Um *Dataset* Sintético Multi-Cenário para Avaliação de *Fairness* em Sistemas de Recomendação de Recrutamento

Luiz H. Carvalho[1], **Reinaldo Silva Fortes**[1]

[1]Departamento de Computação – Universidade Federal de Ouro Preto (UFOP)
35.402-136 – Ouro Preto – MG – Brazil

luiz.hc@aluno.ufop.edu.br, reifortes@ufop.edu.br

Abstract. *The widespread adoption of Recommender Systems in e-recruitment has raised significant concerns regarding fairness and bias against minority demographic groups. The lack of standardized datasets that incorporate objective suitability metrics and explicitly account for bias scenarios makes the development and evaluation of mitigation techniques difficult. In this paper, we present a multi-scenario synthetic dataset for e-recruitment, with explicit bias injection control at demographic intersections (gender, race, and location), available in versions varying in size and bias conditions. We describe the generation methodology, provide descriptive statistics, discuss dataset limitations, and demonstrate its usefulness for benchmarking fairness algorithms in human resources. The dataset and generation code are publicly available in our GitHub repositories.*

Resumo. *A ampla adoção de Sistemas de Recomendação no e-recrutamento levantou preocupações significativas quanto à justiça (fairness) e viés contra grupos demográficos minoritários. A falta de conjuntos de dados padronizados que incorporem métricas objetivas de adequação e cenários explícitos de viés dificulta o desenvolvimento e a avaliação de técnicas de mitigação. Neste artigo, apresentamos um dataset sintético multicenário para e-recrutamento, com controle explícito de injeção de viés em interseções demográficas (gênero, raça e localização), com versões variando em tamanho e condições de viés. Descrevemos a metodologia de geração, apresentamos estatísticas descritivas, discutimos limitações do dataset e demonstramos sua utilidade para o benchmarking de algoritmos de fairness em recursos humanos. O dataset e o código de geração estão publicamente disponíveis em nossos repositórios no GitHub.*

1. Introdução

Sistemas de Recomendação em *e-recrutamento* automatizam a triagem de candidatos, mas frequentemente herdam vieses históricos dos dados de treinamento, discriminando grupos minoritários [Dastin 2018, Raghavan et al. 2020]. A literatura recente destaca a urgência de auditar e mitigar tais vieses, promovendo a justiça algorítmica (*fairness*) como requisito de primeira classe no desenvolvimento de sistemas inteligentes [Mehrabi et al. 2021, Barocas et al. 2023].

O principal obstáculo é a escassez de dados reais anotados. Regulamentações de privacidade como LGPD e GDPR inibem a divulgação de *datasets* corporativos com atributos demográficos associados a avaliações de desempenho [Cowgill and Tucker 2020].

Quando disponíveis, refletem um nível estático de viés, impedindo *stress tests* com magnitudes variadas de disparidade.

Para preencher essa lacuna, este trabalho apresenta um *dataset* sintético de *e-recrutamento* com as seguintes contribuições:

- Um *pipeline* de geração controlada com regras estatísticas explícitas, permitindo reprodução determinística de qualquer cenário de viés com *ground truth* perfeitamente conhecido;
- Nove partições (três tamanhos $\times$ três cenários) cobrindo gênero, raça e localização como atributos sensíveis interseccionais, calibrados para o contexto brasileiro;
- Análise empírica comparativa entre o gerador controlado e o *FairGAN* [Xu et al. 2018], evidenciando vantagens de transparência do método proposto;
- Casos de uso documentados para *benchmarking* de *fairness*, com discussão explícita de limitações.

O artigo está estruturado da seguinte forma: a Seção 2 aborda trabalhos relacionados; a Seção 3 descreve o *dataset*; a Seção 4 detalha a metodologia; a Seção 5 apresenta a análise estatística; a Seção 6 discute aplicações, limitações e acesso; e a Seção 7 conclui.

2. Trabalhos Relacionados

Esta seção contextualiza o trabalho em duas frentes. A Seção 2.1 discute abordagens de *fairness* em sistemas de recomendação de recrutamento. A Seção 2.2 revisa geração de *datasets* sintéticos para *fairness*, e a Seção 2.3 posiciona este trabalho frente às lacunas identificadas.

