COMPUTER SECURITY: Health Care Systems, Democracies and Social Media

Roy Campbell

Week 4: Social Media and their computer security concerns.

Friday Feb 24 9:30-11:00am.

Osher Lifelong Learning Institute

Illinois Classroom

CYBERSECURITY ISSUES

a) Tracking devices, phones, systems
b) Personal/private information
c) Blackmail
d) People trafficking
e) Influencers
f) Popular Systems
g) Legislation
h) Proposed controls
i) Issues.

https://www.cisa.gov/sites/default/files/publications/NCSAM_SocialMediaCybersecurity_2020.pdf

WEEK 4: SOCIAL MEDIA CYBERSECURITY

- Definitions, History, Types, Content Generation
- Types of attacks on Social Media [15]
- 34 Social Media Problems [18]
  Cyber Bullying, Social Media Addiction, Addiction, SEXTING, …
- Does more social media friends mean more friends? Diminishing Returns, Diet [28]
- Benefits of Social Media [31]
- Laws [33]
- Regulation of Social Media and Transparency [38]
- Algorithmic Amplification [45]
- Microtargeting [52]
- Tools: Bad Bots [55]

DEFINITIONS

Dissemination

- Broadcast a message to the public without direct feedback from the audience.

Disinformation

- False information which is intended to mislead, especially propaganda issued by a government organization to a rival power or the media

Conspiracy Theory

- A belief that a conspiracy has actually been decisive in producing a political event of which the theorists strongly disapprove
- Relying on the view that the universe is governed by design, and embody three principles: nothing happens by accident, nothing is as it seems, and everything is connected
- Evolves to incorporate whatever evidence exists against them, so that they become, as a closed system that is unfalsifiable, and therefore "a matter of faith rather than proof." Michael Barkin

DEFINITIONS SOCIAL MEDIA

Social media is a computer-based technology that facilitates the sharing of ideas, thoughts, and information through the building of virtual networks and communities. By design, social media is Internet-based and gives users quick electronic communication of content.

Common features:

1. Interactive Web 2.0 Internet-based applications
2. User-generated content (UGC)—such as text posts or comments, digital photos or videos, and data generated through all online interactions
3. Users create service-specific profiles for the website or app that are designed and maintained by the social media organization
4. Social media helps the development of online social networks by connecting a user's profile with those of other individuals or groups

EXAMPLES

Social media can be used to read or share news whether it is true or false. With no real ability to distinguish between the two it is down to the user of the platform to find the source reliable or not (Wikipedia)

Web-based Apps:

META	WeChat	Weibo	Tieba
Facebook	ShareChat	Twitter	LinkedIn
Facebook Messenger	Instagram	Tumblr	
TikTok	Qzone	Baidu	

Social Media Services:

YouTube	WhatsApp	Pinterest	VK,
QQ	Signal	Viber,	Microsoft Teams
Quora	LINE	Reddit,	
Telegram	Snapchat	Discord,	

Wikis are examples of collaborative content creation

HISTORICAL CONTEXT

The PLATO system was launched in 1960 after being developed at the University of Illinois and subsequently commercially marketed by Control Data Corporation.

It offered early forms of social media features with 1973-era innovations such as:

- Notes,
- PLATO's message-forum application;
- TERM-talk, its instant-messaging feature;
- Talkomatic, perhaps the first online chat room;
- News Report, a crowdsourced online newspaper, and blog and
- Access Lists, enabling the owner of a note file or other application to limit access to a certain set of users, for example, only friends, classmates, or co-workers.

There are as many as 4.76 billion social media users in the world[27] as of January 2023, equating to 59.4% of the total global population.

TYPES OF SOCIAL MEDIA

- Blogs (ex. Huffington Post, Boing Boing)
- Business networks (ex. LinkedIn, XING)
- Enterprise social networks (ex. Yammer, Socialcast)
- Forums (ex. Gaia Online, IGN Boards)
- Microblogs (ex. Twitter, Tumblr)
- Photo sharing (ex. Flickr, Photobucket)
- Products/services review (ex. Amazon, Elance)
- Social bookmarking (ex. Delicious, Pinterest)
- Social gaming (ex. Mafia Wars)
- Social network sites[6] (ex. Facebook, Instagram)
- Video sharing (ex. YouTube, Vimeo)
- Virtual worlds (ex. Second Life, Twinity)

MOBILE SOCIAL MEDIA

1. *Space-timers* (location and time-sensitive): Exchange of messages with relevance mostly for one specific location at one specific point in time (e.g. Facebook Places, WhatsApp, Telegram, Foursquare)
2. *Space-locators* (only location sensitive): Exchange of messages with relevance for one specific location, which is tagged to a certain place and read later by others (e.g. Yelp, Qype, Tumblr, Fishbrain)
3. *Quick-timers* (only time sensitive): Transfer of traditional social media mobile apps to increase immediacy (e.g. posting on Twitter or status updates on Facebook)
4. *Slow-timers* (neither location nor time sensitive): Transfer of traditional social media applications to mobile devices (e.g. watching a YouTube video or reading/editing a Wikipedia article)

•

DEFINITION FOR CONTENT GENERATION

Viral content

- Many social media sites provide specific functionality to help users re-share content, such as Twitter's "retweet" button or Facebook's "share" option.

Bots

- Chatbots and social bots are programmed to mimic natural human interactions such as liking, commenting, following, and unfollowing on social media platforms.
- 'Cyborgs'—either bot-assisted humans or human-assisted bots—are used for a number of different purposes both legitimate and illegitimate, from spreading fake news to creating marketing buzz.
- Cyborgs are also related to sock puppet accounts, where one human pretends to be someone else, but can also include one human operating multiple cyborg accounts.

