



DEEPPFAKES: PREPARA-SE AGORA

RELATÓRIO DA OFICINA (PORTUGUÊS)

BRASIL, JULHO 2019

DEEPFAKES: PREPARA-SE AGORA

Sumário executivo da 1ª reunião de nível nacional no Brasil

+ Principais ameaças e soluções preferenciais

<https://lab.witness.org/brasil-deepfakes-prepara-se-agora/>

'Deepfakes' e a mídia sintética são novas formas de manipulação audiovisual que permitem que as pessoas criem simulações realistas do rosto, de voz ou de ações de outra pessoa. Essas novas formas permitem, de forma fraudulenta, criar uma ilusão de uma pessoa falando ou fazendo algo imoral. Os deepfakes estão mais fáceis de serem produzidos, exigindo poucas imagens para serem construídos e estão sendo, por sua vez, mais comercializados.

Atualmente, os deepfakes frequentemente impactam mulheres, em relação à produção de imagens e vídeos sexuais, manipulados, e não consensuais. Ao mesmo tempo, há temores que essa manipulação audiovisual possa ameaçar várias outras áreas de nossa sociedade. Novas soluções estão sendo propostas para lidar com o uso indevido dessas novas ferramentas, e é fundamental que essa discussão tenha uma perspectiva global, e não de um ponto de vista centrado apenas nos EUA ou na Europa.

A primeira reunião nacional em preparação para deepfakes

No dia 25 de julho de 2019, a WITNESS realizou uma convocação chamada "*Deepfakes e Mídia Sintética Prepare-se Agora*" em São Paulo, Brasil. Em nosso entendimento, foi a primeira discussão multidisciplinar a nível nacional sobre como entender e se preparar pragmaticamente para a ameaça potencial de deepfakes.

O objetivo do encontro foi explorar e priorizar soluções pragmáticas para a prevenção e a defesa contra um futuro de vídeo e áudio alterado por novas técnicas de inteligência artificial (IA), com foco particular nas ameaças identificadas no Brasil e nas soluções desejadas por uma série de grupos interessados. Os participantes do workshop incluíram jornalistas, verificadores de fatos, tecnólogos, ativistas cívicos e outros grupos. Este workshop fez parte de uma iniciativa da WITNESS focada em como proteger e defender vozes marginais, jornalismo cívico e direitos humanos, à medida que tecnologias emergentes como a IA se cruzam com desinformação, manipulação da mídia e crescente autoritarismo. O workshop também contou com o apoio da equipe da WITNESS Brasil. Mais informações sobre o trabalho dos deepfakes da WITNESS, incluindo relatórios de convocações e workshops anteriores, estão disponíveis em inglês em: wit.to/Synthetic-Media-deepfakes

O workshop foi estruturado para ensinar sobre as tecnologias usadas na criação e detecção de mídia sintética; seu uso atual em ataques a mulheres, em violência de gênero e em críticas e sátiras culturais. Em seguida, os participantes colocaram isso no contexto de

desafios existentes sobre a desinformação no Brasil e se concentraram em priorizar ameaças e propostas soluções.

Quais são as ameaças?

Depois que aprenderam sobre as possibilidades tecnológicas dos deepfakes e discutiram sobre a situação atual no Brasil, os participantes priorizaram essas ameaças-chave como áreas em que novas formas de manipulação podem expandir se ameaças existentes continuarem ou se novas forem introduzidas. Por exemplo:

- Jornalistas e ativistas cívicos terão sua reputação e credibilidade atacadas. Isso representa outras preocupações mundiais.
- Figuras públicas enfrentarão imagens sexuais não consensuais e violência de gênero.
- O movimento social enfrentará ataques à credibilidade e segurança de seus líderes e de suas narrativas públicas.
- Haverá ataques contra processos judiciais e o valor probatório do vídeo, tanto para notícias quanto para evidências, pois o vídeo será desacreditado, alegado como não autêntico, e os processos terão um ônus a mais: provar o que de fato e verdadeiro.
- deepfakes será mais uma arma que contribuirá para campanhas de conspiração.
- À medida que os deepfakes se tornam mais comuns e fáceis de serem produzidos em massa, eles tendem a contribuir para uma profusão de falsidades que inundam as agências de verificação de mídia e verificação de fatos, uma vez que estas por ter manipulado o conteúdo quais elas precisam verificar.
- Formas semelhantes de falsidade contribuem para a criação cumulativa de desconfiança nas instituições de uma sociedade com 'confiança zero', onde a verdade é substituída pela opinião.
- A micro segmentação de conteúdo, cada vez mais personalizado pela IA, usará o perfil psicológico de uma pessoa ou de um grupo para realizar uma segmentação muito eficaz com conteúdo falsificado, com a intenção de reforçar uma posição ou opinião existente.

Quais são as soluções que precisamos?

Os participantes discutiram uma série de soluções que estão sendo propostas a nível mundial, mas que são geralmente lideradas pelo Vale do Silício e as ações legislativas em Washington DC e Bruxelas. Isso inclui as seguintes áreas:

1. Podemos ensinar as pessoas como identificar deepfakes?
2. Como construímos a capacidade e coordenação jornalísticas existentes?
3. Existem ferramentas para detecção? (e quem tem acesso?)
4. Existem ferramentas para autenticação? (e quem foi excluído?)
5. Existem ferramentas para ocultar nossas imagens de serem usadas como dados de treinamento?
6. O que queremos das empresas comerciais que produzem ferramentas de síntese?
7. O que os legisladores e as plataformas de mídia social devem fazer?

Após uma discussão sobre o status dos esforços de detecção, a inadequação de treinamentos sobre detecção de deepfakes e os esforços atuais do Facebook nessa luta, os participantes se concentraram nas soluções mais relevantes para o Brasil.

Precisamos de alfabetização midiática contextualizada nos problemas de desinformação, especialmente para comunidades de base.

Em vez de falar sobre o atual algorítmico "calcanhar de Aquiles" de qualquer processo atual de criação de deepfake - que geralmente é uma falha técnica que desaparece à medida que as técnicas são aprimoradas - devemos trabalhar para criar um pensamento crítico que possa fazer as pessoas duvidarem de materiais, verificar fontes e proveniência e corroboração, distingue opinião e propaganda e procure veracidade antes de acreditar e compartilhar. Isso precisa ser sobre o problema mais amplo de desinformação, desinformação e 'notícias falsas', além de descompactar como as narrativas são construídas e compartilhadas, e deve priorizar as comunidades de base e os influenciadores que trabalham com essas comunidades.

Existe uma falta de compreensão pública do que é possível com as novas formas de manipulação de vídeo e áudio. Deveríamos priorizar primeiro o que as pessoas já sabem ou presumem sobre os deepfakes antes de desenvolver esse entendimento sem alarde, por

exemplo, usando o poder de influenciadores como satiristas do deepfake para explicar como funciona. Trabalhar em conjunto com outros projetos, iniciativas, coalizões e agências de verificação de fatos existentes é muito importante não apenas para compartilhar ferramentas e habilidades, mas também para trocar experiências e novas tecnologias.

As ferramentas de detecção precisam ser baratas, acessíveis e aplicáveis para cidadãos e jornalistas

Os participantes, particularmente do mundo do jornalismo e da verificação de fatos, estavam preocupados com a forma como detecção sempre colocaria os jornalistas em desvantagem. Eles já enfrentam as dificuldades de encontrar e desmembrar declarações falsas, especialmente em redes fechadas, sem falar em novas formas de manipulação, como o deepfakes, para as quais não possuem as ferramentas de detecção.

Cada vez mais investimentos estão sendo utilizados para o desenvolvimento de ferramentas para detectar deepfakes, usando novas formas de análise forense de mídia e adaptações dos mesmos algoritmos usados para criar a mídia sintética. Mas existem dúvidas sobre para quem essas ferramentas estarão disponíveis e como os problemas existentes de 'shallowfakes' também serão tratados. Os jornalistas também reiteraram que plataformas como YouTube e WhatsApp não resolverão problemas atuais - você ainda não pode verificar facilmente se um vídeo existente é um 'shallowfake', um vídeo que é simplesmente ligeiramente editado ou apenas renomeado e compartilhado, alegando outra coisa. . Na ausência de ferramentas para detectar o enorme volume existente de "shallowfakes - por exemplo, uma pesquisa de vídeo reversa no WhatsApp -, a detecção de rastreios profundos é um luxo.

Grandes empresas e plataformas como o Facebook devem investir em ferramentas de detecção claras e transparentes e confiáveis além de acessíveis a vários níveis de jornalistas e cidadãos. Quanto mais cedo as falhas e outras falsificações forem detectadas, melhor.

Boa parte da acessibilidade são melhores ferramentas forenses de mídia baratas e disponíveis para todos - que desafiam os incentivos econômicos criados para sintetizar a falsidade e não detectá-la. Isso precisa ser combinado com a capacidade jornalística para novas formas de verificação.

Plataformas como Facebook, YouTube, Google e Whatsapp precisam fazer parte da solução com transparência e apoio para separar a verdade da falsidade.

Plataformas, aplicativos de mensagens fechadas, mecanismos de pesquisa, redes de mídia social e sites de compartilhamento de vídeo também serão os locais onde essas manipulações são compartilhadas. Alguns tópicos e perguntas que devemos discutir são: Qual o papel que as redes sociais e outras plataformas devem desempenhar no combate a

falsificações profundas? Quais devem ser os limites? Como eles devem fornecer acesso aos recursos de detecção? Como eles devem sinalizar aos usuários que o conteúdo é verdadeiro ou falso ou, de alguma forma, manipulado de maneiras que eles não podem ver? Eles devem remover certos tipos de conteúdo manipulado e sob quais critérios?

Os participantes observaram que as plataformas precisam ser mais transparentes no que se refere notícias falsas são distribuídas nelas. Há que se repensar até que ponto as mensagens criptografadas podem chegar e como controlar a disseminação de informações erradas e desinformação.

Para mais informações sobre as recomendações da WITNESS para preparação para deepfakes, consulte: wit.to/Synthetic-Media-deepfakes

Para mais informações sobre as recomendações da WITNESS sobre o que os jornalistas precisam preparar (globalmente), consulte: <https://lab.witness.org/projects/osint-digital-forensics/>

MENU

1. A OFICINA	9
1.1. Quem compareceu?	9
1.2. Como foi a reunião?	10
1.3. Como foi documentado?	11
2. SOBRE A WITNESS E DEEPFAKES	12
2.1. O que é a WITNESS?	12
2.2. Como vídeos são atualmente "falsificados"?	13
2.3. Por que a WITNESS está interessada em deepfakes?	14
2.4. O que a WITNESS está fazendo sobre deepfakes?	14
3. DESINFORMAÇÃO E A VERIFICAÇÃO DE FATOS NO BRASIL	16
3.1. Qual é o cenário de "desinformação" no Brasil?	17
3.2. Como a "desinformação" opera no Brasil?	17
3.3. Como podemos lutar contra a "desinformação"?	17
3.4. Quais são algumas das lições aprendidas pelo Comprova?	18
4. DEEPFAKES: O QUE É ISSO?	20
4.1. O que é um deepfake?	20
4.2. Por que ele é "deep"?	21
4.3. Como são feitos?	21
4.4. É possível fazer um deepfake de um áudio?	21
4.5. Mas como funciona de verdade?	21
4.6. O que é preciso para criar um deepfake?	22
4.7. O que é possível ser feito com a tecnologia por trás dos deepfakes?	22
4.8. Por que devemos nos preocupar com os deepfakes?	22
5. COMO DETECTAR UM DEEPFAKE?	24
5.1. Como detectar deepfakes?	25
5.2. Como o processo de detecção funciona?	25
5.3. Quais são algumas das técnicas para detectar deepfakes?	26
5.4. Quais são as limitações de detecção?	27
6. USANDO ARTE E HUMOR PARA CONSCIENTIZAR	29
6.1. Por que chamar atenção e conscientizar pessoas sobre deepfakes?	29
6.2. Como sensibilizar por meio da arte?	29
6.3. Como deepfakes podem ser usados para conteúdo humorísticos?	31

6.4. Como essa tecnologia está de desenvolvendo nesse cenário?	32
6.5. Qual é o futuro dos deepfakes?	32
7. DEEPFAKES, VIOLAÇÕES DE DIREITOS E MINORIAS	34
7.1. Como os deepfakes chegaram ao grande público?	34
7.2. Como os deepfakes perpetuam a violência de gênero?	34
7.3. Como os deepfakes ameaçam pessoas?	35
7.4. Os deepfakes pode ser usados para o bem?	35
7.5. Qual deve ser o papel das plataformas e redes sociais?	35
7.6. Como lutar contra os deepfakes?	35
8. IMPACTOS PARA ATIVISTAS E COMUNIDADES CARENTES	37
8.1. Como os deepfakes impactam moradores e ativistas de favela?	37
8.2. Algum morador já sofreu consequências de "fake news"?	38
8.3. Como podemos chamar a atenção sobre deepfakes nas favelas?	38
9. LUTANDO CONTRA OS PROBLEMAS DOS DEEPFAKES	39
9.1. O que NÃO devemos fazer na luta contra os problemas dos deepfakes?	39
9.2. O que devemos fazer na luta contra os problemas dos deepfakes?	40
9.3. O que devemos discutir antes de criar uma solução?	40
9.4. Em suma, o que podemos fazer agora?	41
10. AMEAÇAS RELEVANTES E POSSÍVEIS SOLUÇÕES	43
10.1. Quais são algumas das ameaças mais relevantes?	43
10.2. Quais são algumas das ameaças LOCAIS mais relevantes?	45
10.3. Quais são soluções possíveis?	47
A. Como compartilhar conhecimento sobre deepfakes e desinformação?	48
B. Como melhorar a comunicação, principalmente com as comunidades de base?	48
C. Como criar e disseminar ferramentas?	48
D. O que as plataformas devem fazer?	49
11. RESUMINDO	50
12. CRÉDITOS	51

8. A OFICINA

No dia 25 de julho de 2019, a WITNESS produziu uma oficina chamada "Deepfakes: Prepara-se Agora" em São Paulo, Brasil. A oficina teve como objetivo explorar e priorizar soluções pragmáticas para a prevenção e defesa contra um futuro sombrio em que vídeos e áudios elaborados com técnicas de inteligência artificial (AI) estarão amplamente disponíveis para o público.

