v7.n9.2026

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

ARQUITETURA HÍBRIDA DE INTELIGÊNCIA ARTIFICIAL PARA APOIO À SELEÇÃO DE CANDIDATOS COM CRITÉRIOS SOCIOECONÔMICOS E TEXTUAIS

HYBRID ARTIFICIAL INTELLIGENCE ARCHITECTURE TO SUPPORT CANDIDATE SELECTION BASED ON SOCIOECONOMIC AND TEXTUAL CRITERIA

ARQUITECTURA HÍBRIDA DE INTELIGENCIA ARTIFICIAL PARA APOYAR LA SELECCIÓN DE CANDIDATOS BASADA EN CRITERIOS SOCIOECONÓMICOS Y TEXTUALES

Igor Jose Reis dos Reis [1], Paulo Heitor Luz da Cruz [1], Ana Celia Santos da Silva [1], Keventon Guimarães [2], Miguel Ângelo Rodrigues Mocbel [3], Fabricio De Souza Farias [4], Armando José de Sá Santos [5]

e798977
https://doi.org/10.47820/recima21.v7i9.8977

PUBLICADO: 09/2026

RESUMO

A análise de candidatos em processos seletivos de cursos de qualificação profissional envolve informações socioeconômicas e respostas discursivas, tornando necessário estabelecer procedimentos capazes de organizar esses dados de forma padronizada. Neste trabalho, propõe-se e avalia-se uma arquitetura híbrida de Inteligência Artificial (IA) para apoio à análise e classificação de candidatos, integrando três módulos: um Índice de Vulnerabilidade Social (IVS), construído como índice composto sintético, técnicas de Processamento de Linguagem Natural (PLN) baseadas no modelo *Transformer* para análise semântica de respostas discursivas e um algoritmo genético (AG) para suporte à otimização da seleção em cenários de restrições combinatórias. O estudo caracteriza-se como um desenvolvimento tecnológico aplicado, de abordagem quanti-qualitativa, realizado a partir de dados de processos seletivos de cursos de qualificação profissional de uma instituição pública (anos de 2024 e 2025), com simulação de 30 vagas em cada processo. Os resultados evidenciaram diferenças nos indicadores socioeconômicos entre os conjuntos analisados e elevada concentração das pontuações semânticas produzidas pelo módulo de PLN, indicando que a força seletiva do modelo se concentrou predominantemente nas variáveis socioeconômicas. Observou-se também concordância entre a solução evolutiva e a ordenação determinística exata no cenário sem sub-restrições. Os resultados são interpretados como evidências do comportamento da arquitetura nos cenários experimentais analisados, e não como demonstração de superioridade, robustez ou equidade do método. Conclui-se que a integração entre indicadores socioeconômicos, análise textual e otimização constitui uma abordagem com potencial para apoiar comissões avaliadoras, desde que utilizada sob supervisão humana e submetida a avaliações de sensibilidade, auditabilidade e mitigação de vieses.

PALAVRAS-CHAVE: Inteligência artificial. Processos seletivos. Políticas afirmativas. Processamento de linguagem natural. Algoritmo genético.

[1] Bacharelando(a) em Engenharia de Software pela UEPA.
[2] Bacharel em Sistemas de Informação pela UFPA. Professor Substituto na Faculdade de Sistemas de Informação FASI/UFPA.
[3] Bacharel em Sistemas de Informação pela UFPA.
[4] Doutor em Engenharia Elétrica pelo PPGEE/UFPA. Professor Adjunto na Faculdade de Sistemas de Informação FASI/UFPA.
[5] Mestre em Ciência e Engenharia de Materiais pela Universidade do Estado do Pará.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

ABSTRACT

The evaluation of candidates in professional training course selection processes involves socioeconomic data and open-ended responses, requiring standardized and traceable procedures. This study proposes and evaluates a hybrid Artificial Intelligence (AI) architecture to support candidate analysis and ranking, integrating three modules: a synthetic Social Vulnerability Index (IVS), Natural Language Processing (NLP) techniques based on the Transformer model for semantic analysis, and a Genetic Algorithm (GA) to assist optimization under combinatorial constraints. Designed as an applied technological development with a mixed-methods approach, the study utilizes data from professional training course selection processes at a public institution (2024 and 2025), simulating 30 available seats per process. Results demonstrated differences in socioeconomic indicators between datasets and a high concentration of semantic scores generated by the NLP module, indicating that the selective weight relied predominantly on socioeconomic variables. Furthermore, the evolutionary solution converged with the exact deterministic ranking in scenarios without sub-constraints. Findings reflect the architecture's behavior in experimental settings rather than definitive proof of superiority or fairness. The study concludes that integrating socioeconomic indicators, textual analysis, and optimization has potential to assist evaluation committees, provided it operates under continuous human supervision and undergoes further validation regarding sensitivity, auditability, and bias mitigation.

