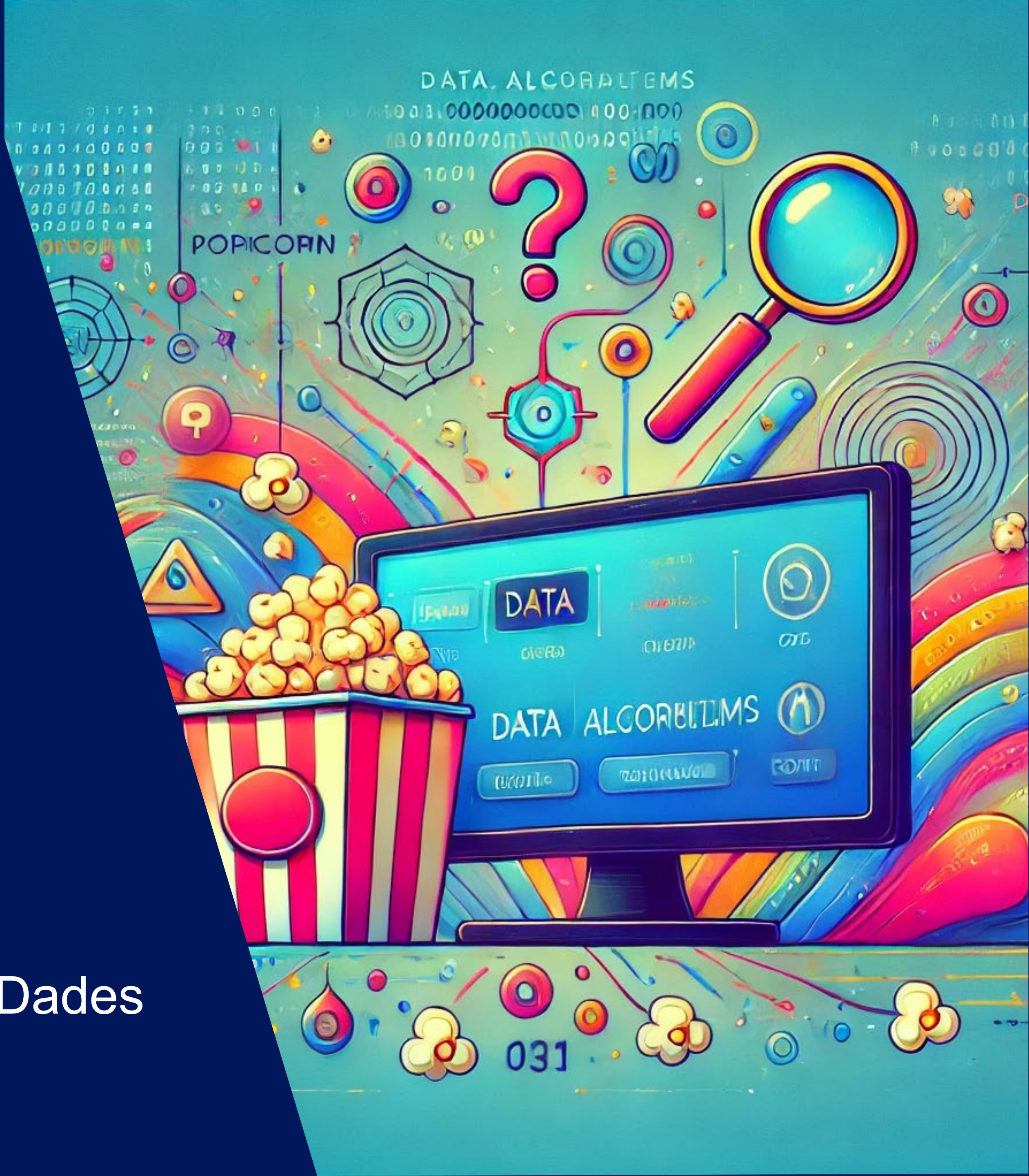
IQS UNIVERSITAT RAMON LLULL

Dades, algoritmes, crispetes i d’altres misteris digitals



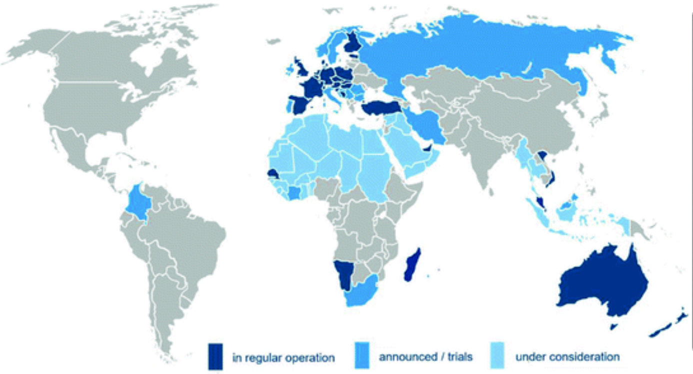
Dr. Francesc Martori
Department de Matemàtiques i Anàlisi de Dades
francesc.martori@iqs.url.edu

Times are changing in the TV industry

Average daily time spent watching television in Spain from 1997 to 2023

