Choices and Consequences in Computing

INFO 1260 / CS 1340

Lecture 7: Section 230 and Personalization

February 5, 2024

Remember the 1990s?

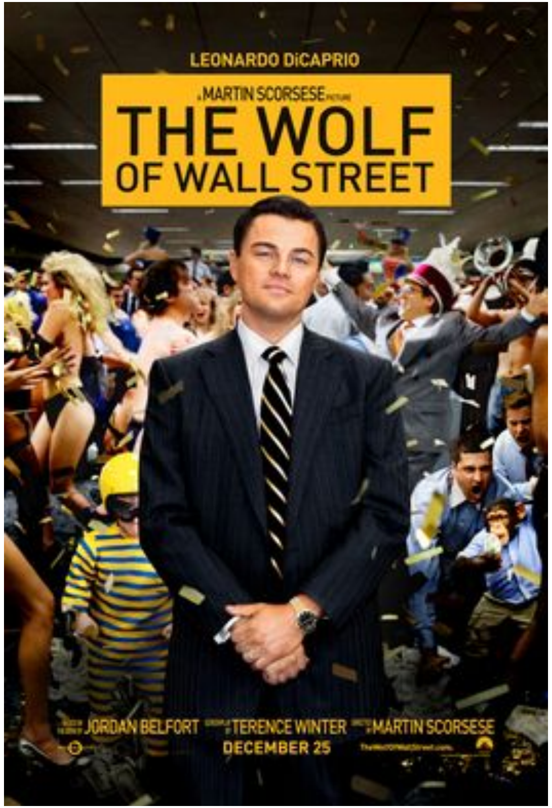
- The web exists; early days of search engines. Social networks as we know them now don’t yet exist. People more commonly getting internet in their homes.
- Early online services like CompuServe, AOL, Prodigy – proto-social networks that included content recommendations, search, bulletin boards, etc.

People start defaming each other on the early internet…

- Cubby v. CompuServe (1991): CompuServe provides a newsletter called Rumorville; it alleges that Cubby's new startup (a competitor rumor site) is a "scam" and that its founder was fired from a previous job
- CompuServe didn't have stated rules or moderate content
- Is CompuServe liable for defamation (assuming the content is defamation)?
- Court: no. CompuServe is like a <u>newsstand</u>; it's just distributing information and can't be expected to censor content

People start defaming each other on the early internet…

- Stratton Oakmont v. Prodigy (1995): a user posts on a financial bulletin board that Stratton Oakmont committed financial fraud
- Prodigy does have community rules and moderate user content
- Is Prodigy liable for defamation (assuming the content is defamation)?
- Court: <u>yes</u>. Prodigy is like a <u>newspaper</u>; it is treated as a “publisher” because of its “own policies, technology, and staffing decisions … to gain the benefits of editorial control.” (remember NYT v. Sullivan)



This doesn't seem ideal!

<u>CompuServe:</u>

No moderation

Treated like a newsstand

Less liability

<u>Prodigy:</u>

Moderation

Treated like a newspaper

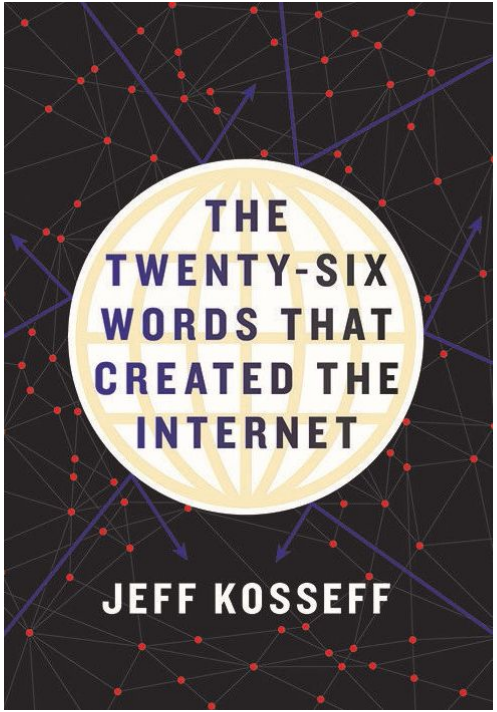
More liability

Clearly this incentivizes platforms <u>not</u> trying to moderate! Not great for the growth of the internet, especially as schools/families begin to try to make use of it as a resource.

