Social Data

May 2, 2018

LIVE

Today

- What is social data?
- How is it being collected and who collects it?
- To what purposes this data is used for?
- Discuss the ethical aspects of the social data usage.

Outline

- Define social data
- Processing pipeline for social data
- Examples and ethical discussion

Social Data

Social data is information that social media users publicly or **privately** share, which includes metadata such as the user's location, language spoken, biographical data, interests, shared links and other user generated content created with the intent to implicitly or explicitly communicate or interact with others.

Source: Social Data [Def. 1]. In Investopedia online. http://www.investopedia.com/terms/s/social-data.asp

Social Data - Examples

- Amazon’s users purchase history
- Criminal record information
- Posts/Likes shared on facebook
- National health security data centers
- Instagram photo albums

Social Data - Goals

- Understanding phenomena specific to social networks, sometimes with the objective of improving them
- Understanding and influencing phenomena beyond the social platforms, seeking to answer questions from sociology, psychology, or other disciplines

Social Data - Processing Pipeline

- Data acquisition and preparation
- Data processing
- Data analysis
- Evaluation and interpretation

Esquire STYLE NEWS POLITICS ENTERTAINMENT FOOD & DRINK

A 'Black Mirror' Episode Is Coming to Life in China

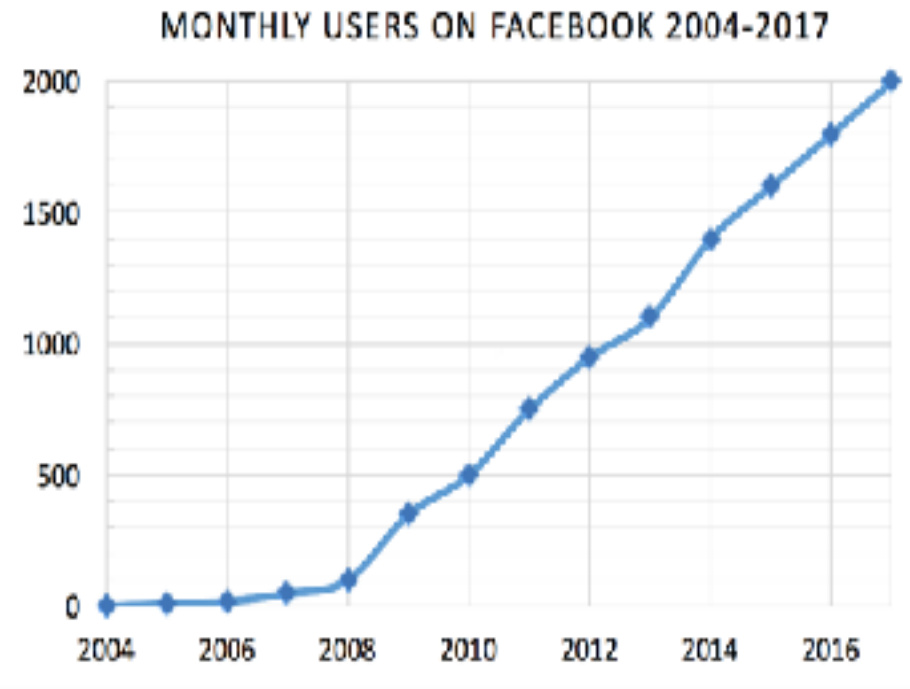
People will be prevented from traveling on trains and planes based on their social credit scores.

