Oversight Board

Public Comment Appendix for

Case 2023-029-FB-UA

Case number

Case description

On October 29, 2022, President Biden went to vote early in-person during the 2022 midterm elections in the United States, accompanied by his adult granddaughter, a first-time voter. After they voted, they exchanged "I Voted" stickers. President Biden placed a sticker on his granddaughter, above her chest, according to her instruction, and kissed her on the cheek. This moment was captured on video.

In May 2023, a Facebook user posted a seven-second altered version of that clip. The footage has been altered so that it loops, repeating the moment when President Biden's hand makes contact with his granddaughter's chest. The altered video plays to a short excerpt of the song "Simon Says" by Pharoahe Monch, which has the lyric "Girls rub on your titties." The caption that accompanies the video states that President Biden is "a sick pedophile" for touching his granddaughter's breast in the way he does. It also questions the people who voted for him, saying they are "mentally unwell."

A user reported the content to Meta, but the company did not remove the post. The reporting user appealed and a human reviewer upheld the decision not to remove the content. As of early September 2023, the post had had fewer than 30 views, and had not been shared. The same user then appealed to the Oversight Board, stating the video was manipulated.

After the Board selected this case, Meta confirmed its decision to leave the content on the platform was correct. According to Meta's assessment, the Manipulated Media Community Standard did not warrant removal of the content because it only applies to videos generated by artificial intelligence or to those in which a subject is shown saying words they did not say. Meta decided

that the Misinformation or Bullying and Harassment Community Standards did not apply in this case either. Additionally, the content was not reviewed by independent fact-checkers as part of Meta's fact-checking program, although Meta did acknowledge that available news coverage indicates the video has been altered.

The Board selected this case to assess whether Meta's policies adequately cover altered videos that could mislead people into believing politicians have taken actions, outside of speech, that they have not. This case falls within the Board's elections and civic space and automated enforcement of policies and curation of content priorities.

The Board would appreciate public comments that address:

- Research into online trends of using altered or manipulated video content to influence the perception of political figures, especially in the United States.
- The suitability of Meta's misinformation policies, including on manipulated media, to respond to present and future challenges in this area, particularly in the context of elections.
- Meta's human rights responsibilities when it comes to video content that has been altered to create a misleading impression of a public figure, and how they should be understood with developments in generative artificial intelligence in mind.
- Challenges to and best practices in authenticating video content at scale, including by using automation.
- Research into the efficacy of alternative responses to political disinformation or misinformation beyond content removal, such as fact-checking programs or labelling (also known as "inform treatments"). Additionally, research on avoiding bias in such responses.

As part of its decisions, the Board can issue policy recommendations to Meta. While recommendations are not binding, Meta must respond to them within 60 days. As such, the Board welcomes public comments proposing recommendations that are relevant to this case.

Oversight Board

Public Comment Appendix for

Case 2023-029-FB-UA

Case number

The Oversight Board is committed to bringing diverse perspectives from third parties into the case review process. To that end, the Oversight Board has established a public comment process.

Public comments respond to case descriptions based on the information provided to the Board by users and Facebook as part of the appeals process. These case descriptions are posted before panels begin deliberation to provide time for public comment. As such, case descriptions reflect neither the Board's assessment of the case, nor the full array of policy issues that a panel might consider to be implicated by each case.

To protect the privacy and security of commenters, comments are only viewed by the Oversight Board and as detailed in the Operational Privacy Notice. All commenters included in this appendix gave consent to the Oversight Board to publish their comments. For commenters who did not consent to attribute their comments publicly, names have been redacted. To withdraw your comment, please email contact@osbadmin.com.

To reflect the wide range of views on cases, the Oversight Board has included all comments received except those clearly irrelevant, abusive or disrespectful of the human and fundamental rights of any person or group of persons and therefore violating the Terms for Public Comment. Inclusion of a comment in this appendix is not an endorsement by the Oversight Board of the views expressed in the comment. The Oversight Board is committed to transparency and this appendix is meant to accurately reflect the input we received.

Oversight Board

Public Comment Appendix for

Case 2023-029-FB-UA

Case number

49

Number of Comments

Regional Breakdown

1	2	7	1
Asia Pacific & Oceania	Central & South Asia	Europe	Latin America & Caribbean
3	35	0	
Middle East & North Africa	United States & Canada	Sub-Saharan Africa	

CASE 2023-029-FB-UA	PC-18000	United States & Canada
Case number	Public comment number	Region
Withheld	Withheld	English
Commenter's first name	Commenter's last name	Commenter's preferred language
Withheld		No
Organization		Response on behalf of organization

Full Comment

It is wrong to mislead people like that. However, the twitter files have shown social media companies have proven they are unable to resist the pressure of government censorship. Facebook, Twitter, etc are no different then the town crier of old atop a wooden box in the middle of the town square. They maybe be private organizations but their platforms are the most public spaces the world has known. People are allowed to lie in a free society. It is up to the PUBLIC PEOPLE to correct each other... not some organization which may be acting on their own agenda or under the influence of the government. Social media companies have too wide a reach and scope to allow the to decide for a mass population what is and is not acceptable.

Link to Attachment

No Attachment

CASE 2023-029-FB-UA	PC-18001	Latin America & Caribbean
Case number	Public comment number	Region
Withheld	Withheld	English
Commenter's first name	Commenter's last name	Commenter's preferred language
Withheld		No
Organization		Response on behalf of organization

Full Comment

While the content does not meet either of the two criteria, which are that someone in the video said words they did not say AND the video is a product of artificial intelligence or machine learning, the video is still not a parody or satire. It intentionally seeks to accuse a public figure with a history of accusations of inappropriate touching, exposing a flaw in the development of the policy. Additionally, it puts at risk the image of what could be considered a minor if it is indeed a minor.

Link to Attachment

No Attachment

CASE 2023-029-FB-UA	PC-18002	United States & Canada
Case number	Public comment number	Region
Marty	Corchero	English
Commenter's first name	Commenter's last name	Commenter's preferred language
DID NOT PROVIDE		No
Organization		Response on behalf of organization

Full Comment

He is just a grandfather nothing else, raising his grandkids doing the best he can. I have researched many of the so called "Photos" and the stories behind them and found no wrong doings concerning the children Getting a bit tired of all the fake pictures and fake news on FB. Some are funny but for the most part I find many inconsistencies. FB needs to start cracking down on these. If its altered then say it is when posted and if its altered to bring harm to a person then it should not be allowed.

Link to Attachment

No Attachment

CASE 2023-029-FB-UA	PC-18003	United States & Canada
Case number	Public comment number	Region
Withheld	Withheld	English
Commenter's first name	Commenter's last name	Commenter's preferred language
Withheld		No
Organization		Response on behalf of organization

Full Comment

The villinization of a public figure such as our President, without cause, is more a reflection on the intents of the publisher to sow mistrust, than on the President. If the sender of the message sees pedophilia in every action of a grandparent's touch of a grandchild, what is going on with the sender? Comments such as these must be rooted out and posted with large text as blatantly false and the sender banned from Meta activity for a specified period of time.

Link to Attachment

No Attachment

CASE 2023-029-FB-UA	PC-18004	United States & Canada
Case number	Public comment number	Region
Julie	Henry	English
Commenter's first name	Commenter's last name	Commenter's preferred language
DID NOT PROVIDE		No
Organization		Response on behalf of organization

Full Comment

There should be NO altered videos allowed AT ALL. I am sick of not having social media blithely accept this!

Link to Attachment

No Attachment

CASE 2023-029-FB-UA	PC-18005	United States & Canada
Case number	Public comment number	Region
Withheld	Withheld	English
Commenter's first name	Commenter's last name	Commenter's preferred language
Withheld		No
Organization		Response on behalf of organization

Full Comment

These fake videos cannot be allowed if we don't want our world destroyed. These are actively harming people and it needs to stop. This is a main reason I rarely visit Facebook anymore…there's so much false crap being pushed on folks for clicks it's destroying too much.

Link to Attachment

No Attachment

CASE 2023-029-FB-UA	PC-18006	United States & Canada
Case number	Public comment number	Region
Timothy	Chauvin	English
Commenter's first name	Commenter's last name	Commenter's preferred language
DID NOT PROVIDE		No
Organization		Response on behalf of organization

Full Comment

While an edited video may be cause for concern, the editing of this particular video is in line with a common meme format, the repetition is to make a point. The point being made coincides with hundreds of hours of publicly available footage showing Biden probing underage children international stage. Because this is the case, this video should be allowed to be maintained, perhaps with a small disclaimer stating that it is edited and that they've used expressed or that of the original poster (OP) and not that of the Biden administration.

Link to Attachment

No Attachment

CASE 2023-029-FB-UA	PC-18007	United States & Canada
Case number	Public comment number	Region
Withheld	Withheld	English
Commenter's first name	Commenter's last name	Commenter's preferred language
Withheld		No
Organization		Response on behalf of organization

Full Comment

Altered photos and videos are a prime source of misinformation. This is not free speech. META needs to be more vigilant. It should not allow or promote demeaning and disreputable photos and videos that are engineered to be offensive and that spread untruths and lies.

Link to Attachment

No Attachment

CASE 2023-029-FB-UA	PC-18008	United States & Canada
Case number	Public comment number	Region
Kristen	Walsh	English
Commenter's first name	Commenter's last name	Commenter's preferred language
DID NOT PROVIDE		No
Organization		Response on behalf of organization

Full Comment

I agree with Meta's decision not to remove the video, but it does seem prudent to have Meta insert a banner that the video content has been altered or manipulated before the user watches the video.

Link to Attachment

No Attachment

CASE 2023-029-FB-UA	PC-18009	United States & Canada
Case number	Public comment number	Region
Withheld	Withheld	English
Commenter's first name	Commenter's last name	Commenter's preferred language
Withheld		No
Organization		Response on behalf of organization

Full Comment

Any video that has been altered to manipulate the public should be removed. It doesn't matter whether it's generated by AI or the speech has been changed. The idea is to counter misinformation, not adhere to an arbitrary standard.

Link to Attachment

No Attachment

CASE 2023-029-FB-UA	PC-18011	Europe
Case number	Public comment number	Region
Withheld	Withheld	English
Commenter's first name	Commenter's last name	Commenter's preferred language
Withheld		No
Organization		Response on behalf of organization

Full Comment

The video is misleading and should be removed because it is clearly reposted in a different context to persuade people to perceive the performed action in a malicious way not present in the original context.

Link to Attachment

No Attachment

CASE 2023-029-FB-UA	PC-18012	Europe
Case number	Public comment number	Region
Ben	Luberti	English
Commenter's first name	Commenter's last name	Commenter's preferred language
DID NOT PROVIDE		No
Organization		Response on behalf of organization

Full Comment

This is an out of the context arranged video and has to be bannen.

Link to Attachment

No Attachment

CASE 2023-029-FB-UA	PC-18014	United States & Canada
Case number	Public comment number	Region
Withheld	Withheld	English
Commenter's first name	Commenter's last name	Commenter's preferred language
Withheld		No
Organization		Response on behalf of organization

Full Comment

Video that is altered to the point of showing an innocent person committing a crime, especially sexual assault, should never be allowed on any platform. I understand that altered video of public figures can be allowed on the basis of free speech, but videos falsely portraying criminal activity go too far. If you must permit such videos to be shown, all altered videos should be clearly identified as altered videos.

Link to Attachment

No Attachment

CASE 2023-029-FB-UA	PC-18015	United States & Canada
Case number	Public comment number	Region
Lynn	Patsiga	English
Commenter's first name	Commenter's last name	Commenter's preferred language
DID NOT PROVIDE		No
Organization		Response on behalf of organization

Full Comment

The altered video of President Biden should be removed because it’s dishonest and harmful and plays into the larger detrimental right wing extremist propaganda machine.

