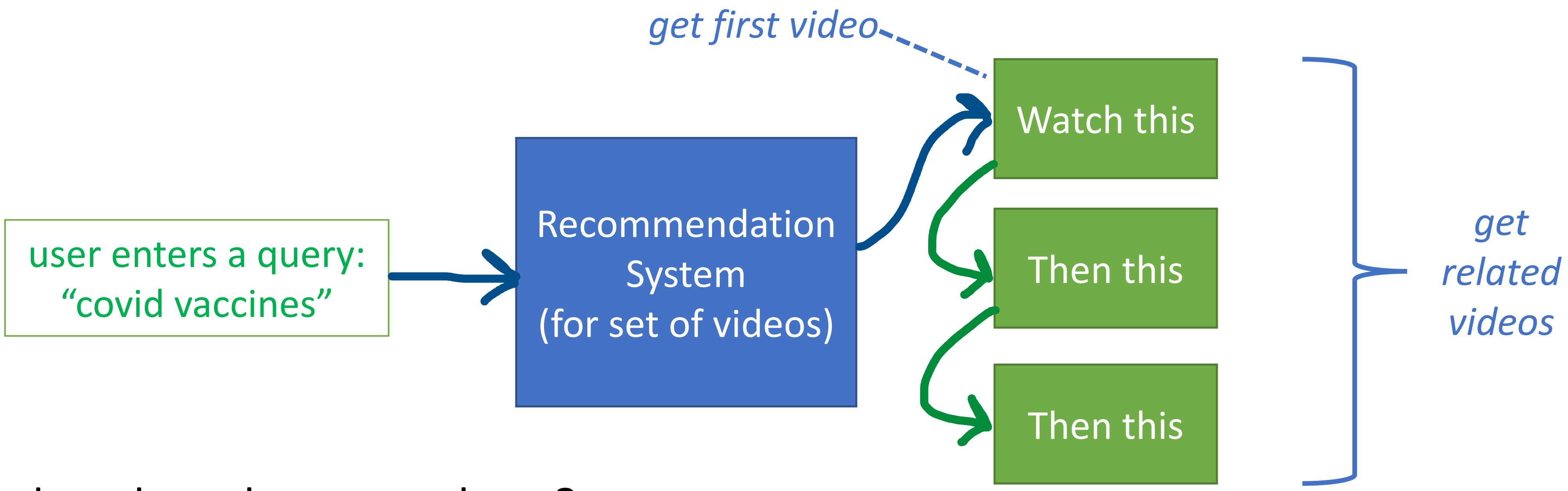
Lecture 19: Designing for Social Responsibility

Goal: Design a video-search/recommendation system


get first video
user enters a query: "covid vaccines"
Recommendation System (for set of videos)
Watch this
Then this
Then this
get related videos

What data does the system have?

- the set of videos

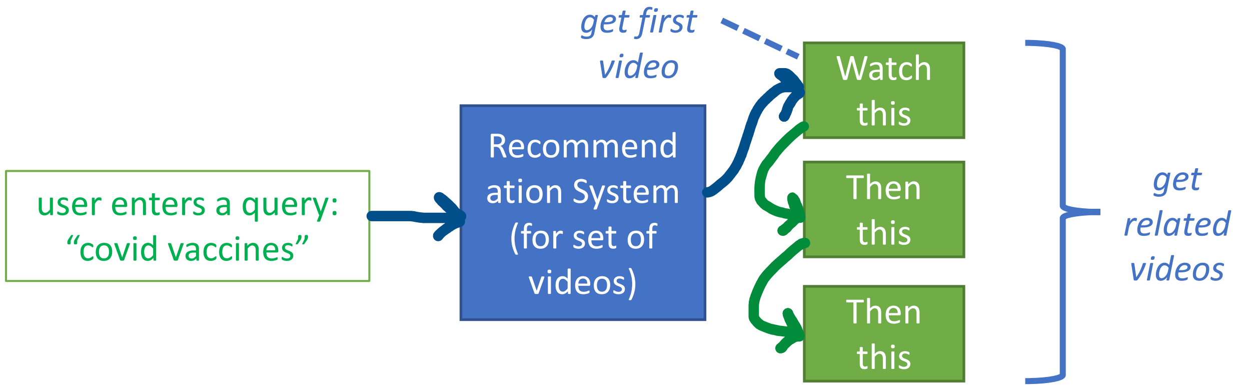
Two proposals for organizing the data:

- decision tree (recommender project)
- hashmap from query terms to videos (hwk4 grep, upcoming search project)

Breakout: sketch out how you could use these two proposals:

- sketch example/types of data structure
- how to find first video
- how to create sequence of related videos?

