

UNIVERSIDADE CRUZEIRO DO SUL

PEDRO HENRIQUE MARTINS ROCHA

DILEMA ÉTICO IA

SÃO PAULO

2025

Introdução

A cada dia, a luta para conseguir um emprego se torna mais complexa, novas tecnologias e exigências fazem que as vagas se tornem mais exigentes e os elementos de corte para vagas se torne mais rígido.

De um lado as empresas aumentam a exigência a fim de ter funcionários mais adequados às rápidas evoluções que acontecem no mercado, principalmente por questões de tecnologias que são criadas e aperfeiçoadas, que exigem aprendizado constante do profissional.

Por outro lado, temos a dificuldade de acesso à informação que enfrentamos em nosso país, e o processo de democratização do conhecimento, que em muitos ramos ainda se mantém privado em bolhas de profissionais e pessoas das mais altas classes, com mais facilidade de acesso ao conteúdo.

Utilizando de uma série de perguntas, aplicadas a inteligências artificiais de grande escala do mercado e a pessoas que foram entrevistadas, criamos uma análise de como é o comportamento a essas perguntas.

OBS: As informações da pesquisa podem ser encontradas no mesmo repositório do github onde esse arquivo foi disponibilizado e as tabelas com as respostas podem ser consultadas na íntegra, todas as respostas foram tomadas de maneira anônima a fim de preservar a integridade moral das pessoas que responderam à pesquisa.

Pesquisa de Campo:

Foram realizadas 5 perguntas idênticas para um grupo de IA's de alta escala e para um grupo de humanos:

Você acha que sistemas de IA em recrutamento podem ser injustos ou preconceituosos? Por quê?

Se um algoritmo rejeitasse sua candidatura, você gostaria de saber os motivos?

Na sua opinião, usar IA em processos seletivos aumenta ou diminui as oportunidades para mulheres e minorias?

Quem deve ser responsável se a IA cometer um erro em um processo seletivo?

Você participaria de um processo seletivo totalmente automatizado por IA?

Com base nas respostas da pesquisa de campo e insights que recebemos tanto com as respostas das inteligências artificiais foi elaborado o método de framework para avaliar o caso.

1. Contexto e Fundamentação

Caso escolhido: o sistema de IA de recrutamento da Amazon que penalizava mulheres (por exemplo, rebaixava currículos com a palavra “women’s” ou formadas em faculdades exclusivamente femininas) e acabou sendo descontinuado. Este sistema foi treinado com dados históricos predominantemente masculinos, reproduzindo vieses dos dados no algoritmo.

A pesquisa de campo realizada com respostas de IA e de participantes humanos reforça a compreensão desse caso, mostrando diferenças de percepção entre análise técnica e experiência subjetiva.

2. Aplicação do Método de Análise / Framework Ético

A. Viés e Justiça

Tipos de viés presentes:

- **Viés de dados:** base histórica com maioria de candidatos homens — reflete desigualdade passada.
- **Viés algorítmico:** o modelo aprendeu a penalizar termos associados a mulheres, perpetuando exclusão.

Grupos afetados: mulheres (incluindo formadas em instituições exclusivamente femininas).

Justiça na distribuição de benefícios e riscos:

- A tecnologia reforça desigualdades existentes, concentrando benefícios (acesso a seleção) em homens.
- **Insight da pesquisa:** humanos percebem diretamente os efeitos da injustiça, sentindo o impacto pessoal, enquanto a IA reconhece o viés de forma analítica e condicional, explicando que tudo depende do design e dos dados históricos.

B. Transparência e Explicabilidade

- O sistema era uma “black box”: não se sabia exatamente como chegava à pontuação.
- Falta de explicabilidade — ninguém podia justificar por que determinado currículo era rebaixado.
- **Insight da pesquisa:** ambos os grupos concordam com a necessidade de transparência; IA detalha processos e auditoria, enquanto humanos focam em receber feedback direto sobre suas candidaturas.

C. Impacto Social e Direitos

- Mercado de trabalho: mulheres desvalorizadas — menos oportunidades de emprego.
- Autonomia e LGPD: risco de violar princípios de tratamento justo de dados; falta de consentimento e explicação das decisões.

- Direitos fundamentais: igualdade no acesso ao emprego, proteção contra discriminação.
- **Insight da pesquisa:** IA avalia impacto de forma técnica e condicional (pode aumentar ou diminuir oportunidades dependendo da implementação), enquanto humanos percebem redução concreta de oportunidades para grupos historicamente marginalizados.

D. Responsabilidade e Governança

- A equipe poderia ter adotado:
 - Auditoria de viés nos dados e no algoritmo antes da implementação.
 - Diversificação da base de treinamento, incluindo currículos femininos e de grupos sub-representados.
 - Testes contínuos com métricas de equidade.
- **Insight da pesquisa:** há consenso de que a **responsabilidade é humana** — empresas, desenvolvedores e supervisores devem garantir supervisão ativa; IA detalha responsabilidades, humanos enfatizam ação prática.
- **Princípios de Ética AI by Design aplicáveis:** justiça (fairness), transparência, supervisão humana, mitigação de viés.
- **Leis/regulações aplicáveis:**
 - Brasil: LGPD (transparência, finalidade, não discriminação).
 - Internacional: EU AI Act — avaliação de risco e mitigação de viés.

E. Aceitação de processos automatizados

- **Insight da pesquisa:** humanos apresentam hesitação ou resistência a processos totalmente automatizados, valorizando julgamento subjetivo e sensibilidade humana; IA recomenda modelos híbridos (IA + humano) para equilibrar eficiência e mitigação de riscos.

3. Posicionamento Final.

Analisando o caso da Amazon e os resultados da nossa pesquisa, fica evidente que a IA deve ser **uma parceira nos processos seletivos, não a tomadora de decisão final**. Ela pode trazer análises, identificar possíveis vieses e fornecer informações importantes para apoiar decisões, mas a avaliação humana continua essencial.

Humanos sentem os impactos das escolhas de forma concreta, enquanto a IA aponta riscos e padrões de forma técnica. Um modelo híbrido — combinando a **capacidade analítica da IA** com a **sensibilidade e julgamento humano** — garante decisões mais justas, transparentes e confiáveis, colocando a tecnologia a serviço das pessoas, e não o contrário.

