hub.packtpub.com

Deepfakes House Committee Hearing: Risks, Vulnerabilities and Recommendations | Packt Hub

Vincy Davis

18-22 minutes

Last week, the House Intelligence Committee held a hearing to examine the public risks posed by “deepfake” videos. Deepfake is identified as a technology that alters audio or video and then is passed off as true or original content. In this hearing, experts on AI and digital policy highlighted to the committee, deepfakes risk to national security, upcoming elections, public trust and the mission of journalism. They also offered potential recommendations on what Congress could do to combat deepfakes and misinformation.

The chair of the committee Adam B. Schiff, initiated the hearing by stating that it is **time to regulate the technology of deepfake videos as it is enabling sinister forms of deception and disinformation** by malicious actors. He adds that “*Advances in AI or machine learning have led to the emergence of advance digitally doctored type of media, the so-called deepfakes that enable malicious actors to foment chaos, division or crisis and have the capacity to disrupt entire campaigns including that for the Presidency*.”

For a quick glance, here’s a TL;DR:

- Jack Clerk believes that governments should be in the business of measuring and assessing deepfake threats by looking directly at the scientific literature and developing a base knowledge of it.
- David Doermann suggests that tools and processes which can identify fake content should be made available in the hands of individuals, rather than relying completely on the government or on social media platforms to police content.
- Danielle Citron warns that the phenomenon of deepfake is going to be increasingly felt by women and minorities and for people from marginalized communities.
- Clint Watts provides a list of recommendations which should be implemented to prohibit U.S. officials, elected representatives and agencies from creating and distributing false and manipulated content.
- A unified standard should be followed by all social media platforms. Also they should be pressurized to have a 10-15 seconds delay in all videos, so that they can decide, to label a particular video or not.
- Regarding 2020 Presidential election: State governments and social media companies should be ready with a response plan, if a fake video surfaces to cause disrupt.
- It was also recommended that the algorithms to make deepfakes should be open sourced.
- Laws should be altered, and strict actions should be awarded, to discourage deepfake videos.

Being forewarned is forearmed in case of deepfake technology

Jack Clerk, OpenAI Policy Director, highlighted in his testimony that he **does not think A.I. is the cause of any disruption, but actually is an "*accelerant to an issue which has been with us for some time***." He adds that computer software aligned with A.I. technology has become significantly cheaper and more powerful, due to its increased accessibility. This has led to its usage in audio or video editing, which was previously very difficult. Similar technologies are being used for production of synthetic media. Also deepfakes are being used in valuable scientific research.

Clerk suggests that interventions should be made to avoid its misuse. He believes that "*it may be possible for large-scale technology platforms to try and develop and share tools for the detection of malicious synthetic media at both the individual account level and the platform level. We can also increase funding.*" *He* strongly believes that governments should be in the business of measuring and assessing these threats by looking directly at the scientific literature and developing a base knowledge. Clerk concludes saying that "***being forewarned is forearmed here.***"

Make Deepfake detector tools readily availaible

David Doermann, the former Project Manager at the Defense Advanced Research Projects Agency mentions that the phrase '*seeing is believing*' is no longer true. He states that there is nothing fundamentally wrong or evil about the technology, like basic image and video desktop editors, deepfakes is only a tool. There are a lot of positive applications of generative networks just as there are

negative ones. He adds that, as of today, there are some solutions that can identify deepfakes reliably. However, Doermann fears that it's only a matter of time before the current detection capabilities will be rendered less effective. He adds that "*it's likely to get much worse before it gets much better.*"

Doermann suggests that **tools and processes which can identify such fake content should be made available in the hands of individuals, rather than relying completely on the government or on social media platforms to police content. At the same time, there should also be ways to verify it or prove it or easily report it.** He also hopes that automated detection tools will be developed, in the future, which will help in filtering and detection at the front end of the distribution pipeline. He also adds that "***appropriate warning labels should be provided, which suggests that this is not real or not authentic, or not what it's purported to be. This would be independent of whether this is done and the decisions are made, by humans, machines or a combination***."

Groups most vulnerable to Deepfake attacks

Women and minorities

Danielle Citron, a Law Professor at the University of Maryland, **describes Deepfake as** "***particularly troubling when they're provocative and destructive*.**" She adds that, we as humans, tend to believe what our eyes and ears are telling us and also tend to share information that confirms our biases. It's particularly true when that information is novel and negative, so the more salacious, we're more willing to pass it on. She also specifies that the

deepfakes on social media networks are ad-driven. When all of this is put together, it turns out that ***the more provocative the deepfake is, the salacious will be the spread virally.*** She also informed the panel committee about an incident, involving an investigative journalist in India, who had her posters circulated over the internet and deepfake sex videos, with her face morphed into pornography, over a provocative article.