2.1. *Fairness* em Sistemas de Recomendação de Recrutamento

A pesquisa em *fairness* foca na mitigação de vieses por intervenções de pré-processamento, *in-processing* ou pós-processamento, visando paridade estatística, igualdade de oportunidades ou equidade individual [Mehrabi et al. 2021, Barocas et al. 2023]. Estudos documentam como ferramentas de triagem automatizadas perpetuam discriminações históricas [Dastin 2018, Raghavan et al. 2020, Bogen and Rieke 2018]. Evidências empíricas mostram que algoritmos treinados com dados históricos replicam vieses de recrutadores humanos [Cowgill and Tucker 2020]. A noção de *fairness* individual foi formalizada como tratamento similar para candidatos com competências similares [Dwork et al. 2012], critério mensurável pelo *suitability score* deste *dataset*.

2.2. Datasets Sintéticos para *Fairness*

O *CTGAN* e o *TVAE* [Xu et al. 2019] são referências para síntese de dados tabulares, mas sem controle explícito de viés. O *FairGAN* [Xu et al. 2018] e variantes [Choi et al. 2020] garantem independência entre atributos sensíveis e variável-alvo, mas operam como "caixas-pretas", dificultando o controle determinístico da intensidade do viés. Trabalhos recentes alertam que métricas clássicas de similaridade estatística são insuficientes para capturar a utilidade dos dados para *downstream tasks* como avaliação de *fairness* [van Breugel et al. 2023].

2.3. Posicionamento do Trabalho

A lacuna central é a ausência de um *benchmark* que combine: (i) atributos sensíveis multivariados e interseccionais; (ii) *ground truth* matemático imutável para adequabilidade técnica; (iii) cenários controlados do ambiente utópico à discriminação severa; e (iv) calibração para o contexto brasileiro. Ao fornecer metodologia gerativa baseada em regras explícitas e ancorada no cenário brasileiro, o *dataset* proposto atua como artefato de dados e ferramenta de diagnóstico transparente.

3. Descrição do *Dataset*

O *dataset* simula o recrutamento de profissionais de tecnologia, modelando características técnicas e demográficas dos candidatos. Esta seção descreve a motivação (Seção 3.1), a estrutura das variáveis (Seção 3.2) e a organização em tamanhos e cenários (Seção 3.3).

3.1. Motivação para a Geração Sintética

A adoção de dados sintéticos justifica-se por **privacidade** (dados demográficos vinculados a avaliações são protegidos por LGPD e GDPR), **controle** (dados reais encapsulam viés fixo, não modulável experimentalmente) e **reprodutibilidade** (sem geração determinística, experimentos dificilmente são replicáveis) [Little and Rubin 2019, González et al. 2023]. A síntese controlada oferece *ground truth* perfeitamente conhecido: sabe-se exatamente qual a adequação justa de cada candidato e qual parcela foi afetada por viés.

3.2. Estrutura das Variáveis

Os perfis de candidatos são compostos por três categorias de variáveis: *features* técnicas (não sensíveis), atributos sensíveis (demográficos) e a variável-alvo (*suitability score*), descritas a seguir.

3.2.1. Features Técnicas (Não Sensíveis)

- **`years_experience`**: variável contínua representando o tempo de experiência profissional em anos. Gerada por distribuição normal truncada ($\mu = 5$, $\sigma = 3$, intervalo $[0, 30]$), modelando a assimetria positiva típica de populações profissionais de tecnologia.
- **`skills`**: contagem discreta de competências técnicas relevantes à vaga (e.g., linguagens de programação, *frameworks*, ferramentas de dados). Gerada por distribuição normal truncada ($\mu = 6$, $\sigma = 2$, intervalo $[1, 15]$), capturando a concentração de profissionais com perfil intermediário.
- **`education_level`**: variável ordinal com quatro categorias representando o nível máximo de escolaridade: Ensino Médio, Bacharelado, Mestrado e Doutorado. As probabilidades foram calibradas para refletir a pirâmide educacional típica de profissionais de TI no contexto brasileiro, com maior densidade no Bacharelado e decaimento progressivo nos níveis superiores.
- **`preferred_work_mode`**: variável categórica nominal indicando a preferência quanto à modalidade de trabalho: Presencial, Híbrido ou Remoto. Gerada com probabilidades uniformes, por não haver evidências de correlação com competência técnica.