Link to Attachment

No Attachment

CASE 2023-029-FB-UA	PC-18016	United States & Canada
Case number	Public comment number	Region
Dan	Neunaber	English
Commenter's first name	Commenter's last name	Commenter's preferred language
DID NOT PROVIDE		No
Organization		Response on behalf of organization

Full Comment

Your misinformation policy appears arbitrary, and designed to help you avoid responsibility. The clip is: 1. Not true, 2. Hurtful and disrespectful , 3. Designed to mislead, 4. Will definitely will influence a significant number of people, and creates more confusion and hatred in a political atmosphere that does not need more of either. 6. In what world do we need more misinformation?

Link to Attachment

No Attachment

CASE 2023-029-FB-UA	PC-18017	United States & Canada
Case number	Public comment number	Region
Withheld	Withheld	English
Commenter's first name	Commenter's last name	Commenter's preferred language
Withheld		No
Organization		Response on behalf of organization

Full Comment

Are you people seriously asking us if you should take down an edited, fake clip? Is that really what you're trying to do here? Is this not the sort of deeply easy call that even laypeople like me can comfortably make?

Jesus tap-dancing Christ, I hope the government breaks Meta into little pieces.

Link to Attachment

No Attachment

CASE 2023-029-FB-UA	PC-18019	United States & Canada
Case number	Public comment number	Region
Gregory	Stanton	English
Commenter's first name	Commenter's last name	Commenter's preferred language
Genocide Watch		Yes
Organization		Response on behalf of organization

Full Comment

This altered image of President Biden is clearly a fraudulent attempt to portray him as immoral. That is called libel. It is such a clear violation of Facebook policy that Facebook must take it down and permanently bar those who posted it from using Facebook.

Link to Attachment

No Attachment

CASE 2023-029-FB-UA	PC-18021	United States & Canada
Case number	Public comment number	Region
Dr Brett	Prince	English
Commenter's first name	Commenter's last name	Commenter's preferred language
DID NOT PROVIDE		No
Organization		Response on behalf of organization

Full Comment

This is NOT Disinformation. What it is is another over-reaching example of FB's ongoing, documented one-sided, Orwellian-level censorship. It is obscene hypocrisy in light of the fact that FaceBook/Meta choose to promote and donate nearly half a billion dollars in the last election to defeat Republicans and President Trump, and then knowingly and deliberately censor vital information that interfered with a US election (The Hujnter Biden Files/Laptop). There is no intellectual or political diversity on the Board. My offer as a doctor and independent thinking humanitarian has not been accepted.

Link to Attachment

No Attachment

CASE 2023-029-FB-UA	PC-18022	Middle East & North Africa
Case number	Public comment number	Region
Victor	Okwara	English
Commenter's first name	Commenter's last name	Commenter's preferred language
DID NOT PROVIDE		No
Organization		Response on behalf of organization

Full Comment

The contents should be removed from Facebook

Link to Attachment

No Attachment

CASE 2023-029-FB-UA	PC-18023	United States & Canada
Case number	Public comment number	Region
Withheld	Withheld	English
Commenter's first name	Commenter's last name	Commenter's preferred language
Withheld		No
Organization		Response on behalf of organization

Full Comment

Meta needs to focus on the malicious intent aspect of its video distortion ban. If a person whose political position can be affected by influencing the public through voting, then videos intended to "catch" the person doing it saying something they didn't do or say, or to manipulate the words or deeds to appear prurient or otherwise negative, should not be circulating on the Meta platform, whether or not AI was utilized.

The public should feel confident that Meta is responsible for removing maliciously manipulated content.

Link to Attachment

No Attachment

CASE 2023-029-FB-UA	PC-18024	United States & Canada
Case number	Public comment number	Region
Clay	Calvert	English
Commenter's first name	Commenter's last name	Commenter's preferred language
DID NOT PROVIDE		No
Organization		Response on behalf of organization

Full Comment

Three minor changes to the “Manipulated Media” policy I recommend are: 1) Considering the question of manipulation through the eyes of the “average adult,” rather than the eyes of the “average person.” This is particularly relevant if one is primarily concerned with voters (i.e., adults) being influenced. I don’t think we want to ask how whether an average three-year-old would be fooled, so making it an “average adult” perspective helps here; 2) Changing the phrase “would likely mislead” to “would be substantially likely to mislead.” This makes the policy more free-speech friendly by narrowing its sweep; 3) Include some language limiting the reach of the policy to “material” falsities, not just any falsities. For instance, falsely manipulating a video to show that President Biden said a person was wearing a “green” shirt when he in fact said it was a “brown” probably is a meaningless falsity.

Link to Attachment

No Attachment

CASE 2023-029-FB-UA	PC-18025	United States & Canada
Case number	Public comment number	Region
Daniel	Vandenberg	English
Commenter's first name	Commenter's last name	Commenter's preferred language
DID NOT PROVIDE		No
Organization		Response on behalf of organization

Full Comment

Look, some of this stuff is photoshopped and you can't believe everything that you see. At the same time people should have a sense of humor and not be so thick skinned. Personally I think that Biden is not a pedophile likely but he is the worst president in history. The country has went straight to hell under his watch. I'm a Christian conservative Patriot in that order. I don't like marxism, socialism of any kind, and stupid economic, domestic and foreign policy. I do believe the election was stolen in 2020. And, I'm worried that it will happen again. If it does this country is doomed. Now please don't kick me off Facebook for my opinion because you guys do it all the time and it's for the dumbest stuff cuz your algorithms need tuning. I posted other people's comments and got kicked off. I had one where I said we should shoot down any balloons that come into our sovereign airspace and they give me a ban for that. What you're doing as you're pushing conservatives away from Facebook and that's sad because we use it for our family and recipes and I have a lot of liberal friends it wouldn't be no fun to be on Facebook of everybody agreed with you. Keep on working on it

guys. And I'll tell you what you ought to put a couple rednecks like me on the board. I can still Factor polynomials. And I have to say I was one of the best ER doctors that ever slung a stethoscope. Just read my post from my nurses. But I'm a conservative and I believe in strict conservation written as the founding fathers wrote it.

Link to Attachment

PC-18025

CASE 2023-029-FB-UA	PC-18026	United States & Canada
Case number	Public comment number	Region
Kalena	Horie	English
Commenter's first name	Commenter's last name	Commenter's preferred language
DID NOT PROVIDE		No
Organization		Response on behalf of organization

Full Comment

Anything warping the character of a president is hugely impactful. Altered video, altered representation of the video makes people's opinion of the president different than what the facts are.

Also, on the forms for reporting, there should be a comment box where an explanation of why the content is being reported. I've reported malware links and all I get back is "It doesn't go against our standards". Making it appear that Facebook doesn't care if its users are being scammed or their identity stolen.

I reported my previous account as hacked and I still can't get it back from someone in Asia. I sure didn't move to Asia.

Link to Attachment

No Attachment

CASE 2023-029-FB-UA	PC-18027	United States & Canada
Case number	Public comment number	Region
James	Baehm	English
Commenter's first name	Commenter's last name	Commenter's preferred language
DID NOT PROVIDE		No
Organization		Response on behalf of organization

Full Comment

ALL Misleading comment in any form should be removed from ALL media !!

Link to Attachment

No Attachment

CASE 2023-029-FB-UA	PC-18028	United States & Canada
Case number	Public comment number	Region
Robert	Kriegar	English
Commenter's first name	Commenter's last name	Commenter's preferred language
DID NOT PROVIDE		No
Organization		Response on behalf of organization

Full Comment

This is an illustration of a pattern that Facebook has I'm allowing false, demeaning, harassing, sexually oriented untrue statements when it is directed at Democrat politicians. Had the same video been made of Donald Trump or any other of the tin foil hat brigade, it would have been removed immediately. This is not only intolerable in either direction, but it is blatant hypocrisy. Moreover it uses controversy to generate traffic for Facebook or Meta as they call themselves now, and there is no way to look at it without coming to the conclusion that it is an abusive lack of application of policy. One could rightfully conclude that it is bias.

This conduct not only my lines and elected official, but also a private citizen. His daughter.

This altered video is not opinion, but it was entirely false. It is misinformation. Misinformation is a lie. No credible organization with any integrity supports or

promotes lies.

After having Facebook wrongfully accuse me of trading drugs or arms on Facebook and completely stripping me of my original profile, I'm well aware of the damage that Facebook causes with their political bias.

Not only does their application of existing policy not adequately addressed this situation, but Facebook intentionally uses it to abuse users on one side of the political spectrum, while they promote and support users on the other side of the political spectrum. This is an intolerable situation no matter which direction it is going in. And that is the true test.

The oversight board has been extremely LAX in allowing so-called liberals to be abused by Facebook's policies and standards in the United States.

Link to Attachment

No Attachment

CASE 2023-029-FB-UA	PC-18031	United States & Canada
Case number	Public comment number	Region
Dahvi	Cohen	English
Commenter's first name	Commenter's last name	Commenter's preferred language
Rep. Adam Schiff (CA-30)		Yes
Organization		Response on behalf of organization

Full Comment

Dear Members of the Oversight Board,

I am writing today regarding the threat posed by altered content, specifically “cheap fakes" on digital platforms, including Meta. I support the Oversight Board’s well-founded decision to review an altered and deeply misleading video posted by a Facebook user of President Biden voting early in-person during the 2022 midterm elections. Additionally, I urge the Board to assess the effectiveness of Meta’s policies to prevent further spread of election misinformation through “cheap fakes.”

As you know, flagging and removing altered content on digital platforms such as Facebook and Instagram is critical to protect users from election misinformation, interference, and intentional manipulation. This is particularly important as millions of Americans increasingly look to these platforms to obtain news and election information.

Quickly evolving artificial intelligence (AI) technology makes it difficult for voters to accurately identify fraudulent video and audio material and has rightfully been the topic of much discussion as the technology continues to advance. However, platforms must also remain wary of any doctored content posted on digital platforms, even if not AI was not used to manipulate the content. These "cheap fakes" are easy to make and can be equally dangerous in the spread of election misinformation.

Throughout my tenure in Congress, I have pushed digital platforms to address realistic falsified videos and images, including manually edited "cheap fakes." Although the video in this case was not altered using AI, I remain concerned that any altered content of elected officials or political candidates poses a threat to our democratic institutions and voters' ability to make accurately informed decisions. This is especially true of content altered with the intent of removing context or distorting reality.

In the case currently being reviewed by the Oversight Board, Meta's initial decision not to remove the altered video of President Biden encourages the spread of other "cheap fakes" and other election misinformation – as we saw with the decision to leave an altered "cheap fake" of Speaker Nancy Pelosi on its platform leading up to the 2020 election. Furthermore, the determination by Meta that the content did not violate the company's policies on manipulated media fails to consider the detrimental impact of the spread of altered content related to elections, regardless of the use of AI.

Meta's policy on misinformation, as laid out in their Facebook Community Standards, includes the removal of "content that is likely to directly contribute to interference with the functioning of political processes and certain highly deceptive manipulated media." As the 2024 Presidential election quickly approaches, particularly with one candidate willing to continuously break the bounds of ethical behavior on social media as we saw with January 6, I urge the Oversight Board to use its authority to recommend that Meta ensure their policies on manipulated media and misinformation include regulation of any altered content on its platform that seeks to spread election misinformation and

disinformation.

Thank you for your attention to this important matter.

Sincerely,

Adam B. Schiff

Member of Congress

Link to Attachment

PC-18031

CASE 2023-029-FB-UA	PC-18032	United States & Canada
Case number	Public comment number	Region
Laurel	Noack	English
Commenter's first name	Commenter's last name	Commenter's preferred language
DID NOT PROVIDE		No
Organization		Response on behalf of organization

Full Comment

Please delete it. This is insane!!