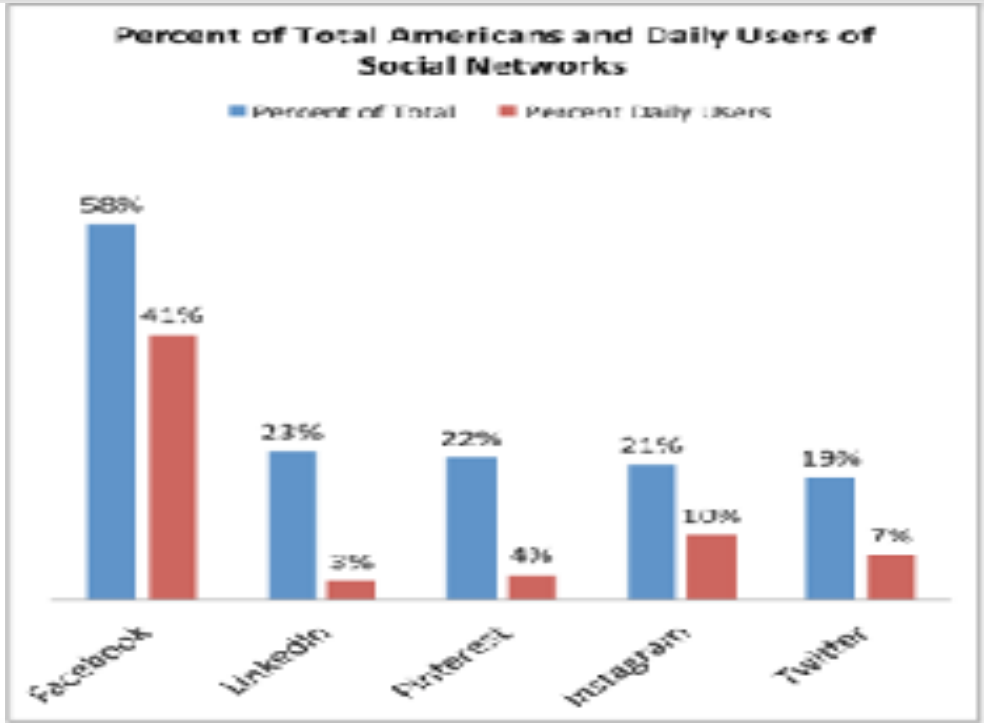
BY GABRIELLE BRUNEY MAR 17, 2018

Facebook

- American online social networking service
- Launched by Mark Zuckerberg on February 4, 2004
- 2.2 billion monthly users active users (January, 2018)
- Estimated net worth of 480 billion dollars
- Owns: Instagram, Whatsapp and Oculus VR

