Explaining Why Fake Photos are Fake: Does it Work?

Margie Ruffin, Gang Wang, Kirill Levchenko

The ACM international conference on
Supporting Group Work 2023

https://mruffin.github.io

The New York Times

Worries Grow That TikTok Is New Home for Manipulated Video and Photos

Misleading edits, fake news stories and deepfake images of politicians are starting to warp reality on the popular video platform.

Give this article 88

Manipulated media is spreading before the midterms on TikTok, whose policies prohibit digital forgeries "that mislead users by distorting the truth." Kiichiro Sato/Associated Press

The New York Times

Worries Grow That TikTok Is New Home for Manipulated Video and Photos

Misleading edits, fake news stories and deepfake images of politicians are starting to warp reality on the popular video platform.

Give this article 88

Manipulated media is spreading before the midterms on TikTok, whose policies prohibit digital forgeries "that mislead users by distorting the truth." Kiichiro Sato/Associated Press

yahoo!

Photo of Joe Biden groping child is digitally altered

648

Bill MCCARTHY, AFP USA

Fri, December 30, 2022 at 3:48 PM CST · 2 min read

Social media users are sharing an image that purports to show Joe Biden touching a child inappropriately below the waist. But the photo has been digitally manipulated; the original shows the child stretching a T-shirt out to show the US president, who points in return.

"Is this a photo of Biden fondling a small child's genitalia?" said far-right radio host Stew Peters, who has previously spread Covid-19 misinformation, in December 30, 2022 posts, sharing the picture to Twitter, Telegram and Truth Social.

The same image spread across Twitter, Facebook and other social media platforms, using manipulated images as part of a baseless narrative -- fueled by the QAnon conspiracy theory and, in some instances, former president Donald Trump -- suggesting inappropriate conduct by the US president.

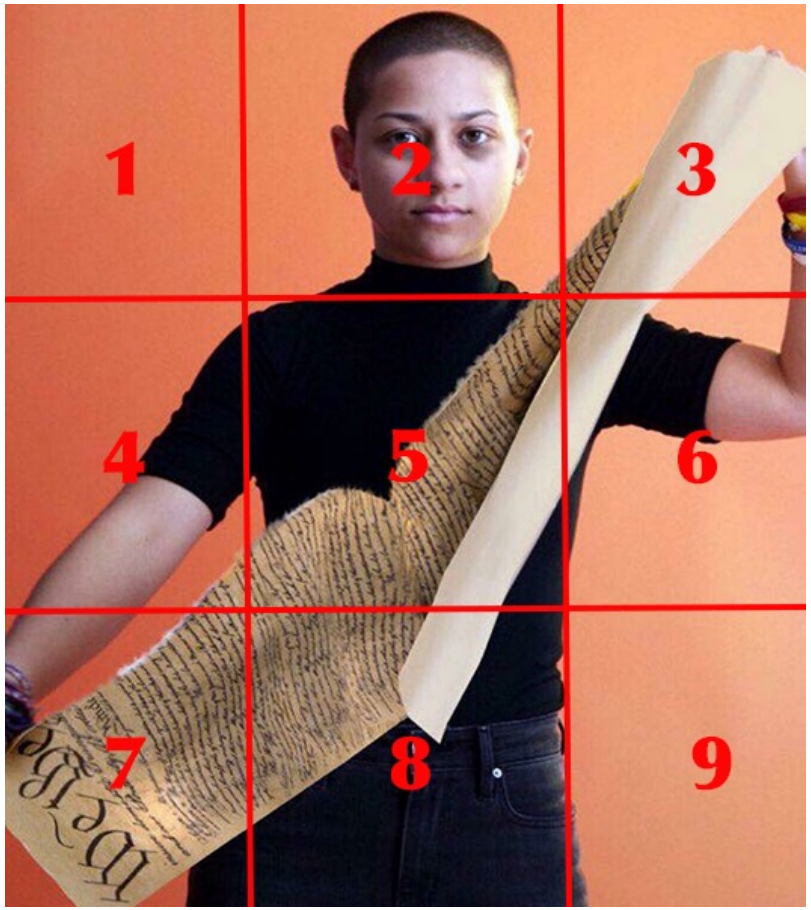
But the photo has been altered to change the position of Biden's hands. A White House spokesperson confirmed to AFP that the version circulating online is "not real."

We Ask the Following Questions…

R1: Can viewers detect manipulated images?

We Ask the Following Questions…

R1: Can viewers detect manipulated images?



We Ask the Following Questions…

R1: Can viewers detect manipulated images?

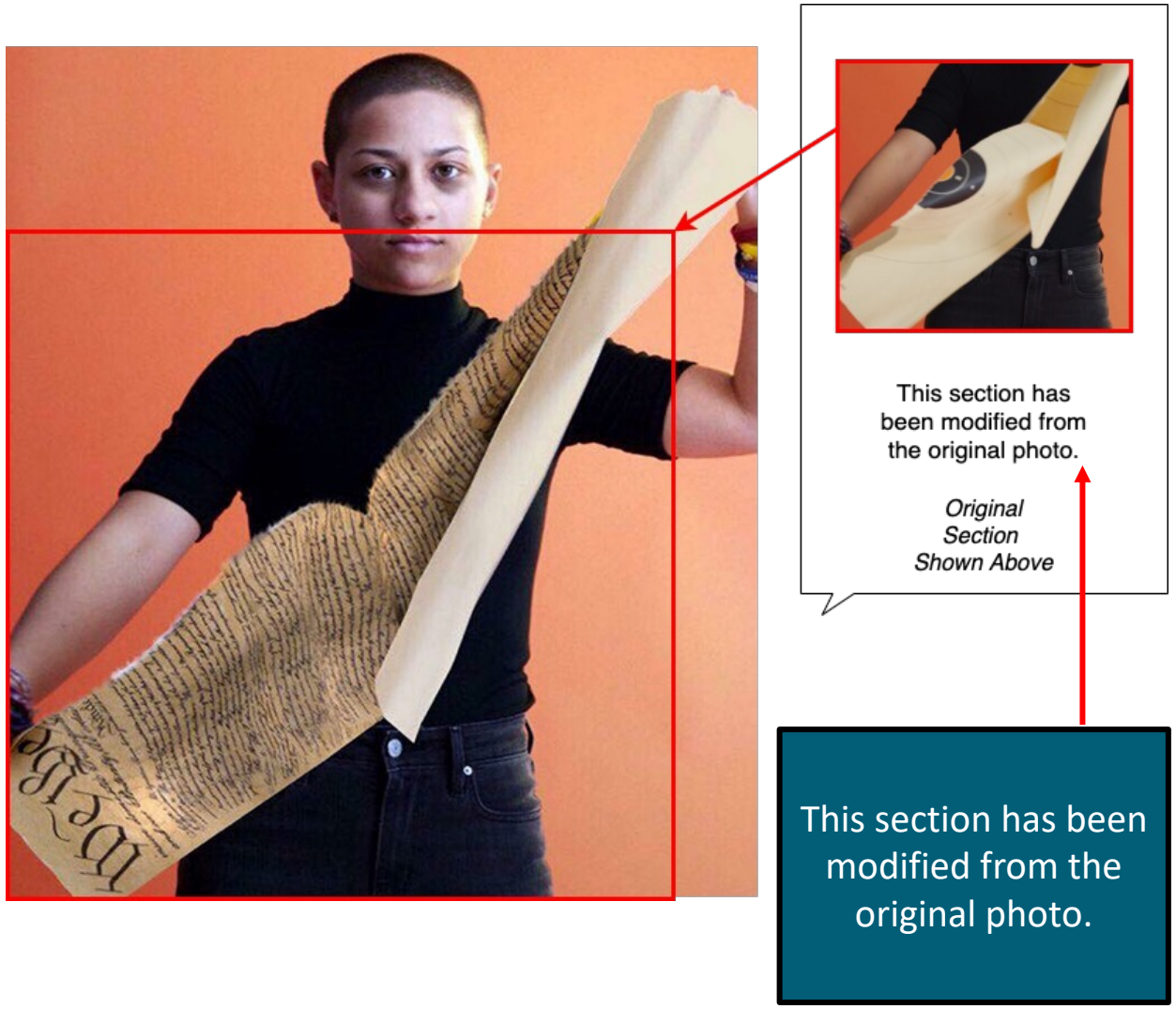
R2: Are viewers' opinions influenced by manipulated images?