KEYWORDS: *Artificial intelligence. Selection processes. Affirmative action. Natural language processing. Genetic Algorithm.*

RESUMEN

El análisis de candidatos en procesos de selección de cursos de cualificación profesional involucra información socioeconómica y respuestas discursivas, siendo necesario establecer procedimientos estandarizados y trazables. Este trabajo propone y evalúa una arquitectura híbrida de Inteligencia Artificial (IA) para apoyar el análisis y clasificación de candidatos, integrando tres módulos: un Índice de Vulnerabilidad Social (IVS) sintético, técnicas de Procesamiento del Lenguaje Natural (PLN) basadas en el modelo Transformer para análisis semántico, y un algoritmo genético (AG) para respaldar la optimización bajo restricciones combinatorias. Se trata de un estudio de desarrollo tecnológico aplicado, con enfoque cuantitativo-cualitativo, utilizando datos de procesos de selección de cursos de cualificación profesional en una institución pública (2024 y 2025), simulando 30 plazas por proceso. Los resultados evidenciaron diferencias en los indicadores socioeconómicos entre los conjuntos y una elevada concentración en las puntuaciones semánticas del módulo PLN, mostrando que la fuerza selectiva recayó principalmente en las variables socioeconómicas. Asimismo, la solución evolutiva coincidió con la ordenación determinista exacta en escenarios sin sub-restricciones. Los resultados reflejan el comportamiento de la arquitectura en escenarios experimentales y no una demostración definitiva de superioridad o equidad. Se concluye que la integración de indicadores socioeconómicos, análisis textual y optimización posee potencial para apoyar a las comisiones evaluadoras, siempre que se utilice bajo supervisión humana continua y se someta a evaluaciones de sensibilidad, auditabilidad y mitigación de sesgos.

PALABRAS CLAVE: *Inteligencia artificial. Procesos de selección. Acción afirmativa. Procesamiento de lenguaje natural. Algoritmo genético.*

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

INTRODUÇÃO

A democratização do acesso à educação constitui um importante instrumento para a redução das desigualdades sociais e para a promoção da justiça educacional no Brasil (BASSO-POLETTO; EFROM; RODRIGUES, 2020). Nesse contexto, as políticas de ações afirmativas ampliaram as oportunidades de acesso de grupos historicamente excluídos, consolidando-se como mecanismos voltados à promoção da equidade e da inclusão social (VIEIRA; VIDEIRA; CUSTÓDIO, 2025). Entretanto, a implementação dessas políticas também exige procedimentos capazes de analisar os critérios estabelecidos nos processos seletivos de maneira organizada, consistente e passível de verificação.

A análise das candidaturas pode envolver informações relacionadas à renda, origem escolar, pertencimento a grupos contemplados pelas políticas afirmativas e outras características socioeconômicas. Esses elementos assumem relevância porque as oportunidades educacionais não são determinadas exclusivamente pelo desempenho individual, sendo também influenciadas pelas condições sociais, econômicas e familiares dos candidatos (SILVA, 2016; TAVARES SANCHES *et al.,* 2019). Dessa perspectiva, a utilização de indicadores capazes de organizar diferentes dimensões socioeconômicas pode auxiliar na sistematização das informações utilizadas pelas comissões responsáveis pela avaliação.

Na prática, parte desse processo depende da análise de documentos, declarações e informações fornecidas pelos candidatos. A operacionalização das políticas afirmativas pode, portanto, demandar esforço administrativo para verificar critérios de elegibilidade e analisar diferentes perfis de candidatos (BASSO-POLETTO; EFROM; RODRIGUES, 2020). Além disso, atividades avaliativas que envolvem informações qualitativas podem estar sujeitas a variações decorrentes da interpretação dos responsáveis pela análise, aspecto relacionado à discricionariedade presente na implementação de políticas públicas (LIPSKY, 2019).