Times are changing in the TV industry

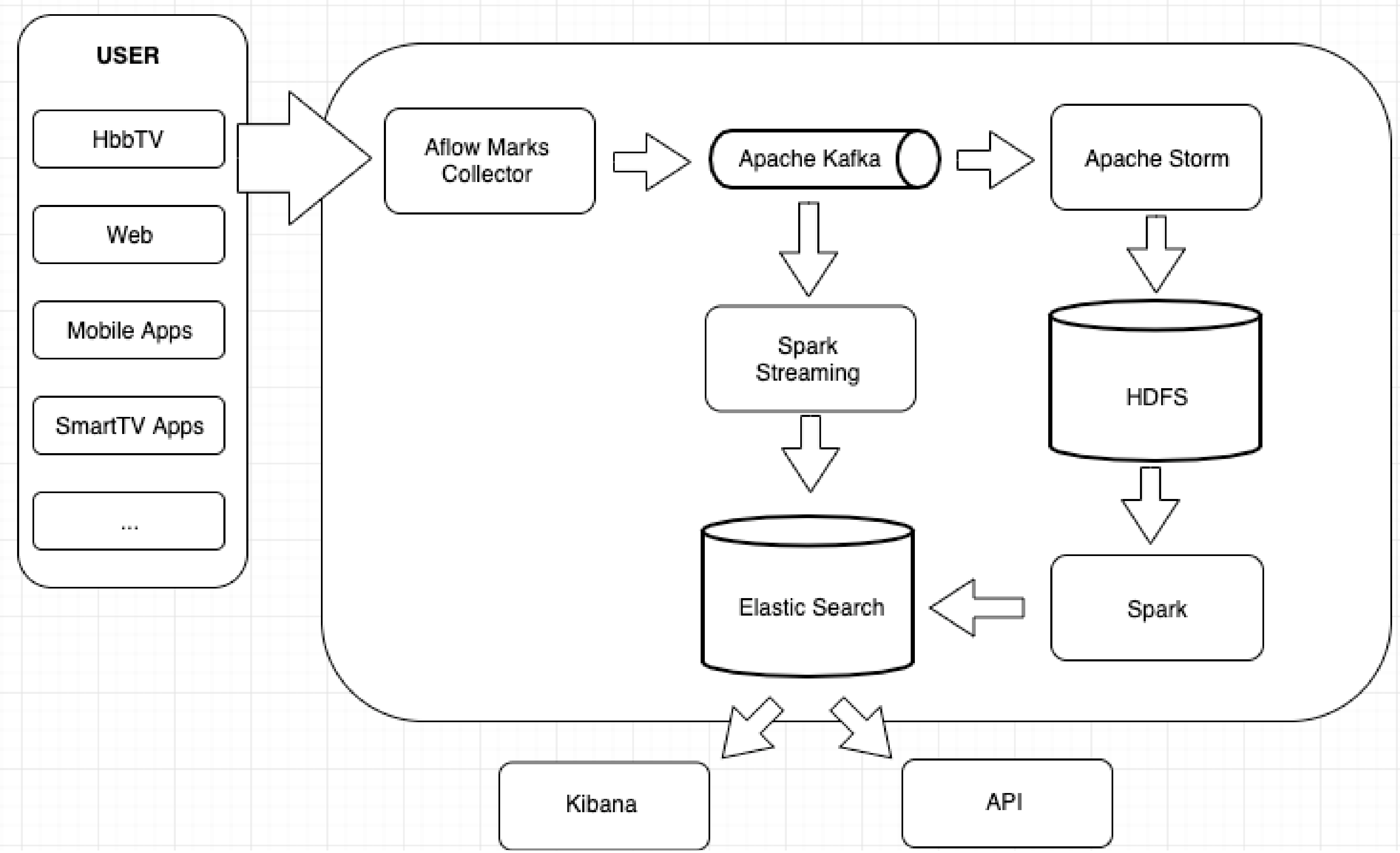
Hybrid Broadcast Broadband TV (**HbbTV**) is both an industry standard and promotional initiative for hybrid digital TV to harmonise the broadcast, IPTV, and broadband delivery of entertainment to the end consumer through connected TVs (smart TVs) and set-top boxes