Goal: Design a video-search/recommendation system


user enters a query: "covid vaccines"
Recommendation System (for set of videos)
get first video
Watch this
Then this
Then this
get related videos

Let's explore the hashmap option:

- keys are search/query terms
- values are sets of videos
- this gets us to a set of videos from which we select a first video (based on views or likes), then generate a sequence of other liked videos

What data does the system have?

- the set of videos
- *info on what people have viewed or liked*

Provided by users

Provided by <u>different</u> users than one who uploaded the video

What if a user enters a query term that hasn't been seen before?

this is a standard decision test/case when using a hashmap (not SRC-specific)

YouTube Search/Recommender Algorithm Design

(danah boyd, “The Fragmentation of Truth”, https://points.datasociety.net/the-fragmentation-of-truth-3c766ebb74cf)

“Search doesn’t return high-quality content when high-quality content doesn’t exist”

- If people search for content that doesn’t exist, YouTube waits for users to upload something to the platform (rather than see what other quick-response sites like Twitter might be reporting)
- YouTube autoplays a “next” video, based on what previous viewers watched next or left comments on
- People find content by searching on keywords. If the search algorithm uses simple keyword match, groups can agree on keywords/phrases to promote content. Then if journalists or influencers promote the phrases ...
- Searching on outdated terms returns outdated results, devoid of context (including newer thinking on the terms)
- If content largely takes a one-sided opinion, all search results reinforce that opinion

“unknown key” question

other users coordinate to fill the gap with quick uploads and likes

people amplify those search terms

end up with a single viewpoint influenced by a coordinated group

A naive hashmap design enables this! Hashmap is designed to narrow down and cluster results

From Background Survey:

Consider a restaurant review website (like Yelp, TripAdvisor, or Dazhong Dianping). What would it mean for the system to be "fair"?

Look at a sample of 12 anonymous responses from classmates (raising a variety of perspectives)

Your breakout group works in slide

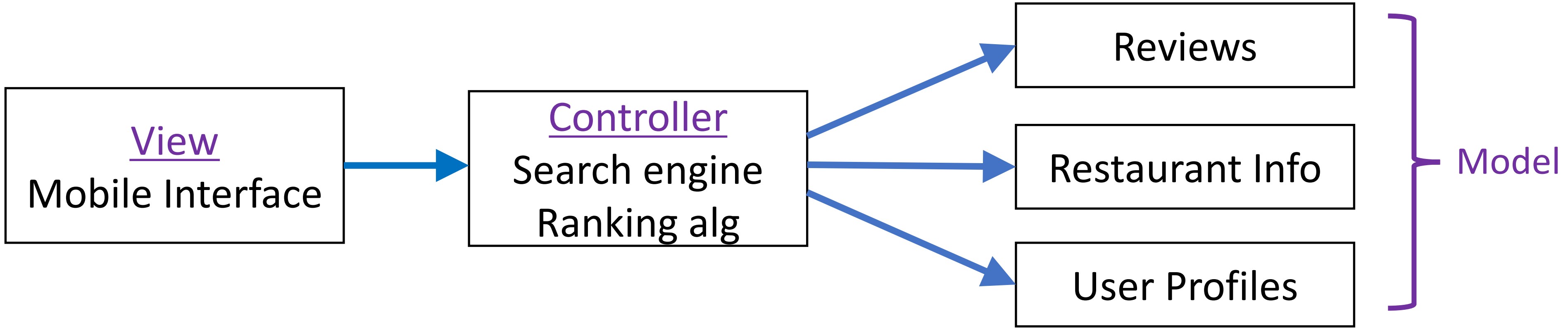
```
breakoutRoom# * 2+1
```

(e.g., room 1 → slide 3,
room 2 → slide 5, ...)