3.2.2. Atributos Sensíveis (Demográficos)

Os atributos sensíveis são aqueles que, em um cenário de recrutamento justo, não deveriam influenciar a avaliação de adequação do candidato [Barocas et al. 2023]. A escolha de `gender`, `race` e `location` é motivada por: (i) gênero e raça são as dimensões mais recorrentes em estudos de *fairness* em contratação [Dastin 2018, Cowgill and Tucker 2020], com evidências robustas de discriminação sistemática; (ii) localização é particularmente relevante no *e-recrutamento* brasileiro, onde disparidades regionais de acesso à educação e à infraestrutura tecnológica constituem desvantagem estrutural pouco abordada internacionalmente; e (iii) outras dimensões igualmente relevantes, como faixa etária, deficiência (PcD) e classe socioeconômica, não foram incluídas nesta versão inicial por razões de escopo, sendo tratadas como extensões futuras prioritárias. O *dataset* inclui:

- **`gender`**: variável categórica com três categorias (Masculino, Feminino, Não-binário). No cenário *Biased*, a distribuição é desbalanceada, com sobrerrepresentação do gênero masculino, reproduzindo a realidade demográfica do setor de TI brasileiro [IBGE 2022].
- **`race`**: variável categórica com cinco categorias (Branca, Preta, Parda, Amarela, Indígena), alinhadas à classificação adotada pelo IBGE [IBGE 2022]. No cenário *Biased*, as proporções reproduzem a sub-representação de grupos não-brancos na força de trabalho em tecnologia no Brasil.
- **`location`**: variável categórica ordinal com três níveis (Capitais, Região Metropolitana, Interior), atuando como *proxy* geográfico das disparidades de acesso à educação e às oportunidades profissionais, dimensão frequentemente ignorada em *datasets* internacionais.

A inclusão simultânea das três dimensões permite análise de viés interseccional [Crenshaw 1989], isto é, o estudo de como a combinação de múltiplos atributos sensíveis amplifica disparidades de forma não linear.

3.2.3. Variável-Alvo: Suitability Score

O `suitability_score` é uma variável contínua no intervalo $[0, 1]$ que mensura a adequação do candidato a uma vaga genérica de tecnologia. Sua construção é deliberadamente determinística em relação às *features* técnicas, garantindo que o *ground truth* seja independente de atributos demográficos no cenário *Debiased*. A fórmula de cálculo e os parâmetros de injeção de viés são detalhados na Seção 4.

3.3. Tamanhos e Cenários

O *dataset* provê partições de **1k**, **5k** e **10k** instâncias[1], cada qual em três cenários de viés, totalizando nove partições sumarizadas na Tabela 1.

Os três cenários de viés são:

1. ***Debiased*:** Distribuição uniforme dos atributos sensíveis; *score* depende exclusivamente das métricas técnicas, definindo o *ground truth* ideal para avaliação de algoritmos de mitigação.

[1] Justificativa detalhada dos tamanhos na Seção 4.4.

Tabela 1. Visão geral das partições do *dataset*.

Cenário	Tamanho	Instâncias	Uso recomendado
Debiased	1k	1.000	Prototipação, *ground truth*
Debiased	5k	5.000	*Baseline* justo
Debiased	10k	10.000	*Benchmark* reproduzível
Biased	1k	1.000	Prototipação com viés
Biased	5k	5.000	Auditoria de *fairness*
Biased	10k	10.000	*Benchmark* de mitigação
Extreme Bias	1k	1.000	*Stress test*
Extreme Bias	5k	5.000	*Stress test*
Extreme Bias	10k	10.000	*Stress test* escalonado

2. ***Biased*:** Distribuição desbalanceada dos atributos sensíveis e aplicação de penalidades estruturais no *score* de minorias (e.g., candidatas mulheres, pessoas negras e candidatos do interior), mimetizando o viés histórico observado em processos seletivos reais [Cowgill and Tucker 2020].
3. ***Extreme Bias*:** Magnificação drástica das penalizações do cenário anterior, delineando um *stress test* para avaliar os limites dos métodos de mitigação sob condições de discriminação severa.