So Congress steps in to clear this up: Section 230 of the Communications Decency Act of 1996.

Section 230: the 26 words that created the internet

- “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.” **(the “shield” provision)**
- “No provider or user of an interactive computer service shall be held liable on account of … any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected.” **(the “sword” provision)**
- In other words: platforms <u>can</u> moderate if they want to, and doing so doesn’t take away their protection against liability!



Why 230?

- To promote innovation and the growth of the internet
- To encourage platforms to moderate content! (a “good Samaritan” law)
- Some narrow exceptions to 230:
  - Federal criminal law, intellectual property law, electronic communications privacy law
  - 2018 amendment (FOSTA-SESTA): exception for material that promotes sex trafficking
    - Many deride this as making sex work much more dangerous because it drives it offline

As a result…

[note: harassment, sexual violence]

As a result…

- Dating platform not liable for defamation when a user made a fake, offensive profile about an actress from Star Trek: Deep Space Nine, including home address; platform would not remove (Carafano v. MetroSplash)
- eBay not liable for sales of forged autographs (Gentry v. eBay)
- Yelp not liable for false and malicious user reviews (Hassell v. Bird)
- Facebook not liable for algorithmically promoting content from a terrorist organization (Force v. Facebook)
- Grindr not liable when a person used it to impersonate his ex, describing rape fantasies; sent 1000+ men to his house to demand sex. Grindr would not remove even though the victim asked them to 100 times (Herrick v. Grindr)
    - Grindr’s lawyer: “a neutral system … open to good users and to bad users”
    - Grindr was really the only party that could stop this

Additional possible drivers of change to 230

- Some dislike from both right and left
- Reasons:
  - Perception of political bias
  - Perception that innovation rationale no longer holds
  - Concern about tech companies' power to control speech
  - Proliferation of ugly speech online
    - Though note much of this is "lawful, but awful" so 230 doesn't necessarily affect it anyway

Social media liability law is likely to be reviewed under Biden

Section 230 has become a favorite target of President Trump. Democrats have their own gripes about it.

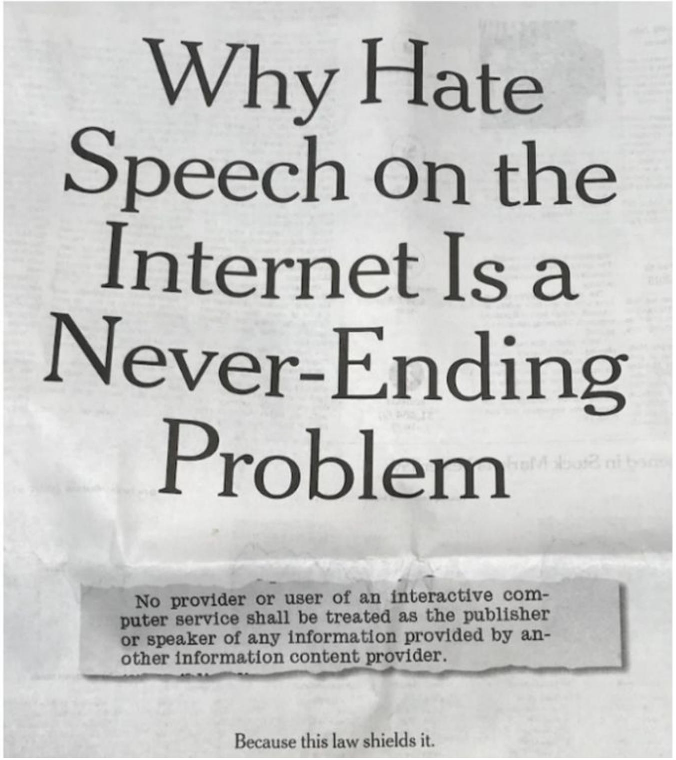
Donald J. Trump
@realDonaldTrump

REPEAL SECTION 230!!!

