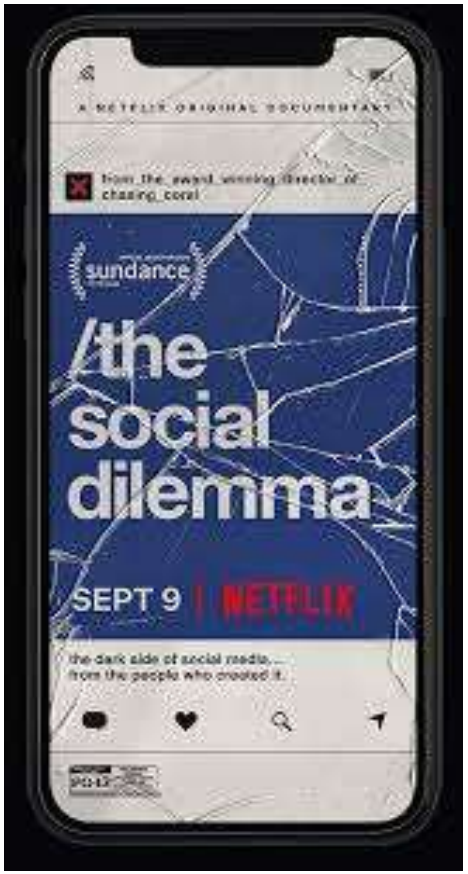
Ethical Issues in IR

Ethical Issues

- IR technology raises a wide variety of ethical issues
  - Privacy
  - Fairness
  - Disinformation
  - Internet addiction
  - Filter bubble
  - Radicalization

Recent Documentaries

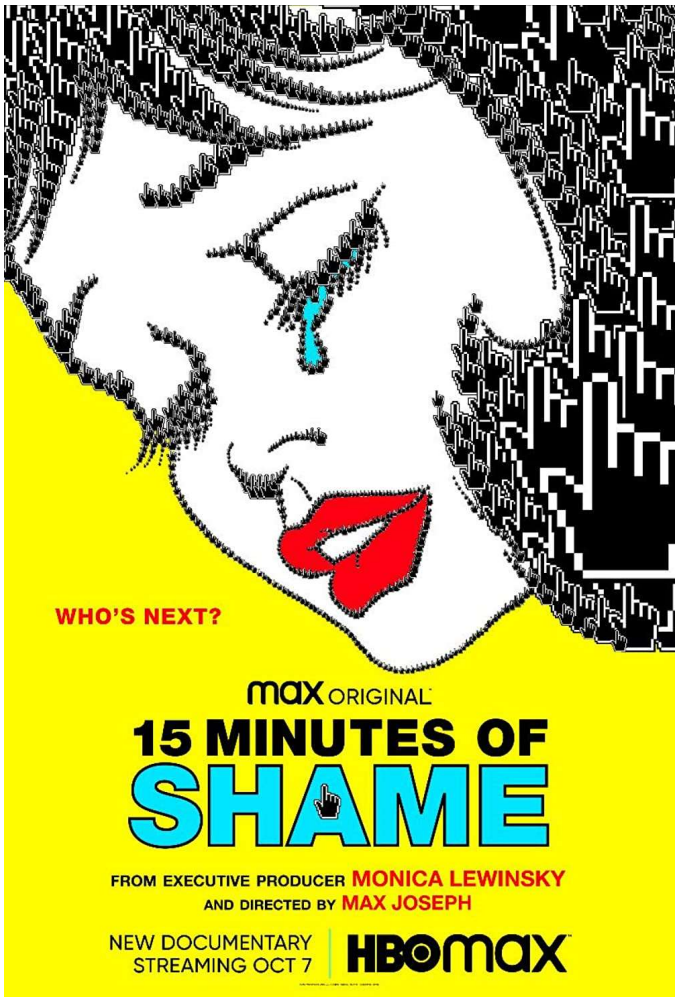
Explores the dangerous human impact of social networking, with tech experts sounding the alarm on their own creations.



