

Encaixe Perfeito: Um Jogo Educacional sobre Vieses Algorítmicos em Sistemas de Recrutamento e Seleção

Barbara Vilma de Melo do Nascimento Emanuel Felipe Duarte

Relatório Técnico - IC-PFG-25-30

Projeto Final de Graduação

2025 - Dezembro

UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE COMPUTAÇÃO

The contents of this report are the sole responsibility of the authors.
O conteúdo deste relatório é de única responsabilidade dos autores.

Encaixe Perfeito: Um Jogo Educacional sobre Vieses Algorítmicos em Sistemas de Recrutamento e Seleção

Barbara Vilma de Melo do Nascimento* Emanuel Felipe Duarte†

Resumo

O uso de sistemas de Inteligência Artificial (IA) para automatizar etapas do processo de recrutamento e seleção tem crescido de forma acelerada nos últimos anos, sendo adotado por empresas globais em larga escala. Entretanto, estudos recentes demonstram que esses sistemas frequentemente reproduzem — e até amplificam — desigualdades históricas presentes nos dados que os treinam, levando à exclusão injustificada de candidatos e à consolidação de vieses estruturais. Apesar da relevância do tema, a compreensão pública sobre como tais vieses emergem e por que persistem ainda é limitada. Este projeto apresenta o desenvolvimento do Encaixe Perfeito, um jogo educacional interativo implementado em Python/Pygame, cujo objetivo é tornar visíveis os mecanismos de viés algorítmico em sistemas de contratação. O projeto combina elementos de narrativa interativa, visualização lúdica de dados e simulação comportamental para promover consciência crítica sobre a automação de decisões sensíveis.

1 Introdução

Nos últimos anos, o uso de Inteligência Artificial (IA) em processos de recrutamento e seleção tornou-se amplamente difundido. Estimativas recentes apontam que 99% das empresas da Fortune 500 já utilizam algum tipo de automação em suas práticas de contratação [1]. A promessa central dessas tecnologias é acelerar etapas do processo seletivo, padronizar avaliações e reduzir vieses humanos. No entanto, evidências acumuladas demonstram que essa expectativa muitas vezes não se concretiza. Em vez de eliminar desigualdades, sistemas automatizados podem perpetuar ou amplificar preconceitos existentes, reproduzindo padrões históricos de discriminação presentes nos dados com os quais são treinados.

Diversos estudos e casos reais recentes evidenciam que sistemas de IA podem perpetuar e até amplificar preconceitos existentes, levando a práticas de contratação discriminatórias se não houver cuidado na sua concepção e uso. Diante disso, o tema do viés algorítmico em ferramentas de recrutamento tornou-se central em debates acadêmicos, jurídicos e éticos, especialmente a partir de 2022, com atenção voltada a como esses sistemas reproduzem preconceitos de gênero, raça, idade, entre outros, e o que pode ser feito para mitigar tais vieses.

Nesse contexto, e em alinhamento com grande desafio “Implicações da Inteligência Artificial na IHC” [2], surge a necessidade de ferramentas pedagógicas que tornem visível a lógica por trás desses sistemas, permitindo que usuários compreendam como suas próprias decisões influenciam o comportamento subsequente de modelos automatizados.

*Instituto de Computação, UNICAMP, 13083-852 Campinas, SP. b167109@dac.unicamp.br

†Instituto de Computação, UNICAMP, 13083-852 Campinas, SP. emanuel1@ic.unicamp.br

1.1 Objetivo

Considerando o contexto apresentado, o objetivo deste trabalho é desenvolver um jogo digital educativo, utilizando Python e a biblioteca Pygame, capaz de simular de forma interativa o funcionamento de sistemas automatizados de recrutamento, evidenciando como decisões humanas enviesadas podem treinar algoritmos e gerar impactos sociais significativos, a fim de promover conscientização crítica sobre viés algorítmico em processos de seleção.

Para atingir esse objetivo geral, definem-se os seguintes objetivos específicos:

1. Simular um processo de contratação no qual o jogador toma decisões individuais sobre candidatos, reproduzindo a dinâmica real de sistemas de triagem automática.
2. Demonstrar, por meio de mecânicas de jogo e feedback narrativo, como escolhas humanas são incorporadas em modelos automatizados e influenciam seu comportamento futuro.
3. Expor as consequências de um algoritmo treinado com dados enviesados, por meio de telas narrativas e “manchetes de jornal” que mostram efeitos sociais amplificados.
4. Proporcionar uma experiência pedagógica acessível, utilizando elementos de gamificação e narrativa interativa, que permita ao usuário compreender fenômenos como viés algorítmico, discriminação automatizada e riscos éticos associados ao uso de IA no recrutamento.
5. Contribuir para discussões acadêmicas e formativas sobre o uso responsável de IA em recursos humanos, reforçando a importância da supervisão humana, transparência e auditoria contínua de modelos automatizados.

2 Revisão Bibliográfica

2.1 IA em recrutamento e potenciais tendências discriminatórias

Atualmente, a inteligência artificial pode intervir em diversas etapas do processo seletivo, desde a divulgação de vagas até as entrevistas. Cada uma dessas aplicações traz riscos potenciais de viés.

Na etapa de divulgação de vagas, plataformas utilizam algoritmos para direcionar anúncios de emprego a públicos específicos. Isso pode limitar quem enxerga determinadas oportunidades — por exemplo, anúncios de vagas tradicionalmente associadas a homens foram exibidos de forma desproporcional para usuários masculinos em redes sociais, mesmo quando o empregador não estipulou segmentação por gênero [3]. Esse tipo de direcionamento algorítmico reforça estereótipos e pode impedir que candidatos qualificados de grupos sub-representados sequer tenham acesso às vagas.

Na triagem de currículos, sistemas de aprendizado de máquina são utilizados para filtrar e ranquear candidatos de acordo com sua adequação à vaga. Contudo, ao serem treinados em dados históricos da empresa, esses modelos podem reproduzir preconceitos passados. Se contratações anteriores privilegiaram homens ou candidatos de certos grupos raciais, o algoritmo tende a perpetuar essas escolhas, penalizando características associadas a mulheres ou minorias.

