WITNESS

SEE IT
FILM IT
CHANGE IT

DEEPFAKES: PREPARE NOW

WORKSHOP REPORT (ENGLISH)

BRAZIL, JULY 2019

DEEPFAKES: PREPARE NOW (Perspectives from Brazil)

Executive Summary: 1st national-level meeting + Key threats and preferred solutions (English)

https://lab.witness.org/brazil-deepfakes-prepare-now/

'Deepfakes' and synthetic media are new forms of audiovisual manipulation that allows people to create realistic simulations of someone's face, voice or actions. They enable people to make it seem like someone said or did something they didn't. They are getting easier to make, requiring fewer images to build them from, and they are increasingly being commercialized.

Currently deepfakes overwhelmingly impact women, targeted with non-consensual sexual images and videos, but there are fears they will impact more broadly across society. Solutions are being proposed for how to handle malicious uses of these new tools, and it is critical that this discussion be informed from a global perspective, rather than a strongly US or European-centric point of view.

The first national-level meeting on deepfakes preparedness

On July 25th 2019, WITNESS held a convening on "Deepfakes and synthetic media: Prepare yourself now" in São Paulo, Brazil. To our understanding it was the first national-level multi-disciplinary discussion on how to pragmatically understand and prepare for this potential threat.

The meeting aimed to explore and prioritize pragmatic solutions for the prevention and defense against a dark future of video and audio made with artificial intelligence (AI) techniques, with a particular focus on the threats identified in Brazil and solutions desired by a range of stakeholders. The workshop participants included journalists, fact-checkers, technologists, civic activists and others. It was part of a WITNESS initiative focused on how to better protect and uphold marginal voices, civic journalism, and human rights as emerging technologies such as AI intersect with disinformation, media manipulation, and rising authoritarianism. The workshop was also supported by the team at WITNESS Brasil. More information on WITNESS' deepfakes work including previous convening and workshop reports is available in English at: wit.to/Synthetic-Media-Deepfakes

The workshop was structured to learn about the technologies of deepfakes and synthetic media creation and detection; their current use in attacks on women and gender-based violence and in cultural critique and satire. Then participants placed this in the context of existing challenges of misinformation and disinformation in Brazil and focused on prioritizing perceived threats and solutions.

What are the threats?

After learning about the technological possibilities of deepfakes and discussing the current situation in Brazil participants prioritized these key threats as areas where new forms of manipulation might *expand* existing threats, *introduce* new threats, *alter* existing threats, and *reinforce* other threats.

- Journalists and civic activists will have their reputation and credibility attacked. This echoes global concerns.

- Public figures will face non-consensual sexual imagery and gender-based violence

- Social movement will face attacks on the credibility and safety of their leaders as well as their public narratives

- There will be attacks against judicial processes and the probative value of video for both news and evidence as video is discredited, claimed to be inauthentic even when it isn't, or processes are overwhelmed by the burden of proving true from false

- Deepfakes will yet be another weapon contributing to conspiracy campaigns

- As deepfakes become more common and easier to make at volume, they will contribute to a firehose of falsehood that to floods media verification and fact-checking agencies with content they have to verify

- Similar volumes of falsehood contribute to cumulative creation of distrust in institutions and a 'zero trust' society where truth is replaced by opinion

- Micro-targeting of increasingly customized AI-generated content will use a person or a group's psychological profile to carry-out a very effective targeting with falsified content in order to reinforce an existing position or opinion they hold.

What are the solutions we need?

Participants discussed a range of the solutions being proposed at a global level, but that often lead out of Silicon Valley and legislative actions in Washington DC and Brussels. These included the following areas:

1. Can we teach people to spot deepfakes?

2. How do we build on existing journalistic capacity and coordination?

3. Are there tools for detection? (and who has access?)

4. Are there tools for authentication? (and who's excluded)

5. Are there tools for hiding our images from being used as training data?

6. What do we want from commercial companies producing synthesis tools?

7. What should platforms and lawmakers do?

After a discussion on the status of detection efforts, the inadequacy of training people to 'spot' deepfakes and the current platform efforts by Facebook and others, participants focused on the which solutions felt most relevant to focus on in Brazil.

We need media literacy contextualized in bigger misinformation and disinformation problem, especially for grassroots communities

Rather than talking about the current algorithmic "Achilles heel" of any current deepfake creation process – which is usually a technical glitch that will disappear as techniques improve -- we should work to create a critical thinking that can make people doubt materials, check sources and provenance and corroboration, distinguish opinion and propaganda, and look for veracity before believing and sharing. This needs to be about the broader problem of disinformation, misinformation and 'fake news' as well as unpacking how narratives are constructed and shared, and must prioritize grassroots communities and influencers who work with these communities

There is a lack of public understanding of what's possible with new forms of video and audio manipulation. We should prioritize listening first to what people already know or presume about deepfakes befor building on this understanding without scaremongering, for example using the power of influencers like deepfake satirists to explain how works. Working together with other existing projects, initiatives, coalitions and fact-checking agencies is very important not only to share tools and skills, but also to exchange experiences and new technologies.

Detection tools need to be cheap, accessible and explainable for citizens and journalists

Participants, particularly from the journalism and fact-checking world, were concerned about how the nature of detection would always put journalists at a disadvantage. They already grapple with the difficulties of finding and debunking false claims especially within closed networks, let alone new forms of manipulation like deepfakes, for which they don't have the detection tools.

More and more investment is going into the development of tools for detecting deepfakes using new forms of media forensics and adaptations of the same algorithms used to create the synthetic media. But there are questions of who these tools will be available to, and how existing problems of 'shallowfakes' will also be dealt with. Journalists also reiterate that platforms like YouTube and WhatsApp haven't solved for existing problems – you still can't easily check whether an existing video is a 'shallowfake', a video that is simply slightly edited or just renamed and shared claiming its something else. In the absence of tools to detect the existing massive volume of shallowfakes – for example, a reverse video search out of WhatsApp – then deepfakes detection is a luxury.

As big companies and platforms like Facebook invest in detection tools they need to build detection tools that are clear, transparent and trustworthy, as well as accessible to many levels of journalists and citizens. The earlier that deepfakes and other falsifications can be spotted the better.

A big part of accessibility is better media forensics tools that are cheap and available to all – and challenging the economic incentives that build for synthesizing falsehood not detecting it -- but this needs to be combined with journalistic capacity in new forms of verification and media forensics.

Platforms like Facebook, YouTube, Google and Whatsapp need to be part of the solution with transparency and support to separate truth from falsehood

Platforms, closed messaging apps, search engines, social media networks and video-sharing sites will also be the places where these manipulations are shared. Some topics and questions we should discuss are: What role should social networks and other platforms play in fighting deep fakes? What should be the limits? How should they provide access to detection capabilities? How should they signal to users that content is true or fake or in some way manipulated in ways they cannot see? Should they remove certain kinds of manipulated content and under what criteria?

As a starting point, participants noted that platforms need to be more transparent on what they learn about how fake news is distributed on them. They need to rethink how far closed messaging can reach and how to control the spread of mis and disinformation.

****

For more information on WITNESS' recommendations for preparation for deepfakes see: wit.to/Synthetic-Media-deepfakes

For more information on WITNESS' recommendations for what journalists need to prepare (globally) see: https://lab.witness.org/projects/osint-digital-forensics/

MENU

8. THE WORKSHOP

On July 25th 2019, WITNESS held a workshop called "Deepfakes and synthetic media: prepare yourself now" at São Paulo, Brazil. The workshop aimed to explore and prioritize pragmatic solutions for the prevention and defense against a dark future with video and audio made with artificial intelligence (AI) techniques.

GOALS OF THE WORKSHOP

1 / Expand the understanding of these new technologies and its implications for journalists, researchers and human rights defender.

2 / Recognize potential positive uses for this technology and start to build a global mapping of innovations in those areas and a common understanding of the malicious uses of images and audios generated by AI against public discourse, trust in the press and human rights documentation, as well as recognize potential answers for those threats.

3 / Identify and prioritize threat models in the usage of these tools in the Brazilian context.

4 / Revise and prioritize potential pragmatic, tactical and normatives answers currently being discussed about detection, authentication, media organization coordination, as well as communication to the public about those news forms of mediatic manipulation based on AI.

5 / Identify priorities for continuous discussions among strategic groups and exchanges between discussions in Brazil and in the world.

The workshop builds on a WITNESS initiative focused on how to better protect and uphold marginal voices, civic journalism, and human rights as emerging technologies such as AI intersect with disinformation, media manipulation, and rising authoritarianism. More information on this work including previous convening and workshops reports is available in English at: https://lab.witness.org/synthetic-media-and-deep-fakes/. The convening in Sao Paulo is the first of a number WITNESS will facilitate worldwide to ensure the understanding of and solutions to deepfakes reflect a global perspective.

8.1. Who attended?

The workshop had the participation of over 25 people from different areas and backgrounds: communicators, academic researchers and technologists, journalists working with fact-checking and human rights defenders. They were all invited to engage in conversations oriented to look for solutions and answers for those challenges put by "deepfake" media.

Organizations with representatives in the meeting were: Agência Lupa, Agência Pública, Agência Reuters, Aos Fatos, Coding Rights, Coletivo Papo Reto, Coletivo Tulipa Negra Direitos Humanos, Facebook, First Draft, Folha, FSB Comunicação, G1 - Rede Globo, Instituto de Computação da Universidade de Campinas, Instituto de Tecnologia e

Sociedade do Rio, Marco Zero Conteúdo, Medialab UFRJ, Politécnico de Milão (Itália), PUC-PE, Revista Época, Secretária Municipal de Educação da Prefeitura de São Paulo, Universidade Federal de Goiás, UNIBES Cultural, UNISINOS, Universidade de Padova (Itália) and WITNESS.

Under Chatham House rules, we are not attributing particular comments to individual speakers other than the presenters.

8.2. How did the meeting go?

The meeting started with a quick presentation by Sam Gregory, the program director of WITNESS and an expert in deepfakes and a quick warm-up in which participants gathered in-circle and presented themselves before positioning themselves in a spectrum in the room accordingly to their answers for this question:

*** Do you understand what “deepfakes” means in our world?**

*** How much are you worried about deepfakes?**

It was clear that although most non-experts were still not very familiar with the technology, those who had knowledge about it had deep worries about them - especially activists.

The content of the day included a series of presentations by seven different speakers about deepfakes and synthetic media creation and different perspectives on inequalities, threats to minorities and grassroots activists. Three different exercises to prioritise and discuss threats and think about potential solutions were also carried out during the day.

AGENDA OF THE DAY

MORNING (09am to 12pm)

* Participants, WITNESS and workshop presentation

* Introduction to deepfakes and "synthetic medias"

* Technical perspectives

* Deepfakes and inequalities

* Deepfakes and desinformation in Brazil

AFTERNOON (2pm to 5pm)

* Group discussion about threats models and vulnerabities

* Interdisciplinary solutions and approaches

* Understanding the Brazilian context

* Finalization and to-dos.