STATISTICS

Social networking services with the most users, January 2022

#	Network Name	Number of Users (in millions)	Country of Origin
1	Facebook	2,910	United States
2	YouTube	2,562	United States
3	WhatsApp	2,000	United States
4	Instagram	1,478	United States
5	WeChat	1,263	China
6	TikTok	1,000	China
7	Facebook Messenger	988	United States
8	Douyin	600	China

USAGE BY MINORS

Apps used by U.S. tweens (ages 9–12), 2019-2020[69]:39–42

Platform	Overall	Boys	Girls	9-year-olds	12-year-olds
YouTube	67%	68%	66%	53.6%	74.6%
Minecraft	48%	61%	35%	43.6%	49.9%
Roblox	47%	44%	49%	41.2%	41.7%
Google Classroom	45%	48%	41%	39.6%	49.3%
Fortnite	31%	43%	20%	22.2%	38.9%
TikTok	30%	23%	30%	16.8%	37%
YouTube Kids	26%	24%	28%	32.7%	22.1%
Snapchat	16%	11%	21%	5.6%	22.3%
Facebook Messenger Kids	15%	12%	18%	19.1%	10.4%
Instagram	15%	12%	19%	3%	28.8%
Discord	8%	11%	5%	0.7%	14.4%
Facebook	8%	6%	9%	2.2%	15%
Twitch	5%	7%	2%	1.0%	9.9%
None of the above	5%	6%	5%	9.6%	3.3%

UNIVERSITY ADMISSIONS

- A survey in July 2017 by the American Association of College Registrars and Admissions Officers found that:
- 11 percent of respondents said they had refused to admit an applicant based on social media content.

(This includes 8 percent of public institutions, where the First Amendment applies.)

- The survey found that 30 percent of institutions acknowledged reviewing the personal social media accounts of applicants at least some of the time.

LoMonte, Frank (2021-12-13). "The First Amendment, Social Media and College Admissions". Inside Higher Ed.

TYPES OF ATTACKS ON SOCIAL MEDIA

- **SOCIAL MEDIA DEFAMATION**
- **CYBER STALKING**
- **SOCIAL MEDIA TROLLING**
- **STEALING OF IDENTITY AND VIOLATION OF PRIVACY**
- 1) Burglary via social networking;
- 2) Social engineering and phishing;
- 3) Malware;
- 4) Identity theft;
- 5) Cyber-stalking;
- 6) Cyber-casing.
- Disaster Fraud
- Credit Card Fraud

CYBER STALKING

social media "stalking" or "creeping" have been popularized over the years, and this refers to looking at the person's "timeline, status updates, tweets, and online bios" to find information about them and their activities.

- Identity thieves are not concerned with the effects of their actions on victim, while cyber-stalkers are well aware and do it deliberately.
- One woman out of twelve and one man out of forty-five will be stalked in their lifetime.
- According to the Bureau of Justice Statistics (BJS) [24], every 14 out of 1000 persons at the age of 18 are victims of stalking and around 1 out of 4 victims complaints about some sort of cyber-stalking in form of e-mail and instant messaging.
- Catfishing is a deceptive activity in which a person creates a fictional persona or fake identity on a social networking service, usually targeting a specific victim.

Bureau Of Justice Statistics, “Stalking”. https://www.bjs.gov/index.cfm?ty=tp&tid=973

CYBER CASING

- a process, which is used to produce real world location using various data available in online resources (e.g. geotagging).

34 SOCIAL MEDIA PROBLEMS

1. Cyberbullying
2. Social media Addiction
3. Content is never deleted
4. Sleep Deprivation
5. Damage to company’s and employer’s public reputation
6. Digital Footprint
7. Discrimination in hiring
8. Social media can lead to FOMO, fear of missing out
9. Sensitive information is leaked
10. Sexting

MORE SOCIAL MEDIA PROBLEMS

11. Legal issues
12. Using social media during work
13. There’s a lot of noise
14. Copyright
15. Obsession with likes and comments
16. Endorsements
17. Social media can lead to multitasking and becoming unproductive
18. Fake Identities
19. It’s incredibly time-consuming
20. Decreases face-to-face interactions

EVEN MORE SOCIAL MEDIA PROBLEMS

21. Social media promotes procrastination
22. Does more social media friends mean more friends?
23. Removes privacy
24. Social media can trigger sadness
25. Makes us compare our lives
26. Proper grammar is forgotten
27. Jealousy
28. There’s a misconception that social media will help solve our problems
29. Shorter attention spans

NOT MORE SOCIAL MEDIA PROBLEMS

31. Limiting thoughts
32. Social media glamorizes drug and alcohol use
33. It prevents us from spending quality time with each other
34. It prevents us to live here and now

1. CYBER BULLYING

- 1 in 4 students report being bullied during the school year
- Most students won't report bullying - 63% of victims don't report bullying
- 90% of students who are bullied offline are also bullied online
- Looks, body shape, and race are the most common reasons for bullying
- Students who are bullied are:
  - Twice as likely to suffer from anxiety, depression, and other health issues
  - More likely to report high levels of suicide-related behavior
- Students that bully are:
  - At a great risk for academic problems, substance abuse, and violent behavior, both now and later in life
- What you can do:
  - Take a stand! More than half of bullying situations end when a peer intervenes
  - Start a program! Schools for bully prevention programs can reduce bullying by 25%
  - Get everyone involved! Effective bully prevention involves everyone - students, parents, teachers, and community members

https://www.raymondgeddes.com/stats-and-effects-of-bullying

2. SOCIAL MEDIA ADDICTION

- spending the majority of time on social media, thinking about it, or creating social media content (outside of work)
- neglecting offline relationships
- inability to focus on things other than social media
- feeling restless, anxious, or agitated when unable to access social media
- using social media to escape reality
- increasing social media use over time to achieve the same gratification
- stopped or reduced access to social media may cause restlessness, irritability, agitation, or distress.

https://www.addictionresource.net

CONSEQUENCES OF SOCIAL MEDIA ADDICTION

- procrastination
- poor time management
- reduced work or academic performance
- poor mental health
- increased risk for substance misuse
- decreased physical activity
- social anxiety
- disrupted sleep patterns
- reduced connection to people in “real” life

WHO'S AT RISK OF SOCIAL MEDIA ADDICTION

Social media overuse is primarily a concern among teenagers and young adults, who are more likely to use social media. However, it may affect people of any age.