Outro exemplo são as entrevistas por vídeo automatizadas. Ferramentas como as da HireVue avaliam expressões faciais, entonação e fala dos candidatos para pontuar seu desempenho. Pesquisas independentes, porém, apontam que esses sistemas podem reproduzir vieses de visão computacional e reconhecimento de voz, prejudicando candidatos com sotaque estrangeiro, pessoas com deficiência auditiva ou indivíduos de minorias raciais. Em casos documentados, a falta de legendas acuradas comprometeu entrevistas com candidatos surdos, gerando avaliações negativas injustas.

Por fim, também surgiram plataformas gamificadas e testes psicométricos automatizados para avaliar habilidades. Embora possam engajar candidatos e padronizar avaliações, existem lacunas que a IA não percebe ou acomoda. Por exemplo, desempenhos fracos em minijogos online podem refletir baixa familiaridade com jogos digitais, não falta de competência para o trabalho visado [3]. Candidatos sem acesso frequente a tecnologia ou com menor letramento digital – muitas vezes de faixas etárias mais altas ou contextos socioeconômicos diversos – podem ser indevidamente desfavorecidos nesses filtros. Além disso, faltas de acessibilidade (como ausência de adaptações para pessoas com deficiência visual, motora etc.) nesses testes podem excluir talentos que apenas necessitariam de um formato diferente para demonstrar suas aptidões.

2.2 Casos emblemáticos e evidências de viés algorítmico

Diversos casos emblemáticos nos últimos anos ilustram como sistemas de recrutamento baseados em IA podem discriminar certos grupos.

Amazon — Um dos primeiros casos divulgados foi o da Amazon, que desenvolveu um algoritmo interno para triagem de currículos. Em 2018, a empresa admitiu ter abandonado o sistema após descobrir que ele sistematicamente desfavorecia candidatas mulheres em vagas técnicas [4, 5]. O modelo havia sido treinado em dados de contratações passadas, marcadas por um histórico majoritariamente masculino, o que fez com que o sistema incorporasse e reproduzisse esse viés de gênero em seu funcionamento. Entre os comportamentos identificados, o algoritmo rebaixava currículos que mencionavam palavras como *women’s* (por exemplo, em “time de rugby feminino” ou nome de faculdade feminina) e privilegiava termos mais comuns em descrições masculinas (“executou”, “liderou”, etc.). Apesar de tentativas dos engenheiros para corrigir o problema, a tendência discriminatória persistiu e o projeto foi cancelado. O caso Amazon tornou-se um símbolo claro de como dados históricos enviesados podem levar uma IA a reproduzir desigualdades de gênero, mesmo sem intenção explícita dos desenvolvedores.

HireVue e o caso Intuit — A HireVue, fornecedora líder de software de entrevistas por vídeo com IA, enfrentou em 2025 uma denúncia de discriminação apresentada pela ACLU do Colorado. No caso, uma funcionária surda e indígena da empresa Intuit alegou ter sido preterida em uma promoção devido à avaliação de um algoritmo da HireVue, que teria interpretado erroneamente sua comunicação por voz e vídeo [6, 7]. Segundo a queixa, a candidata solicitou uma legenda humana como acomodação (devido à sua surdez), mas a empresa negou e fez a entrevista usando apenas transcrição automática — que não conseguiu captar adequadamente sua fala nem seu dialeto, prejudicando seu desempenho. O feedback gerado pela IA destacou “problemas de comunicação” e ela foi rejeitada para o cargo, levando os autores da denúncia a alegarem discriminação por deficiência e raça (violando tanto a Lei dos Americanos com Deficiência – ADA – quanto o Título VII da Lei dos Direitos Civis). Embora a HireVue negue o uso de IA específica nesse caso, ele ilustra preocupações mais amplas sobre viés contra pessoas com deficiência auditiva (que podem ser mal avaliadas por sistemas de reconhecimento de voz) e viés racial/cultural (modelos treinados em padrões de fala majoritariamente brancos tendem a ter menor acurácia com falantes de outras origens). Esse episódio repercutiu como um alerta de que ferramentas automatizadas devem prever acomodações e ajustes para não excluir talentos diversos injustamente.

Workday — Em fevereiro de 2024, uma ação coletiva nos EUA (*Mobley vs. Workday*) acusou a Workday Inc., fornecedora de software de RH com IA, de promover um “padrão de discriminação” contra candidatos negros, pessoas acima de 40 anos e indivíduos com deficiência [3]. O sistema automatizado da Workday — integrado a processos seletivos de várias empresas clientes — teria rejeitado desproporcionalmente candidatos desses grupos, segundo a denúncia, violando leis civis e trabalhistas americanas. Trata-se do primeiro grande processo judicial contra um provedor de IA

de contratação por suposta discriminação algorítmica em larga escala. Ainda que o caso esteja em andamento, ele sinaliza que viés de raça, idade e deficiência em algoritmos de recrutamento pode acarretar responsabilidade legal para fornecedores e empregadores. Ademais, destaca que candidatos impactados estão buscando reparação jurídica, rompendo a dificuldade histórica de contestar decisões opacas de IA em seleções.

iTutorGroup — Um exemplo flagrante de viés programado ocorreu com a iTutorGroup, empresa de tutorias online. Investigação divulgada em 2022 revelou que seu sistema de triagem fora deliberadamente configurado para excluir candidatos acima de certa idade: mulheres com mais de 55 anos e homens com mais de 60 eram automaticamente rejeitados antes mesmo da análise curricular [8]. Esse caso, tornado público via ação da EEOC (Equal Employment Opportunity Commission), evidenciou um caso de discriminação etária automatizada. A empresa acabou enfrentando um acordo legal em 2023, com pagamento de multa e exigência de reconsiderar as candidaturas afetadas. Mais do que um viés “inconsciente” aprendido de dados, aqui houve intenção explícita de discriminar, expondo que a automação também pode ser usada para implementar políticas ativamente preconceituosas — o que é ilegal em muitos países. A repercussão do caso reforçou chamados para maior transparência nos algoritmos: se essa lógica de exclusão não tivesse sido auditada, o filtro etário poderia permanecer oculto.

