CS 4001: Computing, Society & Professionalism

Munmun De Choudhury | Assistant Professor | School of Interactive Computing

Week 10: Trust and Algorithmic Transparency

March 12, 2019

Ethics of Algorithms

Algorithms, Control, and Manipulation

- What happens when powerful communication tools (e.g., social media, digital technologies) are exploited to
  - Manipulate social engagement and discussion
  - Change the public perception
  - Influence what we see, read, consume, and learn
  - Impact real world outcomes and events, e.g., politics

Case Study 1

Reasoning about invisible algorithms in the News Feed

- What we see on social media platforms is not the raw information, but rather curated by “invisible” algorithms
  - E.g., Facebook’s News Feed
- These algorithms shape (even manipulate?) users’ experiences but many users remain unaware of their presence
- Study with 40 users about their perceptions of the News Feed algorithm

Research Questions

- **RQ1.** How aware are users of the News Feed curation algorithm and what factors are associated with this awareness?
- **RQ2.** How do users evaluate the curation of their News Feed when shown the algorithm outputs? Given the opportunity to alter the outputs, how do users' preferred outputs compare to the algorithm's?
- **RQ3.** How does the knowledge users gain through an algorithm visualization tool transfer to their behavior?

Findings

- More than half (62%) were not aware of the presence of the algorithm
- Initial reactions were surprise and anger
- Developed a system FeedViz to show the differences between the raw feed and the curated feed
- Users were upset when content from close friends and family were not shown
- Missing stories attributes to friends' decision to exclude them
- Longitudinal study (2-6 months later), algorithmic awareness led to more more active engagement on the platform

"It's kind of intense, it's kind of waking up in 'the Matrix' in a way. I mean you have what you think as your reality of like what they choose to show you. [...] So you think about how much, kind of, control they have..." (P19).

"I feel like I'm a mouse, a little experiment on us. To me, that's the price I pay to be part of this free thing. It's like we're a part of their experiment and I'm okay with it" (P21).

"I have like 900 and some friends and I feel like I only see 30 of them in my News Feed. So I know that there's something going on, I just don't know what it is exactly" (P26).

"[My friends] all don't get to see everything, and I've always been suspicious of [Facebook], on how they choose who gets to see it, who doesn't" (P28).

"I have never seen her post anything! And I always assumed that I wasn't really that close to that person, so that's fine. What the hell?!" (P3).

"Well, I'm super frustrated [pointing to a friend's story], because I would actually like to see their posts" (P3).

"I think she needs support for that; if I saw it, then I would say something [to support her]" (P8).

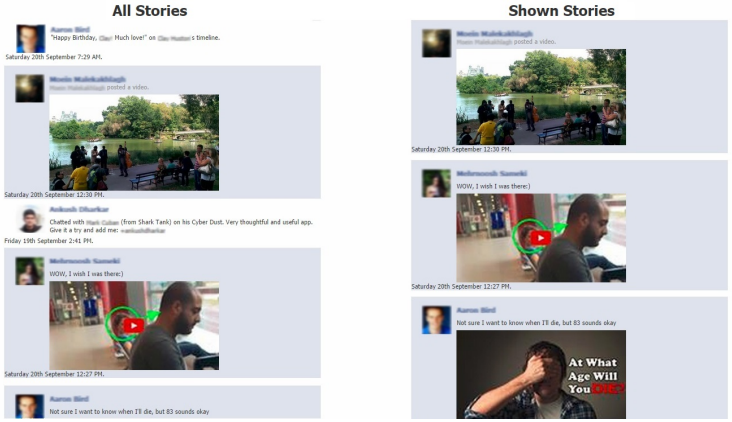


Figure 1. The Content View. Shown stories (in blue) occur across both columns, while the hidden stories (white) appear only in the left column as 'holes' in News Feed. Stories appear in reverse chronological order.