- low self-esteem
- extraversion
- low conscientiousness
- narcissism
- using social media to cope with stress
- family conflicts
- substance misuse of a sibling or parent

SOCIAL MEDIA USE RATES

- About 69 percent of all Americans use Facebook, while 40 percent use Instagram and 21 percent use TikTok. About 72 percent report using at least one social media site.
- The majority of social media users access social media sites regularly, with a smaller percentage using social media multiple times a day.
- People who overuse social media may be at higher risk of poor body image and developing harmful drinking or drug habits.
- Social media use, like drugs or alcohol, may boost ‘feel good’ hormones in the brain, like dopamine, which can reinforce continued social media use for the purpose of feeling pleasure.

SEXTING

- Survey: 948 public high school students (55.9% female) participated. The sample consisted of African American (26.6%), white (30.3%), Hispanic (31.7%), Asian (3.4%), and mixed/other (8.0%) teens (South East Texas)

https://jamanetwork.com/journals/jamapediatrics/fullarticle/1212181

- 28 percent of teens admitted to having sent a sext.
- 76.2 percent of teens who were propositioned to sext admitted to having had sexual intercourse.
- Girls were asked to send a sext (68 percent) more often than boys (42 percent).
- The peak age of sexting is around 16 and 17 years of age.
- Sexting seems to decline in individuals 18 and older.

DOES MORE SOCIAL MEDIA FRIENDS MEAN MORE FRIENDS?

- natural social network sizes may have a characteristic size in humans.
- This is determined in part by cognitive constraints and in part by the time costs of servicing relationships.