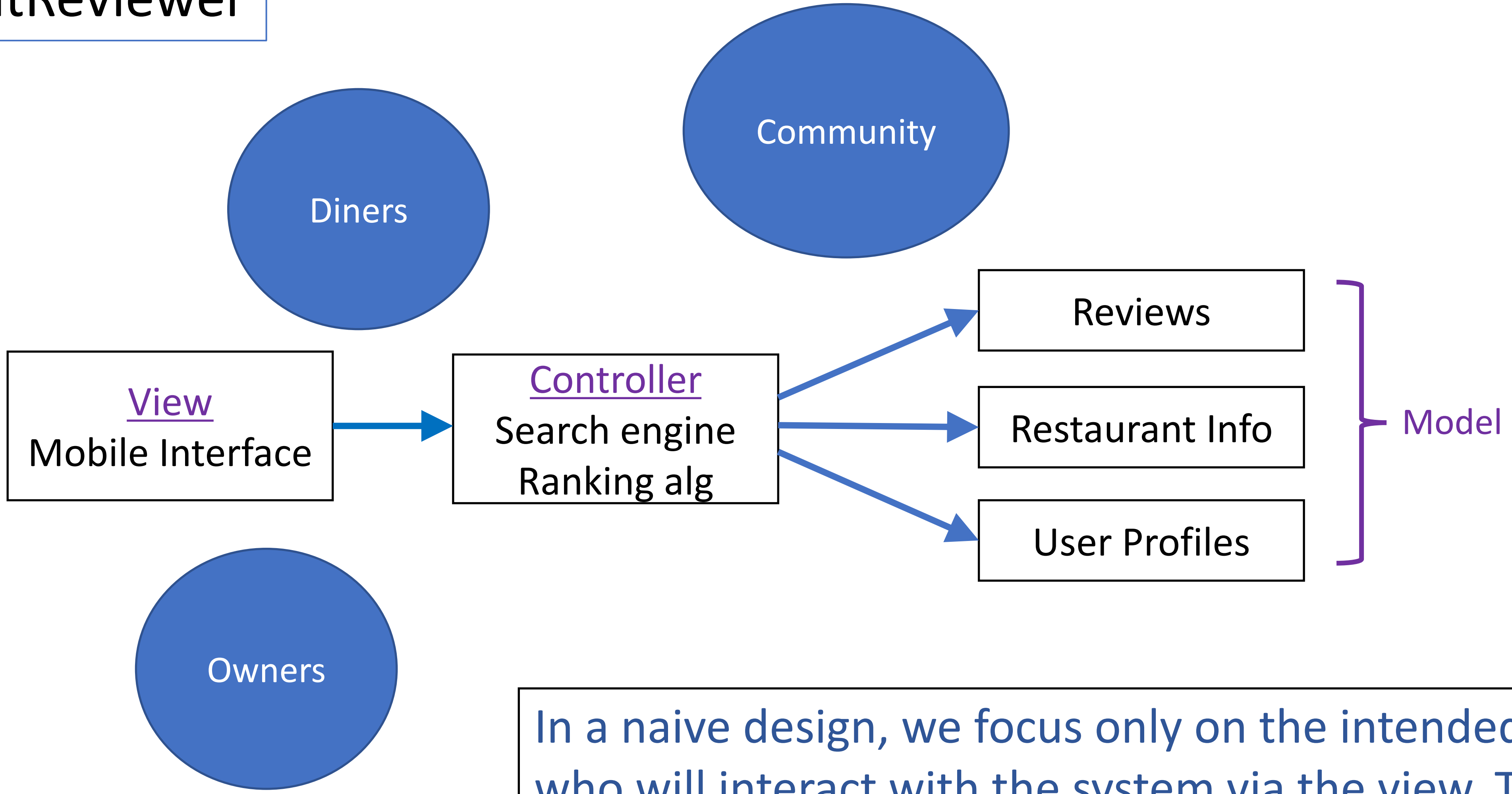
Breakout Exercise:

For each response (will complete on a google slide)

- Who is potentially harmed by this unfairness?
- Which aspects (Model/View/Controller/other) of the app play a role?