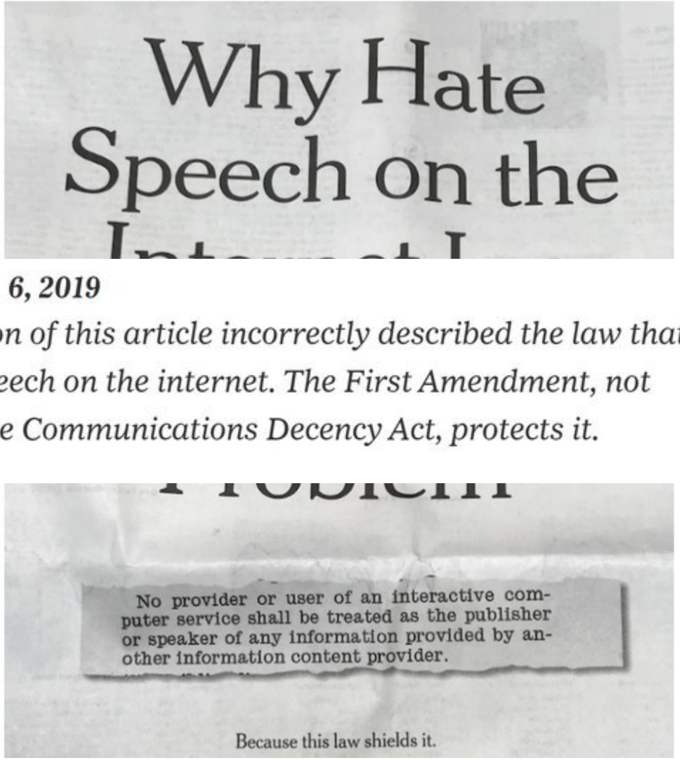
12:08 PM · Oct 6, 2020 · Twitter for iPhone

2.3K Retweets **228** Quote Tweets **7.3K** Likes

230 remains widely misunderstood



230 remains widely misunderstood



Correction: Aug. 6, 2019

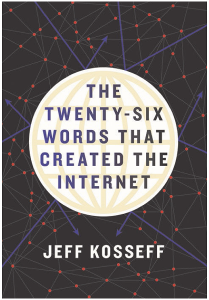
An earlier version of this article incorrectly described the law that protects hate speech on the internet. The First Amendment, not Section 230 of the Communications Decency Act, protects it.

What would the internet look like without (the current version of) 230?

- It depends what replaces it!
- If no protection against liability: how could platforms deal with the flood of user content?
- Some argue that we could make 230 protection contingent on reasonable content moderation policies (oversight, reporting procedures, labor protections for moderators, etc.)
- Next week we'll talk more about what platforms actually do to moderate "bad behavior" by users

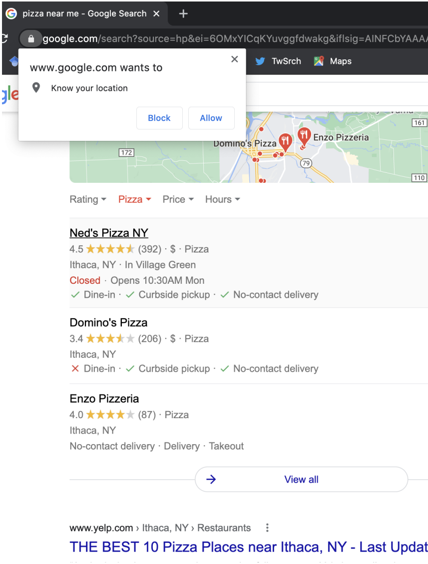
Reflections: Section 230

- The different forces acting on collective attention interact with theories of free speech.
- So far, all of this has treated the audience for the content as mainly homogeneous.
  - Collectively, it's either easier or harder for us to find a piece of content.



