

CMSC 304 Computer Ethics

Dr. Cynthia Matuszek

1

Bookkeeping

- About the paper
- Quick exam overview
 - Just the ones you can't easily look up
 - 24 hour waiting period
- Deepfakes

2

About the paper: Content

- A deeper ethical analysis of a **broad** issue
- Select a topic (e.g., one from the list we went through on the second day of class)
- Identify the issues or ethical questions in this area
- Analyze the ethical, social, economic, and/or legal aspects of these issues
- Identify *at least* one ethical question identified to which you apply the ethical analysis framework
 - Analyzing possible policies, identifying stakeholders and consequences, and drawing and supporting a conclusion

3

About the paper: Content

- This is an extended analysis – go beyond that simple framework in two ways:
- **First**, you should look at the issues from multiple angles
 - For example, if you were to write an extended analysis of cryptocurrency, you might analyze the policies and issues from three perspectives: those of the crypto miners, ransomware attackers, and policymakers
- **Second**, you should also include *at least* two specific cases or incidents
 - For example, if you chose bias in AI, you might consider the cases of Timnit Gebru, the Optum health care algorithm, and Amazon's hiring algorithm, discussing the details of what happened and the ethical issues

4

About the paper: Format

- Approximately 4000-5000 words
- Approximately 8-10 pages
- Single-space 11- or 12-point Times
- You may work individually or in pairs
 - The usual group work rules apply
 - Both are responsible for the content of the entire work
 - It should be done together, not split into pieces that are merged at the last minute

5

About the paper: Integrity

- It goes without saying, but the paper, the topic statement, and anything else must adhere to the class academic honesty policy
- All writing that you submit should be your own, other than clearly delineated quotes with proper references
- All authors are fully responsible for integrity of the work

6

Topic Statement

- Topic statement: a short summary of your intended topic area
- Topic
- Main issues
- Specific cases/incidents if you have identified some
- A single short paragraph, but you are free to write more

7

Codes of Conduct

3. Something you have an obligation to do **per your profession's codes of conduct**, but not a legal obligation
 - Examples:
 - Don't accept tasks you aren't competent to do.
 - Give realistic time estimates for projects.
 - Giving examples that aren't **required by a CoC** (although they may be required by basic professionalism)
 - There's no **ethical** requirement to, e.g., dress professionally.

8

Blameworthiness

5. Blameworthiness (passive responsibility)
 - Wrong-doing – a norm was violated, something wrong happened.
 - Causal contribution – something they did caused the effect.
 - Foreseeability – they should have known that effect would occur.
 - Freedom of action – they must have acted freely, not been coerced.
6. Was Elon Musk blameworthy?
 - Something wrong definitely happened
 - Nobody is coercing him
 - Did he have **knowledge** of the possibility of deaths occurring?
 - Did he have any **control** over the situation?

9

ISPs

7. Making ISPs responsible for hate speech
 - Stakeholders: law enforcement; customers of the ISP; online platforms, e.g. social media; hate speech victims; hate speech perpetrators; ...
 - Sufficient detail – not, e.g., “users”
 - Positive and negative consequences – may reduce hate speech, **but** also may negatively impact users' privacy, may have free speech implications, may be unfair to ISPs (too difficult to implement)

10

Deepfakes

- What is a deepfake?
- “Deepfakes (a portmanteau of “deep learning” and “fake”) are synthetic media in which a person in an existing image or video” (or voice) “is replaced with someone else’s likeness.”^[1]
- Historical difference: using AI makes deepfakes easier and faster to produce than traditional photo retouching
- “Deepfakes have garnered widespread attention for their uses in creating **child sexual abuse material, celebrity pornographic videos, revenge porn, fake news, hoaxes, bullying, and financial fraud**”^[1]

^[1] <https://en.wikipedia.org/wiki/Deepfake>


11

- [Obama/Jordan Peele video]

13

Deepfakes generation

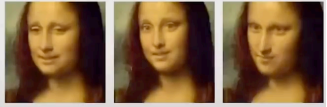
- Existing tools make it very easy
- Primarily for video work – voices are hard



14

First, some positive uses


- Education: Bring historical figures to life
 - JFK's speech on ending the Cold War, which was never delivered
- Entertainment
- Bringing art to "life"
 - Providing support for disabled persons
 - E.g., Deaf people



