01 - https://g1.globo.com/globo-reporter/video/racismo-algoritmo-pesquisadora-estudatema-que-surge-com-avanco-da-ia-13124203.ghtml

A reportagem fala sobre uma mulher negra chamada Nina da Hora, ela é especialista em inteligência artificial e tecnologia e também é conhecida pelos seus trabalhos de conscientização sobre ética em sistemas de IA. Nina encontrou um propósito na área da computação enquanto fazia um trabalho de reconhecimento facial, durante os teste do trabalho Nina acabou que não conseguia testá-lo, pois não era reconhecida pelo sistema. A partir daí ela resolveu focar na área de ética em IA com sua tese de mestrado na faculdade sendo o racismo algorítmico, que também virou uma meta de vida. Segundo Nina alguns sistemas por conta desse viés acaba analisando alguns rostos negros e considerando eles como ameaça, perigo ou alguém que cometeu um crime, ela também cita que o racismo na sociedade tende a ser reproduzido pelo sistema.

02 - https://youtu.be/59bMh59JQDo

O vídeo fala sobre o preconceito no aprendizado de máquina, onde nesse aprendizado as pessoas programam a solução de um problema passo a passo e as máquinas aprendem essa solução encontrando padrões em dados, porém cada ser humano acaba tendo uma preferência em relação a tudo, por exemplo, sapatos, roupas, carros e entre outros, com isso acaba sendo quase impossível separar essas preferências/preconceitos das soluções que são passadas para as máquinas. Esses preconceitos podem ser passados de diferentes formas, o primeiro é através da interação onde por exemplo se pede para desenhar um sapato favorito, com isso pode ocorrer um padrão muito maior de um sapato em relação aos outros, a segunda maneira é pelo viés latente onde por exemplo uma IA pode considerar que todo físico é homem, e por fim tem o viés de seleção onde não são selecionados dados o suficiente para abranger a maioria das características/diferenças dos humanos e acaba que algumas pessoas podem ficar excluídas por não ter tido um treinamento com dados o suficiente (O racismo algorítmico é um exemplo).

03 - https://doi.org/10.1016/j.mlwa.2024.100525

O artigo fala como pequenas e aparentemente insignificantes alterações em sistemas de IA podem levar a consequências graves e imprevisíveis em relação ao viés e justiça. Inspirado na Teoria do Caos, o "Efeito Borboleta" em IA pode se originar de pequenas falhas, como vieses ocultos nos dados de treinamento, pequenos desvios durante o desenvolvimento do

algoritmo ou mudanças na distribuição dos dados entre as fases de teste e aplicação real. Essas alterações que parecem mínimas podem se amplificar, resultando em resultados injustos que impactam grupos marginalizados e expõe desigualdades sociais existentes. Esse fenômeno pode intensificar vieses e aumentar a vulnerabilidade a ataques maliciosos. O artigo defende que é importante examinar rigorosamente qualquer modificações nos sistemas, propondo métodos para detectar, medir e eliminar o Efeito Borboleta.

04 - https://youtu.be/og67qeTZPYs

O vídeo fala sobre o que é o viés algorítmico, que é a tendência de sistemas de IA a produzirem resultados injustos ou preconceituosos, porém como é mostrado o problema não está na máquina em si, mas nos dados que são fornecidos pelos humanos. Existem várias causas disso acontecer, a primeira delas é os dados de treinamento já enviesados, onde eles não representam toda a diversidade da população, outra causa são os erros de programação ou a incorporação de regras subjetivas pelos desenvolvedores que podem acabar introduzindo vieses no sistema, e por fim tem o viés de avaliação que é outra causa onde a interpretação humana dos resultados do algoritmo também pode ser preconceituosas, levando a decisões injustas. O vídeo também apresenta alguns exemplos reais e sugere algumas estratégias para combater esses problemas como: governança de IA com diretrizes éticas, uso de dados mais diversos, detecção contínua de vieses, transparência nos algoritmos e a formação de equipes de desenvolvimento mais inclusivas.