EVENING (7pm to 9pm)

* Public debate in the main auditorium of Mário de Andrade Library

8.3. How was it documented?

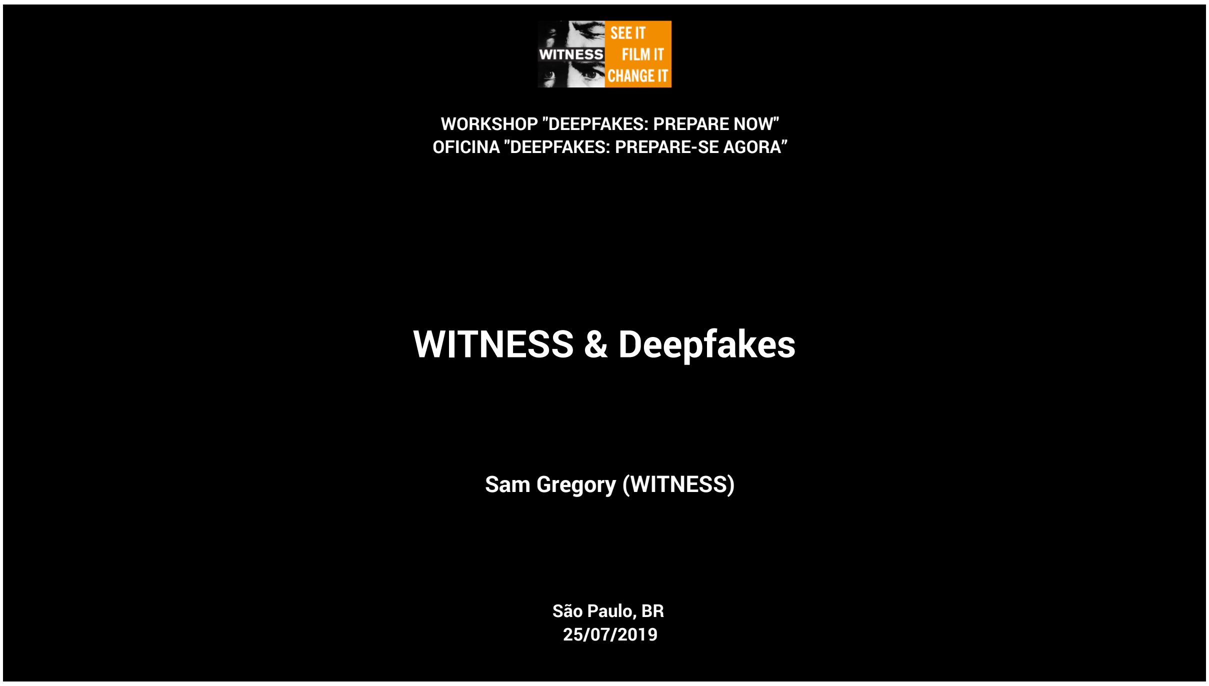
The meeting was documented in video, photos, audio and text and that material was used to create the present report. Here you can find the main ideas presented, created and discussed through-out that meeting, as well as video version of selected talks and materials presented. Please note that this is neither a chronological description of the presentation, nor a full transcription of content mentioned, but a synthesis of its main points by a person hired to document the meeting.

The text follows a “question and answer” structure (like a FAQ), so you can easily browse through the text and find the most relevant content for you. Questions were organised as much as possible following the chronological order presented by each of the speakers, although some of their ideas had its presentation changed inside the same block. All titles (questions) were created by this writer, as well as the footage edition (continuar).

The workshop was conducted under the Chatham House rules, and there was an option for "off-the record talks" as long participants asked. No participant asked for it, so we have identified their ideas and attributed credits for their interventions and work in this report.

9. ON WITNESS AND DEEPFAKES

This first section was presented by Sam Gregory, program director of WITNESS. Sam works with research and solutions development in technology, video, artificial intelligence (AI) and human rights. In WITNESS, he leads initiatives and trainings for civic activists and journalists aroundabout media manipulation using AI, as well as inter-disciplinary development of solutions to current challenges of deepfakes.


WITNESS
SEE IT
FILM IT
CHANGE IT

WORKSHOP "DEEPFAKES: PREPARE NOW"
OFICINA "DEEPFAKES: PREPARE-SE AGORA"

WITNESS & Deepfakes

Sam Gregory (WITNESS)

São Paulo, BR
25/07/2019

[click here to watch this video presentation with slides]

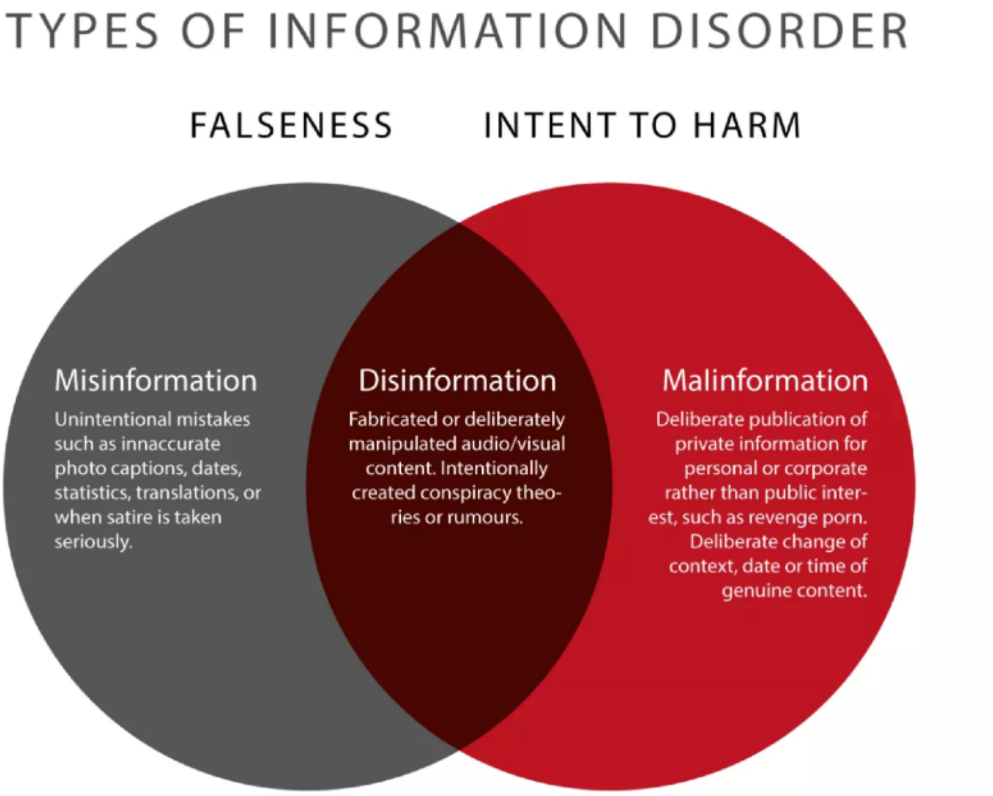
9.1. What is WITNESS?

WITNESS is a global human rights organisation that supports anyone, anywhere to use video and technology to protect and defend human rights. They are fundamentally concerned with how people use video and technology to create more trustworthy and reliable information. WITNESS works across the globe, with team-members in all continents. The organisation was founded 25 years ago after a police violence incident that was filmed by a bystander. Today, it works with different journalists and community activists from all over the world. It has a strong presence in Brazil via WITNESS Brasil (portugues.WITNESS.org).

In a nutshell, WITNESS tries to understand how people are using video and social media tools to share good information and show reality, for example documenting war crimes or police violence. It works close to local actors to learn from them, teach them how to create more trustworthy, compelling and effective information and narratives and how to share that knowledge with those that could use it for good. They create videos on how to document higher quality and relevant information that can be used in trials and in the search of justice and protection of rights (vae.witness.org). WITNESS also works in the infra-structure of technology, engaging with companies based on the experiences and needs of marginalized groups particularly in the Global South and advocating for better policies and products. As an example they advocated to YouTube around building a tool to enable the blurring of faces in videos, in response to threats to vulnerable human rights defenders. For more info, click here.

9.2. How are videos currently faked?

Fake videos, as so-called "fake news", usually do one of three things: misinform, disinform or malinform.


TYPES OF INFORMATION DISORDER
FALSENESS
INTENT TO HARM
Misinformation
Unintentional mistakes such as innaccurate photo captions, dates, statistics, translations, or when satire is taken seriously.
Disinformation
Fabricated or deliberately manipulated audio/visual content. Intentionally created conspiracy theories or rumours.
Malinformation
Deliberate publication of private information for personal or corporate rather than public interest, such as revenge porn. Deliberate change of context, date or time of genuine content.

Source: https://rm.coe.int/information-disorder-report-november-2017/1680764666

Deepfakes are still not used widely for any of these three purposes, but are likely to soon be. Fake videos today use other techniques, like decontextualisation or editing of an existing video, or staging a video.

- DECONTEXTUALIZING CONTEXTS: states that the video is from one place when it's originally from another.
- EDITING: do not show the entire story or has a few phrases or quotations used out completely out of context.

- MANIPULATION: either has their speed reduced or another form of technical manipulation.

- STAGING: videos that are fictional and created to deceive people into thinking they are real (not as common as people assume, but there are some examples)

- FIREHOSE OF FALSEHOOD: creation of multiple contradictory videos, social media and content until people don't know what to believe (as first pioneered by Russia in Ukraine).

9.3. Why is WITNESS interested in deepfakes?

Eighteen months ago WITNESS started hearing about deepfakes. Although they were still very rare, some people started seeing on deepfakes a potential source of an infocalypse or tecnopocalypse where all trust in images and words would collapse This raised WITNESS attention as they rapidly understood that this would be an important topic to be shared with its partners, and to prioritize proactive work on to ensure that potential harms were minimized.

Today we already start to feel a negative impact from deepfakes in relation to gender based violence (especially through deepfake porn videos, where it all started), but still not yet for disinformation - although that just seems a matter of time.

WITNESS decided to go global to discuss this topic and use this window of opportunity to better prepare before we find ourselves in the center of this tornado. Their mantra has been 'prepare, not panic'. In the last few years it has carried out meetings and workshops with experts, researchers, journalists, platforms policy people, among others from different parts of the world (among WITNESS allies and partners are significant global organisations such as the Partnership on AI and BBC).

9.4. What is WITNESS doing about deepfakes?

The rhetoric on deepfakes up to the beginning of the 2019 was overly focused in the US and Europe, as well as on the harms at a national level to politicians. Key to WITNESS is a commitment to including expertise and measures from a non-U.S./Western perspective, and a focus on listening to journalists, disinformation experts, human rights defenders, and vulnerable communities in the Global South – to avoid repeating the mistakes that were made in earlier responses to disinformation crises. So WITNESS decided to not only diversify the countries where they would carry out those meetings and discussions, but also actors invited: now not only the media and journalists, but also activists, fact-checkers and other key-actors.

Two months ago, WITNESS organised a meeting in Rio de Janeiro to carry out a first conversation with Brazilian community activists about their perception of risks of deepfakes. Below you can find some of the conclusions from that meeting about what can we do now about deepfakes and that were used as starting premises for this meeting agenda:

WHAT CAN WE DO NOW ABOUT DEEPFAKES?

- De-escalate the rhetoric
- Identify threat models and desired solutions from a global perspective and particularly highlight the experiences of already vulnerable and marginalized people
- Promote interdisciplinary approaches and multiple solutions
- Build on past experiences and communities, especially in media literacy, journalism and OSINT - and incorporate the discussion on deep fakes into the broader ‘fake news’ discussion
- Establish what we want from platforms and companies that sell those tools
- Strongly highlight the pros and cons of technical infrastructure choice, especially as they impact on people’s overall trust in videos and images.

Deepfakes e mídia sintética:
Prepare-se Agora
@samgregory
https://lab.witness.org/synthetic-media-and-deep-fakes/

[to download slides, click here.]

10. ON “DISINFORMATION”AND FACT-CHECKING IN BRAZIL

This section was presented by Sergio, a journalist with a background in journalistic business management and digital marketing. He is the editor of Comprova, a coalition of 24 organisations (mainly newspapers) created in 2018 to monitor news during Brazilian elections and carry out sort of “social debunking” by monitoring and verifying news produced by anonymous sources (not authorities, these were left for fact-checking agencies) and shared on social media platforms.