https://royalsocietypublishing.org/doi/10.1098/rsos.150292

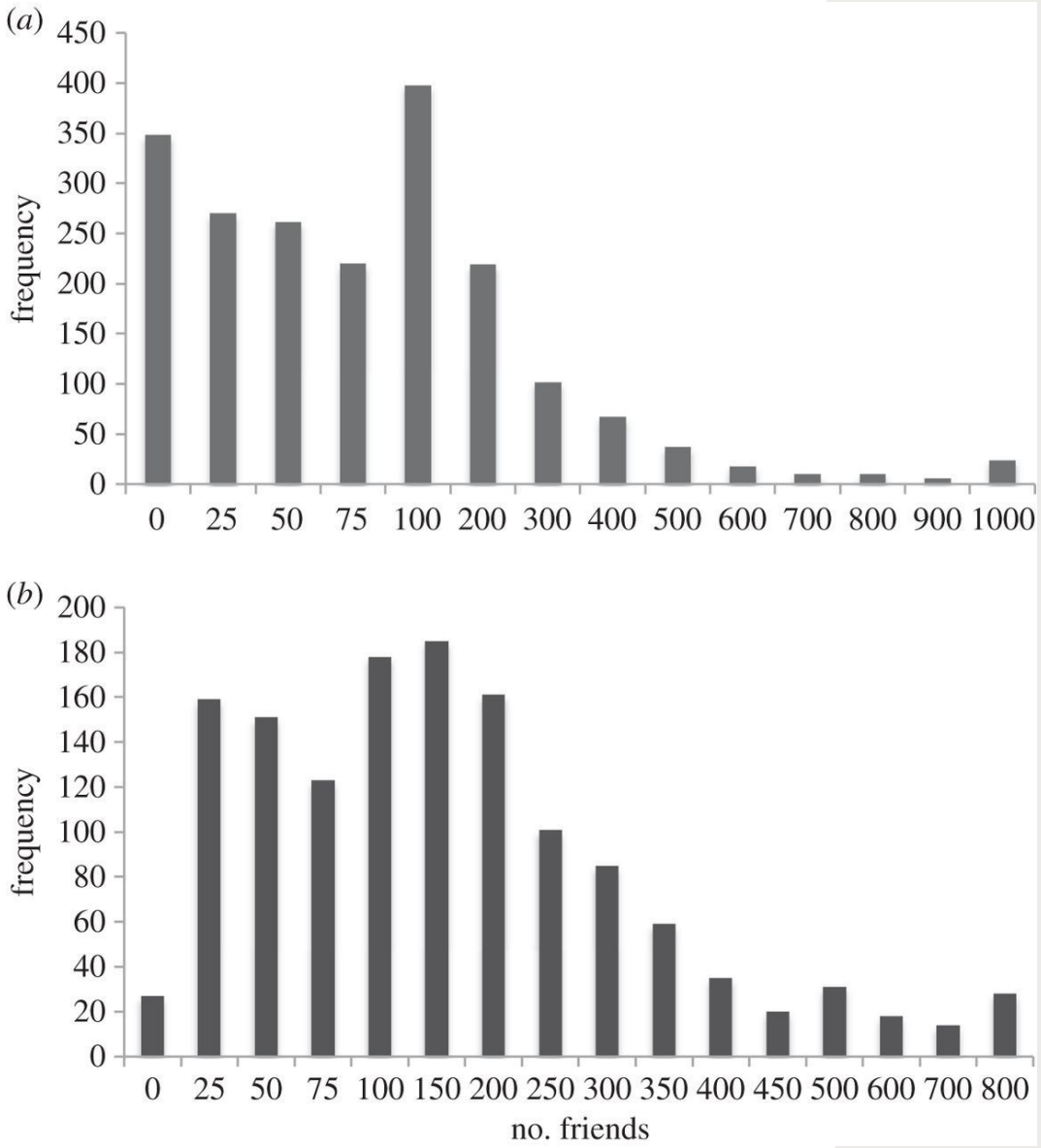
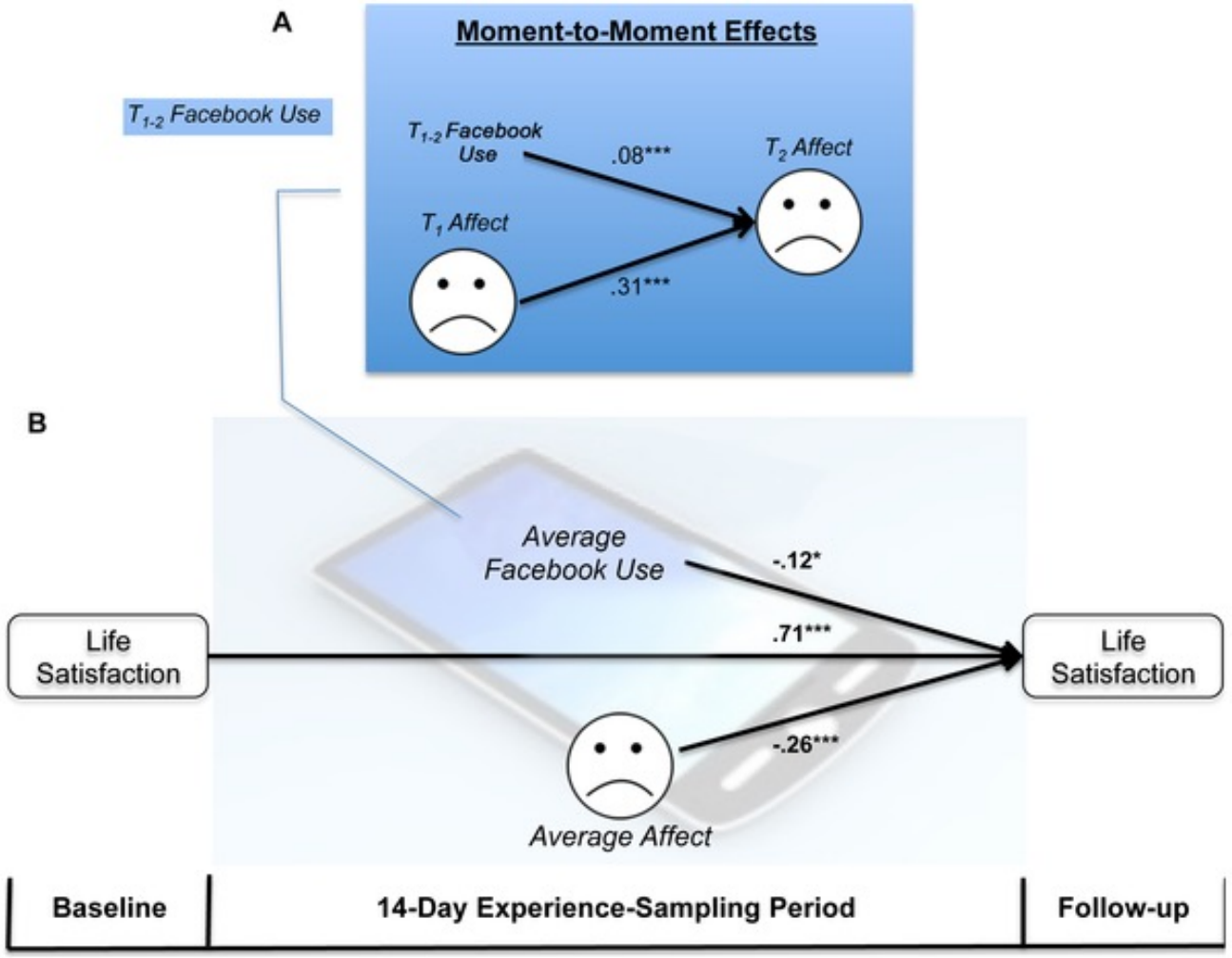


Figure 1. Distribution of network size for (*a*) Sample 1 (social network users: *N*=2000) and (*b*) Sample 2 (business employees: *N*=1375).

Figure 1. Facebook use predicts declines in affect and life satisfaction over time.



Kross E, Verduyn P, Demiralp E, Park J, Lee DS, et al. (2013) Facebook Use Predicts Declines in Subjective Well-Being in Young Adults. PLOS ONE 8(8): e69841. https://doi.org/10.1371/journal.pone.0069841
https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0069841

DISORDERED DIETING

Literature suggests that social media can breed a negative feedback loop of viewing and uploading photos, self-comparison, feelings of disappointment when perceived social success is not achieved, and disordered body perception.[168]

One study shows that the microblogging platform, Pinterest is directly associated with disordered dieting behavior, indicating that for those who frequently look at exercise or dieting "pins" there is a greater chance that they will engage in extreme weight-loss and dieting behavior.[169]

Holland, G.; Tiggerman, M. (2016). "A systematic review of the impact of the use of social networking sites on body image and disordered eating outcomes". *Body Image*. **17**: 101–109. doi:10.1016/j.bodyim.2016.02.008. PMID 26995158.

