PIA - Youtube's Recommender System

Search

Home

Videos

Miscellaneous

Archive

Privacy Impact Assessment

YouTube's Recommender System

KU LEUVEN

Privacy and Big Data

WS 2020/2021, KU Leuven

Carlos Sebastian Michels Alfaro

Milton Ossamu Tanizaka Filho

Youtube

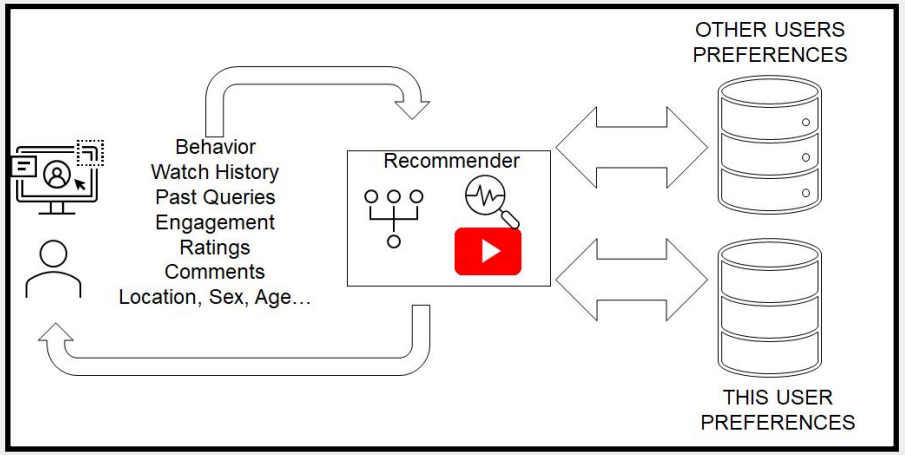
- Video streaming platform owned by Google

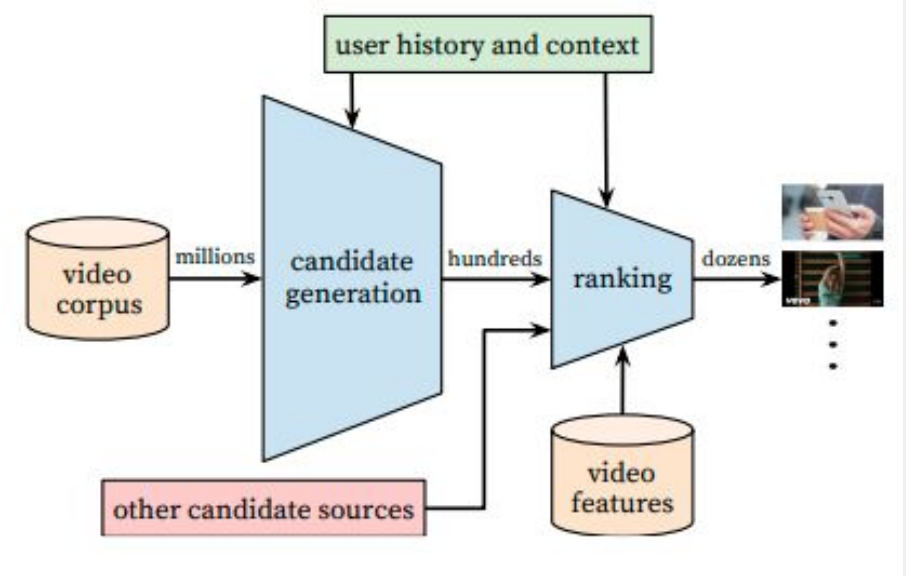
Recommender system

- Offers specific content by comparing the user profile with some reference characteristics
- Suggests videos and advertisements based on individual preferences
- 70% of content is recommended (700 million hours a day)
- 2 neural nets select and rank from video corpus.