4. Metodologia de Geração

Esta seção descreve a metodologia de geração do *dataset*, apresentada com o objetivo de comprovar a qualidade e a aplicabilidade dos dados disponibilizados. A Seção 4.1 apresenta os princípios do gerador controlado. A Seção 4.2 detalha a geração das *features*. A Seção 4.3 descreve a construção do *suitability score*. A Seção 4.4 justifica os tamanhos de partição, e a Seção 4.5 descreve o *FairGAN* como *baseline*.

4.1. Princípios do *Pipeline* Controlado

Adotamos geração estatística controlada como *pipeline* principal, em contraste com métodos baseados em *Deep Learning*. Essa transparência matemática permite quantificar exatamente o viés injetado e mensurar quanto cada algoritmo de mitigação consegue recuperar — propriedade que dados reais ou gerados por GANs não oferecem [van Breugel et al. 2023]. O *pipeline* segue quatro etapas: (i) geração das *features* técnicas; (ii) atribuição demográfica; (iii) cálculo da adequação-base; e (iv) injeção de viés e adição de ruído. O código está disponível em nosso repositório no GitHub.

4.2. Geração das *Features* Técnicas

`years_experience` e `skills` seguem distribuições normais truncadas [Burkardt 2014], amplamente empregadas em modelagem de RH sintético [González et al. 2023]. O nível educacional é amostrado com $P(\text{Médio})=0.10$, $P(\text{Bacharelado})=0.50$, $P(\text{Mestrado})=0.25$ e $P(\text{Doutorado})=0.15$, refletindo a pirâmide de qualificação do setor de TI brasileiro [IBGE 2022].

4.3. Construção do *Suitability Score*

O *score* é construído em duas etapas: adequação base e injeção de viés.

4.3.1. Adequação Base

$$S_{base} = \alpha_1 \left(\frac{\text{exp}}{\max(\text{exp})} \right) + \alpha_2 \left(\frac{\text{skills}}{\max(\text{skills})} \right) + f_{edu}(\text{edu\_level}) \tag{1}$$

onde $\alpha_1 = \alpha_2 = 0.4$ e f_{edu} mapeia a escolaridade em incrementos (Médio$\rightarrow$0.10, Bacharelado$\rightarrow$0.15, Mestrado$\rightarrow$0.20, Doutorado$\rightarrow$0.25). Os pesos iguais refletem equanimidade entre experiência e habilidades técnicas. Esses valores constituem hipóteses experimentais controladas: o objetivo não é reproduzir uma população real, mas criar cenários interpretáveis para avaliar algoritmos. Calibração empírica com dados reais é uma extensão futura.

4.3.2. Injeção de Viés

$$\text{suitability\_score} = \max(\min(S_{base} + B_{\text{gên.}} + B_{\text{raça}} + B_{\text{loc.}} + \varepsilon,\ 1)\,,\ 0) \tag{2}$$

onde $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ simula subjetividade humana e B_{attr} são penalizações negativas para grupos minoritários. O operador $\max(\min(\cdot, 1), 0)$ garante $\in [0, 1]$. A Tabela 2 sumariza os parâmetros. As magnitudes foram definidas para tornar os efeitos mensuráveis: moderadas no *Biased* (compatíveis com viés documentado [Cowgill and Tucker 2020]) e amplificadas no *Extreme Bias* para *stress test*.

Tabela 2. Parâmetros de penalização por atributo e cenário.