OBJETIVOS DA OFICINA

- 1/ Expandir o entendimento acerca dessas novas tecnologias e de suas implicações para jornalistas, pesquisadores e defensores de direitos humanos.
- 2/ Reconhecer os possíveis usos positivos dessa tecnologia, começar um mapeamento global de inovações nessas áreas e um entendimento comum sobre os usos maliciosos de imagens e áudios gerados por IA contra o discurso público, a confiança na imprensa e na documentação de direitos humanos, bem como identificar potenciais respostas para essas ameaças.
- 3 / Identificar e priorizar modelos de ameaça no uso dessas ferramentas para o contexto brasileiro.
- 4 / Revisar e priorizar potenciais respostas pragmáticas, táticas e normativas atualmente em discussão sobre detecção, autenticação, coordenação organizada da mídia, bem como comunicação para o público sobre essas novas de manipulação midiática baseada em IA.
- 5 / Identificar prioridades para discussões contínuas com grupos estratégicos e trocas de discussões entre o Brasil e o mundo.

A oficina é parte de uma iniciativa da WITNESS focada em como melhor proteger e sustentar (UPHOLD) vozes marginais, jornalismo cívico e direitos humanos enquanto tecnologias emergentes como IA intersectam com desinformação, manipulação de mídias e um crescente autoritarismo. Mais informações sobre esse trabalho incluindo prévios eventos e relatórios de oficinas estão disponíveis em inglês: <https://lab.witness.org/synthetic-media-and-deep-fakes/>.

O evento em São Paulo é o primeiro de uma série que a WITNESS irá promover ao redor do mundo para assegurar que o entendimento de- e soluções para os deepfakes reflita uma perspectiva global.

8.1. Quem compareceu?

O workshop contou com a participação de mais de 25 pessoas de diferentes áreas e formações: comunicadores, pesquisadores acadêmicos e tecnólogos, jornalistas que trabalham com verificação de fatos e defensores de direitos humanos. Todos foram convidados a participar de conversas orientadas a procurar soluções e respostas para os desafios colocados pela mídia "deepfake".

Organizações com representantes: 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) e WITNESS.

De acordo com as regras da Chatham House, não atribuímos comentários específicos a oradores individuais que não sejam os apresentadores.

8.2. Como foi a reunião?

A reunião começou com uma rápida apresentação de Sam Gregory - o Diretor de Programas da WITNESS e especialista em deepfakes. Em seguida, um rápido aquecimento no qual os participantes se reuniram em círculo e se apresentaram antes de se posicionarem em um espectro na sala de acordo com suas respostas para estas perguntas:

*** Você entende o que os “deepfakes” significam em nosso mundo?**

*** Quanto você está preocupado com deepfakes?**

Ficou claro que, embora a maioria dos não especialistas ainda não estivesse muito familiarizada com a tecnologia, aqueles que tinham conhecimento sobre ela tinham profundas preocupações com elas - especialmente ativistas.

O conteúdo do dia incluiu uma série de apresentações de sete palestrantes diferentes sobre deepfakes e criação de mídia sintética e diferentes perspectivas sobre desigualdades, ameaças a minorias e ativistas de base. Três exercícios diferentes para priorizar e discutir ameaças e pensar em possíveis soluções também foram realizados durante o dia.

AGENDA DO DIA

MANHÃ (09am to 12pm)

- * Apresentações dos participantes, da WITNESS e do workshop
- * Introdução às deepfakes e “mídias sintéticas”
- * Perspectivas técnicas
- * Deepfakes e desigualdades
- * Deepfakes e desinformação no Brasil

TARDE (2pm to 5pm)

- * Discussão em grupo sobre modelos de ameaças e vulnerabilidades
- * Soluções e abordagens interdisciplinares
- * Aprofundamento sobre o contexto brasileiro

NOITE (7pm to 9pm)

- * Debate público no auditório principal da Biblioteca Mário de Andrade

8.3. Como foi documentado?

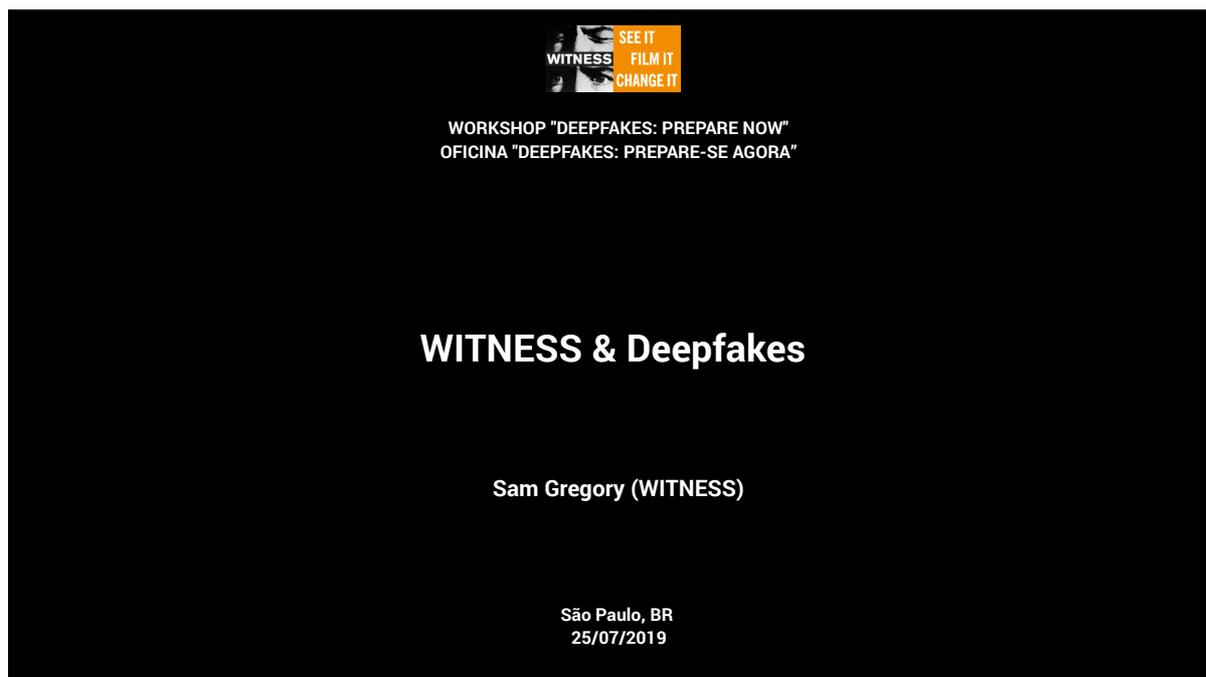
A reunião foi documentada em vídeo, fotos, áudio e texto e esse material foi usado para criar o presente relatório. Aqui você encontra as principais idéias apresentadas, criadas e discutidas durante a reunião, bem como a versão em vídeo das palestras e materiais selecionados apresentados. Observe que esta não é uma descrição cronológica da apresentação, nem uma transcrição completa do conteúdo mencionado, mas uma síntese de seus principais pontos por uma pessoa contratada para documentar a reunião.

O texto segue uma estrutura de "perguntas e respostas" (como uma FAQ), para que você possa navegar facilmente por ele e encontrar o conteúdo mais relevante para você. As perguntas foram organizadas, o máximo possível, seguindo a ordem cronológica apresentada por cada um dos palestrantes, embora algumas de suas idéias tenham mudado de ordem dentro do mesmo bloco. Todos os títulos (perguntas) foram criados pelo relator, bem como os vídeos editados de cada uma das apresentações.

O workshop foi realizado de acordo com as regras da Chatham House e havia uma opção para "conversas fora dos registros", desde que os participantes solicitassem. Nenhum participante solicitou, portanto identificamos suas idéias e atribuímos créditos por suas intervenções e trabalho neste relatório.

9. SOBRE A WITNESS E DEEPPFAKES

Esta primeira parte foi apresentada por Sam Gregory, diretor de programa da WITNESS. Sam trabalha com pesquisa e desenvolvimento de soluções em tecnologia, vídeo, inteligência artificial (IA) e direitos humanos. Na WITNESS, ele lidera iniciativas e treinamentos para ativistas cívicos e jornalistas sobre manipulação da mídia usando IA, bem como o desenvolvimento interdisciplinar de soluções para os desafios atuais dos deepfakes.



[[clique aqui](#) para assistir o vídeo da apresentação com os slides]

9.1. O que é a WITNESS?

A WITNESS é uma organização global de direitos humanos que apoia qualquer pessoa, em qualquer lugar, a usar vídeo e tecnologia para proteger e defender os direitos humanos. Ela está fundamentalmente preocupada com o modo como as pessoas usam o vídeo e a tecnologia para criar informações mais confiáveis. A WITNESS trabalha em todo o mundo e conta com membros em todos os continentes. A organização foi fundada há 25 anos após um incidente de violência policial que foi filmado por um espectador. Hoje, ela trabalha com diferentes jornalistas e ativistas comunitários de todo o mundo e possui forte presença no Brasil via WITNESS Brasil (portugues.WITNESS.org).

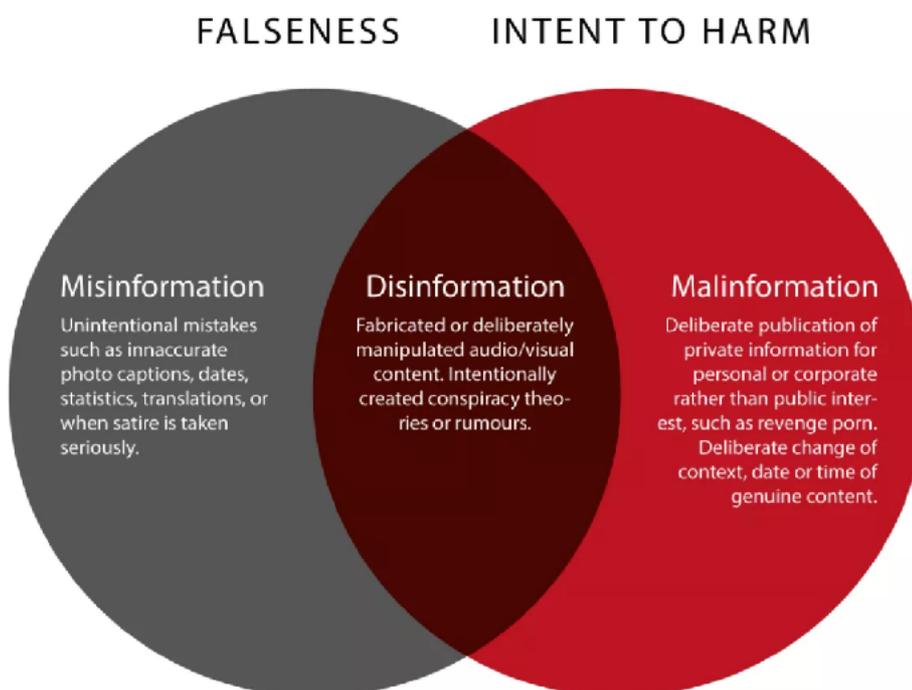
Em poucas palavras, a WITNESS tenta compreender como as pessoas estão usando ferramentas de vídeo e mídia social para compartilhar boas informações e mostrar a realidade - como, por exemplo, documentando crimes de guerra ou violência policial. Ela trabalha junto aos atores locais para aprender com eles, ensiná-los a criar informações e narrativas mais confiáveis, convincentes e eficazes, e também como compartilhar esse conhecimento com aqueles que poderiam usá-lo para o bem. Ela produz vídeos sobre

como documentar informações relevantes e válidas para serem usadas em julgamentos, na busca de justiça e na proteção de direitos (vae.witness.org). A WITNESS também trabalha com a infra-estrutura de tecnologia, interagindo com empresas focadas nas experiências e necessidades de grupos marginalizados - particularmente no Sul Global - e também defendendo melhores políticas e produtos. Como exemplo, eles pressionaram o YouTube na criação de uma ferramenta para permitir o desfoque de rostos em vídeos, em resposta a ameaças contra defensores vulneráveis de direitos humanos. Para mais informações, [clique aqui](#).

