

Indústria trabalha em padrões técnicos para identificar conteúdo gerado por IA: Meta anuncia que rotulará IA em seus apps

Em linha com a preocupação dos associados do Conselho Digital quanto a autenticação de conteúdos gerados por IA, a Meta anunciou no blog da empresa, nessa terça-feira (6), que, nos próximos meses, a empresa começará a rotular imagens postadas pelos usuários no Facebook, Instagram e Threads quando detectar indicadores padrão da indústria de que foram geradas por IA. A empresa está trabalhando **com parceiros da indústria em padrões técnicos comuns** para identificar conteúdo gerado por IA, incluindo imagens, vídeos e áudios. A iniciativa demonstra o quanto o ecossistema está atento às demandas da sociedade e das autoridades.

Para identificar e rotular conteúdo gerado por IA, a Meta usa marcadores visíveis, marcas d'água invisíveis e metadados embutidos nos arquivos de imagem. Isso melhora a robustez dos marcadores invisíveis e ajuda outras plataformas a identificá-los. A Meta já rotula imagens fotorealistas criadas usando sua própria ferramenta de IA para informar as pessoas de que foram "Imaginadas com IA". A empresa busca estender essa prática para conteúdo criado com ferramentas de outras empresas, trabalhando com parceiros para alinhar em padrões técnicos que sinalizem quando um conteúdo foi criado usando IA. A empresa está desenvolvendo ferramentas que podem identificar esses marcadores invisíveis em grande escala, permitindo rotular imagens de empresas como **Google, OpenAI, Microsoft, Adobe, Midjourney e Shutterstock** à medida que implementam planos para adicionar metadados às imagens criadas por suas ferramentas.

Embora as empresas estejam começando a incluir sinais em seus geradores de imagens, a inclusão desses sinais em ferramentas que geram áudio e vídeo ainda não começou na mesma escala. Enquanto a indústria trabalha para desenvolver essa capacidade, a Meta está adicionando um recurso para que as pessoas divulguem quando compartilham vídeo ou áudio gerado por IA, permitindo que a empresa adicione uma etiqueta a esse conteúdo. Haverá penalidades para quem não usar essa ferramenta de divulgação e rotulagem ao postar conteúdo orgânico.

A empresa também está desenvolvendo tecnologias de marca d'água invisível, como a tecnologia "Stable Signature", que integra o mecanismo de marca d'água diretamente no processo de geração de imagem para alguns tipos de geradores de imagem, o que pode ser valioso para modelos de código aberto, impedindo que a marca d'água seja desabilitada.

A empresa enfatiza a importância de considerar vários fatores ao determinar se o conteúdo foi criado por IA, como a confiabilidade da conta que compartilha o conteúdo ou a procura por detalhes que possam parecer ou soar artificiais. A empresa continua colaborando com seus parceiros da indústria e mantendo diálogo com governos e sociedade civil para desenvolver padrões comuns e salvaguardas para o conteúdo gerado por IA.

Clique [aqui](#) para ler a matéria completa.

Category

1. Notícias

Tags

1. inteligência artificial

Date

08/09/2024

Date Created

06/02/2024