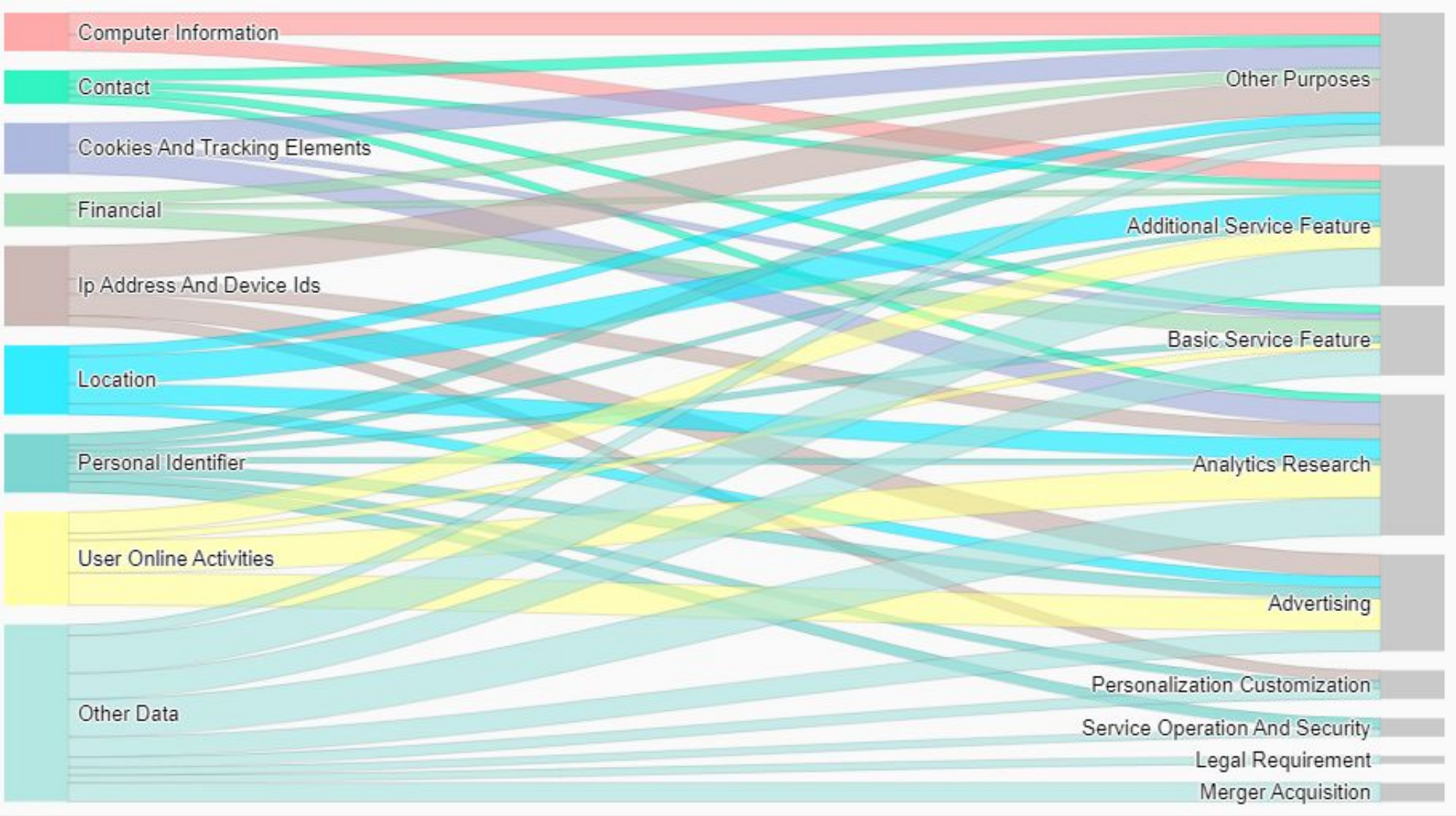
Data Collected

- Watch and search history
- Personal identifiers
- Explicit and implicit preferences

