

The use of Artificial Intelligence to detect harmful practices during elections PT (Portuguese)

[00:00:18] Olá. Meu nome é Osama Aljaber e estou muito feliz de estar aqui com vocês hoje. Sou Especialista em Democracia Digital no Hub Regional do PNUD para os Estados Árabes. Então, no primeiro módulo do MOOC, você ouviu falar sobre questões levantadas pelo uso da IA para mídias sociais. Vou resumir brevemente seu impacto online na sessão de hoje antes de falar sobre o uso da IA para detectar práticas nocivas, como desinformação, discurso de ódio e propaganda online durante as eleições, mas também como as autoridades nacionais podem lutar contra essas práticas nocivas. Então, em 2020, criamos diariamente um banco de dados de 2,5 trilhões de dados. E nos últimos 11 anos, o volume de dados gerados, copiados e consumidos em todo o mundo cresceu quase 5000%. O ambiente de big data é uma das formas de analisar a questão da sobrecarga de informações que estamos experimentando. Resultando um ambiente caótico de notícias e aumentando as chances de ver informações erradas, discursos de ódio, etc. No entanto, também dá às grandes empresas de tecnologia acesso a grandes quantidades de dados para impulsionar sua própria IA, incluindo, mas não limitado ao seu próprio algoritmo de moderação de conteúdo. Algoritmos que podem prever o conteúdo que você se envolverá mais e o direcionará de volta com anúncios direcionados. Embora as plataformas de mídia social dependam de diferentes algoritmos baseados em IA para moderação de conteúdo, elas são limitadas de várias maneiras, principalmente devido à precisão e confiabilidade desses algoritmos e, às vezes, de seu piloto. O desenvolvimento da IA promoveu muitas discussões sobre como as autoridades nacionais utilizam o aprendizado de máquina para combater os desafios que enfrentam, especialmente durante as eleições, como desinformação durante a campanha eleitoral, o armamento do discurso de ódio e o direcionamento do eleitor para contas falsas representam desafios críticos para a integridade eleitoral. Elas também contribuem para prejudicar e distorcer o debate público e a tomada de decisões informadas durante a eleição. Isso pode ser muito perigoso, especialmente para jovens que podem ser manipulados durante as eleições, sobretudo com suas escolhas no dia da eleição. Cuidar das partes interessadas nacionais, incluindo órgãos de gestão eleitoral, organizações da sociedade civil, organizações comunitárias e instituições de mídia, pode avaliar se devem e podem tirar proveito da IA em seu trabalho conjunto e mandato para ajudar a combater distúrbios de informação online, incluindo desinformação, discurso de ódio e violência online contra grupos desfavorecidos. No entanto, uma das principais questões que sempre surgem quando se fala sobre o uso de IA é se podemos realmente confiar no julgamento e nos resultados de um sistema de IA, especialmente com o reconhecimento de que a tecnologia pode herdar os preconceitos humanos devido aos vieses no conjunto de dados de treinamento. E há muitos, muitos exemplos disso, mas também outras questões relacionadas ao que os sistemas de IA podem fazer e o que não podem fazer. Quem deve usar sua IA para lidar com conteúdo nocivo? Se ele pode apoiar o mandato principal de uma instituição ou grupo? Quem tem a capacidade humana e financeira de usar a IA sem impedir outras atividades principais? E, finalmente, como realmente conduzir uma discussão ética em torno da IA. Os órgãos de gestão eleitoral em geral têm tarefas extensas e gerenciam o processo eleitoral e podem ou não estar no melhor lugar para assumir uma tarefa tão extensa e sensível. Ou eles podem não ter nem mesmo o mandato legal para monitorar a campanha eleitoral. No entanto, os órgãos de gestão eleitoral que têm o poder de supervisão para regular este ambiente ou as eleições podem se beneficiar dessa capacidade aprimorada, desde que cumpram rigorosamente as leis internacionais e nacionais. Também é importante que as partes interessadas nacionais examinem todas essas questões antes de decidirem se devem construir seu próprio modelo e, potencialmente, sua própria estratégia de AI. Além disso, também é importante

garantir sistemas de IA seguros, confiáveis e inclusivos, aprimorando a colaboração entre organizações governamentais, industriais, acadêmicas e da sociedade civil. O desenvolvimento de sistemas de IA deve garantir três princípios fundamentais principais, que são justiça, responsabilidade e transparência. Pois um sistema AI livre significa que as partes interessadas nacionais precisam examinar claramente o que podem trazer para a mesa. Se é realmente justo aplicar o aprendizado de máquina no contexto de primeiro lugar e se o sistema pode violar a privacidade e os direitos de uma raça, gênero e grupo individuais específicos. A responsabilidade, por outro lado, é outro princípio essencial, especialmente para garantir que um sistema confiável de IA não vá na direção errada, mas também para quando um resultado de IA contradizer a opinião e os interesses dos cidadãos. Finalmente, quanto a transparência de AI, um cidadão precisa saber realmente quando um sistema de IA estiver sendo usado, mas também é necessário ter transparência sobre os próprios algoritmos. É sempre importante lembrar que a qualidade da saída de um sistema de IA depende muito de seus dados de entrada. Portanto, limpar os conjuntos de dados de treinamento a partir de suposições de preconceitos conscientes e inconscientes permitiu tomar melhores decisões baseadas em dados. Neste módulo, a Comissão Supervisora de Eleições no Líbano apresentará a você o E-Monitor Blog. Que é um algoritmo baseado em IA para que você e as pessoas monitorem a mídia digital como parte de um apoio eleitoral mais amplo como parte do processo eleitoral. Como parte de seu mandato principal, os juízes da Comissão Supervisora de Eleições no Líbano, apoiados por uma equipe de técnicos e monitores, detectaram violações eleitorais, polarização política, desinformação, misinformation, discurso de ódio, violência online contra mulheres e campanhas financiadas online. Além da observação on-line.