DSci529: Security and Privacy In Informatics

Regulation of Content Disinformation

Prof. Clifford Neuman

Lecture 12
9 April 2021
Online

Course Outline

- Overview of Security and Privacy
- What data is out there and how is it used
- Technical means of protection
- Identification, Authentication, Audit
- Reasonable expectation of privacy
- Big Data – Technology and Privacy
- AI and Bias
- The Internet of Things and Security and Privacy
- Social Networks and the use of our Data
- Access to Data by Governments - Privacy in a Pandemic
- Privacy Regulation - GDPR, CCPA, CPRA
- **Influence of Social Media – Free Speech – Disinformation**
- CryptoCurrency - TOR - Privacy Preserving Technologies

Upcoming Presentations Privacy and Finance – April 16th

- Jonathan De Leon – Privacy in Finance
- Sidong Wang – History and Technologies for Cryptocurrencies
- Saurabh Jain – Privacy of Credit Card/Payment card information
- Yifeng Shi -Financial value of data gathered through free services

- 40 minutes

Secure Communication – Privacy Preserving Technologies – April 16th

- Zihuan Ran – Privacy Preserving Database Technologies
- Aziza Saulebay – 5G and Data Privacy
- Carol Varkey – Messaging Application Privacy
- Francisco Ventura – Encryption Technologies and implications

- 40 minutes

Upcoming Presentations
Other Security Topics – April 23rd

- Yo-Shuan Liu – User experience and Multi-Factor Authentication
- Philana Williams – Security for Web App Development
- Haonan Xu – Privacy issues in Cloud Computing
- Pratishtha Singh – Card privacy Concerns in India

Today's Agenda

12:00 – 12:05 Introduction and Announcements
12:05 – 12:30 Security and Privacy – Influence on Elections
12:20 – 12:40 Influence of Social Media on Society
12:40 – 13:30 Disinformation
- 12:40 – 12:50 Adriana Nana – Deep Fakes
- 12:50 – 13:00 Social Bots
- 13:00 – 13:30 Influence on Public Discourse

13:30 – 13:40 Break
13:40 – 14:40 Regulation of Content
- 13:40 – 13:55 Types of Regulated Content
- 13:55 – 14:05 Technical approaches to detecting and regulating content
- 14:05 – 14:15 Resherle Verna – Should platforms have right of censorship
- 14:15 – 14:25 Section 230
- 14:25 – 14:50 Open disuussion Regulation of Content

14:50 – 15:20 Current Event Discussions

Elections and Cyber-Security

- Our discussion of disinformation will focus in part on items in the news recently: Elections.
  - In 2020 the USC Election Cybersecurity Initiative presented to campaigns and election officials in all 50 states on these issues.
  - The slides that follow are an updated presentation provided as part of this activity. My focus was cybersecurity, revising some of lectures 3 and 4 in a different context.
  - I discuss the broader context of information security in elections.
  - There were other presentations on disinformation, but in this class we will focus instead on some of the assigned readings.

USC Annenberg
School for Communication and Journalism

USC University of Southern California

Election Cybersecurity Initiative

Cybersecurity and Cyber Safety

Dr. Clifford Neuman
Director, USC Center for Computer Systems Security
Scientist, USC Information Sciences Institute
Associate Professor of
Computer Science Practice
USC Viterbi School of Engineering

March 25, 2021 | Regional Workshop: DE, MD, NJ, NY, and PA

USC Annenberg
School for Communication and Journalism

USC University of Southern California

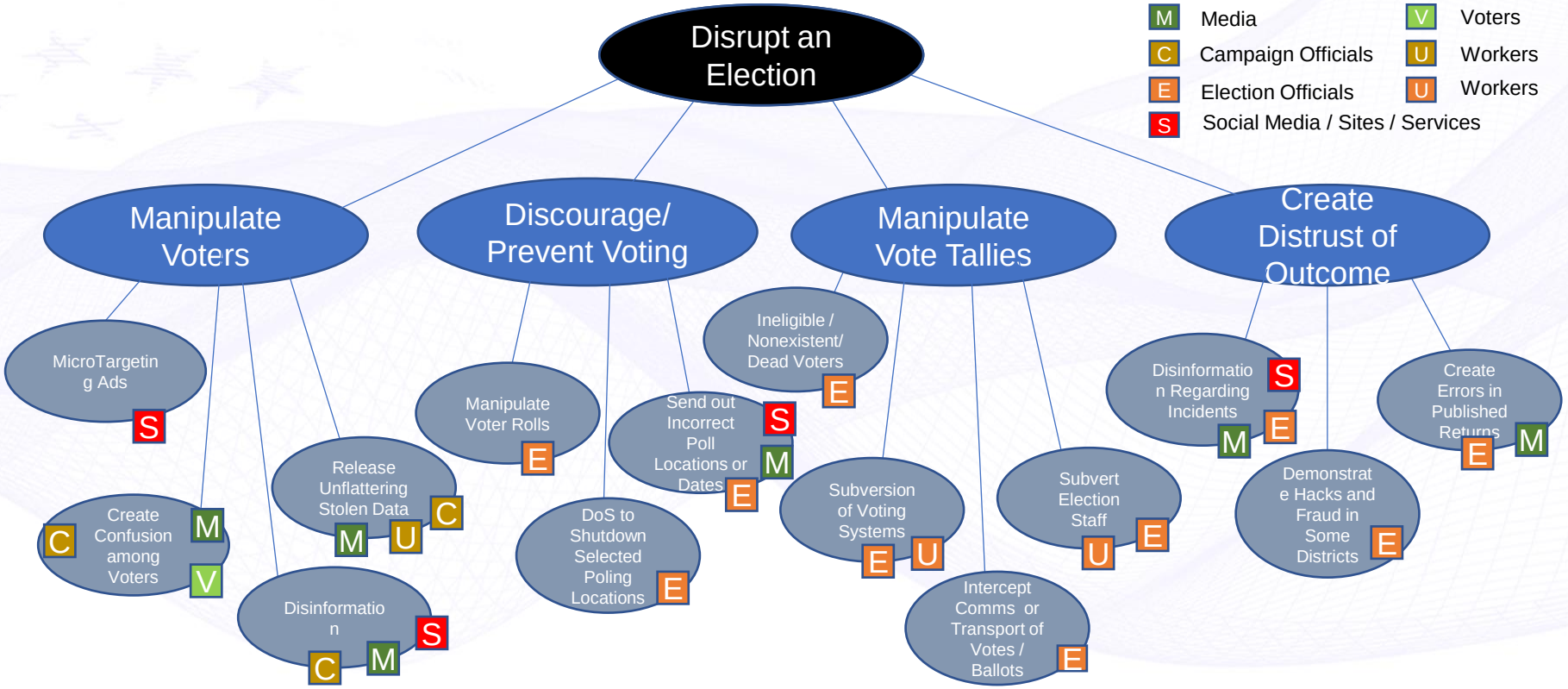
Who's Responsible for Protecting Our Elections?

We are all responsible, but some users have greater impact in defending some kinds of attacks. We will discuss the best defenses throughout the day.

USC Annenberg School for Communication and Journalism

USC University of Southern California

Roadmap



Source Clifford Neuman

USC Annenberg School for Communication and Journalism

USC University of Southern California

What Happened in 2020?

- On March 16th the DOJ and Homeland Security confirmed their findings that they found “no evidence that any foreign government-affiliated actor prevented voting, changed votes, or disrupted the ability to tally votes or to transmit election results in a timely manner; altered any technical aspect of the voting process; or otherwise compromised the integrity of voter registration information of any ballots cast during 2020 federal elections.”

- However, they did find evidence of “several incidents when Russian, Chinese, and Iranian government-affiliated actors materially impacted the security of networks associated with or pertaining to US political organizations, candidates, and campaigns during 2020 federal elections.”