We Ask the Following Questions…

R1: Can viewers detect manipulated images?

R2: Are viewers' opinions influenced by manipulated images?

R3: Does explaining how an image has been altered protect against the effects of image manipulation?



Data Collection + Processing

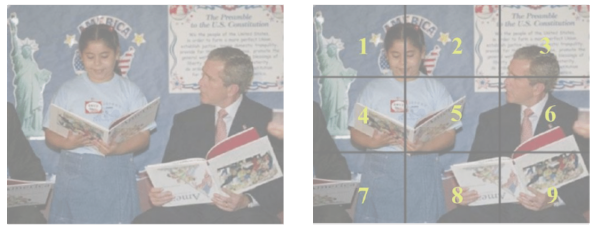
Fact-checking Site

Image Datasets from Prior Work

Social Media

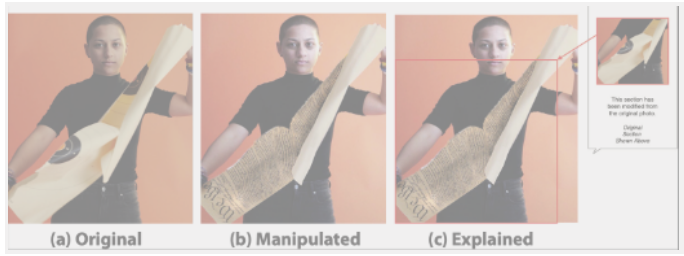
R1 Detecting Manipulation

18 Image Sets



R2 & R3 Mitigating Manipulation

8 Image Sets



Data Collection + Processing

Snopes

Fact-checking Site

Image Datasets from Prior Work

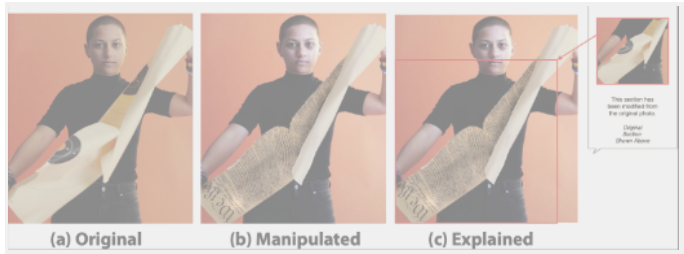
Social Media

R1 Detecting Manipulation

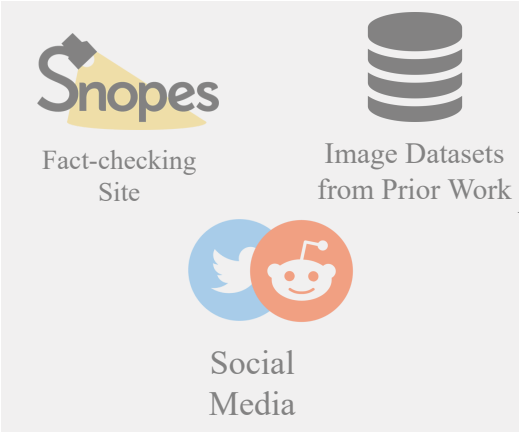
18 Image Sets

R2 & R3 Mitigating Manipulation

8 Image Sets

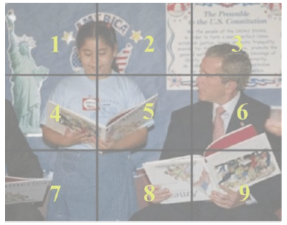


Data Collection + Processing



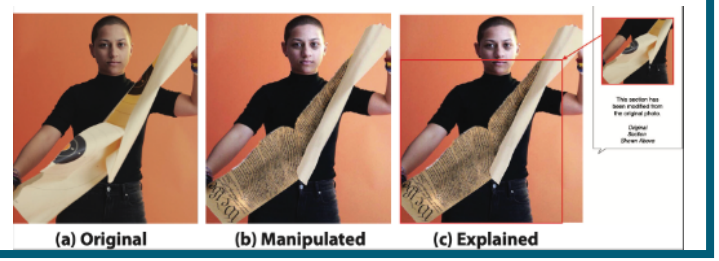
R1 Detecting Manipulation

18 Image Sets



R2 & R3 Mitigating Manipulation

8 Image Sets



Presentation Overview

Detecting Manipulation Survey

Mitigating Manipulation Survey

Summary of Findings

Presentation Overview

Detecting Manipulation Survey

Mitigating Manipulation Survey

Summary of Findings

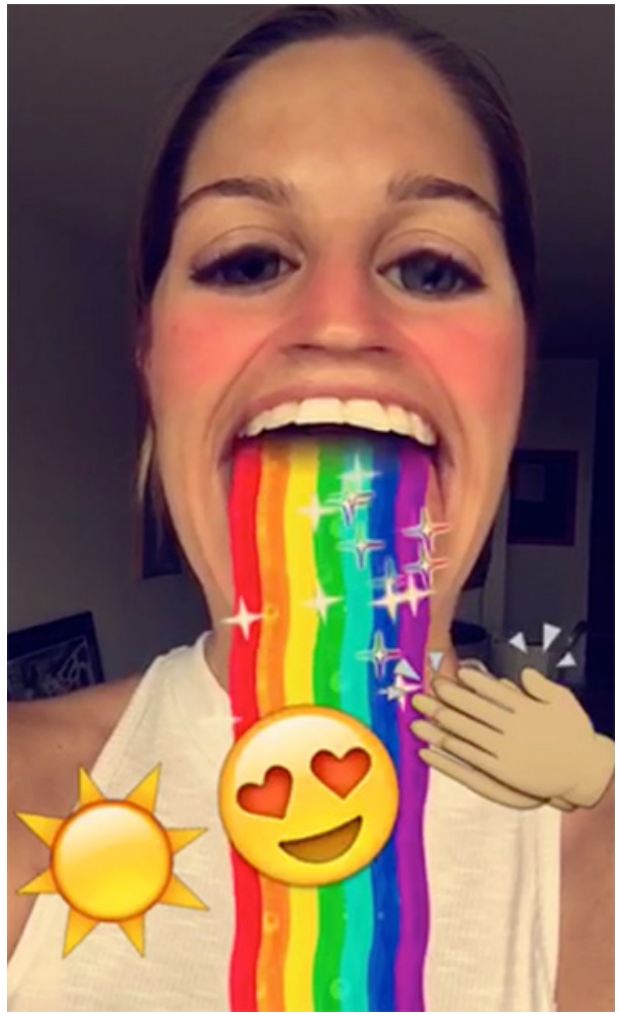
Detecting Manipulation

- 3 batches of survey w/ different images
  - Collected 18 image sets (original and manipulated)
  - Dataset split into 3 batches of 6 image sets

Detecting Manipulation

- 3 batches of survey w/ different images
  - Collected 18 image sets (original and manipulated)
  - Dataset split into 3 batches of 6 image sets

- Each participant was shown
  - 6 images, 3 manipulated & 3 original
  - 1 control image



Detecting Manipulation

- 3 batches of survey w/ different images
  - Collected 18 image sets (original and manipulated)
  - Dataset split into 3 batches of 6 image sets

- Each participant was shown
  - 6 images, 3 manipulated & 3 original
  - 1 control image

- Participants were asked to detect and locate manipulation

Does this photo appear digitally manipulated?

- Yes
- No
- I’m not sure

Does this photo appear digitally manipulated?

- Yes
- No
- I’m not sure

In which region is the alteration? Select all that apply.