WITNESS SEE IT FILM IT CHANGE IT

WORKSHOP "DEEPFAKES: PREPARE NOW"
OFICINA "DEEPFAKES: PREPARE-SE AGORA"

Projeto Comprova e
o estado da desinformação no Brasil

Sérgio Lüdtke (Comprova)

São Paulo, BR
25/07/2019

[click here to watch this video presentation with slides]

10.1.What is the “disinformation” scenario in Brazil?

We live today an era in which opinion comes before information: when information arrives, one already has a formed opinion about the topic. And a lie is always sexier then the truth: it is much easier to produce disinformation as it allows you to create whatever your imagination and technology allows, while information, news, needs to be bounded to facts.

The political, social networks, journalism and trust scenarios are very important to understand where Brazil stands at the moment. Since the protests of 2013, groups started publishing a great number of items of “fake news” that had only its headline read by people and then shared. People started using more WhatsApp and sharing news in a more ”private way”. Furthermore, a lot of newspaper and other website created paywalls for their content and blocked its access for a great number of people. This combination, in a scenario of low trust in the media as its the case, creates a scenario prone for fake news, as these often come from people you trust (a friend, a family member) versus a news item from an organisation that you don’t really trust.

10.2.How does disinformation work in Brazil?

The process reached a peak during the last elections, in which "fake news", micro targeting and bots were widespread. This disinformation trends follow closely the public agenda and many public figures help to create those factoids, share disinformation and raise people's attention in the news. These disinformation usually come in waves, around a given fact (for instance, fake news and rumours about an election poll before its release).

In other to trick other people, fake videos usually have a very low quality, an audio that cannot be fully heard, etc. Not always all the content is fake; sometimes, different bits of true content are put together to construct a narrative that in its whole is fake and misleads people to think whatever they want them to think its true. As people want to believe that their beliefs and opinions are true, they are often fooled by these news even when they are clearly not true.

10.3.How can we fight disinformation?

Comprova created wide monitoring databases capable of identifying who are the agents behind these disinformation, what is in play with that content, who are those capable of disseminating that content, etc. The goal is to discover where its being created and check for this information in its early stages, before it reaches a point where it was already shared with thousands or millions of people.

People either shared this news with them via WhatsApp or they actively looked for in their search. Usually they had a team of journalists (around 5 or 6) working to debunk disinformation. It's complex work in which you need to always check and verify sources and communicate it in an easy-to-understand way to people.

Comprova also distributes releases with the results of their verification and tries to foster a culture of verification in newspapers and newsrooms. Currently, it works with the investigation of disinformation about public policies at the federal level.

A good analogy for their work is the FIFA VAR (Video-Assisted Referee). As with the FIFA VAR, people do understand the importance of Comprova’s work, but often finds it boring, think it makes things slower. And as the VAR, Comprova is trying to be quicker and more transparent about how it made its judgement of a veracity of a given news (how it made the check, where they search for the information, invite people to visit the original link and re-do their path). This may be the only way to make someone change their opinion.

10.4.What are some of the lessons learned by Comprova?

Raising awareness over an existing technology like deepfakes is crucial. People need to know that this technology exists in order to start questioning the truth of other videos received. For that, humor videos can have a good potential, as they can go viral and make a lot of people understand the concept behind it.

Some of the most important lessons of Comprova that should be thought about when developing strategies to deal with deepfake videos are:

- Just because a note or document is official doesn’t necessarily mean that it’s content is true (as they are partial).
- Reading comments left on news items is important as it helps them understand how meaningful that news is and what people are talking about it.
- You cannot bring to people the true version of a fake story without talking and sharing about the fake version. This can have some negative effects as in those cases where the fake news can become more popular after the social debunking than before and people still believe it's true because it matches their preconceived beliefs.
- In order to communicate with the masses, it's important to have allies in places other newspapers and in TV news that can reach an audience that the internet or a printed version of a newspaper simply can’t.
- Most people only read news headlines (that are often sensationalist and disinformative) or even when they read the full version, many times they can get its meaning wrong due to cognitive bias.

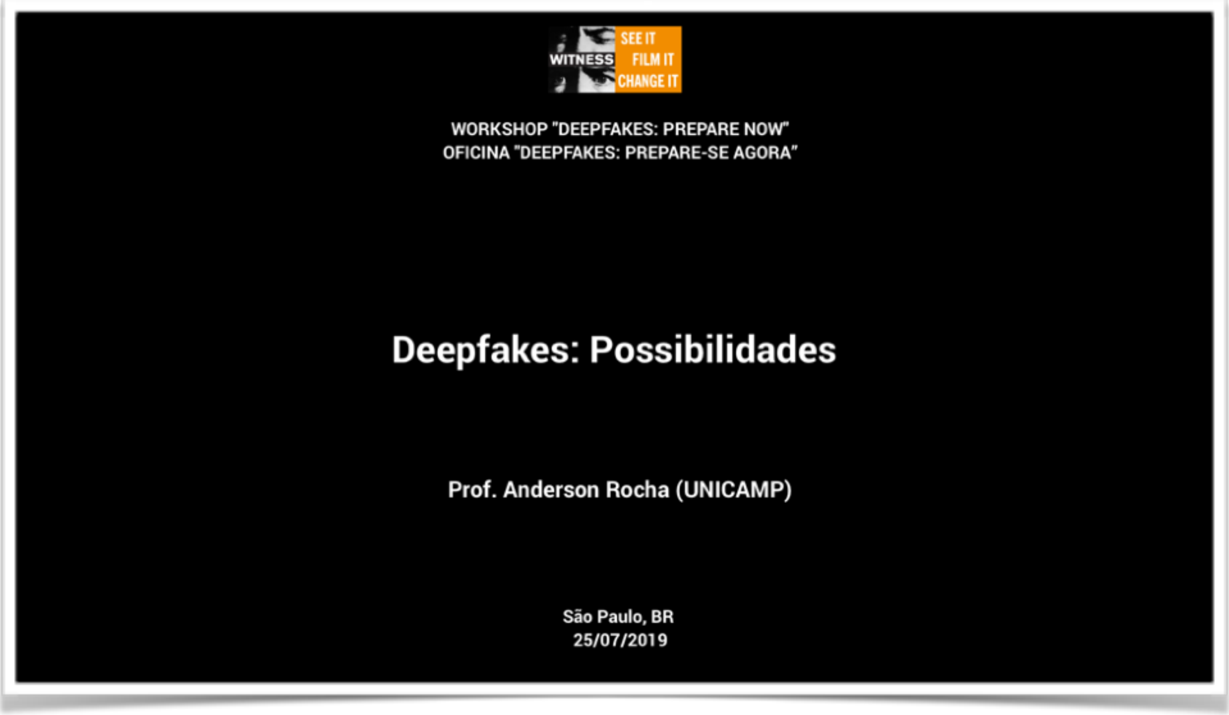
Projeto Comprova

Sérgio Lüdtke | 25 de julho de 2019

[to download slides, click here.]

11. DEEPFAKES: WHAT'S THAT?

This section was presented by Anderson Rocha, professor and Director of the Institute of Computing at the State University of Campinas in two different talks. His main areas of expertise are forensics computing, complex data inference and machine intelligence. He works as associated editor of important international journals and he is a member of the Brazilian Academy of Sciences.


WITNESS
SEE IT
FILM IT
CHANGE IT

WORKSHOP "DEEPFAKES: PREPARE NOW"
OFICINA "DEEPFAKES: PREPARE-SE AGORA"

Deepfakes: Possibilidades

Prof. Anderson Rocha (UNICAMP)

São Paulo, BR
25/07/2019

[click here to watch this video presentation with slides]

11.1.What is a deepfake?

A deepfake video is one in which you have a piece of an origin person (be it a face, a body, lips, etc) inserted or merged in a target person (the one that will have that part inserted on). However, a deepfake is more than a simple "head swap" (like those done in photos with Photoshop or videos with AfterFX): they use artificial intelligence (AI) in its creation process. That is why they are so new: the technology is currently being developed.

The most famous technique behind deepfakes today is a "head-swap", a process that has been done for centuries with statues or paintings, with photos and even with videos in some Hollywood films. Avatar (2009) had the same technology that is being discussed today. In Hollywood movies, sensors were put in people and its movements transformed into other characters. The difference is that now those tools are easier and cheaper to use.

11.2.Why is it "DEEP"?

The term "deepfake" has its origin in the name of a Reddit User that started, in early 2018, to post porn videos with the faces of famous actresses (now deleted by the user from the platform). There is nothing magic behind the word, although later on people started justifying the usage of "deep" because it uses deep neural networks. Today the "deep" is used to distinguish these fakes from fakes that are "shallow" (like shallow fakes, videos that only uses post-production tools, like After FX, or involves the miscontextualization of existing videos) or even "dumb" (dumbfakes, like memes and other content that is a grotesque or obvious fake).

11.3.How is it made?

It’s basically old, good math.To create a deepfake video, you need to go through a process of search, optimisation and distance calculation. You transform the element of the origin that you want to insert in the target (usually a face) and turn it into some sort of “play dough”. You detect several points in this face and create a 3D model. And then what you want to do is to distort those points to adjust the origin face into the target face. Its a mathematics process is which you can get the exact points of origin, the point of the target and do a math mapping diminishing the distance between them. And then they become closer and closer to each other and adjust. You can do that with faces, lips, bodies, etc.

11.4.Can you deepfake an audio?

Of course! To create a deep fake audio, the process is similar: you transform speech into micro segments (something that is not new). However now you break those into micro segments of intonation and desintonation and you have several micro pieces that you can combine with greater and greater precision to provide a really good sample of a target person

11.5.But how does it really work?

In order to create a deepfake, you have two different AI networks competing with each other: one that will try to generate the fake video and another that will try to detect whether its fake or not. They are adversaries. So the first one goes back and makes another creation. At one point in time, they will balance each other and the one that is trying to spot forgeries doesn't see any mistakes. That is why they are called Generative Adversarial Networks.

In this modelling process today one can make as many combinations as they want between those points: you can put the eye point up, down, generate another point between them, explode it. With good computing equipment you can generate a sample in a few hours. But the longer and the bigger the database you have, the better results you will get.

11.6.What does one need to make a deepfake?

In order to create a deepfake one needs a good amount of so-called ‘training data’, usually video footage from the target. That is why politicians and famous actors are the most obvious choices for experiments. The more data you have available online (be it pictures or videos), the easier it is for someone to create a deepfake using your image. You also need a graphic processing unit (GPU), as it allows for the parallels calculation that is necessary to create the deepfake. They are widely available at a reasonable cost at the market. And, of course, enough expertise to download and install the repositories or even use apps like Fakeapp, that don't always have a user-friendly interface.

11.7.What can you do with the technology behind deepfakes?