Why is Cyber Security Important?

- In the same timeframe we learned of significant “successful” attacks on government and corporate systems: the Solar Winds breach, and Microsoft Exchange Server email breaches, which have been blamed on Russian and Chinese hacking groups.
  - These kinds of attacks could have easily spread to election infrastructure, and you can be certain that our those seeking to compromise that infrastructure will keep trying.
  - I believe Solar Winds did not affect our election infrastructure due to the level of isolation afforded to the systems used to count ballots.
  - Such isolation is an important computer security technique.

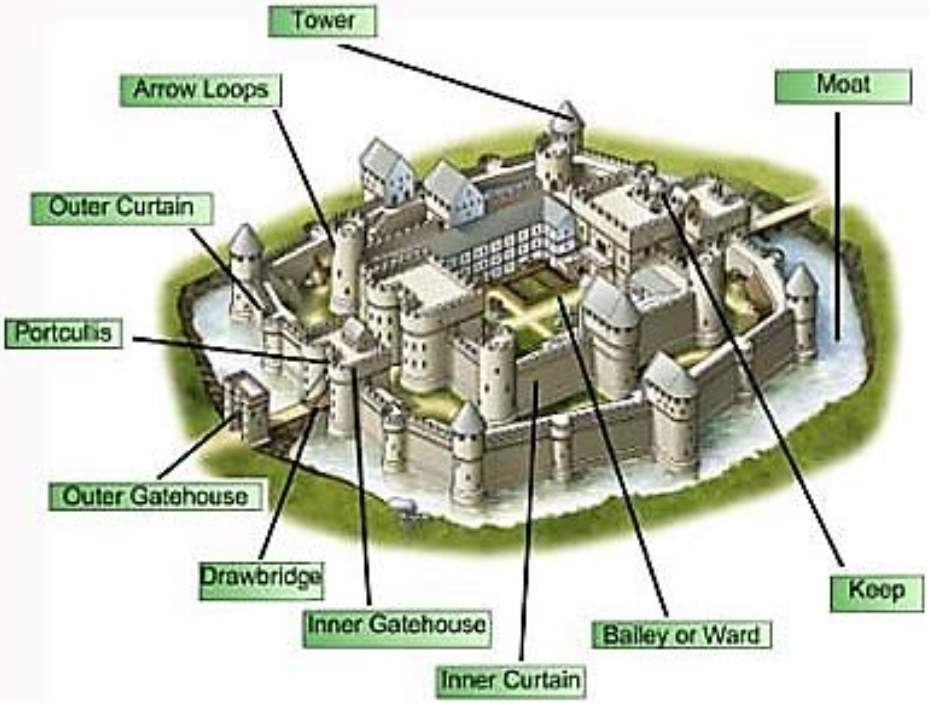
- The successful attacks on “political organizations, candidates, and campaigns” can materially impact the outcome of the elections through influence on voters.

Common Attack Vectors

- Poor Password Management
- Malicious Code (e.g. Viruses)
- Social Engineering
- Unprotected Data
- Disinformation / Misinformation

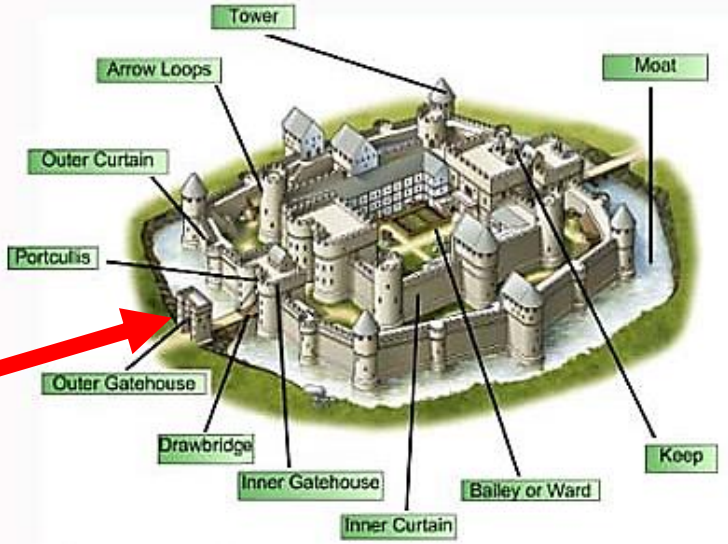
What Can We Do

Defense in Depth



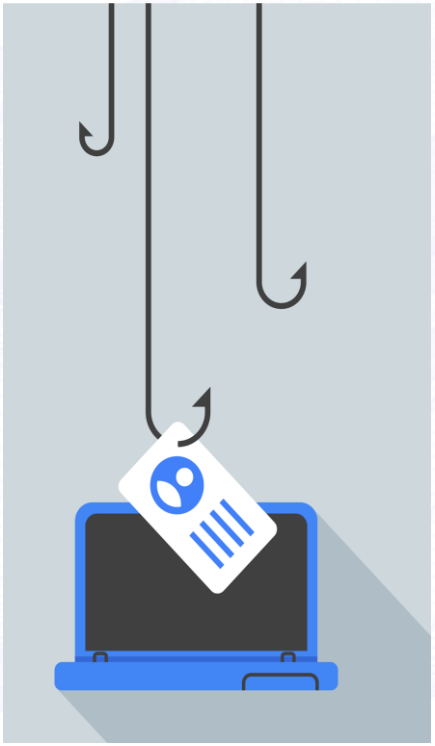
Use Strong Password

- Think of passwords as the first gatehouse in the castle
- Use Passphrases as an easy way to create easy to remember passwords
  - **KeepTh3m0utOfRsystems!**
- Do not reuse passwords between accounts!
- Do not use simple passwords!
- **Do not use the password above.**



Phishing

- Phishing is a method of gathering personal information (e.g. passwords) using deceptive e-mails, websites, apps, text messages, etc.
- Once your click on a link or provide a password, the hacker accesses your account or infects your machine



USC Annenberg
School for Communication and Journalism

USC University of Southern California

Is Phishing effective?

- EVERYONE is a target, including you!
- #1 way hackers gain access to systems

45%

The **most believable** phishing sites trick almost half of the users.