9.2. Como vídeos são atualmente "falsificados"?

Os vídeos "falsos" (fake) usualmente fazem uma de três coisas: misinform, disinform ou malinform (os termos em inglês foram mantidos por não achar equivalentes em português).

TYPES OF INFORMATION DISORDER



Fonte: <https://rm.coe.int/information-disorder-report-november-2017/1680764666>

Os deepfakes ainda não são amplamente utilizados para qualquer um desses três propósitos, mas é provável que em breve os sejam. Hoje, os vídeos falsos usam outras

técnicas, como descontextualização ou edição de um vídeo existente ou encenando um vídeo.

- ▶ **DESCONTEXTUALIZANDO CONTEXTOS:** afirma que o vídeo é de um lugar quando é originalmente de outro.
- ▶ **EDIÇÃO:** não mostra a história inteira ou possui algumas frases ou citações completamente fora de contexto.
- ▶ **MANIPULAÇÃO:** a velocidade é reduzida ou há alguma outra forma de manipulação técnica.
- ▶ **ENCENAÇÃO:** vídeos que são fictícios e criados para enganar as pessoas, passando como verdadeiros (ainda não tão comuns quanto as pessoas supõem, mas existem alguns exemplos)
- ▶ **MANGUEIRA DE FALSIDADE (FIREHOSE OF FALSEHOOD):** criação de vários vídeos contraditórios, mídias sociais e conteúdo até que as pessoas não saibam em que acreditar (como pioneiro da Rússia na Ucrânia).

9.3. Por que a WITNESS está interessada em deepfakes?

Há dezoito meses, a WITNESS começou a ouvir falar em deepfakes. Embora ainda fossem muito raras, algumas pessoas começaram a ver nos deepfakes uma fonte potencial de um “infocalipse” ou “tecnopocalipse”, onde toda a confiança em imagens e palavras entraria em colapso. Isso despertou a atenção da WITNESS, pois ela rapidamente entendeu que esse seria um tópico importante a ser compartilhado com seus parceiros, e priorizar o trabalho proativo para garantir que seus possíveis danos sejam minimizados.

Hoje já começamos a sentir um impacto negativo dos deepfakes em relação à violência de gênero (especialmente por meio de vídeos pornográficos deepfake, o lugar onde tudo começou), mas ainda não por desinformação - embora isso pareça apenas uma questão de tempo.

A WITNESS decidiu discutir este tópico em diferentes lugares do mundo e usar esta janela de oportunidade para se preparar melhor antes que nós nos encontremos no centro de um furacão. O mantra deles foi "prepare-se, não entre em pânico". Nos últimos anos, ela realizou reuniões e workshops com especialistas, pesquisadores, jornalistas, políticos de plataformas, entre outros de diferentes partes do mundo (entre aliados e parceiros da WITNESS e organizações globais importantes, como o Partnership on AI e a BBC) .

9.4. O que a WITNESS está fazendo sobre deepfakes?

A retórica sobre deepfakes até o início de 2019 foi excessivamente focada nos EUA e na Europa, bem como nos danos a nível nacional para políticos. Central para a WITNESS é o compromisso de incluir conhecimentos e medidas a partir de uma perspectiva não americana / ocidental e ouvir jornalistas, especialistas em desinformação, defensores de direitos humanos e comunidades vulneráveis no Sul Global, para evitar repetir os erros que

foram cometidos em respostas anteriores a crises de desinformação. Portanto, a WITNESS decidiu não apenas diversificar os países onde realizariam essas reuniões e discussões, mas também os atores a serem convidados: agora não apenas pessoas da mídia e jornalistas, mas também ativistas, verificadores de fatos (fact-checkers) entre outros atores-chave.

Há dois meses, a WITNESS organizou uma reunião no Rio de Janeiro para realizar uma primeira conversa com ativistas da comunidade brasileira sobre suas percepções acerca dos riscos dos deepfakes. Abaixo, algumas das conclusões dessa reunião sobre o que podemos fazer agora acerca dos deepfakes, conclusões que foram usadas como premissas iniciais para a agenda desta reunião:

O QUE PODEMOS FAZER AGORA A RESPEITO DOS DEEPFAKES?

- ▶ Desescalar a retórica
- ▶ Identificar modelos de ameaças e soluções desejáveis a partir de uma perspectiva global.
- ▶ Promover abordagens interdisciplinares e mais soluções.
- ▶ Construir sobre experiências passadas e comunitárias, especialmente de alfabetização midiática, jornalismo e OSINTs.
- ▶ Estabelecer o que queremos das plataformas e das empresas que vendem ferramentas para criação de deepfakes.
- ▶ Ressaltar fortemente os prós e contras da escolha de infraestrutura técnica, especialmente no momento em que impactam na confiança geral das pessoas nos vídeos e imagens.



[para baixar os slides, [clique aqui](#)]

10. DESINFORMAÇÃO E A VERIFICAÇÃO DE FATOS NO BRASIL

Esta seção foi apresentada por Sergio, um jornalista com experiência em gestão de negócios jornalísticos e marketing digital. Ele é o editor do COMPROVA, uma coalizão de 24 organizações (principalmente jornais) criada em 2018 para monitorar as notícias durante as eleições brasileiras e realizar uma espécie de "desmistificação social", monitorando e verificando as notícias produzidas por fontes anônimas (não autoridades, estas foram



deixadas para agências de verificação de fatos) e compartilhadas nas plataformas de mídia social.

[\[clique aqui\]](#) para assistir o vídeo da apresentação com os slides]

10.1. Qual é o cenário de "desinformação" no Brasil?

Vivemos hoje uma época em que a opinião vem antes da informação: quando a informação chega, a pessoa já tem uma opinião formada sobre o assunto. E uma mentira é sempre mais sexy do que a verdade: é muito mais fácil produzir desinformação, porque ela comporta o conteúdo que sua imaginação e tecnologia permitirem, enquanto as informações, as notícias precisam estar ligadas a fatos.

Os cenários político, de redes sociais, de jornalismo e de confiança são muito importantes para entender onde o Brasil se encontra no momento. Desde os protestos de 2013, alguns grupos começaram a publicar um grande número de "notícias falsas" que tinham apenas o título lido pelas pessoas e depois era compartilhada. As pessoas começaram a usar mais WhatsApp e compartilhar notícias de uma maneira mais "privada". Além disso, muitos jornais e outros sites de informação criaram paywalls para seu conteúdo e bloquearam seu acesso a um grande número de pessoas. Essa combinação, em um cenário de baixa confiança na mídia, por exemplo, criou um cenário propenso a notícias falsas: estas geralmente são provenientes de pessoas em quem você confia (um amigo, um membro da família) enquanto muitas notícias vêm de organizações em que as pessoas não confiam em sua maioria.

10.2. Como a "desinformação" opera no Brasil?

O processo atingiu um pico durante as últimas eleições presidenciais, em que "notícias falsas", micro-direcionamento e o uso de bots foram generalizados. Essas ondas de desinformação acompanham de perto a agenda pública e muitas figuras públicas ajudam a criar factóides, compartilhar desinformação e chamar a atenção das pessoas para notícias falsas. Essas desinformações geralmente ocorrem em ondas em torno de um dado fato (por exemplo, notícias falsas e rumores sobre uma pesquisa eleitoral antes de seu lançamento).

Para enganar as pessoas, os vídeos falsos geralmente têm uma qualidade muito baixa, os áudios não podem ser totalmente ouvidos etc. Nem sempre todo o conteúdo é falso; às vezes, diferentes partes de conteúdo verdadeiro são reunidas para construir uma narrativa que, no seu conjunto, é falsa e induz as pessoas a pensarem o que seu autor quer que elas pensem que é verdade. Como, hoje, as pessoas querem acreditar que suas crenças e opiniões são sempre verdadeiras, muitas vezes elas são enganadas por essas notícias mesmo quando claramente elas não são verdadeiras.

10.3. Como podemos lutar contra a "desinformação"?

O Comprova criou amplos bancos de dados de monitoramento capazes de identificar quem são os agentes por trás dessas desinformações, o que está em jogo com esse conteúdo, quem é capaz de disseminar esse tipo de conteúdo etc. O objetivo é descobrir

onde ele está sendo criado e verificar essas desinformações em seus estágios iniciais, antes de chegar-se a um ponto em que já foi compartilhada com milhares ou milhões de pessoas.

As pessoas compartilharam essas notícias com o Comprova via WhatsApp ou procuraram-nos ativamente em suas pesquisas. O Comprova tinha uma equipe de jornalistas (cerca de 5 ou 6) trabalhando para desmascarar a desinformação. É um trabalho complexo, no qual você sempre deve verificar e checar as fontes e comunicar os resultados da pesquisa de maneira fácil, de fácil entendimento das pessoas.

O Comprova também distribui relatórios com os resultados de suas verificações e tenta promover uma cultura de verificação em jornais e redações. Atualmente, trabalha com a investigação de desinformação sobre políticas públicas em nível federal.

Uma boa analogia para o trabalho deles é o FIFA VAR (árbitro assistido por vídeo). Como no VAR da FIFA, as pessoas entendem a importância do trabalho do Comprova, mas muitas vezes o consideram chato pois acham que isso torna as coisas mais lentas. E, como o VAR, o Comprova está tentando ser mais rápido e transparente sobre a maneira pela qual julgou a veracidade de uma determinada notícia (como fez a verificação, onde buscaram informações, convite para as pessoas a visitar o link original e refazer o caminho deles). De acordo com eles, essa pode ser a única maneira de fazer alguém mudar de opinião.

10.4. Quais são algumas das lições aprendidas pelo Comprova?

Aumentar a conscientização sobre uma tecnologia emergente como a dos deepfakes é crucial. As pessoas precisam saber que essa tecnologia existe para começar a questionar a veracidade de outros vídeos recebidos. Por isso, vídeos de humor podem ter um bom resultado, pois podem se tornar virais e fazer com que muitas pessoas entendam o conceito por trás dele.

Algumas das lições mais importantes do Comprova que devem ser consideradas no desenvolvimento de estratégias para lidar com vídeos deepfake são:

- ▶ Só porque uma nota ou documento é oficial não significa necessariamente que seu conteúdo seja verdadeiro (pois são parciais).
- ▶ A leitura dos comentários deixados nas notícias é importante, pois ajuda a entender o quanto essas notícias são significativas e o que as pessoas estão falando sobre elas.
- ▶ Você não pode trazer para as pessoas a versão verdadeira de uma história falsa sem falar e compartilhar sobre a versão falsa. Isso pode ter alguns efeitos negativos, como nos casos em que as notícias falsas se tornaram mais populares após a checagem social do que antes e as pessoas podem continuar a acreditar que é verdadeira (mesmo dizendo claramente o contrário) porque aquela versão corresponde às suas crenças preconcebidas.