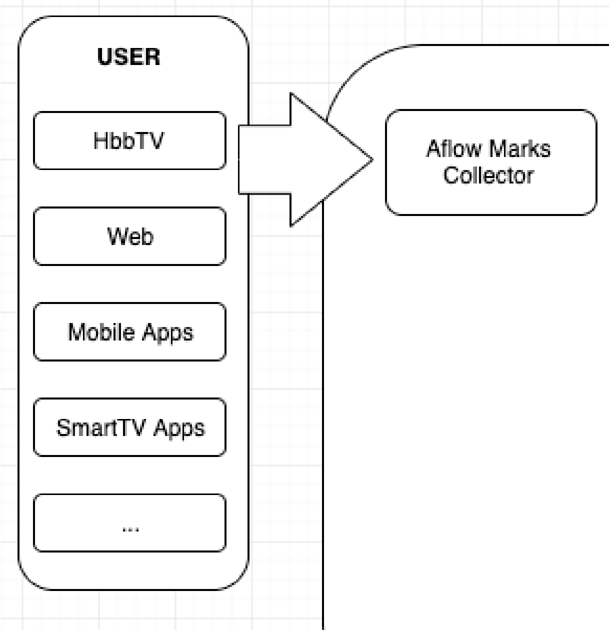
The main change is that TVs are like computers, and broadcasts like websites


TODAY
MENU

The flow of data



How is data obtained?



2 Key concepts

- sdk
- Cookies

Cookies

The EU Cookie Legislation requires 4 actions from website owners who use cookies:

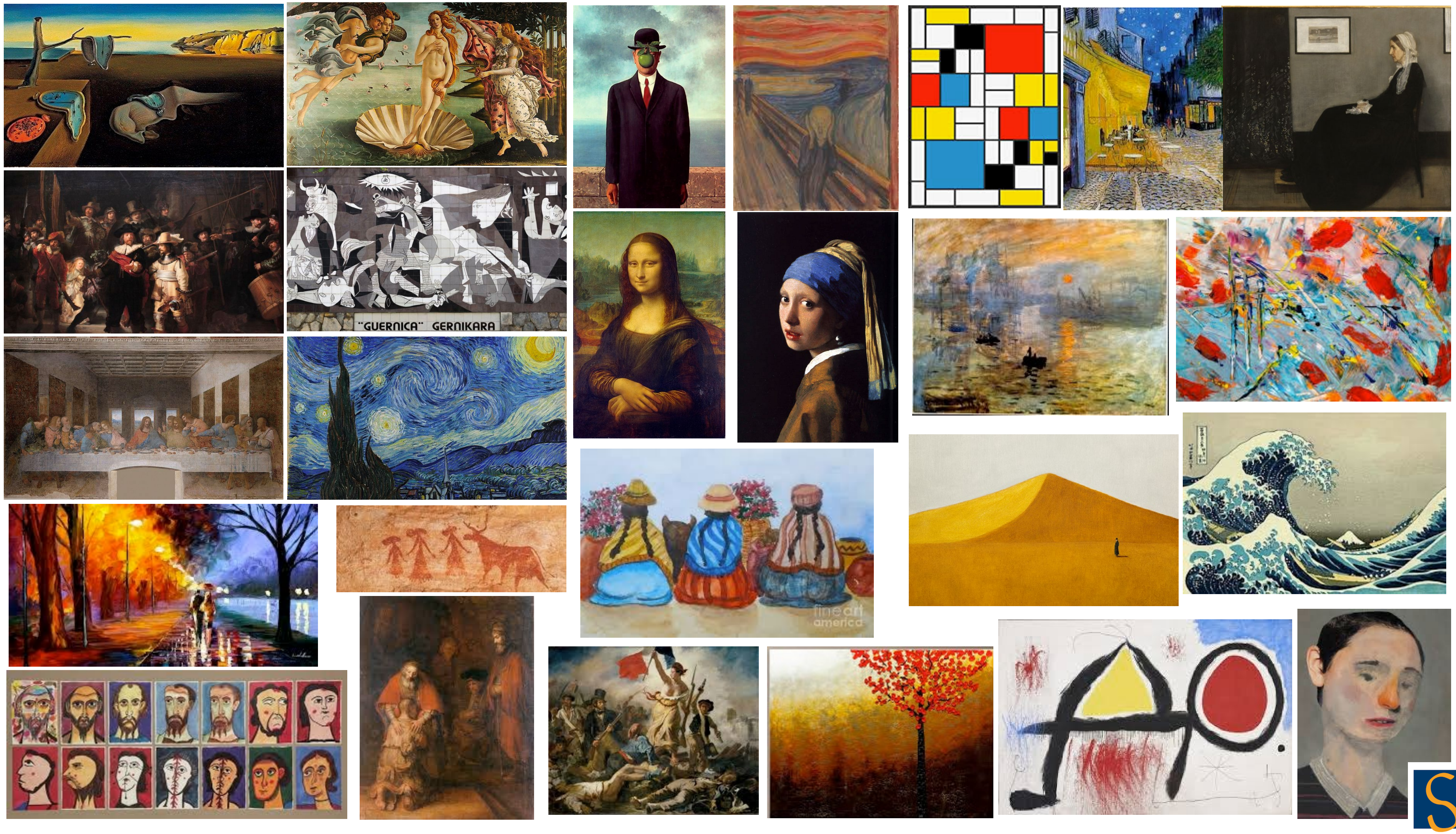
1. When someone visits your website, you need to **let them know that your site uses cookies**.
2. You need to provide detailed information regarding **how** that cookie data will be utilized.
3. You need to provide visitors with some means of **accepting or refusing** the use of cookies in your site.
4. If they refuse, you need to ensure that cookies will not be placed on their machine.

but … some Cookies Don't Need Consent

https://www.privacypolicies.com/blog/eu-cookies-directive/

What can we do with these data?

Choose a painting



Choose a painting

Choice overload

- In 2000, psychologists Sheena Iyengar and Mark Lepper from Columbia and Stanford University published a study about jams. On a regular day at a local food market, people would find a display table with 24 different kinds of jams. Then on another day, at that same food market, people were given only 6 different types of jam choices.

- Guess which display table lead to more sales?

Choice overload

- Iyengar and Lepper found was that while the big display table (with 24 jams) generated more interest, people were far less likely to purchase a jar of jam than in the case of the smaller display (about ten times less likely).