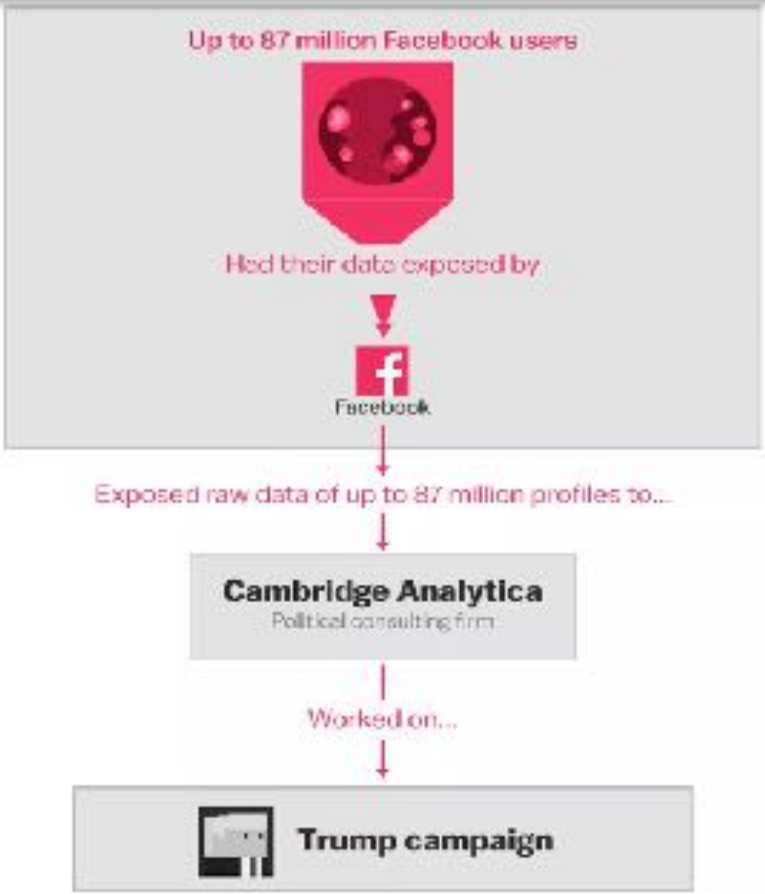
Facebook


MONTHLY USERS ON FACEBOOK 2004-2017
2000
1500
1000
500
0
2004
2006
2008
2010
2012
2014
2016


Percent of Total Americans and Daily Users of Social Networks
58%
41%
23%
3%
22%
4%
21%
10%
13%
7%
Facebook
LinkedIn
Pinterest
Instagram
Twitter

Cambridge Analytica scandal

facebook



Cambridge Analytica

China Social credit system

- Chinese government initiative for developing a national reputation system.
- Intended to assign a social credit rating to every citizen.
- Credit is evaluated from economic and social data gathered by the government.
- Main plan was issued in 2014 and will be operational by 2020.

Processing Pipeline for Social Data - Diagram

- Purchase history
- Flight destinations
- Fuel consumption
- Facebook contacts
- Internet usage
- etc..

- Cleaning data
- Aggregating data
- Data Enrichment

- Inferences
- Predictions
- Causal relations
- Principal Components

- Prove/Disprove hypothesis.
- Generalization beyond given context.

Data Acquisition

- Which kind of data do we seek?
  - Personal information (gender, age, salary...)
  - Metadata (location, browsing history…)
  - User generated content (posts, photos, shared links, music…)

Data Acquisition

- How is the data being collected?
  - Online surveys
  - Web crawlers
  - Social data from social networks
  - Surveillance

Data Acquisition - Issues

- Data collected from a social network may be private and shared without consent.
- Can we collect data about a person without him being aware of it ?
- **China’s social credit system.**
- **Facebook example**

Data Acquisition - China’s SCS

- China collects data from many different sources on its 1.4 billion citizens, with the intent of creating a “trustworthiness” rating for each one.
- Some may say a lot of the information is irrelevant for the cause.
- Examples:
  - Trash recycling
  - Charity work
  - Social media posts about the country
  - Canceling reservations
  - Behaviour in government facilities
  - And much more..

Data Acquisition - Facebook

- How much data does facebook actually keep about us?
- Average of 200MB of data (that is shared with us) for an average facebook user.
- For 2.2 billion users, a total of 440,000,000,000MB of social data is stored on facebook servers.

Facebook Data - Example

- Facebook allows us to download all of the data they accumulated on us and present it on an offline profile.
- Lets see a facebook profile...

Facebook Offline Profile

Facebook Offline Profile

Facebook Offline Profile

Facebook Offline Profile

Profile
Contact Info
Timeline
Photos
Videos
Friends
Messages
Pokes
Events
Security
Ads
Applications

Events

Attending

Data Processing

Data Processing consists of three main operations:

- Data Cleaning - ensure that the data faithfully represent the phenomenon being studied.
- Data Enrichment - adding categorical labels or tagging the data in a more complex way.
- Data Aggregation - structuring and reorganizing the data or transforming the data

Data Processing - Clean Data

- What is the definition of clean data?
- Is the data we collect already cleaned by another party?
- How much bias is inserted while processing the data?

Data Analysis

The main methods for data analysis:

- Qualitative Analyses - Answering questions like “How?” and “Why?” a phenomenon happens.
- Descriptive Statistics - Quantifying social data with numbers or graphs. These analyses allows us to measure certain properties of the data.
- Inference and Predictions - Draw conclusions beyond our dataset and make inferences about unseen or larger populations.

Data Analysis - Considerations

- Minimizing Bias - Learning bias.
- How do we differ between causality and correlation?