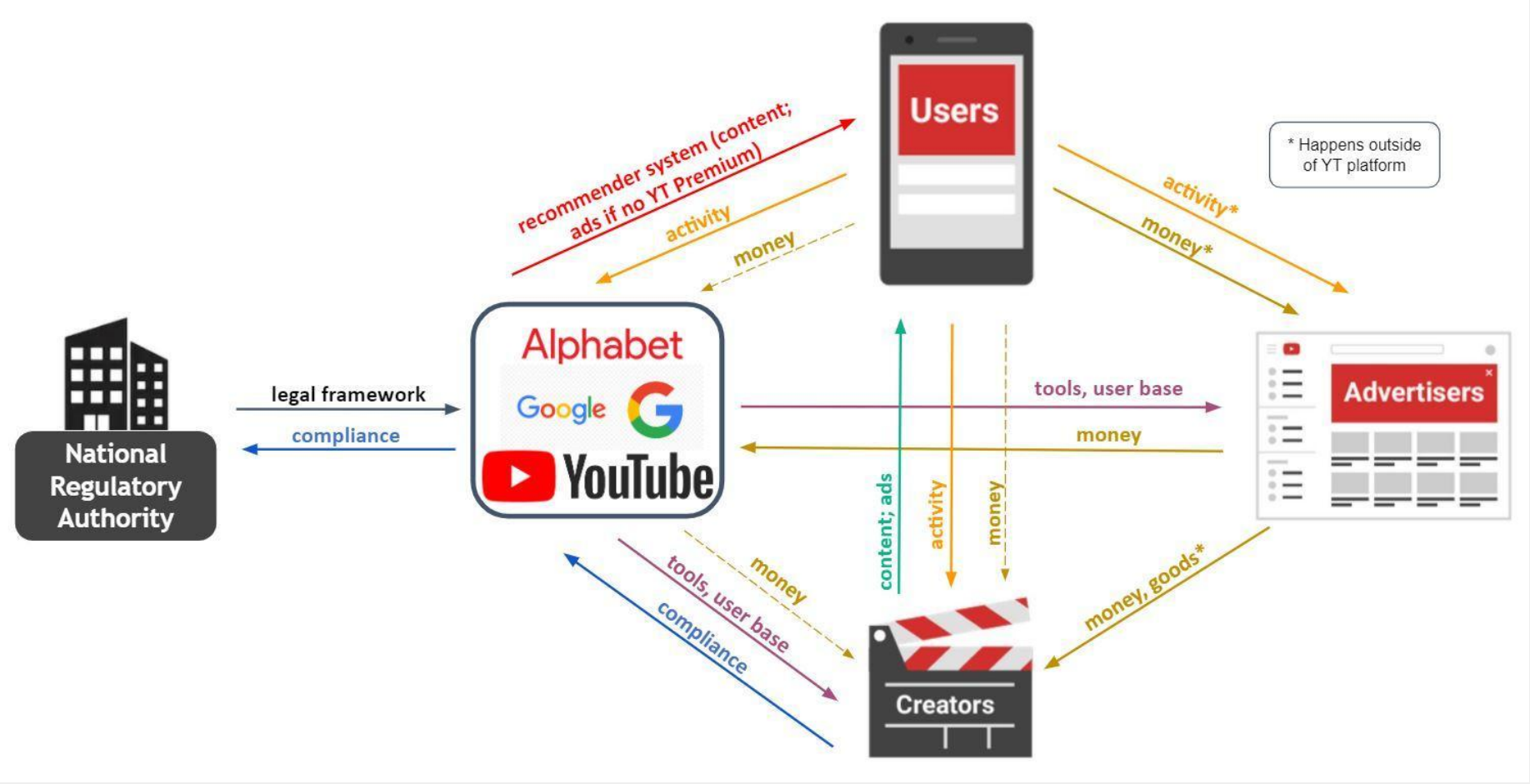




Data Collected and their Use (source: https://pribot.org/polisis)



Stakeholders



Privacy concerns

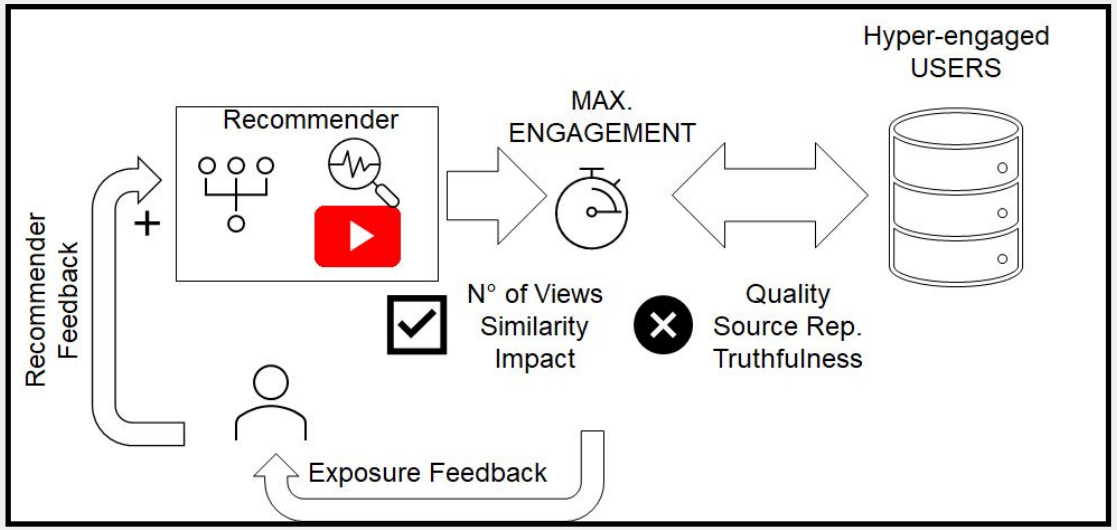
- Privacy: enabling right to human autonomy
- recommender systems infer personal interests and affiliations