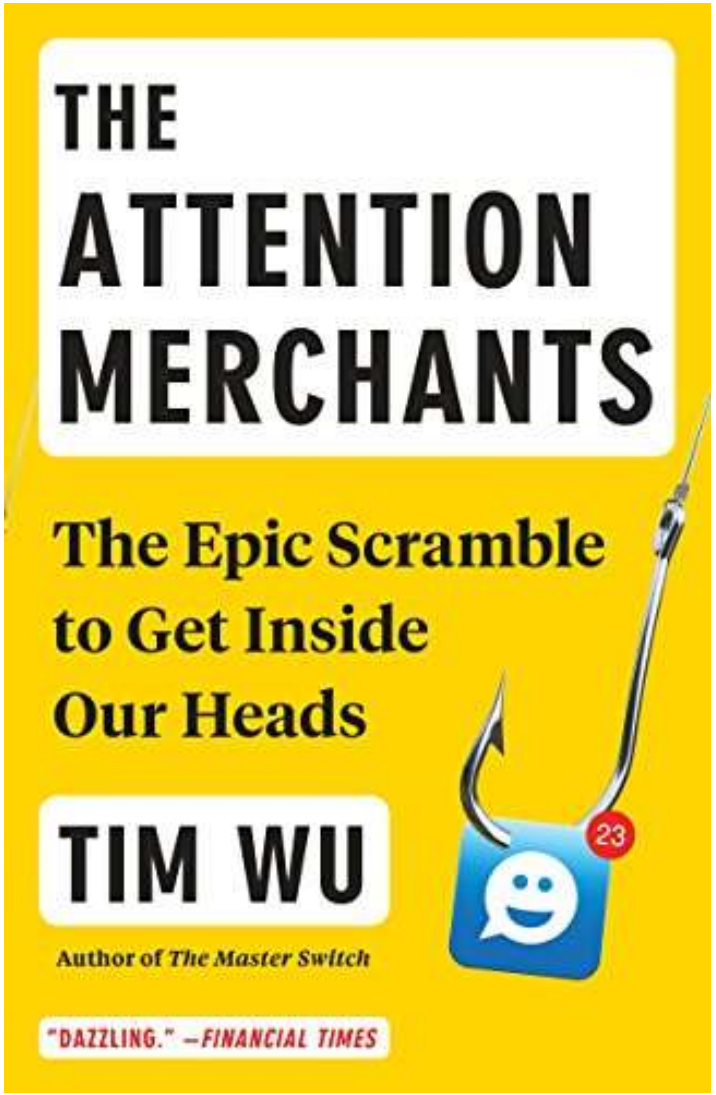
- Explores the Cambridge Analytical company that mined data from Facebook to impact the Brexit and 2016 Presidential elections.