*Lewallen, Jennifer; Behm-Morawitz, Elizabeth (March 30, 2016). "Pinterest or Thinterest?: Social Comparison and Body Image on Social Media". Social Media + Society. **2** (1): 205630511664055. doi:10.1177/2056305116640559.*

Benefits of Social Media

https://www.linkedin.com/

1. Build relationships.
Social media is not just about brands connecting with their customers. In fact, at its root, social media is about connecting people to people.

2. Share your expertise.
Social media gives you an opportunity to talk about what you know and what you want to be known for. Sharing your expertise will attract potential professional and personal connections.
If you share content on topics that you know much about, you can begin to build credibility. This doesn't only go for your online presence. If you live your personal brand and your actions reflect your online presence, it validates that you can be trusted and those relationships you are building will be that much more authentic and valuable.

Benefits of Social Media

https://www.linkedin.com/

3. Increase your visibility.
If you spend time honing in on your expertise, consistently managing your social channels, then you have the potential to greatly increase your visibility and even become a thought-leader in your space. Good content gets shared, so if you are consistently posting quality content, the more people who share it, the more people see it. It's not just about pushing content, however. You also need to be engaging with other people's content. Following people and interacting with them on social media will work to build relationships (we keep coming back to this one!) and will help to get your name out there for people to turn to.

4. Educate yourself.
There is a lot of noise on the Internet. Social media allows you to hone in on what you really care about and what you really want to read. You can create lists that curate content from your favorite people, thought leaders in the space, or media outlets. You can easily learn about current events and things taking place near you.

5. Connect anytime
The advantage of being able to communicate and connect with anyone instantly outweighs the potential negative.

LAWS AND SOCIAL MEDIA

Laws associated with social media litigation include:

- the Digital Millennium Copyright Act
- Communications Decency Act.
- Lanham(TRADEMARK) Act
- Consumers' Privacy Protection Act (COPPA)

LANHAM(TRADEMARK) ACT

The FTC and the Department of Justice –violations of the Lanham Act are punishable as criminal antitrust violations.

- If a company illegally copies another's social media posting, that could be illegal. Under the FTC Act, it is also illegal for a company to fail to disclose to customers what it did to the content that was posted on its social media channels.
- Even if a company is not copying another company's content, but instead produces content for itself, it can still face legal troubles.

Just one instance of using someone else's intellectual property in the process of making content can cause problems for a company.

DIGITAL MILLENNIUM COPYRIGHT ACT

- 1998 United States copyright law implements two 1996 treaties of the World Intellectual Property Organization (WIPO).
- Criminalizes production and dissemination of technology, devices, or services intended to circumvent measures that control access to copyrighted works (commonly known as digital rights management or DRM).
- Criminalizes the act of circumventing an access control, whether or not there is actual infringement of copyright itself.
- Extends the reach of copyright, while limiting the liability of the providers of online services for copyright infringement by their users.

CONSUMERS' PRIVACY PROTECTION ACT (COPPA)

- Consumers' Privacy Protection Act (COPPA) is a United States Federal Trade Commission (FTC) regulation that prohibits marketing communications and direct mail from minors under the age of 13. COPPA applies to social media (social networking sites, cell phones, tablets, and similar devices) and extends the FTC's reach to include all marketers in the US.

COMMUNICATIONS DECENCY ACT

1996 (**CDA**) was the United States Congress's first notable attempt to regulate pornographic material on the Internet. As eventually passed by Congress, Title V affected the Internet (and online communications) in two significant ways.

1) It attempted to regulate both indecency (when available to children) and obscenity in cyberspace.
2) Section 230 of title 47 of the U.S. Code, part of a codification of the Communications Act of 1934 (Section 9 of the Communications Decency Act / Section 509 of the Telecommunications Act of 1996)[3] has been interpreted to mean that operators of Internet services are not publishers (and thus not legally liable for the words of third parties who use their services).
3) Allow States and Victims to Fight Online Sex Trafficking Act (FOSTA) and Stop Enabling Sex Traffickers Act (SESTA) makes it illegal to knowingly assist, facilitate, or support sex trafficking, and amends the Communications Decency Act's section 230 safe harbors (which make online services immune from civil liability for their users' actions) to exclude enforcement of federal or state sex trafficking laws from immunity. The intent is to provide serious legal consequences for websites that profit from sex trafficking and give prosecutors tools to protect their communities and give victims a pathway to justice.

SUPREME COURT AND SOCIAL MEDIA

- Section 230 of the Communications Decency Act protects the companies from liability for content posted by individual users, no matter how discriminatory, defamatory or even dangerous the information may be.
- Oral arguments in the case, Gonzalez v. Google zeroed in on the concept of algorithms, which companies use to recommend third-party content, and whether they deserve protection as an inherent part of publishing information online, which is protected by Section 230.
- "Algorithms are ubiquitous, but the question is what does the defendant do with algorithm," said Gonzalez family counsel Eric Schnapper. "If it uses the algorithm to direct, to encourage people to look at ISIS videos, that's within the scope of [liability].”
- Nearly all the justices appeared sympathetic to legal protection for internet companies as publishers, with many voicing concern about a potential flood of lawsuits that could come from ending immunity for algorithms.

REGULATION OF SOCIAL MEDIA AND TRANSPARENCY

Arguments against just transparency

- Disclosure rules imply that whatever the companies do is fine as long as they are transparent about it
- Transparency requirements written into law wouldn't ensure much useful disclosure.
- Transparency initiatives, the critics say, just distract from the hard work of developing and implementing more effective methods of control
- Substantially increased disclosures wouldn't do much to mitigate the information disorder on social media.

https://www.brookings.edu/blog/techtank/2022/11/01/transparency-is-essential-for-effective-social-media-regulation/

ARGUMENTS FOR TRANSPARENCY REGULATIONS

- Requirement for disclosure will not produce useful disclosures insists on the importance of a regulator
- A dedicated regulatory agency must define and implement them through rulemaking and must have full enforcement powers, including the ability to fine and issue injunctions
- Without transparency, no other regulatory measures will be effective.