Attacks

- Identification of user's preferences
- Behavior influence of the user via targeted content (extreme, biased or untruthful)

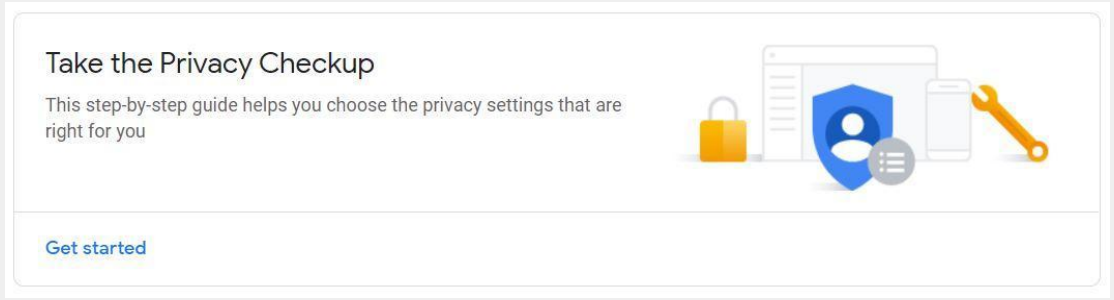
Trust Assumptions

- Several measures are put in place (e.g. encryption, 2 step verification)
- Barriers put in place protecting user privacy and data by YouTube and the user hold true



User Controls

- Youtube and Google level (account integration)
- Control the use search and watch history
- Download data / Delete account



Account Integration

- Policy and data management apply “Google wide”
- Data for different services is collected under single user (huge wealth of data)

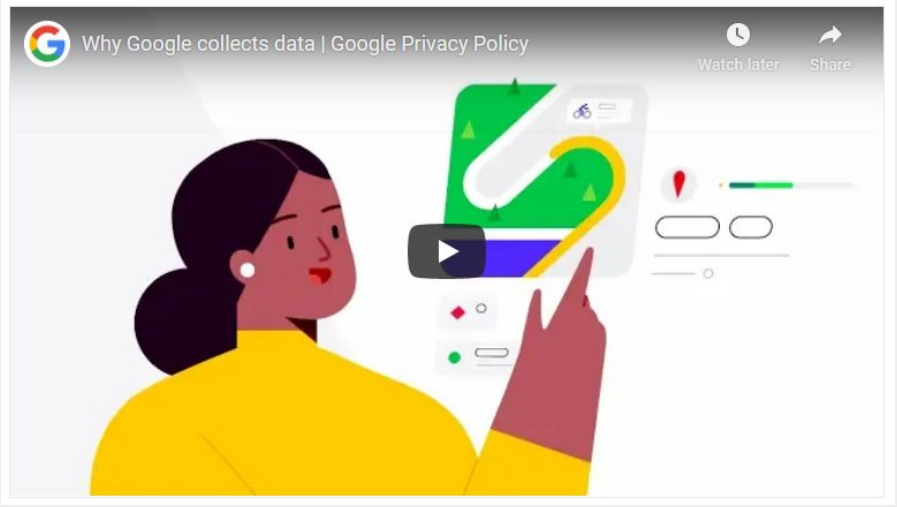
“We keep some data for the life of your Google Account if it’s useful for helping us understand how users interact with our features and how we can improve our services”.

Data Retention Periods and Dissemination

- Most data is kept for life (account deletion)
- Some data is deleted under user control, other data anonymized (k-anonymity and l-diversity)
- Most data stored in remote servers - even recommendations

Third-Party Sharing

- Explicit user consent
- External data processing (compliance)
- Justified Legal Reasons
- non-personally identifiable information (e.g. to advertisers)