Data Analysis - a little example

Examples and Ethical Discussion

Facebook - Cambridge Analytica

- Cambridge Analytica - British political consulting firm which worked with the trump campaign.
- Partnered with researcher Aleksander Kogan which claimed to use facebook data for academic purposes.
- In 2014, 270,000 U.S. Citizens were asked to take a personality quiz online.
- Each quiz taker was paid around 5 dollar to take the quiz

But there was a catch...

Facebook - Cambridge Analytica

- To take the quiz users had to login to facebook.
- Loging in to facebook gave Kogan access to the users profile.
- Kogan combined the quiz results and facebook data (mainly Likes) to create a “personality” profile.
- These profiles, crossed with voter records then passed on to CA.
- Kogan and CA didn't stop there

Facebook - Cambridge Analytica

- By taking the quiz on facebook this also gave Kogan permission to collect data about the user's friends.
- Over 80 million user profiles were compiled.
- CA then used that data to target people with political messaging.
- CA said this helped the trump campaign (Trump's team denied it).

Faceb

So was

- Al … arch b
- … ust between

POLITICS

Facebook's Zuckerberg: I started this place, I run it, I'm responsible

The CEO acknowledges that millions more profiles than initially reported were passed to Cambridge Analytica. In a candid conversation with the press, he says it's all on him.

BY IAN SHERR, RICHARD NIEVA / APRIL 5, 2018 3:17 PM PDT

Facebook - Cambridge Analytica Issues

- Given personal data about a group of people, does it gives us the privilege to conduct a research?
- And if it does, who else can see it? Can it be published or used for political/ business gain?
- Does it matter who conducts the research and for what purpose?
- **After we publish something in the social network, who owns the data?**

Facebook - Emotional Study

- For one week in January 2012 over 600,000 facebook users were randomly selected to unknowingly partake in a study.
- The study's goal was to determine the effect of "emotional alteration" by Facebook posts.
- Facebook manipulated the balance between negative and positive posts seen by the users.

Facebook - Emotional Study

- After a week, the sentimental context of the posts by the users were examined.
- The researchers concluded that they have found evidence that emotions can spread through a network.

Positive Words (per cent)
5.0
5.1
5.2
5.3
5.4
Negativity Reduced
Positivity Reduced
Control
Experimental
Negative Words (per cent)
1.50
1.60
1.70
1.80

Facebook - Emotional Study Ethical Issues

- As of now there is no legislation regarding these actions,

  it's really more a question of ethics.
- **The lead researcher, Adam Kramer said** - "..because we care about the emotional impact of Facebook and the people that use our product, we felt that it was important to investigate the common worry that seeing friends post positive content leads to people feeling negative or left out."

Facebook - Emotional Study Ethical Issues

- Does the cause justify means?

A Possible solution:

- Another level of **informed consent** that Facebook and other platforms would put in front of its users when they enter them to these studies.

Gcogle Terms of Service

Last modified: October 25, 2017 (view archived versions)

Welcome to Google!

Thanks for using our products and services ("Services"). The Services are provided by Google LLC ("Google"), located at 1600 Amphitheatre Parkway, Mountain View, CA 94043, United States.

By using our Services, you are agreeing to these terms. Please read them carefully.

Our Services are very diverse, so sometimes additional terms or product requirements (including age requirements) may apply. Additional terms will be available with the relevant Services, and those additional terms become part of your agreement with us if you use those Services.

Using our Services

You must follow any policies made available to you within the Services.

Don't misuse our Services. For example, don't interfere with our Services or try to access them using a method other than the interface and the instructions that we provide. You may use our Services only as permitted by law, including applicable export and re-export control laws and regulations. We may suspend or stop providing our Services to you if you do not comply with our terms or policies or if we are investigating suspected misconduct.

Using our Services does not give you ownership of any intellectual property rights in our Services or the content you access. You may not use content from our Services unless you obtain permission from its owner or are otherwise permitted by law. These terms do not grant you the right to use any branding or logos used in our Services. Don't remove, obscure, or alter any legal notices displayed in or along with our Services.