05 - https://abpnrevista.org.br/site/article/view/744

O artigo analisa como o racismo algorítmico se manifesta em sistemas de visão computacional que é a tecnologia que permite às máquinas enxergar e interpretar imagens, ele fala que esses sistemas frequentemente reproduzem e amplificam preconceitos raciais, especialmente contra a população negra. O artigo demonstra que o problema está em dois conceitos, o primeiro é a branquitude onde ela se manifesta quando o padrão de normalidade nos dados usados para treinar as IAs é o de pessoas brancas, como consequência, esses sistemas funcionam muito bem para reconhecer rostos brancos, mas falham frequentemente com rostos negros, o segundo conceito que é gerado pela branquitude é a opacidade, que é a invisibilidade de pessoas negras para a tecnologia, onde elas não são reconhecidas ou são classificadas de forma errada e preconceituosa. O artigo aponta que a maneira como os dados são coletados, categorizados e processados reflete relações de poder desiguais da sociedade e conclui que o racismo algorítmico não é um erro técnico isolado, mas uma consequência direta de como as estruturas de poder e o

viés racial são incorporados desde a concepção dos bancos de dados até o funcionamento final dos algoritmos de visão computacional.

06 - https://www.cienciasuja.com.br/temporada-4/pele-negra%2C-m%C3%A1quinasbrancas

O episódio do podcast investiga o grave problema do racismo algorítmico, focando em como tecnologias de reconhecimento facial são falhas, perigosas e resultam na discriminação racial. Ibnicialmente é falado do caso real de Robert Williams, um homem negro de Detroit que foi preso injustamente em sua própria casa, na frente de sua família, após ser identificado erroneamente por um software de reconhecimento facial da polícia. Este caso serve como exemplo para demonstrar as consequências humanas devastadoras de uma tecnologia imperfeita. Para aprofundar a discussão, o episódio conta com a participação do pesquisador Tarcízio Silva (Autor do artigo acima), ele explica que o problema central é a "branquitude" dos bancos de dados, onde os algoritmos são treinados predominantemente com imagens de homens brancos, o que os torna extremamente imprecisos ao tentar identificar rostos negros, especialmente de mulheres. O podcast também contextualiza o problema para a realidade brasileira, citando o uso massivo de câmeras de reconhecimento facial na segurança pública, como por exemplo no carnaval da Bahia, e o alto potencial para prisões injustas. O episódio argumenta que essa tecnologia é uma forma de "ciência suja", com vieses profundos que é aplicada de forma a amplificar o racismo estrutural, transformando preconceito em um suposto fato tecnológico colocando vidas em risco.

07 - https://www.cienciasuja.com.br/temporada-5/publicar-ou-perecer

O episódio do podcast Ciência Suja fala sobre a pressão esmagadora para que cientistas e pesquisadores publiquem artigos de forma incessante, um fenômeno conhecido como "publicar ou perecer". A discussão central gira em torno de como essa cultura, que deveria incentivar a produção de conhecimento, acabou criando um sistema com falhas graves. A carreira de um pesquisador, seu prestígio e até mesmo o financiamento para suas pesquisas são medidos pela quantidade de publicações em seu nome. Essa métrica, no entanto, raramente leva em conta a qualidade, o impacto real, a relevância do que é publicado ou até mesmo a veracidade ou não dos dados utilizados.

É possível entender que, embora a intenção original de incentivar a publicação contínua fosse positiva, visando a rápida disseminação do conhecimento e o avanço científico, na prática, o sistema gerou efeitos colaterais extremamente negativos. Essa pressão por volume abriu as portas para uma sensível queda na qualidade da produção científica. Para

cumprir metas, muitos pesquisadores se veem forçados a dividir um estudo robusto em vários artigos menores ou a publicar trabalhos com metodologia pouco confiável, que oferecem pouca contribuição científica. Pior ainda, essa demanda incessante criou um ecossistema predatório. O podcast destaca o surgimento de um mercado de "revistas predatórias", que exploram a necessidade dos acadêmicos publicarem qualquer artigo mediante pagamento, sem um processo de revisão sério. Em seus casos mais extremos e antiéticos, essa cultura pode levar a fraudes graves, como a manipulação ou até a invenção completa de dados para garantir resultados de maior impacto, mencionando-se a existência de verdadeiras "fábricas de artigos falsos". Dessa forma, aquilo que deveria ser um apoio à ciência se transforma em um risco para sua integridade e credibilidade.