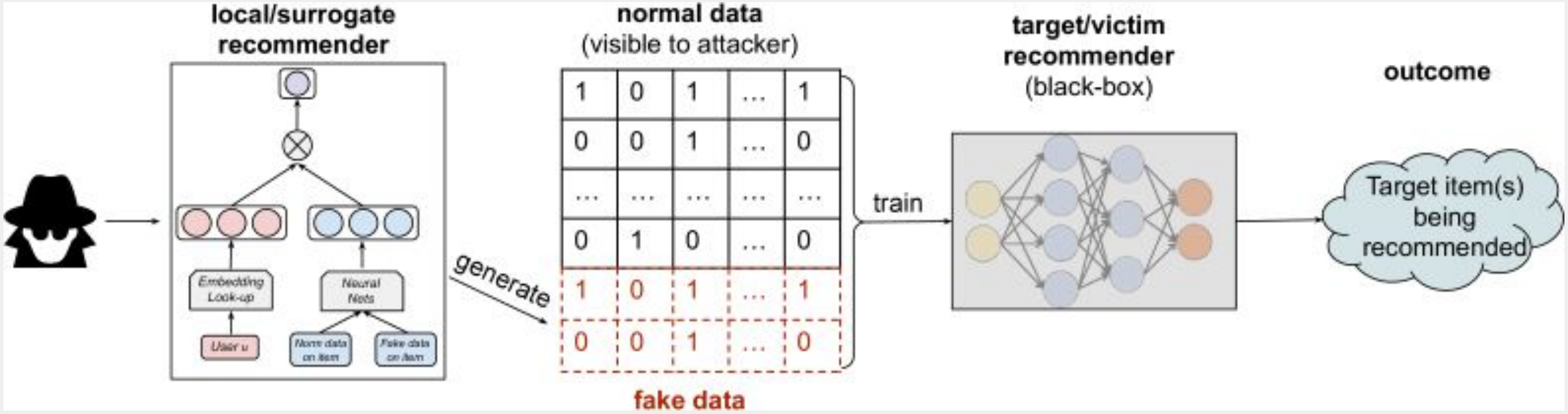
Consent

- Systematically requested in the terms of use of Google services
- Mattan 2019: *"no longer effectively protects personal privacy in our present data-rich world"*

Privacy Policy

- Clear and accessible, well structured, visuals video explanations of key concepts
- However, it takes 30 to 60 minutes to browse completely

Recommender System - Threat models

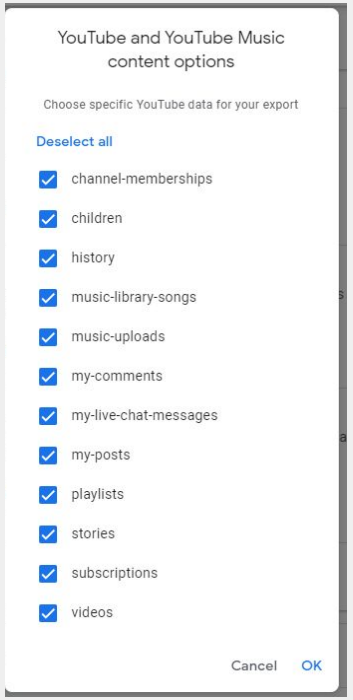


Type of attacks

→ Adversary creates craft fake data to inject them into training data and produce biased outcomes

Legal Aspects (GDPR)

- Broadness of purpose, justified by the wealth of services provided
- Core functionalities, including the RecSys are still possible with less data
- The RecSys can infer data within “special categories” (not disclosed)
- Storage limitation principle - most of the data is retained indefinitely
- All controls in the least private settings (even for disclosed data)
- Policy: “concise, transparent, intelligible and easily accessible form, using clear and plain language”.



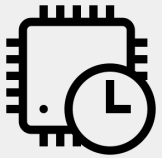
Ethical Aspects

- Maximize user attention can be in direct conflict with user wellbeing (addiction)
- Is the RecSys trying to entertain or deliberately tries to manipulate the user?
- Quality of the content is not taken into account into the RecSys
- Strong ethical concerns and conflict of interest

Recommendation

Incognito mode and privacy by default

→ Make incognito mode available for every devices - by default!



Data retention

→ Allow users automatic deletion of their data by a certain period

Minimum data collection and usage

→ Possibility to allow only minimum data collection to get recommendations - limit to recent data

Objective of the recommendation

→ Enable users to choose recommendations objective. Include user's wellbeing

Mandatory privacy check up

→ Mandatory check up everytime the privacy policy and terms of service change and include didactic content for better comprehension

PIA - Youtube's Recommender System

Search

Home

Videos

Miscellaneous

Archive

Privacy and Big Data

WS 2020/2021, KU Leuven

Carlos Sebastian Michels Alfaro

Milton Ossamu Tanizaka Filho