Whatever else governments might need to do to control social media misbehavior in content moderation, they have to mandate openness, which requires implementing specific rules governing these disclosures.

DIMENSIONS OF TRANSPARENCY REGULATIONS

Disclosures to

1) users,
2) public reporting, and
3) access to data for researchers.

1) TRANSPARENCY TO USERS

- Content moderation standards a social media company has in place
- Its enforcement processes
- Explanations of take downs
- Other content moderation actions
- Descriptions of complaint procedures...

Each of these outputs provide users with opportunities to complain about problematic content and to receive due process when social media companies take action against them.

The level of detail to which this should be done should be through public rulemaking with input from civil society, industry, and academia.

2) TRANSPARENCY REPORTING

- Reports and internal audits of platform content moderation activity
- The risks created through social media company activities
- The role of algorithms in distributing harmful speech
- Assessments of what the companies do about

  Hate speech

  Disinformation

  Material harmful to teens

  Other problematic content.
- Transparency reporting could also include a company's own assessment of whether its activities are politically discriminatory.

METRICS

- Perhaps a social media self-regulatory group, similar to the Financial Industry Regulatory Authority, the broker-dealer industry organization, to define common reporting standards.

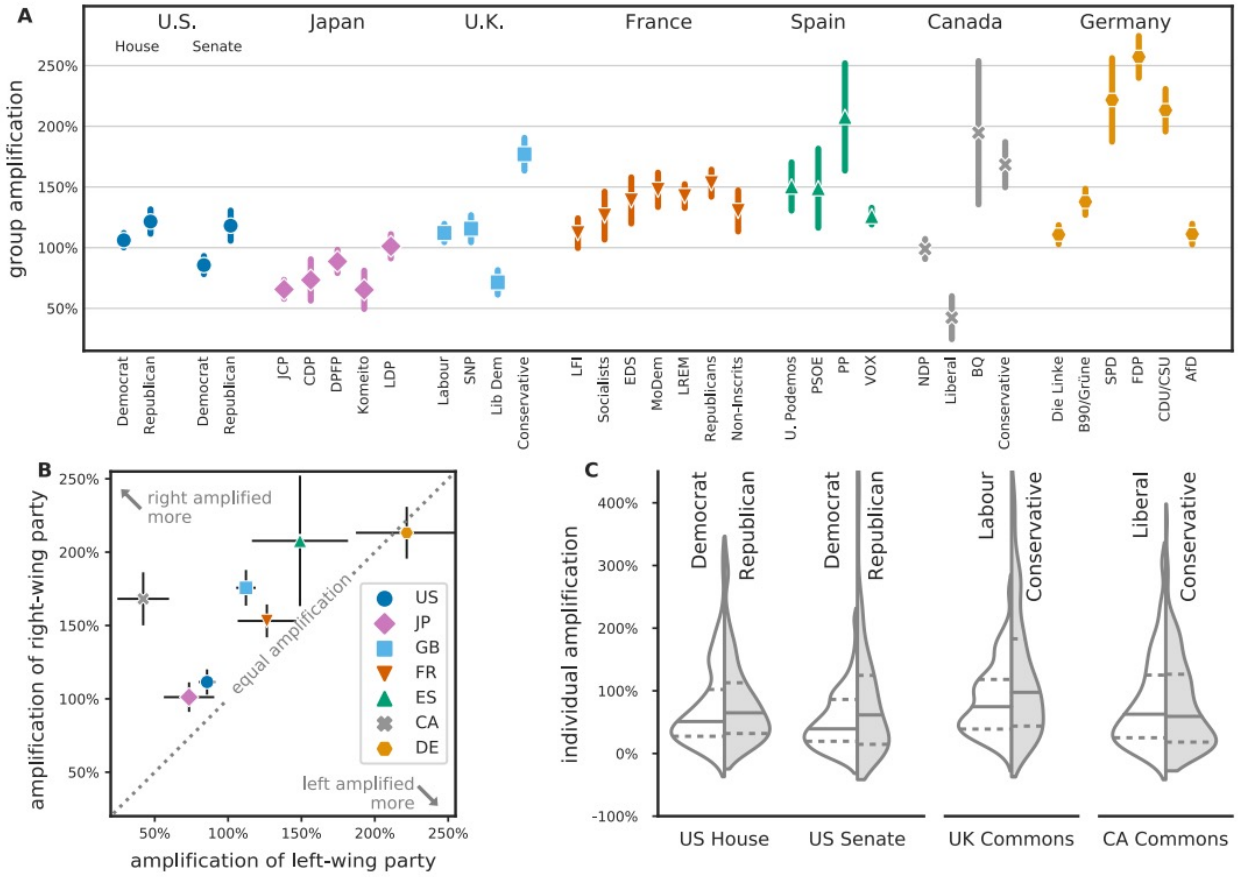
3) ACCESS TO DATA FOR RESEARCHERS

- Access to the internal company data that researchers need to conduct independent evaluations
- Outside evaluations would not be under company control and would assess company performance on content moderation and the prevalence of harmful material
- Vetted researchers to validate internal company studies, such as Twitter's own assessment of political bias
- **Examining algorithmic amplification of political content on Twitter (blog)**
- **Algorithmic amplification of politics on Twitter (PNAS paper)**

EXAMINING ALGORITHMIC AMPLIFICATION OF POLITICAL CONTENT ON TWITTER

- Tweets about political content from elected officials, regardless of party or whether the party is in power, do see algorithmic amplification when compared to political content on the reverse chronological timeline
- Group effects did not translate to individual effects. In other words, since party affiliation or ideology is not a factor our systems consider when recommending content, two individuals in the same political party would not necessarily see the same amplification
- In six out of seven countries — all but Germany — Tweets posted by accounts from the political right receive more algorithmic amplification than the political left when studied as a group
- Right-leaning news outlets, as defined by the independent organizations listed above, see greater algorithmic amplification on Twitter compared to left-leaning news outlets. However, as highlighted in the paper, these third-party ratings make their own, independent classifications and as such the results of analysis may vary depending on which source is used.