Nesse cenário, a Inteligência Artificial (IA) vem sendo incorporada a diferentes atividades de gestão educacional, especialmente para automação de tarefas, análise de dados e processamento de grandes volumes de informações (CUNHA *et al.,* 2026; COSTA JÚNIOR *et al.,* 2023). Entretanto, sua aplicação em processos que podem influenciar o acesso a oportunidades educacionais requer cuidados adicionais. Aspectos como transparência, explicabilidade, supervisão humana e mitigação de vieses devem ser considerados para evitar que sistemas computacionais reproduzam ou ampliem desigualdades presentes nos dados utilizados (HEGGLER; SZMOSKI; MIQUELIN, 2025). Assim, a utilização de IA nesse contexto deve ser compreendida como suporte à atividade avaliativa, e não como substituição da responsabilidade institucional e humana pela decisão final.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Apesar da expansão das aplicações de IA na educação, permanecem lacunas relacionadas à utilização dessas tecnologias em processos específicos de apoio à análise de candidatos, especialmente quando são combinadas informações socioeconômicas e textuais. Estudos apontam a necessidade de maior atenção à transparência, à participação humana e ao uso responsável de sistemas inteligentes em contextos educacionais (HOLMES *et al*., 2019; ZAWACKI-RICHTER *et al*., 2019). Essa lacuna motiva a investigação de arquiteturas capazes de integrar diferentes fontes de informação sem eliminar a participação das comissões avaliadoras no processo decisório.

Diante desse contexto, este trabalho propõe uma arquitetura híbrida composta por quatro módulos integrados: (I) um índice composto denominado Índice de Vulnerabilidade Social (IVS), desenvolvido especificamente para organizar indicadores socioeconômicos dos candidatos; (II) um módulo de Processamento de Linguagem Natural (PLN), baseado em modelos *Transformers*, para análise semântica das respostas discursivas; (III) função de aptidão (*fitness*), que realiza a combinação dos resultados de IVS e PLN; e (IV) um algoritmo genético (AG), empregado como mecanismo de otimização da seleção, explorando combinações de candidatos de acordo com a função de aptidão definida no estudo (VASWANI *et al.*, 2017; DEVLIN *et al*., 2018; HOLLAND, 1992; MITCHELL, 1998).

A questão que orienta esta pesquisa é: como uma arquitetura híbrida baseada em um índice socioeconômico, análise semântica por PLN e otimização por algoritmo genético pode apoiar a análise e seleção de candidatos em processos seletivos de cursos de qualificação profissional, mantendo a supervisão humana sobre a decisão final?

Com base nessa questão, o objetivo deste trabalho é desenvolver e avaliar uma arquitetura híbrida de apoio à seleção de candidatos, integrando indicadores socioeconômicos, análise semântica de respostas discursivas e otimização por algoritmo genético. Especificamente, busca-se analisar o comportamento dos módulos integrados, verificar a contribuição dos indicadores utilizados na classificação e discutir as limitações da abordagem diante de aspectos como sensibilidade dos pesos, capacidade discriminatória do módulo textual, necessidade de otimização evolutiva e possíveis vieses decorrentes das variáveis utilizadas.

A contribuição do estudo consiste na proposição e avaliação experimental de uma arquitetura capaz de integrar informações quantitativas e qualitativas em um mesmo fluxo de análise. A solução é concebida como ferramenta de apoio às comissões avaliadoras, preservando a supervisão humana e permitindo que os resultados produzidos pelo sistema sejam posteriormente analisados e contestados pelos responsáveis pelo processo seletivo.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

1. REFERENCIAL TEÓRICO

1.1. Políticas afirmativas e avaliação da vulnerabilidade socioeconômica

As políticas de ações afirmativas no Brasil resultam de lutas históricas pela democratização do acesso ao ensino público (SOUSA; NASCIMENTO, 2026). Marcos legais como a Lei nº 12.711/2012, a Lei nº 13.409/2016 e a atualização promovida pela Lei nº 14.723/2023 estabeleceram critérios objetivos de reserva de vagas baseados em renda familiar, origem em escola pública, pertencimento étnico-racial e deficiência (BRASIL, 2012, 2016, 2023). A efetividade dessas políticas depende de mecanismos institucionais transparentes e padronizados para a verificação de elegibilidade (VIEIRA; VIDEIRA; CUSTÓDIO, 2025).

A perspectiva do mérito relacional aponta que o desempenho acadêmico deve ser analisado conjuntamente com o contexto social de origem do estudante (SILVA, 2016; WAINER; MELGUIZO, 2018). Indicadores compostos sintéticos cumprem o papel de aglutinar múltiplas dimensões da vulnerabilidade em um escore único e auditável. Contudo, a definição dos pesos e critérios dessas variáveis exige rigoroso alinhamento com os marcos legais para evitar valorações arbitrárias de atributos protegidos.

Nesse sentido, o fortalecimento das políticas de ações afirmativas depende não apenas da existência de critérios legais, mas também da adoção de mecanismos institucionais capazes de garantir maior padronização, transparência e consistência na análise das candidaturas. Esses desafios tornam relevante a investigação de estratégias que possam apoiar as comissões avaliadoras na operacionalização dos critérios estabelecidos pelas políticas de inclusão.

