WITNESS

SEE IT FILM IT CHANGE IT

# RELATÓRIO[1]

## Fortalecendo a verdade na era das mídias sintéticas no Brasil

**ESPM TECH, São Paulo, Brasil, 8-9 de maio de 2024**

[1] Escrito por Caroline Burle, com a colaboração de: Jacobo Castellanos, Inês Menezes, Jessica Ribeiro e Leandro Ramos.

*Dando continuidade às conversas das reuniões 'Deepfakes: Prepare Now' de 2019 a 2021, realizadas na Europa, Brasil, África Subsaariana, Sudeste Asiático e EUA, a WITNESS realizou o workshop "Fortalecendo a verdade na era das mídias sintéticas no Brasil", nos dias 8 e 9 de maio, em São Paulo, na ESPM Tech.*

*O intuito do workshop foi explorar o impacto que as tecnologias para manipular vídeos e áudios utilizando inteligência artificial têm, ou podem potencialmente ter, nos conteúdos criados por defensores dos direitos humanos, meios de comunicação cívicos e comunitários, ativistas e comunidades marginalizadas; identificar e priorizar soluções pragmáticas para preparação e defesa contra um futuro potencialmente perigoso de vídeos e áudios manipulados usando inteligência artificial, incluindo o uso de ferramentas e serviços de proveniência e autenticidade; e explorar o potencial criativo da mídia sintética, da IA generativa e dos deepfakes para promover os direitos humanos e o jornalismo independente.*

*O encontro seguiu a regra da Chatham House, com a opção de contribuições "em off" se solicitadas pela pessoa participante. Pessoas participantes são livres para utilizar as informações que ouvirem durante a reunião, mas a identidade ou afiliação da pessoa palestrante, ou de qualquer outra pessoa participante, deverá ser preservada. No entanto, observe que a WITNESS gostaria de mencionar a sua participação neste workshop em materiais públicos; Por favor, deixe-nos saber se você tiver alguma dúvida sobre isso. Se usarmos citações diretas em nossos relatórios públicos, entraremos em contato com você antes de publicá-los.*

# Sumário

# DIA 1 - O que são os meios sintéticos e que relevância têm para nós?

## Introdução WITNESS & ‘Fortalecer a verdade’

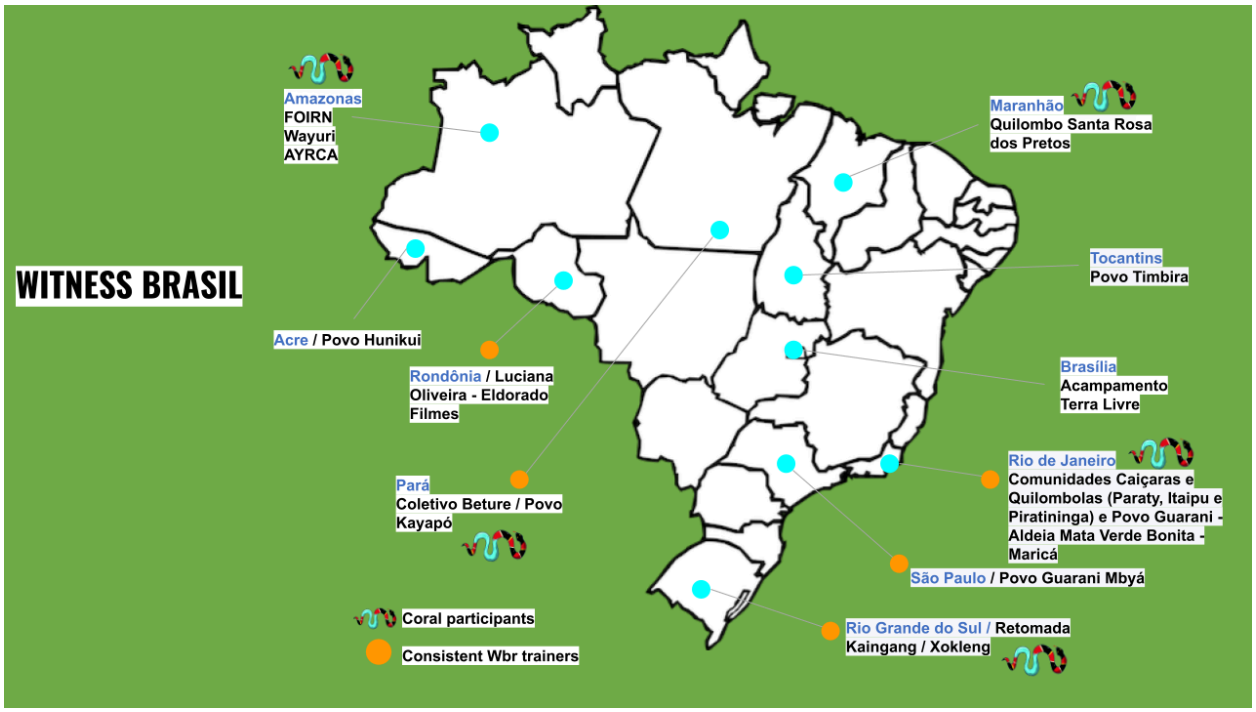
A Witness foi fundada entre 1991 e 1992 e trabalha com pessoas e movimentos para apoiar o uso de vídeo e tecnologia na defesa dos direitos humanos. No início da década de 1990, a questão dos vídeos e equipamentos pessoais e acessíveis estavam tornando-se mais recorrentes. Quando o Rodney King foi agredido brutalmente pela polícia e foi filmado por um morador com a câmara Sony recém comprada, teve início uma onda de protestos nos Estados Unidos. Começou-se, então, a utilizar vídeos como uma ferramenta de direito.

Atualmente, a Witness está nos Estados Unidos, na América Latina e no Brasil, no Oriente Médio e no norte da África, no Pacífico Asiático e na África Subsahariana. Realiza treinamentos, oficinas que reúne comunicadores, ativistas e especialistas. As metodologias criadas pela Witness são acessíveis e estão disponíveis no site da Witness.


A WITNESS está há 30 anos apoiando comunidades e organizações no uso de vídeo e tecnologia de alto risco e alto interesse público, criando conteúdo confiável para o jornalismo cívico e os direitos humanos em todo o mundo. Há 10 anos interagindo com plataformas sobre como elas apoiam usuários e conteúdos de alto interesse público e trabalhando com grandes

volumes de mídia manipulada. E há 6 anos trabalhando em ‘Prepare-se, Não Entre em Pânico’ em torno de novas formas de manipulação como deepfakes.