AMPLIFICATION OF TWEETS FROM MAJOR POLITICAL GROUPS AND POLITICIANS IN 7 COUNTRIES



TWITTER

- Our analysis of far-left and far-right parties in various countries does not support the hypothesis that algorithmic personalization amplifies extreme ideologies more than mainstream political voices. However, some findings point at the possibility that strong partisan bias in news reporting is associated with higher amplification. We note that strong partisan bias here means a consistent tendency to report news in a way favouring one party or another and does not imply the promotion of extreme political ideology.

CONTENT MODERATION EFFORTS

- Content moderation efforts suffer from an externality problem similar to environmental pollution. Polluting companies export the cost of pollution rather than experience it themselves. Social media companies also externalize the harms they exacerbate.

MYANMAR: THE SOCIAL ATROCITY: META AND THE RIGHT TO REMEDY FOR THE ROHINGYA

- https://www.amnesty.org/en/documents/ASA16/5933/2022/en/
- Beginning in August 2017, the Myanmar security forces undertook a brutal campaign of ethnic cleansing against Rohingya Muslims in Myanmar's northern Rakhine State.
- The violence pushed over 700,000 Rohingya – more than 80 per cent of the Rohingya population living in northern Rakhine State at beginning of the crisis – into neighbouring Bangladesh, where most linger in refugee camps to this day.
- In the months and years leading up to and during the 2017 atrocities, Facebook in Myanmar became an echo chamber of virulent anti-Rohingya content. Actors linked to the Myanmar military and radical Buddhist nationalist groups systematically flooded the Facebook platform with incitement targeting the Rohingya, sowing disinformation regarding an impending Muslim takeover of the country and seeking to portray the Rohingya as sub-human invaders.
- The Independent International Fact-Finding Mission on Myanmar concluded that "the role of social media was significant in the atrocities that ensued.

META STAFFING ISSUE

- Meta's wholly inadequate staffing of its Myanmar operations prior to 2017 was a significant factor in the company's staggering failures to remove harmful anti-Rohingya content from the Facebook platform.
- This is symptomatic of the company's broader failure to adequately invest in content moderation across the Global South.
- In mid-2014, Meta staff admitted that they only had one single Burmese-speaking content moderator devoted to Myanmar at the time, based in their Dublin office.
- Meta has never disclosed the precise number of Burmese-language content moderators it employed during the 2017 atrocities, but the company claimed to have hired 'dozens' more in mid-2018.

AMNESTY RECOMMENDATIONS

- Ban targeted advertising on the basis of invasive tracking practices, such as cross-site tracking, and tracking based on sensitive data or other personal data.
- Introduce obligations for platform companies to ensure they address systemic risks to human rights stemming from the functioning and use made of their services.
- Legally require companies, including social media companies, to conduct human rights due diligence and report publicly on their due diligence policies and practices.
- Regulate technology companies to ensure that content-shaping algorithms used by online platforms are not based on profiling by default and must require an opt-in instead of an optout, with the consent for opting in being freely given, specific, informed and unambiguous.

MICROTARGETING

- US Senate Select Committee on Intelligence report, Russia's Internet Research Agency used digital personalisation techniques to interfere in the 2016 US presidential elections.
- This campaign specifically targeted African-Americans, with misinformation used to generate outrage against other social groups, co-opt participation in protests, or even convince individuals not to participate in the elections at all.

CAMBRIDGE ANALYTICA

- Microtargeting is a form of targeting that uses recent technological developments to gather large amounts of online data. The data from people's digital footprints is analysed to create and convey messages that reflect an individual's preferences and personality.
- Research has shown that such digital footprints can be used to accurately and unobtrusively predict psychological traits and states of large groups of people.
- The Republican National Committee's database is called Voter Vault. The Democratic National Committee effort is called VoteBuilder.
- These databases are then mined to identify issues important to each voter and whether that voter is more likely to identify with one party or another. As described by Cambridge Analytica's CEO, their key was to identify people who might be enticed to vote for their client or be discouraged to vote for their opponent.

Isaak, Jim; Hanna, Mina J. (August 2018). "User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection". *Computer*. **51** (8): 56–59. doi:10.1109/MC.2018.3191268. S2CID 52047339.

TOOLS --- BAD BOTS

- Bots are autonomous programmes on a network that can interact with systems or users. On social media, bots can impersonate real people, for example, by automatically writing messages. Multiple bots acting together can create a buzz around a person, product, or topic and push a particular point of view or agenda. Bots can amplify the reach of disinformation by pushing specific messages, hashtags, or accounts, creating the impression that a particular perspective is popular and, therefore, more likely to be true.
- Researchers at Brown University reported that a quarter of tweets on climate change were likely posted by bots to spread climate-denial propaganda.

