

1.1) Vieses algorítmicos são erros sistemáticos e injustos em sistemas de IA que levam a resultados tendenciosos, discriminatórios ou imprecisos para determinados grupos de pessoas. Esses vieses não são intencionais, mas sim um reflexo das falhas e preconceitos presentes no processo de criação desses modelos. Eles podem surgir de diversas formas:

- **Dados de Treinamento Enviesados:** Esta é a principal fonte de viés. Os modelos de IA aprendem com os dados que lhes são fornecidos. Se o conjunto de dados usado para treinar um modelo de reconhecimento facial, por exemplo, contiver predominantemente fotos de pessoas brancas, o modelo aprenderá a reconhecer com mais precisão esse grupo, falhando ou apresentando desempenho inferior com pessoas de pele mais escura. O sistema apenas replica o viés presente nos dados.
- **Viés de Aferição (Measurement Bias):** Ocorre quando a forma como os dados são coletados ou medidos é inconsistente. No caso do reconhecimento facial, a iluminação, a qualidade das fotos ou o ângulo podem variar de forma desigual entre diferentes grupos, afetando a precisão do sistema.
- **Viés Histórico:** O modelo reflete e perpetua preconceitos que já existem na sociedade. Por exemplo, se um modelo de IA for treinado com dados de contratação que historicamente favorecem homens para certas profissões, o modelo pode aprender a recomendar candidatos do sexo masculino com mais frequência para esses cargos.
- **Falta de diversidade nas equipes de desenvolvimento,** o que reduz a percepção de possíveis falhas.

1.2) No documentário *Coded Bias*, é destacado que os sistemas de reconhecimento facial apresentavam desempenho significativamente pior para mulheres negras, com taxas de erro superiores a 30%, enquanto para homens brancos os erros eram próximos de 1%. Um exemplo emblemático é a experiência da pesquisadora Joy Buolamwini, que percebeu que esses sistemas tinham dificuldade em detectar seu rosto quando não usava uma máscara branca. Isso acontecia porque os algoritmos haviam sido treinados majoritariamente com imagens de homens brancos, o que resultou em uma representação desproporcional dos demais grupos e levou a falhas significativas no reconhecimento de rostos de mulheres negras, evidenciando uma discriminação direta contra grupos sub-representados.

1.3) Para essa questão resolvi escolher duas estratégias técnicas e duas regulatórias.

- **Estratégias Técnicas:**

- **Aumento da diversidade dos dados de treinamento:** incluir dados representativos de diferentes etnias, gêneros, idades e contextos culturais, para reduzir a desigualdade de desempenho entre grupos.

- **Métricas de Equidade** (Fairness Audits and Explainable AI): Modelos de IA podem ser submetidos a auditorias regulares para avaliar seu desempenho em diferentes subgrupos populacionais. Ferramentas de IA Explicável (XAI) também podem ser usadas para entender como o modelo chegou a uma determinada decisão, revelando possíveis padrões tendenciosos. Isso permite que os desenvolvedores corrijam o problema antes que o sistema seja lançado.

- **Estratégias Regulatórias:**

- **Transparência e prestação de contas:** exigir que empresas divulguem informações sobre como os algoritmos são treinados, testados e avaliados.

e a partir do primeiro criar o:

- **Selo de Responsabilidade e Certificação:** Criar um selo de "IA Responsável" ou certificações obrigatórias para sistemas de alto risco, como os usados em saúde ou segurança, incentivaria as empresas a demonstrarem que seus modelos foram testados para vies e atendem a padrões de equidade e transparência.