Personalization

- We don't all see the same things, even on the same site or in response to the same queries.
- A useful question to ask yourself regularly on-line:
  - "Why am I seeing this?"
- In many situations, it would seem strange not to personalize.
  - Would it make sense to show everyone the same set of search results for the query, "Pizza near me"?

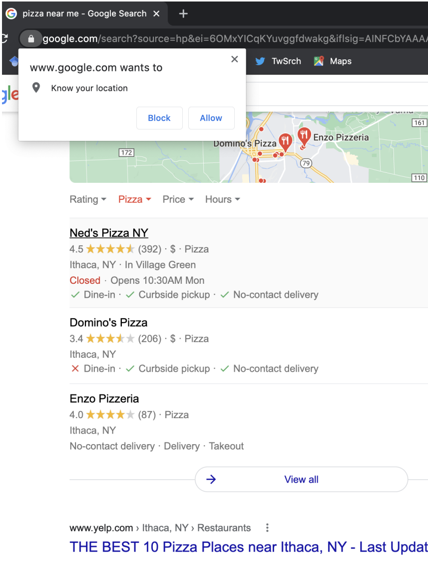


Q: What are some of the basic information sources that Web sites can use to personalize content to you?

Personalization

We all form intuitions about whether the page we're seeing has been personalized to us or not.

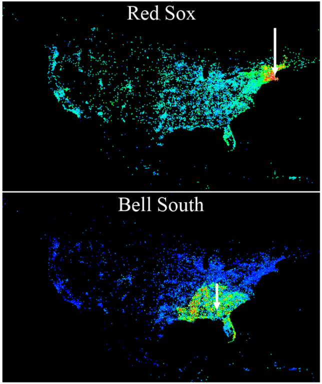
- A search for "pizza near me"
- A search for "New York City"
- Purchase choices on a site that sells iPhone cases
- The Wikipedia page for Cornell
- The r/Cornell Reddit page
- The YouTube main page
- The New York Times main page



Personalization

Useful information for personalizing includes:

- Geographic location of the current access
- Information you've filled out in a profile
- Past history:
    - Searches
    - Clickthroughs
    - Transactions
    - Dwell time on specific pieces of content
- All of this information for other users that you have contacts with (communication, social links)
- All of this information for other users you seem similar to.
- ... ...



Building a Model of the User

	Source 1	Source 2	Source 3	Source 4
User 1	.1	.2	.3	.4
User 2	.08	.16	.24	.32
User 3	.05	.1	.15	.2
User 4	.02	.04	.06	.08
User 5	.01	.02	.03	.04

When user i arrives at the platform, they choose one item to read.

- The probability the item they choose is from source j is $p(i,j)$.
- And with some probability they choose nothing.
- A table of $p(i,j)$ values. Can we summarize it more succinctly?

A latent factor model [from the readings: Bell et al, "The million-dollar programming prize"]

- Suppose we give each user i an activity level u_i: (1.0, .8, .5, .2, .1).
- And we give each source j an appeal s_j: (.1, .2, .3, .4).
- Then notice that $p(i,j) = u_i s_j$.

Building a Model of the User

	Source 1	Source 2	Source 3	Source 4
User 1	.1	.2	.3	.4
User 2	.08	.16	.24	.32
User 3	.05	.1	.15	.2
User 4	.02	.04	.06	.08
User 5	.01	.02	.03	.04

- Suppose we give each user i an activity level u_i: (1.0, .8, .5, .2, .1).
- And we give each source j an appeal s_j: (.1, .2, .3, .4).
- Then notice that $p(i, j) = u_i s_j$.

