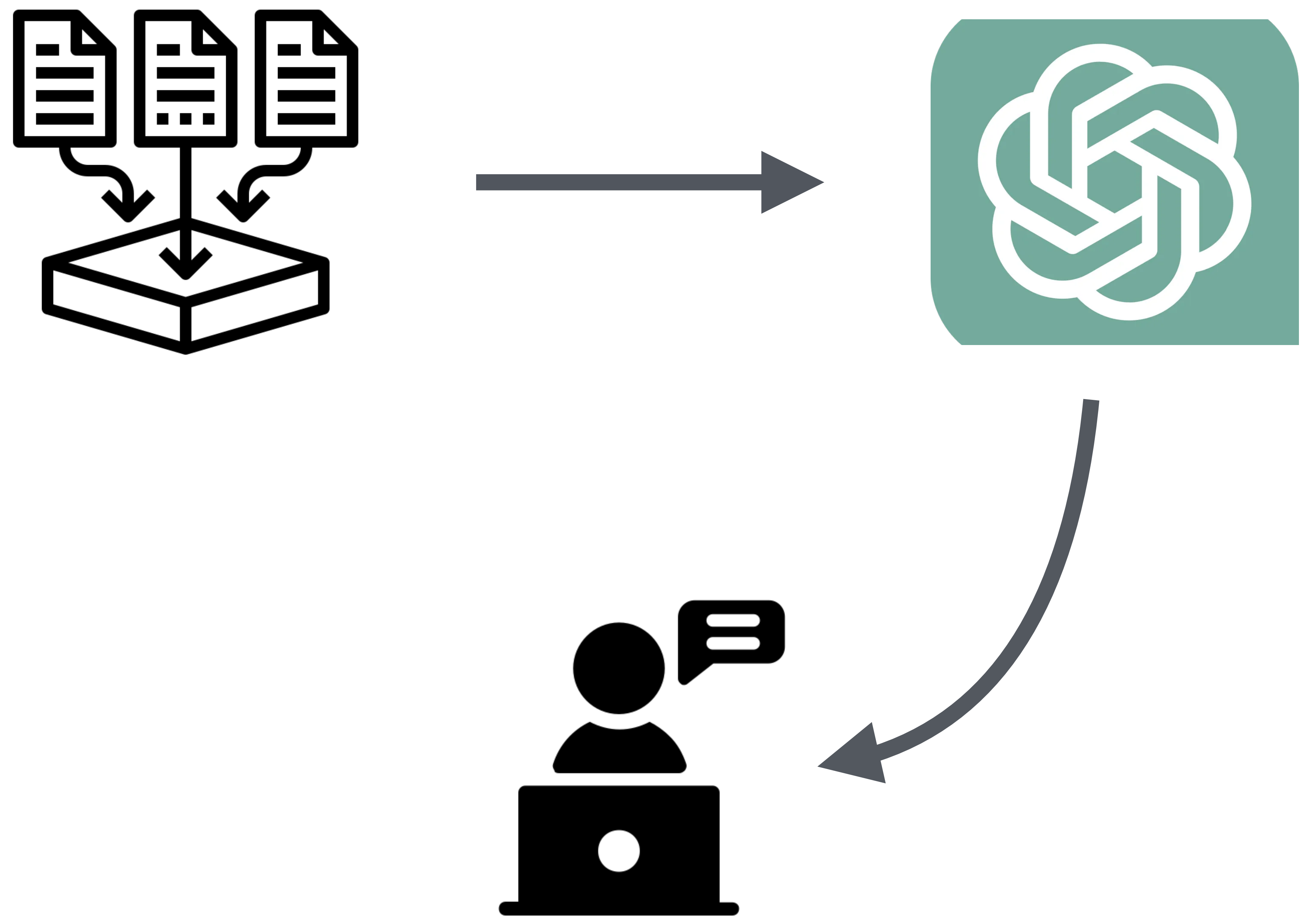
Motivation

How important is it to account for feedback in the data?