20%

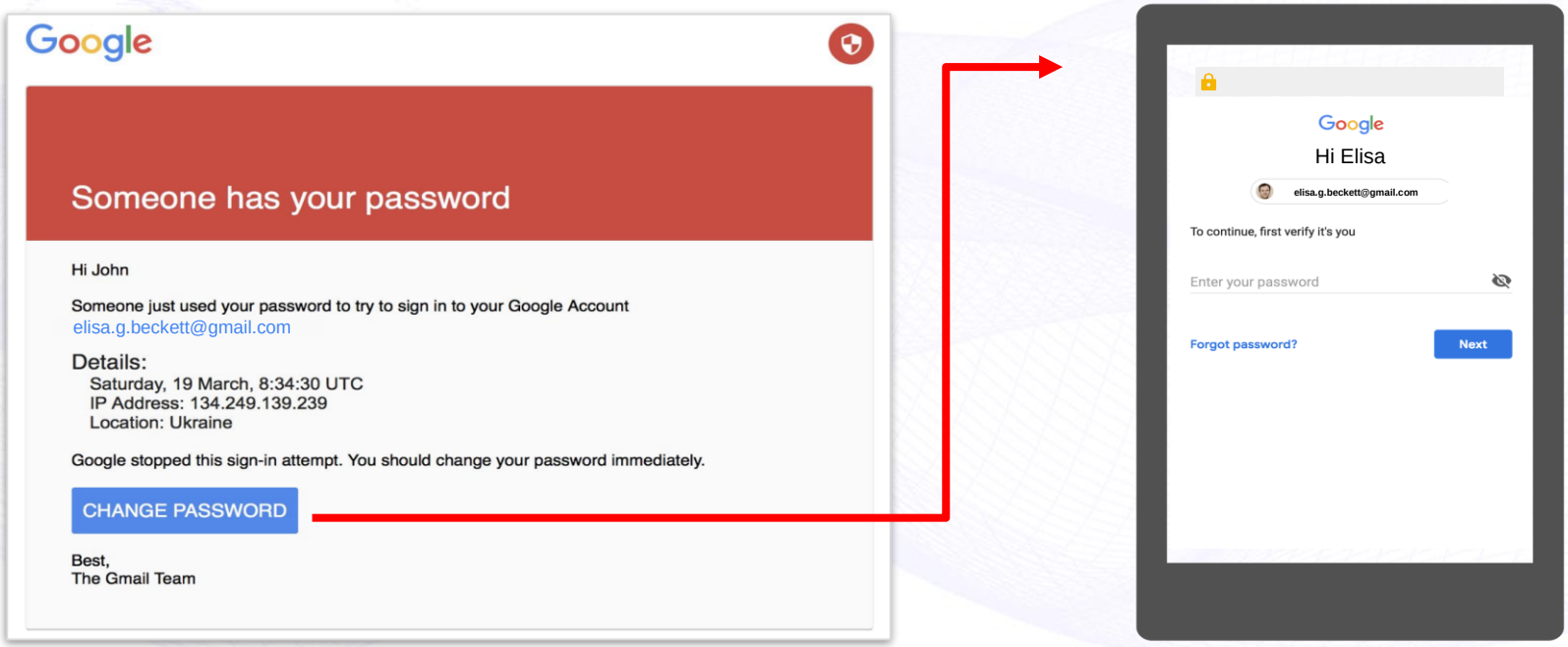
Hackers move fast: ⅕ of the accounts are accessed within 30 minutes after being phished.

THINK/CALL BEFORE YOU CLICK

USC Annenberg
School for Communication and Journalism

USC University of Southern California

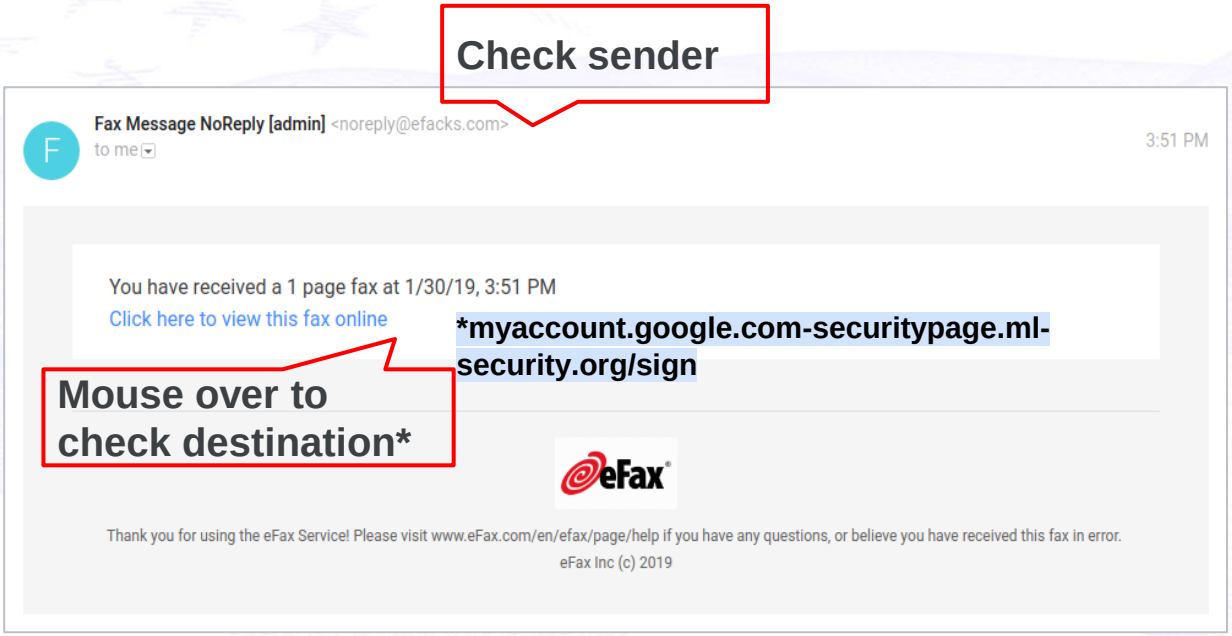
Phishing / Social Engineering



USC Annenberg School for Communication and Journalism

USC University of Southern California

Phishing / Social Engineering



Fraudulent email
Hackers will often send emails that look legit, so it's important to check the sender and the destination of any embedded links.

USC Annenberg
School for Communication and Journalism

USC University of Southern California

Phishing Quiz

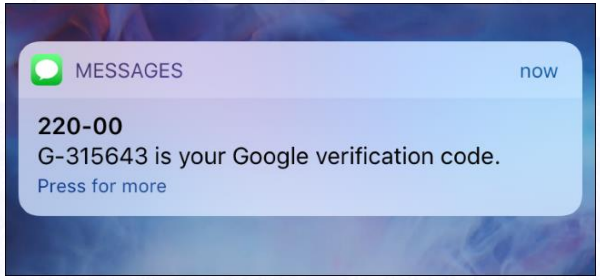
g.co/phishingquiz

Election Cybersecurity Initiative

USC Annenberg School for Communication and Journalism

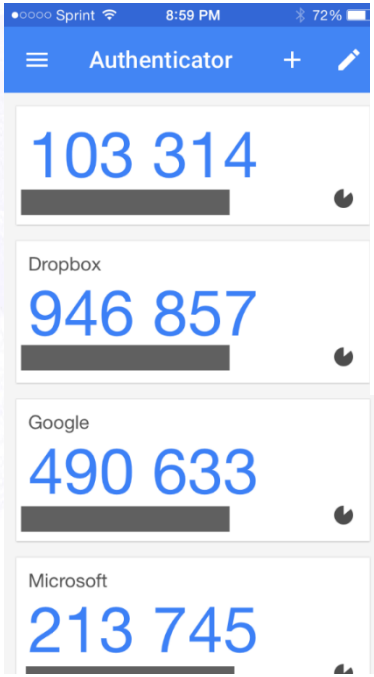
USC University of Southern California

Two-Factor or Multi-Factor Authentication



Better than passwords alone, but text messages can still be intercepted, or your phone account taken over.

- Add PIN/Passcode to your cellphone account (supported by major carriers including ATT, Verizon, Sprint).
- This helps prevent motivated individuals from moving your cell phone number to their phone to intercept texts used by second factors