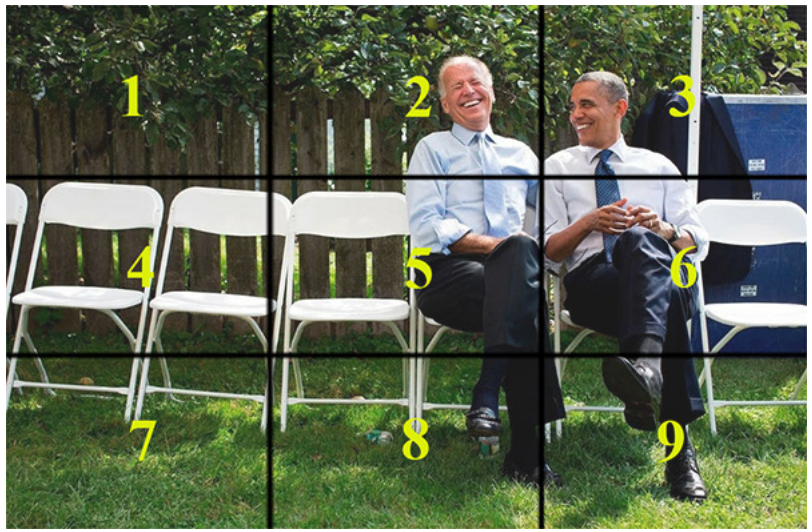
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- I don't think it's manipulated or I'm not sure

Does this photo appear digitally manipulated?

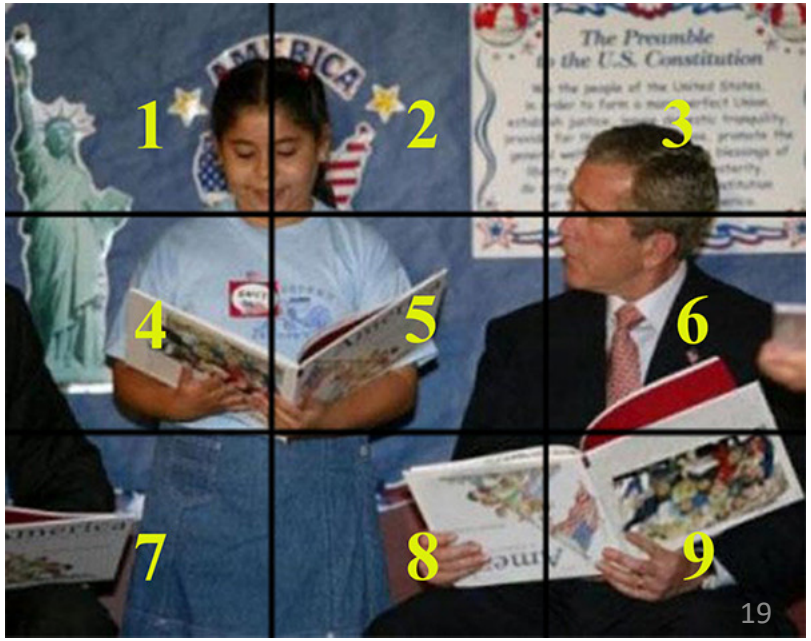
- Yes
- No
- I’m not sure

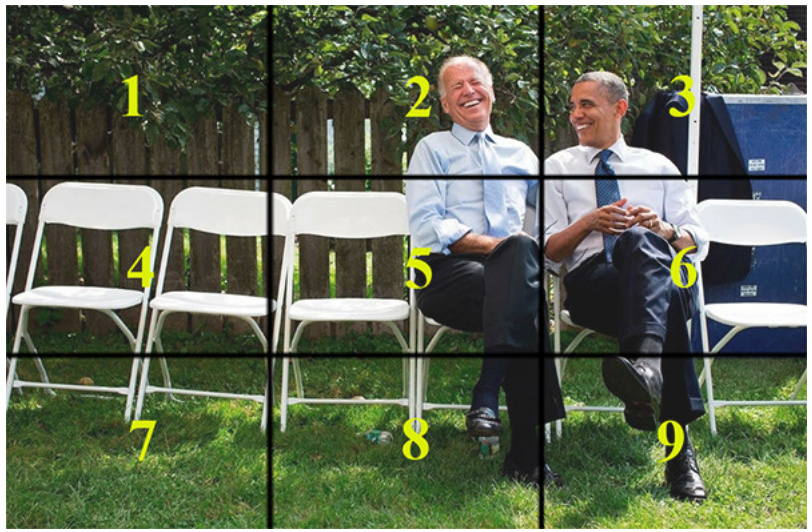
In which region is the alteration? Select all that apply.

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- I don't think it's manipulated or I'm not sure

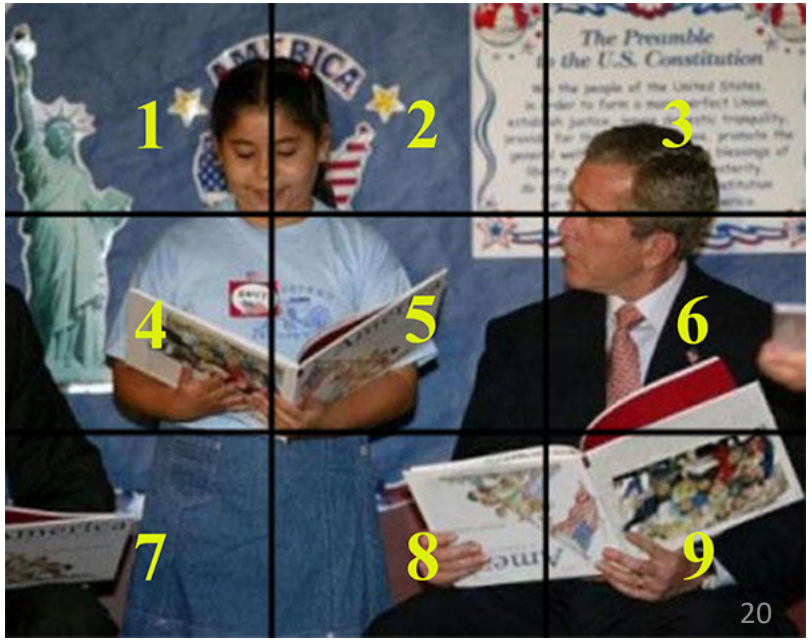

1
2
3
4
5
6
7
8
9

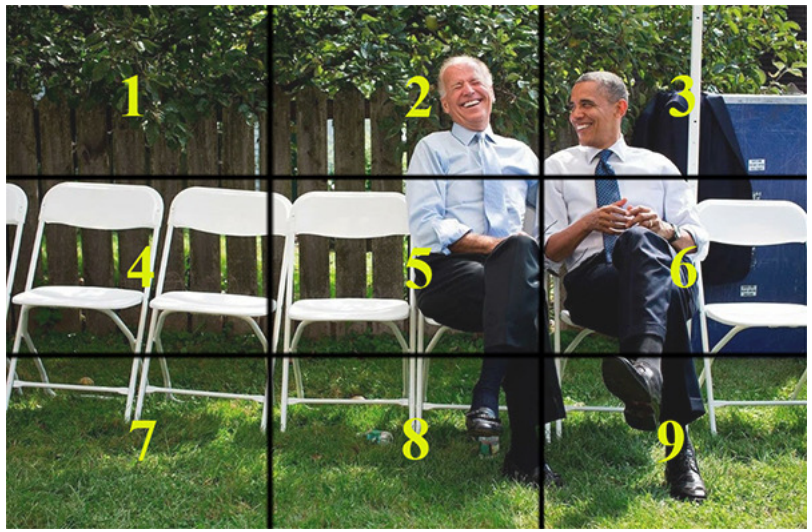
The Preamble
to the U.S. Constitution


The Preamble
to the U.S. Constitution
1
2
3
4
5
6
7
8
9



Answer: Not Manipulated





Answer: Not Manipulated

Answer: Manipulated

Are people good at it?

Manipulated Image?