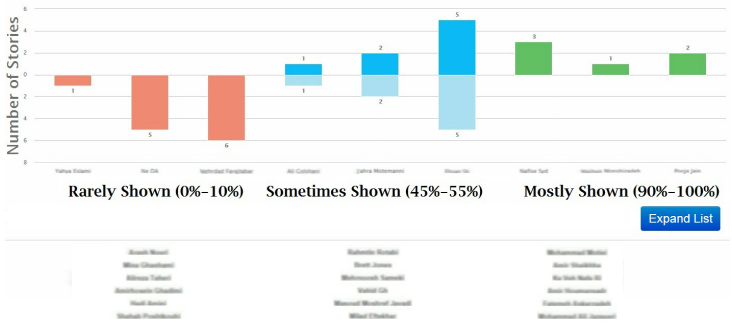


Figure 2. The Friend View. "Rarely shown" includes friends whose stories were mostly hidden (0%-10%) from the user. "Sometimes shown" includes friends who had roughly half of their posts (45%-55%) shown to the user. "Mostly shown" includes those friends whose stories were almost never filtered out (90%-100%) for the user. The number of the shown stories is displayed above the x-axis and the number of hidden stories is below the x-axis. The expand button augments the three category lists below the chart.

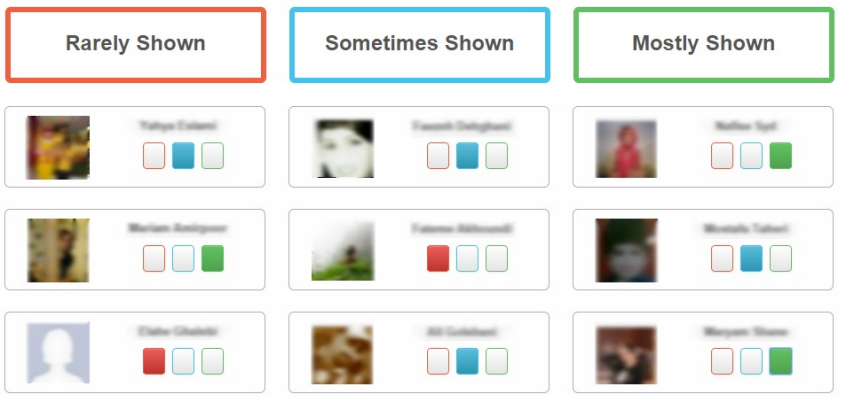


Figure 3. The Friend Rearrangement View. User can move friends between the categories by changing the color of a friend to the destination category's color.

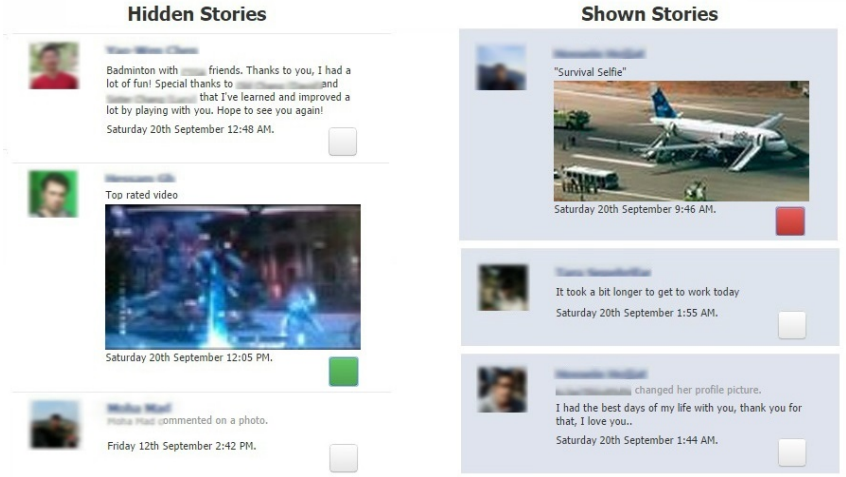


Figure 4. The Content Rearrangement View. User can move a story from its original category to the other by clicking the button beside each story.

A View from **Karrie Karahalios**

Algorithm Awareness

How the news feed on Facebook decides what you get to see.

October 21, 2014

Increasingly, it is algorithms that choose which products to recommend to us and algorithms that decide whether we should receive a new credit card. But these algorithms are buried outside our perception. How does one begin to make sense of these mysterious hidden forces?

The question gained resonance recently when Facebook revealed a scientific study on "emotion contagion" that had been conducted by means of its news feed. The study showed that displaying fewer positive updates in people's feeds causes them to post fewer positive and more negative messages of their own. This result is interesting but disturbing,

Connected by the Cloud

TIME

THE CLOUD | MAGAZINE | VIDEOS

Connected by the Cloud

The Cloud

What Exactly Is The Internet of Things?