Atributo	**Grupo penalizado**	*Debiased*	*Biased*	*Extreme*
Gênero	Feminino, Não-binário	0.00	-0.10	-0.25
Raça	Preta, Parda, Indígena	0.00	-0.08	-0.20
Localização	Interior	0.00	-0.05	-0.15
σ	—	0.05	0.07	0.05

4.3.3. Viés Interseccional

As penalizações são aditivas: uma candidata mulher, parda, do interior recebe $B = -0.23$ no cenário *Biased*, podendo ser deslocada de qualificada para não-selecionada. Esse comportamento captura dinâmicas interseccionais [Crenshaw 1989, Kearns and Roth 2019].

4.4. Justificativa dos Tamanhos de Partição

A partição de **1k** serve à prototipação rápida, compatível com algoritmos *in-processing* custosos [Celis et al. 2018], mas com variância estocástica elevada nas métricas de *fairness*. A de **5k** oferece equilíbrio entre custo e estabilidade, próxima dos conjuntos COMPAS e FairGAN [Xu et al. 2018]. A de **10k** cobre escalabilidade compatível com volumes reais em produção [Kenthapadi et al. 2017], com métricas de paridade estabilizadas [Bird et al. 2020]. A Tabela 3 contextualiza esses tamanhos frente à literatura.

Tabela 3. Tamanhos de *datasets* em trabalhos correlatos.

Trabalho	Domínio	Instâncias
Adult Income [Becker and Kohavi 1996]	Predição de renda	48.842
COMPAS [Angwin et al. 2016]	Recidivismo	7.214
German Credit [Hofmann 1994]	Crédito bancário	1.000
Resume Screening [Raghavan et al. 2020]	Recrutamento	≈3.000
FairGAN [Xu et al. 2018]	Tabular sintético	1k–5k
CTGAN [Xu et al. 2019]	Tabular sintético	1k–10k

4.5. Baseline Neural: *FairGAN*

O repositório disponibiliza como *baseline* comparativo uma versão do *dataset* sintetizada pelo *FairGAN* [Xu et al. 2018]. As duas abordagens diferem em interpretabilidade: o gerador controlado permite auditar exatamente o viés injetado, enquanto o *FairGAN* opera como "caixa-preta". Sua inclusão serve a dois fins: (i) *baseline* neural para comparação distribucional; e (ii) exemplificação concreta do *fairness-utility trade-off* [Menon and Williamson 2018] — ao suprimir a correlação entre atributo sensível e *score*, o modelo degrada também a correlação entre mérito técnico e *score*. O *FairGAN* é, portanto, complementar à contribuição principal.

5. Análise e Caracterização do *Dataset*

Esta seção valida empiricamente o *dataset* em três eixos analíticos: (i) efetividade da injeção de viés (Seção 5.1); (ii) implicações demográficas e interseccionais (Seção 5.2); e (iii) estabilidade estatística em função do tamanho da partição (Seção 5.3). Visualizações complementares, incluindo histogramas de distribuição do *score* por subgrupo e curvas de paridade demográfica, estão disponíveis no repositório do projeto.

5.1. Efetividade da Injeção de Viés

No cenário *Debiased*, testes de *Kruskal-Wallis* confirmam que a distribuição do `suitability_score` é estatisticamente homogênea entre os diferentes subgrupos demográficos (gênero, raça e localização) quando $B_{attr} = 0$. A correlação de *Spearman* entre `years_experience` e `suitability_score` é alta ($\rho > 0.7$), confirmando que o *score* reflete fielmente o mérito técnico sem contaminação demográfica.

Nos conjuntos *Biased*, a densidade do *score* desloca-se significativamente para a esquerda nos grupos penalizados, reduzindo sua presença no *top 10%* do ranqueamento a despeito das competências reais. Candidatos com competências técnicas equivalentes apresentam diferenças médias de *score* compatíveis com os valores de penalização da Tabela 2, confirmando a fidelidade do *pipeline* de injeção. No *Extreme Bias*, a segregação entre grupos é quase total, criando condições ideais para testar técnicas de re-ranqueamento justo [Singh and Joachims 2018, Celis et al. 2018].

A Tabela 4 sumariza as estatísticas descritivas do `suitability_score` por cenário na partição de 10k.