Estudos empíricos recentes — Além dos casos corporativos, pesquisas acadêmicas recentes forneceram evidências quantitativas dos vieses em sistemas de contratação por IA. Um estudo de 2024 da Universidade de Washington auditou três modelos de linguagem de última geração (LLMs) usados para classificar currículos e revelou viés consistente tanto de raça quanto de gênero [1]. Os pesquisadores submeteram mais de 550 currículos reais a esses algoritmos, manipulando apenas os nomes dos candidatos para refletir identidades diferentes (nomes tipicamente associados a homens ou mulheres, brancos ou negros). Os resultados foram alarmantes: currículos com nomes associados a pessoas brancas foram favorecidos em 85% das comparações, enquanto nomes femininos foram favorecidos em apenas 11% dos casos — e currículos de homens negros nunca foram escolhidos em detrimento de candidatos homens brancos. Ou seja, a IA apresentou um forte “viés branco-masculino”. Esse padrão não era simplesmente a soma dos vieses de raça e gênero, mas revelava efeitos interseccionais complexos. Estudos menores já apontavam na mesma direção, mas o estudo da Universidade de Washington, em maior escala, solidificou evidências de que mesmo as mais novas ferramentas de IA podem incorporar vieses sistêmicos, levantando preocupação sobre seu uso desenfreado sem supervisão independente.

Os casos e pesquisas acima confirmam que o viés algorítmico em recrutamento não é hipotético, mas real e multifacetado. Modelos tendem a internalizar e perpetuar preconceitos humanos — como sintetizou uma analista, “essas ferramentas não eliminam o viés humano, apenas o replicam por meio do software” [6]. As consequências vão desde oportunidades perdidas para indivíduos qualificados de grupos sub-representados até possíveis violações legais de igualdade de emprego.

2.3 Questões éticas e respostas regulatórias

Além da aplicação de leis existentes, novas iniciativas regulatórias específicas estão em curso para lidar com IA no recrutamento. Em nível local, uma das legislações pioneiras é a da cidade de Nova York (EUA): desde 2023, entrou em vigor uma lei que exige que empregadores submetam suas ferramentas automatizadas de decisão de contratação a auditorias anuais de viés, além de notificarem candidatos sobre o uso dessas ferramentas [6]. Essa medida – a primeira do tipo nos EUA – busca tornar mais transparente o uso de IA em seleção e garantir que elas passem por verificações de impacto (por exemplo, aferindo se rejeitam desproporcionalmente minorias). Outras jurisdições estaduais, como Illinois e Colorado, também implementaram ou propuseram normas

exigindo transparência e equidade em entrevistas virtuais e avaliações algorítmicas.

Na União Europeia, discute-se um marco regulatório abrangente: o *AI Act* europeu, previsto para vigorar plenamente nos próximos anos, classifica sistemas de IA para emprego (incluindo recrutamento, promoções e demissões) como de “alto risco”, sujeitando-os a requisitos rigorosos de conformidade [9]. Pelo *AI Act*, ferramentas de contratação deverão passar por avaliações de conformidade e gerenciamento de risco, garantindo qualidade dos dados (representatividade, ausência de vieses evidentes) e implementando obrigações de transparência e supervisão humana na tomada de decisão. Práticas consideradas de “risco inaceitável” – como sistemas que inferem características sensíveis (emoções, traços de personalidade) de forma manipulativa ou sem base científica – poderão ser banidas do uso em contexto de trabalho.

Em suma, há um movimento global crescente para regulamentar a IA em recursos humanos, a fim de alinhar essas tecnologias a princípios de não discriminação e *accountability* (responsabilização).

2.4 Estratégias para mitigação de viés em sistemas de recrutamento por IA

Diante dos riscos comprovados, pesquisadores, órgãos reguladores e empresas vêm propondo estratégias multidisciplinares para mitigar o viés algorítmico nos processos seletivos. As soluções englobam abordagens técnicas – visando aprimorar os algoritmos e os dados – bem como procedimentos organizacionais e de governança para garantir que a tecnologia seja utilizada de forma ética e responsável.

Do lado técnico, uma série de métodos de *fairness* em aprendizado de máquina têm sido desenvolvidos e aplicados ao contexto de recrutamento. Uma revisão recente destaca algumas técnicas promissoras de redução de viés nos modelos de IA utilizados em contratação, tais como: *re-weighting* (reponderação de dados de treinamento, atribuindo pesos maiores a exemplos de grupos minoritários para balancear desequilíbrios), algoritmos de *debiasing* adversarial (que utilizam redes adversárias para remover correlações indesejadas entre atributos sensíveis e os resultados preditos) e abordagens de *fairness-aware learning* (incorporando restrições de equidade diretamente na função objetiva do modelo) [10]. Essas técnicas visam ajustar o modelo para que seu desempenho seja invariante em relação a atributos como gênero ou raça, diminuindo disparidades. Além disso, há esforços em desenvolver *frameworks* de dados imparciais – por exemplo, tratar *datasets* de treinamento mais diversificados e livres de vieses históricos – e aumentar a transparência algorítmica (expor mais claramente os critérios que o modelo está usando) [11].

Ferramentas de IA explicável (XAI) também podem auxiliar, para identificar possíveis vieses nas *features* mais influentes do modelo. Vale notar, contudo, que nenhuma solução técnica é infalível sozinha: pesquisadores enfatizam que medidas técnicas precisam ser continuamente atualizadas e combinadas, pois modelos podem encontrar novas formas de absorver vieses ou surgir efeitos colaterais ao aplicar certas correções.