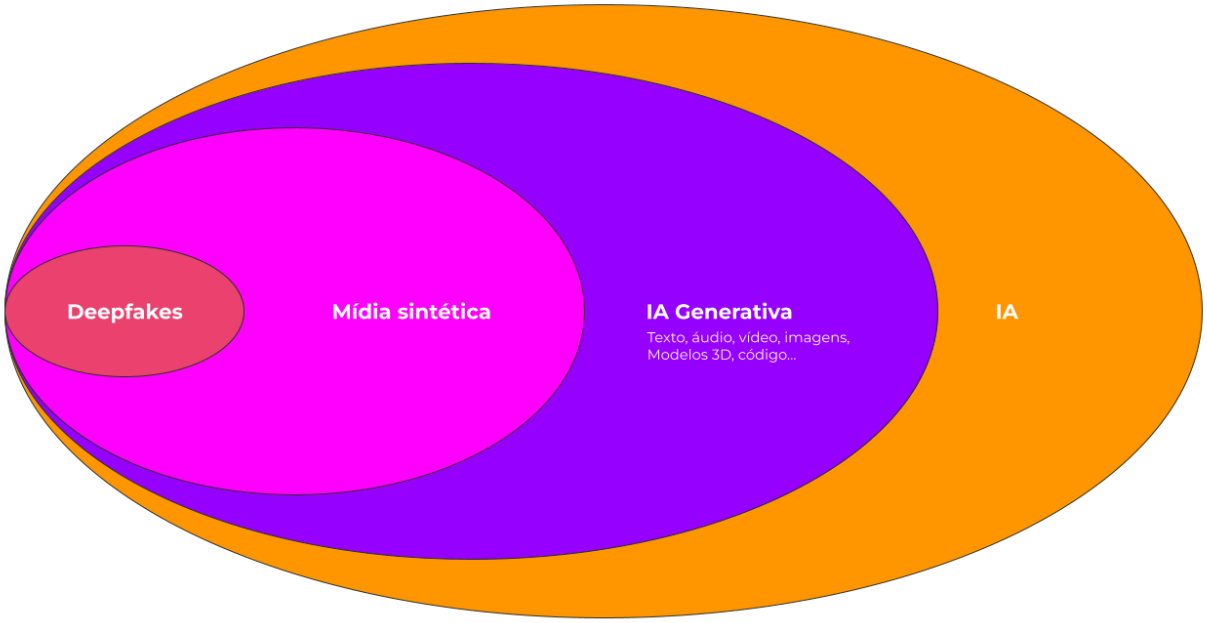
Mapa da Witness no Brasil




Diversos programas estão em andamento, incluindo iniciativas de memória e preservação, justiça climática e o uso de vídeo para a verificação e tecnologia aplicada à justiça. Entre os programas temáticos, destacam-se a manutenção de arquivos e a produção de vídeos como prova documental. Uma atividade notável ocorreu no Quilombo do Bracuí, com a realização de um vídeo sobre o trabalho em Angra dos Reis, abordando temas audiovisuais e investigações subaquáticas sobre a escravidão no Brasil. Enfrentam desafios como a crescente importância do vídeo e da experiência audiovisual na sociedade, enquanto sua integridade é subestimada em meio a uma crise de confiança e credibilidade. Há uma necessidade urgente de "fortalecer a verdade" para defensores dos direitos humanos, jornalistas e a sociedade civil. Olhando para o futuro, é essencial reforçar a veracidade das imagens e vídeos, pois, apesar de serem um poderoso meio de comunicação, a desconfiança em relação ao conteúdo audiovisual tem aumentado.

## Mídias Sintéticas, Democracia e Eleições

Explicação sobre o que são as mídias sintéticas e a IA generativa, como funcionam, qual o estado atual e qual a expectativa para o futuro.




As mídias sintéticas e a IA generativa representam um avanço significativo na criação de conteúdo digital. Mídias sintéticas referem-se a conteúdos gerados ou manipulados por inteligência artificial, enquanto a IA generativa é a tecnologia que permite a criação desses conteúdos a partir de dados existentes. A IA generativa utiliza redes neurais avançadas para produzir texto, imagem, vídeo, modelos 3D e áudio. Atualmente, essas tecnologias estão em rápido desenvolvimento, com crescente acessibilidade e sofisticação, e espera-se que continuem a evoluir, impactando diversos setores.

As técnicas da IA generativa incluem a capacidade de transformar texto em texto, imagem, vídeo, modelo 3D e áudio. Também é possível converter imagem em imagem e vídeo, além de transformar áudio em áudio e fala. Vídeos podem ser transformados em texto e outros vídeos. Exemplos notáveis incluem a geração de imagens e vídeos a partir de texto, e a criação de novas imagens a partir de imagens existentes. Essas capacidades permitem a geração e estilização de praticamente qualquer tipo de conteúdo, a criação de deepfakes, dublagens com sincronização labial em diversos idiomas, clonagem de voz, edição de recursos faciais, transferência de movimento e expressão, remoção e adição de objetos em vídeos e imagens, além de adicionar contexto a esses conteúdos.

Entre 2019 e 2024, as principais mudanças no desenvolvimento da IA generativa audiovisual incluem a comercialização e comoditização das tecnologias, facilitando seu uso. Houve um aumento no volume e na variação dos conteúdos gerados, especialmente em áudio e imagens, além de uma maior personalização e personificação. A multimodalidade, que permite a interação entre diferentes tipos de conteúdo, e a interação ao vivo com deepfakes também se destacam. Exemplos incluem deepfakes usados para satirizar políticos, campanhas políticas

com conteúdo gerado por IA e a criação de avatares de IA para humanizar ou ridicularizar figuras públicas.

Em 2024, ainda reconhecemos a importância de nos preparar e agir conscientemente diante da mídia gerada por IA, evitando o pânico. A mídia gerada e modificada por IA pode ser usada para atacar, comunicar, enganar, negar, humanizar, ressuscitar e ridicularizar. Exemplos de uso incluem a voz de IA de um presidente declarando vitória, campanhas políticas com anúncios gerados por IA, deepfakes satíricos, e conteúdo manipulado para fortalecer narrativas enganosas. O futuro aponta para um aumento no volume, qualidade e personalização do conteúdo gerado por IA, contribuindo potencialmente para um declínio da confiança pública geral.