Users' Answers	Yes	No
Yes	241	150
No	98	189

Identification Precision = 62%

Identification Recall = 71%

Identification Accuracy

60%

mean accuracy

Location Accuracy

38%

mean accuracy

Presentation Overview

Detecting Manipulation Survey

Mitigating Manipulation Survey

Summary of Findings

Mitigating Manipulation

- 3 Survey Groups
  - Original
  - Manipulated
  - Explained

Original

Participants are shown original images

Original

Manipulated

Participants are shown original images

Participants are shown manipulated versions of the images

Original

Manipulated

Explained

Participants are shown original images

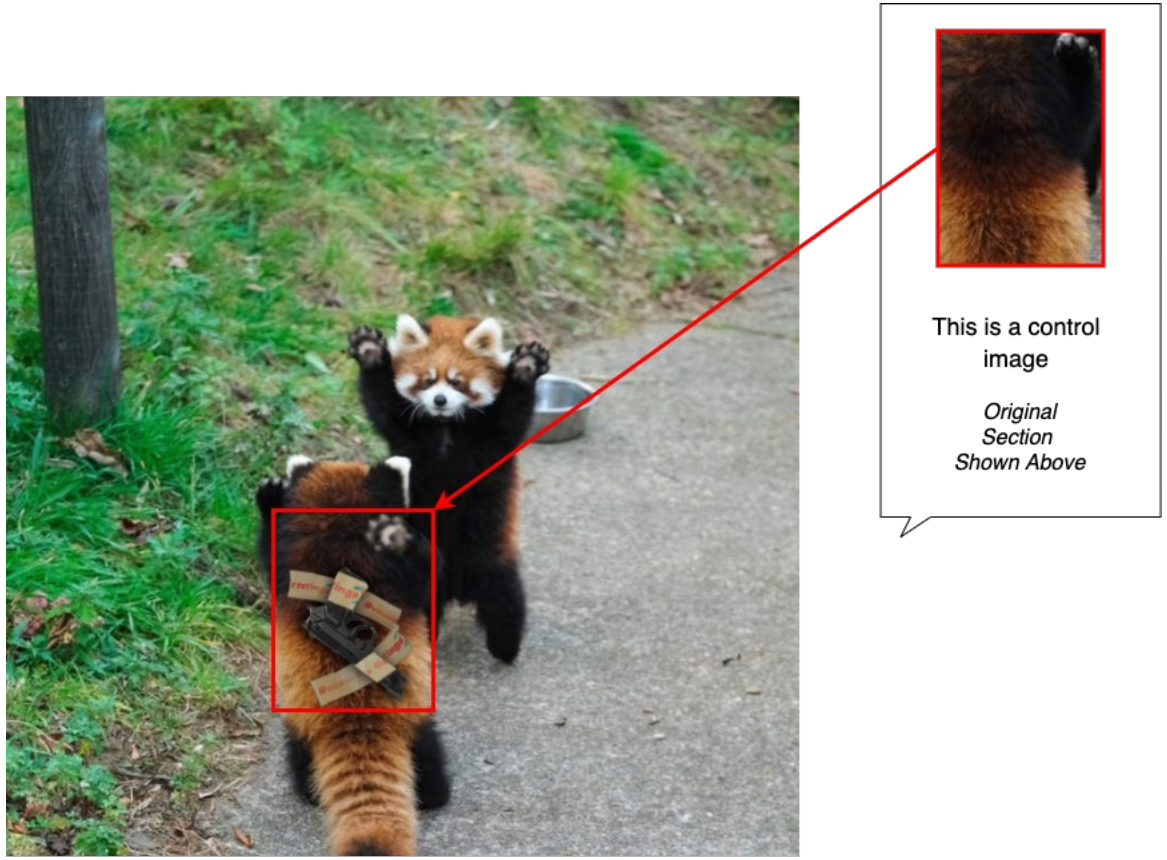
Participants are shown manipulated versions of the images

Participants are shown an image explaining how the manipulated image was produced

Mitigating Manipulation

- 3 Survey Groups
  - Original
  - Manipulated
  - Explained

- Each participant was shown
  - 4 images out of our 8 image sets
  - 1 control image



Mitigating Manipulation

- 3 Survey Groups
  - Original
  - Manipulated
  - Explained
- Each participant was shown
  - 4 images out of our 8 sets
  - 1 control image
- Each image was shown with a caption
  - Caption identifies the person or people pictured
  - Same caption used across images in set

Notable Survey Questions

Q1: Message Agreement

Do you believe that Democratic lawmakers—Ilhan Omar of Minnesota, Alexandria Ocasio-Cortez of New York, Rashida Tlaib of Michigan, and Ayanna Pressley of Massachusetts, also known as The Squad—hold extreme views?

Notable Survey Questions

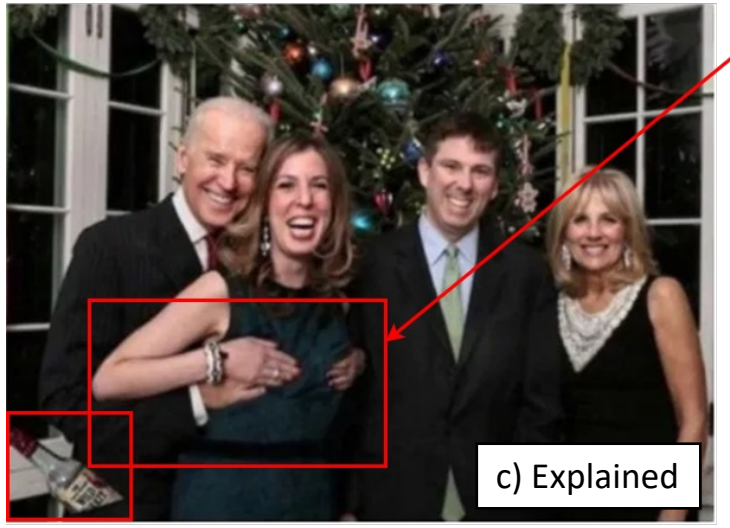
Q1: Message Agreement

Do you believe that Democratic lawmakers—Ilhan Omar of Minnesota, Alexandria Ocasio-Cortez of New York, Rashida Tlaib of Michigan, and Ayanna Pressley of Massachusetts, also known as The Squad—hold extreme views?

Q2: Subject Sentiment

From negative to positive, what is your opinion of the person(s) in this photo?

Do Explanations Work?

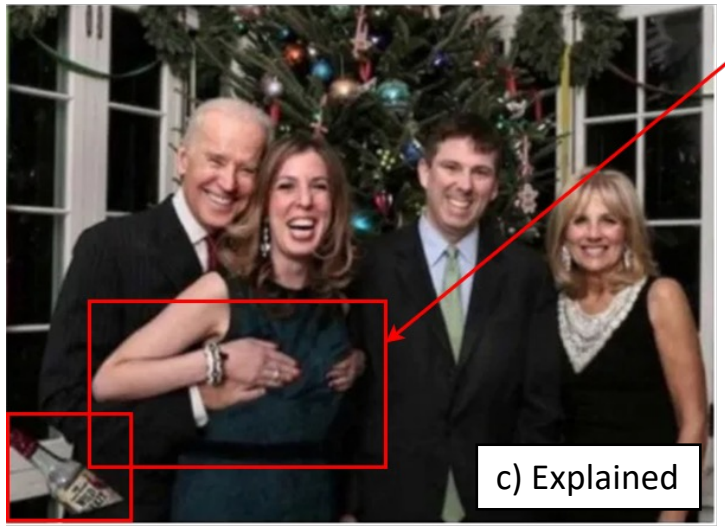


Do you believe United States President Joe Biden inappropriately touches women?

From negative to positive, what is your opinion of United States President Joe Biden in this photo?

Message Agreement		
Variable	β	p
Image (Reference = Manipulated)		
Original	-0.20	0.28
Explained	-0.57	**0.01****

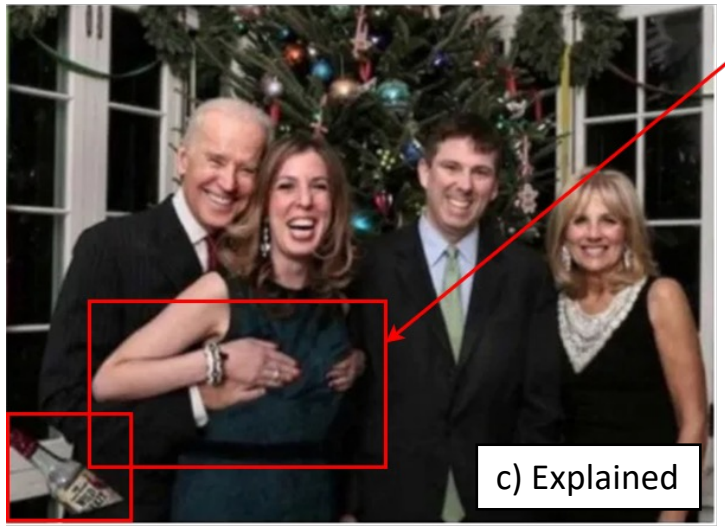
Subject Sentiment		
Variable	β	p
Image (Reference = Manipulated)		
Original	0.03	0.86
Explained	-0.53	**0.01****



Do you believe United States President Joe Biden inappropriately touches women?

