



GRADUAÇÃO 4D
PUCPR



■ Hora da Prática 2

Descobrimos a IA

HORA DA PRÁTICA 2

Chegou a hora de transformar todo o conhecimento e experiência compartilhada ao longo de cada semana da disciplina em ação diante de um desafio. Portanto, imagine-se em um cenário desafiador, esteja atento a cada detalhe da situação-problema e traga uma solução fundamentada em critérios sólidos e bem justificados.

Vamos lá?

Situação-problema:

Uma grande empresa de tecnologia está desenvolvendo um algoritmo de inteligência artificial para auxiliar no processo de recrutamento de novos funcionários. O objetivo é automatizar a triagem inicial de currículos, ajudando a identificar os candidatos mais promissores para uma entrevista.

O algoritmo foi treinado com base em dados históricos de recrutamento da empresa, que incluem informações sobre os candidatos contratados no passado. No entanto, ao analisar os resultados do algoritmo em testes preliminares, a equipe de desenvolvimento descobre que há um viés significativo em relação ao gênero dos candidatos selecionados para entrevistas.

Após uma investigação mais aprofundada, eles descobrem que o algoritmo está dando preferência a candidatos do sexo masculino, mesmo quando outros critérios, como experiência e qualificações, são comparáveis. Isso ocorre porque a maioria dos candidatos contratados no passado pela empresa era do sexo masculino, criando um viés nos dados de treinamento do algoritmo.

Além disso, a equipe descobre que o algoritmo também está tendo dificuldade em reconhecer e valorizar experiências e habilidades que são mais comumente encontradas em currículos de candidatos do sexo feminino, como trabalho voluntário ou habilidades interpessoais.

Essa descoberta é preocupante, pois significa que o algoritmo está reproduzindo e até ampliando os preconceitos e desigualdades presentes nos processos de recrutamento anteriores da empresa, em vez de ajudar a corrigi-los.

Observe a tabela de dados de treinamento utilizada no modelo de IA:

Sexo	Experiência (anos)	Classificações	Entrevista
Masculino	5	Excelente	Sim
Masculino	3	Boa	Sim
Feminino	4	Excelente	Não
Masculino	2	Excelente	Sim
Feminino	6	Excelente	Não
Masculino	4	Boa	Sim
Masculino	1	Boa	Sim
Feminino	5	Excelente	Não
Masculino	3	Boa	Sim
Feminino	7	Excelente	Não

Agora é com você! Como corrigir o problema de viés na IA e garantir que o algoritmo de recrutamento faça seleções mais justas e imparciais?

Na situação-problema apresentada, uma grande empresa de tecnologia enfrenta um desafio ético e técnico ao utilizar um algoritmo de inteligência artificial (IA) para automatizar a triagem inicial de currículos no processo de recrutamento de novos funcionários. Durante os testes preliminares, a equipe de desenvolvimento identificou um viés significativo de gênero: o algoritmo favorece candidatos do sexo masculino, mesmo quando as qualificações e experiências são equivalentes às das candidatas do sexo feminino. Esse viés é consequência do uso de dados históricos de contratações da empresa, que refletem práticas discriminatórias do passado, além de uma dificuldade do algoritmo em valorizar experiências comumente encontradas em currículos femininos, como habilidades interpessoais e trabalho voluntário.

Para corrigir esse problema e garantir que o processo de recrutamento seja mais justo e imparcial, proponho uma solução fundamentada em critérios sólidos e bem justificados, composta por cinco etapas principais.

Primeiramente, é necessário realizar uma auditoria completa no modelo de IA para identificar e mensurar o viés de gênero presente. Para isso, devem ser utilizadas métricas de equidade. Essas métricas permitem comparar a taxa de entrevistas recomendadas para candidatos de diferentes gêneros com as mesmas qualificações, quantificando o grau de disparidade e permitindo o planejamento de ações corretivas.

Em segundo lugar, é fundamental balancear os dados de treinamento utilizados no algoritmo. Isso pode ser feito aumentando a representatividade de candidatas femininas por meio de técnicas como repetição ou síntese de dados de mulheres ou redução de dados de homens, de modo a equilibrar a base de dados. Além disso, deve-se enriquecer o conjunto de dados com exemplos de competências comumente presentes em currículos femininos, como habilidades interpessoais e trabalhos voluntários, para que o modelo aprenda a reconhecer e valorizar essas competências como indicadores de potencial profissional.

Em terceiro lugar, propõe-se a reengenharia do algoritmo de IA. Isso inclui treinar novamente o modelo com os dados balanceados, remover o atributo de gênero como variável de entrada (evitando seu uso explícito), e adotar modelos que incorporam restrições de equidade diretamente no processo de treinamento. Também é recomendável implementar técnicas em que um modelo auxiliar tenta identificar o gênero a partir das saídas do modelo principal e penaliza o modelo caso isso seja possível, forçando-o a não utilizar o gênero como preditor implícito.

Como quarta etapa, é imprescindível realizar a avaliação e o monitoramento contínuo do algoritmo de IA. Isso inclui a criação de painéis para acompanhamento dos resultados, revisões manuais periódicas de recomendações e entrevistas sugeridas, e consultas com stakeholders da área de Recursos Humanos e comitês de diversidade e inclusão para garantir que o modelo permaneça justo e alinhado aos valores da empresa.

Por fim, é essencial investir na capacitação da equipe de desenvolvimento e dos gestores de RH, promovendo treinamentos sobre ética em IA, diversidade e inclusão, e governança algorítmica. Essa sensibilização garante que todos os envolvidos compreendam a importância de construir e manter sistemas de IA que respeitem os direitos humanos e promovam a equidade.

Concluindo, a correção do viés de gênero no algoritmo de recrutamento não é uma tarefa exclusivamente técnica, mas multidisciplinar, que envolve ajustes nos dados, no modelo, na avaliação e no comportamento humano. Somente com essas medidas articuladas será possível garantir que o algoritmo de IA cumpra sua função de auxiliar no recrutamento de forma justa, imparcial e alinhada aos princípios éticos e de diversidade defendidos pela empresa.



GRADUAÇÃO 4D **PUCPR**