08 - https://dl.acm.org/doi/10.1145/3613904.3642116

O artigo apresenta uma investigação sobre uma questão fundamental dos tempos modernos, como a inteligência artificial está, de fato, mudando os riscos à nossa privacidade? Para responder a isso, os pesquisadores analisaram 321 incidentes reais e documentados onde tecnologias de IA causaram problemas de privacidade. O grande ponto positivo e a principal contribuição do trabalho é a criação de uma taxonomia, ou seja, um sistema de classificação claro e organizado para os riscos de privacidade em IA. Ao basear a análise em casos concretos, o estudo oferece uma linguagem comum e um framework com 12 categorias de risco para ajudar desenvolvedores, legisladores e o público a entender e discutir os danos de forma precisa. Por exemplo, o artigo mostra como a IA não apenas agrava problemas antigos, como a Vigilância, ao incentivar a coleta massiva de dados para treinar modelos , mas também cria riscos novos. Um exemplo chocante é o ressurgimento da Frenologia/Fisionomia, uma pseudociência que agora é revivida por algoritmos que afirmam ser capazes de inferir características como orientação sexual ou "criminalidade" a partir de fotos ou vídeos de uma pessoa. Essa organização dos riscos é uma ferramenta poderosa para direcionar os esforços de mitigação de forma mais eficaz. Contudo, a análise aprofundada do artigo revela um cenário bastante negativo e preocupante sobre o estado atual da proteção da privacidade.

A pesquisa conclui que as soluções técnicas mais populares, como a Aprendizagem Federada e a Privacidade Diferencial, são incapazes de resolver todos os problemas apontados. Essas abordagens geralmente se concentram apenas em uma pequena parte dos riscos, como proteger dados brutos durante o treinamento do modelo, mas ignoram completamente os danos causados pelas próprias capacidades da IA. Por exemplo, um sistema pode usar Aprendizagem Federada para prever a "probabilidade de alguém ser um criminoso" a partir de uma foto, mas essa técnica não faz nada para impedir o risco

fundamental da fisionomia algorítmica ou o dano da Exclusão, que ocorre quando os dados das pessoas são usados sem seu consentimento. Da mesma forma, essas ferramentas são inúteis contra riscos de Distorção, como a criação de deepfakes de áudio ou vídeo para espalhar desinformação prejudicial. Assim, o estudo expõe uma falha crítica em que as ferramentas que temos para construir uma IA que preserve a privacidade estão muito atrás das novas e complexas maneiras pelas quais essa mesma IA pode violá-la.

09 - https://www.ted.com/talks/cathy_o_neil_the_era_of_blind_faith_in_big_data_must_end/transcript?subtitle=en

Em sua palestra no TED, a cientista de dados Cathy O'Neil alerta que nossa fé cega em algoritmos como ferramentas justas e objetivas é perigosa. Embora criados com a promessa positiva de eliminar o preconceito humano, muitos se tornam o que ela chama de "Armas de Destruição Matemática", ou seja, modelos secretos e de grande escala que causam danos reais às pessoas. Ela destaca os pontos negativos com exemplos claros, como algoritmos falhos que demitem bons professores injustamente e sistemas de policiamento preditivo que reforçam o racismo ao concentrar ações em bairros de minorias, criando uma cadeia viciosa. O'Neil argumenta que esses modelos não são neutros,são "opiniões embutidas em código", que refletem e amplificam os vieses de seus criadores aumentando cada vez mais o preconceito. Sua conclusão, no entanto, é construtiva. Em vez de abandonar a tecnologia, ela exige responsabilidade e sugere que devemos criar maneiras de auditar os algoritmos para verificar sua justiça e transparência, responsabilizando quem os cria pelos impactos que eles geram na sociedade.