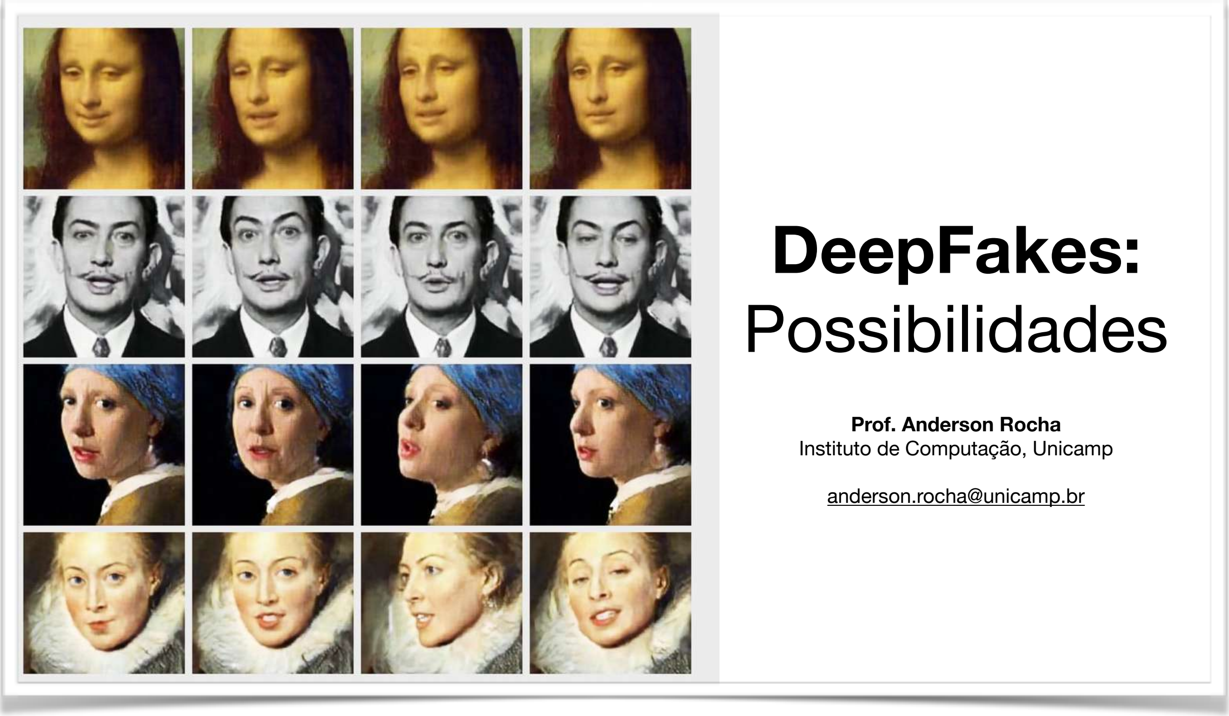
Face-swap is just the tip of the Deepfake iceberg. This technology brings different potential usages, especially in the movies industry, for videos reenactments, creating characters, changing voices in translations, in content based editing, etc. Some day you might be able to have your face and your friends playing in your favorite Netflix show. You can also use it to create apps to identify and classify content, among other usages.

11.8.Why should we worry about deepfakes?

Current threats are not on the technology side, but on disinformation. Although this might seem scary in the near future, a deepfake video today is still easy to be noticed, especially if you use technical tools for detection. The creation process of a deep fake still leaves clues that can be spotted and are easily identifiable - even with bare eyes: lights, pose, shadows, things you cannot do perfectly with the current technology. (NOTE from editors: This is not necessarily as true of some smaller deepfake modifications such as lip movements, rather than the whole face).

The problem however is not the technology itself but the integration into disinformation processes. It's the same as in "Brave New World": with a hose of false content, you don't know what to believe. This can be extremely difficult to deal with in this time of "post-truth" in which people seem to only believe in those news and content that are aligned with their prior beliefs. Widespread fake content creates confusion in people and can create oblivion, a situation in which you cannot understand what is truth or not.

Fake news are severely affecting democracy through micro-targeting, with fake videos (usually porn) sent to blackmail and stop activists and journalists, among other techniques seen around the globe.



DeepFakes:
Possibilidades

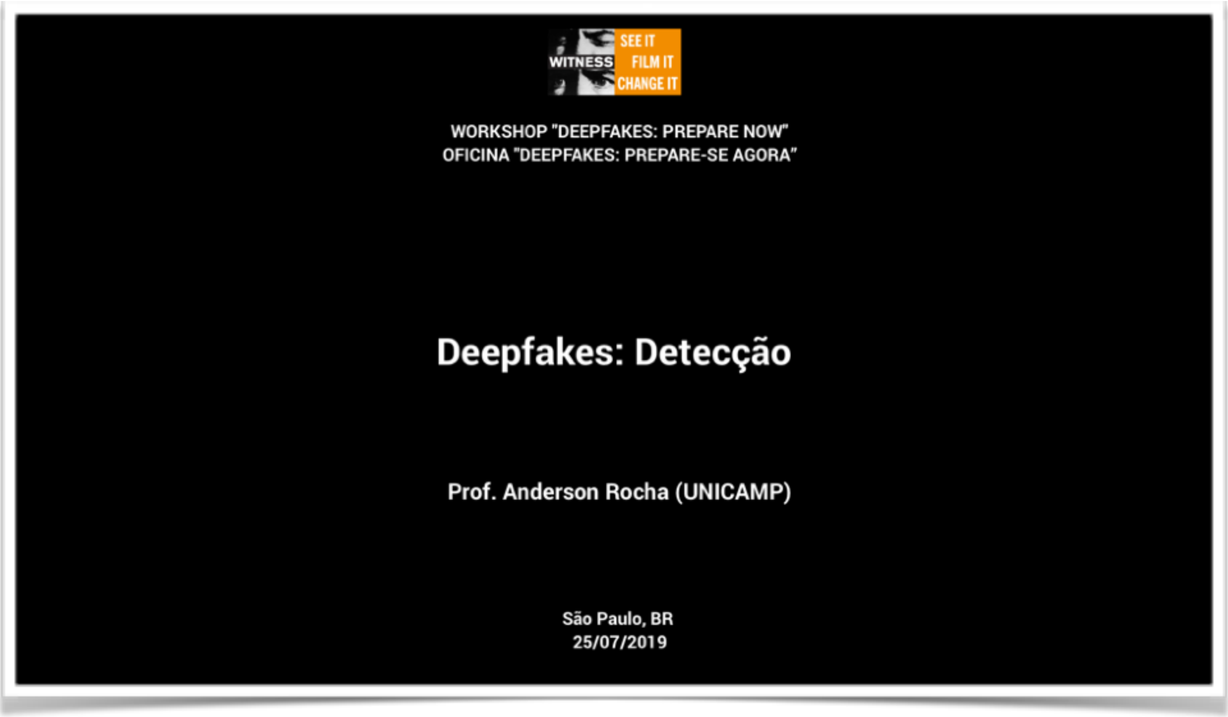
Prof. Anderson Rocha
Instituto de Computação, Unicamp

anderson.rocha@unicamp.br

[to download slides, click here.]

12. HOW TO DETECT A DEEPFAKE?

This section was also presented by Anderson Rocha (already described before).



WITNESS SEE IT FILM IT CHANGE IT

WORKSHOP "DEEPFAKES: PREPARE NOW"
OFICINA "DEEPFAKES: PREPARE-SE AGORA"

Deepfakes: Detecção

Prof. Anderson Rocha (UNICAMP)

São Paulo, BR
25/07/2019

[click here to watch this video presentation with slides]

12.1.How can we detect deepfakes?

To detect a deepfake experts can use the same technology used for its creation. They can use AI against AI by training the AI to detect what is false (instead of generating fake content, as in a deepfake generation process).

Current deepfakes also still face problems with eyes and mouth movements or hair texture, just to name a few. So instead of having to look to the whole video, you can inspect only certain regions (those more prone to imperfections) and look at it more carefully. They usually look for details and imperfections that would be harder to be improved overnight or in a short period of time. Usually, they try to analyse image properties that would be harder to be modified in a perfect way.

12.2.But how does deepfake detection work?

If the generation process consists of two neural networks (one receiving a source image and another generating the target content), the detection process requires training a neural network with two sorts of data (fake and true videos) and giving them clues and elements about what could have been falsified in the video. First, one trains an AI to generate deepfakes and another to detect. And then gets the improved detection AI to use it as a tool to detect deep fake videos using that technology (or other previous technologies learned).

12.3.What are some techniques to detect deepfakes?

Analysing lights: Current deepfake technology still doesn't have the same physical process of image capture. That means that its light distribution gives clues about video authenticity. If one gets a video of someone speaking in a sunny day and tries to swap faces with someone shot on a cloudy day, experts could easily debunk fakes by creating a map of illuminants (the profiles of the light sources) or by looking at the direction of light (light reflected in objects are different if they come from different materials). So if a deep fake didn't use an algorithm to match the lights, this can be a way to spot them.

Multiple phylogeny: One can create an AI to discover the chronological and derivative order of images or videos, so as to find what source images or video they were based on. This can work with photos, videos and even text, although it can sometimes fail (for example, when an author has very different styles, like Portuguese poet Fernando Pessoa with his heteronyms).

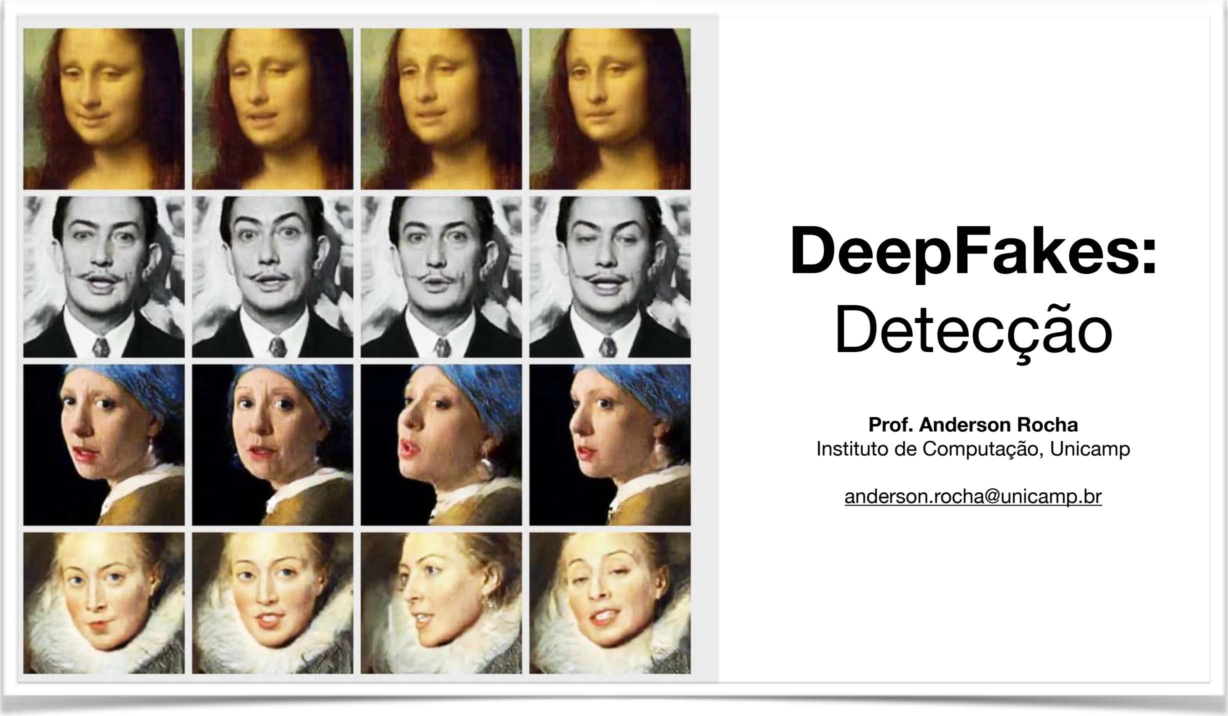
Make AI learn one's style: To detect a fake text, for instance, one can train an algorithm that learns one's writing style by checking usage of capitalizations, slangs, nouns, verb combination, articles, etc. You can define a stylometric profile for each author and detect what is fake or not. However, if someone can train another AI to generate to learn ones

styles, maybe the detector will not be able to figure it out (currently it can identify bots as those have a static writing style).

12.4.What are its limitations?