## IA generativa e mídia sintética no Brasil

A Inteligência Artificial (IA) é um campo de estudo multidisciplinar que abrange diversas áreas do conhecimento, buscando desenvolver sistemas que pensam ou atuam como seres humanos ou de forma racional. Existem várias abordagens, como os sistemas que pensam como seres humanos (Haugland, 1985), sistemas que atuam como seres humanos (Kurzweil, 1990), sistemas que pensam racionalmente (Charniak; McDermott, 1985) e sistemas que atuam racionalmente (Poole et al., 1998).

Redes Neurais são modelos computacionais inspirados no sistema nervoso central dos animais, especialmente o cérebro, capazes de realizar aprendizado de máquina e reconhecimento de padrões. Entre essas, as Redes Adversárias Generativas (GANs) se destacam por serem compostas por duas redes neurais que se contrapõem: uma geradora, que cria novas instâncias de dados, e uma discriminadora, que avalia a autenticidade dessas instâncias, determinando se pertencem ou não ao conjunto de dados de treinamento original.

A criação de vídeos deepfake envolve um banco de dados para reconhecimento e mapeamento do rosto e o treinamento das redes neurais para aprendizado. Essas práticas levantam diversas questões legais, como o direito à personalidade e à imagem, conforme estabelecido no Código Civil, e a proteção contra a exposição da intimidade sexual, de acordo com o Código Penal. Esses direitos visam proteger a honra, boa fama e respeitabilidade das pessoas, inclusive após a morte.

O futuro da IA é promissor em diversas áreas como comunicação, entretenimento e saúde, com exemplos práticos incluindo dublagem automática, geração de imagens a partir de texto (Stable Diffusion/Midjourney), influenciadores artificiais (Fláv.ia Martini), e a criação de vídeos e músicas a partir de texto (Sora da OpenAI e Suno.AI). No entanto, a IA também pode ser usada de forma criminosa, como na produção de pornografia, disseminação de fake news e golpes. Tecnologias como a fotopletismografia, que detecta mudanças no volume sanguíneo em imagens, são desenvolvidas para combater fraudes, determinando a autenticidade de vídeos.

## ‘Espectrograma’ de riscos

A verificação de conteúdos enfrenta desafios gerais como dados, horas ou locais descontextualizados ou inexistentes, conteúdos editados e conteúdos fabricados. As áreas temáticas afetadas incluem a Covid-19, direitos do meio ambiente, culturais e sociais, ataques à sociedade civil e à mídia, e ataques a mulheres e minorias.

No contexto das eleições, a IA tem sido utilizada para comunicação e campanhas políticas. Exemplos incluem a voz de IA e avatares realistas, como o avatar do presidente da Coreia do Sul, e a voz de IA de Khan declarando vitória na prisão durante a eleição de 2024 no Paquistão. Também há imagens de IA gerando campanhas de anúncios, como nas eleições na Argentina, e imagens de K. Chandrashekhar Rao com agricultores antes das eleições em Telangana.

IA também é usada para criar "softfakes", como o áudio lúdico de Modi cantando uma música de Bollywood, e deepfakes para ressuscitar figuras como o ex-presidente Suharto na Indonésia. Avatares de IA foram usados para representar candidatos indonésios Prabowo e Gibran. Para ridicularizar, deepfakes satíricos foram utilizados nas eleições de Taiwan, e uma visão do futuro com Biden declarando um alistamento militar para a Ucrânia.

Para fortalecer narrativas, deepfakes de áudio mostraram erros relacionados à idade de Biden, e imagens de IA mostraram eleitores negros apoiando Trump antes das eleições nos EUA. Para enganar e atacar candidatos, deepfakes foram usados nas eleições de Bangladesh, no Reino Unido e no Paquistão, com áudio falso de Biden pedindo que seus seguidores não votem. Em operações de influência, avatares de IA e vídeos em massa foram usados para espalhar rumores e propaganda, como no caso da Venezuela e de Taiwan. A desculpa de ser ou não ser IA tem sido usada para escapar da responsabilidade, destacando a necessidade de sempre verificar a fonte dos clipes de áudio.

O panorama do ecossistema informativo inclui a comercialização, facilidade de uso e acessibilidade das ferramentas de IA, aumento no volume e variação de conteúdos, e a multimodalidade. A personalização e a interação ao vivo com deepfakes são tendências emergentes que aumentam a complexidade da verificação e a necessidade de novas abordagens para garantir a integridade da informação.

Considerando que as áreas temáticas em que a IA e as mídias sintéticas têm impacto incluem saúde pública, direitos à terra e outros direitos econômicos, sociais e culturais (ESC), propaganda, ataques à sociedade civil e à mídia, ataques a mulheres e minorias, e eleições, foi realizado um exercício coletivo com a seguinte pergunta: De acordo com a sua experiência, qual é a gravidade do risco na sua área de trabalho?

- **Vermelho**: Maior grau de ameaça/dano potencial
- **Laranja**: Grau moderado de ameaça/dano potencial
- **Amarelo**: Grau mínimo alto de ameaça/dano potencial

A reflexão do grupo sobre as eleições é que os riscos são exemplificados pelo uso de deepfakes e outras técnicas de IA para enganar e manipular eleitores. A tecnologia pode ser usada para criar vídeos falsos de candidatos, disseminar fake news, ou mesmo criar áudios falsos para influenciar a opinião pública. Esses riscos são exacerbados pelo fácil acesso a ferramentas de IA e a possibilidade de criar conteúdos convincentes que podem se espalhar rapidamente pelas redes sociais.

Neste contexto, um exemplo é a pesquisa do IBOPE que mostrava Bolsonaro à frente de Lula, levando pessoas a usar vídeos para questionar o sistema eleitoral brasileiro. O acesso à internet na periferia é majoritariamente por internet móvel, com plataformas como Facebook e WhatsApp sendo acessadas mesmo sem pacotes de dados. Isso ilustra como a tecnologia pode ser usada tanto para bem quanto para mal, dependendo da intencionalidade. É crucial identificar os interesses econômicos por trás do uso de deepfakes e outras tecnologias de IA, e implementar medidas preventivas e mitigadoras para proteger a integridade dos processos eleitorais.

O panorama do ecossistema informativo atual inclui a comercialização e acessibilidade das tecnologias de IA, aumento do volume e variação de conteúdos, e a multimodalidade, que permite a combinação de diferentes tipos de mídias. Há um aumento da personalização e da interação ao vivo com deepfakes, o que torna a verificação de informações ainda mais desafiadora.