Election Cybersecurity Initiative

USC Annenberg
School for Communication and Journalism

USC University of Southern California

Malware and Ransomware

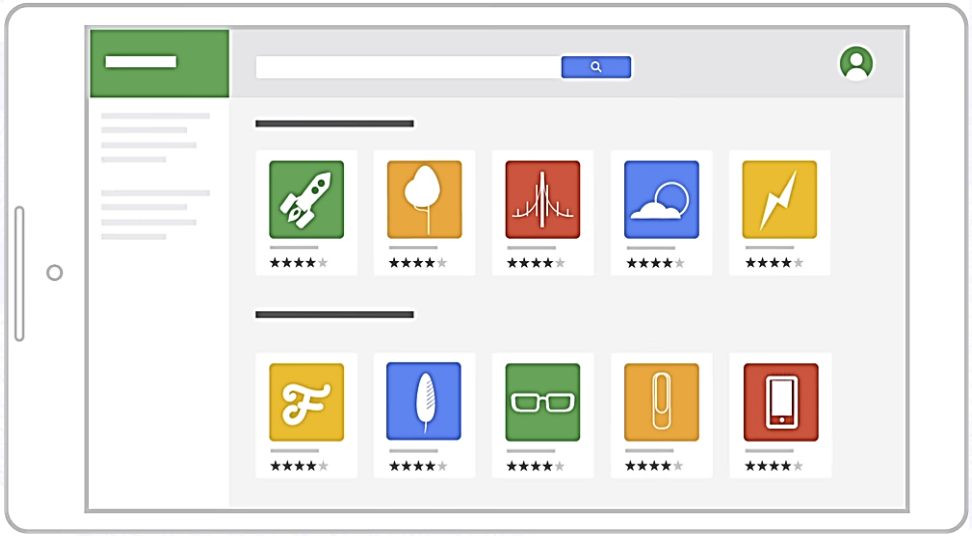
- Ransomware is when the contents of your system is locked and held for ransom
  - Or threaten to leak the data
- Typically “installed” on computers because:
  - Someone clicked on a bad link (phishing)
  - Did not patch their computers/servers
  - Downloaded “Free” software
  - Plugged in “Free” USB Drives
- Always keep a disconnected backup of your data

USC Annenberg
School for Communication and Journalism

USC University of Southern California

Downloads

Always download apps only from **trusted sources**



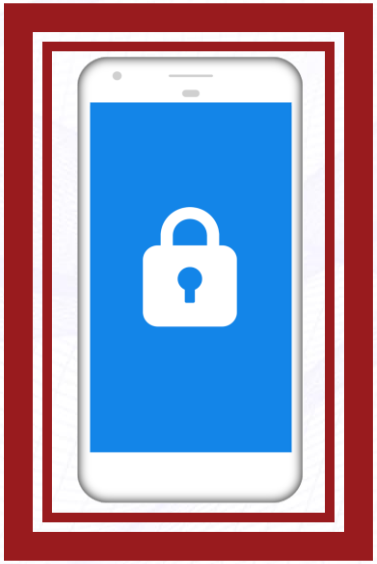
- ✓ Install ONLY the apps you really need.
- ✓ Each new app is potential risk.

USC Annenberg School for Communication and Journalism

USC University of Southern California

Tips to Protect Your Communications, Data, and Systems

- When working from home (or on the road)
  - Use your organizations VPN
  - Don't use same systems for "entertainment"
  - Be conscious of where you store sensitive data
  - Use your organizations IT Resources (email, desktops)
- Web Sites and e-mail, Chat, Voice, Video
  - Use SSL/TLS (https:)
  - Be vigilant about links, software and apps
  - End-to-end encryption
- Data at Rest (on your device)
  - Memory or Whole Disk Encryption (w/ Lockscreen/Passcode)



Today's Agenda

12:00 – 12:05 Introduction and Announcements
12:05 – 12:30 Security and Privacy – Influence on Elections
12:20 – 12:40 Influence of Social Media on Society
12:40 – 13:30 Disinformation
- 12:40 – 12:50 Adriana Nana – Deep Fakes
- 12:50 – 13:00 Social Bots
- 13:00 – 13:30 Influence on Public Discourse

13:30 – 13:40 Break
13:40 – 14:40 Regulation of Content
- 13:40 – 13:55 Types of Regulated Content
- 13:55 – 14:05 Technical approaches to detecting and regulating content
- 14:05 – 14:15 Resherle Verna – Should platforms have right of censorship
- 14:15 – 14:25 Section 230
- 14:25 – 14:50 Open discussion Regulation of Content

14:50 – 15:20 Current Event Discussions

Social Networks and Social Media (review)

Services that Enable us to:

- Share our thoughts and experiences
- Record intricate details of our lives
- Create communities of like-minded individuals
- Manage our relationships with others online.

The intersections of technology with social interaction.

Bulletin Boards, AOL, Myspace, Facebook, Twitter, WeChat, TikTok, Instagram, SnapChat, and many related services.

- But also includes email and the rest of the web.

Threat Vectors – Social Media (review)

Our use of social media – discloses our interests.
Others use of social media – to learn about us.
Monitoring and surveillance of Social Media
False information in social media
Reputation and permanence
Many forms of impersonation
Inferences from network analysis
Social Engineering through Social media

This Week
Social Media Influence on Society

I previously assigned a “viewing”, and it is optional since not all of might have access to Netflix. If you can, please view the NetFlix documentary "The Social Dilemma". The link that follows is to the trailer (you should view the full documentary if possible):

- https://www.youtube.com/watch?v=uaaC57tcci0

Your phone and TV are tracking you, and political campaigns are listening in – Evan Halper - LA Times 2/20/19

It was a crowded primary field and Tony Evers, running for governor, was eager to win the support of officials gathered at a Wisconsin state Democratic Party meeting, so the candidate did all the usual things: He read the room, he shook hands, he networked.

Then he put an electronic fence around everyone there.

The digital fence enabled Evers' team to push ads onto the iPhones and Androids of all those attending the meeting. Not only that, but because the technology pulled the unique identification numbers off the phones, a data broker could also use the digital signatures to follow the devices home. Once there, the campaign could use so-called cross-device tracking technology to find associated laptops, desktops and other devices to push even more ads.

Welcome to the new frontier of campaign tech — a loosely regulated world in which simply downloading a weather app or game, connecting to Wi-Fi at a coffee shop or powering up a home router can allow a data broker to monitor your movements with ease, then compile the location information and sell it to a political candidate who can use it to surround you with messages. (more online)

November 20, 2018 - Study analyzes the impact of targeted Facebook advertising on U.S. elections

https://phys.org/news/2018-11-impact-facebook-advertising-elections.html

by Carlos III University of Madrid

A study analyzes the effectiveness of microsegmented political advertising on social networks such as Facebook. Republican Donald Trump's team spent $44 million on Facebook, running 175,000 different adverts during the 2016 election campaign, compared to a spend of $28 million by Democrat Hillary Clinton. These campaigns target Facebook users based on factors such as gender, location or political allegiance. This micro-targeted advertising on social media was highly effective in persuading undecided voters to support Trump, as well as in convincing Republican supporters to turn out on polling day.

In particular, it increased the probability that a non-aligned voter would decide to vote for candidate Trump by at least five percentage points, according to the results of the study. In Trump's case, the impact of the campaign was strongest among voters who used Facebook regularly, those who used it as their main source of news, and among voters without university or college-level education. Specifically, political micro-targeting was particularly effective when based on ideology, gender or educational level, much less so when based on race or age. "Our results show that learning about politics on Facebook does not make voters more informed, but does make them less likely to change their voting choice, which is very in line with the concept of political polarisation.

According to the authors, the paper contributes to an incipient body of literature that is using Facebook data in a completely privacy-preserving manner. The platform represents a novel and highly valuable data source to address important socio-economic questions. "Thanks to predictive analytics, companies like Facebook offer a toolkit for targeting voters at an extremely granular level based on their previous online behaviour. These online campaign channels are potentially very powerful political instruments. It is therefore vital that we understand how political campaigns on social media work, their impact on voter behaviour, and, ultimately, on election results," saysco-author Michela Redoano, associate professor at the University of Warwick Department of Economics.

Antonio Russo, another of the researchers from this multidisciplinary team, points out that Facebook's impact on turnout "suggests that social media has great potential for stimulating the political participation of people who would otherwise have lost interest in politics. In a world where confidence in democracy is dwindling, I believe this is good news. However, we still have much to learn about whether the information that voters are exposed to on social media really helps them make informed choices."

Six Ways the Media Influence Elections

https://journalism.uoregon.edu/news/six-ways-media-influences-elections

- Ask Donald Trump and he'll tell you journalists wield a lot of power over the U.S. political process. It's true that the media have played an important role in politics since the First Amendment established freedom of the press as a cornerstone of American democracy. Voters need information to make educated decisions, and it's journalists' job to give it to them.