Citron thus states that "*the economic and the social and psychological harm is profound*". Also based on her work in cyber stalking, she believes that ***this phenomenon is going to be increasingly felt by women and minorities and for people from marginalized communities.*** She also shared other examples explaining the effect of deepfake on trades and businesses. Citron also highlighted that "*We need a combination of law, markets and really societal resilience to get through this, but the law has a modest role to play*." She also mentioned that though there are laws to sue for defamation, intentional infliction of emotional distress, privacy torture, these procedures are quite expensive. She adds that criminal law offers very less opportunity for the public to push criminals to the next level.

National security

Clint Watts, a Senior Fellow at the Foreign Policy Research Institute provided insight into how such technologies can affect national security. He says that "***A.I. provides purveyors of disinformation to identify psychological vulnerabilities and to create modified content digital forgeries advancing false narratives against Americans and American interests.***"

Watts suspects that Russia, "*being an enduring purveyor of*

disinformation is and will continue to pursue the acquisition of synthetic media capability, and employ the output against adversaries around the world." He also adds that China, being the U.S. rival, will join Russia "*to get vast amounts of information stolen from the U.S. The country has already shown a propensity to employ synthetic media in broadcast journalism. They'll likely use it as part of disinformation campaigns to discredit foreign detractors, incite fear inside western-style democracy and then, distort the reality of audiences and the audiences of America's allies.*" He also mentions that deepfake proliferation can present a danger to American constituency by demoralizing it. Watts suspects that the U.S. diplomats and military personnel deployed overseas, will be prime target for deepfake driven disinformation planted by adversaries.

Watts provided a list of recommendations which should be implemented to "***prohibit U.S. officials, elected representatives and agencies from creating and distributing false and manipulated content***."

- The U.S. government must be the sole purveyor of facts and truth to constituents, assuring the effective administration of democracy via productive policy debate from a shared basis of reality.
- Policy makers should work jointly with social media companies to develop standards for content and accountability.
- The U.S. government should partner with private sectors to implement digital verification designating a date, time and physical origination of the content.
- Social media companies should start labeling videos, and forward the same across all platforms. Consumers should be able to

determine the source of the information and whether it's the authentic depiction of people and events.

- The U.S. government from a national security perspective, should maintain intelligence on capabilities of adversaries to conduct such information. The departments of defense and state should immediately develop response plans, for deepfake smear campaigns and mobilizations overseas, in an attempt to mitigate harm.
- Lastly he also added that public awareness of deepfakes and signatures, will assist in tamping down attempts to subvert the U.S. democracy and incite violence.

Schiff asked the witnesses, if it's "*time to do away with the immunity that social media platforms enjoy*", Watts replied in the affirmative and listed suggestions in three particular areas.

- If social media platforms see something spiking in terms of virality, it should be put in a queue for human review, linked to fact checkers, then down rate it and don't let it into news feeds. Also make the mainstream understand what is manipulated content.
- Anything related to outbreaks of violence and public safety should be regulated immediately.
- Anything related to elected officials or public institutions, should immediately be flagged and pulled down and checked and then a context should be given to it.

Co-chair of the committee, Devin Nunes asked Citron what kind of filters can be placed on these tech companies, as "*it's not developed by partisan left wing like it is now, where most of the time, it's conservatives who get banned and not democrats*". Citron

suggested that proactive filtering won't be possible and hence companies should react responsibly and should be bipartisan. She added that "*but rather, is this a misrepresentation in a defamatory way, right, that we would say it's a falsehood that is harmful to reputation. that's an impersonation, then we should take it down. This is the default I am imagining.*"

How laws could be altered according to the changing times, to discourage deepfake videos

Citron says that laws could be altered, like in the case of Section 230 C. It states that "*No speaker or publisher — or no online service shall be treated as a speaker or publisher of someone else's content*." This **law can be altered to "*No online service that engages in reasonable content moderation practices shall be treated as a speaker or publisher of somebody else's content.***" Citron believes that avoiding reasonability could lead to negligence of law. She also adds that "*I've been advising Twitter and Facebook all of the time. There is meaningful reasonable practices that are emerging and have emerged in the last ten years. We already have a guide, it's not as if this is a new issue in 2019. So we can come up with reasonable practices.*"

Also Watts added that if any adversary from big countries like China, Iran, Russia makes a deepfake video to push the US downwards, we can trace them back if we have aggressive laws at our hand. He says it could be anything from an "***arrest and extradition, if the sanction permits, response should be individually, or in terms of cyber response***", could help us to discourage deepfake.

How to slow down the spread of videos

One of the reasons that these types of manipulated images gain traction is because it's almost instantaneous – they can be shared around the world, shared across platforms in a few seconds. Doermann says that these social media platforms must be pressurized to have a **10-15 seconds delay**, so that it can be decided whether to label a particular video or not. He adds that "*We've done it for child pornography, we've done it for human trafficking, they're serious about those things. This is another area that's a little bit more in the middle, but I think they can take the same effort in these areas to do that type of triage.*" This delay will allow **third parties or fact checkers to decide on the authenticity of videos and label them**. Citron adds that this is where labelling a particular video can help, "*I think it is incredibly important and there are times in which, that's the perfect rather than second best, and we should err on the side of inclusion and label it as synthetic*."