Os desafios gerais de verificação incluem a dificuldade em identificar dados, horas ou locais descontextualizados ou inexistentes, conteúdos editados e conteúdos fabricados. Além disso, a perspectiva da periferia destaca que muitos riscos considerados menores para a classe média já foram experimentados intensamente nas comunidades periféricas, como o encarceramento em massa e a ameaça de aumento dessa população com o uso de IA. A pesquisa por nomes femininos frequentemente resulta em pornografia, enquanto nomes masculinos mostram usos criativos, evidenciando um viés na representação de mulheres.

## Perspectivas sobre IA Generativa e Mídia Sintética – Diálogos de Grupo

### Grupo Gênero

O Instituto AzMina é uma organização dedicada à defesa e promoção dos direitos das mulheres, desenvolvendo projetos que abordam questões de gênero de forma propositiva e impactante. A história do Instituto é marcada por iniciativas inovadoras que buscam a igualdade de gênero e o empoderamento feminino. Suas ações incluem campanhas de conscientização, produção de conteúdo educativo e projetos voltados para a proteção e o fortalecimento das mulheres.

No contexto da mídia sintética, o Instituto AzMina tem se dedicado a discutir os impactos negativos que essas tecnologias podem causar às vítimas, especialmente no caso de vídeos pornográficos. A disseminação de deepfakes e outras formas de mídia sintética pode intensificar problemas preexistentes de violência e abuso contra mulheres, pessoas negras, indígenas e outras minorias. A questão central não é a tecnologia em si, mas como ela potencializa as desigualdades e violências que já existem na sociedade.

Há um debate importante sobre a diferença entre uma imagem pornográfica produzida por IA e uma manipulada por Photoshop. A técnica utilizada pode variar, mas o foco deve estar no dano causado às vítimas. Ambas as tecnologias podem ser usadas para criar conteúdo falso que pode prejudicar a reputação, a segurança e o bem-estar das mulheres.

O letramento digital é crucial para combater esses problemas. As pessoas precisam estar cientes da existência da mídia sintética e entender como ela pode ser usada de forma maliciosa. A educação sobre essas tecnologias é fundamental para capacitar a sociedade a reconhecer e reagir adequadamente às ameaças que elas representam. O Instituto AzMina continua a trabalhar para informar e proteger as mulheres contra os perigos da mídia sintética, promovendo um uso responsável e ético da tecnologia.

### Grupo Meios comunitários e vídeo para a defesa da terra

A inteligência artificial (IA) está presente em nossa sociedade e, muitas vezes, precisamos encontrar soluções sem entender completamente como a tecnologia funciona. É crucial compreender como a IA existe, quais recursos estão sendo usados e onde está sendo desenvolvida. É importante considerar se, no futuro próximo, veremos pessoas negras, da periferia e indígenas discutindo sobre IA.

A utilização de deepfakes como provas de crime, como vídeos de câmeras acopladas em fardas policiais, levanta preocupações sobre a manipulação de evidências. A ansiedade climática também é um problema significativo. Boatos sobre desastres naturais, como a abertura ou colapso de barragens, podem causar pânico e levar a sérios problemas de saúde, como AVCs e infartos, decorrentes das fake news.

A rede de colaboração criada em encontros comunitários deve ser mantida e fortalecida. Precisamos de espaços de cocriação para verificar a veracidade das informações e conhecer as soluções existentes para replicá-las. O tempo para provar se uma história é real ou não é

limitado, e o cotidiano na periferia é tão difícil que as pessoas muitas vezes não temem o que não veem. Explicar que a IA pode causar danos significativos é um desafio, pois a tecnologia não elimina a responsabilidade humana. Alguém projetou e direcionou essa tecnologia.

Não podemos nos individualizar na luta contra a desinformação. A tecnologia deve estar em mãos que possam usá-la para o bem comum e para regionalizar questões locais. Por exemplo, imaginar um indígena na Times Square com pintura facial pode ser uma maneira de levar informações que as pessoas possam visualizar antes que problemas ocorram. Casos de celebridades que sofreram ataques à moral devido a deepfakes mostram que a periferia não tem os mesmos recursos para detectar falsificações. Portanto, a IA precisa ser democrática.

Os eixos principais são: quem controla a tecnologia, como ela pode ser usada para beneficiar as comunidades locais e a necessidade de fornecer informações e recursos para que todos possam se preparar e responder aos desafios futuros. A IA deve ser uma ferramenta para empoderar as comunidades e não para perpetuar desigualdades.

## Grupo Desinformação

A ansiedade tecnológica é uma preocupação crescente, especialmente no contexto das eleições e da disseminação de desinformação. A Resolução TSE 23732/2024 é um indicativo de regulação, mas existem limites e preocupações que precisam ser abordados. As próximas eleições serão dispersas e polarizadas, com diferentes mecanismos de disseminação de informações e desinformações.

Há uma necessidade urgente de um debate mais crítico focado no conteúdo e na produção dele, considerando a desinformação. As redes de desinformação são altamente profissionalizadas e bem financiadas, existindo desde 2014. Essas redes são capazes de produzir e distribuir conteúdo de forma eficiente, o que torna essencial uma abordagem estratégica para combatê-las.

Outro ponto de preocupação é a autocensura que pode impactar a liberdade de imprensa. Uma solução a longo prazo é a educação midiática, que visa desenvolver melhor as competências das pessoas para lidar com informações e desinformações. No entanto, a transparência é problemática, especialmente com a questão das APIs, que pode dificultar o avanço nesse campo.

A desinformação gerada pela IA não precisa ser perfeita para causar danos. Desde 2018, fake news mal feitas viralizam por outros motivos que não apenas a qualidade do conteúdo. O maior perigo da desinformação feita pela IA é criar ceticismo na realidade e poluir o debate público. Se as pessoas não têm noção do que é desinformação, podem rejeitar informações verdadeiras que não se alinham com suas diretrizes políticas. Nesse contexto, é crucial entender o papel das big techs e como lidar com a estrutura tecnológica que facilita a disseminação de desinformação.

## Privacidade e direitos digitais

A Conversa foi baseada em três perguntas:

1. Como podem os meios sintéticos ajudar a garantir os direitos digitais
2. Quais seriam os perigos associados à mídia sintética?
3. Qual é a perspectiva sobre parametrizar. Como seria uma forma de parametrizar para diferentes áreas de conhecimento, diferentes perspectivas para exemplos mais tangíveis?