We are still not able to create a universal detector that would work with any algorithm and technology. Current detection technology requires examples of fakes in order to be able to detect them (they actually need two classes, real and fakes).To create a universal detector, we would have to understand the distribution of natural images, something that is still far from reality.

Therefore, each detector is only useful to detect the technologies they were created for. That can be a big problem in a scenario that technology used is rapidly improving. It’s a cat-and-mouse game and it's hard to say who will win simply by using technology.



DeepFakes:
Detecção

Prof. Anderson Rocha
Instituto de Computação, Unicamp

anderson.rocha@unicamp.br

[to download slides, click here.]

13. USING ART AND HUMOUR TO RAISE AWARENESS

This first part was presented by Marlus Araújo, a designer and coder interested in the convergence of art, design and technology and working with surveillance and technology.

WORKSHOP "DEEPFAKES: PREPARE NOW"
OFICINA "DEEPFAKES: PREPARE-SE AGORA"

Deepfakes & Arte

Marlus Araújo (Medialab UFRJ)

São Paulo, BR
25/07/2019

[video not available by author request]

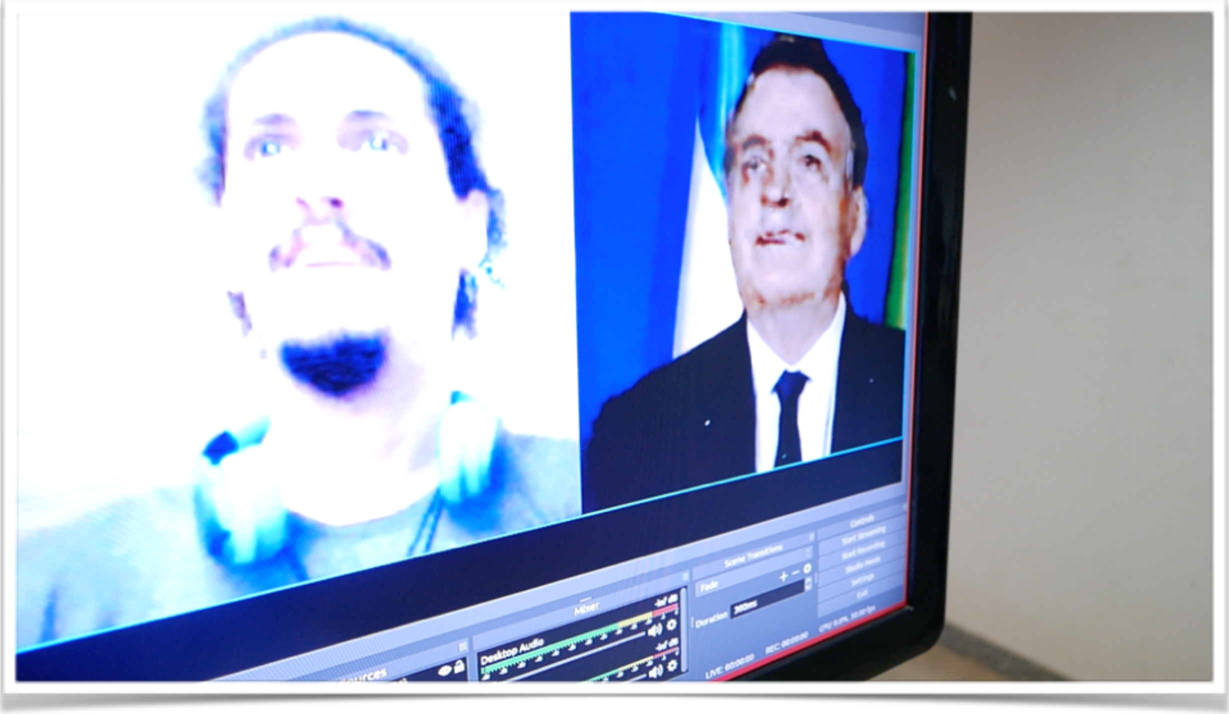
13.1.Why should we raise awareness on deepfakes?

Marlus argued that civil society should learn from their past mistakes and not adopt a strategy of not talking about deep fakes or their risks so as to not raise attention over them. He thinks it's the moment to make deepfakes more popular and known to people because if (or when) the avalanche of deepfake videos come then people will be aware that it is now easy to fake videos, swap faces, create fake dialogues, etc. Deepfake awareness is key for surviving a future avalanche of fake videos.

13.2.How can we raise awareness using art?

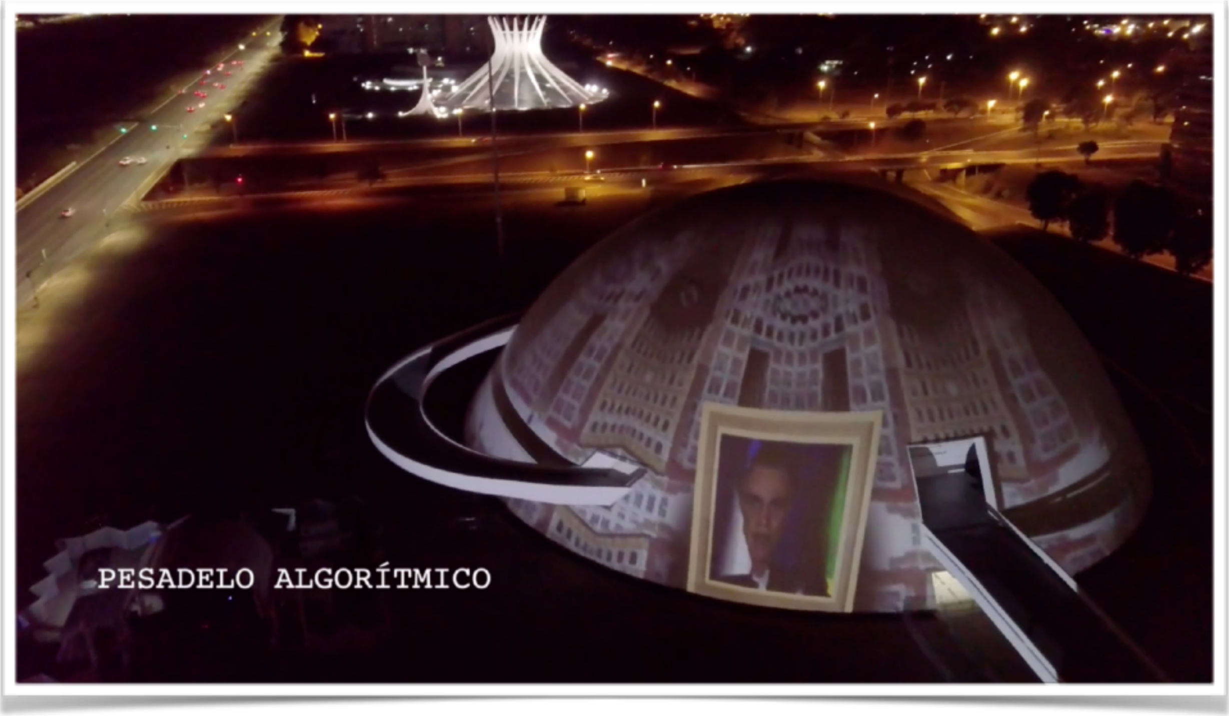
During his brief talk, he presented two different videos in which he wished to visually show people the potential risks and applications of deepfake technology. His projects mix video and art.

The first video shows an art installation he made in a Brazilian museum in which the public could control the Brazilian President's face live. In order to create it, he got a video from Bolsonaro that had him in different angles, making different movements, with one solid background color. He then cropped his face and made a 3 hours training with an AI machine. He then did the reverse process for himself: got his webcam that recorded with face and the computer could synthesize in real-time the image. The quality was still low and the fake grotesque, but it raised awareness over what can happen in the near future or during the next elections.



[to watch this video, click here]

The second video he made with a more complex AI training process: he not only cropped images, but also zooms in and outs, adding other data from different angles and positions, made more interactions and created a better fake. He brought this project to the Brasilia video mapping festival in the Brasilia National Museum. People could control Bolsonaro's face as projected on an outdoor scene and he later used that video to create a video that was seen as a protest video.



[to watch this video, click here]

13.3.How can deepfakes be used in humour?

This second part was presented by Bruno Sartori - a humorist from a midsize town in Brazil that discovered deep fakes right in the start and saw great potential in that technology for his own videos and he started studying it.

WORKSHOP "DEEPFAKES: PREPARE NOW"
OFICINA "DEEPFAKES: PREPARE-SE AGORA"

Deepfakes & Humour

Bruno Sartori (@brunnosarttori)

São Paulo, BR
25/07/2019

[click here to watch this video presentation with slides]

Bruno played four different deepfakes techniques/demos that he has already worked with in order to answer:

- Face-swap from politicians using doubles: Bruno has used a video with a professional comedian pretending to be President Bolsonaro, in which he inserted the actual Brazilian President's face. He considers this video one of his best deep fakes, especially because he put in a lot of work in After Effects to make it more realistic.

[to watch this video, click here]

- Video and voice to create memes: Bruno produces video memes that uses deepfake technology to make the image (face) and voice resemble to the target. The video shows a deepfake of the President singing a silly song that went viral in a YouTube video a few years back.

[to watch this video, click here]

- Live deepfakes of politicians in humor shows: Bruno produced a video that shows what can be done in real-time (live) with image and voice. It is a deepfake of former judge and current Minister Justice Sergio Moro for a famous Brazilian show. In the video, the show host makes a live online video-call with Moro's deepfake that is being controlled by the artist in real-time. It appears very realistic and despite its humorous content it is scary.

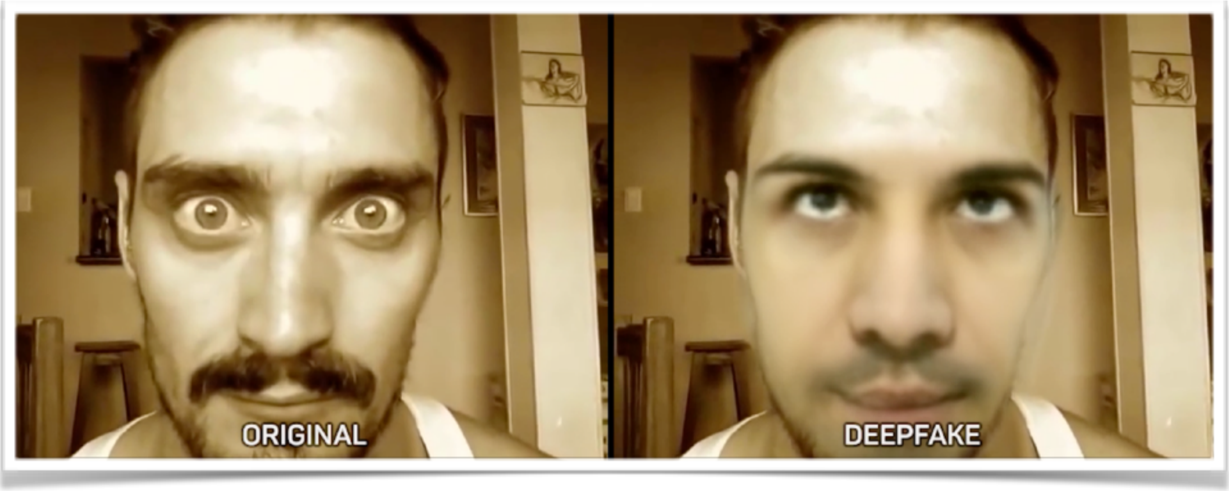
[to watch this video, click here]

13.4.How is this technology developing?