10 - https://dl.acm.org/doi/pdf/10.1145/3375627.3375820

O artigo mergulha em uma camada mais profunda e complexa do debate sobre inteligência artificial, questionando não apenas as falhas da tecnologia de reconhecimento facial, mas os próprios dilemas éticos do processo de auditoria. Os autores partem de um ponto de vista positivo, reconhecendo que as auditorias algorítmicas, como o estudo "Gender Shades" , são ferramentas essenciais para expor o desempenho enviesado de sistemas. De fato, a pesquisa demonstra que a auditoria pública funciona, ao testar APIs de empresas como Microsoft e Amazon, os pesquisadores notaram que as disparidades na classificação de gênero, uma tarefa pela qual ambas já haviam sido criticadas, eram significativamente menores em comparação com concorrentes que nunca haviam sido auditados publicamente. Isso confirma que a auditoria pode forçar melhorias. No entanto, o trabalho revela um lado profundamente negativo e problemático, argumentando que essas

auditorias, mesmo bem-intencionadas, podem prejudicar as populações que buscam proteger. Um dos principais problemas é que as empresas podem ser incentivadas a corrigir apenas o problema específico apontado, ignorando falhas sistêmicas. O estudo descobriu, por exemplo, que as mesmas empresas que melhoraram na classificação de gênero ainda apresentavam enormes disparidades de desempenho em outras tarefas, como a estimativa de idade. Além disso, o artigo expõe tensões éticas inerentes ao processo. Para criar um banco de dados de auditoria que seja representativo, é preciso coletar imagens de grupos marginalizados, o que entra em conflito direto com o direito à privacidade e consentimento dessas mesmas pessoas. Outra questão crítica é que, ao focar em aumentar a precisão de uma tarefa como a classificação de gênero, a auditoria corre o risco de normalizar e legitimar uma tecnologia que é, em sua essência, excludente para indivíduos trans e não-binários. Por fim, os autores sugerem que é preciso lidar com as auditorias de forma receptiva. Elas são parte da solução, mas não a solução inteira e não devem servir como uma forma de aprovação para justificar o uso de uma tecnologia. Em vez disso, seu verdadeiro papel é expor aspectos escondidos e dificultar ou barrar a adoção de sistemas perigosos. A auditoria, portanto, não é uma meta a ser alcançada, mas sim um requisito mínimo que não se pode deixar de cumprir.

11 - https://mitsloan.mit.edu/ideas-made-to-matter/unmasking-bias-facial-recognition-algorithms

O artigo relata a experiência da pesquisadora Joy Buolamwini, que descobriu que um software de reconhecimento facial só a identificava quando ela usava uma máscara branca. Este incidente expõe a tese central do texto, a promessa de uma IA objetiva esbarra em uma realidade de profundo preconceito. Buolamwini introduz o conceito de "Sombras de Poder" para explicar como os vieses da nossa sociedade, como o patriarcado e a supremacia branca, são codificados nos dados que treinam os algoritmos. Isso cria sistemas que, por exemplo, aprendem a discriminar mulheres em processos de contratação ou falham em reconhecer rostos de pele escura. O texto conclui que a tecnologia apenas reflete o mundo que lhe apresentamos. Para que a IA seja verdadeiramente benéfica para todos, é preciso um esforço intencional para construir sistemas com dados mais justos e representativos, em vez de simplesmente usar os dados mais convenientes, que apenas perpetuam as desigualdades existentes.

12 - https://dl.acm.org/doi/pdf/10.1145/3442188.3445922

O artigo questiona a tendência da inteligência artificial em construir modelos de linguagem cada vez maiores. Os autores apontam os benefícios dessa tendência em que modelos