This Budding Startup Is Changing How You Buy Flowers

Why This Is the Most Important New Technology in Decades

BUSINESS FACEBOOK

Here's How Facebook's News Feed Actually Works

Victor Luckerson @VLuck | July 9, 2015

How a controversial feature grew into one of the most influential products on the Internet

There are two very important rooms that will help determine the

Why News Feed?

- Over the past nine years, the product, which was initially controversial, has evolved into the most valuable billboard on Earth—for brands, for publishers, for celebrities and for the rest of us.
  - The feed must be completely personalized but still highly engaging to Facebook's users so they'll keep coming back and seeing more ads from the company's 2 million advertisers.
  - Facebook says the average user has access to about 1,500 posts per day but only looks at 300.

Why News Feed?

- For years, the News Feed has been fueled by automated software that tracks each user's actions to serve them the posts they're most likely to engage with.
  - That proved successful in helping News Feed generate more revenue for Facebook than any other part of the site.
- But it's also led to a growing anxiety about how much Facebook knows, and how the company can use that knowledge to influence what users buy, how they vote, even how they feel.

In defense of algorithmic curation

Class Discussion Point 1

All platforms use algorithmic curation that is invisible to the user – Netflix was the first. Do you expect such shock if the same study was done for the algorithms in Netflix and Amazon? Why or why not?

Class Activity 1

Google's Face Recognition Controversy

- 2015: Google Photos Tags Two African-Americans As Gorillas Through Facial Recognition Software
- The incident: When Brooklyn-native Jacky Alcine logged onto Google GOOGL +1.75% Photos on Sunday evening, he was shocked to find an album titled "Gorillas," in which the facial recognition software categorized him and his friend as primates.
- Others 1: It is important to note that African-Americans are not the only group mislabeled by Google Photos. Until recently, Google Photos was confusing white faces with dogs and seals.

Face Recognition Controversy

- Others 2: Flickr's facial recognition software labeled both black and white people as “animals” and “apes” (these tags were promptly removed).
- Others 3: Many Native American dancer photos were tagged with the word “costume,” which added great insult to the community.
- Others 4: Back in 2009, Nikon's face-detection cameras were accused of being “racist.” Many times, when an Asian face was photographed, a message flashed across the screen asking, "Did someone blink?” — even when their eyes were wide open.

Google's Solution

- Google said it was "appalled" at the mistake, and promised to fix the problem.
- The company blocked its image recognition algorithms from identifying gorillas altogether — preferring, presumably, to limit the service rather than risk another miscategorization.
- Google had restricted its AI recognition in other racial categories. Searching for "black man" or "black woman," for example, only returned pictures of people in black and white, sorted by gender but not race.

Class Discussion

- Is it really a fix?

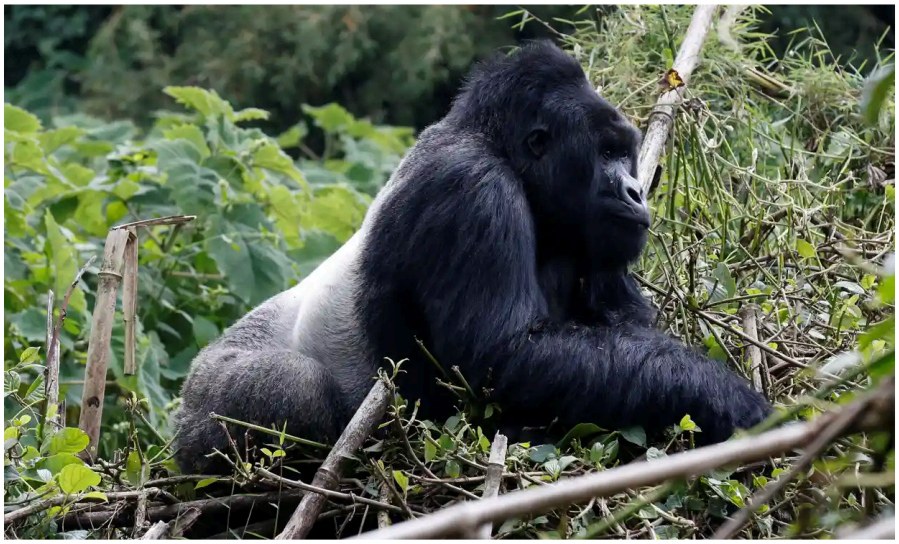
Google

Google's solution to accidental algorithmic racism: ban gorillas

Google's 'immediate action' over AI labelling of black people as gorillas was simply to block the word, along with chimpanzee and monkey, reports suggest

Alex Hern
@alexhern

Fri 12 Jan 2018 11.04 EST



A silverback high mountain gorilla, which you'll no longer be able to label satisfactorily on Google Photos. Photograph: Thomas Mukoya/Reuters

After Google was criticised in 2015 for an image-recognition algorithm that auto-tagged pictures of black people as "gorillas", the company promised "immediate action" to prevent any repetition of the error.