RestaurantReviewer



In a naive design, we focus only on the intended users who will interact with the system via the view. To design for social responsibility, need to also consider non-users who will be affected by the system (the “stakeholders”, which includes the users)

A mature design might have model elements that mitigate impacts for owners and community as well

Social Threat Modeling

Identify the stakeholders in your system. Then think through these issues:

Think through issues for each one

Information from which system makes decisions

Who can use or influence the data or computation?

Process by which decisions are made

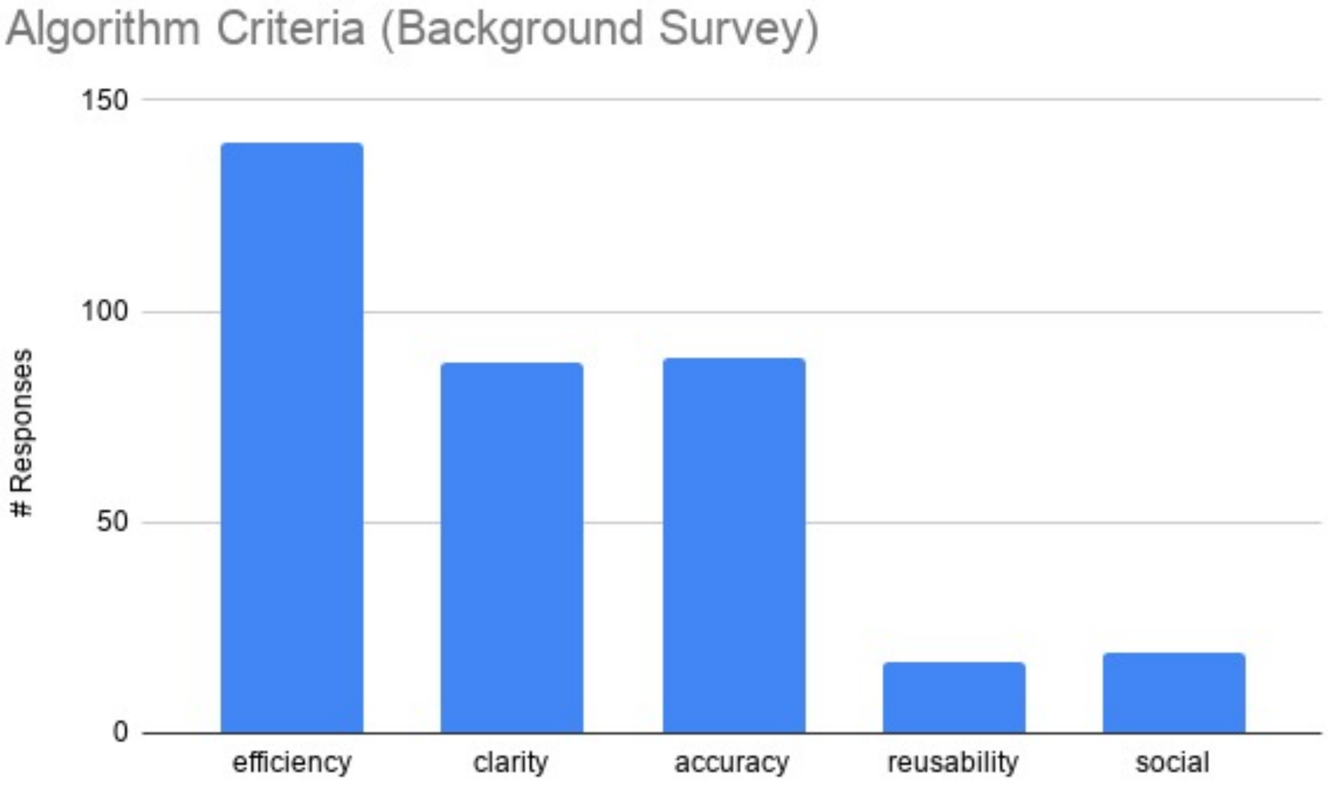
Stakeholder	Data	Agency	Algorithm
Fairness/Inclusivity			
Privacy/Reputation			
Sustainability			

Prompting questions for each cell are in the table at this link (which we will use on upcoming assignments)

Of system, environment, communities

From the Course Background Survey:

What criteria do you feel are important to consider when deciding whether an algorithm is good?



Social impacts are as important as efficiency: you should be able to design for both