1.2. Processamento de linguagem natural e algoritmos de otimização

O avanço das arquiteturas *Transformer* permitiu a conversão de textos em representações vetoriais densas (*embeddings*), capazes de capturar nuances semânticas e contextuais (DEVLIN *et al.,* 2018; VASWANI *et al.,* 2017). No contexto educacional, o PLN possibilita incorporar informações qualitativas (como motivações e expectativas) aos processos de avaliação (HOLMES *et al.,* 2019). No entanto, o uso de pontuações textuais em processos seletivos exige a verificação de sua capacidade discriminatória, evitando que a baixa variância das notas transforme o componente textual em um elemento irrelevante ou ruidoso.

Para a seleção de candidatos sob múltiplos critérios, os algoritmos genéticos (AG) oferecem mecanismos de busca inspirados na evolução biológica para maximizar funções de aptidão (*fitness*) (HOLLAND, 1992; MITCHELL, 1998). Enquanto problemas simples de seleção sem sub-restrições podem ser resolvidos de forma exata por ordenação determinística (algoritmos *greedy*), a utilização de AGs justifica-se em cenários de otimização combinatória

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

restrita, nos quais coexistem sub-limites de vagas e concorrências simultâneas por modalidades de cotas (RUSSELL; NORVIG, 2022).

Estudos de Wainer e Melguizo (2018), baseados em dados do Exame Nacional de Desempenho dos Estudantes (ENADE), evidenciam que as diferenças de desempenho entre cotistas e não cotistas estão mais relacionadas às condições socioeconômicas dos estudantes do que à modalidade de ingresso. De forma complementar, Silva (2016) destaca que fatores como vulnerabilidade econômica, contexto familiar, dificuldades de permanência e desigualdades estruturais continuam exercendo influência significativa sobre a trajetória acadêmica dos estudantes, reforçando a importância da avaliação socioeconômica nas políticas de ações afirmativas.

Essas evidências demonstram que a avaliação socioeconômica desempenha papel central na efetividade das políticas de inclusão, tornando necessária a utilização de indicadores capazes de sintetizar diferentes dimensões da vulnerabilidade social. Por essas razões, a construção de indicadores compostos pode contribuir para organizar informações socioeconômicas de maneira padronizada, permitindo sua utilização como suporte à análise e à comparação entre diferentes perfis de candidatos.

1.3. Trabalhos relacionados

A utilização de técnicas de Inteligência Artificial em contextos educacionais tem sido investigada sob diferentes perspectivas, abrangendo desde o apoio ao processo de ensino-aprendizagem até a personalização, o monitoramento e a gestão de atividades educacionais. Costa Júnior *et al.* (2023) discutem a utilização da IA como ferramenta de apoio no ensino superior, destacando seu potencial para auxiliar diferentes atividades educacionais e contribuir para a melhoria dos processos institucionais. De forma semelhante, Cunha *et al.* (2026), ao analisarem estudos sobre a utilização da IA no ensino superior, identificam aplicações relacionadas ao ensino, à aprendizagem e à avaliação, evidenciando a expansão dessas tecnologias no contexto educacional.

Algumas pesquisas apresentam aplicações práticas da IA voltadas ao acompanhamento e à melhoria do desempenho dos estudantes. Matos *et al.* (2026) avaliaram uma plataforma educacional baseada em IA e algoritmos adaptativos, comparando estudantes submetidos a trilhas personalizadas de aprendizagem com um grupo submetido à metodologia tradicional. Os resultados apontaram avanços nos níveis de proficiência e um ganho médio relativo de aproximadamente 49% no desempenho dos estudantes acompanhados pela plataforma. Os autores ressaltam, entretanto, que os resultados devem estar associados à mediação pedagógica ativa e ao acompanhamento contínuo dos estudantes.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Em uma perspectiva de estruturação de soluções baseadas em Inteligência Artificial, Santos (2025) desenvolveu um modelo conceitual para Ambientes Virtuais de Ensino e Aprendizagem, integrando técnicas de IA nas dimensões de personalização, engajamento, monitoramento e análise de desempenho. A pesquisa combinou revisão sistemática da literatura, *Design Science Research* e validação qualitativa com especialistas, resultando também no desenvolvimento da plataforma web InteliEduca. O trabalho demonstra a possibilidade de organizar diferentes funcionalidades baseadas em IA em uma arquitetura integrada, orientada ao suporte de processos educacionais.