Wired's Investigation in 2018

- Wired tested more than 40,000 images of animals on the service.
- Photos accurately tagged images of pandas and poodles, but consistently returned no results for the great apes and monkeys – despite accurately finding baboons, gibbons and orangutans.

Class Discussion

- What else could have been done by Google?

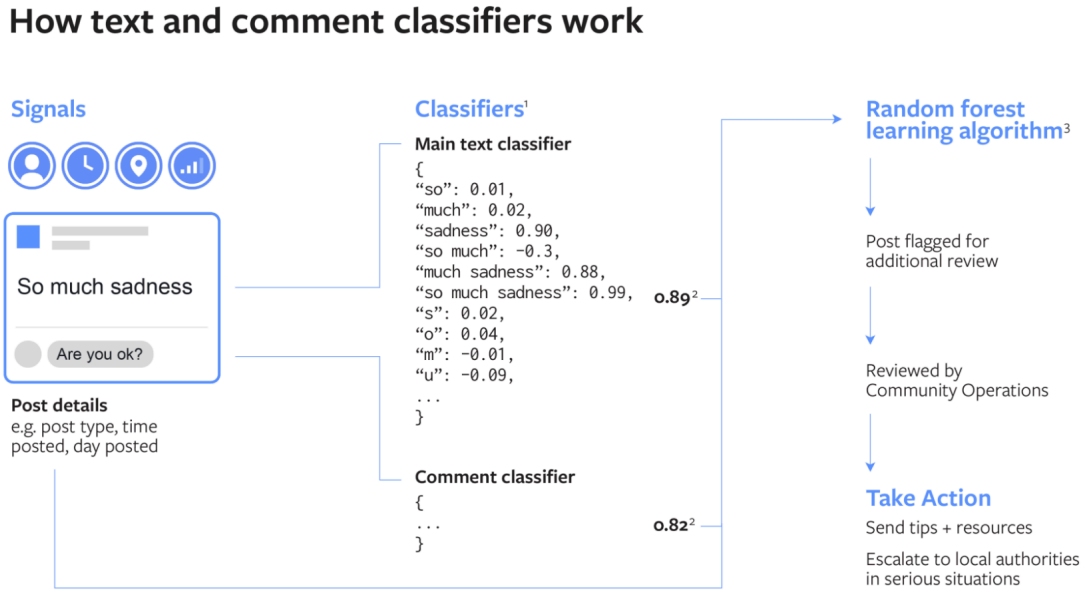
Such technologies are frequently described as a “black box”, capable of producing powerful results, but with little ability on the part of their creators to understand exactly how and why they make the decisions they do.

Technology used in self-driving cars has a racial bias that makes autonomous vehicles more likely to drive into black people, a new study claims.

Main Findings

- The researchers said they undertook the study after observing higher error rates for identifying certain demographics by existing image recognition systems.
- Tests on eight image-recognition systems found this bias held true, with their accuracy proving five per cent less accurate on average for people with darker skin.
- To prove the hypothesis, the scientists divided a large pool of pedestrian images into two groups of lighter and darker skin using the Fitzpatrick scale – a scientific way of classifying skin color.
- Even when changing the time of day or obstructing the image-detection systems view, the average accuracy remained the same.

Class Activity 2: Facebook Suicide AI



(1) Various combinations of words can have positive or negative impact on the classifier's confidence that the content includes a suicidal thought.
(2) The classifiers score the content based on how closely correlated the main text and comments are to previously confirmed cases of suicidal expression.
(3) The classifier scores, combined with other numerical details (e.g. time of day, day of week) are inputted into a "random forest learning algorithm," a type of machine learning that specializes in numerical data. It uses many decision trees and outputs the mean prediction of the individual trees.