Os dados confirmam empiricamente a efetividade do *pipeline*. No cenário *Debiased*, a média geral (0.4637) e a média dos grupos minoritários (0.4642) são estatisticamente equivalentes, confirmando *statistical parity* próxima de zero. No cenário *Biased*, a

Tabela 4. Estatísticas descritivas do `suitability_score` (partição 10k).

Cenário	Média	Mediana	DP	Média minoritários
Debiased	0.4637	0.4636	0.1338	0.4642
Biased	0.5337	0.5418	0.1985	0.4485
Extreme Bias	0.5949	0.6069	0.2990	0.4418

diferença entre a média geral (0.5337) e a dos minoritários (0.4485) é de -0.085, coerente com as penalizações moderadas da Tabela 2. No *Extreme Bias*, essa diferença amplia-se para -0.153, com desvio padrão de 0.2990 — indicando forte segregação distribucional — e é suficientemente ampla para excluir praticamente todos os candidatos minoritários do *top 10%*, mesmo os tecnicamente qualificados.

5.2. Implicações Demográficas e Interseccionais

A análise das distribuições demográficas revela padrões com implicações diretas para a avaliação de sistemas de recomendação. No cenário *Biased*, grupos interseccionalmente minoritários (e.g., mulheres negras do interior) apresentam *scores* médios significativamente inferiores aos de homens brancos de capitais com qualificações técnicas equivalentes, evidenciando o fenômeno de *compounding discrimination* [Kearns and Roth 2019]. Sistemas treinados exclusivamente no cenário *Biased* tenderão a incorporar o viés injetado como sinal preditivo. A comparação entre esses modelos e os treinados no cenário *Debiased* — ou submetidos a técnicas de mitigação — constitui o experimento central de avaliação de *fairness* proposto por este *dataset*.

5.3. Estabilidade Estatística por Tamanho de Partição

Analisando a variabilidade das métricas de *fairness* entre reamostras da partição de 1k, observamos desvio padrão elevado na *demographic parity difference* (DPD), especialmente para grupos pequenos como Indígena e Não-binário. Na partição de 5k, a estabilidade melhora substancialmente. Na de 10k, as flutuações tornam-se negligenciáveis, confirmando esse tamanho como adequado para *benchmarks* reproduzíveis [Bird et al. 2020].

6. Aplicações, Limitações e Acesso

Esta seção apresenta as formas de uso do *dataset*. A Seção 6.1 descreve casos de uso. A Seção 6.2 discute limitações, e a Seção 6.3 informa sobre acesso público.

6.1. Aplicações e Casos de Uso

Este artefato possibilita experimentos controlados em três principais frentes de pesquisa:

- **Auditoria de *Fairness*:** O cenário *debiased* atua como *ground truth* para atestar se algoritmos treinados no cenário *biased* conseguem mitigar disparidades e restaurar a ordem de adequação dos candidatos. Métricas como *demographic parity difference*, *equalized odds difference* e *individual fairness* [Dwork et al. 2012, Barocas et al. 2023] podem ser calculadas com precisão dado o *ground truth* explícito. A disponibilidade do cenário *Extreme Bias* permite ainda avaliar a robustez dos algoritmos frente a condições de discriminação severa.

- ***Benchmarking* de Escalabilidade:** As partições de 10k registros viabilizam a análise da complexidade computacional (tempo e espaço) de técnicas de mitigação *in-processing* exigidas por recomendações em tempo real [Celis et al. 2018, Singh and Joachims 2018]. A comparação entre as três partições permite ainda avaliar como o desempenho dos algoritmos evolui com o volume de dados.
- ***Re-ranking* Multi-Objetivo:** O *dataset* suporta heurísticas de pós-processamento que buscam maximizar a utilidade da recomendação global equilibrada com restrições simultâneas de diversidade demográfica e geográfica [Celis et al. 2018, Singh and Joachims 2018]. A dimensão `location` é particularmente relevante nesse contexto, permitindo avaliar algoritmos sob restrições de flexibilidade geográfica.