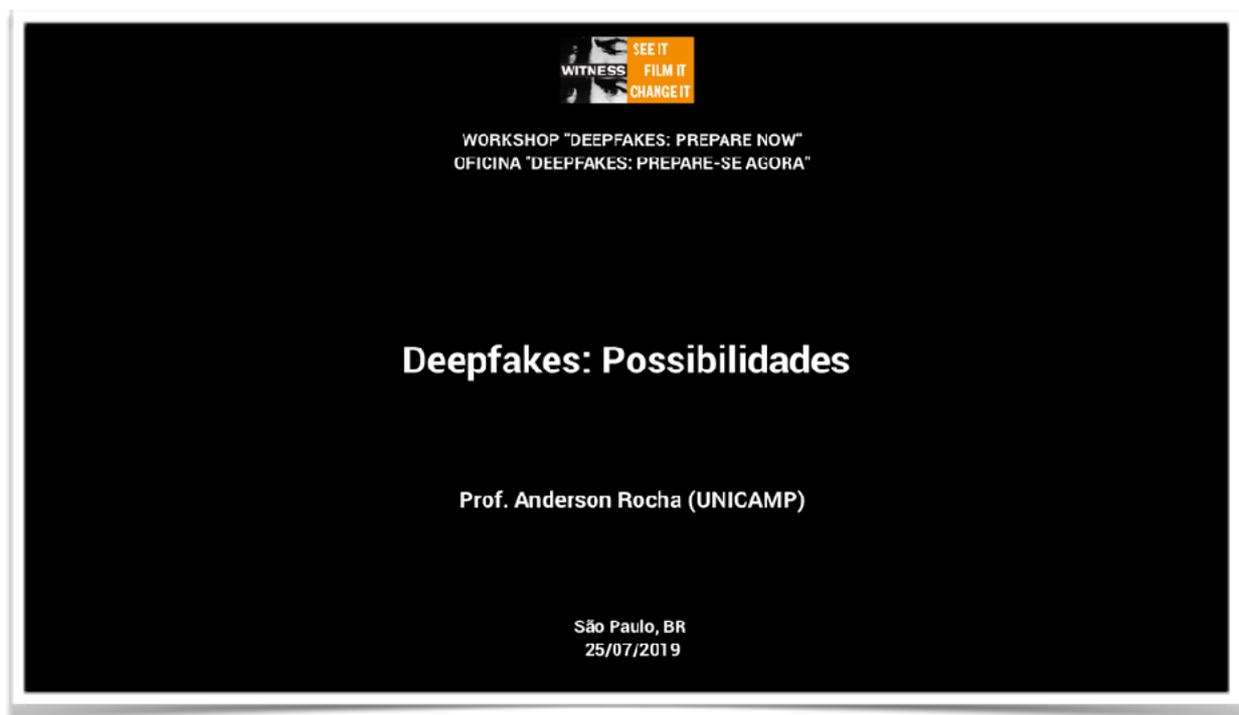
- ▶ Para se comunicar com as massas, é importante ter aliados em outros jornais e em noticiários de TV que possam alcançar um público que a Internet ou uma versão impressa de um jornal simplesmente não consegue.
- ▶ A maioria das pessoas lê apenas as manchetes das notícias (que geralmente são sensacionalistas e desinformativas) ou mesmo quando lêem a versão completa, muitas vezes podem entender errado o significado devido ao viés cognitivo.



[para baixar os slides, [clique aqui.](#)]

11. DEEPFAKES: O QUE É ISSO?

Esta seção foi apresentada por Anderson Rocha, professor e diretor do Instituto de Computação da Universidade Estadual de Campinas em duas palestras diferentes. Suas principais áreas de especialização são computação forense, inferência de dados complexos e inteligência computacional. Ele trabalha como editor associado de importantes revistas internacionais e é membro da Academia Brasileira de Ciências.



[clique aqui](#) para assistir o vídeo da apresentação com os slides]

11.1. O que é um deepfake?

Um vídeo deepfake é aquele em que você tem um elemento de uma pessoa-origem (seja um rosto, um corpo, lábios etc.) inserido ou mesclado em uma pessoa-destino (aquele em que essa parte será inserida). No entanto, um deepfake é mais do que uma simples "troca de cabeças" (como as feitas em fotos com Photoshop ou vídeos com AfterFX): eles usam inteligência artificial (IA) em seu processo de criação. É por isso que eles são tão novos: a tecnologia por de trás deles está sendo desenvolvida atualmente.

A técnica mais famosa por trás dos deepfakes hoje é a "troca de cabeças", um processo que tem sido feito há séculos com estátuas ou pinturas, com fotos e até com vídeos em alguns filmes de Hollywood. Avatar (2009) tinha a mesma tecnologia que está sendo discutida hoje. Nos filmes de Hollywood, sensores foram colocados nas pessoas e seus movimentos transformados em personagens. A diferença é que, agora, com a IA, essas ferramentas são mais fáceis e baratas de usar.

11.2. Por que ele é "deep"?

O termo "deepfake" tem origem no nome de um usuário do Reddit que começou, no início de 2018, a postar vídeos pornográficos com rostos de atrizes famosas (agora excluídos pelo próprio usuário). Não há nada mágico por trás da palavra, embora mais tarde as pessoas tenham começado a justificar o uso de "deep" porque a tecnologia usa "redes neurais profundas". Hoje, o "deep" é usado para distinguir essas vídeos falsos de outros (como "shallow fakes", vídeos que usam apenas ferramentas de pós-produção, como After FX, ou envolve a falta de contextualização dos vídeos existentes) ou até mesmo "dumb-fakes" (como memes e outros conteúdos que são farsas grotescas ou evidentes).

11.3. Como são feitos?

Basicamente por meio de equações matemáticas. Para criar um vídeo deepfake, você precisa rodar um processo de pesquisa, otimização e cálculo de distância. Você transforma o elemento de origem que deseja inserir no alvo (geralmente um rosto) e o transforma em um tipo de "massinha de modelar". Você detecta vários pontos nessa face e cria um modelo 3D dela. E, a partir disto, é preciso distorcer esses pontos para ajustar a face de origem na face de destino. É um processo matemático, no qual você pode obter os pontos de origem exatos, o ponto do alvo e fazer um mapeamento matemático, diminuindo a distância entre eles. E então eles se tornam cada vez mais próximos um do outro e se ajustam. Você pode fazer isso com rostos, lábios, corpos, etc.

11.4. É possível fazer um deepfake de um áudio?

Claro! Para criar um áudio falso ou deepfake, o processo é semelhante: você transforma a fala em microssegmentos (algo que não é novo). No entanto, agora você os divide em microssegmentos de entonação e desentonação. Você tem vários micropartes que podem ser combinados com uma precisão cada vez maior para fornecer uma amostra realmente boa de uma pessoa-alvo.

11.5. Mas como funciona de verdade?

Para criar um deepfake, você geralmente tem duas redes de IA diferentes competindo entre si: uma que tentará gerar o vídeo falso e outra que tentará detectar se aquele produto é falso ou não. Eles são adversários. Então, se o segundo detecta, o primeiro volta e faz outra criação. Em determinado momento, eles se equilibrarão e a IA que está tentando detectar falsificações não verá nenhum erro. É por isso que eles são chamados de Redes Adversárias Generativas.

Hoje, neste processo de modelagem, é possível fazer quantas combinações quiser entre esses pontos: você pode colocar o olho para cima, para baixo, gerar outro ponto entre eles, explodi-lo. Com um bom equipamento de computação, você pode gerar uma amostra em poucas horas. Porém, quanto mais longo e maior o banco de dados que você tiver, melhores serão os resultados obtidos.

11.6.O que é preciso para criar um deepfake?

Para criar um deepfake, é necessário uma boa quantidade dos chamados "dados de treinamento" - geralmente imagens de vídeo do alvo. É por isso que políticos e atores famosos são as escolhas mais óbvias para experimentos. Quanto mais dados alguém tiver disponível on-line (sejam fotos ou vídeos), mais fácil será para alguém criar um deepfake usando sua imagem.

Você também precisa de uma unidade de processamento gráfico (GPU), pois ela permite o cálculo paralelo necessário para criar o deepfake. As GPUs estão amplamente disponíveis a um custo razoável no mercado.

E, é claro, é também necessário experiência suficiente para baixar e instalar os repositórios ou até mesmo usar aplicativos como o Fakeapp, que nem sempre têm uma interface amigável.

11.7.O que é possível ser feito com a tecnologia por trás dos deepfakes?

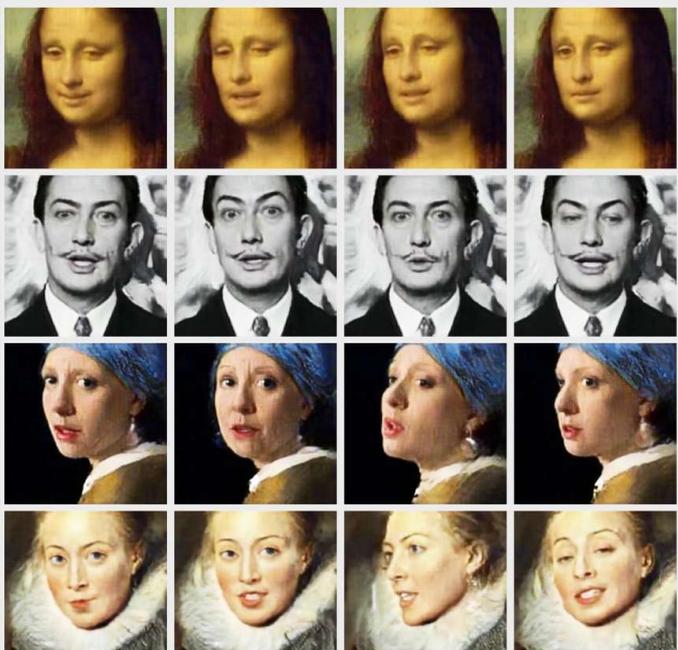
A troca de rosto é apenas a ponta do iceberg Deepfake. Essa tecnologia traz diferentes usos em potencial, especialmente na indústria cinematográfica, para reconstituições de vídeos, criação de personagens, mudança de voz em traduções, edição baseada em conteúdo etc. Um dia você poderá ter seu rosto e seus amigos passando no sua série favorita do Netflix. Você também pode usar a tecnologia para criar aplicativos para identificar e classificar conteúdo, dentre outros usos.

11.8.Por que devemos nos preocupar com os deepfakes?

As ameaças atuais não estão no lado da tecnologia, mas na desinformação. Embora isso possa parecer assustador num futuro próximo, hoje é ainda fácil notar um vídeo deepfake, especialmente se você usar ferramentas técnicas para detecção. O processo de criação de um vídeo deepfake ainda deixa pistas que podem ser identificadas e facilmente identificáveis - mesmo a olho nu: luzes, poses, sombras, coisas que você não pode fazer perfeitamente com a tecnologia atual. (NOTA dos editores: isso não é necessariamente verdade em algumas modificações menores do deepfake, como movimentos dos lábios, e não em toda a face).

O problema, no entanto, não é apenas o uso desta para desinformar. É o mesmo que em "Admirável mundo novo": com uma mangueira de conteúdo falso, você não sabe em que acreditar. Isso pode ser extremamente difícil de lidar nesse período de "pós-verdade", no qual as pessoas parecem acreditar apenas naquelas notícias e conteúdos alinhados com suas crenças pré-estabelecidas. Conteúdo falso pode criar confusão nas pessoas e pode também criar esquecimento (oblivion), uma situação em que você não consegue entender o que é verdade ou não.

As fake-news estão afetando gravemente a democracia por meio da segmentação múltipla e específica, com vídeos falsos (geralmente pornô) enviados para chantagear e impedir ativistas e jornalistas, dentre muitas outras técnicas vistas em todo o mundo.



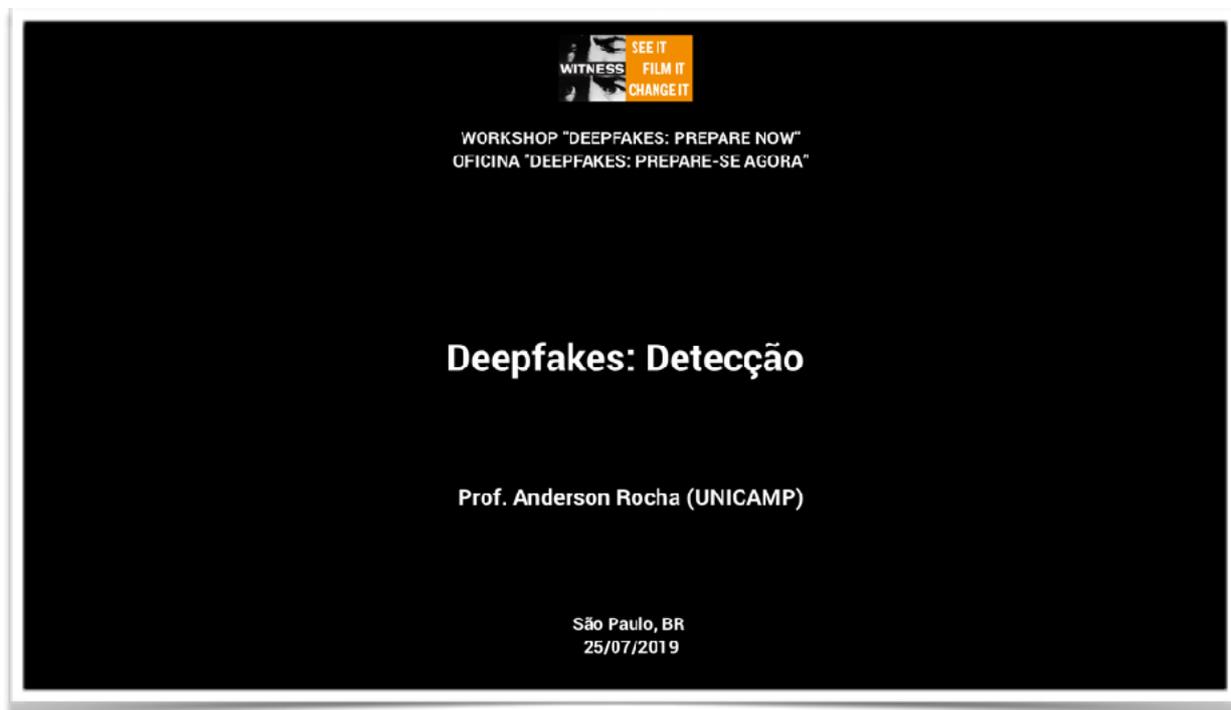
DeepFakes: Possibilidades

Prof. Anderson Rocha
Instituto de Computação, Unicamp
anderson.rocha@unicamp.br

[para baixar os slides, [clique aqui](#)]

12. COMO DETECTAR UM DEEPPFAKE?

Essa seção também foi apresentada por Anderson Rocha, do Instituto de Computação da UNICAMP (já caracterizado acima).