Message Agreement		
Variable	β	p
Image (Reference = Manipulated)		
Original	-0.20	0.28
Explained	-0.57	**0.01****

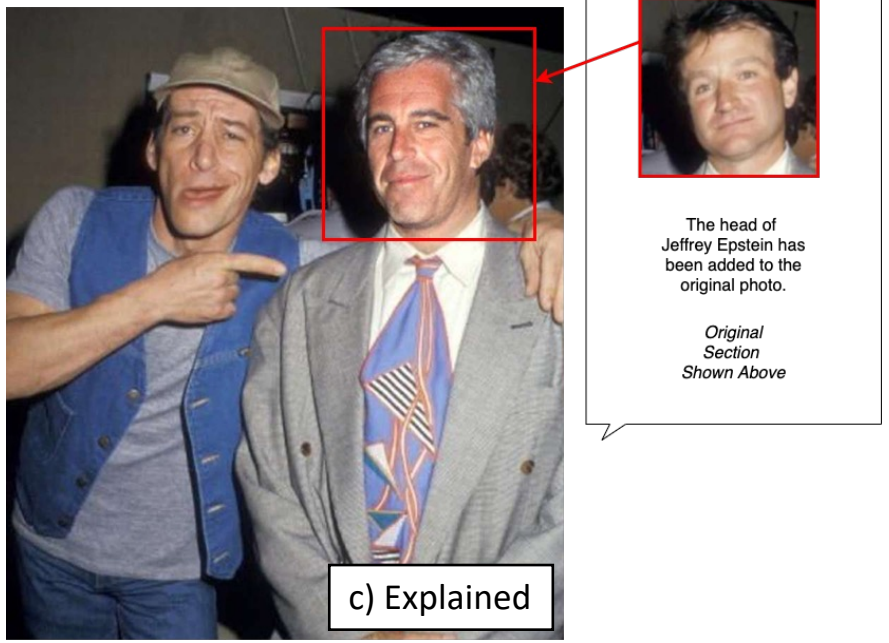
Viewers who saw Explained image were **LESS AGREEING** with the above statement.



However, those who saw the Explained image held a stronger **NEGATIVE** feeling about the subject.

From negative to positive, what is your opinion of United States President Joe Biden in this photo?

Subject Sentiment		
Variable	β	p
Image (Reference = Manipulated)		
Original	0.03	0.86
Explained	-0.53	**0.01****



Do you believe that actor Jim Varney may be guilty of the sex trafficking of minors?

From negative to positive, what is your opinion of the person on the left in this photo?

Message Agreement

Variable	β	p
Image (Reference = Manipulated)		
Original	-0.44	**0.02***
Explained	-0.58	**0.01****

Subject Sentiment

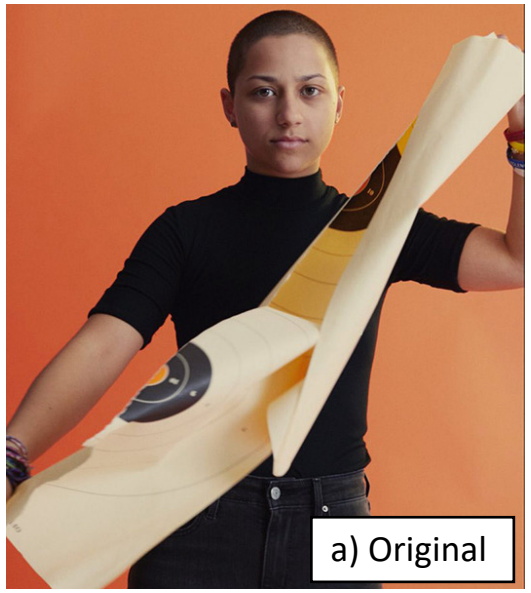
Variable	β	p
Image (Reference = Manipulated)		
Original	0.21	0.10
Explained	-0.40	**0.01****

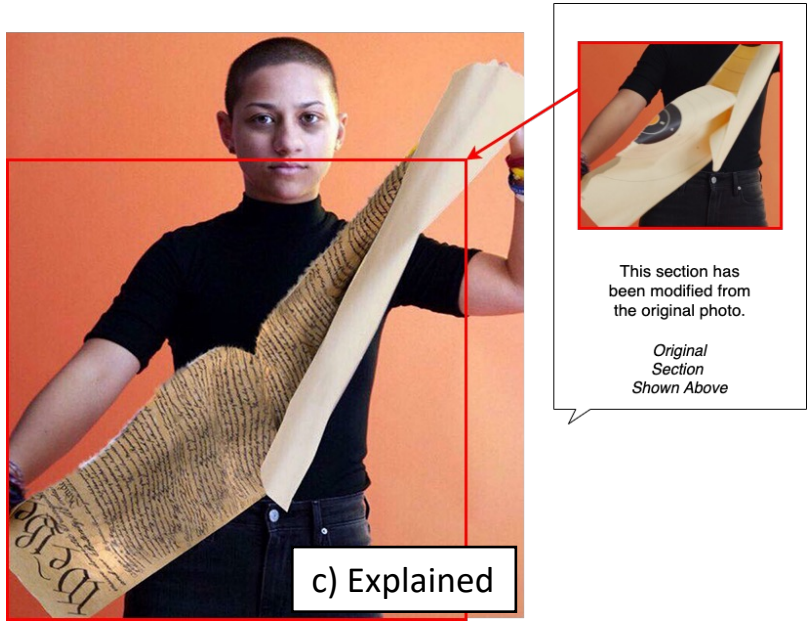
Do you believe that Democratic lawmakers … known as the squad, hold extreme views?

From negative to positive, what is your opinion of the person(s) in this photo?

Message Agreement		
Variable	β	p
Image (Reference = Manipulated)		
Original	0.12	0.53
Explained	-0.09	0.60

Subject Sentiment		
Variable	β	p
Image (Reference = Manipulated)		
Original	-0.06	0.64
Explained	-0.27	**0.03***





Do you believe that those who campaign against guns do not respect the Constitution of the United States?

From negative to positive, what is your opinion of the person in this photo?

Message Agreement		
Variable	β	p
Image (Reference = Manipulated)		
Original	-0.62	**0.01****
Explained	-0.44	**0.02***

Subject Sentiment		
Variable	β	p
Image (Reference = Manipulated)		
Original	0.20	0.14
Explained	0.01	0.93

Presentation Overview

Detecting Manipulation Survey

Mitigating Manipulation Survey

Summary of Findings

Embracing the Expected and Unexpected

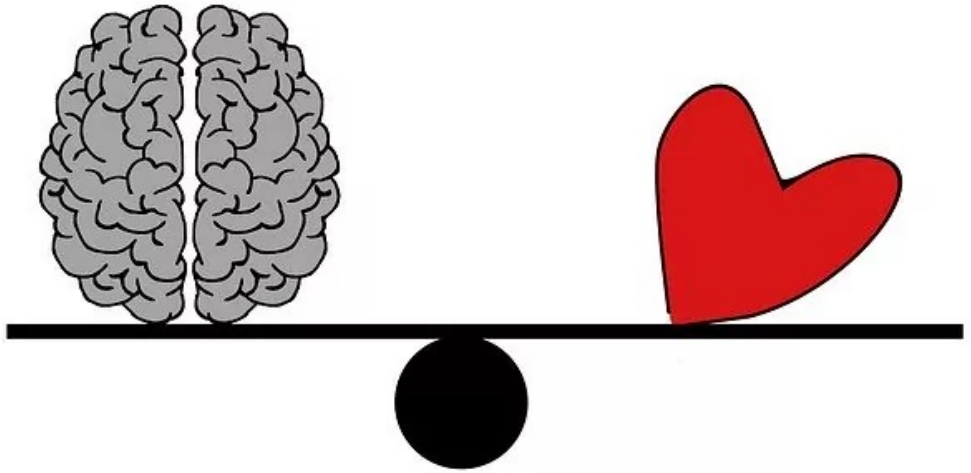
Detecting and Locating Manipulation

- Confirmed prior research:
  Users are not effective at determining whether an image was manipulated or not