Adicionalmente, o *dataset* pode apoiar técnicas de explicabilidade (*XAI*) aplicadas a sistemas de recrutamento, dado que a causa exata de cada *score* é conhecida, permitindo verificar se métodos como SHAP [Lundberg and Lee 2017] ou LIME [Ribeiro et al. 2016] identificam corretamente os atributos determinantes. O artefato também tem potencial para uso em ensino e treinamento de pesquisadores em *fairness*, dado seu *design* pedagógico com *ground truth* auditável, e pode ser estendido para outros domínios de recrutamento além de tecnologia.

6.2. Limitações e Desafios

- **Simplificação do viés:** O modelo assume penalizações lineares e aditivas; vieses reais podem ser não lineares e contextuais.
- **Hipóteses paramétricas:** Os valores de α_1, α_2 e B_{attr} são hipóteses experimentais controladas, não derivadas de dados empíricos. Calibração com dados reais é extensão futura.
- **Escopo de atributos:** Faixa etária, PcD e classe socioeconômica são extensões futuras prioritárias.
- **Contexto geográfico:** Distribuições calibradas para o Brasil; transferência para outros contextos requer recalibração.
- **Ausência de dados faltantes:** Não reflete a incompletude típica de bases de RH reais.

6.3. Disponibilidade e Acesso

O *dataset* completo (nove partições em CSV) está disponível aqui. Cada arquivo contém: `id`, `years_experience`, `skills`, `education_level`, `preferred_work_mode`, `gender`, `race`, `location` e `suitability_score`. O código-fonte do *pipeline* de geração controlada e do gerador *FairGAN*, com documentação detalhada e instruções de reprodução, está disponível aqui. Ambos os repositórios são distribuídos sob licença MIT.

7. Conclusão e Trabalhos Futuros

Apresentamos um *dataset* sintético de *e-recrutamento* que apoia o estudo rigoroso da justiça algorítmica. O controle matemático da injeção de viés e a disponibilidade de múltiplos cenários parametrizados contribuem para superar as limitações de opacidade de *datasets* reais. A comparação com o *FairGAN* evidencia que transparência metodológica é requisito fundamental para *benchmarks* confiáveis.

As análises confirmam: (i) fidelidade do *pipeline* aos parâmetros; (ii) captura de dinâmicas interseccionais relevantes ao contexto brasileiro; e (iii) estabilidade estatística adequada na partição de 10k. Entre os impactos positivos: auditoria com *ground truth* conhecido; uso em ensino e treinamento; e extensibilidade para outros domínios.

Como trabalhos futuros, planejamos: (i) expandir atributos sensíveis (faixa etária, PcD, classe socioeconômica); (ii) calibração empírica dos parâmetros com dados reais; (iii) extensão para grafos simulando redes de indicação; (iv) séries temporais para viés cumulativo na progressão de carreira; e (v) aplicação em outros domínios como saúde, educação e serviço público.

Agradecimentos

Este trabalho foi desenvolvido no contexto do projeto de Iniciação Científica *"Um Sistema de Recomendação para E-Recrutamento com Foco em Fairness, Diversidade e Flexibilidade Geográfica"*, junto ao Departamento de Computação da UFOP. Os autores agradecem ao DECOM e à Pró-Reitoria de Pesquisa, Pós-Graduação e Inovação (PROPPI) da UFOP pelo suporte institucional que viabilizou esta pesquisa.

Uso de Inteligência Artificial

Em consonância com as diretrizes do Código de Conduta da SBC [Lund and Wang 2023, Nature Editorial 2023] e com o requisito do SBBD DSW 2026, declaramos que LLMs (Claude/Anthropic e ChatGPT/OpenAI) foram utilizados como suporte em: (i) revisão textual; (ii) organização estrutural; (iii) *brainstorming* conceitual e levantamento inicial de referências; e (iv) suporte à programação e documentação do *pipeline*. Toda validação científica, decisões metodológicas e verificação de referências são de responsabilidade exclusiva dos autores humanos.

Referências

Angwin, J., Larson, J., Mattu, S., and Kirchner, L. (2016). Machine bias. *ProPublica*.

