

Ética & Inteligência Artificial (Trabalho 1)

Aprendizado de Máquinas e Reconhecimento de Padrões (MO886)

Randerson A. Lemos (103897)

Introdução

Os avanços das tecnologias computacionais dos últimos anos somados ao desenvolvimento e ao aprimoramento de novos algoritmos de inteligência artificial possibilitaram a resolução de problemas e automatização de processos e tomadas de decisões em diversos campos da vida cotidiana tanto dentro dos espaços privados quanto públicos. A consolidação dos algoritmos de aprendizado de máquinas - com destaque para as redes neurais - e a maior disponibilidade de dados nos seus mais diversos formatos (categóricos, numéricos, fotográficos, auditivos) impulsionaram diversas áreas sensíveis a sociedade moderna como por exemplo: do transporte com os veículos autônomos, da agricultura com a identificação de pragas, da saúde com o diagnóstico de doenças, da segurança com sistemas anti-fraude e de reconhecimento facial, do entretenimento e lazer com os algoritmos de sugestão de conteúdos e produtos personalizados, entre outros [2].

Diante de tantas possibilidades de aplicação e uso, é possível afirmar que os algoritmos de inteligência artificial firmaram uma posição bastante próxima junto à sociedade moderna de maneira geral e na vida das pessoas de maneira particular. A interação próxima entre ser humano e máquina ou melhor entre ser humano e algoritmos de inteligência artificial, cuja tendência é se tornar cada vez mais íntima com o passar do tempo, suscita diversas discussões e questionamentos sobre os efeitos colaterais advindos da relação estabelecida entre esses dois atores. Uma dessas discussões, é centrada na ética na inteligência artificial e busca entender como esse conceito, que foi primordialmente concebido para operar sobre as relações de homem para homem pode ser aplicado nas relações de homem para máquina [2].

Com o objetivo de ser uma introdução à discussão da ética na inteligência artificial, esse trabalho traz a seguir os seguintes tópicos:

- uma definição de ética na inteligência artificial;
- apresentação de uma notícia recente (2022) de um problema de ética na inteligência artificial; e
- uma breve descrição dos problemas causados pelas ‘Armas de Destruição em Massa’ do capítulo 4 do livro ‘Algoritmos de Destruição em Massa’.

Definição de Ética na Inteligência Artificial

De acordo com a versão online do dicionário Aurélio [1], o termo ética é primeiramente definido como uma área de estudo da filosofia que se dedica a entender e analisar as motivações que ocasionam, alteram ou orientam o comportamento humano, principalmente as que estão conformadas por regras, preceitos ou normas sociais. Dessa maneira, ainda de acordo com o dicionário Aurélio [1], a ética pode ser entendida como a reunião de normas de juízo de valor moral presentes nas pessoas, em uma sociedade ou grupo social.

Com a consolidação das soluções de inteligência artificial em diversos âmbitos da vida humana, o assunto sobre ética na inteligência artificial vem atraindo cada vez mais interesse da sociedade e hoje é uma parte a mais da área de estudo da ética, sendo que um dos focos da ética na inteligência artificial é o de entender a ética das máquinas e suas condutas para que a interação delas com os seres humanos, ou seres vivos de maneira mais ampla, não seja danosa.

De maneira simplista, a inteligência artificial se dedica a desenvolver máquinas e algoritmos que sejam capazes de tomarem decisões coerentes e satisfatórias de acordo com informações de entrada e um contexto. Decisões coerentes e satisfatórias podem ser entendidas como aquelas que seriam igualmente tomadas por seres humanos capacitados para tal. Aqui, novamente, o assunto da ética salta aos olhos e com ele as chamadas Leis da Robótica. As três leis da robótica foram idealizadas pelo escritor Isaac Asimov e tem como objetivo balizar certos comportamentos das máquinas de modo a proteger o próprio ser humano. Essas regras são:

- Primeira lei: um robô não pode ferir um ser humano ou, por inação, permitir que um ser humano sofra algum mal;

- segunda lei: um robô deve obedecer as ordens que lhe sejam dadas por seres humanos exceto nos casos em que tais ordens entrem em conflito com a Primeira lei; e
- Terceira lei: um robô deve proteger sua própria existência desde que tal proteção não entre em conflito com a Primeira ou a Segunda lei.