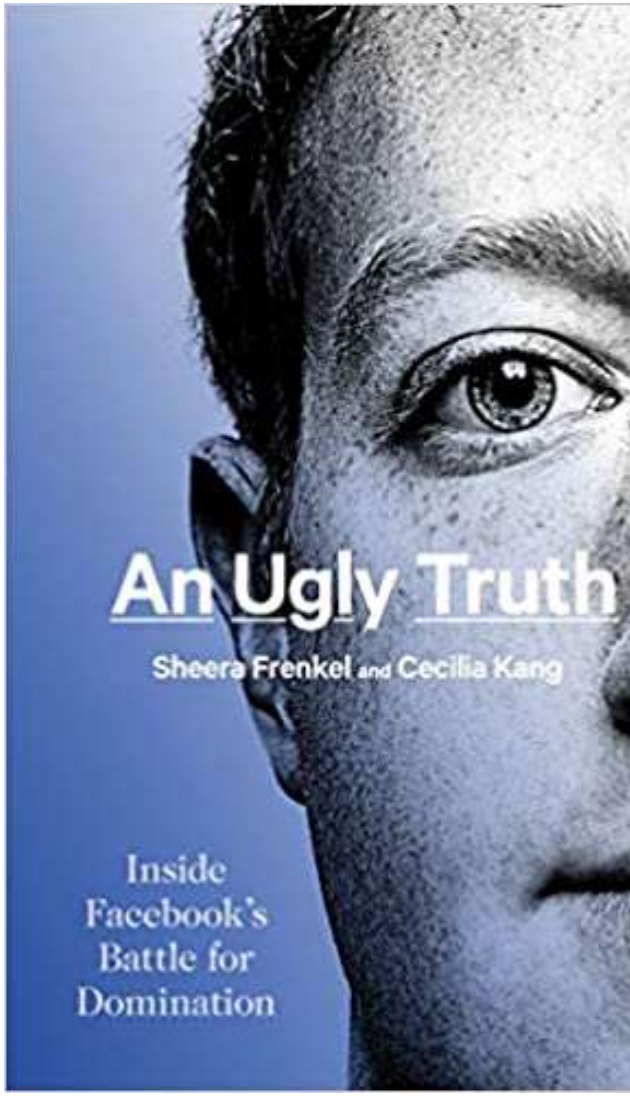
Recent Documentaries

Explores how
ocial media has
xacerbated
ssues with
cancel culture”

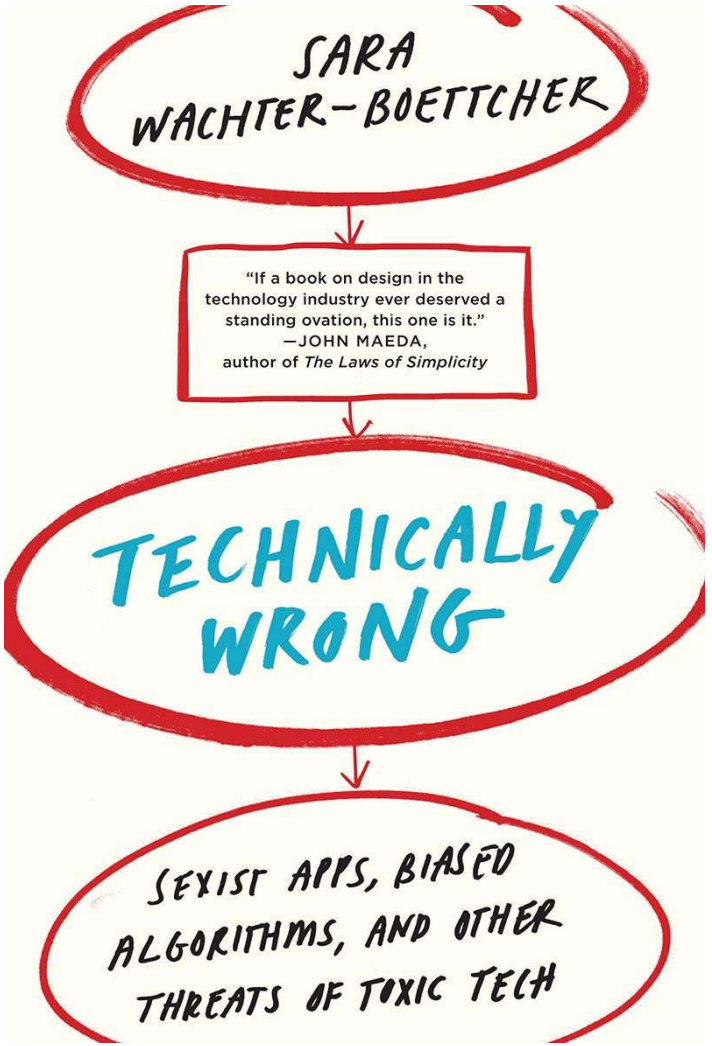


Relevant Recent Popular Science/Tech Books





Relevant Recent Popular Science/Tech Books



Privacy

- Search engines, social media sites, etc. collect enormous amounts of person information about their users.
- Data is sold to advertisers in various forms to allow “micro targeting.”
- Very few people read “terms of use” or change privacy settings from defaults.
- Some sites like DuckDuckGo do not collect user data nor personalize results.

Deanonymization

Even if user IDs are removed, they can frequently be inferred from other disclosed information.