[[clique aqui](#) para assistir o vídeo da apresentação com os slides]



12.1. Como detectar deepfakes?

Para detectar um deepfake, os especialistas podem usar a mesma tecnologia usada para sua criação. Eles podem usar a IA contra a IA treinando a AI para detectar o que é falso ao invés de gerar conteúdo falso, como em um processo de geração do deepfake.

Os deepfakes atuais também enfrentam problemas com os movimentos dos olhos e da boca ou na textura do cabelo, só para citar alguns. Portanto, em vez de precisar olhar para o vídeo inteiro, você pode inspecionar apenas determinadas regiões (aquelas mais



propensas a imperfeições) e analisá-lo com mais cuidado. Especialistas geralmente procuram detalhes e imperfeições que seriam mais difíceis de serem melhorados da noite para o dia ou em um curto período de tempo. Geralmente, tentam analisar as propriedades da imagem que seriam mais difíceis de serem modificadas de maneira perfeita.

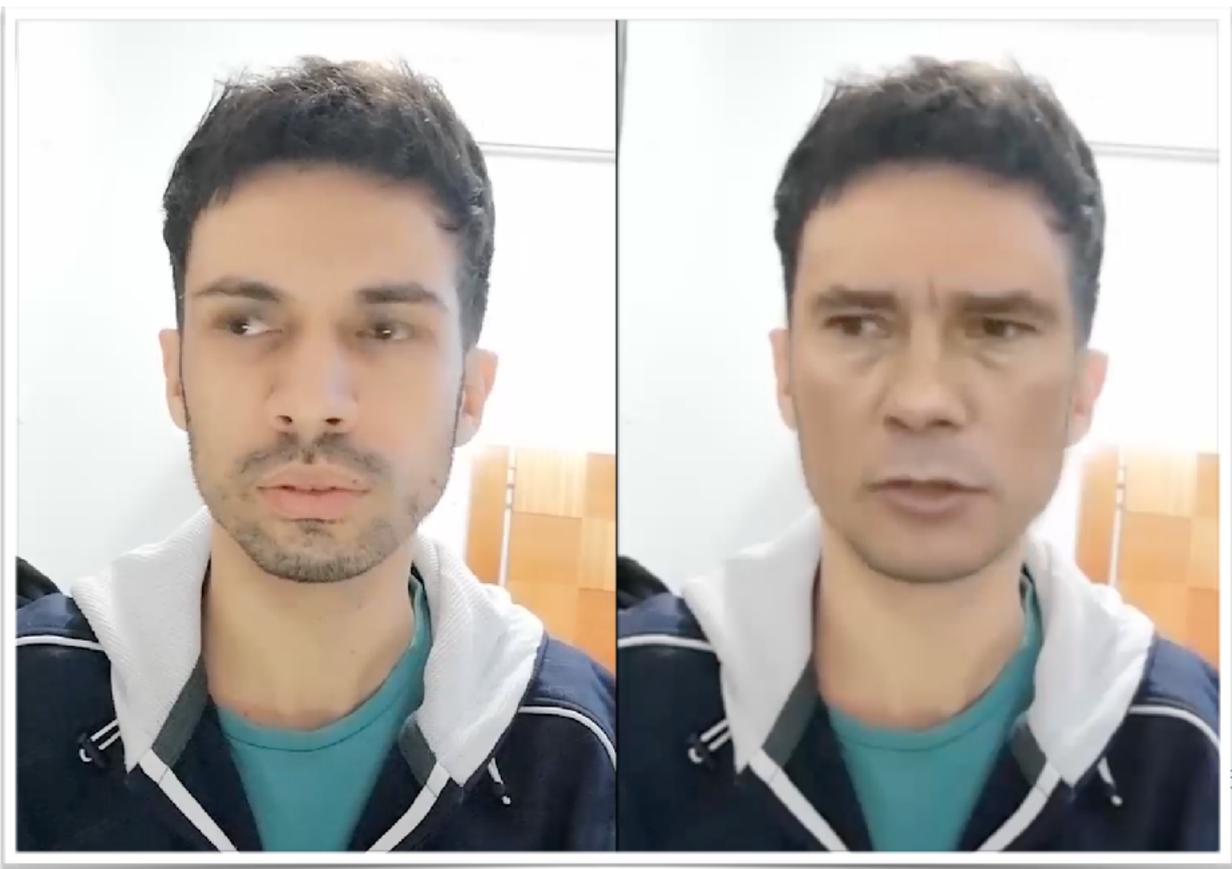
12.2. Como o processo de detecção funciona?

Se o processo de geração consiste em duas redes neurais (uma recebendo uma imagem de origem e outra gerando o conteúdo de destino), o processo de detecção consiste no treinamento de uma rede neural com dois tipos de dados (vídeos falsos e verdadeiros) e na disponibilização de pistas e elementos sobre o que poderia ter sido falsificado no vídeo. Resumindo: primeiro, é preciso treinar uma IA para gerar deepfakes e outra para detectar. E, em seguida, obtém-se dessa última IA a capacidade de detecção aprimorada e usá-la como uma ferramenta para detectar vídeos falsos profundos usando essa tecnologia (ou outras tecnologias anteriores aprendidas).



12.3. Quais são algumas das técnicas para detectar deepfakes?

Analisar luzes: a tecnologia deepfake atual ainda não possui o mesmo processo físico de captura de imagem. Isso significa que sua distribuição de luz fornece pistas sobre a autenticidade do vídeo. Se alguém obtiver um vídeo de alguém falando em um dia ensolarado e tentar trocar de rosto com alguém gravado em um dia nublado, os





DeepFakes: Detecção

Prof. Anderson Rocha
Instituto de Computação, Unicamp

anderson.rocha@unicamp.br

especialistas poderão facilmente desmascarar falsificações criando um mapa de iluminantes (os perfis das fontes de luz) ou olhando a direção de luz (a luz refletida nos objetos é diferente se eles vierem de materiais diferentes). Portanto, se uma farsa profunda não usou um algoritmo para combinar com as luzes, essa pode ser uma maneira de identificá-las.

Filogenia múltipla: pode-se criar uma IA para descobrir a ordem cronológica e derivada de imagens ou vídeos, para descobrir de quais imagens ou vídeos de origem eles se basearam. Isso pode funcionar com fotos, vídeos e até texto, embora às vezes possa falhar (por exemplo, quando um autor tem estilos muito diferentes, como o poeta português Fernando Pessoa, com seus heterônimos).

Fazer com que a IA aprenda o estilo de alguém: para detectar um texto falso, por exemplo, é possível treinar um algoritmo que aprenda o estilo de escrita, verificando o uso de maiúsculas, gírias, substantivos, combinação de verbos, artigos etc. Você pode definir um perfil estilométrico para cada autor e detecte o que é falso ou não. No entanto, se alguém puder treinar outra IA para gerar para aprender os estilos, talvez o detector não consiga descobrir (atualmente ele pode identificar bots como aqueles que possuem um estilo de escrita estática).

12.4. Quais são as limitações de detecção?

Ainda não conseguimos criar um detector universal de deepfakes que funcionaria com qualquer algoritmo e tecnologia. A tecnologia de detecção atual requer exemplos de falsificações para poder realizar posterior detecção (elas realmente precisam de duas classes, reais e falsas). Para criar um detector universal, no entanto, teríamos que entender a distribuição de imagens naturais, algo que ainda está longe de ser realidade.

Portanto, cada detector é útil apenas para detectar as tecnologias para as quais foram criados. Isso pode ser um grande problema em um cenário em que a tecnologia usada está

melhorando rapidamente. É um "jogo de gato e rato" e é difícil dizer quem vencerá simplesmente usando a tecnologia.

[para baixar os slides, [clique aqui](#)]

13. USANDO ARTE E HUMOR PARA CONSCIENTIZAR

Esta seção foi apresentada por Marlus Araújo, designer e programador interessado na convergência da arte, design e tecnologia e trabalhando atualmente com vigilância e tecnologia.



[vídeo não disponível a pedido do autor]

13.1. Por que chamar atenção e conscientizar pessoas sobre deepfakes?

Marlus argumentou que a sociedade civil deveria aprender com seus erros passados e não adotar uma estratégia de não falar sobre deepfakes ou seus riscos para não chamar a atenção para eles (como fizeram durante as últimas eleições com Bolsonaro). Ele acha que agora é o momento de tornar os deepfakes mais populares e conhecidos pelas pessoas, porque se (ou quando) surgir a avalanche de vídeos deepfakes, as pessoas estarão cientes de que agora é possível e relativamente fácil falsificar vídeos, trocar rostos, criar diálogos falsos, etc. Para ele, a conscientização é essencial para sobreviver a uma futura avalanche de vídeos falsos.

13.2. Como sensibilizar por meio da arte?

Durante sua breve palestra, Marlus apresentou dois vídeos diferentes nos quais desejava mostrar visualmente às pessoas os riscos e aplicações potenciais da tecnologia deepfake. Seus projetos misturam vídeo e arte.

O primeiro vídeo mostra uma instalação de arte que ele fez em um museu brasileiro no qual o público poderia controlar o rosto do presidente brasileiro ao vivo. Para criá-lo, ele conseguiu um vídeo de Bolsonaro que o tinha em diferentes ângulos, fazendo movimentos diferentes, com uma cor sólida de fundo. Ele então cortou o rosto e fez um treinamento de 3 horas com uma máquina de IA. Ele então fez o processo inverso: pegou uma webcam para gravar o rosto do participante e o computador sintetizava em tempo real a imagem. A qualidade ainda era baixa e era uma falsificação grotesca, mas ajudou a aumentar a conscientização sobre o que pode acontecer no futuro próximo ou durante as próximas eleições.



[para assistir esse vídeo, [clique aqui.](#)]

O segundo vídeo foi feito com um processo de treinamento de IA mais complexo: ele não apenas cortou imagens, mas também ampliou e reduziu elas, adicionou outros dados de diferentes ângulos e posições, fez mais interações e criou uma farsa melhor. Ele apresentou esse projeto para o festival de video-mapping de Brasília no Museu Nacional. As pessoas podiam controlar o rosto de Bolsonaro ao vivo e ele mais tarde usou esse vídeo para criar um vídeo que foi visto como um vídeo de protesto.



[para assistir esse vídeo, [clique aqui.](#)]

13.3. Como deepfakes podem ser usados para conteúdo humorísticos?

Esta seção foi apresentada por Bruno Sartori - um humorista de uma cidade de médio porte no Brasil que descobriu os deepfakes logo no início e logo viu um grande potencial nessa tecnologia para criação de seus próprios vídeos de humor e começou a estudá-la.

[[clique aqui](#) para assistir o vídeo da apresentação com os slides]

Bruno apresentou quatro técnicas / demos diferentes de deepfakes com as quais ele já trabalhou para responder:

- ▶ **Troca de rosto de políticos usando dublês:** Bruno usou um vídeo com um comediante profissional fingindo ser o presidente Bolsonaro, no qual inseriu o rosto do atual presidente brasileiro. Ele considera este vídeo um dos seus melhores falsos profundos, especialmente porque ele trabalhou bastante no After Effects para torná-lo mais realista.

[para assistir esse vídeo, [clique aqui.](#)]

- ▶ **Vídeo e voz para criar memes:** Bruno produz memes em vídeo que usam a tecnologia deepfake para fazer com que a imagem (rosto) e a voz se pareçam com o alvo. O vídeo mostra um deepfake do presidente brasileiro cantando uma música boba que viralizou em um vídeo do YouTube há alguns anos atrás.

[para assistir esse vídeo, [clique aqui.](#)]

- ▶ **Deepfakes ao vivo de políticos em shows de humor:** Bruno produziu um vídeo que mostra o que pode ser feito em tempo real (ao vivo) com imagem e voz. É um deepfake do ex-juiz e atual ministro da Justiça Sergio Moro para um famoso talk-show de humor brasileiro. No vídeo, o apresentador do programa finge fazer uma videochamada online ao vivo com o deepfake de Moro, que está sendo controlado pelo artista em tempo real. Parece muito realista e, apesar do seu conteúdo humorístico, é assustador.