Notícia envolvendo Ética e Inteligência Artificial

Em uma busca rápida pela internet sobre o assunto de ética e inteligência artificial, diversas notícias, somente deste ano de 2022, foram encontradas. Se, por um lado, a quantidade de notícias encontradas não foi algo surpreendente, dado que esse assunto é realmente atual, por outro lado, o conteúdo das notícias foi, sem dúvidas, surpreendente. Isso porque muitas dessas notícias abordavam experiências não satisfatórias no campo da ética entre homens e máquinas. Uma das notícias encontradas foi veiculada pelo Jornal O Estado de São Paulo (O ESTADÃO) e apresentava o seguinte título: Uma inteligência artificial sem ética pode arruinar a sociedade [3]. No início, há na notícia a afirmação de que os sistemas de inteligência artificial aprendem conosco e, na sequência, a pergunta “Mas será que estamos sendo bons professores?”. A explicação da pergunta, aparentemente sem muito nexo, vem em seguida: a ideia é contextualizar uma discussão sobre ética na inteligência artificial voltada para os riscos que estão suscetíveis os sistemas inteligentes de aprenderem comportamentos preconceituosos e não mais toleráveis dentro da nossa sociedade. Risco que é agravado quando levado em conta que muitos desses sistemas inteligentes interagirão com multidões e, assim, poderão passar adiante os comportamentos nocivos aprendidos para outras pessoas. Como forma de ilustrar as razões da discussão levantada, na notícia há a apresentação do caso da ferramenta Tay, que foi lançada pela Microsoft. Essa ferramenta ficou por trás de uma conta no Twitter e buscava simular uma adolescente criada para conversar com os usuários. A ferramenta não ficou no ar por mais de 24 horas. Isso porque nesse pequeno tempo, depois de interagir e conversar com diversas pessoas, a ferramenta desenvolveu uma personalidade racista, xenofóbica e sexista. A Tay, por exemplo, se tornou fã de Adolf Hitler e seus ideias. Na mesma notícia um outro problema associado às soluções de inteligência artificial e ética que acontece nos sistemas de recrutamento foi apresentado. Esses sistemas analisam milhares de currículos para uma vaga e, se não estiverem bem treinados, podem deixar de fora bons profissionais devido a algum viés desenvolvido com relação à idade, gênero ou raça dos candidatos.

Armas de destruição em massa e Corrida armamentista: indo à universidade

Em concordância com o assunto de ética e inteligência artificial suscitado neste trabalho, o capítulo “Corrida armamentista: indo à universidade” (capítulo 4) do livro “Algoritmos de destruição em massa” [4] traz uma discussão juntamente com um exemplo que são muito ilustrativos dos problemas das Armas de destruição matemáticas (ADMs). A discussão é desenvolvida e sustentada pelo exemplo dado que é apresentado ao longo do capítulo. Para dar início ao fio da meada, descreve-se no capítulo uma situação hipotética na qual todas as pessoas norte americanas optariam por ter uma dieta baseada principalmente em carne: nesse cenário hipotético, as consequências, no mínimo do ponto de vista econômico, seriam dramáticas. Haveria um grande desbalanceamento nas relações de oferta e demanda de diversos produtos agrícolas com grandes aumentos nos preços das carnes e uma significativa redução nos preços dos demais produtos. Após a descrição do cenário, o autor elucida que a correlação entre ele, o cenário, e as ADMs é a escala. Nas palavras do livro: “uma fórmula, quer seja uma dieta ou um código tributário, pode ser perfeitamente inofensiva em teoria. Mas se cresce para se tornar um padrão nacional ou global ele cria a sua própria economia distorcida e distópica”.

Na sequência, há no capítulo a apresentação de um caso real de uma ADMs nascida em 1983 nos Estados Unidos da América. Essa caso se passou no âmbito do projeto de elaboração de um rank nacional destinado a avaliar e classificar as universidades norte americanas empreendido pela revista semanal U.S. News & World Report. Essa iniciativa durou até 2010 e durante esse período cumpriu um dos seus principais objetivos: ajudar a revista com suas responsabilidades financeiras que vinham sendo cumpridas com bastante dificuldade. A parte das motivações relacionadas às finanças, a ideia do rank parecia ótima uma vez que ajudaria milhões de jovens na decisão de onde estudar. No entanto, junto com esse rank, vieram efeitos colaterais danosos sobre todo um ecossistema formado pelos jovens aspirantes a uma vaga em curso superior e seus familiares, pelos professores e dirigentes das universidades, e por alguns grupos de profissionais que se especializaram em prestar consultorias às universidades.

A elaboração dos primeiros ranks foi baseada principalmente em critérios subjetivos. Esses ranks foram alvos de críticas de diversos dirigentes das universidades que questionavam a colocação das suas instituições no rank e, por consequência, a confiabilidade do rank em si. Como forma de aprimorar o rank, os profissionais da revista buscaram indicadores quantificáveis que fossem capazes de capturar as características de uma boa

universidade. Alguns desses indicadores foram: SAT (o teste de aptidão escolar), razão aluno/professor e taxas de aprovação entre outros. Mas, como medir a felicidade, aprendizagem, confiança, amizade dos alunos? Esses indicadores não foram quantificados. Os novos ranks foram feitos com essas medidas objetivas tomando conta de 75% da nota final e sem nenhum auxílio de profissionais especializados em educação. Eles foram um sucesso e a maioria dos jovens que desejavam entrar em uma universidade baseavam suas decisões nos ranks da revista U.S. News & World Report e a maioria das universidades que almejavam maior prestígio acompanhavam de perto suas posições no rank bem como as medidas e os pesos utilizados para composição de suas notas. Assim, com esse rank, os alunos poderiam fazer escolhas mais conscientes quanto a universidade que estudar e as universidades poderiam orientar com eficiência seus recursos em políticas de melhoria da qualidade de ensino.