Os meios sintéticos, como a inteligência artificial (IA) generativa e as mídias sintéticas, têm o potencial de garantir os direitos digitais de várias maneiras. Eles podem facilitar a tradução automática, tornando informações acessíveis a uma audiência mais ampla, atravessando barreiras linguísticas e culturais. Além disso, essas tecnologias podem aumentar a inclusão digital, proporcionando ferramentas que ajudam na comunicação de pessoas com deficiências, como a geração de texto a partir de áudio para deficientes auditivos. Esses benefícios, no entanto, precisam ser explorados e implementados de forma ética e responsável.

Os perigos da mídia sintética são significativos e multifacetados. A desinformação gerada por deepfakes e outros conteúdos sintéticos pode criar ceticismo em relação à realidade, dificultando a distinção entre informações verdadeiras e falsas. Além disso, há riscos associados à privacidade e à segurança, como a manipulação de imagens ou vídeos para fins maliciosos, incluindo a criação de pornografia não consensual e o ataque à reputação de indivíduos. Esses perigos são exacerbados pela falta de regulamentação adequada e pela possibilidade de as tecnologias serem utilizadas por atores mal-intencionados para enganar e manipular o público.
A parametrização dos riscos associados à mídia sintética requer uma abordagem multidisciplinar e contextual. É importante diferenciar entre "perigo" (a potencialidade de causar dano) e "risco" (a probabilidade de que o dano ocorra). Para diferentes áreas de conhecimento, os riscos podem variar. Por exemplo, no campo da saúde pública, a desinformação pode ter consequências graves para a segurança e o bem-estar das pessoas, enquanto no setor de entretenimento, os riscos podem estar mais relacionados à violação de direitos autorais e à manipulação de imagem.

A discussão sobre a regulação da mídia sintética envolveu a análise do arcabouço legal existente no Brasil. O país possui diversas leis setoriais que regulam a tecnologia e a proteção de dados pessoais, além da Lei Geral de Proteção de Dados (LGPD). O consenso é que a regulação é importante, mas deve ser acompanhada de uma estrutura robusta para garantir que as normas sejam cumpridas. É necessário envolver diversos atores da sociedade em um debate plural para evitar a captura regulatória por grandes empresas e lobbies.

A educação midiática é fundamental para capacitar a sociedade a lidar com as novas tecnologias e a desinformação. As autoridades e organizações devem promover o letramento digital, orientando a população sobre como identificar e combater a desinformação. Equipes multidisciplinares, incluindo profissionais de direito, tecnologia, comunicação e outros setores, são essenciais para abordar os desafios complexos apresentados pela mídia sintética.

O grupo entendeu que a implementação eficaz da regulação deve ir além da criação de leis, exigindo mecanismos de fiscalização e execução robustos. A educação e a capacitação contínuas da sociedade são essenciais para mitigar os riscos associados à mídia sintética. O debate sobre esses temas deve ser inclusivo e plural, envolvendo diferentes setores e perspectivas para criar soluções equilibradas e eficazes.

## DIA 2 - Plano de Ação: Fortalecendo a Verdade na Era da Mídia Sintética

## Resumo de Riscos - *Framing do dia*

O atual cenário de preocupações com desinformação e personalização em campanhas políticas destaca o crescente papel da inteligência artificial (IA) e a necessidade urgente de regulamentação eficaz. Enquanto a IA se integra à cultura dominante, surgem questões cruciais sobre o controle e a autonomia das comunidades indígenas e outras minorias no uso dessas tecnologias. A eficácia das regulamentações existentes, especialmente no Brasil, é questionada, exigindo uma revisão rigorosa de como essas leis são aplicadas e sua capacidade de mitigar a desinformação e promover práticas éticas, especialmente em relação ao financiamento de campanhas e decisões judiciais como as do Tribunal Superior Eleitoral (TSE).

Para enfrentar esses desafios complexos de maneira significativa, é fundamental ouvir e integrar perspectivas diversas na formulação de políticas. Isso inclui desenvolver abordagens que não apenas garantam conformidade regulatória, mas também promovam equidade e proteção para grupos marginalizados, fortalecendo assim os princípios democráticos diante das rápidas transformações tecnológicas e sociais.

## Regulação da IA no Brasil

No contexto jurídico brasileiro, o uso de inteligência artificial (IA) e mídias sintéticas levanta uma série de questões complexas que envolvem direitos da personalidade, direito autoral, e regulamentações como a Lei Geral de Proteção de Dados Pessoais (LGPD). Enquanto é claro que o uso comercial da imagem ou para ferir a honra de alguém pode ser alvo de processos legais, situações como a postagem de fotos não comerciais em redes sociais ou seu uso em piadas complicam a aplicação dessas leis. A mídia sintética, especialmente o uso de vídeos e

imagens manipuladas, é um fenômeno relativamente novo que desafia interpretações tradicionais de direitos e regulamentações existentes.

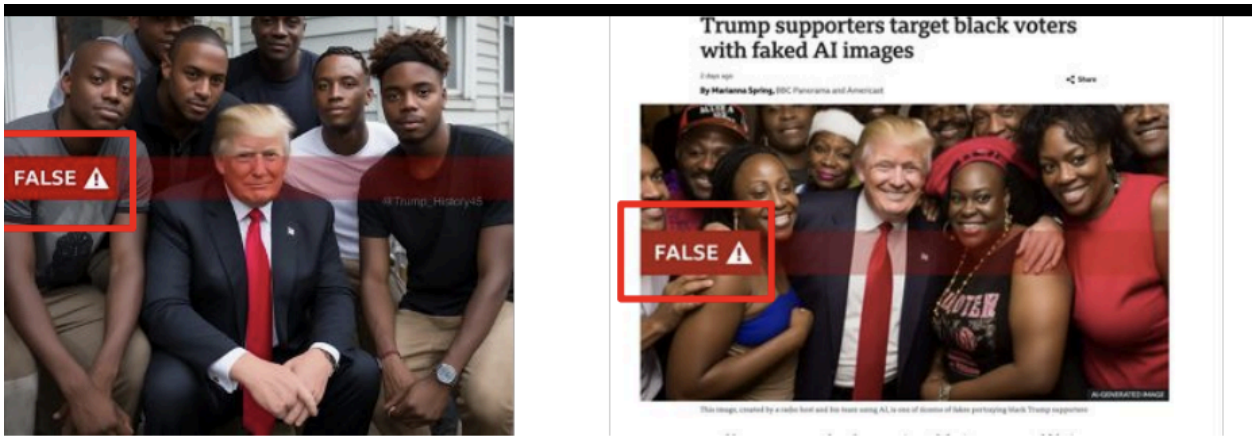
O direito autoral emerge como uma ferramenta crucial para bloquear o uso não autorizado de mídias sintéticas, especialmente quando envolve bases de dados ou material protegido por direitos autorais de terceiros. A LGPD, por sua vez, está no centro de um intenso debate sobre como equilibrar a liberdade de expressão com a proteção da privacidade e dos direitos pessoais diante do avanço tecnológico.