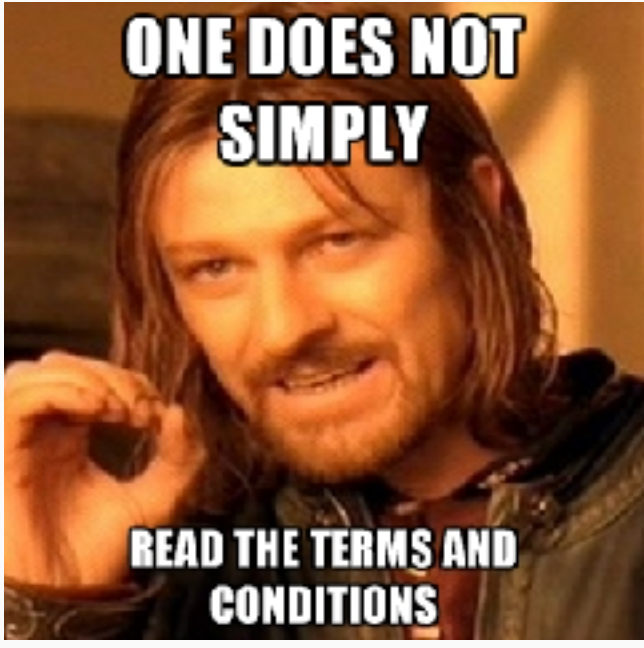
Our Services display some content that is not Google's. This content is the sole responsibility of the entity that makes it available. We may review content to determine whether it is illegal or violates our policies, and we may remove or refuse to display content that we reasonably believe violates our policies or the law. But that does not necessarily mean that we review content, so please don't assume that we do.

Terms of service

- **Terms of service** (also known as **terms of use**) - rules by which one must agree to abide in order to use a service (Wikipedia).
- Terms of service acts as an agreement between the user and the platform.
- It’s also there to provide legal coverage and protection to the platform.
- Written in a legal language and often is too long to read.

Terms of service

- ToS;dr, a temporary solution - link



Facebook

Share review: https://tosdr.org/#facebook

Very broad copyright license on your content Discussion

The copyright license that you grant to Facebook goes **beyond the requirements** for operating the service. For instance, it includes the right for Facebook to transfer the license or to license it others on their terms ("sublicense"). Also, the copyright license **does not end when you stop using the service** unless your content has been deleted by everyone else.

This service tracks you on other websites Discussion

This service uses cookies to track you even if you are not interacting with them directly. Amazon for instance, use cookies to track your device and serve targeted advertisements on other websites (Amazon associates, websites using Amazon Checkout). They "obtain certain types of information when your Web browser accesses Amazon.com or advertisements and other content served by or on behalf of Amazon.com on other Web sites".

Facebook automatically shares your data with many other services Discussion

Facebook automatically shares your information with Bing, Pandora, TripAdvisor, Yelp, Rotten Tomatoes, Clicker, Scribd, and Docs, unless you manually opt-out.

Facebook uses your data for many purposes Discussion

Including: data analysis, testing, service improvement, control of the effectiveness of the personal ads, and location features and services.

Amazon

No Class Yet

- Terms may be changed any time at their discretion, without notice to the user
- This service tracks you on other websites
- Amazon enables third-party advertisers to target you by default
- Amazon may sell user data as part of a business transfer

More details

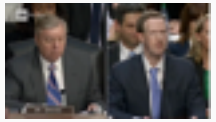
Google

Class C

- Google keeps your searches and other identifiable user information for an undefined period of time
- This service tracks you on other websites
- Google can use your content for all their existing and future services
- Partial archives of their terms are available
- Google may collect your device fingerprint.

More details

Terms of service - Ethical Issues



- Most companies admit users often skip reading the ToS.
- The implications of agreeing with the ToS are not understood by the average consumer.
- How far can the ToS expand the platform’s control over their users data?