- But can the media really alter the outcome of an election?
  1. **To cover or not to cover**
  2. **Bias, scripts and the polarization of America**
  3. **Social media: Echo chamber and direct line to the masses**
  4. **A picture is worth 1,000 words**
  5. **Data journalism: Fact-checking, polls and the self-perpetuating cycle**
  6. **Watchdogs of democracy**

Brexit and Social Media

- https://www.teamlewis.com/uk/magazine/this-week-in-social-brexit-and-social-media-its-not-eu-its-me/

- Firstly, and most obviously, social media has been a forum for users to share their opinions on the controversial and increasingly heated topic of whether the referendum decision and departure plan are right for the county. When it comes to this serious political issue, this could be potentially damaging as users are likely to be influenced by whichever message appears most often or most prominent rather than necessarily searching for the true facts. It has been found that during the time leading up to the referendum in 2016, the number of people supporting the 'Leave' campaign outnumbered the 'Remain' campaign on Twitter by 7 to 1 and the most used hashtags were #Brexit, #Beleave and #VoteLeave.

- As a result of this, it's more than likely that this played a part in the final decision about Brexit which led to the UK now having to leave the EU. Since then, new analysis has found that 'Remainers' are winning the most attention on social media with 61 in 100 Brexit stories shared with negative sentiment. Although it can be argued that social media is a reflection of the general feeling of the public towards the issue, it could also be quite troubling to consider that people not necessarily knowledgeable about politics are able to make huge decisions for the future of our country based on hashtags and memes they have seen online.

2017 French Presendential Election

- Seeing the very big picture of the French 2017 presidential election: Social media, fake news, and political communities http://www.cnrs.fr/en/seeing-very-big-picture-french-2017-presidential-election-social-media-fake-news-and-political

- CNRS and EHESS researchers analyzed nearly 60 million political tweets posted during the 2017 presidential election in France. They noted that fake news flagged by the *Le Monde* Decodex fact-checking website accounted for only 0.1% of all Twitter content, and that 73% of the bogus information was spread by two political communities. Their findings are published in *PLOS ONE* (September 19, 2018).

False information in Social Media

- Tweets from hacked accounts can affect markets, cause panic, or instigate conflict.
  - False Tweet on AP News Feed
  - Pro-Gun Russian Bots Flood Twitter After Parkland Shooting

Today's Agenda

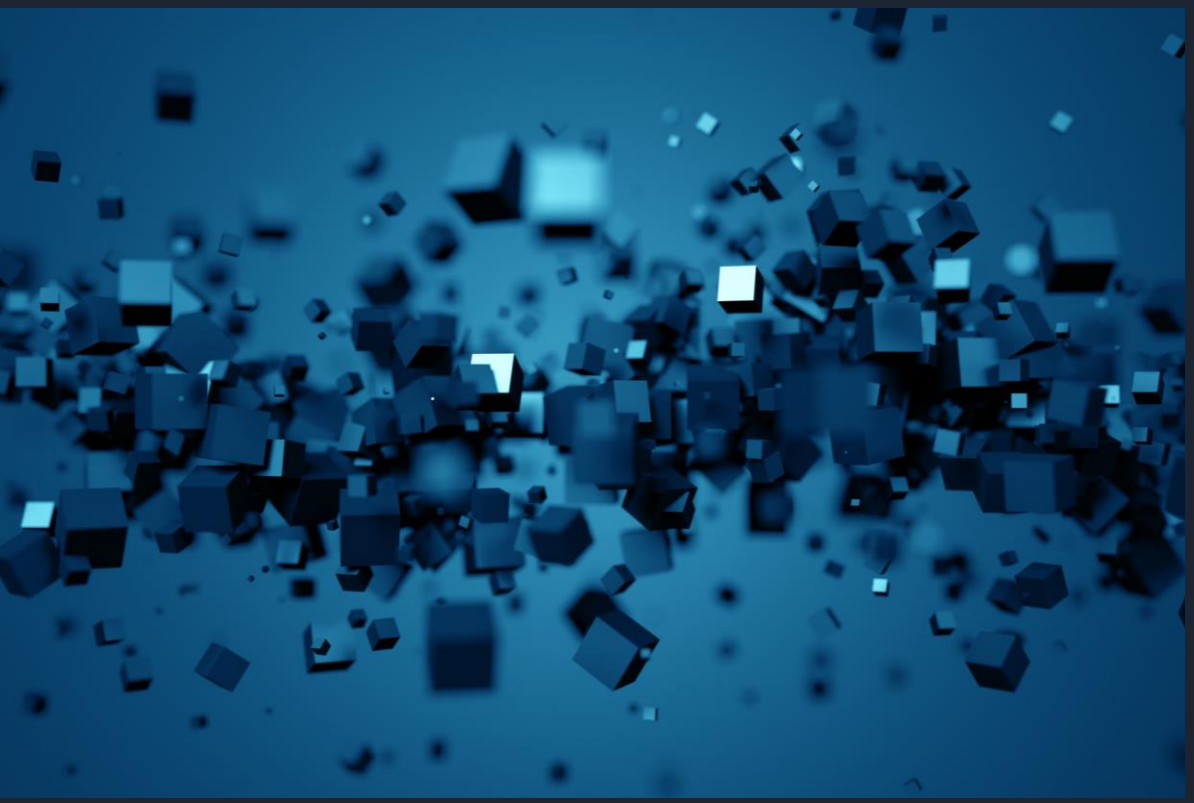
12:00 – 12:05 Introduction and Announcements
12:05 – 12:30 Security and Privacy – Influence on Elections
12:20 – 12:40 Influence of Social Media on Society
12:40 – 13:30 Disinformation
- 12:40 – 12:50 Adriana Nana – Deep Fakes
- 12:50 – 13:00 Social Bots
- 13:00 – 13:30 Influence on Public Discourse

13:30 – 13:40 Break
13:40 – 14:40 Regulation of Content
- 13:40 – 13:55 Types of Regulated Content
- 13:55 – 14:05 Technical approaches to detecting and regulating content
- 14:05 – 14:15 Resherle Verna – Should platforms have right of censorship
- 14:15 – 14:25 Section 230
- 14:25 – 14:50 Open disuussion Regulation of Content

14:50 – 15:20 Current Event Discussions

Technical Aspects of Disinformation

- New Techniques Deep Fakes
  - Deepfake ON SECURITY and Privacy - Nana Andriana
- Social Bots



DEEPFAKE ON SECURITY AND PRIVACY

Nana Andriana

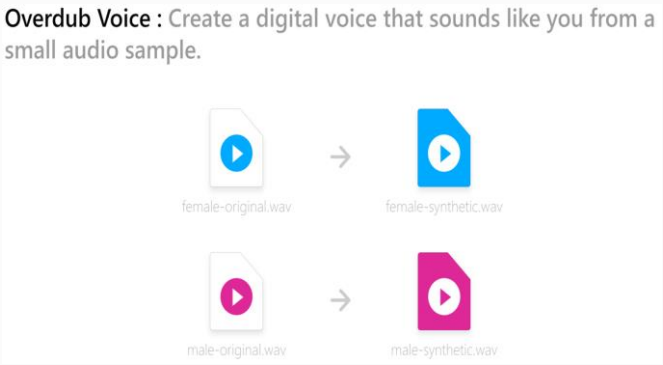
DSCI-529, April 09th 2021

Deepfake—a combination of the words ‘deep learning’ and ‘fake’—refers to an AI-based technology used to create or alter images, audio, and video resulting in synthetic content that appears authentic.