Link to Attachment

No Attachment

CASE 2023-029-FB-UA	PC-18033	United States & Canada
Case number	Public comment number	Region
Withheld	Withheld	English
Commenter's first name	Commenter's last name	Commenter's preferred language
Withheld		No
Organization		Response on behalf of organization

Full Comment

This response seeks to provide the Oversight Board with feedback on:

- The challenges with expanding Meta's content policies to cover editing of videos in simple ways that may be misleading in certain contexts.

- Problems with expanding fact-checking, especially into contentious social and political issues.

- Alternative paths forward for ascertaining truth and dealing with misinformation and manipulated media.

This case entails a piece of content in which a video of a President Biden is simply edited with a loop. In this sense, the video is not false or fake, but clipped at a specific time and then allowed to play back again. This is a common and basic editing technique that is part of most tools on phones and social media platforms (including Instagram's boomerang feature). This basic editing technique is not much different from clipping a small portion of a longer video, nor is it substantively different from taking a snapshot

of a video to portray one moment from a video. It is also similar to gifs that can be created and shared by anyone through readily available websites and technology. [From here on out I will refer to loops, clips, gifs, and screenshots as "simple editing."] Such simply edited clips are ubiquitous and used in countless political communications:

- Gifs and clips of President Trump making strange and exaggerated faces or hand motions.

- Jeb Bush's "please clap" clip.

- Clips of Mitt Romney talking about "binders of women".

- Other simply edited imagery that show President Biden awkwardly touching women.

This piece of content makes an additional claim regarding President Biden as a pedophile, as his hands come close to his granddaughter's breast. But, as noted above, a mere screenshot or clipping of this video could have accomplished a similar result by showing the President apparently touching near her breast. Thus, the content presents a broad challenge: should Meta remove — or otherwise suppress —any simply edited imagery that could be misleading? For multiple reasons, the answer should be no.

Starting with the scale, the sheer amount of content that could be considered misleading because of simple editing is near endless. Creating a broad policy to cover this sort of content would stifle countless expressions of political or non-political speech. The basic photo features of my iPhone and the boomerang feature built into Instagram itself could be violating if posted to Meta's platforms. Even if not proactively enforced (but only enforced upon report), that amount of content could still be massive. This is especially true for Instagram, where every post is a picture or video and many, if not most, have been simply edited.

A new policy in this space will also likely silence true criticisms and create opportunities to game the system. In this case, President Biden has faced accusations of being too touchy with women and children. Many of these criticisms are made via short video clips or screenshots of videos. Regardless of what one thinks about President Biden's mannerisms, there will be other individuals who may engage in malicious behavior on camera but, when simply edited, could run afoul of a policy in this area,

silencing victims and protecting aggressors. Satirical and humorous efforts that rely on simply edited imagery to criticize public figures could similarly be silenced as manipulated misinformation. The whole point of such content is often to use exaggeration or absurdity and so could be considered misleading or misinformation. Such a policy, especially if enforced upon report by users, would also create opportunities for bad actors to game the reporting system, such as in the Board's "Russian bot" case. Political campaigns could report any simply edited content that portrays their candidate in compromising or awkward position. This would empower rich and powerful users as well as trolls to silence various types of speech on the platform.

It is critical that criticism of public figures and leaders be allowed and not hampered in the name of countering misinformation (indeed this is why Meta has different standards under its Bullying and Harassment policy for attacks on public figures vs private individuals). The harm done to expression from such a policy far outweighs any vague and undefined harm of misinformation. Certainly, in this case, the harm is approaching zero, with no violence or offline actions taken and the mistake easily remedied with footage of the event by anyone who wants to look deeper.

This leads to the fundamental challenge of determining what is "true" and what is "false." We do not hesitate to venture that it is not possible for Meta's reviewers or automated system to enforce a policy in this area at scale. This is why, currently, Meta only removes specific, enumerated harmful claims of misinformation or the use of deepfakes where the editing is so severe and advanced that it is difficult for a user to know if the video is false. For all other content, Meta uses external fact-checkers in a more ad hoc manner rather than the determination of its own systems or reviewers. Even assuming no biases, asking reviewers or automated systems to accurately adjudicate an endless number of claims based on simply edited content is an endeavor bound to fail. How will reviewers or automated systems search out the truth and know when a simply edited piece of content is "too misleading"? And what about evolving situations where the full facts are just not known? And then we must add biases that are also extremely likely to emerge when dealing with political and social issues. As a result, beyond the negative impact of the sheer size of this censorial effort, it would also further undermine faith in the fairness in Meta's adjudications.

This brings us to the question of fact-checking itself and the efforts to label content accordingly. The same problems described above also exist for efforts to label or

inform. While labeling is a less severe action, and certainly preferably to removal of content, figuring out when content that has been simply edited is misinformation versus sufficiently truthful will be a herculean task with massive negative impacts to user expression. Rather than the challenge of internal accuracy and bias, using fact-checkers raises the same trouble, but with external groups. The public failures of misinformation and fact-checking described below have already significantly undermined faith in such efforts and expanding their remit will only worsen the trust deficit.

When discussing social media fact-checking, it is important to underscore from the start an ideological cleavage that has not been overcome, and only deepened over time. A major challenge is that certain political and philosophical viewpoints, especially those focused on freedom of expression, have little interest in serving as formal fact-checkers that suppress and remove speech. So, there is an inherent selection bias built into any fact-checking system that enforces truth — only those that want to suppress content and want to work with social media companies join such programs. Given this ideological lack of diversity, "expert" driven fact-checking is likely only to further undermine faith in Meta and media fact-checkers as overly censorial. Whether it be a political bias, bias toward certain government or social narratives, or simply a different view of what should be considered misinformation, many examples illustrate the failure of the existing system. Many claims about COVID — ranging from vaccines to lockdown and masking policies —made by governments and experts turned out to be wrong, or at least not as ironclad as they asserted. They could not be questioned, even by other experts legitimately inquiring about conclusions or public policies that did not appear to be well supported in the data. Moreover, social media companies relied on governments and fact-checkers to determine these narratives. Add to this many fact-checking incidents on social or political issues where the tone, perspective of poster, concerns about how others may use factual information, making a "prediction we can't fact-check," the fact-checker's or a social media company's assumptions, or just inaccuracy are used to suppress opinions or otherwise truthful arguments.

Worse still is the fact that appeals of fact-checks can only be made to the organization that made the fact-check and fact-checkers can act with near impunity. Meta rarely acts to remove a fact-checker and absolves itself of responsibility by saying that any fact-check must appealed to the original independent fact-checker —an appeals system that 85% of American users think needs to change. Meanwhile, fact-checkers can do whatever they want with an appeal and claim to be a neutral cog in the system Meta set

up. With a lack of accountability for those in the fact-checking program on top of a program inherently drawing organizations that are supportive of suppression, the ancient problem of quis custodiet ipsos custodes — "Who watches the watchers?" — is just as relevant today as it was 2,000 years ago.

Ultimately, the epistemological challenge of determining truth must involve a system that is open to challenge. As Jonathan Rauch has noted, the two fundamental rules of liberal inquiry are that 1) no one gets the final say and 2) no one has personal authority. Methods based on iterative consensus that allow divergent views to challenge one another are the best way to reach truth and meaningfully correct those in error. Appeals to authoritative fact-checkers or governments that cannot be meaningfully appealed fail both of those principles. As a private company, Meta is not under an obligation to subject itself to such a standard, but it has indicated that it values both free expression and consumer trust. The Oversight Board is effectively tasked with helping the company handle difficult content moderation questions where there may be conflicting views or values at hand and suggesting best practices, including such frameworks for handling difficult or disputed decisions around the content available.

Thus, a better approach in line with such a framework could be something along the lines of X or Twitter's community notes, which require some level of consensus by diverse types of users. By merely adding context instead of more aggressive suppression, "good speech" is provided to users to counter bad information. This kind of notification would likely have been effective with this Biden content, as the full video of the event would have given people the information they needed to decide for themselves what President Biden was doing. Building consensus among different users to find key facts is also more likely to positively move people towards being better consumers of information that hold institutions to account, rather than current appeals to authority that are undermining trust in various institutions.

Another example of how community consensus over the factual nature of information may be provided is the Wikipedia editorial process. While open to all and certainly subject to vandalism and misleading claims at times, the wisdom of the crowd has yielded general factual consensus — even on disputed facts. It also creates a labeling system that indicates when claims are not backed up by citations or sufficient support, but does not itself decide on the replacement or appropriate citation. Such continual refinement may provide an alternative approach on how consensus may be reached while still highlighting uncertainty.

One simple proposed solution may initially appear to be to tag or flag all "manipulated" media or media that contains AI. However, this may actually result in a meaningless label if applied broadly or risk the appearance of selective application if applied narrowly. A broad interpretation of what qualifies could result in labeling any photo with a filter being labeled as "manipulated" media. Similarly, even in the political advertising realm, an overly broad interpretation could result in labeling ads as manipulated if they have accessibility features like autogenerated captioning, AI voices translating an ads voiceover to Spanish, or the use of AI to remove background noise. The result could be that the labeling of that media as potentially manipulated results in fatigue for the consumer such that it no longer triggers an appropriate scrutiny of what they are consuming. If a manipulated, distorted, or AI-generated media label were to be developed, the rules of what will and will not receive this label should be clear to consumers, content creators, and advertisers, and the application should be such that it does not target only particular viewpoints.

In sum, efforts to add a new policy to counter simply edited videos that may be considered by some to be misinformation could significantly harm both political and non-political expression, be abusable by those with more resources and internet trolls, present a problem that will be impossible to handle at scale with any semblance of fairness, and further undermine faith in Meta's fairness or the fairness of its fact-checking enterprise. A better approach would be to utilize methods of providing greater context that rely on iterative consensus and are open to being meaningfully challenged by different facts and viewpoints. For AI-generated content, labels could be applied to notify users of such content, but such labels also risk fatiguing users if broadly applied or being viewed as selective bias if narrowly applied.

David Inserra, Fellow for Free Expression and Technology, Cato Institute

Jennifer Huddleston, Technology Policy Research Fellow, Cato Institute

The experts make these comments in their individual capacities.

Link to Attachment

PC-18033

CASE 2023-029-FB-UA	PC-18035	Central & South Asia
Case number	Public comment number	Region
hira	saleem	English
Commenter's first name	Commenter's last name	Commenter's preferred language
Digital Rights Foundation		Yes
Organization		Response on behalf of organization

Full Comment

DIGITAL RIGHTS FOUNDATION PUBLIC COMMENT ON META'S OVERSIGHT BOARD CASE(S) ON ALTERED VIDEO OF JOE BIDEN

Submission Author: Hira Saleem
Edited by: Maryam Saeed and Seerat Khan
Submission Date: October 17th, 2023

*To read the Oversight Board's full decision on this case

**To see all submitted Public Comments

The rapid advancement of digital technology has transformed the way information is created, consumed, and exchanged. In the last two decades, the internet has made knowledge and information more accessible to a wider section of society by removing barriers to access across the globe. However, it is

concerning to see how technology is being used to manipulate public opinion by spreading disinformation, misinformation, and malinformation. There is a worrying trend of manipulating facts and videos, which has disrupted public communication and democratic processes in societies.

Recently, President Joe Biden's video of 29th October 2022 has surfaced on Meta's platforms, Instagram and Facebook, which included an altered clip of the President placing an "I Voted" sticker on his granddaughter's chest and kissing her cheek during last year's midterm elections. This video footage was altered and showed President Biden's motion of touching the girl's chest and a short excerpt of the song by Pharoahe Monch's "Simon Says" which has the lyrics "Girls, rub on your titties" was in that altered video. It was posted on Facebook on May 2023 with a caption calling Biden "a sick pedophile." for touching his granddaughter's breast. Comments under the post questioned the people who voted for him, saying they are "mentally unwell" .

