

Treinando sistemas de AI que reflitam a realidade e respeitem a privacidade dos usuários

A inteligência artificial ajuda bastante as pessoas a melhorar algumas atividades de suas vidas. Isso está levando diversos pesquisadores a utilizarem os dados para desenvolver sistemas de IA reflitam o mundo real, protegidos contra preconceitos não intencionais e tratados de maneira transparente e respeitosa com privacidade e segurança. Esses dados, como alimento para esses modelos de machine learning, precisam refletir a sociedade que está sendo modelada. Esses dados são fornecidos por humanos, que possuem uma história frágil e com preconceitos, por isso é muito importante que os envolvidos tenham realidades distintas, para não correr o risco de resolver problemas que alguns grupos homogêneos de pessoas acham que são importantes.

As pessoas estão mais conscientes sobre esse tema, se preocupando mais com o que esses algoritmos estão realmente fazendo e onde obtém seus dados. Por isso, esses modelos de aprendizagem devem ser transparentes o suficiente para que especialistas que não sejam de AI entendam como funciona a previsão daquele sistema e não apenas uma caixa preta. Como, na maioria dos casos, os dados fornecidos estão relacionados com contextos da sociedade, é importante que ética, transparência e privacidade sejam levados em consideração.

Tecnologias autônomas são invasivas à vida dos usuários, tanto em relação a testes feitos para melhoria de algum serviço, quanto para divulgar propagandas. É muito importante que a privacidade dos usuários seja preservada, justamente porque muitos usuários não sabem que seus dados estão sendo fornecidos para algum agente externo, ocasionando muitas vezes em descobertas sobre fatos confidenciais de usuários ou que violem a sua privacidade. Sobre ética, é necessário que cada usuário concorde primeiramente para que seus dados sejam usados por terceiros, além disso as pesquisas não deveriam prejudicar as pessoas, mas beneficiá-las em alguma atividade, como também todos os usuários devem ser tratados de maneira justa e igualitária, evitando injustiças e preconceitos.

Há um ponto cego na pesquisa de IA

Em 2016 a casa publicou um relatório de 4 eventos que ocorreram em 2016 sobre o futuro da inteligência artificial, nesses eventos tiveram dezenas de apresentações sobre as técnicas de IA para melhorar a performance a execução de tarefas desde de câncer no estado inicial à redução dos custos de energia em data centers.

Pedro domingos falou em 2015 que as pessoas acham que o computadores vão mudar o mundo, mas o real problema é que eles são muito estúpidos e já conquistaram o mundo. E muitos pesquisadores rejeitam a perspectiva da singularidade tecnológica, dizendo que esse campo é muito jovem ainda.

Segundo a CEO da IBM Ginni Rometty o investimento em IA vai chegar a 2 trilhões na próxima década.

Para reduzir impactos sociais no algoritmo existe 3 técnicas: implantar e cumprir, values in design e experimentos mentais.

Em implantar e cumprir o algoritmo é lançado e ele vai sendo aperfeiçoado em produção, como o caso que aconteceu com o google dele reconhecer um casal de afro descendentes como gorilas, sendo o erro encontrado apenas depois que o algoritmo já estava em produção e assim corrigido.

Em values in design são usados frameworks para identificar grupos de interesse e seus valores, e consequentemente são utilizados técnicas para estabelecer as opiniões das pessoas sobre privacidade pessoal, o ambiente entre outras coisas.

Em experimentos mentais são tratados os casos de sobre situações hipotéticas, tratando casos como o de um carro que deve decidir se vai matar 5 pessoas em uma via ou desviar e matar uma outra pessoa, entre outros casos.

Uma análise prática e amplamente aplicável dos sistemas sociais pensa em todos os possíveis efeitos dos sistemas de IA em todas as partes. Também se envolve com impactos sociais em todas as fases - concepção, projeto, implantação e regulação.