- Having too many options makes the decision more difficult for your brain and so, NOT making a decision becomes the easiest way out

Choice overload

- In an extensive study published in the Journal of Consumer Psychology in 2015, researchers analyzed a total of 99 ‘choice studies’ and specifically looked at those cases in which reducing choices helped to boost sales.

- They found four criteria that motivate consumers to buy:
  - When people want to make a quick and easy choice
  - When the product is complex (so fewer choices help the consumer make a decision)
  - When it’s difficult to compare alternatives
  - When consumers don’t have clear preferences

When do online recommendations make business sense?

Mostly, 3 business cases

- Netflix or my income comes from my subscribers
- Amazon or my income comes from my catalog
- My income comes from publicity

Users more time engaged = more ads

Users engaged = happy to continue paying

Personalized catalog = better engagement

Personalize ads = more likely to respond to it

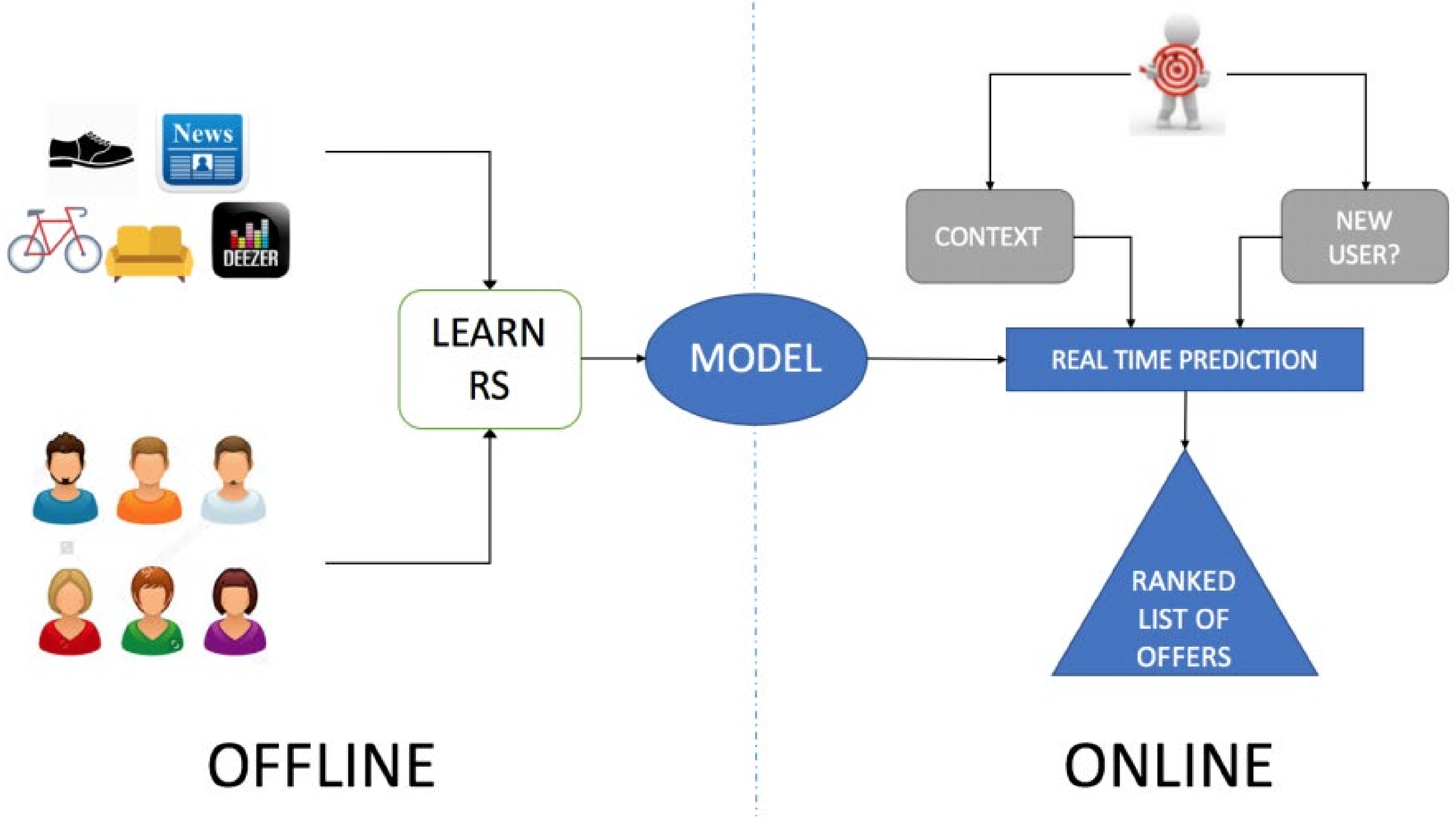
What is a recommendation?

