CSCI 1800 Cybersecurity and International Relations

Social Media and Propaganda

John E. Savage

Brown University
Outline

• Information Ecosystem
• Simple examples of information manipulation
• Information apocalypse scenarios
• Advanced AI technologies
• AI-supported impersonation
• Coping with the threat
• Politicization of information technology
• Coping with misuse of social media
Information Ecosystem

• Do social media prioritize clicks, shares, likes and ads over quality information?
  – Does this result in amplification of “fake news?”

• Has a critical threshold of addictive and toxic misinformation been reached?
  – Will faith in facts be lost on the altar of technology?

• Is technology addiction producing a loneliness epidemic?
Aviv Ovadya*

- In 2016 Aviv Ovadya “warned of an impending crisis of misinformation” in his Infocalypse talk!
- “We are so screwed it's beyond what most of us can imagine,” he said. “We were utterly screwed a year and a half ago and we're even more screwed now. And depending how far you look into the future it just gets worse.”

Impact of the Information Ecosystem

• **Technology addiction** identified in 1995 paper*
  – Clinical criteria for addiction:
    • Salience, Euphoria, Tolerance, Withdrawal symptoms, Conflict, Relapse
  – Signs and symptoms of technology addiction**

• **The loneliness epidemic†**
  – Nearly half of Americans say they are lonely! Global!
  – It makes people sick! Equivalent to 15 cigs/day!
  – Young people most at risk!
  – Amplified by social media

† https://theweek.com/articles/815518/epidemic-loneliness
Dangers of Information Ecosystem

• Social media tools coupled with AI allow us to
  – Enhance and distort reality
  – Create new realities
  – Launch influence campaigns
  – Amplify reactions, e.g. via trolls and bots
  – To do all of this at a distance

• Consequence:
  – Misinformation
  – Propaganda
Trolls and Bots

• New York Times video explaining trolls & bots (5:03):

• **Active measures** involve the following seven steps*:
  1. Find cracks in public issues that can be used to exploit divisions
  2. Create the Big Lie
  3. Wrap the big lie around a truth
  4. Conceal your hand
  5. Find a useful idiot who will promote the Big Lie
  6. Deny, deny, deny
  7. Play the long game

---

Lect19 4/9/20
© JE Savage
Examples of Artificial Outputs

• Changing audio in a video*
  – [https://youtu.be/9Yq67CjDqvw](https://youtu.be/9Yq67CjDqvw) (8:00)
  – Authors learn mouth shapes from speech
  – Superimpose shapes on stock footage

• Videos with full animated superimposed faces†
  – [https://www.youtube.com/watch?v=ohmajJTcpNk](https://www.youtube.com/watch?v=ohmajJTcpNk) (6:35)


2016 Predictions of Aviv Ovadya*

- Many slick, easy-to-use, and powerful tools to manipulate perception and falsify reality coming.
- These tools will provide the ability to distort truth at will and put core institutions at risk!
- Ovadya holds positions at U. Michigan Center for Social Media Responsibility and Columbia University Center for Digital Journalism.

Ovadya’s Disruptive Scenarios*

- **Diplomacy manipulation**
  - Create video with Kim Jung un declaring war
  - This would precipitate a diplomatic crisis

- **Polity simulation**
  - Create a fake grass roots campaign
  - Bombard offices of legislators with realistic pleas

- **Laser (targeted) phishing**
  - Spam users w. realistic fake messages from friends

* https://www.buzzfeed.com/charliewarzel/the-terrifying-future-of-fake-news
Assault on FCC Public Comment System*

• In ‘17 FCC solicited public input on net neutrality
  – Required under 1946 Administrative Procedure Act
• 23 million comments were received
  – More than all previous government feedback!
• Many comments generated by AI-driven bots.
  – Some fake, some not. Number fake is unknown
• Open platforms can be subverted by bots
  – Congress wants an investigation

* https://www.wired.com/story/bots-broke-fcc-public-comment-system/
Wired’s Analysis of FCC Comments

• Helped by FiscalNote, it studies public comments on behalf of corporations.
• Large-scale analysis* estimates > 1 million fakes
• Small-scale analysis† of 39 Nicholas Thompson
  – 6 confirmed bots
  – 11 form letters
  – 3 real Nicholas Thompsons
  – 19 unknown source
• This analysis suggests > half of comments are fake

Audio Impersonation

• Realistic impersonation of individuals using Adobe’s #VoCo

• What nefarious uses could be made of this technology?