In particular:

1. What are the implications for learning if the platform *fails* to account for the **feedback** in the data-generating process?

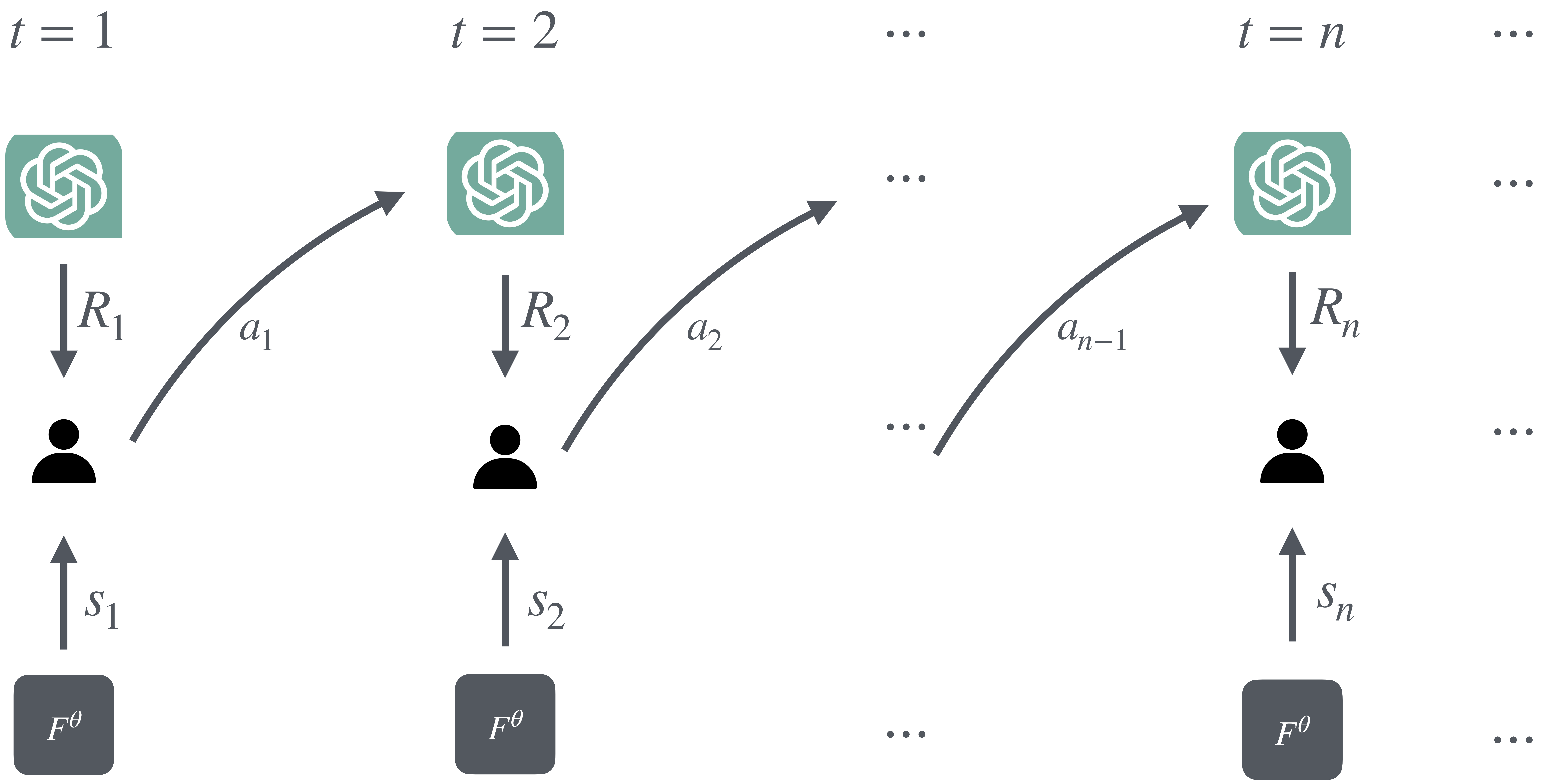
2. If learning fails in 1., is there a policy which corrects the failure?



