

WITNESS

O QUE É PRECISO SABER
SOBRE O USO DE
FERRAMENTAS
DE DETECÇÃO
DE IA

SOBRE ESTE FOLHETO

A IA facilita a manipulação da realidade. As alegações sobre o uso da IA também facilitam a negação ou o desprestígio plausível da realidade.

Nesse cenário em constante mudança, ter uma compreensão clara das ferramentas de detecção de IA para áudio, imagens e vídeo, considerando suas capacidades, limitações e melhores práticas, é um passo fundamental para “fortalecer a verdade” e apoiar a compreensão sobre como distinguir imagens reais de imagens feitas por IA (sintética).

Este livreto serve como um guia de referência rápida para jornalistas, comunicadores, verificadores de fatos e outros defensores de direitos humanos, fornecendo-lhes dicas essenciais para o uso eficaz e responsável de ferramentas de detecção de IA em seu trabalho.

5 COISAS IMPORTANTES QUE O PÚBLICO
PRECISA SABER SOBRE

FERRAMENTAS DE DETECÇÃO DE IA

1

A precisão não é garantida

As ferramentas de detecção por IA podem oferecer informações úteis, mas não são infalíveis. As taxas de precisão variam, geralmente entre 85% e 90%, o que significa que sempre há margem para erros.

2

Mídias diferentes exigem ferramentas diferentes

As ferramentas de detecção de IA costumam ser projetadas para analisar tipos específicos de conteúdo, como texto, imagens, vídeo ou áudio. Nenhuma ferramenta funciona igualmente bem para todos os formatos de mídia. As diferenças linguísticas podem afetar a precisão da detecção de texto e fala, já que alguns idiomas podem ser mais difíceis de analisar. Além disso, a detecção de áudio é geralmente mais complexa e menos desenvolvida em comparação com a detecção de imagens e vídeos, o que torna mais difícil identificar áudio gerado por IA.

3

A síntese de IA está evoluindo rapidamente

O conteúdo gerado por IA (deepfakes, imagens geradas por IA, textos escritos por IA, áudio sintético, etc.) está evoluindo rapidamente, e as ferramentas de detecção podem ter dificuldade em acompanhar os avanços mais recentes.

4

A qualidade dos dados de entrada é importante

A qualidade do conteúdo que está sendo analisado afeta significativamente os resultados da detecção. Arquivos de mídia de baixa resolução ou altamente compactados podem ser mais difíceis de serem avaliados com precisão pelas ferramentas de detecção.

5

É possível que ocorram falsos positivos e falsos negativos

As ferramentas de detecção de IA podem classificar erroneamente um conteúdo real como gerado por IA (falsos positivos) ou deixar de reconhecer um conteúdo gerado por IA (falsos negativos); portanto, os resultados devem sempre ser interpretados com cautela.

ANTES DE UTILIZAR FERRAMENTAS DE DETECÇÃO

1 Comece com o PIER ou uma abordagem semelhante de alfabetização midiática

Comece aplicando o método PIER (Pare, Investigue a fonte, Encontre mais informações, Rastreie as alegações, citações ou mídias até sua fonte original) ou uma estratégia semelhante para avaliar a credibilidade do conteúdo antes de recorrer às ferramentas de detecção por IA.

2 Realizar análises de falhas e revisar metadados

Verifique se há falhas visuais ou de áudio, distorções ou inconsistências que possam indicar adulteração. Além disso, examine os metadados em busca de pistas sobre a origem e a autenticidade do material.

3 Adote uma abordagem avançada de OSINT

Utilize técnicas de inteligência de fontes abertas (OSINT), como a comparação cruzada de conteúdos. A verificação de fontes, a busca reversa de imagens e a análise do contexto e da linha do tempo das mídias, para obter um panorama mais completo e contextualizado.

4 Use ferramentas de detecção baseadas em IA como ponto de referência

Considere as ferramentas de detecção como uma etapa do processo de verificação, tendo em mente suas limitações. Utilize-as como evidência de apoio, mas não as transforme na única base para a tomada de decisões e conclusões.

5 Esteja ciente das deficiências das ferramentas de detecção de IA

Esteja ciente das limitações inerentes às ferramentas de detecção, como menor precisão com determinados formatos de mídia, variações linguísticas ou técnicas sintéticas emergentes. Utilize os resultados com cautela e responsabilidade.

AO UTILIZAR FERRAMENTAS DE DETECÇÃO

1 Obtenha a versão original e da mais alta qualidade

Certifique-se sempre de estar trabalhando com a melhor versão possível do conteúdo para melhorar a precisão dos resultados da detecção por IA.