Facebook admits it did not read terms of the app that harvested data of 87 million

- "We did not read all of the terms and conditions," CTO Mike Schroepfer told U.K. lawmakers at a parliamentary committee hearing.
- Schroepfer said that Facebook did not notify the U.K.'s data protection watchdog after it learned of the sharing of data with Cambridge Analytica.
- He said it was a "mistake" that Facebook hadn't notified users of the data leak earlier than it did.

Ryan Browne | @Ryan_Browne_
Published 6:41 AM ET Thu, 26 April 2018 | Updated 8:07 AM ET Thu, 26 April 2018

CNBC

A Terrifying Story

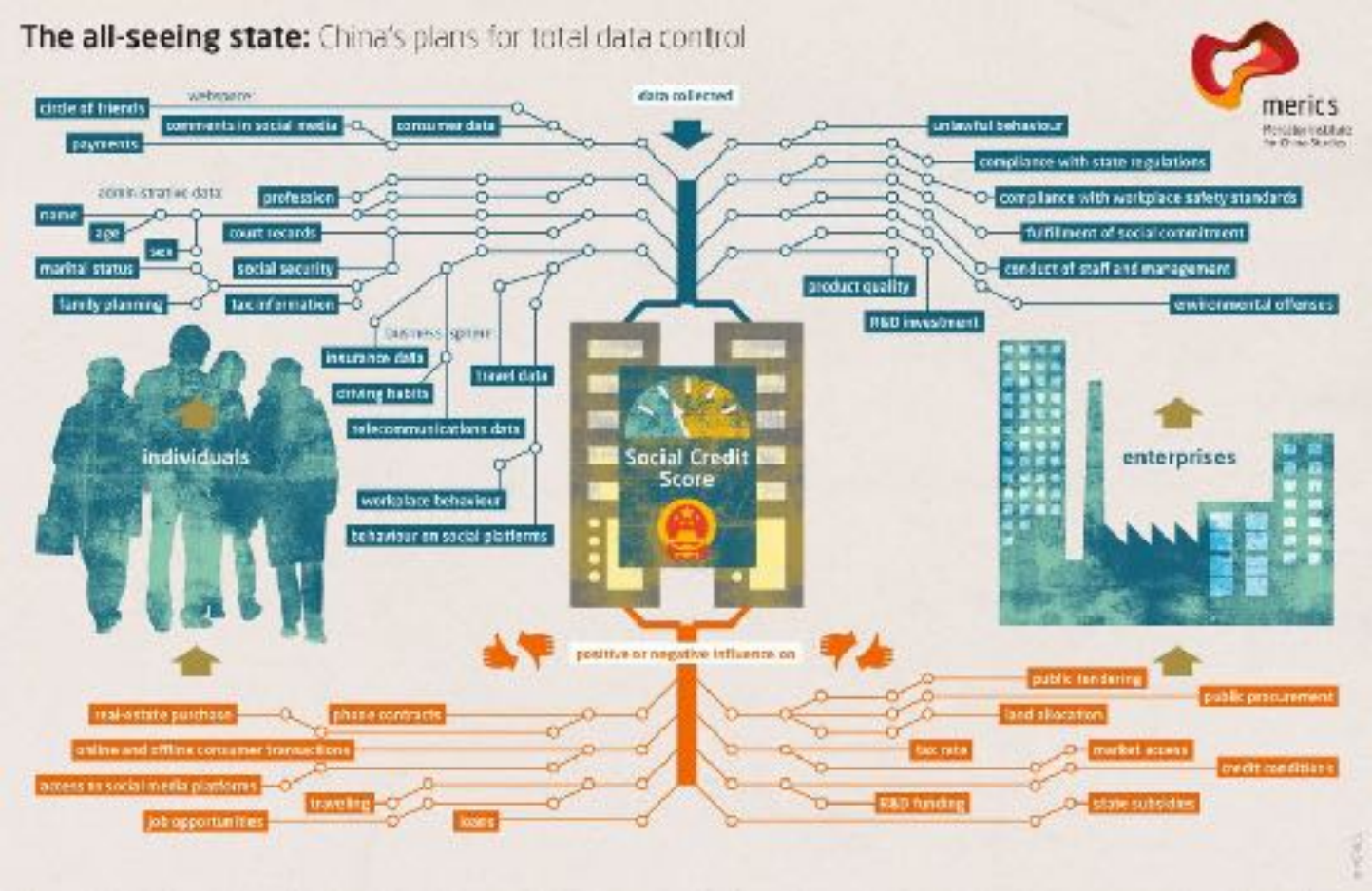
- In 1949, George Orwell published a dystopian novel called 1984.
- The novel is set in a province of the superstate Oceania.
- Residents are victims of omnipresent government surveillance and public manipulation.
- The Government persecutes individualism and independent thinking, which are regarded as "thoughtcrimes".
- The tyranny is overseen by a mysterious leader known as **Big Brother**.