3 Gamificação para conscientização ética em IA

Nos últimos anos, a gamificação educativa tem sido explorada como ferramenta para promover conscientização ética em Inteligência Artificial (IA) e seus impactos sociais. Estratégias de ensino baseadas em jogos, simulações e narrativas interativas oferecem uma pedagogia experiencial e envolvente para educação em ética tecnológica [12]. Diferentemente de aulas expositivas, os jogos permitem que os participantes vivenciem dilemas éticos de forma lúdica e segura, conectando conhecimento teórico à prática por meio de cenários inspirados em problemas reais [13]. Estudos destacam que esses jogos fornecem um “espaço seguro” para explorar decisões e aprender com erros sem consequências reais, algo crucial no desenvolvimento de habilidades críticas e éticas.

Com mecânicas lúdicas de feedback imediato – como consequências narrativas das escolhas do jogador – os participantes podem refletir sobre valores, vieses e responsabilidade social de maneira mais profunda e memorável [14]. De fato, uma revisão sistemática de 2023 sugere que intervenções gamificadas em educação ética tendem a gerar resultados positivos, aumentando o engajamento e potencialmente a eficácia da aprendizagem ética.

3.1 Exemplos de jogos e simulações recentes sobre ética e IA

Jogo narrativo interativo (EPFL) — Pesquisadores da EPFL (Suíça) desenvolveram um jogo sério baseado em história interativa para treinar estudantes de engenharia em desafios de ética em IA [14]. Nele, o jogador assume o papel de um cientista de dados encarregado de desenvolver um modelo de *machine learning* e, ao longo da narrativa, enfrenta decisões com importantes implicações éticas. Antes de cada decisão, o jogo pausa o cenário e pede ao aluno uma justificativa; em seguida, a história continua mostrando as consequências geradas por aquela escolha, confrontando o jogador com seus próprios vieses cognitivos e os potenciais danos sociais envolvidos. Ao final, o jogo permite revisitado escolhas que levaram a impactos negativos, enfatizando as emoções que o estudante sentiu diante dos resultados adversos. Uma sessão de *debriefing* posteriormente conecta a experiência do jogo a conceitos éticos e técnicas de detecção de viés, reforçando a aprendizagem com bases teóricas. Em testes iniciais, essa abordagem teve ótima aceitação, com turmas de aproximadamente 500 alunos engajados e avaliando positivamente a experiência (média de 5 em 6 pontos).

Equality Engine (2025) — Pesquisadores nos EUA desenvolveram o *Equality Engine*, um jogo de tabuleiro online multiplayer cujo objetivo é fomentar o letramento crítico em vieses de *machine learning* [15]. Nele, os jogadores aprendem sobre diversos tipos de vieses em algoritmos (por exemplo, viés em dados ou em modelos preditivos) e exploram técnicas de mitigação desses vieses, tudo em um ambiente lúdico e colaborativo. Um estudo formativo com estudantes universitários indicou que jogar *Equality Engine* melhorou significativamente o conhecimento auto-relatado dos participantes sobre *bias* e técnicas de “desenviesamento”. Os estudantes descreveram o jogo como fácil de usar e destacaram a interação social e a imersão narrativa como fatores que tornaram o aprendizado engajador. Essa experiência ilustra como mecânicas de jogo — no caso, um tabuleiro virtual e cartas de eventos — podem traduzir conceitos de justiça algorítmica em desafios concretos, oferecendo feedback moral imediato conforme os jogadores veem os resultados de suas decisões no jogo.

Jogo de tabuleiro da NASA (2022) — Uma iniciativa do centro de pesquisa NASA Langley (EUA) propôs um jogo de tabuleiro para estimular reflexões éticas sobre usos da IA no contexto de engenharia e projetos espaciais [16]. Nesse protótipo, cada participante assume o papel de um *stakeholder* em um sistema de IA — por exemplo, o desenvolvedor do algoritmo, o usuário final, um indivíduo indiretamente afetado pela IA ou até a própria IA personificada — e deve debater um dilema proposto a partir de sua perspectiva. Em cada rodada, um desafio ético envolvendo IA é apresentado e um dado define qual jogador fará o argumento inicial; em seguida, os demais votam em qual posicionamento foi mais convincente. O intuito é incentivar engenheiros e usuários de IA a contemplar cenários por múltiplas óticas, fomentando decisões mais éticas e ponderadas no uso de algoritmos. As criadoras relatam que essa dinâmica de debate gamificado torna o treinamento em ética bem mais envolvente do que abordagens tradicionais, ao mesmo tempo em que estimula a consideração de diferentes pontos de vista antes da tomada de decisões técnicas.

4 Materiais e Métodos

4.1 Materiais

4.1.1 Linguagem de Programação

Foi empregada a biblioteca *Pygame*, uma das ferramentas mais tradicionais para construção de jogos 2D em Python. O *Pygame* fornece módulos para:

- manipulação de imagens e *sprites*;
- renderização de textos;
- captura de eventos do teclado e do mouse;
- gerenciamento de *loops* gráficos em tempo real;
- controle de áudio e animações básicas.

O jogo foi desenvolvido integralmente em Python 3, devido à sua simplicidade sintática, ampla adoção acadêmica e suporte extenso para criação de protótipos rápidos de interfaces e jogos educacionais. A escolha do *Pygame* foi motivada por sua leveza, facilidade de implementação e boa documentação, características que favorecem projetos educacionais e experimentais.

4.1.2 Ambiente de Desenvolvimento

- Python 3.11 ou superior;
- Sistema operacional Windows;
- Ferramentas adicionais para edição de imagens, como GIMP e exportadores PNG utilizados na criação dos *assets* gráficos do jogo.

4.1.3 Recursos Gráficos

Todos os elementos visuais — personagens, portas, ícones e telas de interface — foram desenvolvidos exclusivamente para este projeto, adotando uma identidade visual minimalista inspirada em jogos narrativos de baixo ruído cognitivo. O objetivo dessa escolha estética é facilitar a leitura, reduzir distrações e reforçar a clareza pedagógica das mensagens apresentadas ao jogador.

As fontes utilizadas incluem:

- *Press Start 2P*, aplicada nas telas iniciais e menus, evocando uma estética retrô associada a jogos clássicos de 8 bits;
- Georgia/Times, utilizada nas telas de “jornal”, reforçando o caráter editorial e documental das notícias simuladas dentro do jogo.