2.1)

- o **Dartmouth Summer Research Project on Artificial Intelligence** foi conferência, considerada o "nascimento" da IA como campo de estudo, ocorreu em Dartmouth College, em Hanover, New Hampshire, durante o verão de 1956, e durou aproximadamente seis a oito semanas.
- Os principais participantes e organizadores da Conferência de Dartmouth, em 1956, foram John McCarthy, Marvin Minsky, Nathaniel Rochester e Claude Shannon, considerados os fundadores da área de Inteligência Artificial. Também estiveram presentes outros nomes importantes, como Ray Solomonoff, Allen Newell, Herbert Simon e Oliver Selfridge, entre outros. A proposta central apresentada no evento afirmava que “toda característica da inteligência pode, em princípio, ser descrita de forma tão precisa que uma máquina pode ser construída para simulá-la”. Assim, os pesquisadores se reuniram para discutir como as máquinas poderiam usar linguagem, formar conceitos, resolver problemas e até mesmo se autoaperfeiçoar. Foi nesse contexto que John McCarthy introduziu formalmente o termo

“Inteligência Artificial”, estabelecendo o marco inicial desse campo como disciplina acadêmica.

2.2)

- O reencontro em celebração ao cinquentenário da conferência original ocorreu de 13 a 15 de julho de 2006, novamente no Dartmouth College, em Hanover, New Hampshire. Essa conferência, batizada de AI@50 (“The Dartmouth Artificial Intelligence Conference: The Next Fifty Years”), foi organizada por James H. Moor, que também atuou como diretor do evento Wikipedia . O encontro reuniu cinco dos participantes originais de 1956 (John McCarthy, Marvin Minsky, Ray Solomonoff, Oliver Selfridge e Trenchard More) o que permitiu um resgate histórico direto dos pioneiros que ajudaram a fundar a disciplina.
- O evento comemorativo de 50 anos, chamado **AI@50** ou "Dartmouth Artificial Intelligence Conference: The Next Fifty Years", ocorreu de 13 a 15 de julho de 2006, novamente no Dartmouth College, o local histórico da conferência original.

Nesse reencontro, as discussões e reflexões sobre o futuro da IA foram marcadas por uma perspectiva mais pragmática em comparação ao otimismo de 1956. Os participantes reconheceram que, embora a promessa de uma "inteligência geral" ainda não tivesse se concretizado, a área havia progredido significativamente em domínios específicos. Foi observado que a **robótica** havia evoluído de autômatos fixos para máquinas que executavam tarefas complexas, como explorar Marte. O debate sobre qual abordagem dominará o futuro da IA, se seria **baseada em lógica ou probabilidade**, também foi um tema central. Além disso, as discussões destacaram o **crescimento de novas abordagens como as redes neurais** e a onipresença da IA em dispositivos e sistemas do cotidiano. A conferência serviu como um momento para refletir sobre o passado e inspirar a próxima geração de pesquisadores, com a esperança de que eles pudessem alcançar as metas que ainda estavam em aberto para o centenário da IA.

3) Em minha opinião, a maior diferença entre a visão dos pioneiros da IA em 1956 e a realidade de 2006 é que eles pareciam ter subestimado a imensa complexidade do desafio que tinham pela frente. A aposta inicial era de que, usando lógica e

simbolismo, a inteligência humana seria simulada em apenas algumas décadas. Em 1956, havia um otimismo muito grande, uma crença de que a inteligência artificial geral seria rapidamente alcançada.

Quando a gente olha para a realidade de 2006, é notável o quanto a trajetória da IA se mostrou diferente. Em vez de uma única "máquina pensante" capaz de tudo, o avanço veio de forma mais focada, com a IA se destacando em domínios específicos. As discussões no reencontro de 2006 refletiram bem essa nova realidade. Em vez de celebrar a chegada da inteligência geral, os debates se concentraram em como as abordagens baseadas em dados e probabilidades (como as redes neurais, que estavam ganhando força) poderiam resolver problemas complexos que os métodos originais não conseguiram.

A conferência de 2006 não foi sobre a vitória da IA, mas sim sobre a constatação de que o caminho era muito mais longo e desafiador do que se imaginava. As medidas discutidas foram, na verdade, sobre abrir espaço para novos pesquisadores e sobre a necessidade de abordar os problemas em aberto, inspirando a próxima geração a continuar a jornada, agora com a clareza de que não há uma solução única e rápida, mas sim um processo de evolução contínua.