Literature

- Social learning: Banerjee (1992); Bikhchandani, Hirshleifer and Welch (1992); Smith and Sørenson (2000); Acemoglu et al. (2011) and many others.
- Welfare in social learning models: Vives (1997); Smith, Sørenson and Tian (2021).
- Naive social learning: DeMarzo, Vayanos and Zweibel (2003); Golub and Jackson (2010), **Eyster and Rabin (2010)**
- (Social) learning with experimentation: Kremer, Mansour and Perry (2014); Che and Hörner (2018);
- AI Errors and Model Collapse: Kobak et al. (2024), Chawla (2024), Shumailov et al. (2024)
- Learning from Reviews: **Acemoglu et al. (2022)**

Model



$t = 1$
$t = 2$
$t = n$
R_1
R_2
R_n
a_1
a_2
a_{n-1}
s_1
s_2
s_n
F^θ
F^θ
F^θ

Model

Players and timing

- There is an unknown state of the world $\theta \in \{L, H\}$.
- Players: A long-lived **platform** and short-lived **agents** $1,2,\ldots$
- In period t :
  - The platform provides a **recommendation** $R_t \in \{0,1\}$ to the agent.
  - The agent observes an exogenous signal $s_t \sim F^\theta$ and chooses an **action** $a_t \in \{0,1\}$.
  - The platform observes the action of the agent but not the signal.

Model

Technical aside

- Failures of learning occur when the distribution over private signals induces **bounded posterior beliefs** (Smith and Sørenson; 2000)
- Failure of learning in my setting happens for a different reason.
- We will assume F^L and F^H induce **unbounded private beliefs.**

Model

Utility

- Agent t receives a payoff of 1 if their action matches the state, and 0 otherwise,

$$u_t(a_t) = \mathbb{I}[a_t = a^\theta],$$

- where $a^L = 0$ and $a^H = 1$.

- Platform derives utility from making correct recommendations to agents.

- Simplifying assumption: platform always prefers to make correct recommendations.

Model

Strategies

- Agent t chooses an action $a_t \in \{0,1\}$ after observing R_t and s_t.
- Importantly, the platform is more likely to be well informed at later times t, so a strategy for agent t is a function: $a_t = \alpha_t(R_t, s_t)$.
- The platform observes the full history of actions and recommendations, so a strategy maps histories into recommendations at time t, for every t.

Model

Strategies

- Agent t chooses an action $a_t \in \{0,1\}$ after observing R_t and s_t.
- Importantly, the platform is more likely to be well informed at later times t, so a strategy for agent t is a function: $a_t = \alpha_t(R_t, s_t)$.
- The platform observes the full history of actions and recommendations, so a strategy maps histories into recommendations at time t, for every t.
- Every history induces a posterior distribution: $Q_t = \mathbb{P}(\theta = H \mid a_1, \ldots, a_{t-1}, R_1, \ldots, R_{t-1})$.

Model

Naïve recommendation platforms

- A **naïve recommendation platform** is one which does not account for the effect of recommendations on learning from the data.
- Platform treats the data as i.i.d.
- A strategy for a naïve recommendation platform is a function $R_t = \rho(Q_t)$.
- There's something a little bit subtle here.

Model

Equilibrium

- I characterise learning outcomes in perfect Bayesian equilibria of the game.

- An equilibrium strategy for agents must satisfy:

$$\alpha_t(R_t, s_t) \in \arg\max_a \mathbb{E}_t[u_t(a) \mid R_t, s_t] .$$

- I focus on equilibria in which the platform's recommendations are always informative (i.e. no babbling).