- The purpose of a recommender system is to suggest relevant items to users
- Meaning (in math)

- Estimate a utility function that predicts (automatically) how much will a user like an item

- Using
  - Past behaviour (if it exists)
  - Relations to other users
  - Similarity between items
  - …

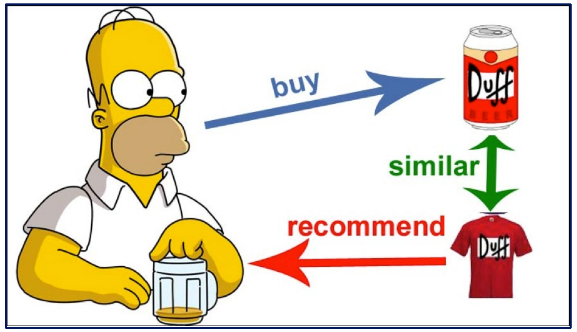
Two-step process



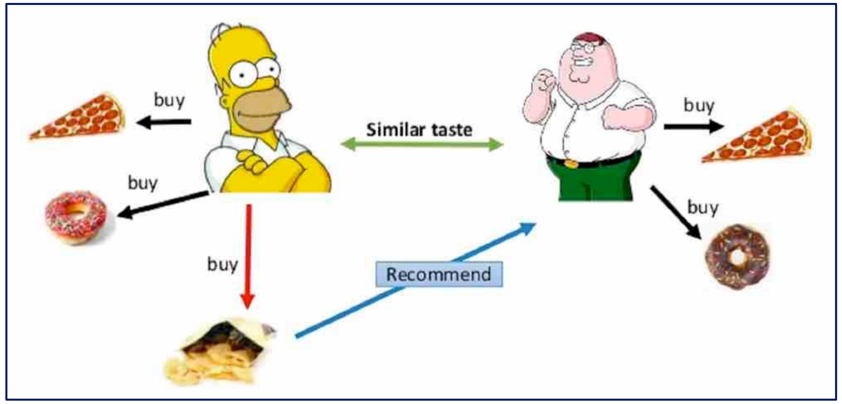
Sumit Sidana. Recommendation systems for online advertising. Computers and Society [cs.CY]. Université Grenoble Alpes, 2018. English.

Classical approaches to Recommendation

- **Content Based**: Recommendations are based on the attributes (content features) of items.



- **Collaborative Filtering**: Recommendations are based on other users' preferences and behavior.



Content based

- Recommendations are based on the attributes (content features) of items.
- The attributes of a movie could include the movie's genre, actors, producer, musical score, year, and/or topic words or keywords from its title and description.
- For example, a user is recommended other action movies if she liked the latest James Bond film.

Content based

Typical approach:

1 – Build a feature space (define which attributes will play)

2 – Represent items and users in that feature space

3 – Generate a similarity function between users and items

- Euclidian distance
- Manhattan distance
- Cosine similarity
- Tanimoto similarity
- …

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}},$$

Content based

PROS

- With this method you can make recommendations with sparser datasets.
- It and can easily describe why a particular item is being recommended.
- There is no cold start problem.

CONS

- Some domain specific knowledge may be required
- Quality of the input data is a must (No attributes, no item)
- There are some serendipity problems
- It can be easy to overfit a model

Collaborative Filtering

- Recommendations are based on other users' preferences and behavior. The key to this algorithm is determining the similarity between users through their ratings history, and then identifying other items these similar users liked.