Além das aplicações diretamente relacionadas ao ensino e à aprendizagem, técnicas de IA também têm sido utilizadas para apoiar processos de decisão e otimização no contexto educacional. Nogueira (2025) desenvolveu um sistema para auxiliar professores na formação de equipes de estudantes para projetos de *software*, utilizando um algoritmo genético para considerar habilidades dos alunos, preferências individuais e requisitos dos projetos. O autor comparou o desempenho do AG com uma abordagem de Força Bruta e observou tempos de execução significativamente menores para o algoritmo evolutivo, principalmente em cenários com maior quantidade de estudantes. Esse resultado evidencia a aplicabilidade de técnicas de otimização na resolução de problemas educacionais que envolvem múltiplos critérios e diferentes possibilidades de combinação.

Sob uma perspectiva crítica, Machado (2026) analisou o uso da IA como ferramenta de apoio ao ensino-aprendizagem na Educação Profissional e Tecnológica. O estudo destaca a IA como um suporte cognitivo capaz de auxiliar o estudante, mas ressalta que sua utilização deve ocorrer mediante a participação do docente. A autora também chama atenção para riscos relacionados à terceirização cognitiva e à reprodução de vieses algorítmicos, reforçando a necessidade de preservar a atuação humana na utilização de sistemas inteligentes. Nessa mesma perspectiva, Moreira *et al.* (2025) discutem aspectos éticos relacionados ao uso da IA na educação, destacando a necessidade de considerar questões como transparência, privacidade e possíveis vieses na adoção dessas tecnologias.

Os estudos apresentados demonstram que a IA já vem sendo aplicada em diferentes dimensões da educação, incluindo personalização da aprendizagem, acompanhamento do desempenho, apoio à prática docente e otimização de processos de decisão. Entretanto, observa-se que essas aplicações se concentram predominantemente em atividades pedagógicas, de acompanhamento acadêmico ou de gestão educacional. Paralelamente, pesquisas relacionadas às condições socioeconômicas no contexto educacional evidenciam a influência desses fatores sobre as trajetórias e os resultados dos estudantes (FERNANDES; PASSADOR, 2023; TAVARES SANCHES *et al.*, 2019), enquanto estudos sobre políticas de ações afirmativas têm discutido questões relacionadas ao acesso, à permanência e à equidade

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

(VIEIRA; VIDEIRA; CUSTÓDIO, 2025). Contudo, essas perspectivas ainda não são exploradas, de forma integrada, em arquiteturas inteligentes destinadas ao apoio à análise e classificação de candidatos.

Nesse sentido, identifica-se uma lacuna na integração entre as duas perspectivas. Embora existam estudos que utilizem IA e AG em diferentes problemas educacionais, ainda são limitadas as propostas que combinam indicadores de vulnerabilidade social, análise semântica de informações discursivas e técnicas de otimização para apoiar processos seletivos relacionados às políticas de ações afirmativas. É nesse espaço que se insere o presente trabalho, ao propor uma arquitetura híbrida destinada a apoiar as comissões avaliadoras na análise e classificação dos candidatos, mantendo a supervisão humana e buscando ampliar a padronização, a rastreabilidade e a eficiência do processo.

2. METODOLOGIA

Esta pesquisa caracteriza-se como um estudo de desenvolvimento tecnológico aplicado, de abordagem quanti-qualitativa, voltado ao desenvolvimento de um sistema inteligente de apoio à análise socioeconômica de candidatos às políticas de ações afirmativas. O ambiente experimental foi estruturado para simular um processo seletivo institucional, utilizando dados com identificadores diretos removidos, em conformidade com os princípios estabelecidos pela Lei Geral de Proteção de Dados (Lei nº 13.709/2018; BRASIL, 2018).

Foram utilizadas duas bases provenientes dos formulários socioeconômicos do Polo Cametá – Forma Cametá, referentes aos anos de 2024 e 2025, contendo, respectivamente, 227 e 160 registros. Para a avaliação experimental, foram consideradas 30 vagas em cada conjunto. Antes do processamento, foram removidos os identificadores diretos, como nome, telefone, WhatsApp, endereço e e-mail. Os demais atributos (como deficiência, raça/cor e renda) foram mantidos para a realização das análises previstas no modelo. As demais informações foram organizadas conforme sua utilização nos módulos da arquitetura.