The influence of disinformation on the outcome of elections has been widely discussed since the Brexit referendum and the 2016 U.S. presidential election. It's now clear that disinformation has become a significant concern in numerous countries around the world. The exploitation of personal user data to manipulate politics gained public attention after the Cambridge Analytica scandal.

Thus, this issue is not limited to any specific region or political system, and it poses a significant threat to the integrity and fairness of democratic processes. The growing impact of disinformation is among the many reasons why democracies worldwide are under pressure.

The rise of AI-powered technologies has brought about an alarming increase in the use of these tools for disinformation campaigns. Among the most concerning forms of disinformation are multimedia content such as deepfake videos or images, voice cloning, and generative text, all of which rely on AI algorithms to create highly convincing fabrications [for more, see Note 1 in Additional Notes].

Multimedia content can be easily manipulated, which can pose significant challenges to video authentication and verification. To address this issue, various techniques have been developed by researchers to detect any discrepancies in a video that could indicate tampering or alteration [For more, see Note 2].

The Misinformation Policy of Meta clearly states that any audio, photos, or videos, regardless of whether they are deepfakes or not, that violate any of Facebook's Community Standards will be removed from the platform. These standards include those that regulate nudity, graphic violence, voter suppression, and hate speech. However, in the recent case, it appears that Meta has not adhered to these policies.

It is worth noting that the video in question was only altered and did not use AI or machine learning to manipulate the footage as confirmed by an assessment by AP News and acknowleged by Meta . Therefore, it did not meet the criteria for removal. This highlights a gap and loophole in Meta's manipulated media policies . For instance, in Slovakia , a fake audio recording was recently circulated on Facebook in which one of the country's prominent politicians seemed to discuss rigging the elections. The creators were able to exploit a loophole in Meta's manipulated media policies, which do not cover faked audio. This raises serious questions about the policy on how to deal with such matters, given the potential for their harmful impact.

Freedom of expression is a fundamental value in democracies and is protected by Article 19(2) of the International Covenant on Civil and Political Rights (ICCPR). However, the advent of digitalization and social media has opened up new ways for human rights violations, which the UN Human Rights Council has confirmed must be applied online and offline. In this digital era, people are more exposed to hate speech and false information, giving state and non-state actors more power to undermine freedom of expression. Disinformation poses a serious threat to various human rights and democratic principles .

Firstly, it can harm the privacy and reputation of the person it targets. False

information about an individual can lead to negative consequences such as job loss or social ostracism. Disinformation can also infringe on the privacy of individuals in its intended audience. For instance, targeted ads or messages based on personal data collected without consent can violate an individual's privacy.

In the digital age, the right to privacy faces new and complex challenges. Social media platforms and other digital spaces have become breeding grounds for personal attacks that can have real-world repercussions for individuals. Furthermore, the collection of personal data for micro-targeting messaging has made it easier for disinformation to spread and influence people's opinions and behaviors.

Social media companies have been under scrutiny in the face of growing concerns over foreign interference in elections. As a result, they have taken steps to self-regulate and implement stricter policies. This increased awareness has led to take measures to prevent future electoral interference. However, ensuring that their platforms are safe and secure for users poses significant challenges for technology companies.

One of the most significant challenges is content moderation. Meta has a duty to ensure that the platforms is free from harmful, illegal, or offensive content. However, it can be challenging to determine where to draw the line between free speech and harmful content.

Identifying perpetrators of cyberattacks and data breaches is another challenge. In recent years, many high-profile cyberattacks have been attributed to foreign actors seeking to disrupt businesses, governments, and even entire countries. It is crucial to identify and hold these individuals or groups accountable to maintain trust in the technology industry. Currently, Meta's transparency reports inform us of the number of fake accounts removed every year, but we still do not know what these accounts were trying to achieve or the effects they had before identification.

To address these challenges, there needs to be a common safety standard and standardized reporting that allows watchdog agencies to compare reports from different social media companies. Tech companies should share technical data and know-how to develop best practices and optimize capabilities for tracking and removing antagonists across platforms. Additionally, independent auditing should be considered to build and maintain trust in social media companies' reports.

Finally, technology companies such as Meta, must consider the impact of their actions on society. Tweaking algorithms for the public good can have far-reaching consequences, and it is essential to consider the impact on individuals and communities before making changes. The responsibility of technology companies is enormous, and they must balance innovation with social responsibility delicately.

DRF strongly believes by recognizing these threats and taking appropriate measures to address them, we can better safeguard our human rights and democratic values in the digital realm.

Additional Notes :
Note 1:
Deepfake videos are a result of machine learning algorithms that alter the facial expressions and movements of an individual in a video. This manipulation is done to make it appear as if the person is doing things or saying things that they never actually did. Voice cloning, on the other hand, is used to create convincing audio recordings of individuals that sound like they are saying things they never actually said.

Deepfake images can be created by AI algorithms that mimic an individual's facial features and expressions, and then superimpose them onto another person's body, creating a highly convincing image that looks like the original individual is in a different location or situation.

Generative text is another form of disinformation that has become increasingly popular. It involves the use of AI algorithms to create entirely fabricated content that appears to be written by a human. This type of content can be used to spread false information, manipulate public opinion, and create division, posing a significant threat to society. All of these techniques can be used to spread false information, manipulate public opinion, and sow discord. This type of content is particularly dangerous during elections, which are a highly vulnerable period.

Note 2:
One such technique is intelligent authentication, which is particularly effective in cases where videos lack the attributes that can help confirm their authenticity. Intelligent authentication involves analyzing each frame of a video and using statistical features such as edges or corner points to establish a frame-by-frame relation. By comparing the processed video with the original, it is possible to identify any differences that could indicate tampering. This technique can also help distinguish between authentic and manipulated footage, making it an essential tool in forensic investigations, legal proceedings, and other applications that require video authentication. However, the drawback with the intelligent techniques is that, for even a single kind of attack, they need a sufficient large amount of databases of tampered and authentic video sequence to learn.

Oversight Board announces two cases: Altered video of President Biden and weapons post linked to Sudan's conflict. October 2023. Available at: https://www.oversightboard.com/news/698422811785085-oversight-board-announces-two-cases-altered-video-of-president-biden-and-weapons-post-linked-to-sudan-s-conflict/

Polarisation and the use of technology in ... - european parliament. Available at: https://www.europarl.europa.eu/RegData/etudes/STUD/2019/634414/EPRS_STU(2019)634414_EN.pdf (Accessed: 17 October 2023).

https://www.investopedia.com/terms/c/cambridge-analytica.asp

Artificial Intelligence, deepfakes, and disinformation - rand corporation. Available at: https://www.rand.org/pubs/perspectives/PEA1043-1.html (Accessed: 17 October 2023).

(PDF) video authentication-an overview - researchgate. Available at: https://www.researchgate.net/publication/265439700_Video_Authentication-An_Overview (Accessed: 17 October 2023).

https://transparency.fb.com/en-gb/policies/community-standards/misinformation/

Video distorts Biden placing voting sticker on granddaughter. Available at:

https://apnews.com/article/fact-check-biden-granddaughter-sticker-voting-495345266413?fbclid=IwAR3Zg6AEyWGxLx54D6y--4yo1wtnkqvJE6bFxWIRKfitxEL0OOY63RfNkgg

Oversight Board announces two cases: Altered video of President Biden and weapons post linked to Sudan's conflict. October 2023. Available at: https://www.oversightboard.com/news/698422811785085-oversight-board-announces-two-cases-altered-video-of-president-biden-and-weapons-post-linked-to-sudan-s-conflict/

Manipulated Media | Transparency Center - FacebookTransparency Center. Available at: https://transparency.fb.com › community-standards

Meaker, M. (2023) Slovakia’s election deepfakes show AI is a danger to democracy, Wired. Available at: https://www.wired.com/story/slovakias-election-deepfakes-show-ai-is-a-danger-to-democracy/ (Accessed: 17 October 2023).

Ibid

https://www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-civil-and-political-rights#:~:text=opinions%20without%20interference.-,2.,other%20media%20of%20his%20choice.

Ibid

(PDF) video authentication-an overview - researchgate. Available at: https://www.researchgate.net/publication/265439700_Video_Authentication-An_Overview (Accessed: 17 October 2023).

Ibid

Ibid

Link to Attachment

PC-18035

CASE 2023-029-FB-UA	PC-18036	Europe
Case number	Public comment number	Region
Sarah	Fisher	English
Commenter's first name	Commenter's last name	Commenter's preferred language
UCL Digital Speech Lab		Yes
Organization		Response on behalf of organization

Full Comment

DID NOT PROVIDE

Link to Attachment

PC-18036

CASE 2023-029-FB-UA	PC-18037	United States & Canada
Case number	Public comment number	Region
Laurie	Woodward Garcia	English
Commenter's first name	Commenter's last name	Commenter's preferred language
People Power United		Yes
Organization		Response on behalf of organization

Full Comment

October 23, 2023

Oversight Board
META

Dear Board Member:

We the American public are deeply troubled by recent cases of misleading political advertisements and posts. Misleading voters is anti-American. Even more troubling, however, is knowing that nothing is being done to mitigate these attempts to spread mis and disinformation.

As a group that stands up for the right to vote each and every day, People Power United fully supports META clarifying their policies that apply to campaign ads or posts formulated with the express intention of misleading voters about other

candidates or political parties.

Clarifying META policies would have no impact on freedom of speech. Campaign ads or posts would still be able to use parody or make statements critical of other candidates or political parties by way of text or images.

We, therefore, are requesting restrictions on the use of MANIPULATED audio and MANIPULATED video, created without the expressed written consent of the individual or entity being represented for a political purpose.

Here is an example of what we are recommending:

An audio or video of President Biden can be shared by anyone. Comments, graphics, sound, and video can be used with the video, however the actual audio or video of President Biden can not be manipulated. A manipulated audio or video of President Biden can only be shared if President Biden provides written consent.

This is a matter of utmost importance to the American public, and we request your prompt attention on the matter, to demonstrate to the nation that the META platform fully supports the security and credibility of the American election process.

Sincerely,

Laurie Woodward Garcia

People Power United

Link to Attachment

PC-18037

CASE 2023-029-FB-UA	PC-18038	United States & Canada
Case number	Public comment number	Region
Tommy	Tucker	English
Commenter's first name	Commenter's last name	Commenter's preferred language
DID NOT PROVIDE		No
Organization		Response on behalf of organization

Full Comment

The soulution is simple. Require that the post include a comment that is and altered post or video or that the post is satire and not accurate. If the post isn't labeled accurately then remove it, problem solved, without censoring free speech!

Link to Attachment

No Attachment

CASE 2023-029-FB-UA	PC-18040	United States & Canada
Case number	Public comment number	Region
nan	nan	English
Commenter's first name	Commenter's last name	Commenter's preferred language
Institute for Strategic Dialogue (ISD)		Yes
Organization		Response on behalf of organization

Full Comment

Please see attached for ISD's full comment with citations. Below is the text:

Thank you to the Oversight Board for the opportunity to comment on case 2023-029-FB-UA, regarding an altered video of President Biden. With the rapid development of deepfake technology and artificial intelligence (AI), it is crucial for social media platforms to have policies and safeguards in place to protect not only public figures but private individuals from malicious uses of new technology – especially for the many critical upcoming elections in 2024.

The Institute for Strategic Dialogue (ISD) is an independent, non-profit organization dedicated to safeguarding human rights and reversing the rising tide of polarization, extremism, and disinformation worldwide. Our work includes in-depth research and analysis identifying and tracking online

manipulation, mis- and disinformation, hate, and extremism in real time. We also formulate, advocate and deliver evidence-based policy approaches and programming.