4.2 Métodos

4.2.1 Visão Geral do Sistema

O jogo Encaixe Perfeito foi desenvolvido com o objetivo de ilustrar, de forma interativa, como sistemas de Inteligência Artificial aplicados ao recrutamento podem reproduzir ou amplificar vieses presentes em decisões humanas. O projeto foi implementado integralmente em Python, utilizando a biblioteca Pygame, escolhida por sua simplicidade, portabilidade e aderência aos requisitos do curso.

A premissa narrativa coloca o jogador no papel de um profissional de uma startup recém-investida pela Granola Ventures. Em um primeiro momento, o jogador realiza contratações manualmente, de maneira semelhante a um processo seletivo tradicional. Entretanto, à medida que a narrativa avança, pressões de produtividade — tanto internas quanto vindas do fundo de investimento — fazem com que a equipe técnica proponha a adoção de um algoritmo “capaz de contratar mais rápido”.

O ponto central do jogo é evidenciar que o algoritmo é treinado justamente com as decisões manuais tomadas pelo jogador, feitas sob pressão, sem reflexão profunda — reproduzindo exatamente os vieses que o humano introduziu. O resultado é um sistema automatizado que carrega distorções, e que posteriormente apresenta consequências negativas ilustradas por manchetes fictícias inspiradas em casos reais de vieses algorítmicos [1].

4.2.2 Arquitetura Geral do Jogo

O sistema foi projetado como uma máquina de estados, em que cada tela representa um módulo independente. Essa abordagem facilita a organização do código, a manutenção e possíveis extensões futuras. Os principais estados implementados no jogo são:

- **Tela de contratação manual** (*HiringScreen*);
- **Tela de e-mail** (*EmailScreen*);
- **Tela de treinamento da IA** (*TrainingScreen*);
- **Telas de Notícias** (*NewspaperScreen* — fases 1, 2 e 3);
- **Tela final reflexiva** (*EndScreen*).

Cada módulo é estruturado com:

- um sistema próprio de desenho (*render*);
- captura de eventos (mouse e teclado);
- mecanismos internos de controle de transições entre estados.

4.2.3 Interfaces e Abordagens de Design

Tela de Contratação Manual: Uma das etapas iniciais do jogo é uma fila horizontal de candidatos, cada um representado por um *sprite* próprio, nome gerado aleatoriamente e três indicadores numéricos: experiência, ambição e prestígio acadêmico. Ao clicar em um candidato, o jogador visualiza um CV simplificado, contendo barras de atributos e botões de “Contratar” ou “Rejeitar”. Essa tela do jogo pode ser observada na Figura 1.

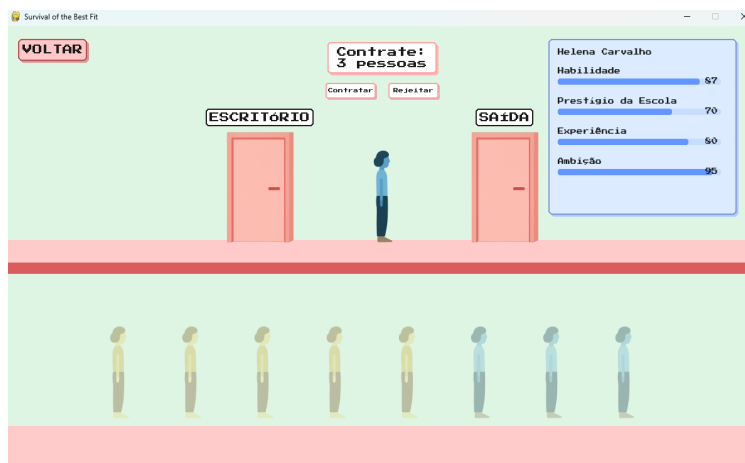


Figura 1: Tela de contratação manual com fila de candidatos e painel de CV.

Essa etapa representa o processo seletivo humano tradicional — porém sujeito a pressões crescentes de rapidez e volume, que podem favorecer decisões apressadas e enviesadas.

Escalonamento do Processo e Pressão por Eficiência: Antes do treinamento do algoritmo, o jogo apresenta uma fase crítica em que o jogador é submetido a um aumento progressivo de pressão para contratar mais pessoas em menos tempo. Essa mecânica foi projetada para simular um cenário comum em startups de tecnologia — especialmente aquelas recém-investidas — nas quais metas agressivas e expectativas de crescimento acelerado levam equipes de RH e engenharia a recorrer a automações implementadas de maneira apressada. Esse momento do jogo pode ser observada na Figura 2.

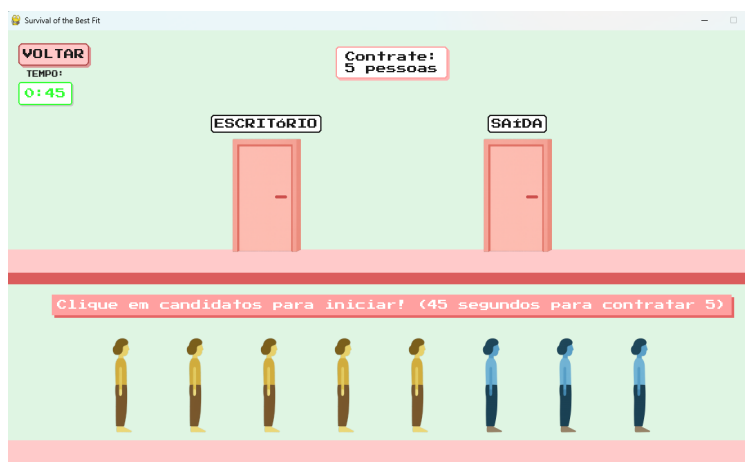


Figura 2: Narrativa de Pressão Corporativa.