The evolution of deep fakes in this 18 months was incredible. As Bruno reports (see below), last year he would take around 30 days to have an acceptable face with a 64X64 pixel image pixels training. With the current technology (256X256 pixel pixels training), he uses around 4.000 images to create a face (that's the ideal number according to his test) and it takes around 4 days to have a face formed. Once this face is "trained", when he needs to insert it in other videos, it's quicker to have the final product - sometimes less than 2 hours. He also mentions that he has already seen some people using smartphone hardware to create the deepfakes and the results were very satisfactory.

13.5.What is the future of deepfakes?

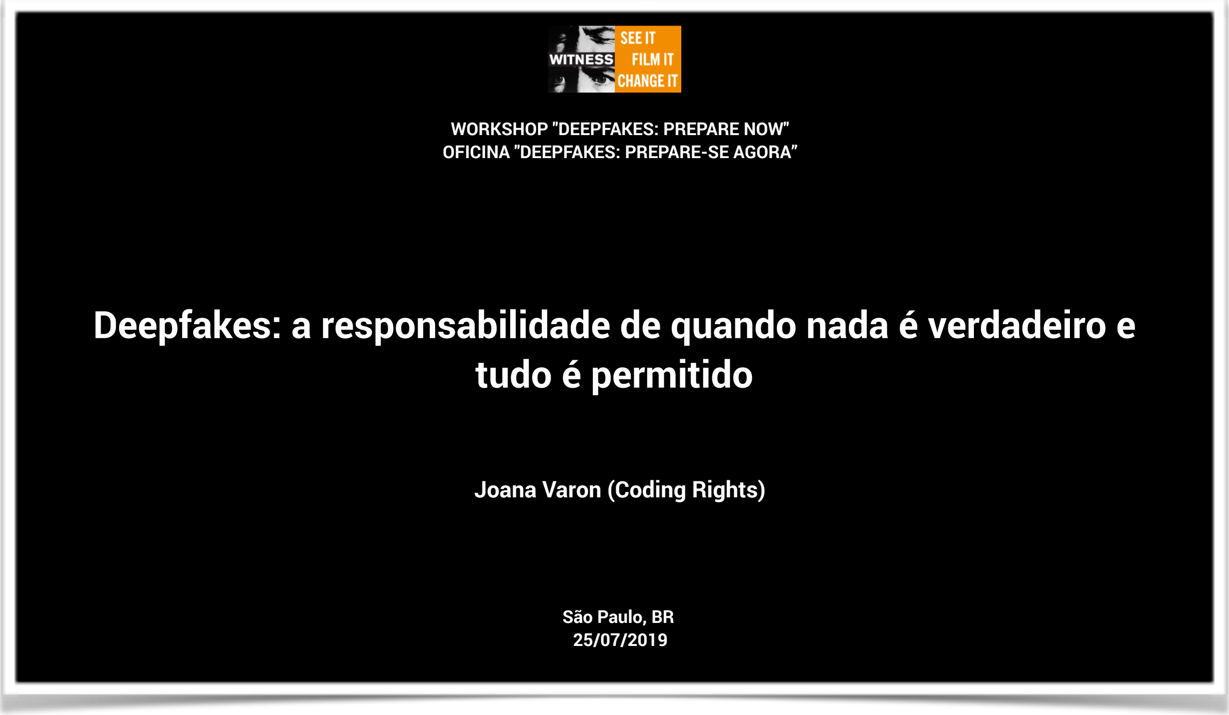
Deepfakes also open possibilities for you to act with your whole family in your favourite Netflix show. The last video shows one of his current works in which he tries to put his own face into a video by a very expressive actor



[to watch this video, click here]

14. DEEPFAKES, RIGHTS VIOLATIONS AND MINORITIES

This section was presented by Joana Varon. Joana is the director and founder of Coding Rights and Transfeminist Network of Digital Care. She is affiliated to the Berkman Klein Center for Internet and Society at Harvard University and has developed creative and advocacy projects that work at the intersection of law, art and technology.



WITNESS SEE IT FILM IT CHANGE IT

WORKSHOP "DEEPFAKES: PREPARE NOW"
OFICINA "DEEPFAKES: PREPARE-SE AGORA"

Deepfakes: a responsabilidade de quando nada é verdadeiro e tudo é permitido

Joana Varon (Coding Rights)

São Paulo, BR
25/07/2019

[click here to watch this video presentation with slides]

14.1.How did deepfakes reach the public?

Joana brought some cases to show that visual technology behind deep fakes were already being used in the cinema and television for quite some time. Her first memory of one is a Björk 1998 videoclip in which the singer turns into a bear. The years passed and in 2015, Instagram filters that allowed you to play with one's face, just like Björk.

In 2019, Snapchat released filters in which you could look older and even change your gender. There is a famous case of a man that took several photos of himself as a woman and created a Tinder profile to see how people would interact with him. In the virtual visual world, almost anything is possible now.

14.2.How do deepfakes perpetuate gender-based violence?

Now deepfakes arise and get popular with a sort of democratisation of this knowledge: what was before just discussed in academic circles, got its first application and recognition worldwide. And it was for porn. Deepfakes came to show that now we are "fucked" and can have a sex video of us publicised without even thinking about engaging in that practice. There are even apps, forums and repositories to help one create a sex deep fake of one's crush or their ex-girlfriend.

Not only deepfakes started as a way of producing degrading content of women, but also one of the first apps using the technology was an "X-Ray Glass" that could make any woman seem naked: yes, the app only worked for women. The app is currently banned although still available in unofficial websites.

14.3.How do deepfakes threaten people?

Although porn is still the main use of deepfakes today (a Google search of the term will result in pages with porn, but not yet fake news), there are already cases in which we see a mix of violence, deep fake and freedom of the press. In India, a journalist had a deep fake sex video of her made in order to silence her. In Brazil, the current governor of São Paulo had a fake video (it was more of a shallow-fake than a deep-fake) in a sex orgy shared on social media during elections. The damages of the video weren't enough to stop him being elected. However, what could have been the damages if the candidate was a woman?

14.4.Can deepfakes be used for good?

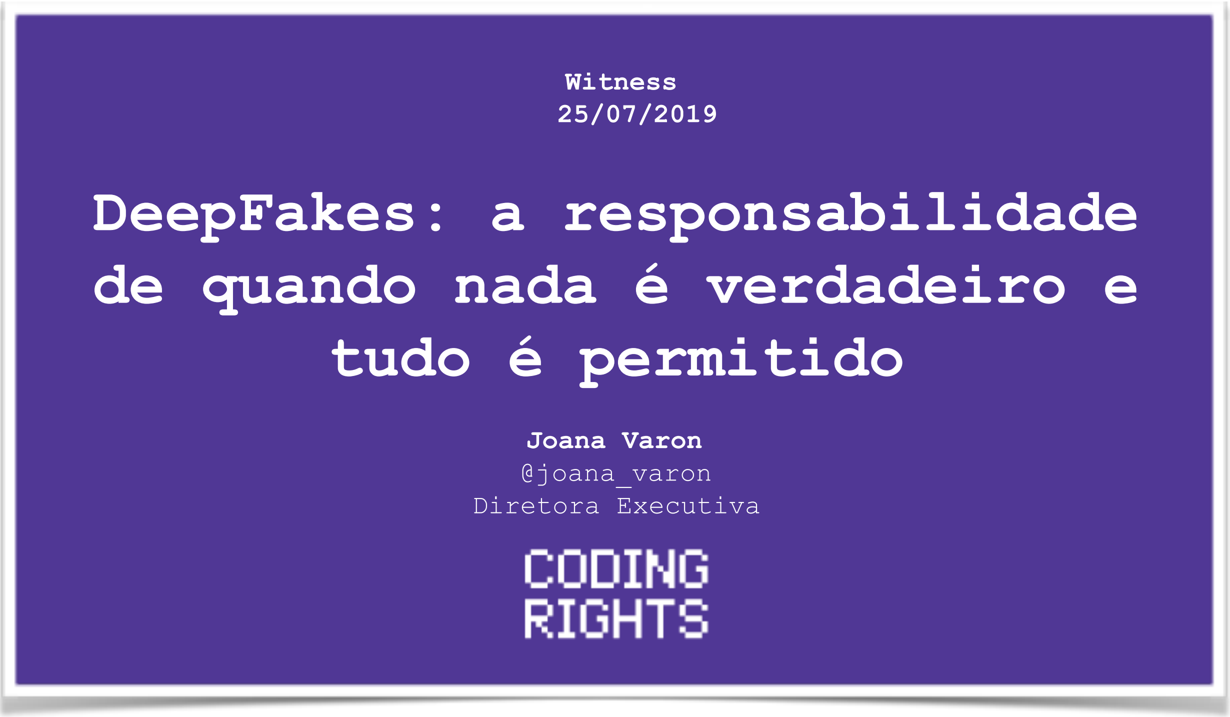
Despite those negative effects, deepfakes can also be used as a means of parody and gathering attention for a certain topic. As it can be used to enlarge gender asymmetries, it can also be used to reduce them. The technology is neutral, the usages that one give it will guide its effects. However, how can we ensure that?

14.5.What should be the roles of platforms?

In a world where most of what we consume come from platforms, what are their roles in making clear what content is fake or true? The case of Nancy Pelosi video that had its speed reduced for her to look drunk showed an unwillingness of social platforms to remove content of this nature. Should the same standards be applied to deepfakes? Limits are important to be established to ensure protection without severely compromising freedom.

14.6.How to fight deepfakes?

In Brazil, the legislation that could be used to regulate deepfakes would protect one against fake porn deepfakes, but not a video with fake content. As the regulation of the Marco Civil da Internet is from 2016, it was discussed and created in an era where the effects of disinformation and fake news was yet not clear. Should this legislation be amended? How much should we allow internet providers to moderate content before news get widespread? Should we criminalize all this conduct? These are some important questions that we should keep in mind.



Witness
25/07/2019

DeepFakes: a responsabilidade de quando nada é verdadeiro e tudo é permitido

Joana Varon
@joana_varon
Diretora Executiva

CODING RIGHTS

[to download slides, click here.]

15. IMPACTS ON ACTIVISTS AND POOR COMMUNITIES

This section was presented by Lana Souza, a journalist, resident of the Complexo do Alemão (one of Rio's biggest favelas complexes) and co-founder of Coletivo Papo Reto, an independent communications collective that acts in Rio's favelas and uses communication as a tool to reaffirm rights and dispute narratives about favelas. Lana was also a participant in an early discussion that WITNESS convened on this topic with community-based activists. She shared some of her impressions on the topic.

WORKSHOP "DEEPFAKES: PREPARE NOW"
OFICINA "DEEPFAKES: PREPARE-SE AGORA"

Deepfake: Impactos para ativistas

Lana Souza (Coletivo Papo Reto)

São Paulo, BR
25/07/2019

[click here to watch this video presentation with slides]

15.1.How can deepfakes impact favela residents and activists?

Earlier in 2019, WITNESS conducted a deepfake workshop in Rio, with community activists to better understand their impressions of this technology. Fear, anguish and worries were some of the main feelings at that meeting as they know they will once more suffer violence with this technology, even if it still takes some time (today they are still fighting against much more basic fake news).

In the favelas, activists and residents do not really care about the technical issues of deep fakes. What they care about is getting back home safe and alive and on their reputation. In a world where you can have a fake video of you created easily, this content can be used to put them in difficulty with their community drug-dealers, police and other actors that could even kill them for that. Not being recognised in its own territory gives them less access to the people and in protecting their rights.

15.2.Has anyone from those communities suffered due to “fake news”

Raull, an activist from Rio, had a big fight with Facebook to verify his profile as a lot of fake profiles were trying to damage his reputation inside his own territory by sharing fake news under his name. Sometimes this news are not entirely fake, but they are decontextualized. Sometimes, the content is just fake and then they have to work hard to rebuke it.