Artificial intelligence technology can produce synthetic images and audio of candidates, generating wholly artificial scenarios to further a political agenda. Deepfakes are now so realistic that it is becoming nearly impossible for the average voter to discern fact from fiction. More concerningly, they are already being deployed by candidates and political parties in the US and beyond, demonstrating the need for regulation and clear social media policies. As shown by recent events across the world -- from the January 6 insurrection in Washington, DC to the Brazilian Congress attack in Brasília -- elections are already vulnerable to online disinformation and could become even more so with disinformation driven by AI. With some already referring to the upcoming 2024 US presidential election as "the most online election ever," social media platforms are in a unique position to be at the forefront of preventing harmful online narratives that will likely rapidly emerge and proliferate.

Our submission seeks to address the Oversight Board's request for comments on research into online trends of using altered or manipulated video content and Meta's policies and enforcement practices regarding manipulated media:

1. Research into online trends of using altered or manipulated video content to influence the perception of political figures, especially in the United States.

Previous research published by ISD has found that social media product features are already amplifying election disinformation, harming candidates, and assisting the organizing efforts of those disseminating false and harmful claims. It is now easier than ever for users and networks (such as foreign influence operations) to generate altered or manipulated content containing disinformation or misleading narratives and use social media product features to spread it more widely and efficiently. Past research has shown that women are targeted most often by deepfake videos, particularly with non-consensual intimate images – a trend that will undoubtedly affect women candidates and

politicians, who are already targeted by online gendered and sexualized disinformation campaigns.

With the rise of AI-generated deepfakes that anybody can now create, the Federal Election Commission (FEC) is already considering regulating AI-generated deepfakes in political ads ahead of the 2024 election. Google is also implementing a new policy that political ads using artificial intelligence must be accompanied by a prominent disclosure. Civil society organizations have published reports with policy suggestions on protecting democracy and warnings on how AI could affect voters in 2024. The time is ripe for Meta to participate in these discussions and policy work to ensure its platforms are ready for such a pivotal year.

Recommendations:

- Meta's policy teams need to be responsive to these kinds of online trends, especially when it comes to emerging technology, and adapt policies and enforcement quickly and accordingly.

- Meta should be aware of how generative AI affects different users of different backgrounds. For example, women are likely to be more targeted with deepfake non-consensual intimate images than men.

- Many generative AI technologies are also capable of producing content in languages other than English. Meta already allocates fewer resources towards moderating non-English language content. Meta should ensure that adequate resources are being put towards understanding and responding to generative AI content in non-English languages.

2. The suitability of Meta's misinformation policies, including on manipulated media, to respond to present and future challenges in this area, particularly in the context of elections.

While Meta made several strides in combatting election disinformation in past

years, the recent backsliding of misinformation and other policies does not bode well for the upcoming 2024 election. For example, in late August, Meta reportedly started allowing users to opt out of its fact-checking program, which gave users access to reliable information – particularly related to elections. ISD found that the criteria laid out in Meta's Manipulated Media Policy do not do enough to address emerging online trends with not only AI-generated deepfakes, but also regular deceptively edited videos, as proven in case 2023-29-FB-UA. Even though the manipulation was apparent and recognizable in this case, it still pushed a narrative containing misinformation about President Biden.

The current Manipulated Media Policy, which is also repeated in the Misinformation Policy, bans videos that have been edited or synthesized in ways that "are not apparent to an average person" and would mislead them to believe that: 1) a subject of the video said words they did not say; and 2) the content is a product of artificial intelligence or machine learning. The addition of "and" automatically does not include content that was deceptively edited by a user, which is likely part of the reason why the video in this case was not removed under this policy.

Although Meta did not consider this post for removal under their Hate Speech Policy or Bullying and Harassment Policy, it is important to note that the tactic of calling a public figure (or private individual) a pedophile, whether on the basis of a protected characteristic or not, has been increasingly weaponized by violent and conspiratorial movements and is the sort of language that is more likely to lead to actual violence. ISD's recent report on online and offline anti-drag mobilization highlights how this narrative is used to justify hate and violence against drag performers and members of the LGBTQ+ community. The accusation made in the video against Biden is clearly not a credible allegation and does not provide any sort of evidence or personal testimony, and the furthering of this narrative could lead to harmful online and offline threats against Biden.

Recommendations:

- Meta must recognize that content does not have to be the product of artificial intelligence or machine learning to be manipulated media. It also does not have to be unapparent or particularly well-edited.

- If Meta acknowledges and recognizes that a video is altered or purposefully misleading, it should remove the video from its platforms, no matter how many views it receives. As the Manipulated Media Policy states, this should not extend to content that is parody or satire, or is edited to omit words that were said or change the order of words that were said.

- Meta should include a clause in its Manipulated Media Policy about manipulated actions, not just manipulated speech. In this case, there was no speech involved, just manipulated actions.

- Meta should better protect public figures in its policies and ban disproven false narratives about public figures (in this case, President Biden being a pedophile) from being posted on its platforms.

3. Meta's human rights responsibilities when it comes to video content that has been altered to create a misleading impression of a public figure, and how they should be understood with developments in generative artificial intelligence in mind.

In the US, every individual has a right to commercialize aspects of themselves – including personalities. These rights are protected by the right of publicity (ROP), but in the age of generative AI, the "theft of personality" is a real threat. When looking at the issue from a commercial lens, the misuse of generative AI could damage a public figure's earning potential and brand. But non-consensual deepfakes or content that has been altered to create a misleading impression of a public figure could have reputational damages, too. These damages might not be enough to be clearly labelled as defamation or privacy violations and could therefore fall under the right of publicity, which protects a person's "likeness."

Meta has a responsibility to all the public figures on its platforms to protect their personality rights, or ROP.

Link to Attachment

PC-18040

CASE 2023-029-FB-UA	PC-18041	United States & Canada
Case number	Public comment number	Region
Tyler	Takemoto	English
Commenter's first name	Commenter's last name	Commenter's preferred language
American Civil Liberties Union		Yes
Organization		Response on behalf of organization

Full Comment

Dear Oversight Board,

The American Civil Liberties Union ("ACLU") submits these comments in response to the Oversight Board's deliberation regarding the above-referenced matter, which the Board has selected "to assess whether Meta's policies adequately cover altered videos that could mislead people into believing politicians have taken actions, outside of speech, that they have not." The Oversight Board requests public comments that address, among other issues:

- "The suitability of Meta's misinformation policies, including on manipulated media, to respond to present and future challenges in this area, particularly in the context of elections."

- "Meta's human rights responsibilities when it comes to video content that has been altered to create a misleading impression of a public figure, and how

they should be understood with developments in generative artificial intelligence in mind."

- "Research into the efficacy of alternative responses to political disinformation or misinformation beyond content removal, such as fact-checking programs or labelling . . ."

Meta's manipulated media policy applies to videos "edited or synthesized, beyond adjustments for clarity or quality, in ways that are not apparent to an average person, and would likely mislead an average person to believe [that] [a] subject of the video said words that they did not say." The policy applies only to "video [that] is the product of artificial intelligence or machine learning," and specifically exempts "content that is parody or satire" from removal.

The ACLU writes to emphasize that manipulated media—including AI-generated media—is not categorically harmful. To the contrary, there are uses of manipulated media that add value to public discourse—including parody and satire, which the current policy rightly excludes, as well as humorless, avowedly false speech that is nevertheless illustrative or thought-provoking, which the policy should also seek to protect.

Manipulated media can serve as a novel medium to entertain, process grief, teach, or engage with history and current events. For example, in March 2023 when AI-generated images falsely depicting the arrest of former President Donald Trump proliferated online, observers noted that the images, irrespective of their misleading nature, carried illustrative or—as one commentator put it—"cinematic" and "cathartic" value. In 2019, filmmakers affiliated with the Massachusetts Institute of Technology released a deepfake film of President Nixon delivering the speech written in case the Apollo 11 mission had ended in failure. The film invites the viewer to consider an alternate history, providing a more visceral and memorable interpretation of an existing historical document. The Dalí Museum in St. Petersburg, Florida features an interactive deepfake of the famous surrealist to provide visitors with a more immersive experience, allowing them to engage more directly with the artist and his work. Creators

have used deepfake videos of former President Barack Obama and Meta Co-Founder Mark Zuckerberg to comment on deepfake technology itself—using a self-referencing medium to spark conversation and raise awareness about the challenges and risks raised by the evolving technology. Creative expression is particularly important when it concerns public figures, including political candidates. An essential premise of democracy is that public officials may find themselves subject to "vehement, caustic, and sometimes unpleasantly sharp attacks." N.Y. Times v. Sullivan, 376 U.S. 254, 271 (1964). And as U.S. courts have reasoned in applying the First Amendment to protect false and even harmful expression: speech involving public figures should be given more lenience because public figures may access greater platforms and resources to rebut falsehoods compared to members of the general public. See Gertz v. Robert Welch, Inc., 418 U.S. 323, 323–24 (1974).

Even without some clear high-minded purpose, manipulated content—including content potentially subject to removal by Meta's existing policy—may provide the simple benefit of an evolving medium for users to express themselves and interact with one another. What if characters from different film franchises could exist in the same cinematic universe? What if current and former U.S. Presidents played live videogames together? What if every role in every film was recast with Nicolas Cage? Generative tools have enabled internet users to connect through these and other speculative questions.

It is inarguable that there are harmful uses of manipulated media. One can easily imagine a scenario where a bad faith actor seeking to stop people from exercising their right to vote posts a doctored video featuring a public official who falsely announces an emergency curfew on election day, shares misleading instructions about voting logistics or eligibility, or provides otherwise false information with the intent to deceive voters. These are straightforward cases where moderation of the doctored video would rightly prevent harm to the public.

At the same time, as the examples listed above illustrate, not all uses of manipulated media are in fact, or even potentially, harmful. For that reason,

Meta's policy should target the harms it is concerned could arise from manipulated media, rather than the technology used to create the posts or their mere falsity. In addition, because of the risks to speech, Meta should employ measures short of removal—such as embedded fact checks—where possible. And, where removal is used, Meta should leave ample space for users to appeal potentially wrongful removals.

Relying on fact checks and labels, rather than removals, and enabling appeals are all the more important here because the questions of what false speech might mislead the average person, and what counts as parody or satire are inherently subjective, and often impossible to determine neutrally—particularly at scale. A recent legal dispute involving content on Meta's platform illustrates these difficulties. In 2019, a U.S. Federal Court of Appeals addressed the question of whether a Facebook user's page impersonating an Ohio police department—and posting a series of false content that some users found misleading or in poor taste—might qualify as parody entitled to protection under the First Amendment. Novak v. City of Parma, 932 F.3d 421, 427–28 (6th Cir. 2019). In reaffirming the longstanding First Amendment principle that parody is protected speech, the Court rejected arguments that the fake page, which contained no disclaimer, was unentitled to protection because it actually misled users or because it was not in fact funny. Id. at 424. As the Court explained: "[the user's] page delighted, disgusted and confused. Not everyone understood it. But when it comes to parody, the law requires a reasonable reader standard, not a 'most gullible person on Facebook' standard. The First Amendment does not depend on whether everyone is in on the joke. Neither is it bothered by public disapproval, whether tepid or red-hot." Id. Whether or not something counts as parody is a difficult determination that U.S. Courts have repeatedly said is best left to a jury. See id. at 428; Hustler Magazine, Inc. v. Falwell, 485 U.S. 46, 57 (1988). At scale, it is almost certainly unworkable. For that reason, robust notice and appeals processes for users who believe their content was erroneously removed are essential.

Malicious and harmful use of manipulated media to deceive the public may be inevitable on Meta platforms, but wholesale removal of manipulated and AI-

generated content deprives users of the ability to joke, imagine, serve as fact-checkers, refute falsehoods, and develop the literacy needed to identify increasingly sophisticated misinformation on social media. Targeting Meta's policy to specific harms—rather than the AI used to create a post, or the falsity of speech, relying on measures short of removal to avoid confusion or deception, and ensuring robust appeals processes will ensure users are afforded the crucial space to exercise the creativity and self-expression enabled by new technologies.