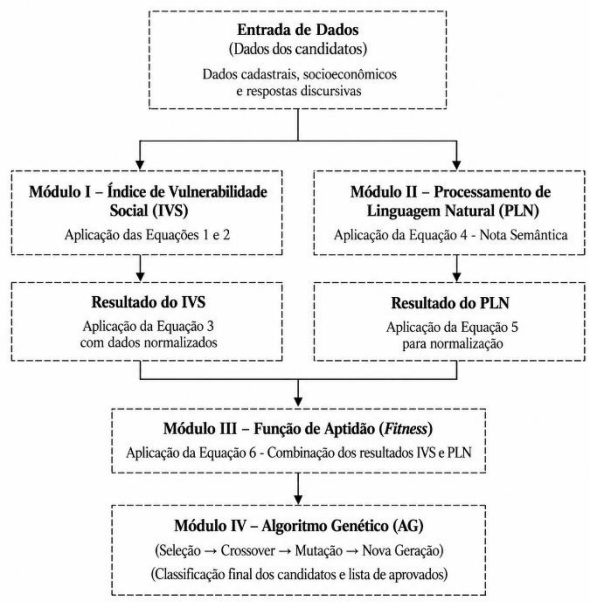
Para validar a proposta, foi desenvolvido um sistema modular baseado em Programação Orientada a Objetos (POO), utilizando Python 3.11.9 e bibliotecas específicas para PLN e AG. O sistema é composto por quatro módulos: (I) Índice de Vulnerabilidade Social (IVS); (II) Processamento de Linguagem Natural (PLN); (III) função de aptidão (*fitness*); e (IV) algoritmo genético (AG), responsável pela otimização da seleção.

Conforme a Figura 1, o sistema inicia seu funcionamento com a entrada dos dados fornecidos pelos candidatos durante o processo de inscrição, incluindo informações cadastrais, socioeconômicas e respostas discursivas. Os dados cadastrais e socioeconômicos são direcionados ao Módulo I, responsável pelo cálculo do IVS. Paralelamente, as respostas

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

discursivas são encaminhadas ao Módulo II, responsável pelo processamento textual e pelo mapeamento semântico, baseado em modelos *Transformers.*

Figura 1. Arquitetura geral e fluxo de funcionamento do sistema



Fonte: Elaborado pelos autores (2026).

Os resultados produzidos pelos Módulos I e II são posteriormente integrados pelo Módulo III, correspondente à função de aptidão (*fitness*), que realiza a combinação ponderada das notas de IVS e PLN. O resultado dessa função é então utilizado pelo Módulo IV, correspondente ao algoritmo genético, que realiza o processo de otimização da seleção, considerando as etapas de seleção, crossover, mutação e formação de novas gerações, resultando na classificação final dos candidatos e na respectiva lista de candidatos selecionados pelo modelo.

2.1. Módulo I – Índice de Vulnerabilidade Social (IVS)

O Módulo I é responsável pelo cálculo do IVS, que analisa as informações socioeconômicas dos candidatos considerando os critérios de elegibilidade e os grupos contemplados pelas políticas de ações afirmativas, conforme estabelecido pela Lei nº 14.723/2023 (BRASIL, 2023) e pelos regulamentos institucionais. Entre os atributos considerados encontram-se a renda familiar per capita, a origem escolar, o pertencimento aos grupos contemplados pelas ações afirmativas e outros indicadores relacionados à vulnerabilidade social. Cada atributo recebe uma Pontuação (vk) previamente definida pelo Quadro 1, sendo posteriormente normalizada para produzir um índice compreendido entre 0 e 1,

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

no qual os valores mais elevados representam maior grau de vulnerabilidade social. O Quadro 1 detalha a operacionalização dos critérios categóricos que compõem o Score Social (S*social).

Quadro 1. Operacionalização e pontuação dos atributos sociais categóricos

Atributo Categórico	Condição / Categoria	Pontuação (vk)	Referência normativa do critério
Pessoa com Deficiência (PCD)	Sim	1,0	Lei nº 13.409/2016 e Lei nº 14.723/2023
	Não	0,0	
Origem Escolar	Escola Pública Integral	1,0	Lei nº 12.711/2012
	Escola Pública Parcial	0,7	
	Privada com Bolsa Integral	0,4	
	Privada Ampla	0,0	
Comunidades Tradicionais	Indígena / Quilombola / Ribeirinho	1,0	Constituição Federal e Lei nº 14.723/2023
	Não pertencente	0,0	
Localização Geográfica	Zona Rural / Áreas Isoladas	1,0	Diretrizes do PNAES
	Periferia Urbana	0,5	
	Zona Urbana Central	0,0	
Raça / Cor / Etnia	Pretos, Pardos e Indígenas (PPI)	1,0	Lei nº 14.723/2023
	Demais grupos / Não declarado	0,0	
Composição Familiar	3 ou mais dependentes no lar	1,0	Critérios Socioeconômicos do PNAES
	1 a 2 dependentes	0,5	
	Nenhum dependente	0,0	

Fonte: Elaborado pelos autores (2026).