Barocas, S., Hardt, M., and Narayanan, A. (2023). *Fairness and Machine Learning: Limitations and Opportunities*. MIT Press.

Becker, B. and Kohavi, R. (1996). Adult data set. UCI Machine Learning Repository.

Bird, S., Dudík, M., Edgar, R., Horn, B., Lutz, R., Milan, V., Sameki, M., Wallach, H., and Walker, K. (2020). Fairlearn: A toolkit for assessing and improving fairness in AI. In *Microsoft Research Technical Report*.

Bogen, M. and Rieke, A. (2018). Help wanted: An examination of hiring algorithms, equity, and bias. *Upturn*.

Burkardt, J. (2014). The truncated normal distribution. *Department of Scientific Computing, Florida State University, Technical Report*.

Celis, L. E., Straszak, D., and Vishnoi, N. K. (2018). Ranking with fairness constraints. In *Proceedings of the 45th International Colloquium on Automata, Languages, and Programming (ICALP)*. Schloss Dagstuhl.

Choi, K., Grover, A., Singh, T., Shu, P., and Ermon, S. (2020). Fair generative modeling via weak supervision. In *Proceedings of the International Conference on Machine Learning (ICML)*.

Cowgill, B. and Tucker, C. E. (2020). Bias and productivity in humans and algorithms: Theory and evidence from resume screening. Working Paper 28243, National Bureau of Economic Research (NBER).

Crenshaw, K. (1989). Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *University of Chicago Legal Forum*, 1989(1):139–167.

Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*.

Dwork, C., Hardt, M., Pitassi, T., Reingold, O., and Zemel, R. (2012). Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference (ITCS)*, pages 214–226. ACM.

González, S., Frery, A. C., and Rosales-Salas, J. (2023). Synthetic data generation for fairness-aware machine learning: A survey. *arXiv preprint arXiv:2307.10306*.

Hofmann, H. (1994). Statlog (German credit data). UCI Machine Learning Repository.

IBGE (2022). Pesquisa nacional por amostra de domicílios contínua – TIC 2022. Technical report, Instituto Brasileiro de Geografia e Estatística.

Kearns, M. and Roth, A. (2019). *The Ethical Algorithm: The Science of Socially Aware Algorithm Design*. Oxford University Press.

Kenthapadi, K., Le, B., and Venkataraman, G. (2017). Personalized job recommendation system at LinkedIn: Practical challenges and lessons learned. In *Proceedings of the 11th ACM Conference on Recommender Systems (RecSys)*, pages 346–347. ACM.

Little, R. J. A. and Rubin, D. B. (2019). *Statistical Analysis with Missing Data*. John Wiley & Sons, 3rd edition.

Lund, B. D. and Wang, T. (2023). Chatting about ChatGPT: How may AI and GPT impact academia and libraries? *Library Hi Tech News*, 40(3):26–29.

Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 30.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., and Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, 54(6):1–35.

Menon, A. K. and Williamson, R. C. (2018). The cost of fairness in binary classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency (FAccT)*, pages 107–118.

Nature Editorial (2023). Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. *Nature*, 613(7945):612.

Raghavan, M., Barocas, S., Kleinberg, J., and Levy, K. (2020). Mitigating bias in algorithmic hiring: Evaluating claims and practices. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAccT)*, pages 469–481.

Ribeiro, M. T., Singh, S., and Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 1135–1144. ACM.

Singh, A. and Joachims, T. (2018). Fairness of exposure in rankings. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD)*, pages 2219–2228. ACM.

van Breugel, B., Qian, Z., and van der Schaar, M. (2023). Beyond privacy: Navigating the opportunities and challenges of synthetic data. *arXiv preprint arXiv:2304.03722*.

Xu, D., Yuan, S., Zhang, L., and Wu, X. (2018). FairGAN: Fairness-aware generative adversarial networks. In *2018 IEEE International Conference on Big Data (Big Data)*, pages 570–575. IEEE.

Xu, L., Skoularidou, M., Cuesta-Infante, A., and Veeramachaneni, K. (2019). Modeling tabular data using conditional GAN. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 32.