China Social credit system

- Chinese government initiative for developing a national reputation system.
- Intended to assign a social credit rating to every citizen.
- Credit is evaluated from economic and social data gathered by the government.
- Main plan was issued in 2014 and will be operational by 2020.

China Social credit system

China Social credit system



China Social Credit System - Sesame Credit app

Sesame Credit app

China Social credit system - Any good?

“Credit will ensure that the **bad** people in society don’t have a place to go, while **good** people can move freely and without obstruction.” **Lucy Peng, Alibaba chief executive**.

- A tool of governance to steer the behavior of citizens in the chinese market.
  - Prevent fraudulence and counterfeit goods.
  - Fight food safety issues.
  - Incentivize a government defined “good behavior” in public space.

China SCS - The Good, the Bad and the Ugly

- Is it so easy to decide who is a good person?
- Can you decide?
- What are the “components” of a good individual?
- An algorithm to decide who is good and who is bad.

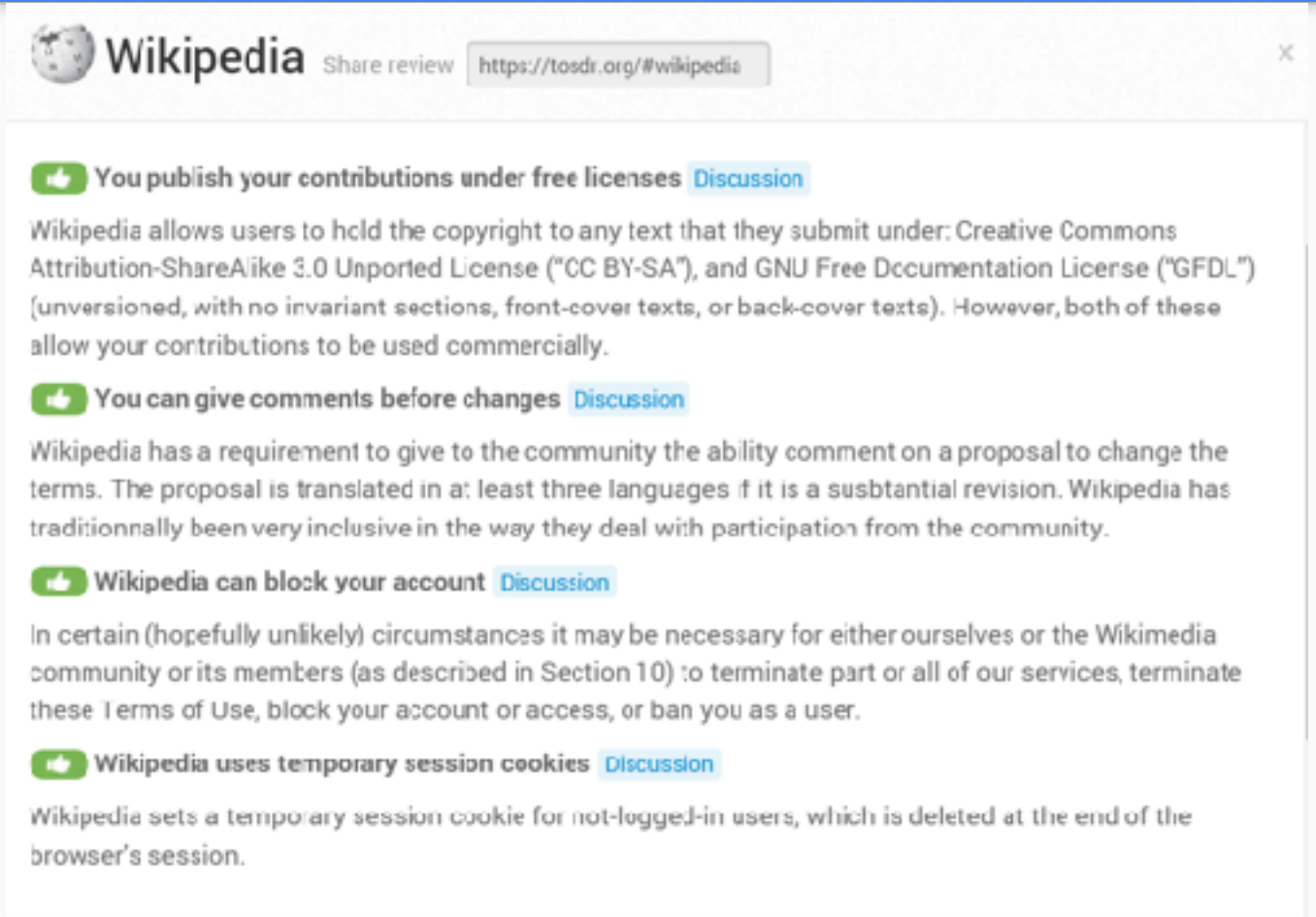
China Social Credit System- Ethical issues