  They are even worse when it comes to locating the manipulation

Attempting to Mitigate Manipulation

- Attempting to explain the effect of misinformation is trickier than we thought
- Showing a manipulated image with an explanation may still produce it's intended negative effect



System 2: Slow

Conscious Reasoning | Slow | Effortful

Message Agreement Questions invoked ***critical thinking***

System 1: Fast

Subconscious Emotion | Impulsive

Sentiment Questions invoked ***feelings***

Exploration Question:

Can we counteract the influence of manipulated images more effectively by nudging users to perform more deliberate thinking and reasoning? If so, how to facilitate this process?

[1] Daniel Kahneman. 2011. Thinking, fast and slow. Farrar, Straus and Giroux.

What does all this mean?

Attempting to explain how an image was manipulated should be done with caution

Showing a manipulated image, even in the context of an explanation, may still produce a negative effect

Thank You!

Summary:

- Users generally are not good at identifying manipulated images
  - They are even worse at locating the manipulated regions
- Simply highlighting and explaining the manipulation to users was not always effective
- Highlighting and explanation led viewers to hold less favorable opinions about the subjects pictured

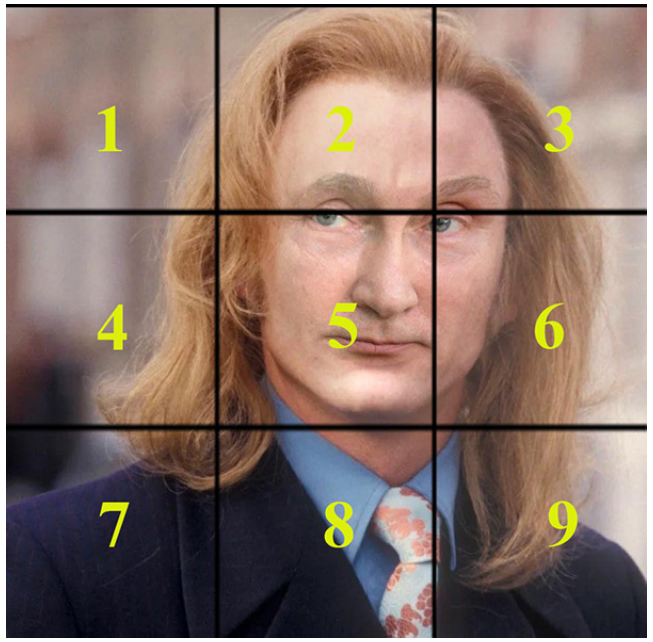
Margie Ruffin, Gang Wang, Kirill Levchenko

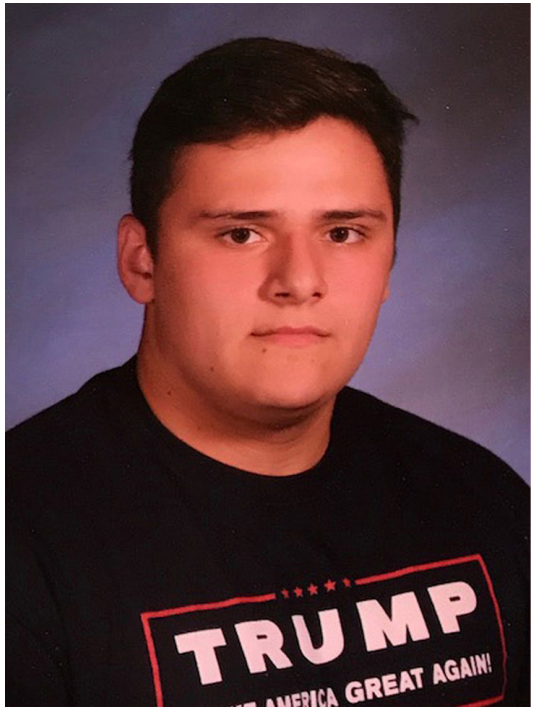
Thank You

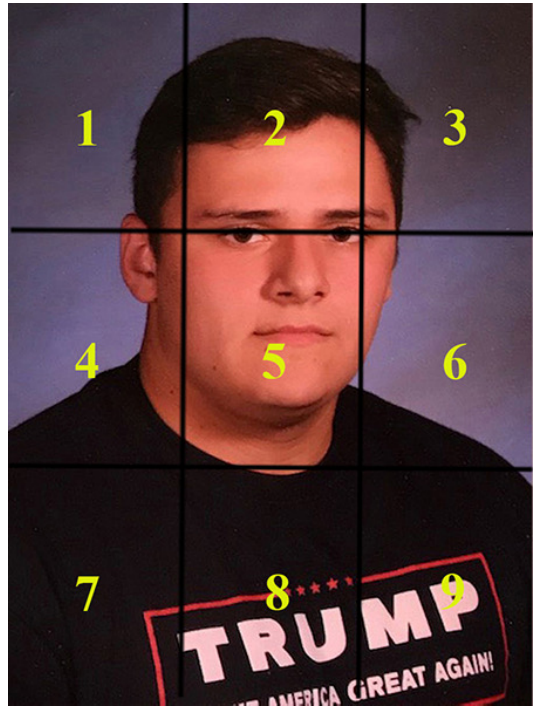
	Survey 1	Survey 2			
Attributes	**Total**	**Original**	**Manipulated**	**Explained**	**Total**
Gender					
Male	76	114	115	111	340
Female	37	58	71	68	197
Other	0	1	4	1	6
Age					
Under 18	0	0	0	0	0
18-24	9	17	16	4	37
25-34	16	83	83	69	235
35-44	7	46	57	62	165
45-54	62	15	6	20	41
55-64	19	9	17	18	44
65+	0	2	1	7	10
Not Disclosed	0	1	0	0	1
Education					
< High school	0	2	0	0	2
High school	5	12	8	15	35
Some college	54	12	9	9	30
Associate	30	9	15	10	34
Bachelor	13	114	118	117	349
Graduate	9	22	40	29	91
Not Disclosed	2	2	0	0	2
Total	**113**	**173**	**190**	**180**	**543**

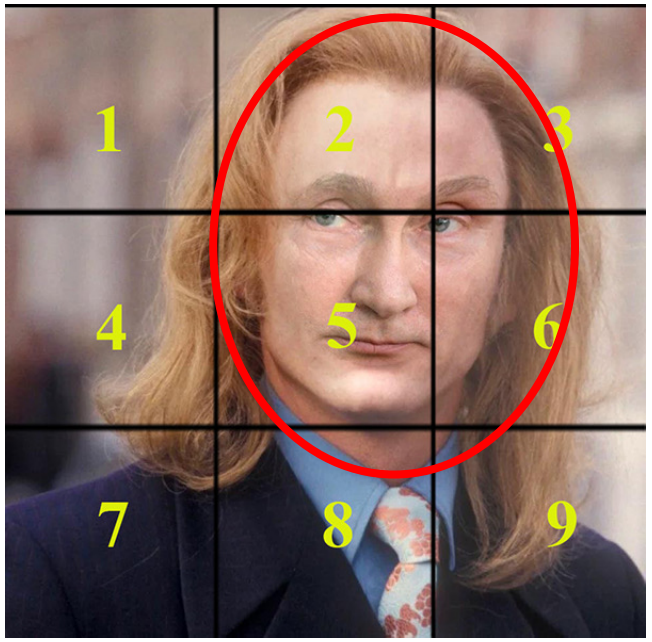
Table 1. Demographics— We show the demographics information of participants of Survey 1 (Detecting Manipulation) and Survey 2 (Mitigating Manipulation).