Sincerely,

Tyler Takemoto

Vera Eidelman

Speech, Privacy & Technology Project

American Civil Liberties Union Foundation

125 Broad Street, 18th Fl.

New York, NY 10004

Link to Attachment

PC-18041

CASE 2023-029-FB-UA	PC-18042	United States & Canada
Case number	Public comment number	Region
Saurav	Ghosh	English
Commenter's first name	Commenter's last name	Commenter's preferred language
Campaign Legal Center		Yes
Organization		Response on behalf of organization

Full Comment

DID NOT PROVIDE

Link to Attachment

PC-18042

CASE 2023-029-FB-UA	PC-18043	Europe
Case number	Public comment number	Region
Chris	Idema	English
Commenter's first name	Commenter's last name	Commenter's preferred language
DID NOT PROVIDE		No
Organization		Response on behalf of organization

Full Comment

The video is not deceptively altered and therefore shouldn't be labeled altered. It also isn't Bullying or Harassment.

With deceptively altered I mean removing or adding audio or video in a way that seems like it is genuine with the intent of libel.

In this case it is clear neither the audio track nor the caption is from the original footage.

Therefore the audio track and the caption have to be considered commentary on the events in the original footage.

The commentary isn't Bullying or Harassment since no specific user is targeted by the "mentally unwell" charge.

And an opinion about a political candidate's touching is protected speech even if the opinion takes the form of an insult.

Also the opinion about Biden's touching isn't just based on this footage, but in the context of many publicly available videos of him touching people in a controversial way.

Banning such videos will set a bad president for censoring political viewpoints and will inhibit public commentary of public events.

Link to Attachment

No Attachment

CASE 2023-029-FB-UA	PC-18044	United States & Canada
Case number	Public comment number	Region
Chand	Rajendra-Nicolucci	English
Commenter's first name	Commenter's last name	Commenter's preferred language
Initiative for Digital Public Infrastructure		Yes
Organization		Response on behalf of organization

Full Comment

October 24, 2023

Chand Rajendra-Nicolucci, Initiative for Digital Public Infrastructure

In response to your call for public comment in case 2023-029-FB-UA, I will offer some thoughts on alternative responses to political disinformation and misinformation beyond content removal. In short, community annotation is a promising avenue for responding to political disinformation and misinformation beyond content removal, with advantages over traditional fact-checking and labeling programs.

Community annotation is a form of community governance where users of a

social media platform contribute annotations to posts which—after a selection process—are displayed alongside those posts on the platform. X's Community Notes initiative is an early and prominent example of community annotation that illustrates its promise. X reports that people who see a note are 20-40% less likely to agree with the substance of a potentially misleading post, compared to someone who sees the post alone. X also reports that people who see a note are 15-35% less likely to Like or Retweet a post, compared to someone who sees the post alone. And X reported that notes were informative regardless of a person's political party—there was no statistically significant difference across party ID.

Community annotation offers two main advantages over traditional fact-checking and labeling programs.

First, community annotation is more scalable. Traditional fact-checking and labeling is limited by the small set of professional fact-checkers available to review posts. In contrast, community annotation is able to take advantage of a much larger pool of users. And scaling doesn't necessarily mean sacrificing accuracy. Research has shown that crowdsourced annotations can attain similar outcomes as professional fact-checkers. Moreover, in practice, Wikipedia serves as a testament to the ability of crowds to achieve accuracy at scale.

Second, community annotation may be less polarizing than traditional fact-checking and labeling. Today's environment of broad mistrust in institutions likely favors corrections from peers rather than experts. This is further exacerbated in the U.S. by Republican skepticism of mainstream media. Studies in the U.S. have found that self-identified Republicans are less likely to trust and support traditional fact-checking and labeling than Democrats and independents. For example, one study found that 70% of Republicans think fact-checkers will favor a particular side. Another found a widening gap between Democrats and Republicans regarding whether platforms should take steps to restrict false information. In contrast, X found no statistically significant difference in the effect of notes for Democrats and Republicans and found that for both parties more than 70% of people found notes helpful or neither helpful nor unhelpful. Along the same lines, research has suggested that community

governance mechanisms such as community annotation may be perceived as more legitimate than governance undertaken by platforms themselves.

Community annotation is not a cure-all and there are open challenges. For example, though community annotation is more scalable than traditional fact-checking and labeling, it still is only able to address a subset of the questionable content on a platform. Exploring avenues for expanding that subset is important (e.g, applying notes to similar or duplicate content). Additionally, particularly in rapidly developing situations, community annotation may not be quick enough to reach the majority of consumers of questionable content. Further, community annotation is likely to be a target for manipulation by actors with interests in shaping the digital public sphere.

Link to Attachment

PC-18044

CASE 2023-029-FB-UA	PC-18045	United States & Canada
Case number	Public comment number	Region
Sabhanaz	Rashid	English
Commenter's first name	Commenter's last name	Commenter's preferred language
Tech Global Institute		Yes
Organization		Response on behalf of organization

Full Comment

PUBLIC COMMENTS ON CASE 2023-029-FB-UA ALTERED VIDEO OF PRESIDENT BIDEN

Tech Global Institute (https://techglobalinstitute.com) is a policy lab with a mission to reduce equity and accountability gaps between technology platforms and the Global Majority. In this submission, we respond to the Oversight Board's request for public comments on the Altered Video of President Biden with specific reference to the following issues.

SUBMISSION

Research into online trends of using altered or manipulated video content to influence the perception of political figures, especially in the United States.

Online trends around using altered or manipulated video content to influence

the perception of political figures, especially in the United States, include:

Deepfakes and AI-driven manipulation: Advances in artificial intelligence have made it possible to create hyper-realistic but entirely fake content. Deepfake videos use machine learning algorithms to generate fabricated videos of real people, saying or doing things they never actually said or did. Examples include:

Multiple deepfake videos of Nancy Pelosi appearing intoxicated have been widely shared on social media. The videos were later revealed to be fake, but still caused significant damage to Pelosi's reputation.

In 2018, a deepfake video of former US President Barack Obama calling Donald Trump “a total and complete dipshit” was shared online. The video was quickly debunked, but it highlighted the potential for deepfakes to be used to interfere with elections.

In 2019, a deepfake video of Boris Johnson and Jeremy Corbyn endorsing each other for prime minister was posted online in an attempt to show the potential of so-called 'deepfake' videos to undermine democracy.

In 2022, a deepfake video of Ukrainian President Volodymyr Zelenskyy surrendering to Russia was circulated online. The video was intended to demoralize Ukrainian troops and civilians, but it was quickly exposed as a fake.

In 2022, a deepfake video of Russian President Vladimir Putin appeared online claiming that Russia has won the war and that Ukraine has recognized Crimea as Russian territory. In 2023, another video shows Putin announcing that Russia was under attack and declared martial law with a full-scale mobilization plan.

An audio recording surfaced online in October 2023 in which opposition leader Sir Keir Starmer was heard berating party staffers in a profanity-laden tirade on the first day of Labour Party conference, while in another recording he is heard saying he “hated” the city of Liverpool where the conference was held.

Several altered images of former President Donald Trump surfaced online showing him hugging and kissing scientist Dr. Anthony Fauci, in altercation with policemen, posing for a mugshot and in orange prison overalls, and leading a rally.

Several altered videos of Bush appeared online. One shows him explaining future of generative AI, while another features him in Harold & Kumar.

British Prime Minister Rishi Sunak: an image appeared showing the prime minister pulling a sub-standard pint at the Great British beer festival while a woman looks on with a derisive expression, while the original photo shows Sunak appearing to have pulled a pub-level pint while the person behind him has a neutral expression.

Other examples include: Jim Acosta, Jennifer Lawrence and Steve Buscemi, David Beckham Anti-Malaria PSA, World Leaders Sing "Imagine", Dali Museum, Bill Hader impressions, Mark Zuckerberg, Joe Rogan, Nixon and a moon landing, Queen's Christmas speech, Tom Cruise TikToks, Pennsylvania cheerleader case, and an Anthony Bourdain documentary.

More information on this phenomenon is here.

Shallowfakes or Cheapfakes: This term refers to videos that have been manipulated using more basic methods than deepfakes. For example, they might involve editing out context or slowing down footage to make it look like someone is slurring their words.

Misleading edits: Some video manipulations are simple edits made to take statements out of context, clip segments that can be misconstrued, or combine unrelated pieces of footage to create a misleading narrative. This can happen by:

Splicing video clips together to create a false narrative: This is a common tactic used in political campaigns and propaganda videos. For example, a video might be edited to make it appear as if a politician said something they never actually

said.

Adding or removing audio from a video: This can be used to make it appear as if someone is saying something they never actually said, or to make them sound more or less enthusiastic about something.

Changing the speed or pitch of audio: This can be used to make someone sound more or less intelligent, or to make them sound more or less emotional.

Manipulating images with Photoshop or other editing software: This can be used to make someone look older, younger, healthier, or sicker than they actually are. It can also be used to change the background of an image or to add or remove objects from an image.

Memes and satire: While not always malicious, comedic or satirical videos that distort reality can sometimes be shared out of context, leading viewers to misconstrue the intent or believe in the content's veracity.

Political usage: While many politicians and their supporters condemn the use of manipulated videos, there have been instances where altered videos were shared either knowingly or unknowingly by political figures or their affiliates, leading to controversy.

For example, Rudolph Giuliani accidentally shared a deepfake video of Speaker Nancy Pelosi that made it seem as if she were stumbling over and slurring her words on Twitter.

While we provide examples predominantly from the U.S., manipulated media, including deepfake, have become increasingly prevalent in influencing perceptions about political figures in other parts of the world. Some examples include:

India: An Indian politician is using deepfake technology to win new voters

Nigeria: How Deepfake Audio was used to Frame Atiku, Okowa, Others in 2023 Nigeria Elections

Pakistan: PTI Chairman Imran Khan posted AI-generated image to mislead people, according to Pakistan's Minister for Information and Broadcasting

The suitability of Meta's misinformation policies, including on manipulated media, to respond to present and future challenges in this area, particularly in the context of elections.

While admitting that its misinformation policy cannot, and does not, articulate a comprehensive list of prohibited content due to the ever-evolving nature of the definition of "misinformation", Meta categorizes certain types of content that it treats as misinformation. This includes "content that is likely to directly contribute to interference with the functioning of political processes and certain highly deceptive manipulated media."

While the policy states that it aims "to promote election [...] integrity, [and] remove misinformation that is likely to directly contribute to a risk of interference with people's ability to participate in those processes," it primarily concentrates on addressing straightforward logistical misinformation, such as voting dates, locations, and eligibility, as well as participation in the census. However, the policy falls short in addressing more nuanced and insidious threats. For instance, it does not cover the sophisticated manipulation of audio and video media – such as deepfakes and synthetic media – which can fabricate speeches and actions by candidates, potentially sowing confusion and manipulate voters' perception about a candidate. It also does not adequately address other forms of misinformation and disinformation campaigns that may include false narratives about candidates, thereby failing to tackle deliberate attempts to undermine the democratic process. Additionally, the policy could further strengthen its stance on false claims related to election integrity, voter fraud, and the legitimacy of election results, which have been critical issues in recent elections.

Regarding the digitally altered media, both the misinformation policy and manipulated media policy acknowledges that content is removed because "it can go viral quickly and experts advise that false beliefs regarding manipulated media often cannot be corrected through further discourse." However, it requires the media to be words-based video content crafted with advanced AI/ML tools, effectively excluding content manipulated using less sophisticated tools or conduct-based content.