Detection of Bots in Social Media: A Systematic Review,
Mariam Orabi, Djedjiga Mouheb, Zaher Al Aghbari, Ibrahim Kamel
https://www.sciencedirect.com/science/article/abs/pii/S0306457319313937

BOTS AND ONLINE CLIMATE DISCOURSES: TWITTER DISCOURSE ON PRESIDENT TRUMP'S ANNOUNCEMENT OF U.S. WITHDRAWAL FROM THE PARIS AGREEMENT

https://www.scientificamerican.com/article/twitter-bots-are-a-major-source-of-climate-disinformation/

- 6.8 million tweets sent by 1.6 million users between May and June 2017. Trump made his decision to ditch the climate accord on June 1 of that year.
- A random sample of 184,767 users run through the Botometer, a tool created by Indiana University's Observatory on Social Media. Nearly 9.5% of the users in their sample were likely bots.

(https://iuni.iu.edu/projects/botometer Botometer, formerly called BotOrNot, is a machine-learning algorithm that rates how likely a Twitter account is to be a bot, based on tens of thousands of labeled examples.)

- Researchers also categorized the 885,164 tweets those users had sent about climate change during the two-month study period.
- Bots accounted for 25% of the total tweets about climate change on most days.

TOOLS FOR SOCIAL MEDIA

- https://arxiv.org/pdf/1601.05140.pdf The DARPA TWITTER BOT CHALLENGE
- Tweet Syntax.
- Tweet Semantics
- Temporal Behavior
- User Profile
- Network Features

TWEET SYNTAX.

- This included the following features:
- Does the user post tweets whose syntax is similar to the natural language generation program Eliza and auto-generation of language [8]
- Average number of hashtags, user mentions, links, and special characters in tweets.
- Average number of retweets by the user.
- Are the tweets geo-enabled?
- Percentage of tweets ending with punctuation, hashtag, or link – the intuition is that such tweets might be automatically generated.

TWEET SEMANTICS

- This category included the following features:
- Number of user posts related to vaccination.
- User's average sentiment score (on the topic "topic") in topic-related tweets.
- Measures of contradiction in the user's posts on topic-related tweets using functions such as Contradiction Rank which measures variation between the sentiment of the user across a set of topics and the sentiments of his neighbors on the same topics.
- Positive (resp. Negative) Sentiment Strength measuring the average sentiment strength of the user's positive (resp. negative) tweets.
- Most frequent topics that the user tweets about.
- Number of languages in which tweets were generated. The rationale was that accounts posting tweets in many languages may be bots.
- Sentiment inconsistency. Paybots often copy a link from a popular Twitter user and then replace the micro-URLs from the original Twitter post with a spurious link to a site where the paybot owners were paid for generating views. This feature analyzes whether sentiment in the tweet content varied significantly (w.r.t. the topic) as compared to sentiment in a URL embedded within the tweet.

TEMPORAL BEHAVIOR FEATURES

- Did the user's sentiment flip-flop over time?
- Variance in tweet sentiment over time. This identified users who had an explicit infiltration strategy of posting anti-topic tweets to engage the anti-topic community and later switching to a pro-topic stance.
- Entropy of inter-tweet time distribution. The rationale was that algorithmic tweeting should have some temporal regularities that are reflected in relatively low entropy of the corresponding distribution.
- Predictability of tweet timing based on a transfer entropy approach.
- The duration of the longest session by a user without any short (5-minute or 10-minute) breaks – clearly, users that have a session lasting a day without any breaks are not likely to be humans.
- Average number of tweets per day – as in the previous case, if this number is large, it increases the probability that the user is a bot.
- Percentage of dropped followers. What was the percentage of "unfollows" compared to the percentage of "follows"? For instance, a user who dropped a lot of followers compared to the number of people he was following may be anomalous.
- Signal-to-noise ratio (SNR). The ratio of mean to standard deviation, min, max and entropy of these values to detect abrupt changes in users' metadata (followers, followees, posts, etc.).

USER PROFILE FEATURES

- Did the user's profile have a photo? If so, was it from a stock image database?
- Did the user's profile have an associated URL? If so, did the URL have a clone elsewhere? A URL that was a clone of some other URL increased the level of suspicion of the user.
- Did the user's Twitter name look auto-generated? Several heuristics for such tests include by comparing screen-names with user-names after splitting on spaces/underscores and looking for common substrings.
- Number of posts/retweets/replies/mentions.
- Number of followers/followings.
- Number of sources used by the user such as: mobile applications, desktop browsers, or "null" for missing sources.
- GPS coordinate availability for user's tweets.
- Similarity of user profile to known bots (measured using Jaccard similarity or Cosine Similarity)

NETWORK FEATURES

- Average Deviation of user sentiment scores from those of his followers and followings.
- In and Out degree centrality.
- Average clustering coefficient of retweet and mention network associated with each user.
- Pagerank and between-ness centrality of users in both retweet and mention networks.
- Variables related to star and clique networks associated with users.
- Number of known bots followed by a user – a user following several known bots is more likely to be a bot.
- Number/Percentage of bots in the cluster that a user belonged to –if a clustering algorithm places the user in a cluster with many bots, he is more likely to be a bot.

CHATBOTS AND CHATGPT

- Chatbot: a software application used to conduct an online chat conversation via text or text-to-speech, in lieu of providing direct contact with a live human agent. (Wikipedia)
- Think of ChatGPT as a blurry jpeg of all the text on the Web. Ted Chiang

  ChatGPT – a generative pre-trained transformer (GPT) – was fine-tuned (an approach to transfer learning[5]) on top of GPT-3.5 using supervised learning as well as reinforcement learning

  Unlike most chatbots, ChatGPT remembers previous prompts given to it in the same conversation.

  Its not clear how such generative AI will alter social media or its cyber security


Thank
you!!