[para assistir esse vídeo, [clique aqui.](#)]

13.4. Como essa tecnologia está de desenvolvendo nesse cenário?

A evolução das falsificações profundas nesses 18 meses foi incrível. Como Bruno relata, no ano passado, ele levaria cerca de 30 dias para ter uma face aceitável com um treinamento de pixels de imagem de 64X64 pixels. Com a tecnologia atual (treinamento de 256X256 pixels), ele usa cerca de 4.000 imagens para criar um rosto (esse é o número ideal de acordo com seu teste) e leva cerca de 4 dias para formar um rosto. Quando esse rosto é "treinado", quando ele precisa inseri-lo em outros vídeos, é mais rápido o produto final - às vezes menos de 2 horas. Ele também menciona que já viu algumas pessoas usando hardware de smartphone para criar os deepfakes e os resultados foram muito satisfatórios.

13.5. Qual é o futuro dos deepfakes?

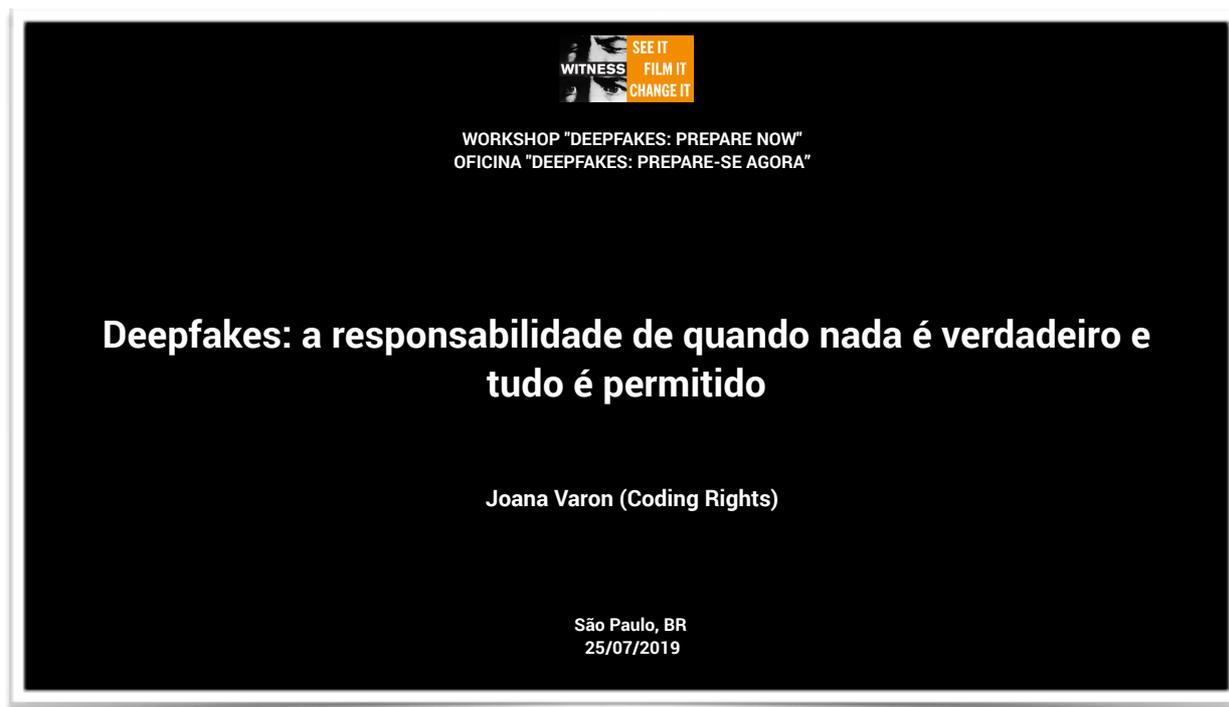
Os deepfakes também abre possibilidades para você agir com toda a sua família em seu programa favorito da Netflix. O último vídeo mostra um dos seus trabalhos atuais de Bruno em que ele tenta colocar seu próprio rosto em um vídeo de um ator muito expressivo



[para assistir esse vídeo, [clique aqui.](#)]

14. DEEPPAKES, VIOLAÇÕES DE DIREITOS E MINORIAS

Esta seção foi apresentada por Joana Varon. Joana é diretora e fundadora da Coding Rights e da "Transfeminist Network of Digital Care". Ela é afiliada ao Berkman Klein Center for Internet and Society da Universidade de Harvard e desenvolveu projetos criativos e de defesa de direitos que trabalham na interseção de direito, arte e tecnologia.



[[clique aqui](#) para assistir o vídeo da apresentação com os slides]

14.1. Como os deepfakes chegaram ao grande público?

Joana trouxe alguns casos para mostrar que a tecnologia visual por trás dos deepfakes já estava sendo usada no cinema e na televisão há algum tempo. Sua primeira lembrança de um deles é um videoclipe do **Björk de 1998**, no qual a cantora se transforma em urso. Os anos se passaram e, em 2015, surgiram filtros do Instagram que permitiram que você brincasse com o rosto de alguém, assim como no clipe da Björk.

Em 2019, o Snapchat lançou filtros nos quais você pode parecer mais velho e até mudar de sexo. Há um caso famoso de um homem que tirou várias fotos de si mesmo como mulher e criou um perfil do Tinder para ver como as pessoas interagem com ele. No mundo virtual visual contemporâneo, quase tudo é possível.

14.2. Como os deepfakes perpetuam a violência de gênero?

Agora os deepfakes surgem e se popularizam com uma espécie de democratização desse conhecimento: o que antes era discutido nos círculos acadêmicos, obteve sua primeira

aplicação e reconhecimento em todo o mundo. E foi para pornografia. Os deepfakes vieram mostrar que agora estamos “fodidos” e que um vídeo falso de sexo envolvendo a sua pessoa pode ser divulgado sem você nem sequer um dia ter pensado em se envolver em tal prática sexual. Existem até aplicativos, fóruns e repositórios para ajudar a criar um deepfake sexual de seu crush ou ex-namorada.

Não apenas os deepfakes começaram como uma maneira de produzir conteúdo degradante contra as mulheres, mas também um dos primeiros aplicativos que utilizaram a tecnologia foi um app estilo “óculos de raio-x” que fazia qualquer mulher ficar “nua”: sim, o aplicativo só funcionava para mulheres. Atualmente, o aplicativo está proibido, embora ainda esteja disponível em sites não oficiais.

14.3. Como os deepfakes ameaçam pessoas?

Embora a pornografia ainda seja a principal área de uso dos deepfakes hoje (uma pesquisa no termo no Google resultará em páginas com pornografia, mas ainda não em notícias falsas), já existem casos em que vemos uma mistura de violência, deepfakes e violação da liberdade de expressão e de imprensa. Na Índia, uma jornalista teve um vídeo de sexo falso atribuído a ela numa tentativa de silenciá-la. No Brasil, o atual governador de São Paulo também foi vítima de um vídeo falso (era mais um shallow fake do que um deepfake) em que aparecia em uma orgia sexual e que foi compartilhado amplamente nas redes sociais na véspera das últimas eleições. Os danos do vídeo não foram suficientes para impedir que ele fosse eleito. No entanto, quais poderiam ter sido os danos se o candidato alvo fosse uma mulher?

14.4. Os deepfakes pode ser usados para o bem?

Apesar desses efeitos negativos, os deepfakes também podem ser usados como um meio de paródia e para chamar a atenção para um determinado assunto. Assim como pode ser usado para aumentar as assimetrias de gênero, também pode ser usado para reduzi-las. A tecnologia é neutra, os usos que se dão é que guiarão seus efeitos. No entanto, como podemos garantir isso?

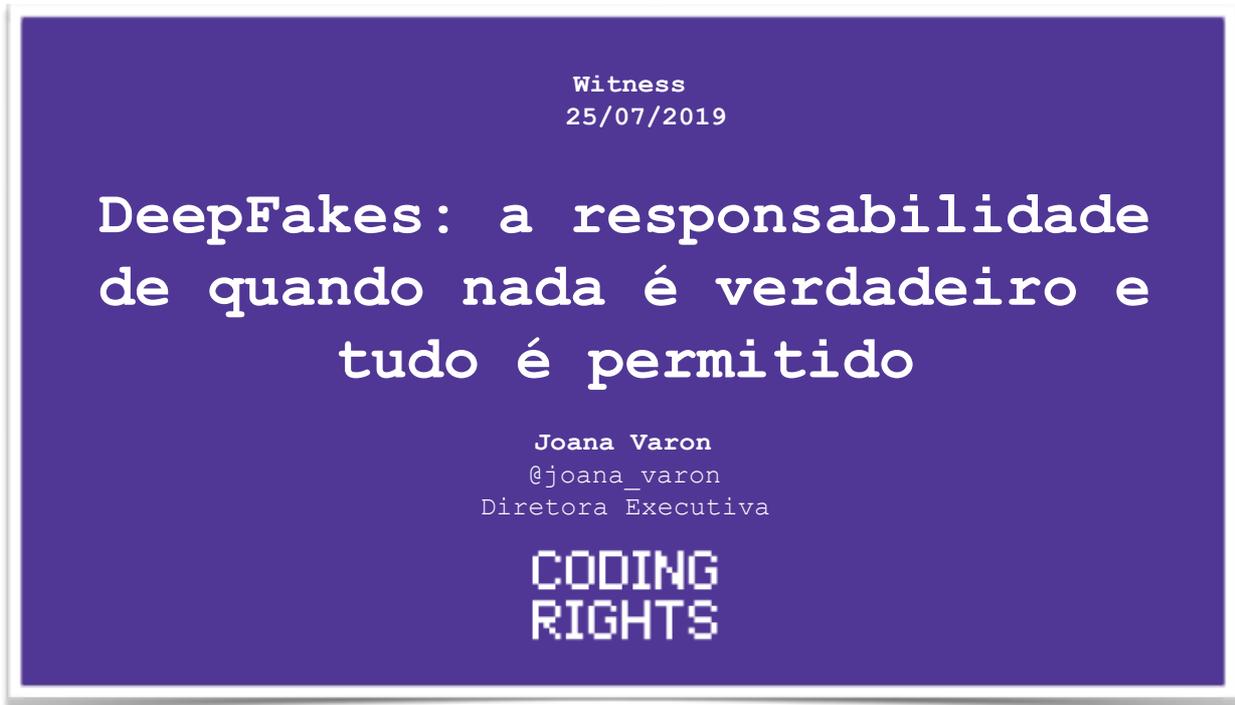
14.5. Qual deve ser o papel das plataformas e redes sociais?

Em um mundo em que a maioria do que consumimos provém de plataformas, quais são os papéis delas em deixar claro que um determinado conteúdo é falso ou verdadeiro? O caso do vídeo de Nancy Pelosi que teve sua velocidade reduzida para ela parecer bebada mostrou uma falta de vontade das plataformas sociais em remover conteúdo dessa natureza. Os mesmos padrões devem ser aplicados aos deepfakes? É importante estabelecer limites para garantir a proteção individual sem comprometer severamente a liberdade de expressão.

14.6. Como lutar contra os deepfakes?

No Brasil, a legislação que poderia ser usada para regular os deepfakes protegeria apenas contra deepfakes pornográficos, mas não contra vídeos com conteúdo falso apenas. Como o regulamento do Marco Civil da Internet é de 2016, ele foi discutido e criado em uma

época em que os efeitos da desinformação e notícias falsas ainda não eram claros e não possui previsão expressa. Esta legislação deve ser alterada? Quanto devemos permitir que os provedores de Internet moderem o conteúdo antes que as notícias se espalhem? Devemos criminalizar toda essa conduta? Estas são algumas questões importantes que devemos ter em mente.



Witness
25/07/2019

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

Joana Varon
@joana_varon
Diretora Executiva

**CODING
RIGHTS**

[para baixar os slides, [clique aqui](#)]

15. IMPACTOS PARA ATIVISTAS E COMUNIDADES CARENTES

Esta seção foi apresentada por Lana Souza, jornalista, moradora do Complexo do Alemão (um dos maiores complexos de favelas do Rio) e co-fundadora do Coletivo Papo Reto, um coletivo de comunicação independente que atua nas favelas do Rio e usa a comunicação como ferramenta para reafirmar direitos e disputar disputas sobre o que acontece nas favelas. Lana também participou da discussão inicial que a WITNESS convocou sobre esse tópico com ativistas comunitários no Rio de Janeiro há alguns meses atrás. Durante o encontro, ela compartilhou algumas de suas impressões sobre o assunto.



[[clique aqui](#) para assistir o vídeo da apresentação com os slides]

15.1. Como os deepfakes impactam moradores e ativistas de favela?

No início de 2019, a WITNESS realizou um workshop de deepfake no Rio, com ativistas de diferentes comunidades para melhor entender suas impressões sobre essa tecnologia. Medo, angústia e preocupações foram alguns dos principais sentimentos presentes naquele encontro, pois os participantes sabiam que, mais uma vez, sofrerão violência com o uso dessa tecnologia, mesmo que ainda demore algum tempo (hoje eles ainda estão lutando contra notícias falsas muito mais básicas do que deepfakes).