15

Positive uses

- Allowing anonymity for endangered people, e.g., journalists in autocratic regimes
- Reach and message amplification
 - E.g., Beckham's malaria message
- Innovation: allowing for new technologies
 - Trying actions or clothes
 - Decorating spaces



16

Deepfakes and politics

- The "post truth" era: "Relating to or denoting circumstances in which objective facts are less influential in shaping public opinion than appeals to emotion and personal belief"
- Truth is less important than emotional response/personal belief
- News sources are increasingly less held accountable for factual accuracy
- Spread of news obtained via social media increases the speed of misinformation
- "Once a political narrative is shifted, it's almost impossible to bring it back to its original trajectory"

en.wikipedia.org/wiki/Post-truth
www.theguardian.com/technology/interactive/2019/jun/22/the-rise-of-the-deepfake-and-the-threat-to-democracy

17

Imagine a deepfake pre-election

- [Trump voice video]

19

Deepfakes and pornography

- 96% of deployed deepfakes are nonconsensual pornography
- Revenge porn, blackmail and extortion, and public discrediting of women
 - Generally, technology doesn't currently work on men
 - Does, however, work on children 🟡
- Face-swap onto existing video
- Deepnudes – generate an undressed version of an image

20

Deepnudes

- [Fake nude and clothed pic side by side]

22

Deepfakes and pornography

- [Ryan Reynolds/Gene Wilder face swap video]

24

Results of sexually explicit media online

- Barriers to or loss of employment
- Harassment
- Social isolation
- Threats or acts of violence
- Inherent mental trauma

cyberforum.slate.edu/blog/2021/7/20/deepfake-pornography-beyond-defamation-law

25

Legal recourse

- Very limited – as usual, law has not kept pace with technology
- Revenge porn is banned in 46 states, but only 2 include faked porn
- If a copyrighted photo is used, IP law may apply
- If the victim can prove the perpetrator's intent to harm, it's possible to use harassment law
 - This is a very high bar
- Defamation law requires intent to deceive
- Right now, for the most part, deepfakes are legal

26

Deepfakes and scams

SCAMMERS USE AI-GENERATED VOICE DEEPPFAKE TO STEAL ALMOST \$250,000 FROM CEO

By Kristin Austin | September 6th, 2019



An energy firm is out almost \$250,000 from a high-tech phone scam that used artificial intelligence to mimic the boss's voice.

27

Deepfakes and scams

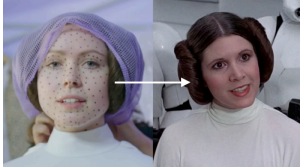
- Scammers posted on social media a deepfake video featuring Oleg Tinkov, the founder of Tinkoff Bank, which is one of the 15 largest banks in Russia. The “clone” of Tinkov called on people to use the bank's investment tools, promising to give every client 50 percent of the amount of investment as a bonus. Once users clicked on the link mentioned in a video, they were taken to a fake website of the bank and asked to share their names, emails, and phone numbers.
- The second scammer posed as an American who, after months of communication, told her that he had been imprisoned overseas. The woman asked “Buck” to help to release the prisoner, and “the admiral” made her transfer thousands of dollars “to pay the lawyer.”

slate.com/technology/2021/09/deepfake-video-scams.htm

28

Deepfakes and death

- Synthetically generated responses to questions
- Posthumous use of images
- Creation of new interactions



www.kidmoozers.com/2018/03/exploring-deepfakes.htm

29

Deepfakes and ethics

- **Control of identity**
- Lack of **consent**
- Obvious wrongs
 - Use to intimidate, humiliate, or blackmail
 - Scams
 - Misinformation and impact on people's beliefs
- Less obviously wrong
 - Posthumous use of people's identity
 - Deliberate use to extend reach (may still be deceptive)

30

Discussion

- What are the biggest positive and negative uses of deepfakes that you see/foresee?
- Given those: what policies should govern the use cases?
 - Should all deepfakes be banned? Probably not!
 - Is the current state unregulated state ideal? Also probably not!
 - What policies and laws should exist to **allow** positives but **prevent** negatives?

<http://tiny.cc/deep-fakes>

31