The representative of Ohio, Brad Wenstrup added that we can have **internal extradition laws**, which can punish somebody when "*something comes from some other country, maybe even a friendly country, that defames and hurts someone here*". There should be **an agreement among nations that "*we'll extradite those people and they can be punished in your country for what they did to one of your citizens.*"**

Terri Sewell, the Representative of Alabama further probed about the current scenario of detecting fake videos, to which Doermann replied that currently we have enough solutions to detect a fake video, however with a constant delay of 15-20 minutes.

Deepfakes and 2020 Presidential elections

Watts says that he's concerned about deepfakes acting on the eve of election day 2020. Foreign adversaries may use a standard disinformation approach by "using an organic content that suits their narrative and inject it back." This can escalate as more people are making deepfakes each year. He also added that "*Right now I would be very worried about someone making a fake video about electoral systems being out or broken down on election day 2020*." So **state governments and social media companies should be ready with a response plan** in the wake of such an event.

Sewell then asked the witnesses for suggestions on campaigns to political parties/candidates so that they are prepared for the possibility of deepfake content. Watts replied that the most important thing to counter fake content would be a **unified standard**, that all the social media industries should follow. He added that "*if you're a manipulator, domestic or international, and you're making deep fakes, you're going to go to whatever platform allows you to post anything from inauthentic accounts. they go to wherever the weak point is and it spreads throughout the system.*" He believes that this system would help counter extremism, disinformation and political smear campaigns. Watts added any sort of lag in responding to such videos should be avoided as "*any sort of lag in terms of response allows that conspiracy to grow*."

Citron also pointed out that firstly all candidates should have a clear policy about deep fakes and should commit that they won't use them or spread them.

Should the algorithms to make deepfakes be open

sourced?

Doermann answered that the algorithms of deepfakes have to be absolutely open sourced. He says that though this might help adversaries, but they are anyway going to learn about it. He believes this is significant as, "*We need to get this type of stuff out there. We need to get it into the hands of users. There are companies out there that are starting to make these types of things.*" He also states that people should be able to use this technology. **The more we educate them, more the tools they learn, more the correct choices people can make.**

On Mark Zuckerberg's deepfake video

On being asked to comment on the decision of Mark Zuckerberg to not take down his deepfake video from his own platform, Facebook, Citron replied that Mark gave a perfect example of "*satire and parody*", by not taking down the video. She added that private companies can make these kinds of choices, as they have an incredible amount of power, without any liability, "it seemed to be a conversation about the choices they make and what does that mean for society. So it was incredibly productive, I think."

Watts also opined that he likes Facebook for its consistency in terms of enforcement and that they are always trying to learn better things and implement it. He adds that he really like Facebook as its always ready to hear *"from legislatures about what falls inside those parameters. The one thing that I really like is that they're doing is identifying inauthentic account creation and inauthentic content generation, they are enforcing it, they have increased the scale,and it is very very good in terms of how they have scaled it*

up, it's not perfect, but it is better."

Read More: *Zuckberg just became the target of the world's first high profile white hat deepfake op. Can Facebook come out unscathed?*

On the Nancy Pelosi doctored video

Schiff asked the witnesses if there is any account on the number of millions of people who have watched the doctored video of Nancy Pelosi, and an account of how many of them ultimately got to know that it was not a real video. He said he's asking this as according to psychologists, people never really forget their once constructed negative impression. Clarke replied that "*Fact checks and clarifications tend not to travel nearly as far as the initial news.*" He added that its becomes a very general thing as "*If you care, you care about clarifications and fact checks. but if you're just enjoying media, you're enjoying media. You enjoy the experience of the media and the absolute minority doesn't care whether it's true."*

Schiff also recalled how **in 2016, "*some foreign actresses, particularly Russia had mimicked black lives matter to push out continent to racially divide people***." Such videos gave the impression of police violence, on people of colour. They "*certainly push out videos that are enormously jarring and disruptive*."

All the information revealed in the hearing was described as "***scary and worrying***", by one of the representatives. The hearing was ended by Schiff, the chair of the committee, after thanking all the witnesses for their testimonies and recommendations.

For more details, head over to the full Hearing on deepfake videos by the House Intelligence Committee.

Read Next

Worried about Deepfakes? Check out the new algorithm that manipulate talking-head videos by altering the transcripts

Lawmakers introduce new Consumer privacy bill and Malicious Deep Fake Prohibition Act to support consumer privacy and battle deepfakes

Machine generated videos like Deepfakes – Trick or Treat?