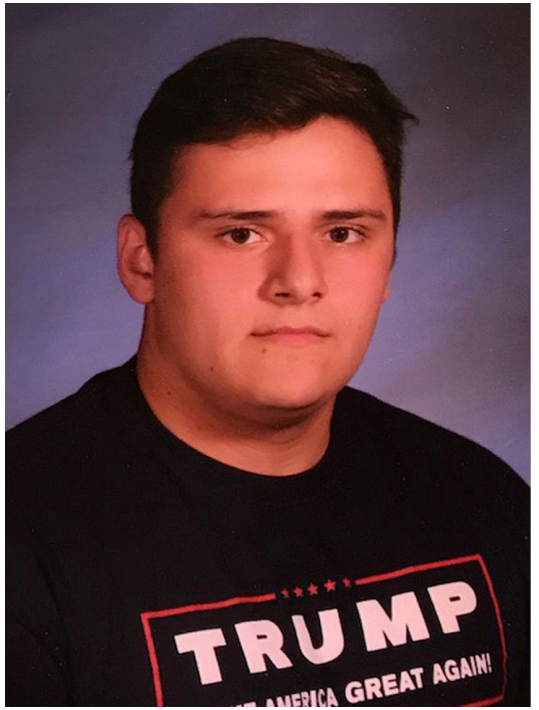
Detecting Manipulation Examples

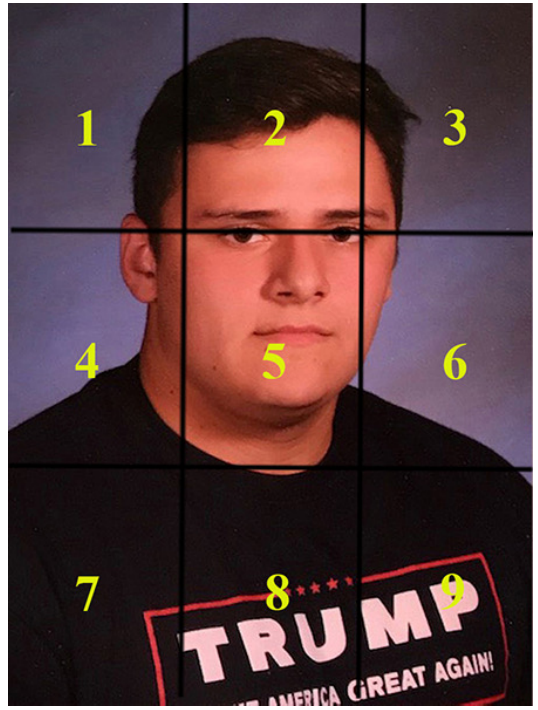

1
2
3
4
5
6
7
8
9


TRUMP
GREAT AGAIN!

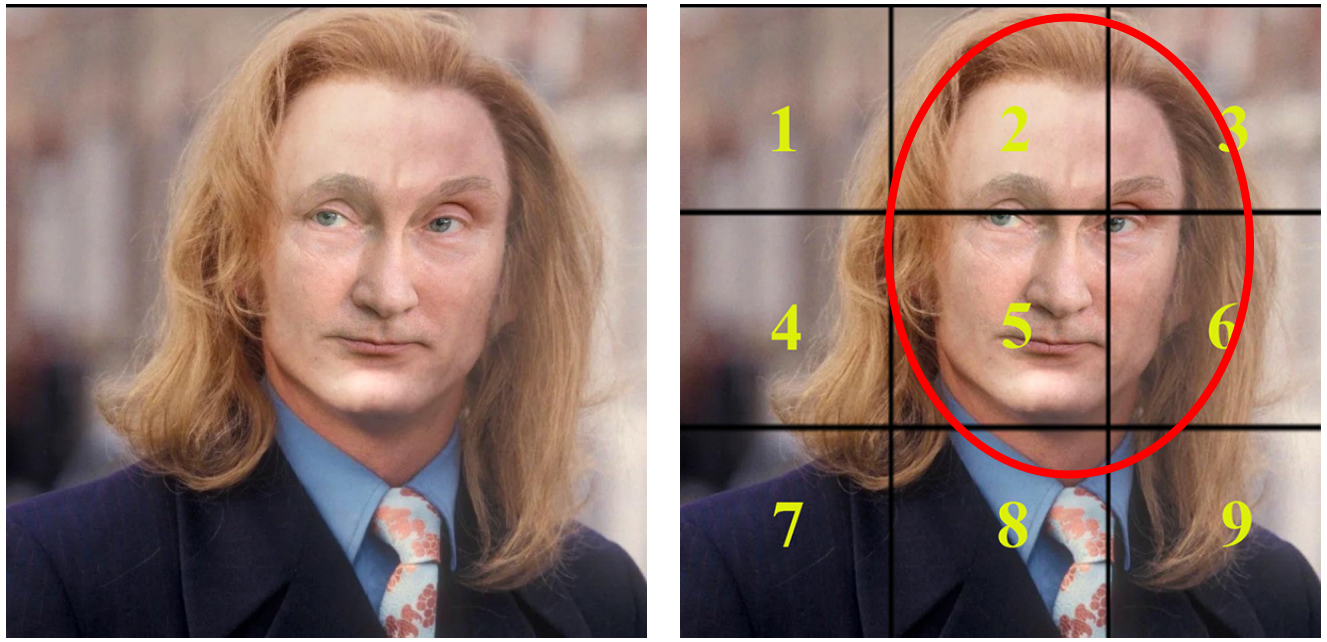

1
2
3
4
5
6
7
8
9
TRUMP
GREAT AGAIN!



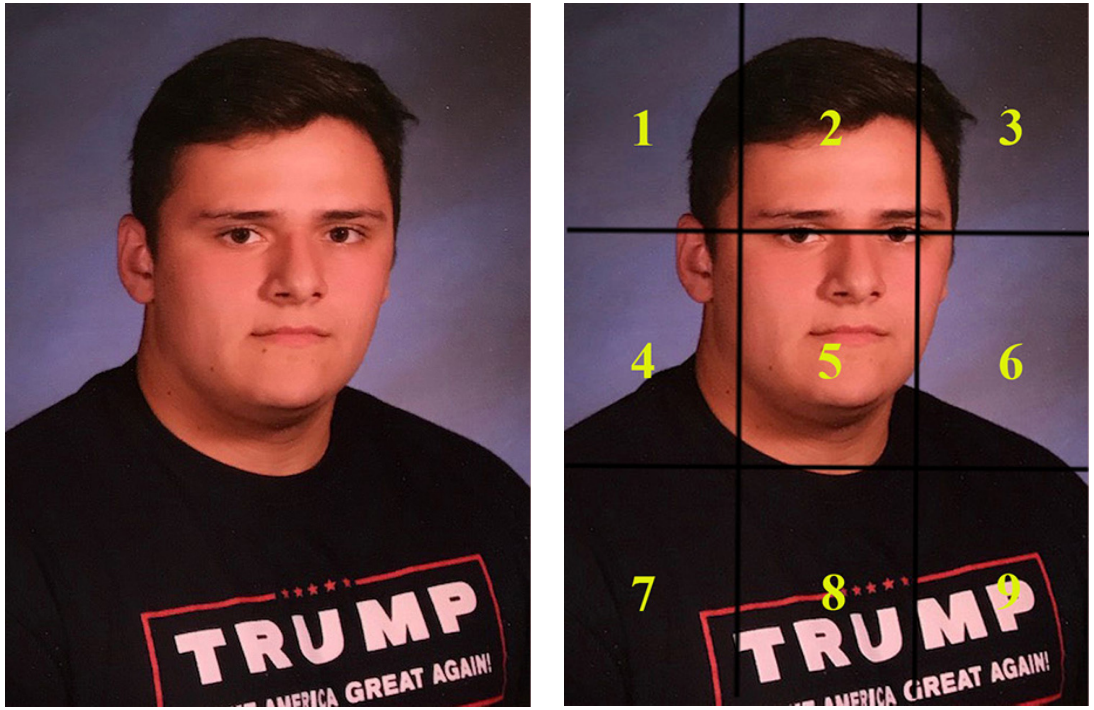




Answer:
Manipulated

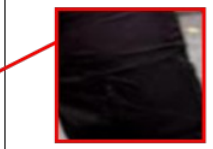


Answer:
Manipulated



Answer:
Not Manipulated

Mitigating Manipulation Examples



The Antifia symbol has been added to the original photo.