O histórico de regulação de IA no Brasil, exemplificado pelo PL 2338/2023 e resoluções como as do Tribunal Superior Eleitoral (TSE) sobre conteúdo sintético, mostra um esforço contínuo para adaptar a legislação às novas realidades digitais, embora desafios significativos permaneçam em relação à implementação eficaz e à interpretação consistente dessas leis em um ambiente em rápida evolução tecnológica.

## Métodos de Transparência para Mídia Sintética

Como evitar e mitigar os riscos da IA generativa? Quais são as soluções ou as respostas que temos que ter? Dados estes riscos, que respostas priorizar?

Exemplo: Donald Trump com pessoas negras




**Quais são os métodos de transparência? Como podemos garantir que eles funcionem?**

Para evitar e mitigar os riscos associados à inteligência artificial generativa, como os gerados por deepfakes e outras mídias sintéticas, é crucial considerar diversas estratégias de transparência e proteção. Uma abordagem importante é a comunicação direta, que envolve a inclusão de marcas d'água visíveis nos conteúdos gerados pela IA, como logotipos ou identificadores claros, para indicar sua artificialidade. Essas marcas podem facilitar a identificação rápida e direta de conteúdos sintéticos, proporcionando uma maneira acessível de alertar o público sobre sua natureza não genuína. No entanto, é essencial garantir que essas

marcas sejam robustas o suficiente para resistir a tentativas de remoção ou edição, o que requer métodos de implementação e verificação cuidadosos.

Além disso, a comunicação indireta por meio de marcas d'água invisíveis oferece uma camada adicional de proteção, pois essas marcas são incorporadas de maneira que não são perceptíveis a olho nu, mas podem ser detectadas por ferramentas especializadas. Isso dificulta a manipulação do conteúdo sem degradar sua qualidade, exigindo habilidades técnicas avançadas para contornar eficazmente essas medidas de segurança.

Outra abordagem eficaz é o fingerprinting, que envolve atribuir um código único (hash) a cada instância de conteúdo gerado pela IA. Esse código permite a identificação do conteúdo mesmo após modificações, garantindo sua integridade ao longo do tempo. O fingerprinting é eficiente em termos de processamento e requer apenas recursos computacionais mínimos para indexação e recuperação, mas depende da qualidade do algoritmo de hashing e levanta questões sobre privacidade devido ao armazenamento centralizado dos dados.

Por fim, a inclusão de metadados verificáveis diretamente nos arquivos de mídia pode fornecer informações sobre o ciclo de vida do conteúdo, como data e local de criação, tornando mais fácil rastrear sua origem e autenticidade. Esta abordagem é relativamente simples de implementar, mas requer adoção generalizada para garantir sua eficácia em diferentes dispositivos e plataformas.

Cada uma dessas estratégias oferece benefícios distintos na mitigação dos riscos da IA generativa, ajudando a promover um uso responsável e transparente das tecnologias sintéticas. O desenvolvimento contínuo e a adoção de soluções integradas, combinadas com educação pública sobre os desafios associados à mídia sintética, são fundamentais para fortalecer a resiliência da sociedade diante dessas inovações tecnológicas.

## Protegendo o que é verdadeiro, Detectando o que é falso - Workshop de Divulgação Indireta

### Grupo #1: divulgação indireta: marca d'água invisível

No cenário hipotético onde se documenta um protesto contra o governo e se deseja proteger identidades removendo rostos e pessoas do enquadramento, mas esse processo deixa rastros invisíveis de marcas d'água, várias preocupações surgem:

Primeiramente, a identificação precisa saber distinguir quais informações foram alteradas e quais permaneceram intactas. A capacidade de recuperar informações apagadas também é crucial, pois pode ser necessário revisitar o conteúdo original para fins de verificação ou análise forense.

Além disso, entender os motivos por trás da edição é fundamental. Enquanto a marca d'água invisível pode ser uma ferramenta para proteger a identidade de pessoas vulneráveis, como manifestantes, é importante garantir que essa proteção não comprometa a integridade ou autenticidade do conteúdo documentado.

A possibilidade de excluir uma marca d'água invisível levanta questões sobre a segurança e a privacidade, especialmente em contextos sensíveis como protestos políticos. As pessoas expostas podem enfrentar riscos adicionais, como monitoramento por dispositivos de rastreamento ou vigilância policial.

A diversidade de ferramentas de proteção, como pintura facial e roupas que evitam reconhecimento facial, pode ser crucial para uma proteção efetiva das identidades em situações de alto risco. Contudo, é essencial decidir que tipo de marca d'água invisível é mais apropriada, considerando o tipo de manipulação detectável e a área específica da imagem afetada.

No contexto mais amplo, a identificação de pessoas em protestos pode ser usada tanto para defender quanto para incriminar, dependendo das agendas políticas e jurídicas envolvidas. O uso de tecnologia pelo Estado, como câmeras de segurança, pode ser uma forma de contestar narrativas ativistas, evidenciando a importância de acessar e preservar imagens de maneira transparente e imparcial.

Portanto, a implementação de marcas d'água invisíveis deve ser cuidadosamente considerada dentro do contexto ético e legal para garantir que as ferramentas de proteção não se tornem instrumentos de vigilância ou opressão contra manifestantes e ativistas.

### Grupo #2: divulgação indireta: impressão digital

No cenário atual de mídias digitais, a utilização de hashes como identificadores únicos de imagens desempenha um papel crucial na verificação da autenticidade e na detecção de manipulações. No entanto, a falta de preparo dos educadores para lidar com essas tecnologias emergentes é uma lacuna preocupante. Para mitigar os efeitos da desinformação, é essencial promover a educação midiática desde cedo nas escolas, ensinando os alunos a identificar e questionar conteúdos falsos.