The impact of manipulated content on the political landscape is strikingly evident in recent instances. A case in point is a suspected deepfake video, reportedly created by an opposition party leader, depicting the Malaysian economic affairs minister (and potential successor to Prime Minister Mahathir Mohamad) in a sexual tryst with a party staffer. A mere suspicion of sodomy resulted in several arrests and nearly thwarted the succession plan. Another example involves the arrest of Jakarta's then-governor, an ethnic-Chinese Christian, after a video of campaign event where he said voters should not be swayed by those "using the Koran as a political tool" was edited to omit the word "use", which left plenty of room for ambiguity. During the 2019 Indonesian election, an online video depicted the seizure of millions of pre-marked ballot papers sent from China. Despite being debunked, the video had already been featured in approximately 17,000 tweets, creating doubts in the minds of voters. Current policies, particularly those addressing misinformation and manipulated media, are ill-equipped to effectively combat the rising tide of false and manipulated content during election campaigns. While the limitations of existing AI/ML tools may at present allow for the identification of manipulated content through discrepancies in, for instance, facial expressions and eye movements, this status quo is shifting. With technology evolving rapidly, the creation of hyper-realistic deepfakes, virtually indistinguishable from genuine content, is on the horizon. This poses an imminent threat to the integrity of democratic institutions and processes, demanding a more inclusive and comprehensive policy framework.

Meta's human rights responsibilities when it comes to video content that has been altered to create a misleading impression of a public figure, and how they

should be understood with developments in generative artificial intelligence in mind.

As a preliminary matter, content moderation by social media intermediaries in general, and Meta in particular, relies on constitution-esque content policies, exogenous human rights instruments, and independent commitments as normative benchmarks.

Although Meta does not have the obligations of governments under the ICCPR, their wide-ranging social and political impact necessitates them to assess the same kind of questions about protecting their users' right to freedom of expression. Previously, intermediaries moderated almost entirely without reference to the human rights implications. Now, Meta's Corporate Human Rights Policy, which serves as the foundation of the company's human rights commitments, reaffirms the company's commitments to the United Nations Guiding Principles on Business and Human Rights (UNGPs) and its interpretive guide as well as the International Covenant on Civil and Political Rights (ICCPR). These frameworks provide a principled and pragmatic model that is well-suited to the fast-paced, uncertain and complex landscape of the twenty-first century technological advancements. Commitments are implemented applying human rights policies and maintaining oversight, governance and accountability, prioritizing the most salient human rights issues in each context based on severity (scope, scale, remediability) and likelihood.

Central to Meta's human rights responsibilities, therefore, is the application of Article 19 of the ICCPR, which recognises the right of every individual,without discrimination, to freedom of expression, which includes the "freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice." It includes political dissent, discourses, and commentary on public affairs, as well as offensive expressions and fake news. However, this right is not absolute and may be subject to certain restrictions imposed by law, provided it is necessary and proportionate with respect to the rights or reputations of others. Significantly, the chilling effect that the restrictive

measures may have on expression and free flow of information necessitates its protection to be the rule, and the interference, properly justified, to remain an exception. Furthermore, Article 5(1) of the ICCPR restrains interpretation that allows actions that could destroy or excessively limit the recognized rights and freedoms. It is in these contexts that our assessment will elaborate Meta's human rights responsibilities.

With respect to content of political discourse, the value placed upon uninhibited expression is particularly high, and indeed expression considered insulting and offensive to a public figure (including heads of state and government) is insufficient to legally justify restriction. General comment No. 34 (on the right to freedoms of opinion and expression under Article 19 of the ICCPR) and General comment No. 25 (on participation in public affairs and the right to vote under Article 25 of the ICCPR) notes that the "free communication of information and ideas about public and political issues between citizens, candidates and elected representatives is essential," and must therefore be fully protected. Similarly, the Joint Declaration on Media Freedom and Democracy states that large online platforms should privilege "public interest content on their services in order to facilitate democratic discourse." A functioning democracy and freedom of expression are mutually reinforcing and complementary. However, restrictions may be imposed where the expression is not a legitimate criticism or political opposition, and instead immoderately attacks the reputation of the individual.

Here, the digitally altered content showing, on a loop, Biden placing a sticker on or around his granddaughter's chest with a suggestive song (containing the lyrics "girls rub on your titties"), accompanied by a caption calling him "a sick pedophile" for touching her breast, does not appear to be a genuine political critique or expression of dissent. It transgresses beyond the permissible limits of allowable insults that can reasonably be directed at a political leader, and instead creates misleading impressions about his character, family values, personal morals and intentions. It is a well-settled position of the Human Rights Committee that false portrayal of individuals, or deliberately spreading false rumors about them to generate public aversion, damages honour and reputation. It is also contrary to the recognitions in the preamble of the ICCPR

that “inherent dignity and of the equal and inalienable rights of all members of the human family is the foundation of freedom, justice and peace in the world” and that “these rights derive from the inherent dignity of the human person.” Given the visceral immediacy and virality of a short-form video and the ease of its dissemination using social media and encrypted messaging services to a potentially unlimited and illimitable audience, reinforced by the long-accepted truism that seeing is believing, it is likely to have greater impact on his reputation than, for instance, a still image. With more than 3.5 billion daily active users, Meta has reached a “scale of connectedness [that] is unprecedented in human history,” and this only increases the severity and likelihood of the impact. Ahead of the election, this and other videos will resurface, and social media users will either amplify or attenuate their spread. As a result, in our view, the altered content is likely to constitute an impermissible personal attack on Biden, who, notwithstanding his standing as a public and political figure, is entitled to protection of his honour, reputation and dignity under the ICCPR.

Secondly, restrictions are allowed to safeguard rights of an individual, which includes human rights and more generally in international human rights law. In this context, Article 1 of the ICCPR confers the right of self-determination to every individual, so that they are freely able to determine their political status. This right has variously been described as the “right to authentic self-government, that is, the right of a people really and freely to choose its own political regime.” Additionally, in relation to Article 25 of the ICCPR, General comment No. 25 asserts the importance of enabling individuals to freely and independently support or oppose government, and to vote, without undue influence, coercion, inducement, or manipulative interference of any kind.

Characterizing voters “mentally unwell” in this context seemingly draws a correlation between the conduct and his presidency, which could compromise his overall integrity, trustworthiness, respectability and credibility as a candidate for the 2024 presidential election, and therefore the ability of the voters to choose their next president freely and independently, without manipulative interference. Plausibly, this could be a part of broader misinformation campaign against Biden: several faux clips emerged showing

him singing the opening lyrics of Baby Shark after announcing that he will sing the national anthem, publicly admitting he is old and may have dementia, using profanities and acknowledging that he knows he "was not [the voters'] first choice in 2020", acting disoriented and asking his wife whether he took his medicine, recommending troop deployment due to Russo-Ukraine war and the impending Chinese blockade of Taiwan, and admitting to getting his salary and pension paid in "in MILFs, orgies and top-tier f**king ice-cream flavors that will make your tiny little maggot lizard brain melt faster than the polar ice caps." Considering videos often have a much more immediate and powerful effect than traditional media, and individuals tend to accept video content at face value as evidence of truth, there is an increased susceptibility for the content to change perceptions of the voters and the outcome of the upcoming elections.

In this context, it is worth noting that it is well-documented that social media can influence the outcome of elections and events surrounding it. At a colloquium organized by UNESCO and the Global Network Initiative in 2018, the use of social media and technologies to spread misinformation, disinformation and hate speech during elections was recognized. For instance, a video was shown by President Recep Tayyip Erdoğan during a political rally of his main challenger, Kemal Kılıçdaroğlu, receiving an endorsement from a designated terrorist organization. In Slovakia, inauthentic content against the leader of the Progressive party Michal Šimečka, discussing vote-rigging, was released during a 48-hour moratorium ahead of the polls opening. Another video shows an advertisement by an opposition party in which Šimečka's voice has been used to say that he "used to believe in 70 genders and pregnant men." Ultimately, he lost the election. Malicious content that overwhelmed the 2020 US presidential race and seeded distrust about the legitimacy of Biden's victory culminated in the storming of the Capitol Building on 6 January by the supporters of then-President Donald Trump who believed his lies that the election was stolen from him. A survey by Brookings found that 57% of those surveyed have seen misinformation during the 2018 US elections and 19% believe it has influenced their vote. Hany Farid observed that altered media is resulting in stolen elections, which has "real-world consequences for individuals, for societies and for democracies." Hence, there is a strong argument to be made that the

removal of altered media that creates a significantly misleading impression of Biden in particular, and public and political figures in general, is warranted.

Aligned with Meta's voluntary human rights commitments, any restriction imposed must adhere to the three-part requirements of legality, legitimate aim, and necessity (and proportionality).

With respect to the principle of legality, Meta's existing policies on manipulated media, misinformation, adult nudity and sexual activity and coordinating harm and promoting crime appear insufficient and imprecise to effectively address the significant impacts on rights and reputation. For instance, the manipulated media policy creates an unnecessary dichotomy, by conditioning takedown for words-based video content crafted with advanced AI/ML tools that may mislead an ordinary person, and thereby excluding content manipulated using less sophisticated raster or vector graphics editors, or conduct-based inauthentic media. Similarly, the sexual activity policy is too constrictive, requiring a content to be "advertisements and recognised fictional images or with indicators of fiction [that shows] [s]queezing female breasts, defined as a grabbing motion with curved fingers that shows both marks and clear shape change of the breasts." Moreover, the misinformation policy fails to address manipulated media's impact on election integrity and outcomes, despite its stated objective to promote elections and commitment to "remove misinformation that is likely to directly contribute to a risk of interference with people's ability to participate in those processes." It is a fundamental requirement that the rules must "provide sufficient guidance to those charged with their execution … to enable them to ascertain what sorts of expression[s] are properly restricted and what sorts are not," so that these rules do not confer unfettered and arbitrary discretion to Meta and provide users with adequate guidance to enable them to regulate their conducts accordingly. It is also worth noting that under Article 2(2) of the ICCPR, Meta should take necessary steps to adopt clear and specific measures to address these policy gaps in alignment with the ICCPR.

However, the bullying and harassment policy sets a more suitable tone by disallowing content with severe sexualized commentary and derogatory attacks.

Distinction is drawn between public figures and private individuals, with content related to public figures only removed where the attacks are severe, taking into account relevant context and intent. Commitments to international human rights standards require contextual assessment of the relevant historical, political, linguistic and social nuances, as well as the context within which it was made, as content is not language- or context-agnostic. Thus, identification of actual and potential human rights impacts should start at a granular level, undertaking multi-faceted analyses of the specific user (or user category), as well as the geographic region and contexts in which use may lead to adverse impacts. Thus, the higher threshold for public figures notwithstanding, the incestuous connotation and unsubstantiated allegations of paedophilia against Biden, coupled with a sexually explicit song, is sufficiently severe to warrant removal of the at-issue content, especially considering Meta's intolerance towards such behavior and its stated position to "strive to create a more inclusive and equitable online environment for all users through our Community Standards and Community Guidelines, which prohibit hate speech, bullying, and harassment."

Any restriction on expression should pursue one of the legitimate aims listed in the ICCPR, including the rights or reputation of others. For reasons mentioned above, this requirement appears to be satisfied in this context.

Finally, the principle of necessity and proportionality provides that any restrictions on expression "must be appropriate to achieve their protective function; they must be the least intrusive instrument amongst those which might achieve their protective function; [and] they must be proportionate to the interest to be protected." Any restriction should substantiate in specific and individualized fashion the precise nature of the harm, and should ideally be limited to that specific content and not on the operation of entire sites and systems. Here, the significance of the internet and altered media in the context of the principle of necessity and proportionality is worth highlighting.