- An equilibrium strategy for the naïve platform satisfies

$$\rho(Q_t) \in \arg\max_{R_t} \mathbb{P}(R_t = a^\theta \mid \mathscr{H}_t)$$

- Where $\mathscr{H}_t = (a_1, \ldots, a_{t-1})$ is the history of actions up until time t.

Analysis

Analysis

Learning

- Suppose (WLOG) that $\theta = H$.
- Say that *asymptotic learning occurs* (in equilibrium) if

$$\liminf_{t \to \infty} \mathbb{P}(\theta = H \mid R_t, s_t) = 1.$$

and that *learning fails* otherwise.

- Two observations:
- Despite the fact that the platform is naïve, it is not ex-ante obvious whether agents can correct any failures to learn in the limit.
- Asymptotic learning can occur iff the platform eventually "learns" the true state with certainty.

Analysis

Failure of learning for naïve platforms

Proposition 1

For any naïve recommendation platform, learning fails.

- **Proof:** (sketch, by contradiction).
  - Suppose $\theta = H$, and the platform's belief converges to the correct belief with certainty.
  - Then with probability 1 there is a time T such that the planner chooses $R_t = 1$ for all $t > T$.
  - But consider a sequence of private signals up to time t under which agents choose $a_t = 0$ for all $t \leq T$.
  - This must induce $R_T = 0$ and occurs with strictly positive probability. A contradiction.

Analysis

Herding

- Say that correct herding occurs (in equilibrium) if for $\theta \in \{L, H\}$,

$$\limsup_{t\to\infty} \mathbb{P}(a_t \neq a^{\theta}) \to 0$$

where $a_t = \alpha_t(R_t, s_t)$.

- An immediate corollary of **Proposition 1** is that under a naive recommendation system, herding on the correct state does not occur.

- But what is the limiting probability that an agent takes the wrong action?

Analysis

Contrarianism

- In the limit the platform either always recommends 1 or always recommends 0.
- Let $\pi = \mathbb{P}(R_\infty = 1 \mid \theta = H) > \frac{1}{2}$
- In particular, there is an infinite stream of agents who take contrarian actions in the limit regardless of the recommendation or the state.

Corollary 1 (Contrarian agents)

If the platform recommends $R_\infty = 0$ and $\theta = H$, then the limiting probability that an agent takes the incorrect action is $\mathbb{P}(s_t \leq \pi) = F^H(\pi)$.

Analysis

Just how bad is the failure of learning?

- Short answer: it's virtually impossible to say exactly, but I can produce a lower bound.
  - The platform essentially uses a majority rule.
  - There is an important tipping point.
  - At this tipping point, the probability of converging to either limit-recommendation is the same (i.e. $1/2$).
- So the probability of failure is $\frac{1}{2} \times$ the probability of arriving at the tipping point.

Policy

Policy

- Some natural things we might do...
  - **Make the platform smarter**
  - Platform gives agents more information
  - Agents give the platform more information
  - Make agents' information better
    - Allow agents to (partially) observe each other

Policy

Can we fix the failure of learning?

- Yes. In some sense it's easy.
  - (But in another sense it's very hard).
- Suppose the platform is cognisant of the feedback in the data-generating process.
- The key thing that this platform does which a naïve platform does not is adopting a version of the "overturning principle".

Policy

Can we fix the failure of learning?

Proposition 2 (Cognisance of feedback is sufficient for learning)

If the platform correctly updates beliefs, then both learning and correct herding occur.

Conclusion

- Recommendation systems suffer from the problem that realizations of the data affect the data-generating process.
- I develop a model of intermediated social-learning through a recommendation platform.
- Naïve platforms do not facilitate asymptotic learning, even when agents' information is "arbitrarily good."
  - Equilibrium behavior in the limit includes an infinite stream of contrarian actions.
- Platforms which do account for feedback facilitate asymptotic learning.