Technology:

- Neural network called an autoencoder.
- Generative Adversarial Network (GAN) attached to the decoder
- Constantly evolving, difficult to combat

- **In 1997, a paper "Video Rewrite: Driving Visual Speech with Audio" by Christoph Bregler, Michele Covell, and Malcolm Slaney.**
- **In 2016, a paper "Face2Face: Real-time Face Capture and Reenactment of RGB Videos" by Technical University of Munich.**
- **In 2017, a paper "Synthesizing Obama: Learning Lip Sync from Audio" by University of Washington. Add wrinkles and dimples and changed colors to better match lighting and skin tone.**
- **In 2017, a Reddit user named Deepfake spread doctored-pornographic video of celebrities. Deepfake went viral.**
- **In 2017, voice mimic by Lyrebird using as little as 1 min of voice recording.**



- **In 2018, a research "Everybody Dance Now" by UC Berkeley, expands the application of deepfakes to the entire body.**

Deepfake Development

Commercial development:

- Desktop app FakeApp (2018)
- Open source Faceswap (2019)
- Command line-based DeepFaceLab (2019)
- Web-based apps DeepfakesWeb.com (2019)
- Mobile app Zao (2019)
- Audio Deepfake
- Mobile app Impressions (2020)

- Most AI-based detections only **work best for celebrities**.
- Provenance of media using blockchain online ledger system (**not foolproof but tamperproof**). White-list evidence concept: if originality can not be proven then may assume fake. Weak approach!
- In 2018, a paper "In Ictu Oculi: Exposing AI Generated Fake Face Videos by **Detecting Eye Blinking**" by University at Albany.
- Current deepfake **now can blink!**
- In 2021, University at Buffalo computer scientists in paper "Exposing GAN-Generated Faces Using Inconsistent Corneal Specular Highlights", **detect deepfake photos by analyzing eye's light reflections. Claim to be 94% accurate**.
- New deepfake evolution? Maybe soon~

No, Tom Cruise isn't on TikTok. It's a deepfake

Deepfake Detection

Cases

- **Financial fraud in 2019:** UK-based energy firm. Loss €220,000 or approx. $243,000 (WSJ)
- **Mischief-making:**
  - Elon Musk smoked → Tesla stock crashed.
  - Donald Trump flew home early from a Nato meeting.
- **Common citizen:** Mother 'used deepfake to frame daughter’s cheerleading rivals’ (BBC News, 2021)
- **Possible crimes:** Video call asking for money, pretending to be family

Benefits and Risks

Benefits:

- **Cheaper movie production (no need for real actor/actress to shot the movie)**
- **Better education experience (using historical figure to explain their event)**
- **Therapeutical treatment (for Alzheimer or post traumatic disease)**
- **Avatar usage for disabled people**

Risks:

- **Harm to Individuals or Organizations**
  - Exploitation (using identity theft to extract financial or command activities)
  - Reputational Sabotage (in some instances, debunking the fake may come too late to remedy the initial harm)
- **Harm to Society**
  - Encourage public distrust (by faking public officials taking bribes, saying or doing things that they did not).
  - Manipulate political campaign.
  - Instigate chaos to national order (by showing soldier killing innocents, shouting racist, or religious mocking).
  - Spread panic and terror (by showing news anchor/public figures declaring war or emergent pandemic).
- **The Liar's dividend :** Give base for a lying figure to avoid admitting authentic videos (claiming them as fake)

Disinformation in General

- **The Ohio State University study:** Effect on 2016 election, although its effect may be small it may affect result on a very close election.
- **Journal Science Advances (January 2019):**
  - <10% Americans share from fake news domains
  - Age >65yo are 7 times higher than <30yo (who less likely to consume news)
- **Study on Facebook users:**
  - Individuals reading fact-checking articles had not originally consumed the fake news at issue.
  - Individuals who consumed fake news in the first place almost never read a fact-check that might debunk it.
- **People fall back on debunked information in the absence of alternative causal explanations**

Snowballs on Fake News

Fake news is defined as information that is invented by people or governments that are “fictions deliberately fabricated and presented as nonfiction with the intent to mislead recipients into treating fiction as fact or into doubting verifiable fact.”

- **Human minds have tendency to be attracted more toward negative news than positive facts.**
- **NPR, comparison of 20 most popular fakes vs real:**
  - False stories (8.7 million engagements).
  - Real news stories (7.3 million engagements).
- **Need to understand “The Information Cascade” dynamic**
  - People stop paying attention to their own information →
  - Depend more on other’s judgement (assuming others know better than themselves) →
  - More confidence in passing along what others think →
  - More passing → More credibility

As cycle repeats, the cascade strengthens

Policy

Facebook's (and Instagram's) early 2020 policy in the run-up to the 2020 US election:

- Banned deepfake videos that are likely to **mislead viewers** into thinking someone "**said words that they did not actually say**".
- Does not extend to deepfakes meant as **parody or satire**
- Does not cover **"shallowfake"** (videos that are either presented out of context or are doctored with simple editing tools). Examples: Nancy Pelosi's slurred speech and Jim Acosta's microphone aggressivity.

Regulations

- **First Amendment** of US' freedom of speech protection. Does not include false speech nor obscenity.
  - Pornographic content's victims may sue for defamation.
  - Online platforms are **protected by Section 230(c)(1)** of the Communications Decency Act.
  - Section 230 **includes an intellectual property exception.**

- In 2018, The Malicious Deep Fake Prohibition Act ("MDFPA"). **May impose criminal offense**. Expired by the end of 2018.
- In 2019, The DEEPFAKES Accountability Act ("DAA"). **Watermark requirement**. Likely to be ignored.

- **California Law (2020):**
  - AB-730 Elections: Deceptive Audio or Visual Media. Forbid political influencing deepfakes within 60 days of an election. Exceptions for satire and parody.
  - AB-602 for pornographic deepfake victim's **cause**.

Privacy and Social Media

- **Your face becomes your new privacy!** Limiting freedom of expression.
- **When you speak publicly, you make yourself vulnerable.**
- **Danger of video-based social media** (such as Tiktok, Instagram, etc) that also record user's device information (breached of data would increase risk of social engineering attack).
- Tiktok and Facebook might ban use of deepfake videos, but this may also tell you to share your own real face (as deepfake resources).
- How about Zoom recording?

Future Developments (Opportunity)

- Promotion of deepfake debunking technologies.
- Alibi providing service (privacy-destructive life logging).
- Other form of identifications (using DNA-based identification.... or even body-implanted-666 chip as prophesied in Bible).
- Avatar-based social media.
- Facial and voice masking in communication app.

Thank You!

References:

1. https://medium.com/@songda/a-short-history-of-deepfakes-604ac7be6016
2. https://en.wikipedia.org/wiki/Deepfake
3. https://grail.cs.washington.edu/projects/AudioToObama/
4. https://www.youtube.com/watch?v=H153u186OGE
5. https://neurosciencenews.com/deepfake-eyes-ai-18029/
6. https://www.allerin.com/blog/can-blockchain-help-in-our-fight-against-deepfakes
7. https://www.bbc.com/news/technology-56404038
8. https://kslnewsradio.com/1934259/fake-news-what-it-is-and-how-it-can-influence-politics/?
9. https://www.californialawreview.org/print/deep-fakes-a-looming-challenge-for-privacy-democracy-and-national-security/
10. https://advances.sciencemag.org/content/6/14/eaay3539
11. https://instituteforpr.org/how-effective-is-providing-alternative-explanations-when-challenging-misinformation/
12. https://www.businessinsider.com/facebook-just-banned-deepfakes-but-the-policy-has-loopholes-2020-1
13. https://cardozoaelj.com/2020/01/19/deepfakes-deep-trouble/

Technical Aspects of Disinformation

- New Techniques Deep Fakes
  - Deepfake ON SECURITY and Privacy - Nana Andriana
- Social Bots

Social Bots

The first of our readings is by a researcher here at USC on the use of "bots", automated processes that post to social media or like posts ot influence the spread of information in social media.