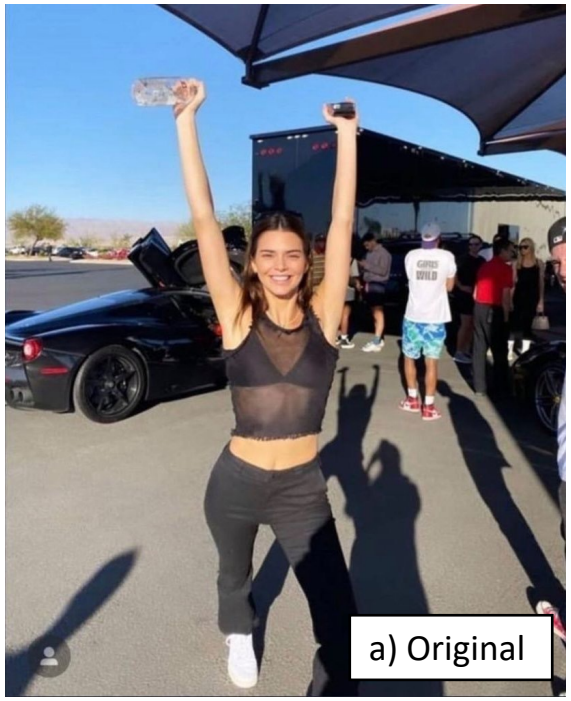
Original Section Shown Above

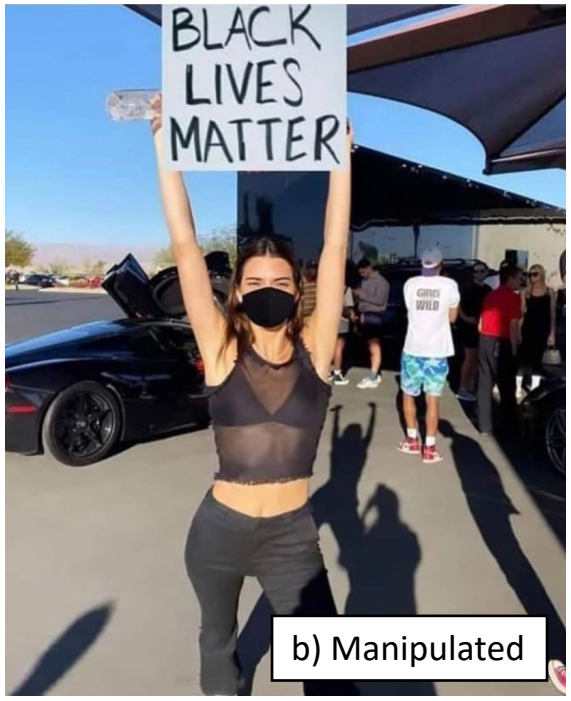
Do you believe Antifa protesters are violent?

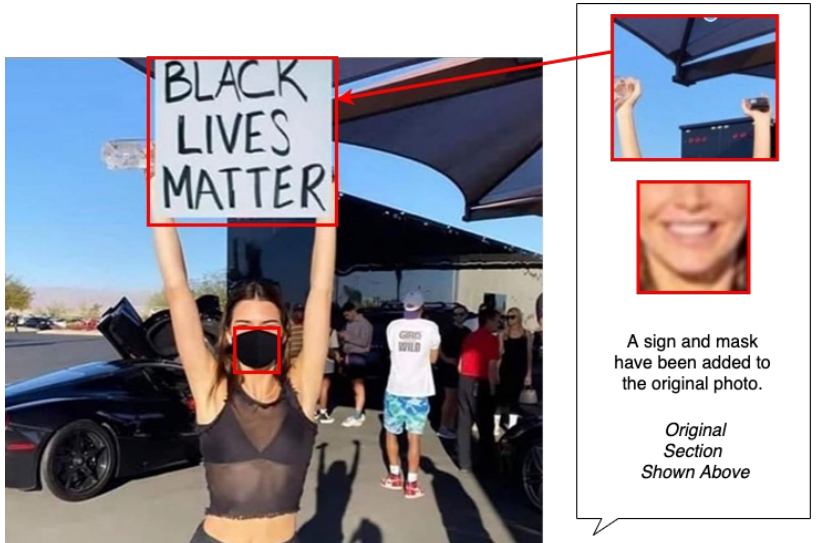
From negative to positive, what is your opinion of the person assaulting law enforcement in this photo?

Message Agreement		
Variable	β	p
Image (Reference = Manipulated)		
Original	-0.19	0.26
Explained	-0.25	0.14

Subject Sentiment		
Variable	β	p
Image (Reference = Manipulated)		
Original	-0.20	0.22
Explained	0.13	0.42





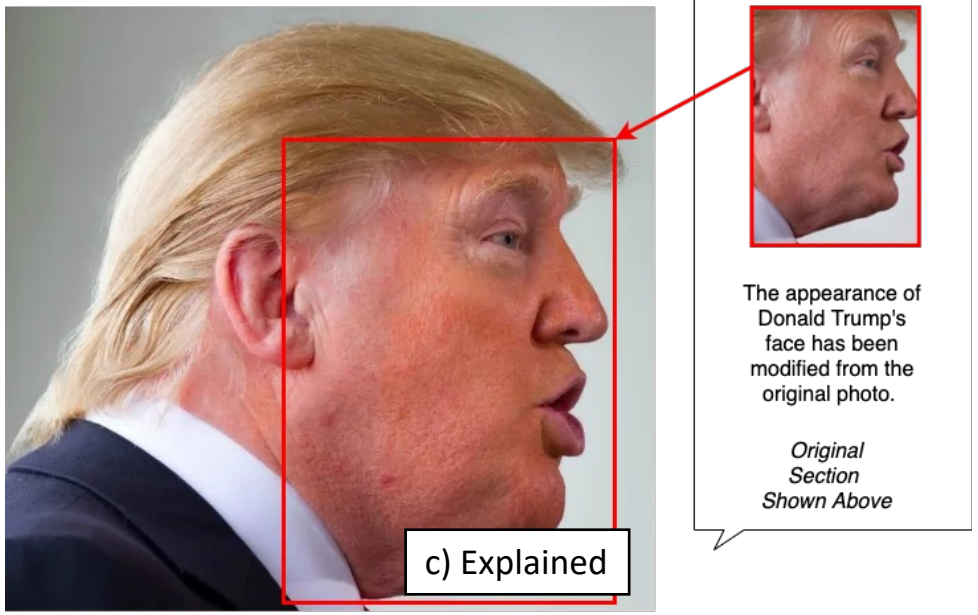


Do you believe that Kendall Jenner supports wearing protective face masks and supports the Black Lives Matter Movement?

From negative to positive, what is your opinion of the person in this photo?

Message Agreement		
Variable	β	p
Image (Reference = Manipulated)		
Original	-0.50	**0.001*** **
Explained	-0.15	0.31

Subject Sentiment		
Variable	β	p
Image (Reference = Manipulated)		
Original	-0.02	0.90
Explained	-0.21	0.10



Do you believe that former United States President Donald Trump is an unattractive person?

From negative to positive, what is your opinion of the person in this photo?

Message Agreement		
Variable	β	p
Image (Reference = Manipulated)		
Original	-0.04	0.85
Explained	-0.08	0.65

Subject Sentiment		
Variable	β	p
Image (Reference = Manipulated)		
Original	0.17	0.30
Explained	-0.08	0.62

a) Original

b) Manipulated

c) Explained
The appearance of Donald Trump has been modified from the orignal.
Original Section Shown Above