On a balance, we consider the removal of the video to serve the protective function of safeguarding honour, dignity and reputation of Biden (and his

granddaughter) as well as the integrity of the elections, and a content-specific restriction to be the least intrusive and most proportionate response to the situation.

Link to Attachment

PC-18045

CASE 2023-029-FB-UA	PC-18046	United States & Canada
Case number	Public comment number	Region
Gregory	Angelo	English
Commenter's first name	Commenter's last name	Commenter's preferred language
New Tolerance Campaign		Yes
Organization		Response on behalf of organization

Full Comment

Members of the Oversight Board:

The New Tolerance Campaign (NTC) is a watchdog organization whose mission is to ensure institutions consistently apply their stated policies and values, particularly when it comes to fostering free speech and open dialogue. As such, we welcome the opportunity to submit comment in response to Facebook's request seeking guidance from the Oversight Board regarding the "Altered Video of President Biden" case.

Despite repeated entreaties by a Facebook user asking that the video in question be taken down, Meta did the right thing in refusing to do so. Determining whether or not posted media constitutes "misinformation" is ultimately the responsibility of the platform user. Caveat videntium — viewer beware.

The Meta Misinformation Policy states that Meta removes videos that have

"been edited or synthesized, beyond adjustments for clarity or quality, in ways that are not apparent to an average person, and would likely mislead an average person to believe a subject of the video said words that they did not say." Meta does not, however, define what the platform believes to be an "average person," and as a result its Misinformation Policy is entirely subjective.

The video at the heart of this case seeks to portray President Biden — and those who voted for him — as deviants. One might not agree with the sentiment, but the presumption cannot be made that this video was created to sow "political disinformation." Assigning motive to manipulated media is, once again, subjective absent a clear assertion by the media's manipulator.

NTC offers the following recommendations regarding the broader issues for which the Oversight Board seeks third-party input on Meta's manipulated media policy:

Research into online trends of using altered or manipulated video content to influence the perception of political figures, especially in the United States, and the suitability of Meta's misinformation policies, including on manipulated media, to respond to present and future challenges in this area, particularly in the context of elections.

While manipulated video is still relatively crude, manipulated audio presents immediate issues in the political landscape that Meta will need to address soon. In the current election cycle, fabricated audio has already become a source of controversy. A political advertisement released in September by Courageous Conservatives used AI-generated audio to give the impression that Republican presidential candidates Nikki Haley and Senator Tim Scott vocally support a "woke" agenda. The policies criticized in the ad were indeed the political positions of the respective candidates, but the voices in the ad were created using artificial intelligence. In July, Never Back Down, a Super PAC supporting Republican presidential candidate Governor Ron DeSantis, released a video using an AI-generated voice of President Donald Trump. The quotes used in the advertisement featured text from Mr. Trump's Truth Social account that

appeared to be “spoken” by the man himself. If an AI-generated voice is used to vocalize something a political figure wrote but never said, would that constitute “misinformation” or “manipulated” media?

Research into the efficacy of alternative responses to political disinformation or misinformation beyond content removal, such as fact-checking programs or labelling (also known as “inform treatments”). Additionally, research on avoiding bias in such responses.
While imperfect, the “Community Notes” feature implemented on X (the social media platform formerly known as Twitter) shows promise as a way to more expeditiously address posts on the platform missing context or containing outright falsehoods. According to the X website, Community Notes was created to “empower” users “to collaboratively add context to potentially misleading posts. Contributors can leave notes on any post and if enough contributors from different points of view rate that note as helpful, the note will be publicly shown on a post.” Rather than giving authority to a select group of fact-checkers, Community Notes employs a hive-mind approach to questionable content. In September, X gave users the ability to add Community Notes to video content, which the platform automatically applies to matching videos posted by other users. The result, according to X, is “[a] highly-scalable way of adding context to edited clips, AI-generated videos, and more.” More generally, expanding the ability for human beings to have a say in content policing will become increasingly important in an age of algorithms, bots, and AI.

NTC thanks the Oversight Board for the opportunity to submit this comment for your consideration, and would be glad to engage with you further on this matter should you see fit.

Link to Attachment

PC-18046

CASE 2023-029-FB-UA	PC-18047	Europe
Case number	Public comment number	Region
Hakim	Charles	English
Commenter's first name	Commenter's last name	Commenter's preferred language
Nexus Horizon		Yes
Organization		Response on behalf of organization

Full Comment

DID NOT PROVIDE

Link to Attachment

PC-18047

CASE 2023-029-FB-UA	PC-18048	United States & Canada
Case number	Public comment number	Region
Kayla	Gogarty	English
Commenter's first name	Commenter's last name	Commenter's preferred language
Media Matters For America		Yes
Organization		Response on behalf of organization

Full Comment

The Board has asked respondents for comments and recommendations on a case related to Meta’s content moderation policies and enforcement practices regarding manipulated media on its platforms. Meta’s manipulated media policies, first introduced in 2020, are narrow, plagued by loopholes (including one that exempts politicians from fact-checking), and inconsistently enforced. Coupled with an inadequate labeling system, these deficiencies have resulted in widespread political misinformation and manipulated media across Meta's platforms. If left unaddressed, these deficiencies in Meta’s manipulated media policies are likely to have great consequences during the 2024 U.S. election cycle.

The Board’s current case on manipulated media — involving a video of President Joe Biden that was misleadingly edited to make it seem like he inappropriately touched his granddaughter’s chest — exemplifies a growing problem on Meta’s platforms that requires meaningful policy changes.

Meta has repeatedly chosen profit and positive press over the safety of its users, seemingly out of fear of the relentless but false claims from conservatives that attempts at moderation and policy enforcement amount to censorship of right-wing viewpoints. As a result, Meta has repeatedly bent its rules, giving preferential treatment to and carving out exemptions for right-wing media and politicians, allowing inaccurate and harmful content to proliferate on the platform.

In order to meaningfully address the problem of political misinformation on its platforms, Meta should expand and robustly enforce its manipulated media policies, particularly as they relate to political figures and issues.

Misleading and altered videos proliferated across Meta platforms during the 2020 and 2022 election cycles

Facebook developed a manipulated media policy in January 2020 following various instances in which misleading clips of Democrats circulated on its platforms, but there were still numerous instances during the 2020 and 2022 election cycles where the platform failed to adequately enforce these policies.

For instance, in March 2020 a misleadingly edited video circulated on Facebook of then-Democratic candidate Joe Biden making what seemed to be an accidental endorsement of then-President Donald Trump: “Excuse me. We can only re-elect Donald Trump.” Facebook did not remove the deceptively edited clip of Biden and only labeled it as “partly false” after one of its partner fact-checking organizations noted that the video was misleading.

In July 2020, Facebook let a digitally altered video purporting to show Biden appearing at a dinner with a man in blackface circulated on its platform.

In summer 2020, a video, which had been altered to make it look and sound as though then-House Speaker Nancy Pelosi (D-CA) was drunk and slurring her words during a press conference, spread across Facebook. The video earned

millions of views before the platform put a fact-check label on it, even though the video was seemingly eligible for removal under Facebook's manipulated media policy. In May 2019 — before the platform established its manipulated media policy — Facebook treated a similarly doctored video of the former speaker that had proliferated on the platform the same way, only applying a fact-check label to it.

Additionally, in July 2022, a misleading manipulated video of Biden awarding the Medal of Honor to a veteran spread among users on Facebook.

And in June of this year, Facebook failed to curb the spread of a deceptively edited clip of Biden that showed him claiming to have "sold a lot of state secrets and a lot of very important things." In reality, Biden opened a meeting with Indian Prime Minister Narendra Modi and the CEOs of Apple, Google, and Microsoft, joking, "I started off without you, and I sold a lot of state secrets and a lot of very important things that we shared." Biden followed up by acknowledging that the comment had been a joke: "Now all kidding aside, look, we're teaming up to design and develop new technologies that are going to transform the lives of our people around the world." This context was omitted from the clip that went viral on Meta's platform.

Facebook even allowed a post from Rep. Marjorie Taylor Greene (R-GA) that contained the clip to earn over 1.2 million views and be shared over 33,000 times.

In addition to allowing manipulated media to spread, Facebook has also profited from it. Trump's reelection campaign ran thousands of Facebook ads in the months leading up to the 2020 election that featured manipulated photos of Biden. Over 2,500 ads featured images in which Biden's coloring and facial features have been edited to make him appear older, while more than a thousand other ads were edited to add an earpiece to images of Biden while accusing him of cheating during the debate.

The prevalence of manipulated media is a product of Meta's inadequate policy,

loopholes, and enforcement failures

Meta's narrow manipulated media policy is the main issue presented in the Oversight Board's current case on the deceptively edited video of Biden. The manipulated media policy's sole focus on content that "would likely mislead an average person" to think "a subject of the video said words that they did not say" or "the video is the product of artificial intelligence or machine learning" is insufficient and will allow political misinformation to proliferate.

The policy also has loopholes that users can exploit since it does not prohibit content that has been misleadingly edited "to omit words that were said or change the order of words that were said." Meta should expand this policy to include videos that have been deceptively edited to omit words or actions because in practice this content leads viewers to believe a subject did something that did not happen.

What's more, Meta's fact-checked misinformation policy exempts "speech from politicians," creating another loophole for political figures to push manipulated media on the platform. As they currently stand, Meta's policies enable political figures to post manipulated media without consequence — which could result in dire consequences to our nation's political systems, as they have before.

Meta's labeling system has failed to adequately control the spread of misinformation on its platforms

The case currently before the Board is just one of Meta's many failures to prevent the spread of misinformation on its platforms. This failure is not without consequences — misinformation on Meta's platforms has contributed to real-world violence and many other harms.

Media Matters has documented the trend of Meta's misinformation failures over the years. According to a May 2022 report that examined posts on Meta's platforms from May 1, 2021, through April 30, 2022, Media Matters found over 13,500 violations of Meta's policies on Facebook and Instagram, with much of

the violative content lacking appropriate labels. Of the more than 13,500 violations, nearly 10,000 were from posts and ads that allowed Trump to evade his then-ban from the platform, nearly 1,500 were from posts that violate the platform's harmful health information, and nearly 1,000 were violations of Meta's policy against hate speech. The remaining posts appeared to violate the platform's election misinformation, dangerous individuals and organizations, and debunked content policies.

One of the primary tools that Meta uses to combat misinformation on its platforms is labeling. However, internal and external data shows that labels have been ineffective and the platforms' application of them has been inconsistent at best. For example, in May 2021, a statement from Trump that contained a debunked claim about Maricopa County's election database being deleted circulated on Facebook. Although Facebook labeled some posts with the debunked statement as "false information," Media Matters identified dozens of posts that had no such label.

A crucial component of misinformation is the "continued influence effect," which is the well-documented phenomenon of misinformation continuing to influence people even after it has been corrected. Meta's labeling system does not account for this effect, and allows users to be exposed to misinformation with a label applied to it — instead of removing posts that contain harmful misinformation altogether.

A June 2021 Media Matters report further demonstrated the inadequacy of labels — showing that on average, posts from Trump that received labels earned over two times more interactions per post than his overall posts.

Meta's manipulated media policy is an insufficient solution to political misinformation on the platform

In the current case before the Board, the manipulated video of Biden would lead viewers to believe he took actions that he did not take. In reality, Biden did not repeatedly and inappropriately touch his granddaughter's chest. Moreover, the

accompanying caption of the post in question that claims Biden is “a sick pedophile” for touching his granddaughter’s chest is blatantly false, but the manipulated media seems to substantiate it.

Meta’s policies should prohibit this content. As the 2024 election cycle approaches, Trump and his network of right-wing politicians and media figures are set to spread political misinformation on Meta’s platforms. It is crucial that Meta better enforces and expands its manipulated media policy to prevent political misinformation on its platforms ahead of 2024.

Link to Attachment

PC-18048