- https://m-cacm.acm.org/magazines/2016/7/204021-the-rise-of-social-bots/fulltext?mobile=true
- Today's social bots are sophisticated and sometimes menacing. Indeed, their presence can endanger online ecosystems as well as our society.
- https://dl.acm.org/citation.cfm?id=2818717

The Future of Free Speech, Trolls, Anonymity and Fake News Online

https://www.pewinternet.org/2017/03/29/the-future-of-free-speech-trolls-anonymity-and-fake-news-online/

- **Many experts fear uncivil and manipulative behaviors on the internet will persist – and may get worse. This will lead to a splintering of social media into AI-patrolled and regulated 'safe spaces' separated from free-for-all zones. Some worry this will hurt the open exchange of ideas and compromise privacy**

New Readings

- [Levemore] Saul Levemore, "The Offensive Internet: Speech, Privacy, and Reputation"
- [Nissenbaum] Helen Nissenbaum, "Privacy in Context: Technology, Policy, and the Integrity of Social Life"

Today's Agenda

12:00 – 12:05 Introduction and Announcements
12:05 – 12:30 Security and Privacy – Influence on Elections
12:20 – 12:40 Influence of Social Media on Society
12:40 – 13:30 Disinformation
- 12:40 – 12:50 Adriana Nana – Deep Fakes
- 12:50 – 13:00 Social Bots
- 13:00 – 13:30 Influence on Public Discourse

13:30 – 13:40 Break
13:40 – 14:40 Regulation of Content
- 13:40 – 13:55 Types of Regulated Content
- 13:55 – 14:05 Technical approaches to detecting and regulating content
- 14:05 – 14:15 Resherle Verna – Should platforms have right of censorship
- 14:15 – 14:25 Section 230
- 14:25 – 14:50 Open disuussion Regulation of Content

14:50 – 15:20 Current Event Discussions

Reasons for Limiting Content

- Privacy
  - E.g. Right to be forgotten, Libel and Slander
- Legal Proceedings
- Hate Speech
- Community Standards
- Controlling discussion
- Religious Doctrine
- Influence, Propaganda
  - Fake News
- Criminal Activity
- Any other terms of service

This Week
Regulating Content

- Some questions to discuss?
  - What is the technology available for filtering content?
  - What are the limitations of these technologies?
  - What is the importance of unregulated speech on the Internet?
  - What issues arise because rules vary across jurisdictions?

- Some additional questions:
  - Is it possible to automatically scan content to determine whether it is appropriate for posting or reading/viewing?
    - A) Yes – we can use automated tools.
    - B) No – we need a human in the loop to make determinations
  - Who should be responsible for enforcing policies on appropriate content?
    - A) Governments
    - B) Tech companies who provide the platform where material is posted?
    - C) It should be based on flagging data (i.e. complaints) from users?
  - What kind of content should be restricted from distribution? (check all that apply):
    - A) Hate speech
    - B) Inaccurate content (e.g. fake news)
    - C) Content that is about an individual without their consent
    - D) Content that violates copyright or commercial rights of owners
    - E) Anything else

Techniques for Limiting Content

- Blocked by Platform
  - Moderated
  - Flagged by users
  - Automated Content Scanning
- Content Firewalls
  - On ingress to a region (company, network, country)
- Trolling
  - Harassment of those posting content not in line with ones views
- Taking down servers
  - May include seizing domain names
- Legal process
  - Slander and libel suits
- Political Correctness
  - Leads to self censorship of ideas

This Weeks Readings

- Facebook went to war against white supremacist terror after Christchurch
  - https://www.vice.com/en_us/article/vb5yk3/facebook-went-to-war-against-white-supremacist-terror-after-christchurch-will-it-work
- Right to be forgotten decisions (outside ones borders)
  - https://www.npr.org/2019/09/24/763857307/right-to-be-forgotten-only-applies-inside-eu-european-court-says
- Friction on access to information not deemed in line with “public” views
  - https://www.uscc.gov/sites/default/files/Molly%20Roberts%20May%204th%202017%20USCC%20testimony.pdf
- Pressure on companies that post content or that provide a platform for such content
  - https://www.vox.com/2019/10/7/20902700/daryl-morey-tweet-china-nba-hong-kong

Prohibited Content Types

Take-down of pirated material (e.g DMCA Take-down Notices)

- https://www.saklaw.net/continuing-dmca-abuse/

Libelous content (England, Germany/Austria)

- https://www.theguardian.com/books/2000/apr/11/irving.uk
  https://www.techdirt.com/articles/20190909/18005242954/hotel-owner-files-libel-suit-against-reviewer-calling-nazis-nazis-gets-support-austrian-court.shtml

Court proceedings (England)

- https://www.bbc.co.uk/academy/en/articles/art20130702112133630

Prohibited Content Types

Hate speech (Europe)

- https://www.politico.eu/article/germany-hate-speech-netzdg-facebook-youtube-google-twitter-free-speech/

Statements against or not in accordance with particular religions

- https://www.theguardian.com/world/2012/sep/19/muhammad-cartoons-freedom-expression

Regulations regarding addictive content targeted to children in China

- http://www.xinhuanet.com/english/2020-10/18/c_139448042.htm

Social Media Companies' Right of Censorship

Resherle Verna

DSCI 529 | Spring 2021

Overview

- Examples of Content Moderation
- Leading Opinions
- Section 230
- First Amendment Protection
- Clarence Thomas’ Essay

Examples of Content Moderation

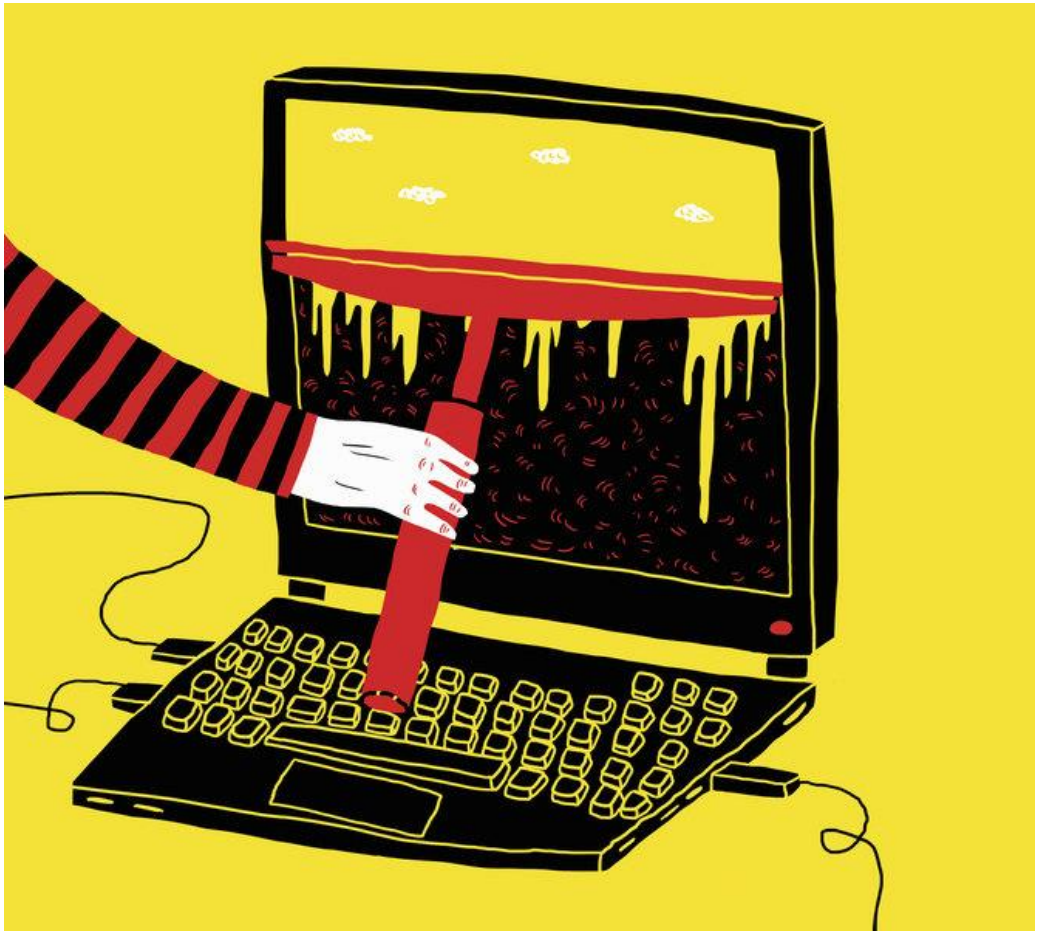
- Twitter banning former President Trump
- Amazon suspending services to Parler
- Alex Jones being banned from Facebook, YouTube, and Apple
- Laura Loomer being banned from numerous social media platforms, payment processors, vehicles for hire, and food delivery mobile apps