15.3.How could we raise awareness about deepfakes in favelas?

To communicate what is a deepfake to people in the favelas can be tricky. The term "fake news" was already complex for them to understand. Using humor and art can be a way to introduce those contents, as well as creating products that “speak their language”. Constantly assessing the impact of ones’ intervention on the least privileged territories, and what they are contributing to creating in terms of political and public safety narratives is also important, as well as ensuring the inclusion of those spaces in discussions.

16. FIGHTING THE DEEPFAKE PROBLEM

The workshop also intended participants to actively contribute to the topic, recognizing the importance of prioritizing the situation in Brazil both in terms of threats from deepfakes and the solutions that are being developed in response to malicious deepfakes. Two exercises were proposed: one on threats that participants prioritized as important or significant and another on solutions that they prioritized as journalists, civic activists, technologists and stake-holders from other sectors in Brazil. But before that, Sam gave an intro to the problem.

WORKSHOP "DEEPFAKES: PREPARE NOW"
OFICINA "DEEPFAKES: PREPARE-SE AGORA"

Deepfake: Áreas de soluções

Sam Gregory (WITNESS)

São Paulo, BR
25/07/2019

[click here to watch this video presentation with slides]

16.1.What we shouldn't do to fight the deepfake problem?

Many people say that spotting a deep fake is easy because they don't blink. This is a conclusion made based on widely publicized piece of research published a year ago. However, two weeks after publishing the paper, its author got sent a deep fake video that blinked as people worked hard to develop an approach that would eliminate this weakness in the algorithm and the training data. Nevertheless, (not) blinking is still cited by many as this info got stuck in their minds. This brings us an important lesson: we should not focus on the algorithmic "Achilles heel" of the moment, because technology develops quick and what is true now might not be in 6 months - but people can still remember it as being true and this can mislead them to believe a fake video is actually true. This is particularly relevant with deepfakes, which are developed via adversarial processes that are inherently structured around a process of competing to improve the quality of the fakery.

16.2.What should we do to fight the deepfake problem?

More than talking about the current algorithmic"Achilles heel" of any current deepfake creation process, we should work to create a critical thinking that can make people doubt materials, check sources and provenance and corroboration and look for its veracity before believing and sharing it. The role of Youtubers and social media influencers (especially in Brazil) can be key for this education and "popularization" of the topic. Furthermore, humor is an important tool for it. Working together with other existing projects, initiatives, coalitions and fact-checking agencies is very important not only to share tools and skills, but also exchange experiences and new technologies. Other methods include creating invisible (to humans) interventions in images - so-called 'adversarial perturbations' - in order to trick the computer vision process and avoid having your data used for AI training purposes.

16.3.What should we discuss before creating a solution?

Blockchain technology combined with other mechanisms to authenticate the source through so-called 'verified capture' are often cited as potential "solutions" for easily identifying a deepfake by tracking the source of images and videos to see if a manipulation has occurred. However, such a fundamental shift in how we assess the veracity of video poses important questions about who should we trust for that, which tools should be allowed or required to use, and what are the risks of this surveillance, of data access, of privacy. The implications of these technical infrastructure questions will have profound implications for how we think about trust issues in our society. WITNESS has an upcoming report on this area.

Platforms, closed messaging apps, search engines, social media networks and video-sharing sites will also be the places where these manipulations are shared. Some topics and questions we should discuss are: What role should social networks and other platforms play in fighting deep fakes? What should be the limits? How should they provide access to detection capabilities? How should they signal to users that content is true or fake or in some way manipulated in ways they cannot see? Should they remove certain kinds of manipulated content and under what criteria? There are a lot of questions in this area that should be addressed before reaching any final conclusions on how to deal with this topic.

16.4.In summary what should we do now?

Key perspectives suggested in the presentation:

- De-escalate the rhetoric around deepfakes, and search for solutions
- Identify threat models and solutions from a global perspective and based on the experiences of people who have already faced threats from mis/disinformation
- Promote interdisciplinary approaches and multiple solutions
- Build on past experiences and existing expertise, particularly from media literacy, open source investigation and journalism
- Determine what we want and don't want from platforms and companies that commercialize creation tools or manage the flow of synthetic media

- Highlight clearly the pros and cons of tech infrastructure choices, especially those that will impact on public trust in images and video

For more details on potential solutions including these areas below, see the document below:

1.Can we teach people to spot these?

2.How do we build on existing journalistic capacity and coordination

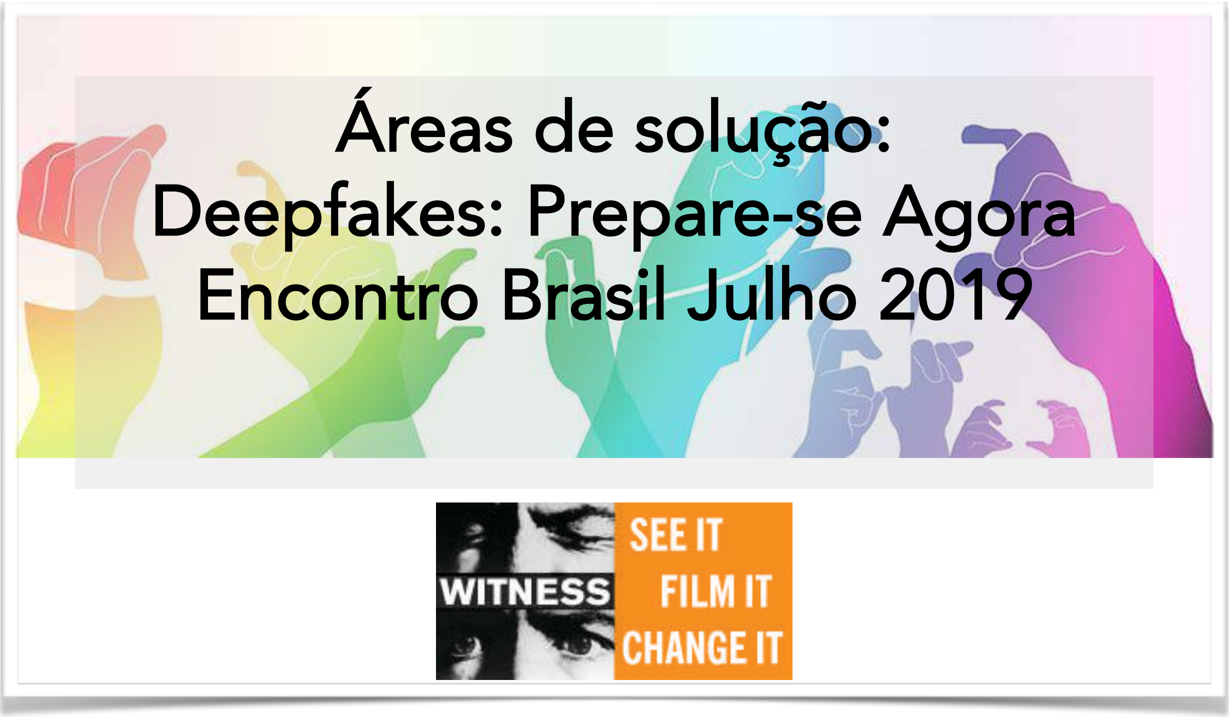
3.Are there tools for detection? (and who has access?)

4.Are there tools for authentication? (and who's excluded)

5.Are there tools for hiding our images from being used as training data?

6.What do we want from commercial companies producing synthesis tools?

7.What should platforms and lawmakers do?



Áreas de solução:
Deepfakes: Prepare-se Agora
Encontro Brasil Julho 2019

WITNESS SEE IT FILM IT CHANGE IT

[to download slides, click here. Or in English only: https://blog.witness.org/2019/06/deepfakes-synthetic-media-updated-survey-solutions-malicious-usages/]

17. RELEVANT THREATS AND POTENTIAL SOLUTIONS

WITNESS has mapped a series of perceived threats across a series of workshops they have held. The first part of the exercise consisted in a dot-sticker voting over these existing threats pasted on the wall. Potential threats were grouped according to the type of problems, like "attacks on journalists and activists" or "attacks that are driving towards a zero trust society, among others.

17.1.What are some of the most relevant threats?

Participants identified the following as the threats that were most relevant to them (voted by participants from fixed range of possibilities):

- Reputation and credibility attacks against journalists and civic activists with enhanced attacks (often on the basis of gender) and more sophisticated falsifications

- Non-consensual AI-generated sexual imagery and gender-based violence against public figures and for bullying

- Attacks on social movements' and movements leaders' credibility/safety and public narratives as well as attacks on marginalized groups via false allegations of behaviour/actions

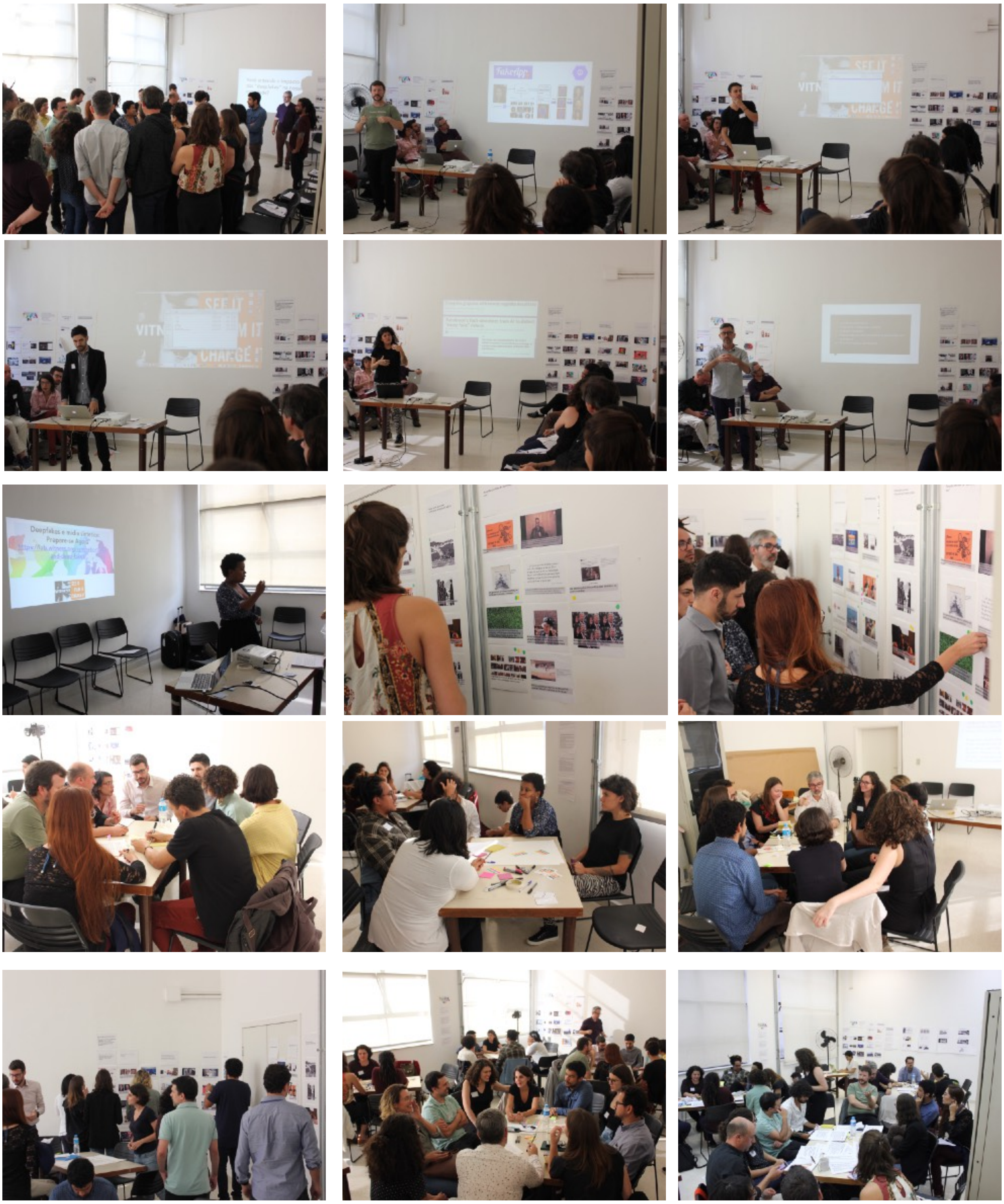
- Attacks against judicial processes and the evidentiary value of video as video is discredited or processes are overwhelmed by burden of proving true from false

- Integration into conspiracy campaigns

- Cumulative creation of distrust in institutions

- Micro-targeting of increasingly customized AI-generated content, including by political figures

- A firehose of falsehood: When you have a lot of fake content, you cannot know what is true and both societal systems, media verification processes and search engines are overwhelmed.