Nas favelas, ativistas e moradores realmente não se importam ainda com as questões técnicas dos deepfakes. Hoje, suas preocupações ainda são focadas em voltar para casa seguros, vivos e com sua reputação intacta. No entanto, em um mundo em que você pode

criar facilmente um vídeo falso, esse conteúdo pode ser usado para colocá-los em dificuldade com os traficantes locais, policiais e outros atores que podem até matar os ativistas por isso. Não serem reconhecidos em seu próprio território lhes dá menos acesso ao povo e menos condições na proteção de seus direitos.

15.2. Algum morador já sofreu consequências de "fake news"?

Raull, um ativista carioca, teve uma grande briga com o Facebook para verificar seu perfil, pois muitos perfis falsos tentavam prejudicar sua reputação dentro de seu próprio território, compartilhando notícias falsas em seu nome. Às vezes, essas notícias não são inteiramente falsas, mas são descontextualizadas. Às vezes, o conteúdo é apenas falso e eles precisam trabalhar duro para repreendê-lo.

15.3. Como podemos chamar a atenção sobre deepfakes nas favelas?

Comunicar o que é um deepfake às pessoas nas favelas pode ser complicado. O termo "fakenews" já era complexo para eles entenderem. Usar humor e arte pode ser uma maneira de apresentar esses conteúdos mais amigáveis, além de criar produtos que "falam o idioma deles". Avaliar também constantemente o impacto da intervenção de alguém nos territórios menos privilegiados e o que eles estão contribuindo para criar em termos de narrativas políticas e de segurança pública, além de garantir a inclusão desses espaços nas discussões é mais do que necessário.

16. LUTANDO CONTRA OS PROBLEMAS DOS DEEPFAKES

O workshop também pretendeu que os participantes contribuíssem ativamente ao tema, reconhecendo a importância de priorizar a situação no Brasil, tanto em termos de ameaças de deepfakes quanto nas soluções que estão sendo desenvolvidas em resposta a deepfakes maliciosos.

Para tanto, foram propostos dois exercícios: um sobre ameaças que os participantes priorizaram como importantes ou significativos e outro sobre soluções, mas antes disso, Sam deu uma introdução ao problema.



[\[clique aqui\]](#) para assistir o vídeo da apresentação com os slides]

16.1.O que NÃO devemos fazer na luta contra os problemas dos deepfakes?

Muitas pessoas dizem que detectar uma farsa profunda é fácil porque elas não piscam. Esta é uma conclusão feita com base em uma pesquisa amplamente divulgada e publicada há um ano atrás. No entanto, duas semanas após a publicação do artigo, seu autor recebeu um deepfake que piscava, resultado de um trabalho duro para desenvolver uma abordagem que eliminasse essa fraqueza no algoritmo e nos dados de treinamento. No entanto, "deepfakes não piscam" ainda é citado constantemente por pessoas, pois essa informação colou.

Isso nos traz uma lição importante: não devemos nos concentrar no algoritmo "calcanhar de Aquiles" do momento, porque a tecnologia se desenvolve rapidamente e o que é verdade agora pode não ser daqui a 6 meses - mas mesmo as pessoas ainda podem se lembrar como verdade e isso pode induzi-las a acreditar que um vídeo falso é realmente verdade. Isso é particularmente relevante com os deepfakes, que são desenvolvidos por meio de processos contraditórios que são inerentemente estruturados em torno de um processo de competição para melhorar a qualidade da falsificação.

16.2.O que devemos fazer na luta contra os problemas dos deepfakes?

Mais do que falar sobre o atual elemento "calcanhar de Aquiles" de qualquer processo atual de criação de deepfake, devemos trabalhar para criar um pensamento crítico que possa levar as pessoas a duvidar de materiais tendenciosos, verificar fontes, proveniência e confirmação das informações e procurar sua veracidade antes de acreditar e compartilhá-la. O papel de Youtubers e influenciadores de mídia social (principalmente no Brasil) pode ser fundamental para a educação e "popularização" do tópico. Além disso, o humor é uma ferramenta importante para isso. Trabalhar em conjunto com outros projetos, iniciativas, coalizões e agências de verificação de fatos existentes é muito importante não apenas para compartilhar ferramentas e habilidades, mas também para trocar experiências e novas tecnologias. Outros métodos incluem a criação de intervenções invisíveis (para seres humanos) em imagens - as chamadas 'perturbações contraditórias' - para enganar o processo de visão por computador e evitar que seus dados sejam usados para fins de treinamento em IA.

16.3.O que devemos discutir antes de criar uma solução?

A tecnologia Blockchain combinada com outros mecanismos para autenticar a fonte por meio da chamada "captura verificada" é frequentemente citada como uma potencial solução para facilmente identificar um deepfake, rastreando a origem de imagens e vídeos para verificar se ocorreu alguma manipulação. No entanto, uma mudança fundamental na maneira como avaliamos a veracidade dos vídeos coloca questões importantes sobre em quem devemos confiar, quais ferramentas devem ser permitidas ou necessárias para usar e quais são os riscos dessa vigilância, do acesso a dados e da privacidade. As implicações dessas questões de infraestrutura técnica terão implicações profundas na maneira como pensamos sobre questões de confiança em nossa sociedade. A WITNESS terá um relatório futuro sobre esta área.

Plataformas, aplicativos de mensagens fechadas, mecanismos de pesquisa, redes de mídia social e sites de compartilhamento de vídeo são os locais virtuais onde essas manipulações são compartilhadas. Alguns tópicos e perguntas que devemos discutir são: Qual o papel que as redes sociais e outras plataformas devem desempenhar no combate a falsificações profundas? Quais devem ser os limites? Como eles devem fornecer acesso aos recursos de detecção? Como eles devem sinalizar aos usuários que o conteúdo é verdadeiro ou falso ou, de alguma forma, manipulado de maneiras que não podem ver? Eles devem remover certos tipos de conteúdo manipulado e sob quais critérios? Há muitas perguntas nessa área

que devem ser abordadas antes de se chegar a conclusões finais sobre como lidar com este tópico.

16.4. Em suma, o que podemos fazer agora?

Principais perspectivas sugeridas na apresentação:

- ▶ Desescalar a retórica em torno de deepfakes e procurar por soluções;
- ▶ Identificar modelos e soluções para ameaças a partir de uma perspectiva global e com base nas experiências de pessoas que já enfrentaram ameaças de desinformação / desinformação;
- ▶ Promover abordagens interdisciplinares e soluções múltiplas;
- ▶ Desenvolver experiências passadas e conhecimentos existentes, principalmente relacionadas a alfabetização midiática, investigação open-source e jornalismo.
- ▶ Determinar o que queremos e o que não queremos de plataformas e empresas que comercializam ferramentas de criação ou gerenciam o fluxo de mídia sintética.
- ▶ Destacar claramente os prós e contras das opções de infraestrutura tecnológica, especialmente aquelas que terão impacto na confiança do público em imagens e vídeos.

Para obter mais detalhes sobre possíveis soluções, incluindo estas áreas abaixo, consulte o documento abaixo:

1. Como podemos ensinar as pessoas a identificá-las?
2. Como podemos construir em cima da capacidade e coordenação jornalísticas existentes?
3. Existem ferramentas para detecção? (e quem tem acesso?)
4. Existem ferramentas para autenticação? (e quem é excluído?)
5. Existem ferramentas para ocultar nossas imagens para serem usadas como dados de treinamento?
6. O que queremos das empresas que produzem ferramentas de síntese?

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

WITNESS SEE IT
FILM IT
CHANGE IT

7.O que as plataformas e os legisladores devem fazer?

[para baixar os slides, [clique aqui](#). Ou leia mais aqui (apenas em inglês): <https://blog.witness.org/2019/06/deepfakes-synthetic-media-updated-survey-solutions-malicious-usages/>]

17. AMEAÇAS RELEVANTES E POSSÍVEIS SOLUÇÕES

A WITNESS mapeou uma série de ameaças identificadas em workshops prévios realizados por ela. A primeira parte do exercício consistiu em uma votação com adesivo de ponto sobre essas ameaças existentes coladas na parede. As ameaças potenciais foram agrupadas de acordo com o tipo de problemas, como "ataques a jornalistas e ativistas" ou "ataques que estão levando a uma sociedade de confiança zero, entre outros.

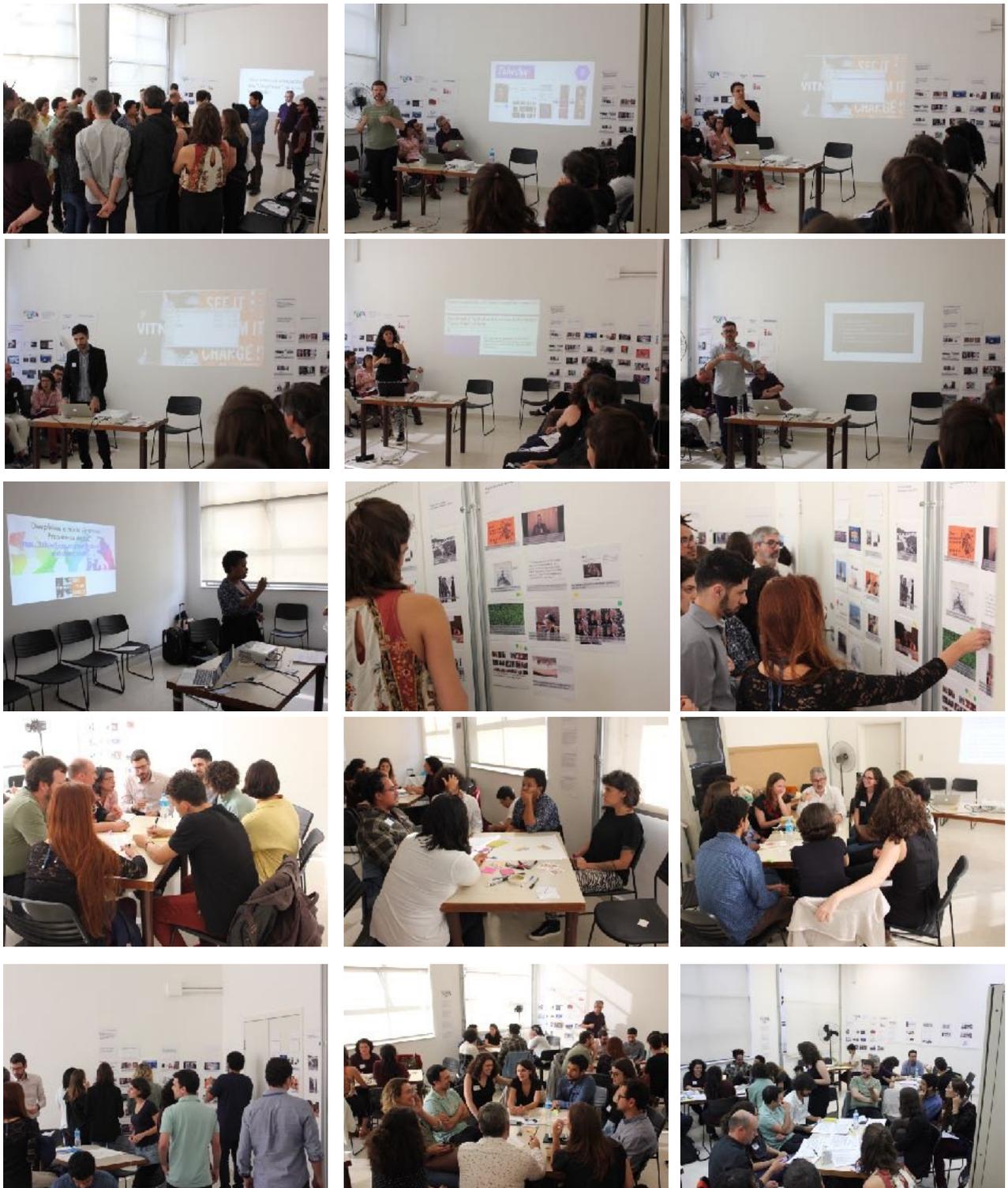
17.1. Quais são algumas das ameaças mais relevantes?

Os participantes identificaram as seguintes ameaças como as mais relevantes para eles (votadas pelos participantes a partir de um rol fixo de possibilidades):



- ▶ Ataques contra a reputação e credibilidade de jornalistas e ativistas da sociedade civil com ataques mais aprimorados (geralmente com base no gênero) e falsificações mais sofisticadas.
- ▶ Imagens sexuais não consensuais geradas por meio de IA, violência de gênero contra figuras públicas e bullying.
- ▶ Ataques à credibilidade / segurança e narrativas públicas de movimentos sociais e líderes destes movimentos, bem como ataques a grupos marginalizados, por meio de falsas alegações sobre seus comportamentos / ações.
- ▶ Ataques contra processos judiciais e o valor probatório do vídeo, uma vez que o vídeo se torna desacreditado ou nos processos tem-se o ônus de se provar se ele é verdadeiro ou falso.
- ▶ Integração em campanhas conspiratórias.
- ▶ Criação cumulativa de desconfiança nas instituições.