- We need some sort of how did anyone liked anything?
- We can rely on item similarity or user similarity

- With either of these methods you do not need to determine the attributes of items and can have a higher diversity of recommendations, but you need a large number of user ratings and you may over recommend popular items.

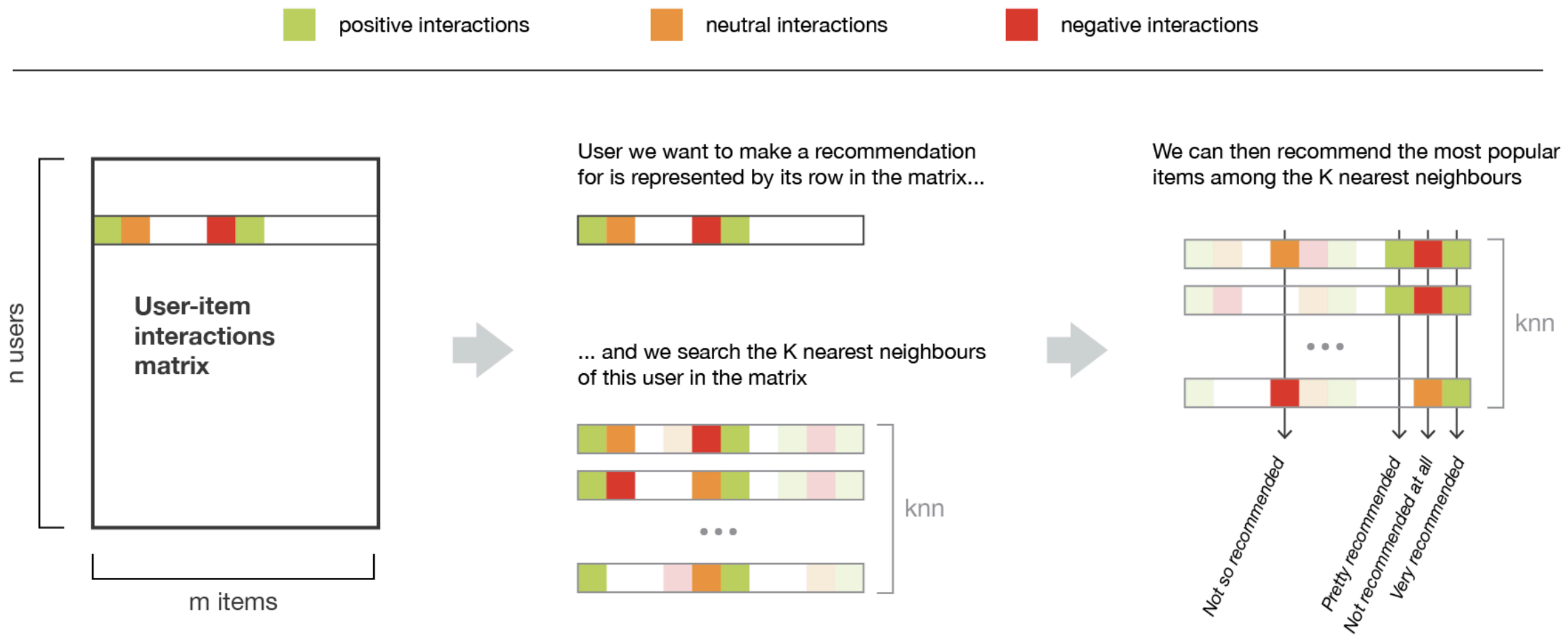
IQS UNIVERSITAT RAMON LLULL

How does it work?

- Consider a recommendation system in which the training data consists of a feedback matrix in which:
- Each row represents a user. Each column represents an item (a movie).
- The feedback about movies falls into one of two categories:
- **Explicit**— users specify how much they liked a particular movie by providing a numerical rating.
- **Implicit**— if a user watches a movie, the system infers that the user is interested.