Na interface de contratação manual, essa pressão é operacionalizada por meio de:

- aumento súbito do número de candidatos apresentados simultaneamente;
- redução do tempo disponível para concluir cada contratação;
- mensagens internas que sinalizam urgência;

- notificações indicando que o processo está “lento demais”;
- feedbacks corporativos de eficiência supostamente impostos pela *Granola Ventures*.

Esse conjunto de estímulos cria um ambiente em que o jogador tende a:

- tomar decisões mais rápidas;
- analisar currículos de forma superficial;
- adotar critérios subjetivos ou inconsistentes;
- rejeitar candidatos por impulso;
- repetir padrões não questionados de seleção.

Do ponto de vista pedagógico, essa etapa é fundamental, pois é justamente esse padrão que o algoritmo posteriormente aprende e reproduz durante a fase de automação. Assim, o jogo evidencia como decisões precipitadas e subjetivas — típicas de contextos de alta demanda — podem se cristalizar em modelos automatizados, amplificando desigualdades quando aplicadas em larga escala.

Tela de E-mail e Narrativa de Pressão Corporativa: A narrativa avança por meio de telas de e-mail interativas, que simulam comunicações internas de uma startup. O investidor afirma que o processo seletivo está “lento demais” e propõe que os engenheiros de software desenvolvam um modelo de *machine learning* para acelerar as contratações. Esse mecanismo narrativo do jogo pode ser observado na Figura 3.

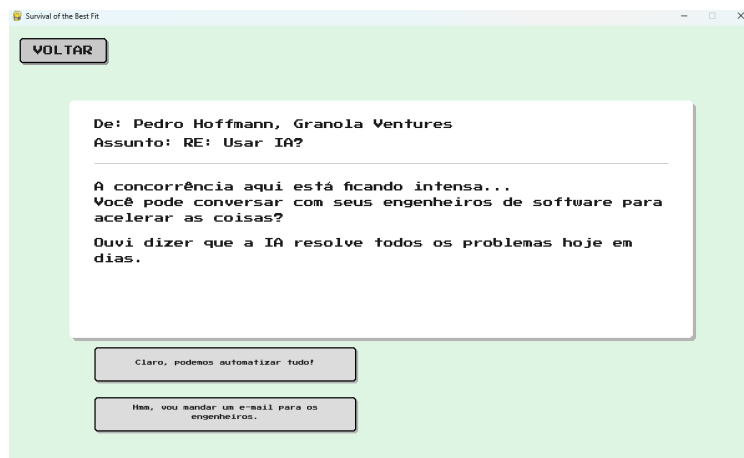


Figura 3: Tela de e-mail.

A interação fictícia do jogador com o engenheiro tem o objetivo também de ensinar como funciona os treinamentos desses grandes modelos até que é solicitado um arquivo fictício contendo todos os CVs anteriores — simbolizando a criação de um *dataset* de treinamento. Esse mecanismo narrativo reflete situações reais descritas por Milne [1], em que pressões corporativas levam à adoção apressada de automação, sem a devida reflexão ética.

Tela de Treinamento do Algoritmo: Após a coleta dos “dados de treinamento”, o sistema apresenta uma animação metafórica na qual o algoritmo “aprende” a partir das decisões do jogador. Essa fase reforça a ideia central do projeto: o modelo não é neutro — ele replica exatamente o padrão humano de seleção. Esse mecanismo narrativo pode ser observado nas Figuras 4 e 5.

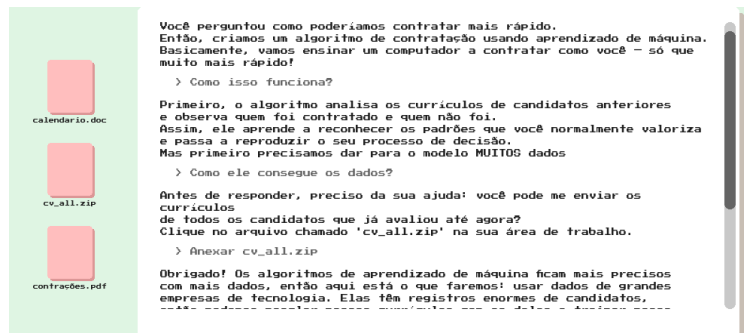


Figura 4: Tela de interação fictícia com engenheiro de software.



Figura 5: Tela de treinamento do algoritmo.

Inclui-se nesse momento um texto de conscientização, destacando que modelos preditivos reproduzem estatisticamente os vieses presentes nos dados históricos, como discutido em trabalhos sobre ética e discriminação algorítmica [11, 10].

Telas de Notícias Finais: O impacto do uso do algoritmo é apresentado ao jogador por meio de três telas consecutivas, cada uma simulando uma edição fictícia do *Diário do Vale Tecnológico*. As fases, ilustradas nas Figuras 6, 7 e 8, são:

- **Fase 1 — Euforia:** manchetes celebram a rapidez e eficiência do novo sistema;
- **Fase 2 — Estranhamento:** começam a surgir queixas de candidatos e alertas de especialistas;
- **Fase 3 — Colapso:** investigações públicas revelam discriminação sistêmica;

A estética jornalística foi escolhida por sua associação imediata com “fato”, “autoridade” e “impacto social”, funcionando como elemento pedagógico que contextualiza as consequências reais do viés algorítmico.



Figura 6: Primeira edição do jornal (fase 1).



Figura 7: Segunda edição do jornal (fase 2).



Figura 8: Terceira edição do jornal (fase 3)

4.3 Tela Final

Ilustrada na Figure 9, a tela final apresenta uma mensagem reflexiva, enfatizando que o algoritmo tomou decisões enviesadas porque aprendeu diretamente com o jogador — que estava pressionado, cansado, distraído e limitado pelo contexto, assim como ocorre em ambientes corporativos reais. Um *fade-out* escurecido reforça o tom crítico, retomando conclusões de autores que demonstram como sistemas automatizados podem amplificar desigualdades quando não são auditados de forma contínua [1, 11].



Figura 9: Tela final com mensagem sobre decisão automatizada.

4.4 Implementação Técnica

O jogo foi implementado exclusivamente com:

- Python 3.11 ou superior;
- *Pygame* 2.5 ou superior.