A crítica informada na análise de mídias digitais assume uma importância crescente diante dos desafios apresentados por deepfakes e fake news, especialmente em contextos sociais e políticos vulneráveis. No Brasil, as políticas educacionais e regulatórias ainda apresentam deficiências significativas quando se trata de abordar adequadamente essas questões emergentes. É crucial promover a transparência e conscientização em todas as faixas etárias, incentivando o uso de tecnologias como marcas d'água para garantir a autenticidade das informações disseminadas.

Além de educar sobre a verificação de fatos, é igualmente importante fornecer orientação sobre o controle emocional diante de informações falsas, visando fortalecer a resiliência dos indivíduos frente à manipulação digital. Propostas como cartilhas educativas dedicadas a deepfakes e fake news, adaptadas para diferentes grupos demográficos, podem desempenhar um papel crucial na disseminação de conhecimentos críticos e na promoção de comportamentos responsáveis online.

Enfrentar efetivamente os desafios das mídias digitais requer uma colaboração estreita entre setores público, privado e acadêmico. A implementação de iniciativas conjuntas para desenvolver políticas robustas, promover a educação e fortalecer as capacidades de resposta é essencial para proteger a integridade das informações e fortalecer a resiliência da sociedade diante das ameaças digitais em constante evolução.

### Grupo #3: Divulgação indireta: metadados

No cenário hipotético de documentação e compartilhamento de um incidente de abuso policial, várias preocupações éticas e práticas emergem. Primeiramente, a proteção da privacidade dos envolvidos é crucial, especialmente ao compartilhar o vídeo anonimamente com meios de comunicação. Conforme as diretrizes da LGPD, evitar a divulgação de dados pessoais identificáveis é essencial para mitigar riscos legais e proteger a segurança dos indivíduos afetados pelo incidente.

Além disso, a gestão cuidadosa dos metadados é fundamental para preservar a integridade e a veracidade do vídeo como evidência. Incluir informações essenciais como data e hora da gravação é necessário para fins de verificação, enquanto proteger a localização específica pode ser crucial para evitar a exposição indevida dos envolvidos. Isso garante que o vídeo possa ser utilizado eficazmente em processos legais sem comprometer a segurança dos participantes.

Questões éticas mais amplas também surgem, especialmente em relação ao monitoramento de comunidades vulneráveis, como quilombolas, e ao uso de tecnologias de reconhecimento facial. É essencial considerar o impacto dessas tecnologias na privacidade e nos direitos individuais, equilibrando a necessidade de segurança pública com a proteção dos direitos humanos.

Por fim, promover transparência sobre a origem e a manipulação do vídeo é crucial para combater a desinformação e assegurar a integridade da informação compartilhada. Isso envolve divulgar claramente como o vídeo foi captado, editado e compartilhado, garantindo que sua utilização sirva para promover justiça e responsabilidade, respeitando ao mesmo tempo os direitos individuais dos envolvidos.

## Detectando as mídias sintéticas

No cenário atual de evolução tecnológica, a perícia digital desempenha um papel crucial na coleta e análise de evidências digitais, especialmente em contextos criminais, seguindo o princípio de Locard, que afirma que todo contato deixa um rastro digital. Essa abordagem inclui a análise detalhada de artefatos em evidências digitais, como a investigação de estruturas de arquivos e a análise de compressão e iluminação em imagens e vídeos para detectar falsificações.

No entanto, o surgimento das realidades sintéticas, impulsionadas pela IA, representa um desafio significativo. Estima-se que até 90% do conteúdo online será sintético nos próximos anos, transformando drasticamente a percepção e o uso da informação. Diante disso, é crucial capacitar métodos avançados de detecção para identificar conteúdos sintéticos, explorando propriedades invisíveis a olho nu e promovendo a cooperação entre grupos de pesquisa para lidar com a evolução rápida das tecnologias de falsificação.

Para enfrentar esses desafios emergentes, recomenda-se educar o público sobre os riscos das mídias sintéticas e desenvolver padrões tecnológicos robustos, como assinaturas digitais em formatos de imagem, para ajudar na detecção e prevenção de manipulações. Além disso, a implementação de políticas e regulamentações é essencial para controlar a criação e disseminação de mídias sintéticas, garantindo que essas tecnologias sejam usadas para fortalecer os princípios democráticos e proteger a liberdade individual, enquanto se combate seu uso mal-intencionado.

## Deepfake Rapid Response Force (Força de resposta rápida sobre Deepfakes)

A detecção de deepfakes representa um desafio crescente e complexo, conforme evidenciado por casos recentes ao redor do mundo. Um laboratório na Itália desenvolveu métodos específicos para autenticar áudios, destacando a necessidade de investimentos em modelos personalizados semelhantes para líderes comunitários e políticos. No Sudão, em julho de 2023, foi relativamente fácil provar a autenticidade de um áudio, mas o vídeo apresentava maior dificuldade, especialmente em regiões onde a tecnologia de inteligência artificial ainda não está amplamente desenvolvida, como no mundo árabe.

Um caso notório no México envolveu uma imagem falsa que supostamente mostrava o presidente López Obrador com El Chapo, ilustrando como a detecção de manipulações visuais pode ser enganosa e desafiadora. Ferramentas computacionais como Deepware e Optic são comumente usadas para detecção, porém enfrentam limitações significativas, incluindo altos índices de falsos positivos e negativos. A detecção computacional é complexa e requer habilidades específicas para interpretar resultados precisos, além de enfrentar desafios de equidade no acesso a tecnologias confiáveis e na capacidade de explicar a autenticidade das mídias.

# Construindo o futuro da mídia sintética no Brasil - Diálogos de Grupo

## Preparação para as eleições

Ganhos positivos da criação do Comitê de Combate à Desinformação do TSE foram evidentes. O esforço conjunto envolvendo um canal de denúncias com conexão rápida às plataformas e o compromisso dessas plataformas em responder às denúncias foi crucial para uma moderação de conteúdo mais qualitativa e eficiente durante as eleições. No entanto, para assegurar um processo eleitoral mais transparente e justo em 2024, é fundamental pensar estrategicamente nas quatro forças regulatórias necessárias.