• How can we protect ourselves?
Neural Networks

- Artificial neural networks (ANNs) trained to recognize various types of data, e.g. cats, words
- Inputs represented by bits, such as pixels for images or frequencies and intensities for audio.
- Outputs associated with categories
Neural Networks

- **Node** values are integers, **edges** have weights
- Values multiplied by weights, passed through non-linear function, giving integer outputs
- Weights adjusted to improve recognition
  - **Adjustments** made via backpropagation of errors

Multiple layers possible
Generative Adversarial Networks (GANs)

- GANs are pairs of competing neural nets
  - One net generates examples
  - Second net evaluates the examples
- Competition drives each net to improve
  - E.g. counterfeiters versus police
- GANs were invented by Ian Goodfellow in 2014 to make machine-learning systems smarter
  - The method is unsupervised
Comments on the Threat

• Anyone could make it “appear as if anything happened, regardless of whether it did or not.”
  – Aviv Ovadaya*

• GANs have both “imagination and introspection” They could set news consumption back 100 years
  – Ian Goodfellow

• Computational propaganda is now a reality!

* https://www.buzzfeed.com/charliewarzel/the-terrifying-future-of-fake-news
What Do Technologies Put at Risk?

- Erosion of authenticity
- Integrity of official statements
- Electoral outcomes
- Potential breakdown of society
Politicization of Information Technologies

• Russians saw dangers.* In 1999 invited views on
  – “Advisability of developing international principles that would enhance the security of global information and telecommunications systems & help to combat information terrorism and criminality”

• Russians meddled in US 2016 elections†
  – Launched “blend of hacking, public disclosures of private emails, and use of bots, trolls, targeted advertising”. In'16 Russia believed at war w. West**

* http://undocs.org/A/RES/53/70
† https://www.cfr.org/report/countering-russian-information-operations-age-social-media
Lect19 4/9/20 © JE Savage
Russian Motivation for Meddling*

- Goal is to weaken adversaries, i.e. neighbors, NATO and US, by any means
- Planted & spread false stories in 20th century
- Social media has made it easier & more effective
- **French thwarted Russian attempt in May 2017**
  - Law: can’t report on campaigns in last 48 hours
  - Citizens largely get news via traditional outlets
- Russians tried to exploit NFL protests in 2017

* https://www.cfr.org/report/countering-russian-information-operations-age-social-media
Characteristics of Social Media

• Citizens reveal their social/political preferences
  – Profiles easily assembled and analyzed
  – Example: Cambridge Analytica

• Easy to target individuals and groups
  – Messages personalized or received from friends
  – Different groups can get contradictory messages
  – False reports can be inserted into genuine outlets

• Lies spread faster than truth*

* http://science.sciencemag.org/content/359/6380/1146.full
What Can Be Done About It?*

• Improve the tools for authenticating documents
  – Images & audio can **cryptographically authenticated**
• Identify authoritative sources and networks
  – Major news outlets have high standards
  – Use recognized trusted international sources
• Transparency and discussion are effective
  – Quantify impact of influence campaigns
  – Identify and publish sources of misinformation
• **Regulation of exploitable social networks**
  – Study effects of social networks on human psychology

Review

• Information Ecosystem
• Simple examples of information manipulation
• Information apocalypse scenarios
• Advanced AI technologies
• AI-supported impersonation
• Coping with the threat
• Politicization of information technology
• Coping with misuse of social media