2 Separar o áudio e o vídeo para uma análise independente

Se for o caso, analise os componentes de áudio e vídeo individualmente para aumentar a precisão da detecção em diferentes formatos de mídia.

3 Selecione a ferramenta de detecção adequada para o tipo de mídia

Utilize ferramentas de detecção projetadas especificamente para o formato de mídia que você está analisando, pois as ferramentas otimizadas para imagens, vídeos ou texto podem apresentar diferentes níveis de eficácia.

4 Mantenha-se informado sobre os avanços na síntese por IA

Acompanhar continuamente os avanços na criação de conteúdo sintético para compreender melhor as capacidades atuais e as tendências emergentes.

5 Acompanhe as tendências da tecnologia de detecção de seguimento, mas esteja ciente de que há limitações

Acompanhe os avanços nas ferramentas de detecção baseadas em IA, reconhecendo, ao mesmo tempo, suas limitações inerentes, que geralmente atingem uma precisão de 85% a 90%.

UMA COBERTURA JORNALÍSTICA RESPONSÁVEL SOBRE IA

1

Divulgue seu processo de verificação

Seja transparente quanto aos métodos e ferramentas utilizados para verificar conteúdos gerados por IA ou manipulados. Explique claramente como se chegou às conclusões e quais medidas foram tomadas para garantir a precisão.

2

Sensibilizar o público sobre os métodos de detecção

Forneça explicações claras sobre aspectos desconhecidos do processo de verificação de IA, ajudando o público a compreender as complexidades da detecção de conteúdo gerado por IA e por que determinadas ferramentas ou métodos foram utilizados.

3

Enquadrar o conteúdo dentro de seu contexto e momento

Certifique-se de que as reportagens relacionadas à IA apresentem claramente o contexto e o momento em que o conteúdo foi criado ou compartilhado, a fim de evitar narrativas enganosas e contribuir para uma maior conscientização do público sobre as capacidades atuais da IA.

4

Reconhecer as questões éticas e as limitações das ferramentas de detecção baseadas em IA

Sempre informe o público sobre as possíveis imprecisões ou limitações das ferramentas de detecção por IA utilizadas em seu relatório, especialmente no que diz respeito a falsos positivos ou negativos e às dependências dos conjuntos de treinamento.

5

Promover a alfabetização midiática e o pensamento crítico

Incentivar o público a abordar o conteúdo gerado por IA de forma crítica, oferecendo orientações sobre como avaliar a credibilidade da mídia digital e reconhecer os sinais de alerta de conteúdo manipulado.

METADADOS CRIPTOGRÁFICOS

1

É como os metadados “tradicionais”, só que permite que você verifique as informações

O que os metadados criptográficos significam é que as informações neles contidas estão vinculadas àquela versão específica do conteúdo digital e que alguém — seja quem for o signatário — atesta a veracidade dessas informações.

2

Ainda cabe a você, o espectador, decidir se confia ou não no conteúdo, independentemente de ele possuir ou não metadados criptográficos

Informações falsas ou enganosas podem, mesmo assim, conter metadados verificáveis, enquanto conteúdos verdadeiros podem não conter.

3

Os metadados criptográficos podem estar se tornando comuns

Estão em andamento iniciativas para padronizar em todo o ciclo de vida da mídia, desde a criação inicial até a edição e a publicação.

4

Os metadados criptográficos devem se referir ao “como”, e não ao “quem”

Para garantir a privacidade e o anonimato, os metadados verificáveis devem se concentrar em como um conteúdo foi produzido, e não em quem o criou, publicou ou compartilhou.

5

Use-o, mas leve em consideração, antes de tudo, as implicações de segurança e éticas

Os metadados criptográficos podem ajudar a fortalecer a verdade, mas leve em consideração a sua segurança e a das pessoas ao seu redor, bem como a forma como isso poderia moldar nosso ecossistema de informação.



A WITNESS é uma organização global da sociedade civil dedicada a tornar a verdade inegável. Na África, na Ásia-Pacífico, no Oriente Médio e Norte da África e na América Latina, capacitamos defensores que atuam na linha de frente para documentar, verificar e preservar evidências audiovisuais, e trabalhamos para garantir que os sistemas e a infraestrutura relacionados a essas evidências permitam que elas sobrevivam, sejam confiáveis e promovam a prestação de contas.

Encontre mais recursos em nossa biblioteca online



Veja a versão digital deste livreto aqui



feedback@witness.org



[@witnessorg](https://twitter.com/@witnessorg)



[@witness_org](https://www.instagram.com/@witness_org)



[WITNESS](https://www.facebook.com/WITNESS)

A produção deste material contou com o apoio de:



Microsoft