- ▶ Uso de micro-segmentação de conteúdo gerado por IA cada vez mais personalizada, inclusive por figuras políticas.
- ▶ Mangueira da falsidade (firehose of falsehood): quando se tem muito conteúdo falso, não se consegue saber o que é verdade e os sistemas sociais, os processos de verificação de mídia e os mecanismos de busca tornam-se sobrecarregados.



17.2. Quais são algumas das ameaças LOCAIS mais relevantes?

Os participantes foram convidados a se dividir em quatro grupos de trabalho diferentes e selecionar um deles para ser relator de cada um.

- ▶ Grupos de base e comunitários (1)
- ▶ Organizações de mídia e verificadores de fatos (2)
- ▶ ONGs / defensores de direitos humanos (3)
- ▶ Especialistas em tecnologia (4)

Eles tiveram tempo para discutir em grupo as potenciais ameaças em suas áreas e compartilhar algumas delas com todos os participantes (e também colá-las na parede ao lado de uma ameaça semelhante). As principais perguntas foram:

- ▶ De que forma as novas formas de manipulação expandem as ameaças existentes, introduzem novas ameaças, alteram ameaças existentes ou reforçam outras ameaças?
- ▶ Como poderíamos estar preparados para isso?



Abaixo, você encontra todas as ameaças organizadas por tópico, mencionando qual grupo relatou cada um. Cada tópico diz respeito a um interesse diferente:

A. Ameaças à segurança pessoal:

- ▶ Falsas evidências: Prisão de líderes locais com base em evidências falsas de vídeos deepfake. (1)

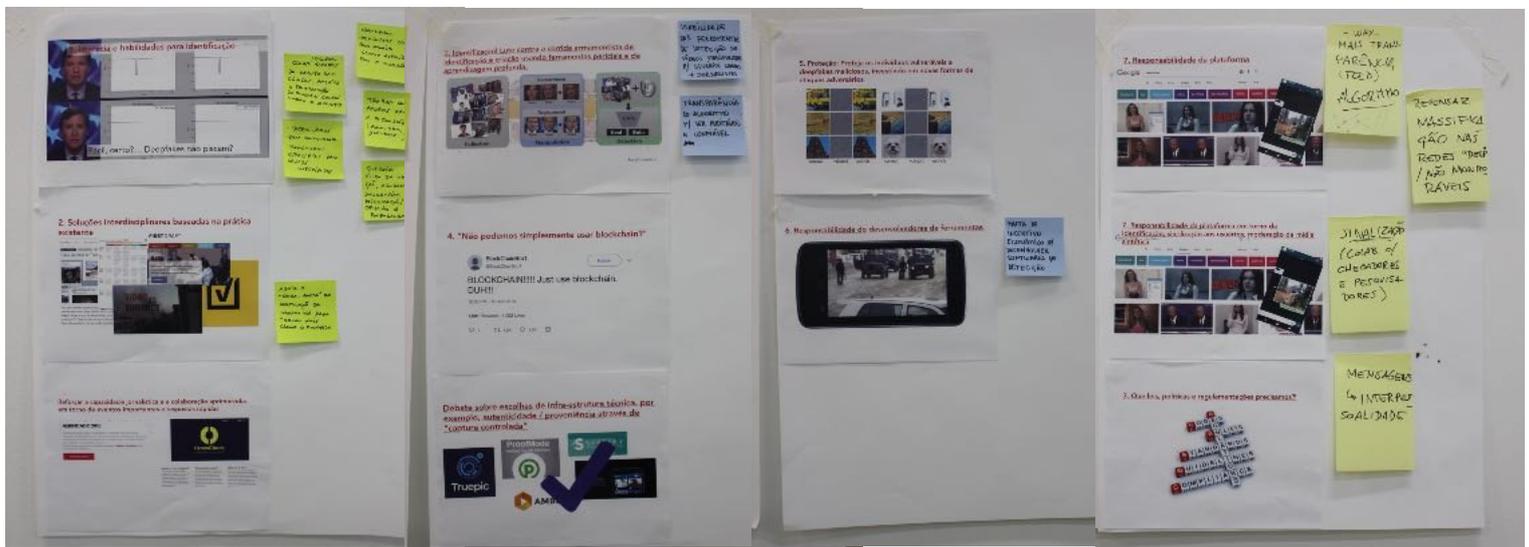
B. Ameaças ao trabalho de ativistas, jornalistas e verificadores de fatos:

- ▶ Credibilidade dos ativistas cívicos: perda da credibilidade dos ativistas devido a vídeos ou conteúdo falsos que prejudicam gravemente o trabalho, a mobilização da comunidade e até sua segurança pessoal. (1)
- ▶ Vídeo como prova: o valor do vídeo como prova será comprometido. (1)
- ▶ Credibilidade da mídia: risco da imprensa e a mídia perderem a credibilidade se vídeos não puderem mais ser usados como prova (para avaliar a verdade e notícias). (1, 2)
- ▶ Mídia sobrecarregada: distribuição em massa de vídeos falsos para inundar jornais e agências de verificação de fatos com o conteúdo que precisa ser verificado, tornando as pessoas incapazes de identificar o que é verdade ou não. 2)
- ▶ Manipulação de imagens de arquivo: usando vídeos deepfakes com baseados eventos passados para caracterizar eventos atuais (por exemplo, manifestantes com proposta distintas) (3, 4)

C. Dificuldades para detecção:

- ▶ Velocidade: haverá dificuldades para detectar e bloquear fakenews, rumores desonestos ou reivindicar "notícias de última hora" com tecnologias deepfake (especialmente se essa detecção precisar ocorrer em tempo real). (4)
- ▶ Acesso: notícias falsas e vídeos manipulados são frequentemente distribuídos em redes fechadas e em mensagens micro-segmentadas às quais os verificadores de fatos muitas vezes não têm acesso direto. 4)
- ▶ Compreensão de como detectar: Muitas pessoas são leigas em tecnologia e não podem executar nenhum procedimento complexo de verificação de fatos; eles simplesmente acreditam no conteúdo sem questionar e os espalham facilmente. (1)
- ▶ Falta de compreensão do que é possível: as pessoas podem ser enganadas com mais facilidade por vídeos falsos, pois ainda não sabem se essa tecnologia pode ser manipulada (ao contrário das fotos, em que a maioria está familiarizada com a ideia de que pode ser manipulada). (3,4)
- ▶ Falta de recursos: os processos de detecção são caros, pois é difícil treinar e financiar equipes para usar ferramentas para realizar essas verificações (e os jornalistas geralmente não são especialistas em TI). 2)
- ▶ Uso de áudio deepfake: a discussão sobre o uso de áudio é limitada, que pode ser altamente eficaz e difícil de detectar.

D. Uso incorreto de deepfakes em uma escala mais ampla:



- ▶ Contribuição para os problemas existentes numa "sociedade de confiança zero": os deepfakes podem ser uma ferramenta poderosa de desinformação em uma "sociedade de confiança zero", onde a verdade é substituída pela opinião. (3)
- ▶ Complementa a micro-segmentação: os riscos associados à tecnologia por trás do deepfake vão além da criação de conteúdo falso: com a IA, você pode descobrir o perfil psicológico de uma pessoa ou de um grupo e usá-lo para realizar uma segmentação muito eficaz com conteúdo falso, a fim de reforçar posição ou opinião existente. (4)

E. Quem está em risco com os deepfakes:

- ▶ Segurança dos dados e imagem pessoal: riscos aumentados de que as pessoas "roubem" e usem dados pessoais para melhorar as IAs, a fim de criar vídeos manipulados de indivíduos. (3)

17.3. Quais são soluções possíveis?

Para o exercício de soluções, os participantes foram divididos em 3 grupos. Cada grupo teve uma ou duas perguntas norteadoras:

- ▶ Educar o público / "alfabetização" midiática
- ▶ 'Colaboração / coordenação necessárias' + 'habilidades e ferramentas jornalísticas necessárias' + 'o que é necessário nas soluções de detecção e autenticação'
- ▶ O que as plataformas devem fazer?

Os participantes tiveram tempo para discutir possíveis soluções para cada pergunta em seu grupo e depois compartilhar em voz alta as soluções mais interessantes. As principais perguntas norteadoras foram:

- ▶ Quais soluções parecem mais relevantes para você? O que você precisa dessas soluções para torná-las viáveis? Quais seriam as formas de abordar isso no Brasil? O que o preocupa com essas soluções?
- ▶ Qual seria o próximo passo concreto nesta área que você gostaria de ver?

A. Como compartilhar conhecimento sobre deepfakes e desinformação?

- ▶ Ensinar ética da informação: discutir a ética da informação para ajudar as pessoas a diferenciar o que é informação do que é opinião ou propaganda. (1)
- ▶ Mostrar como narrativas podem ser usadas para enganar: abrir a caixa preta da construção de narrativas: mostrar como as pessoas e os grupos constroem narrativas para manipular a opinião pública e como a micro-segmentação está sendo usada atualmente para isso, bem como acompanhamento de suas estratégias e fórmulas. (1)

B. Como melhorar a comunicação, principalmente com as comunidades de base?

- ▶ Criar "espaços de escuta": ocupe e crie "espaços de escuta", espaços para ouvir o que as pessoas querem saber ou já sabem sobre deepfakes. (1)
- ▶ Recorrer a outros influenciadores: trazer outros atores para essa discussão, como as garotas do Slam ou Funk ou YouTubers e não deixar toda a responsabilidade para jornalistas e comunicadores. (1)
- ▶ Criar mensagens customizadas: identificar e criar mensagens relevantes para atrair grupos específicos para a discussão. (1)
- ▶ Grupos de micro-targeting: usar o micro-direcionamento para comunicar mensagens específicas a grupos específicos: por exemplo, se queremos conversar com adolescentes, precisamos usar um certo tipo de idioma e canais. (1)

C. Como criar e disseminar ferramentas?

- ▶ Criar ferramentas e procedimentos críveis e explicáveis: criar ferramentas de detecção claras, transparentes e confiáveis, para que as pessoas possam entender todas as etapas do método de verificação e auditar seus resultados e tornar essas ferramentas ainda mais críveis (2).
- ▶ Desenvolver ferramentas acessíveis: reunir ferramentas para jornalistas e cidadãos para verificar a autenticidade de vídeos deepfake (apesar da falta de incentivos econômicos para disponibilizar essas ferramentas). (2)

- ▶ Questionar incentivos econômicos: questionar os incentivos econômicos para criação mas não para detecção (2).
- ▶ Monitorar deepfakes e notícias duvidosas: usar painéis de métricas de engajamento (como o CrowdCompass) para identificar notícias duvidosas e verificá-las antes que elas se tornem virais. (3)

D. O que as plataformas devem fazer?

- ▶ Tornar o que/porque está no seu feed mais transparente: ter um feed mais transparente nas redes sociais, para facilitar a detecção de distribuição de notícias falsas. (3)
- ▶ Repensar as mensagens em redes fechadas: repensar como alcançar pessoas deve funcionar em redes privadas não monitoradas como o WhatsApp, discutindo o limite do número de participantes em grupos, o papel dos bots etc. (3)

18. RESUMINDO

A reunião foi muito produtiva e os participantes compreenderam bem como deepfakes são feitos, como podem ser detectados e como podemos aumentar a conscientização sobre eles e nos prepararmos para um potencial uso desta tecnologia na criação de notícias falsas e desinformação no futuro.

As **AMEAÇAS** destacadas incluem ameaças à segurança pessoal, como prisão de líderes locais; ao trabalho de ativistas, jornalistas e verificadores de fatos, principalmente devido à perda de credibilidade; aumento das dificuldades de detecção em um jogo de gato e rato e num contexto em que as pessoas têm "alfabetização digital" muito baixa; mau uso de falsificações profundas para criar notícias falsas ou um cenário em que você não pode identificar o que é verdadeiro ou não, além de riscos associados à tecnologia relacionada que vão além do vídeo (como o uso como conteúdo micro-direcionado).

As **SOLUÇÕES** propostas giraram em torno de compartilhar conhecimento, particularmente com comunidades mais populares, melhorando a comunicação entre os atores e a sociedade; criar e disseminar ferramentas acessíveis e de fácil uso, e; promover mudanças nos feeds das plataformas de redes sociais e discutir a abordagem sobre disseminação em redes fechadas de mensagens.

19. CRÉDITOS

Este relatório foi criado por Bruno Paschoal, com a ajuda de Sam Gregory.

Todas as fotos, vídeos e materiais neste relatório são licenciados sob Licença Creative Commons CC BY-SA. Para mais materiais, contatos ou informações, entre em contato com a WITNESS.