- In mid-90's Mass. released anonymized data on state employee health records, and MIT PhD student identified the governor's unique record by combining info on birthday, sex, and zip code.
- When AOL released search query records in 2006 for anonymized users, specific users were identified thru information in their queries.
- Netflix Prize data was deanonymized in 2006 by UTCS researchers by matching ratings vectors with public film ratings on IMDB (*Narayanan, Arvind; Shmatikov, Vitaly "How To Break Anonymity of the Netflix Prize Dataset")*

Randomized Response

oal is to produce accurate population statistics without ompromising privacy of individuals.

xample:

- Ask a man whether he has ever had sex with a prostitute. Before he answers, ask him to flip a coin. Instruct him to answer "yes" if the coin comes up tails, and truthfully, if it comes up heads.
- Half of people will answer "yes" regardless of whether they have done The other half will answer truthfully. Whatever proportion of the group said "no", the true number who did not have sex with a prostitute is do that, because of the law of large numbers we assume the two halves are probably close to the same as it is a large randomized sampling.

Differential Privacy

dd random noise to any computation from private data to ensu
at no individual's data can be reverse engineered from the res

nsure that when using released information the probability of a
onclusion about any individual does not change much if their d
removed from the original dataset.

Even if Bob's data was not included in the Netflix data we could still u
the data to infer that if he watched Star Wars Episode II he probably
watched Episode I. But we could not use it to accurately infer that he
probably watched a pornographic film. If we do release info that allow
us to infer this otherwise uninferable fact, we violate differential priva

Fairness

- Search results or recommendations could exhibit various types of bias, such as political or cultural bias.
  - A search for a politically sensitive topic such as “abortion” or “gay marriage” could present results all from a particular perspective.
- Including come sort of diversity component to the ranking score could be useful to ensure that each subsequent result is significantly different from previous results.
- How do we measure “fairness” and how can algorithms prevent inadvertent biases?

Fairness in Machine Learning Systems

- Many machine learning systems exhibit inadvertent biases based on unrepresentative training data.
  - Face recognition systems have a higher error rate for black individuals because training data did not contain a representative fraction of black faces.
- Mechanisms for statistically guaranteeing similar test error rates on particular "protected classes" such as sex, race, religion, and national origin.

Disinformation

The Internet contains a wide variety of information, some of which is ju
naccurate and unsupported by reliable facts, and some is intentional
nisinformation intended to mislead people for political or other purpose

How can we automatically detect "fake news" and "fairly" filter it or
nnotate it with reliable warnings.

Given human-expert labeled training data, could use text categorization
etect some types of misinformation.

- Might detect know types of misinformation represented in the training data such "vaccines cause autism" or ones that use exaggerated expressions.
- Hard to detect novel forms of misinformation such as those analyzed by "fact checking" organizations such as PolitiFact and Snopes.

Section 230 of the Communications Decency Act

- Provides immunity from liability for providers and users of an "interactive computer service" who publish information provided by third-party users:
  - No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider.
- But when recommendation algorithms “push” certain sites that attract attention (and “fake news” tends to attract attention), aren’t they behaving more like publishers who should be held responsible for the material they promote?

Fake Review Detection

- One form of deception detection that has been widely studied is classifying product/hotel/etc. reviews as real or fake.
- Hard to get labelled data, some studies use solicited crowdsourced fake reviews.
- Automated text classification does fairly well at identifying fake reviews by finding linguistic patterns in fake and real reviews, fake reviews tend to be very general and abstract and real ones mention very specific, concrete details.

Internet Addiction

- Effective search and recommendation methods encourage continued engagement and screen time and can lead to addictive behavior.
- Machine learning over massive user data is used to optimize every aspect of information systems to maximize continued use.
- Internet companies use A/B testing to optimize every aspect of information systems to maximize engagement.

Filter Bubble

- Recommendation algorithms and personalized search and newsfeeds reinforce individual biases.
- Users become separated from information that disagrees with their viewpoints, effectively isolating them in their own cultural or ideologic bubbles.
- *Bill Gates 2017: (*Technology such as social media) “lets you go off wi like-minded people, so you're not mixing and sharing and understandin other points of view ... It's super important. It's turned out to be more of problem than I, or many others, would have expected.”
- Algorithms that encourage presentation of a more diverse set of information might help to battle this effect.

Radicalization

- Extreme political/religious beliefs can be reinforced by recommendation and personalization.
- Can lead to "self radicalization" and potentially violent behavior.
  - Pizza gate conspiracy (predecessor to QAnon) led to Edgar Maddison Welch, a 28-year-old man from Salisbury, North Carolina, on December 4, 2016 going to Comet Ping Pong and firing three shots from an AR-15 style rifle.