An interpretation of this formulation:

- User i arrives at the platform and decides to consume something with probability u_i. (Otherwise they leave.)
- If they decide to consume something, the source they choose is j with probability s_j.
- Therefore, the probability they read source j is $u_i s_j$.

Most tables of values won't have such a simple structure.

Different Preferences

	Source 1	Source 2	Source 3	Source 4
User 1	.4	.4	.2	0
User 2	.2	.2	.1	0
User 3	.2	.3	.3	.2
User 4	0	.1	.2	.2
User 5	0	.2	.4	.4

Often, preferences can't be explained using just activities u_i and appeals s_j.

- What's going on in the table above?
- Are Sources 1 and 4 popular or not?

Maybe there are two kinds of users: those who like Sources 1 and 2, and those who like Sources 3 and 4.

- But there's some spillover: User 2 mainly likes Sources 1 and 2, but sometimes looks at Sources 3 and 4.
- And what about User 3, who seems to like everything?

Multiple Dimensions

	Source 1	Source 2	Source 3	Source 4
User 1	.4	.4	.2	0
User 2	.2	.2	.1	0
User 3	.2	.3	.3	.2
User 4	0	.1	.2	.2
User 5	0	.2	.4	.4

A more general model.

- The sources are organized into two *genres*.
- When you select according to one of the genres, you sample at random using a probability distribution specific to that genre.

For example:

- Probabilities for Genre 1 might be (.4, .4, .2, 0).
- Probabilities for Genre 2 might be (0, .2, .4, .4).
- Note: some items can be chosen under either genre.
- $s_j[1]$ will denote the probability Source j is chosen when selecting according to Genre 1; $s_j[2]$ for Genre 2.

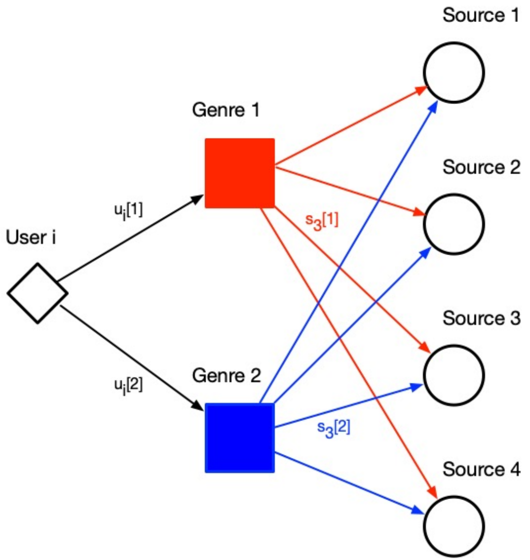
Multiple Dimensions

	Source 1	Source 2	Source 3	Source 4
User 1	.4	.4	.2	0
User 2	.2	.2	.1	0
User 3	.2	.3	.3	.2
User 4	0	.1	.2	.2
User 5	0	.2	.4	.4

How about users?

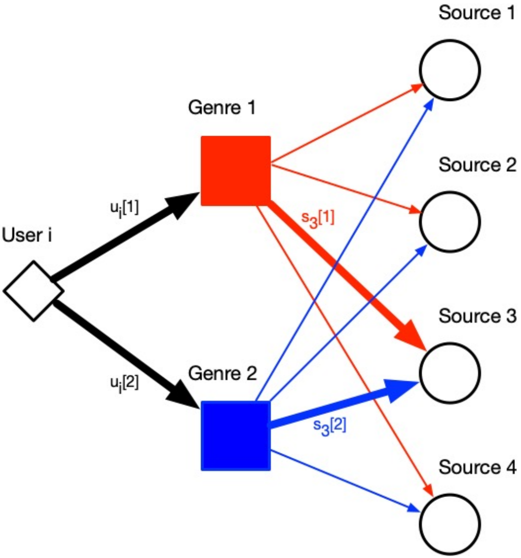
- When User i wants to select a source, they first choose genre at random.
- Chooses Genre 1 with probability $u_i[1]$, Genre 2 with $u_i[2]$.
- Then select a source according to the probabilities in that genre.
- $u_i[1] + u_i[2]$ can be less than 1; they choose nothing with remaining prob.