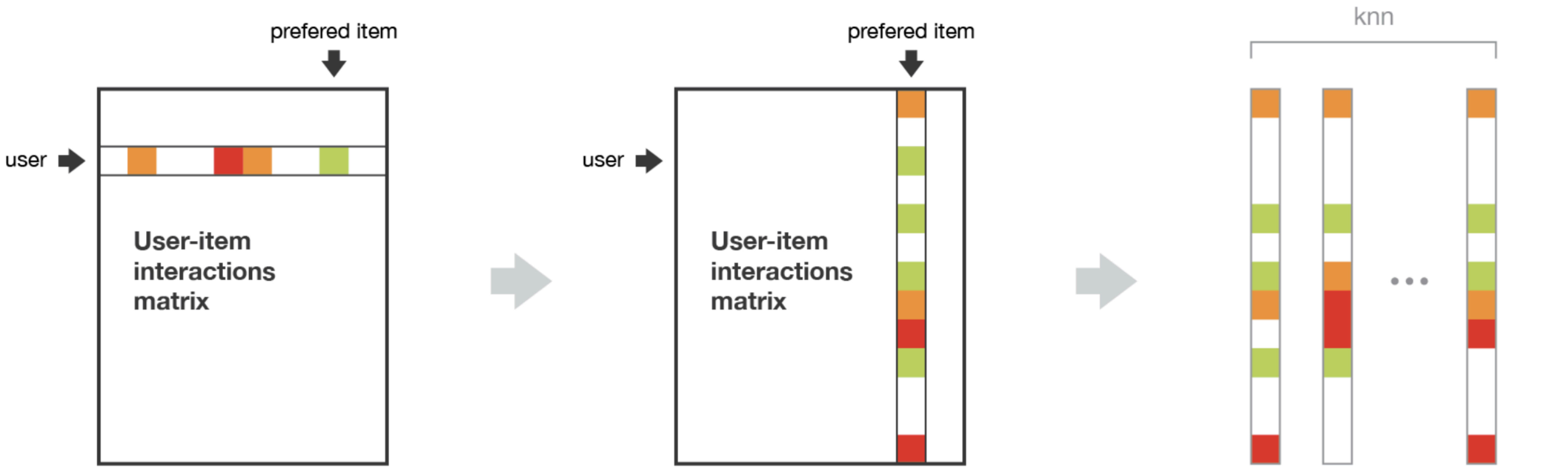
	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
User 1	X		X		X	
User 2		X	X			
User 3				X		X
User 4					X	
User 5	X	X		X		X
User 6			X	X		
User 7	X	X	X		X	X

User - User



https://towardsdatascience.com/introduction-to-recommender-systems-6c66cf15ada

Item - Item



We identify the prefered item of user we want to make recommendation for.

The prefered item is represented by its column in the matrix.

We can search and recommend the K nearest items to this "prefered item"

https://towardsdatascience.com/introduction-to-recommender-systems-6c66cf15ada

Collaborative Filtering

PROS

- Easy to generate good results
- Knowledge agnostic
- Can be easily tweaked into more complex situations (context, user profile)

CONS

- Cold start (and in general it needs lots of data)
- Prone to favorites
- Assumes that prior behavior determines current behavior

Is this enough? Business Rules

- Recommendation coming from statistics might not always capture the required logic in the daily activity.
- Business rules are a way to alter the results in order to enhance certain aspects of the recommendations
- i.e. bias towards new items, bias towards items of the same category, do not present adult content to kids …

Evaluation metrics in recommender systems

- **Accuracy**: Ratio of the number of recommendations used over the number of recommendations generated
- **Coverage**: Number of recommended items over the number of items in the catalog
- **Trust**: Number of users/devices that used recommendations over the number of users/devices who received recommendations
- **Novelty**: Recommendations involving items that the user/device didn’t know
- …

TOPs and friend recommendations are the real challenge

IQS

UNIVERSITAT RAMON LLULL