- Law abiding citizen=”good” person.
- Doesn’t leave room for people to make mistakes.
- Gathers private information and analyzes it without any consent.
- Can the government compel companies to share their data?
- Where is the line between privacy and security?

Crowdsourcing Advantages

- Waze - “Get the best route, every day, with realtime help from other drivers”
- Wikipedia - The Free Encyclopedia.
- Crowdsourcing disease discovery and research

Terms of Service - Wikipedia



Summary

- Data - The new black gold.
- Social data - Huge amounts of it are available and free to use without any consent.
- As a developer - Data is one of the most valuable resources you have.
- As users - No free meals.

Summary - No Free Meals

- An average person (25-30) daily uses: Google, Waze, YouTube, Facebook, Instagram and Whatsapp.
- Google's revenue for 2017: 109.65 Billion dollars.
- Facebook's revenue for 2017: 40.6 Billion dollars.
- How much do you pay for these services?
- **"If you're not paying for the product, you're the product"**

Possible Solutions

- ToS - A better understanding of which data is collected (in tl;dr way).
- Transparency - How is the data being analyzed/shared and for what purposes.
- Shared revenue policy - If users data/generated content is used for business purposes, they should make a profit as well (YouTube Monetization Program).

Questions?

Further Reading

- Experimental evidence of massive-scale emotional contagion through social networks, http://www.pnas.org/content/111/24/8788.full
- https://www.nytimes.com/2018/04/08/us/facebook-users-data-harvested-cambridge-analytica.html
- https://www.theguardian.com/technology/2014/jun/30/facebook-emotion-study-breached-ethical-guidelines-researchers-say

Further Reading

- https://www.cnbc.com/2018/04/26/facebook-cto-admits-firm-didnt-read-terms-of-aleksandr-kogans-app.html
- https://digitalanalyticscourse.nl/2018/02/20/big-brother-is-watching-you/
- Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries. Alexandra Olteanu, Carlos Castillo, Fernando Diaz, Emre Kıcıman.