Em primeiro lugar, é imperativo focar em estratégias de comportamento, como a educação midiática direcionada para diferentes grupos etários e segmentos sociais, incluindo jovens, idosos e comunidades religiosas. Além disso, enfrentamos o desafio significativo do compartilhamento de dados eleitorais, especialmente diante dos recentes fechamentos das APIs por parte das plataformas digitais. Estratégias eficazes para compartilhamento de bases de dados são essenciais para aumentar a transparência e a integridade das informações disponíveis durante o processo eleitoral.

Outro ponto crucial é o fortalecimento dos ecossistemas de comunicação, especialmente para grupos historicamente vulneráveis, como mulheres, mulheres negras, indígenas e LGBTQIA+. Capacitar candidatos dessas minorias é igualmente fundamental, considerando o suporte substancial que agendas de extrema direita e anticlimática têm recebido na criação e disseminação de conteúdo. Essas medidas não apenas ajudarão a mitigar a desinformação, mas também promoverão uma participação eleitoral mais inclusiva e informada, fortalecendo assim a integridade democrática do processo eleitoral no Brasil.

Essas estratégias refletem um compromisso coletivo em garantir que as eleições de 2024 sejam conduzidas com integridade e transparência, abordando os desafios emergentes e promovendo uma participação cidadã mais robusta e consciente.

## Letramento Digital

A Educação Midiática, especialmente em IA e mídia sintética, precisa ser contemplada em Projetos de Lei no contexto brasileiro. Isso envolve não apenas ações legislativas, mas também responsabilidades claras atribuídas à imprensa, ao governo e aos criadores de conteúdo. Para efetivar essas medidas, é essencial capacitar os docentes para integrar essa educação de forma transversal nas escolas. Além de habilidades técnicas, os professores precisam desenvolver a capacidade de promover reflexão crítica entre os alunos. Essa formação não deve focar apenas na tecnologia, mas também em como ela impacta a sociedade, garantindo que não seja uma fonte de dominação ou exclusão social.

É importante adaptar a educação midiática ao dia a dia escolar, especialmente considerando as limitações de ferramentas e acesso à internet. Ela sugere que os alunos, muitas vezes mais familiarizados com a mídia sintética do que seus professores, podem ser agentes ativos na integração dessa educação no currículo. Integrar a expertise dos alunos não apenas fortalece a literacia midiática, mas também promove o pensamento crítico sobre tecnologias digitais. Assim, capacitar os professores de todas as áreas não é apenas uma necessidade educacional, mas uma estratégia vital para fortalecer a cidadania digital e preparar os estudantes para um uso ético e informado do ambiente digital.

Portanto, políticas públicas devem não apenas fornecer recursos materiais e infraestrutura, mas também garantir formação contínua para educadores, adaptando estratégias às necessidades locais. Isso inclui uma abordagem colaborativa que envolva a comunidade educacional na definição de diretrizes educacionais que promovam o uso responsável e crítico das tecnologias digitais.

## Usos práticos e criativos de IA

Foram mostrados exemplos inovadores de como a IA generativa pode ser aplicada de maneira criativa. Inicialmente, destacou como a mídia sintética pode preservar a identidade das pessoas sem perder a autenticidade emocional. Demonstrou-se casos práticos onde a IA foi utilizada para proteger a identidade de indivíduos em situações vulneráveis, além de explorar seu potencial na criação de jogos educativos e na preservação de narrativas históricas.

Além da criatividade, a discussão abordou os desafios de segurança e ética associados ao uso da IA generativa. Reconheceram que, embora a tecnologia ofereça oportunidades significativas em processos criativos, é crucial estabelecer limites claros para sua aplicação ética e segura. Discutiram a importância de adaptar essas ferramentas para atender às necessidades específicas de comunidades, como a adaptação para línguas indígenas e o desenvolvimento de ferramentas que promovam a autonomia cultural e tecnológica.

A conversa também enfatizou a necessidade de garantir o acesso equitativo a essas tecnologias, especialmente em contextos onde recursos básicos são limitados. Propuseram a criação de plataformas e recursos educativos que não apenas protejam identidades e preservem culturas, mas também capacitem comunidades marginalizadas a utilizar a IA de forma construtiva e significativa. Dessa forma, a integração responsável e inclusiva da IA generativa pode não apenas ampliar as possibilidades criativas, mas também fortalecer a autonomia e a representação cultural das comunidades ao redor do mundo.

## Cenários Futuros

Os cenários vislumbrados em 2026 revelaram um planejamento estratégico abrangente para as eleições. Foi considerada a criação de mensagens informativas sobre locais de votação, dados sobre candidatos e conteúdo personalizado para diferentes segmentos da população. Houve

um foco significativo na implementação de capacidades avançadas de detecção para apoiar jornalistas, comunidades indígenas e moradores de favelas, além de discussões sobre a necessidade de regulamentação mais rigorosa das redes sociais. Investimentos em pesquisa para entender e mitigar manipulações, bem como preparar respostas rápidas e confiáveis durante o período eleitoral, também foram contemplados.

Foi enfatizada a importância de utilizar ferramentas existentes para analisar rapidamente casos de desinformação que ganham destaque, com organizações sem fins lucrativos desempenhando um papel crucial ao fornecer verificações de fatos e informações confiáveis à população.

Outro aspecto relevante foi o fortalecimento de líderes locais como multiplicadores de informações precisas. Capacitá-los não apenas para disseminar informações confiáveis, mas também para identificar e combater desinformações, foi considerado essencial para aumentar a resiliência contra campanhas de desinformação durante o processo eleitoral.

Essas estratégias refletem um esforço coordenado para proteger a integridade das eleições, capacitando diversos setores da sociedade com as ferramentas e conhecimentos necessários para enfrentar os desafios emergentes na era digital.

## Próximos passos

Os próximos passos incluem compartilhar as informações discutidas neste workshop com todos os participantes, garantindo que todos estejam atualizados sobre as estratégias e insights desenvolvidos. Além disso, é essencial levar as informações cruciais deste workshop para outros locais no Brasil e no mundo, promovendo o compartilhamento de conhecimento e fortalecendo iniciativas globais contra a desinformação.

Quanto ao treinamento de pessoas, a WITNESS pode explorar a possibilidade de desenvolver programas específicos para capacitar indivíduos em temas relacionados à segurança digital, identidade e uso ético da IA generativa. Isso não apenas promoveria uma maior conscientização sobre esses temas, mas também equiparia mais pessoas com as habilidades necessárias para navegar de maneira segura e responsável no cenário digital em constante evolução.