As categorias e dimensões apresentadas no Quadro 1 foram utilizadas como referências para representar aspectos socioeconômicos e sociais considerados pertinentes às políticas

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

afirmativas. Entretanto, os valores numéricos atribuídos a cada categoria constituem escolhas de modelagem específicas deste estudo, não correspondendo a pesos ou pontuações diretamente estabelecidos pela legislação. Dessa forma, os valores adotados devem ser interpretados como parâmetros experimentais, sujeitos a análise de sensibilidade e revisão conforme o contexto institucional.

2.1.1. Cálculo do Score Social

O Score Social (S^*_{social}) é obtido pela soma normalizada das pontuações (vk) dos seis atributos categóricos do Quadro 1, aplicado na Equação 1:

$$S^*_{social} = \frac{\sum_{\mathcal{K}}^{6} = 1\ vk}{6} \qquad (1)$$

Em que vk representa a pontuação atribuída ao candidato em cada um dos seis atributos categóricos (PCD, origem escolar, comunidades tradicionais, localização geográfica, raça/cor/etnia e composição familiar), e o denominador (6) corresponde à pontuação máxima possível quando todos os atributos assumem o valor 1,0. O resultado é, portanto, um índice contínuo no intervalo [0, 1].

2.1.2. Cálculo do Índice de Renda

O Índice de Renda (I_{renda}) representa a condição econômica do candidato de forma inversa à renda familiar per capita declarada: quanto menor a renda per capita, maior o índice de vulnerabilidade associado. A normalização adotada na Equação 2:

$$I_{renda} = 1 - \min\left(\frac{rendaPerc}{rendaLimite}, 1\right) \qquad (2)$$

Em que $rendaPerc$ representa a renda familiar per capita declarada pelo candidato, enquanto $rendaLimite$ corresponde ao valor de referência de 1 (um) salário-mínimo per capita, adotado com base nos critérios de elegibilidade estabelecidos pela Lei nº 12.711/2012, com as alterações introduzidas pela Lei nº 14.723/2023. A função $\min$ limita a razão entre a renda per capita e o limite de referência ao valor máximo de 1, garantindo que o índice permaneça no intervalo [0,1]. Dessa forma, candidatos cuja renda familiar per capita seja igual ou superior ao valor de $rendaLimite$ recebem ($I_{renda} = 0$), enquanto candidatos com renda familiar per capita igual a zero recebem ($I_{renda} = 1$). Valores intermediários de renda familiar per capita resultam em ($I_{renda} = n$) valores proporcionalmente decrescentes, conforme visto na Equação 2.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

2.1.3. Equação do Índice de Vulnerabilidade Social (IVS)

A Equação 3 apresenta a formulação do Índice de Vulnerabilidade Social (IVS), que estabelece uma combinação linear ponderada entre o Índice de Renda (I_{renda}) e o Score Social (S^*_{social}):

$$IVS = 0{,}40 \cdot I_{renda} + 0{,}60 \cdot S^*_{social} \quad (3)$$

O Índice de Renda recebe peso de 0,40 na composição final, enquanto o Score Social recebe peso de 0,60, refletindo a decisão de dar maior peso ao componente social multidimensional. Ressalta-se que esses pesos foram atribuídos a priori como escolha de modelagem experimental, e não estimados a partir de calibração estatística. A opção por atribuir maior peso ao componente social decorre do objetivo declarado do estudo de representar vulnerabilidade em múltiplas dimensões; portanto, essa ponderação deve ser tratada como hipótese de modelagem, e não como valor universalmente válido. A análise de sensibilidade apresentada nos resultados é utilizada para verificar a influência dessa escolha sobre a classificação dos candidatos.

2.2. Módulo II – Processamento de Linguagem Natural (PLN)

Paralelamente ao cálculo do IVS, o sistema realiza a análise das respostas discursivas fornecidas pelos candidatos. Para essa etapa, foi empregado o *checkpoint* paraphrase-multilingual-MiniLM-L12-v2, disponibilizado por meio da biblioteca *sentence-transformers*. Trata-se de um modelo multilíngue baseado na arquitetura *Transformer*, compatível com textos em língua portuguesa, que gera vetores de *embedding* com 384 dimensões. Como parte do procedimento experimental, foram utilizadas duas perguntas padronizadas: “Quais as suas expectativas em relação ao curso?” e “Quais as expectativas profissionais após terminar o curso?”. Para cada pergunta, os candidatos forneceram respostas discursivas, em campo configurado para receber até 250 caracteres. Após o pré-processamento textual, as respostas foram convertidas em representações vetoriais e comparadas com os textos de referência, fornecidos pela instituição e utilizados como representação dos perfis de motivação e alinhamento considerados pertinentes ao processo seletivo. O grau de similaridade obtido nessa comparação é posteriormente convertido em uma Nota Semântica (NS), conforme a Equação 4.

$$NS = cos\,(\theta\,) = \left(\frac{u \cdot v}{||u|| \cdot ||v||}\right) \quad (4)$$

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

A Equação 4 apresenta a métrica de similaridade de cosseno utilizada para calcular a Nota Semântica (NS). O cálculo avalia o grau de proximidade semântica entre o vetor de *embeddings* das respostas discursivas dos candidatos (u) e o vetor do perfil de referência (v), dividido pelo produto de suas respectivas normas euclidianas ($||u|| \cdot ||v||$).