Com o sucesso do rank em âmbito nacional, isto é em escala, vieram atrás muitos problemas. Um ciclo de retroalimentação que reforçava a ordem classificatória apresentada no rank se estabeleceu. Afinal, os “melhores” alunos intencionavam estudar nas “melhores” universidades, mas um dos motivos dessas universidades serem as “melhores” era devido a qualidade do seu corpo discente. Outra consequência do rank recaiu sobre os próprios professores e gestores das universidades. Muitos deles ficaram desmotivados e desengajados como se viam trabalhando em instituição de ensino superior mal classificadas no rank.

Na tentativa de melhorar a posição de suas instituições de ensino, alguns gestores não mediram esforços. Cientes que parte da nota final do rank era baseada em 15 indicadores, muitos gestores empreenderam políticas agressivas e sem um coerente controle financeiro para aumentar a pontuação de suas universidades nesses indicadores. Muitos gestores foram bem sucedidos nessa empreitada, mas as custas de um aumento significativo das mensalidades cobradas dos alunos para balancear as contas novas.

Na realidade, rapidamente, os gestores universitários entenderam a matemática por trás do rank (as chamadas *proxies*) e se dedicaram, a partir daí, dia e noite, a encontrar formas de aumentar as notas de suas instituições baseadas no comportamento dessas *proxies*. Esse processo, mesmo pautado em intensões honestas, é quase uma trapaça, onde o objetivo é burlar as *proxies*. Isso porque essas equações são demasiado simples para serem utilizadas na representação na realidade complexo de uma quantificação de qualidade de ensino superior e, logo, facilmente enganadas.

Uma caso emblemático do processo de burla das *proxies* se passou junto com o departamento de matemática da Universidade King Abdulaziz, da Arábia Saudita. Esse departamento em um rank de 2014 ficou na sétima posição. O departamento tinha sido criado há apenas dois anos e já desbancava diversos instituições renomadas no mundo da matemática como Cambridge e MIT. De um investigação conduzida por um professor de Berkeley, descobriu-se que essa universidade tinha contratado diversos matemáticos que dispunha de trabalhos altamente citados. A esses professores foi oferecido um salário altíssimo mediante a contra-partida deles trabalharem apenas três semanas na Universidade King Abdulaziz e mudassem sua filiação para essa instituição. Com essa mudança de filiação, a universidade da Arábia Saudita ganhou instantaneamente muitos artigos e, de um dia para outro, alcançou a sétima posição em um rank de classificação de universidades.

Já no final, o capítulo “Corrida armamentista: indo a universidade”, retoma os efeitos nocivos das ADMs e, em particular, com as universidades, os alunos e os ranks. Mostrando que mesmo os alunos que conseguiram aprovações em boas universidades depois sofreriam com as pesadas parcelas do generoso empréstimo estudantil que normalmente precisam fazer para completar seus estudos de nível superior. Claro que, se os ranks considerassem outros fatores de qualidade das universidades que estivessem mais alinhados com as demandas da classe média como, por exemplo, de mensalidade das universidades ou de localização ou empregabilidade, o resultado final poderia ser mais satisfatória. Mas, ainda assim, esses novos ranks estariam suscetíveis a trapaça ou burla.

Assim, nesse capítulo, é apresentando uma série de problemas reais causados por uma ADM reais. Sendo que o grande problema dessas ADMs é de fato a escala e o efeito que causam nas massas populacionais de retroalimentação que geralmente caminha para um direção nociva às próprias pessoas.

Referências

- [1] Dicionário Aurélio. Dicionário online de português. <https://www.dicio.com.br/etica/>, ago. 2022.
- [2] Teresa Bernarda Ludermir. Inteligência artificial e aprendizado de máquina: estado atual e tendências. *Estudos Avançados*, 35(101):85–94, abr. 2021.
- [3] Jornal O Estado de São Paulo. Uma inteligência artificial sem ética pode arruinar a sociedade. <https://brasil.estadao.com.br/blogs/macaco-eletrico/uma-inteligencia-artificial-sem-etica-pode-arruinar-a-sociedade/>, ago. 2022.
- [4] Cathy O’Neil. *Algoritmos de destruição em massa*. Editora Rua do Sabão, 2021.