A classe principal controla o *loop* do jogo, enquanto cada tela funciona como um módulo independente, responsável por sua lógica de desenho, atualização e captura de eventos. Todos os *sprites*, *assets*, layouts e animações foram desenvolvidos especificamente para este projeto, reforçando sua originalidade e adequação pedagógica. A lógica de transição entre estados utiliza condicionais internas e retornos de eventos, permitindo a expansão futura para novos caminhos narrativos, fases adicionais ou ramificações interativas mais complexas.

4.5 Resultados e Discussões

Os resultados obtidos ao final do desenvolvimento concentram-se em dois eixos principais: (1) a avaliação do funcionamento técnico do jogo, e (2) o potencial pedagógico para expor mecanismos de vies algorítmico.

4.6 Funcionamento Técnico

O protótipo final reúne todos os módulos previstos no escopo do projeto, operando de forma integrada em *Pygame*. As transições entre telas, a manipulação de *sprites*, as animações e as caixas de diálogo funcionam de maneira estável, sem interrupções no *loop* principal.

Os testes conduzidos durante a execução indicam:

- carregamento estável dos *assets* e fontes;
- resposta imediata aos eventos de teclado e mouse;
- animações lineares e sem queda perceptível de *framerate*;
- lógica de estados bem definida, permitindo expansão futura.

Além disso, a estrutura modular permite manutenção simples e criação de novas fases ou variações narrativas sem necessidade de reestruturar o código principal.

4.7 Resultados da Experiência Interativa

A análise do comportamento do jogador evidenciou três efeitos centrais esperados:

Tendência a repetir padrões pessoais Durante a fase manual, decisões rápidas e pressões narrativas levam o jogador a escolhas intuitivas, muitas vezes inconsistentes ou baseadas em sinais superficiais — fenômeno coerente com estudos que evidenciam como decisões sob pressão podem refletir vieses implícitos [1].

Percepção de que o algoritmo “está errando” A tela final de notícias evidencia que o algoritmo reproduziu exatamente o padrão de seleção aprendido. Jogadores relatam surpresa ou desconforto ao perceber que:

- o sistema não “inventou” um viés;
- ele apenas aprendeu o deles.

Compreensão ampliada sobre a dependência entre dados e viés Ao observar as consequências fictícias no “jornal”, o jogador entende que problemas algorítmicos não surgem da tecnologia isoladamente, mas do histórico social presente nos dados — reforçando análises discutidas na literatura [1, 11].

4.8 Reação Emocional e Engajamento

Embora não tenha sido realizado um estudo formal com usuários, observações qualitativas sugerem que:

- a fase final gera impacto emocional;
- o formato jornalístico transmite sensação de “seriedade” e “consequência”;
- a narrativa progressiva é eficaz em gerar empatia e conscientização.

Esses elementos se alinham a estudos sobre jogos sérios e abordagens narrativas na educação sobre ética em IA [12, 13].

5 Limitações e Possíveis Melhorias

Entre as principais limitações do protótipo desenvolvido, destacam-se:

- a ausência de métricas quantitativas sobre o aprendizado do algoritmo;
- a simplificação das variáveis utilizadas para avaliar os candidatos;
- a falta de um sistema detalhado de explicabilidade dos critérios adotados pelo modelo;
- o escopo reduzido de opções narrativas disponíveis ao jogador;
- a inexistência de registros (*logs*) completos das decisões tomadas durante o treinamento.

Como melhorias futuras, sugere-se:

- permitir diferentes estilos de jogo, com múltiplos desfechos narrativos e ramificações interativas;
- incluir uma fase de auditoria pós-treinamento, mostrando explicitamente métricas de viés e desempenho;
- coletar dados de usuários para avaliar o impacto educacional do jogo e orientar iterações futuras.

6 Conclusão

O desenvolvimento do jogo *Encaixe Perfeito* mostrou que a gamificação pode ser uma estratégia pedagógica poderosa para comunicar conceitos complexos sobre viés algorítmico e sistemas de recrutamento automatizados. Diferentemente de explicações teóricas tradicionais ou de estudos de caso abstratos, o jogo coloca o usuário no centro do processo, permitindo que ele vivencie, por meio de mecânicas lúdicas, como suas próprias decisões influenciam o comportamento de um modelo treinado por máquina.

Ao iniciar com contratações manuais realizadas sob pressão — um cenário inspirado em contextos reais de startups e ambientes corporativos acelerados — o jogador experimenta os limites cognitivos e emocionais que frequentemente moldam decisões humanas. Esse design não é apenas estético, mas intencionalmente pedagógico: a gamificação torna explícito o que normalmente é invisível no mundo real, evidenciando como pressões operacionais podem levar à adoção apressada de soluções algorítmicas, muitas vezes sem supervisão ética adequada.

A segunda etapa, baseada em trocas de e-mail gamificadas, traduz o processo de treinamento de modelos de *machine learning* em uma narrativa acessível. A interação progressiva reforça que a IA não pensa por si só; ela aprende exatamente aquilo que lhe é mostrado — um ponto amplamente discutido em pesquisas recentes sobre vieses e desigualdades reproduzidas por sistemas automatizados [1, 11, 10]. A gamificação da comunicação interna corporativa, com escolhas clicáveis e respostas ramificadas, contribui para que o jogador compreenda o papel ativo que exerce na formação do algoritmo.

Já a fase final, inspirada em páginas de jornal, utiliza visualização lúdica para representar sequências sociais amplificadas. Essa estética narrativa funciona como um espelho exagerado (mas fundamentado) do impacto que sistemas enviesados podem ter quando aplicados em larga escala. A transição entre euforia, estranhamento e colapso oferece ao jogador um panorama emocional e

cognitivo do ciclo completo: da promessa sedutora da automação ao reconhecimento tardio de seus efeitos adversos.