No need to limit the model to two genres (though our examples will).



User first selects genre, then source according to genre.

- Probability User i chooses Source j is $u_i[1]s_j[1] + u_i[2]s_j[2]$.



User first selects genre, then source according to genre.

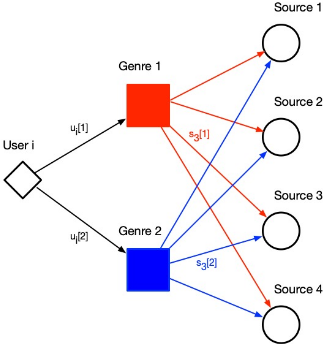
- Probability User i chooses Source j is $u_i[1]s_j[1] + u_i[2]s_j[2]$.

Multiple Dimensions

	Source 1	Source 2	Source 3	Source 4
User 1	.4	.4	.2	0
User 2	.2	.2	.1	0
User 3	.2	.3	.3	.2
User 4	0	.1	.2	.2
User 5	0	.2	.4	.4

In our example:

- Genre probabilities are (.4, .4, .2, 0) and (0, .2, .4, .4).
- User probabilities are (1,0), (.5,0), (.5,.5), (0, .5), (0,1).

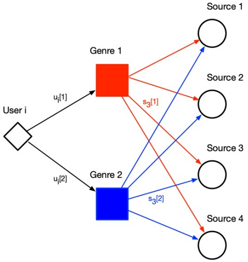


Latent factor models

	Source 1	Source 2	Source 3	Source 4
User 1	.4	.4	.2	0
User 2	.2	.2	.1	0
User 3	.2	.3	.3	.2
User 4	0	.1	.2	.2
User 5	0	.2	.4	.4

Multiple terms for this type of model.

- Latent factor model, mixture model, topic model
- Closely related: principal components analysis, multidimensional scaling.
- Linear algebra provides the main set of methods for fitting the model (outside this course).



Subtle aspect of the model: Who exactly chose the genres?
They emerge from the data and from the answer to the following question:

- Choose two numbers for each user, and two numbers of each source, so the resulting selection process matches the table as well as possible.
- And again: In general, we'd look for more than two genres. (Perhaps 20-50 to fit table with millions of users, thousands of sources.)

Latent factor models

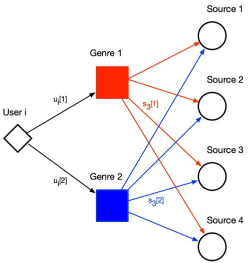
Q: Should we think of the platform as having chosen the genres or not?

Not clear-cut:

- The engineers working for the platform can't necessarily explain where the genres came from.
- But they made a set of decisions along the way:
  - Choosing to use this model at all;
  - Choosing the number of genres: parsimony versus fit.
- An interpretability problem: if you're working with content producers, you might need to explain to them what the different genres represent.
- A feedback loop problem: once system is running, the data you're using to build your model came from recommendations made by your model.

Each user and piece of content described by a small sequence of numbers.

- Recommending entertainment
- Prioritizing search results
- Prioritizing product displays
- Personalizing streams of news



Personalizing content can help users focus on what's most important to them.

- Without personalization, we might all just see what's globally popular.
- Personalization similarly helps creators of niche content get seen.

Q: Counterbalancing this, what are some of the risks in personalizing?

- Loss of opportunities for serendipitous discovery.
  (Foster and Ford, "Serendipity and information seeking," 2003.)
- The model might be wrong about you;
  some of your preferences might not be visible in the data;
  your preferences might change over time.
- Tension with the democracy theory of free speech.