17.2.What are some of the most relevant local threats?

Participants were then invited to divide themselves into 4 different working groups and select one rapporteur for each.

- Grassroots and community-based groups (1)

- Media organizations and fact-checkers (2)

- NGOs/human rights defenders (3)

- Technology experts (4)

They then discussed in groups potential threats in their areas and share a few of them with all participants (and also pasting it to the wall next to a similar threat). Main questions were:

- Where do new forms of manipulation expand existing threats, introduce new threats, alter existing threats, reinforce other threats?

- How could we be prepared for it?

Below you can find all threats organised by
topic, mentioning which group reported each. Each topic concerns a different interest:

A. Threats to personal safety:

- False evidence: Imprisonment of local leaders based on fake evidence from deepfake videos. (1)

- Threats to activists, journalists and fact-checkers' work:

- Civic activist credibility: Loss of activists' credibility due to fake videos or content, as that can severely damage community work and mobilization and even their personal safety. (1)

- Video as evidence: The value of video as evidence will be compromised. (1)

- Media credibility: Risk of losing press and media credibility if videos cannot be used for evidence anymore (for assessing the truth and reporting). (1, 2)

- Media overwhelmed: Mass-distribution of fake videos to flood newspapers and fact-checking agencies with content they have to verify and make people unable to identify what its true or not. (2)

- Manipulation of archive images: Using deepfake videos based on past events to implicate people in current events (e.g. protestors. (3, 4)

B. Difficulties for detection:

- Speed: There will be difficulties for detection and blocking fakenews, dodgy rumours or claimed "breaking news" with deepfake technologies (especially if this detection needs to occur in real time). (4)

- Access: Fake news and manipulated videos are often distributed in closed networks and in micro-targeted messages that fact-checkers don't have access to. (4)

- Understanding of how to detect: Many people are technology laypeople and cannot perform any complex checking procedures; they simply believe the content without questioning and spread them easily. (1)

- Lack of understanding of what's possible: People can be more easily fooled by fake videos as they still don't know if this technology can be manipulated (unlike with photos where most are familiar with the idea that they can be manipulated). (3,4)

- Lack of resources: Detection processes are costly as its difficult to train and finance teams to use tools to carry out those checks (and journalists are usually not IT experts). (2)

- Use of audio: There's limited discussion of the use of audio, that can be highly effective and harder to detect

C. Mal-usages of deepfakes at a broader scale:

- Contribution to existing 'zero confidence society' problems: Deepfakes could be a powerful tool of disinformation in a "zero confidence society" where truth is replaced by opinion. (3)

- Complements micro-targeting: Risks associated with the technology behind deep fake goes beyond fake content creation: with AI you can discover a person or a group's psychological profile and use this to carry-out a very effective targeting with fake content in order to reinforce an existing position or opinion they hold. (4)

D. Who's at risk from deepfakes:

- Personal image data security: Increased risks that people will "steal" and use personal data to improve AIs in order to create manipulated videos of individuals. (3)

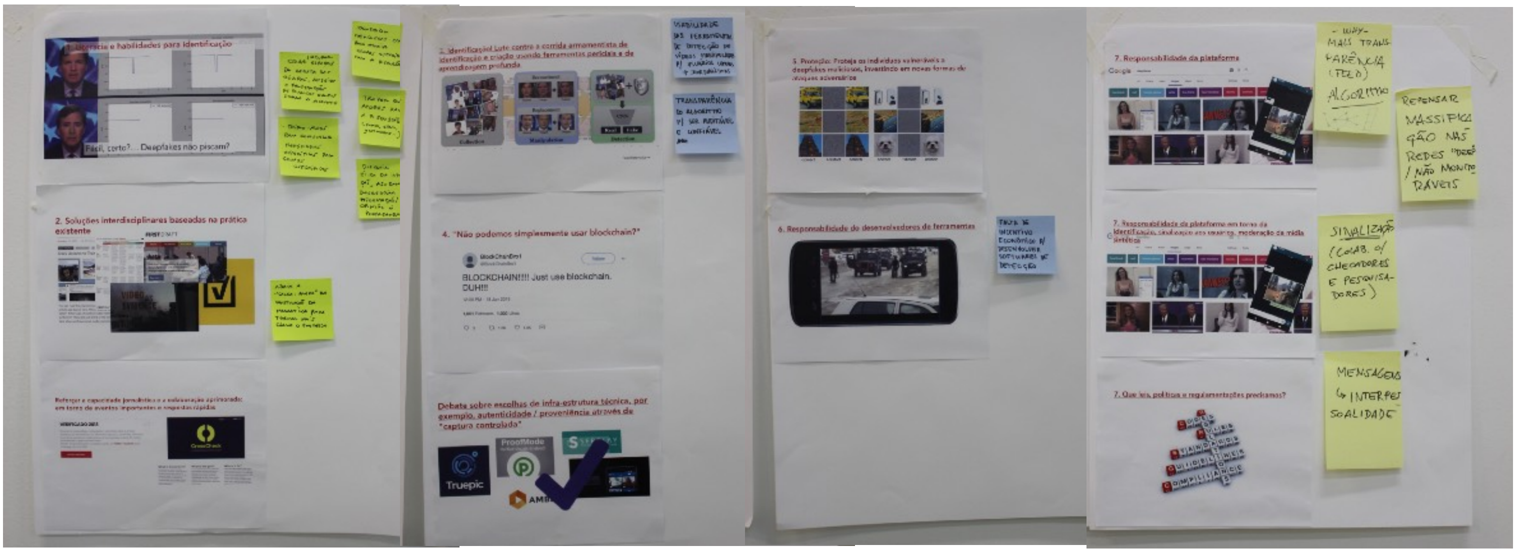
17.3. What are some of the potential solutions?

For the solutions exercise, participants were divided into 3 groups. Each group had one or two guiding questions:

- Educating the public/Media literacies
- 'Necessary collaboration/coordination' + 'needed journalistic skills and tools' + 'what is required from detection and authentication solutions'
- What should the platforms do?

Participants then discussed potential solutions for each question in their group and then sharing it out loud the most interesting solutions. Main guiding questions were:

- What solutions feel most relevant to you? What would you need from those solutions to make them viable? What would be ways to approach this in Brazil? What worries you about these solutions?
- What would be a concrete next step in this area that you'd like to see?


1. Literacia e habilidades para identificação
Fácil, certo?... Deepfakes não piscam?
2. Soluções interdisciplinares baseadas na prática existente
Reforçar a capacidade jornalística e a colaboração aprimorada: em torno de eventos importantes e respostas rápidas
3. Identificação! Lute contra a corrida armamentista de identificação e criação usando ferramentas periciais e de aprendizagem profunda
4. "Não podemos simplesmente usar blockchain?"
BLOCKCHAIN!!!! Just use blockchain. DUH!!!
Debate sobre escolhas de infra-estrutura técnica, por exemplo, autenticidade / proveniência através de "captura controlada"
Truepic
ProofMode
TRANSPARÊNCIA
5. Proteção: Proteja os indivíduos vulneráveis a deepfakes maliciosos, investindo em novas formas de ataques adversários
6. Responsabilidade do desenvolvedores de ferramentas
7. Responsabilidade da plataforma
7. Responsabilidade da plataforma em torno da identificação, sinalização aos usuários, moderação da mídia sintética
7. Que leis, políticas e regulamentações precisamos?
RULES
STANDARDS
GUIDELINES
COMPLIANCE
- WHY-
MAIS TRANSPARÊNCIA
(FEED)
ALGORITMO
REPENSAR MASSIFICAÇÃO NAS REDES / NÃO MONETIZÁVEIS
SINALIZAÇÃO
(COLAB. c/ CHECADORES E PESQUISADORES)
MENSAGENS
INTERPESSOALIDADE

A. How to share knowledge about deepfakes and related mis/disinformation?

> Teach information ethics: Discuss the ethics of information to help people differentiate what is information from opinion or propaganda. (1)

> Show how narratives are used to deceive: Open the narrative building black box: show how people and groups build narratives to manipulate public opinion and how micro-

targeting is currently being used for that, and the accompanying their strategies and formulas. (1)

B. How to improve communication, particularly with grassroots communities?

> Create “hearing spaces”: Occupy and create "hearing spaces", spacs to listen to what people want to know or already know about deepfakes. (1)

> Draw on other influencers: Bring other actors to this discussion by using the girls from Slam or Funk ou YouTubers and not leaving all responsibility for journalists and communicators. (1)

> Develop messaging: Identify key messages to attract specific groups into the discussion. (1)

> Micro-target groups: Use micro-targeting to communicate specific messages to specific groups: for instance, if we want to talk with teenagers, we need to use a certain type of language and channels. (1)

C. How to create and disseminate tools?

> Build credible and explainable tools and procedures: Build detection tools that are that are clear, transparent and trustworthy, so that people can understand all steps of the verification method and audit their results, so making those tools more credible. (2)

> Develop tools that are accessible at all economic levels: Put together tools for journalists and citizens to verify authenticity of deepfake videos (despite the lack of economic incentives to make these tools available). (2)

> Challenge economic incentives that are to build for creation not for detection (2)

> Track deepfakes and dubious news early: Using engagement metrics dashboards (like CrowdCompass) to spot dubious news and check them before before they become viral. (3)

D. What should platforms do?

> Make what is in your feed/why more transparent: Having a more transparent feed in social networks that can be easier to spot fake news distribution. (3)

> Rethink closed messaging: Rethink how reaching many people should work in private non-monitored networks like Whatsapp, such as discussing the cap on the number of participants in groups, the role of bots, etc. (3)

18. TO SUM-UP

The meeting was very productive and participants understood very well how deepfakes are made, how they can be detected, and how we can raise awareness about them so as to prepare for a potential use of the technology for deepfake news in the near future.

THREATS highlighted included: Threats to personal safety, like imprisonment of local leaders; to the work of activists, journalists and fact-checkers, especially due to credibility loss; increasing difficulties for detection in a cat-and-mouse game and in a context where people have very literal digital literacy; mal-usages of deep fakes to create fake news or a scenario in which you cannot identify what is true or not, as well as risks associated with related technology that goes beyond the video (like usage as micro targeted content).

SOLUTIONS proposed revolved in sharing knowledge particularly with grassroots communities; improving communication between actors and to society; creating and disseminating accessible and explainable tools, and pushing for change in social networks platforms feed and approach to dissemination in closed messaging.

19. CREDITS

This report was created by Bruno Paschoal, with the help of Sam Gregory.