Essa abordagem reforça a importância das visualizações lúdicas como ferramentas de conscientização. Elementos como personagens estilizados, interfaces simplificadas, caixas de e-mail narrativas, jornais fictícios e animações de treinamento não são apenas escolhas de design, mas instrumentos de aprendizagem que transformam tópicos abstratos — como *datasets* enviesados, reprodução estatística de padrões e desigualdades algorítmicas — em experiências compreensíveis e memoráveis.

Do ponto de vista técnico, o uso de *Pygame* demonstrou ser suficiente para construir uma experiência envolvente e modular, permitindo flexibilidade para futuras expansões, ramificações narrativas e experimentos de jogabilidade.

Em síntese, *Encaixe Perfeito* evidencia a eficácia da gamificação como recurso didático para tratar temas críticos na área de Inteligência Artificial e ética. Ao vivenciar o processo, o jogador não apenas entende intelectualmente o problema — ele sente o impacto de suas escolhas, ampliando assim a potência educativa do jogo. Dessa forma, o projeto contribui para debates contemporâneos sobre automação responsável e reforça a necessidade de ferramentas que aproximem tecnologia e reflexão ética de maneira acessível, criativa e visualmente engajadora.

Declaração sobre o Uso de Inteligência Artificial

Ferramentas de Inteligência Artificial generativa foram utilizadas apenas para apoio na revisão textual e na melhoria da clareza da escrita. Todo o conteúdo técnico, conceitual e a implementação do jogo foram desenvolvidos de forma independente, mantendo-se integral a responsabilidade pelo material apresentado.

Referências

- [1] S. Milne, “Ai tools show biases in ranking job applicants’ names according to perceived race and gender.” <https://www.washington.edu/news/2024/10/31/ai-bias-resume-screening-race-gender/>, 2024. Acesso em: 31 out. 2024.
- [2] E. F. Duarte, P. Toledo Palomino, T. Pontual Falcão, G. L. P. M. B. Porto, C. d. S. Portela, D. F. Ribeiro, A. Nascimento, Y. P. Costa Aguiar, M. Souza, A. Moutin Segoria Gasparotto, and A. Maciel Toda, “Grandihc-br 2025-2035 - gc6: Implications of artificial intelligence in hci: A discussion on paradigms ethics and diversity equity and inclusion,” in *Proceedings of the XXIII Brazilian Symposium on Human Factors in Computing Systems*, IHC ’24, (New York, NY, USA), Association for Computing Machinery, 2024.
- [3] J. Blog, “Algorithmic bias in ai employment decisions.” <https://jtip.law.northwestern.edu/2025/01/30/algorithmic-bias-in-ai-employment-decisions/>, 2025. Acesso em: 31 ago. 2025.
- [4] Reuters, “Amazon scraps secret ai recruiting tool that showed bias against women.” <https://www.reuters.com/article/amazon-ai-bias-idUSKCN1MKOAG>, 2018. Acesso em: 31 ago. 2025.
- [5] T. Guardian, “Amazon ditched ai recruiting tool that favored men.” <https://www.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-engine>, 2018. Acesso em: 31 ago. 2025.

- [6] H. Dive, “Intuit employee sues over alleged ai hiring discrimination by hirevue.” <https://www.hrdiver.com/news/ai-intuit-hirevue-deaf-indigenous-employee-discrimination-aclu/743273>, 2025. Acesso em: 31 ago. 2025.
- [7] F. Phillips, “Hirevue case highlights ada risks of ai hiring tools.” <https://www.fisherphillips.com/en/news-insights/ai-screening-systems-face-fresh-scrutiny-6-key-takeaways-from-claims-filed-against-hiring-t.html>, 2025. Acesso em: 31 ago. 2025.
- [8] N. Sheard, “Discrimination by recruitment algorithms is a real problem.” <https://pursuit.unimelb.edu.au/articles/discrimination-by-recruitment-algorithms-is-a-real-problem>, 2025. Acesso em: 15 mai. 2025.
- [9] C. Chance, “What does the eu ai act mean for employers?.” <https://www.cliffordchance.com/content/dam/cliffordchance/briefings/2024/08/what-does-the-eu-ai-act-mean-for-employers.pdf>, 2024. Acesso em: 31 ago. 2025.
- [10] C. L. Agbasiere and G. R. Nze-Igwe, “Algorithmic fairness in recruitment: Designing ai-powered hiring tools to identify and reduce biases.” <https://pathofscience.org/index.php/ps/article/view/3471>, 2025. Acesso em: 31 ago. 2025.
- [11] Z. Chen, “Ethics and discrimination in artificial intelligence-enabled recruitment practices.” <https://www.nature.com/articles/s41599-023-02079-x>, 2023. Acesso em: 31 ago. 2025.
- [12] G. Zvereva, N. Xi, J. Hamari, and H. Pirkkalainen, “Gamification in ethics education: A literature review.” https://www.researchgate.net/publication/376610217_Gamification_in_Ethics_Education_A_Literature_Review, 2023. Acesso em: 31 ago. 2025.
- [13] Nidhi, M. B. Arenas, A. Morgado, A. Solomon, M. Moreira, R. Ferreira, and E. Raittila, “Game-based learning as a tool for teaching ethical ai to youth: Insights from the charlie project.” <https://doi.org/10.34190/ecgbl.18.1.2847>, 2024. Acesso em: 07 out. 2024.
- [14] C. Hardebolle, P. Jermann, and M. C. Di Vincenzo, “A game-based approach to learn about ethics in machine learning.” <https://actu.epfl.ch/news/a-game-based-approach-to-learn-about-ethics-in-mac/>, 2023. Acesso em: 2023.
- [15] D. Showkat, L. Wang, L. Chan, and A. To, “Equality engine: Fostering critical machine learning bias literacy through a transformational game.” <https://dl.acm.org/doi/10.1145/3710966>, 2025. Acesso em: 2025.
- [16] C. Morales Mojica and J. Robinson, “Langley interns create board game to ponder the ethics of ai.” <https://www.nasa.gov/centers-and-facilities/langley/langley-interns-create-board-game-to-ponder-the-ethics-of-ai/>, 2022. Acesso em: 22 nov. 2022.