Leading Opinions

- Some conservatives say that social media companies are violating the first amendment when they moderate content
- When it comes to the general public, some believe the companies should have the right, whereas others believe otherwise
- Main reasons for content moderation:
  - The spread of misinformation
  - Content inciting violence

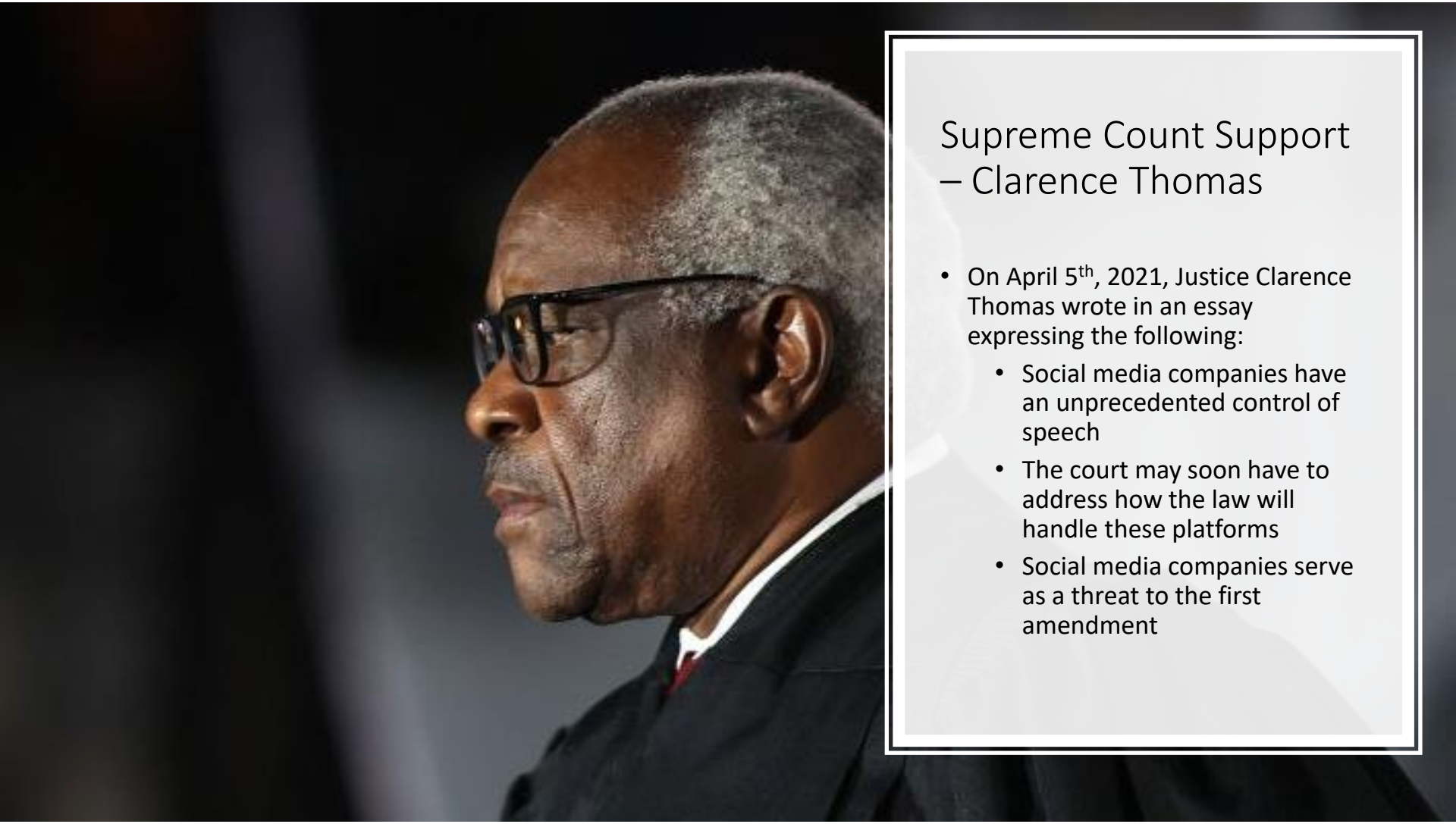
Section 230

- Internet legislation, passed in 1996, that frees internet companies from the responsibilities of traditional publishers
- Companies can’t be sued for content they host for which they haven’t assumed responsibility



First Amendment Protection

- The First Amendment applies only to the government, not to private actors
- Facebook, et al., *cannot* violate anyone's constitutional rights by removing their speech
- Social media companies themselves hold First Amendment rights, including the right to disassociate with expression they do not wish to host



Supreme Count Support – Clarence Thomas

- On April 5th, 2021, Justice Clarence Thomas wrote in an essay expressing the following:
  - Social media companies have an unprecedented control of speech
  - The court may soon have to address how the law will handle these platforms
  - Social media companies serve as a threat to the first amendment

Conclusions

- Social media companies are well within their right to perform content moderation
- Section 230 is too broad and needs stricter definitions
- Clarence Thomas' remarks may signal a new chapter in regulating social media companies
    - The fight over social media's power be a legal one, rather a legislative one

Works Cited

- https://www.forbes.com/sites/petersuciu/2021/01/11/do-social-media-companies-have-the-right-to-silence-the-masses--and-is-this-censoring-the-government/?sh=5b307b5848e2
- https://slate.com/technology/2021/04/clarence-thomas-social-media-first-amendment-censorship.html
- https://slate.com/technology/2019/02/cda-section-230-trump-congress.html
- https://www.nbcnews.com/tech/tech-news/court-online-justice-clarence-virginia-thomas-hit-big-tech-rcna611
- https://www.npr.org/sections/alltechconsidered/2018/03/21/591622450/section-230-a-key-legal-shield-for-facebook-google-is-about-to-change

This Week
Social Media Influence on Society

There is one more viewing that is related to our discussion last week, and section 230 as it protects certain organizations from liability regarding user posted content. This is a pretty good discussion, although I don't agree with all of the points made.
Censorship in Social Media: Section 230 and Free Speech

- https://www.youtube.com/watch?v=_DTFz61lmjY

FCC considering changes to section 230 rules, affecting responsibility for content posted online

- https://www.vox.com/recode/21519337/section-230-trump-fcc-twitter-facebook-social-media-ajit-pai

Here is a story from yesterday's Register that is relevant to our discussion in tomorrow's class.

- https://www.theregister.com/2020/10/27/facebook_nyu_ads/

Current Event Discussion

- http://csclass.info/USC/INF529/s21-lec12-ce.html