como BERT e GPT-3 alcançaram um desempenho impressionante em diversas tarefas. No entanto, o texto foca nos graves pontos negativos e nos riscos dessa abordagem. O argumento central é que esses modelos são, essencialmente, "papagaios estocásticos", ou seja, sistemas que unem sequências de texto de forma estatística, sem qualquer compreensão real do significado por trás das palavras. Eles criam uma perigosa ilusão de coerência, pois nós, humanos, temos a tendência de atribuir intenção e sentido a qualquer texto que pareça fluente. Essa dinâmica gera múltiplos danos. Primeiramente, o custo ambiental e financeiro para treinar esses LLMs é gigantesco, exigindo um consumo de energia gigante. Esses custos impactam desproporcionalmente as comunidades marginalizadas, que raramente se beneficiam dessas tecnologias. Em segundo lugar, os dados de treinamento, coletados de forma massiva e sem curadoria da internet, são um espelho de nossos piores vieses. Os LLMs aprendem a reproduzir e amplificar estereótipos racistas, sexistas e outras ideologias nocivas. A escala desses dados torna a documentação adequada impossível, gerando o que os autores chamam de "dívida de documentação". A conclusão do artigo é um apelo por uma mudança de foco. Em vez de uma busca incessante por tamanho, os autores recomendam que a pesquisa priorize a eficiência energética, invista recursos na curadoria e documentação cuidadosa de conjuntos de dados e explore abordagens que não dependam de quantidades imensuráveis de dados para avançar.

Opinião Pessoal

A análise conjunta das obras apresentadas revela um paradoxo central na era da Inteligência Artificial, embora a tecnologia ofereça um potencial imenso, como no auxílio ao combate a crimes, sua aplicação atual se mostra profundamente falha e não confiável. A principal razão para essa desconfiança, como fica claro em múltiplos exemplos, é que a IA herda e, por vezes, intensifica os preconceitos e vieses presentes nos dados utilizados para o seu treinamento. A experiência de pesquisadoras como Nina da Hora, que não foi reconhecida por um sistema de reconhecimento facial, expõe de forma prática e pessoal essa falha fundamental. Nesse sentido, a IA atua como um espelho da sociedade. Se os dados que a alimentam estão carregados de preconceitos estruturais, como o racismo e a discriminação de gênero, é natural que o sistema reflita essa realidade distorcida, pois foi exatamente isso que ele aprendeu a fazer. O erro, portanto, não reside na máquina, mas na mão humana que a guia. A responsabilidade recai sobre os desenvolvedores e as corporações que treinam e implementam esses sistemas, muitas vezes sem o devido cuidado no refinamento dos dados e sem a realização de testes exaustivos que garantam sua imparcialidade antes de impactarem a vida das pessoas. Quando algoritmos se tornam

"Armas de Destruição Matemática" , capazes de demitir professores injustamente ou intensificar o policiamento em bairros de minorias, a falha é humana e sistêmica. Apesar da gravidade do cenário, há um caminho claro para a correção, e a crença de que as IAs podem ser consertadas é perfeitamente plausível. As auditorias surgem como uma ferramenta indispensável nesse processo. Elas são cruciais para expor as falhas de sistemas comerciais e pressionar as grandes empresas a tomarem consciência de seus erros e agirem para repará-los, como demonstrado em estudos que forçaram melhorias em tecnologias de gigantes como a Microsoft e a Amazon. Contudo, como aponta a obra "Saving Face", essas auditorias precisam ser abrangentes e vistas com a devida complexidade, pois até mesmo o processo de auditoria possui seus próprios dilemas éticos. Elas não devem servir como um selo de aprovação definitivo, mas como um requisito mínimo para a responsabilização contínua. Em última análise, a responsabilidade por um futuro tecnológico mais justo é compartilhada. As grandes corporações têm o dever de garantir que suas IAs sejam treinadas e utilizadas de forma ética e segura. A comunidade científica, por sua vez, deve questionar modelos de incentivo como o "publicar ou perecer", que podem valorizar a velocidade em detrimento do rigor e da ética. E a nós, como sociedade, cabe um papel ativo e crítico, não devemos aceitar passivamente a tecnologia que nos é imposta. É nosso dever duvidar, exigir transparência, questionar o uso de nossos dados e demandar regras claras para o desenvolvimento e a aplicação da IA, garantindo que seu avanço sirva ao bem-estar coletivo e não à perpetuação de injustiças.