Na formulação convencional, a similaridade do cosseno varia no intervalo −1 a 1. Para que o resultado seja expresso como Nota Semântica Normalizada (PLN), aplica-se a Equação 5, para a normalização em uma escala de 0 a 1.

$$PLN = \frac{NS + 1}{2} \qquad (5)$$

2.3. Módulo III – Função de Aptidão (Fitness)

Após a obtenção do IVS e da Nota Semântica normalizada, ambas as informações são utilizadas pela função de aptidão (*fitness*), responsável por combinar esses indicadores em um único valor de avaliação. Esse valor é então utilizado pelo módulo de otimização baseado em AG. Nesse modelo, cada indivíduo da população representa uma possível combinação de candidatos selecionados, enquanto a função de aptidão (*fitness*) determina a qualidade de cada indivíduo a partir da integração entre a vulnerabilidade social e o alinhamento semântico, conforme representado na Equação 6.

$$FITNESSi = \alpha \cdot IVSi + \beta \cdot PLNi \quad (6)$$

A Equação 6 apresenta a função de aptidão ($FITNESSi$) responsável por orientar a avaliação de cada indivíduo i na população do algoritmo. A equação realiza a fusão multicritério, somando a multiplicação do $IVSi$ pelo coeficiente de peso α (Alpha) com a multiplicação da Nota Semântica normalizada ($PLNi$) pelo coeficiente de peso β (Beta), a soma dos pesos deve ser igual a um. Foi adotada na implementação experimental a distribuição com pesos iguais (α=0,50 e β=0,50). Essa parametrização estabelece contribuição inicial equivalente entre os dois critérios, sendo posteriormente avaliada por análise de sensibilidade.

2.4. Módulo IV – Algoritmo Genético (AG)

Durante a etapa de execução do Módulo IV, a população inicial é gerada aleatoriamente e submetida às operações de seleção, cruzamento e mutação. Para a reprodução do experimento, foram adotados os seguintes parâmetros: tamanho da população de 100 indivíduos; número máximo de 100 gerações; seleção por torneio, com cinco indivíduos participantes em cada disputa; *crossover* uniforme, aplicado em 100% dos cruzamentos, com

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

probabilidade de 50% por gene para a escolha do progenitor; mutação do tipo *bit flip*, com taxa de 5% por gene; elitismo correspondente a 10% da população, preservando os 10 melhores indivíduos de cada geração; critério de parada definido pelo atingimento de 100 gerações, sem critério antecipado de convergência; e mecanismo de reparo dos cromossomos para garantir exatamente 30 candidatos selecionados, correspondente ao número de vagas disponíveis. Não foi definida uma semente aleatória específica, sendo utilizada a inicialização padrão do gerador aleatório. O experimento foi executado uma única vez.

Ao final de cada geração, os indivíduos são avaliados pela função de aptidão (*fitness*), que combina IVS e PLN, sendo os melhores preservados pelo elitismo e os demais submetidos às operações evolutivas. Após 100 gerações, a solução de maior *fitness* é selecionada. Por se tratar de um problema de otimização, e não de classificação supervisionada, foram utilizados como indicadores o melhor *fitness* e o *fitness* médio, sua evolução ao longo das gerações, o tempo de execução, a composição dos selecionados, a distribuição dos escores de IVS e PLN e a sensibilidade aos pesos da função de aptidão, além da comparação entre a seleção baseada somente no IVS e aquela que incorpora o PLN.

Para verificar possíveis diferenças entre grupos, foram comparadas as proporções dos candidatos em diferentes categorias sociais na população total e entre os selecionados. Essa análise tem caráter diagnóstico e não é interpretada, isoladamente, como prova de equidade ou ausência de viés.

A arquitetura foi concebida como ferramenta de apoio, mantendo a decisão final sob responsabilidade da comissão avaliadora. Os dados foram tratados mediante remoção dos identificadores diretos e utilizados exclusivamente para fins experimentais. Considerando a presença de informações socioeconômicas e atributos potencialmente sensíveis, o tratamento dos dados deve observar os princípios aplicáveis da Lei nº 13.709/2018, incluindo finalidade, necessidade, segurança e proteção dos titulares.

3. RESULTADOS E DISCUSSÃO

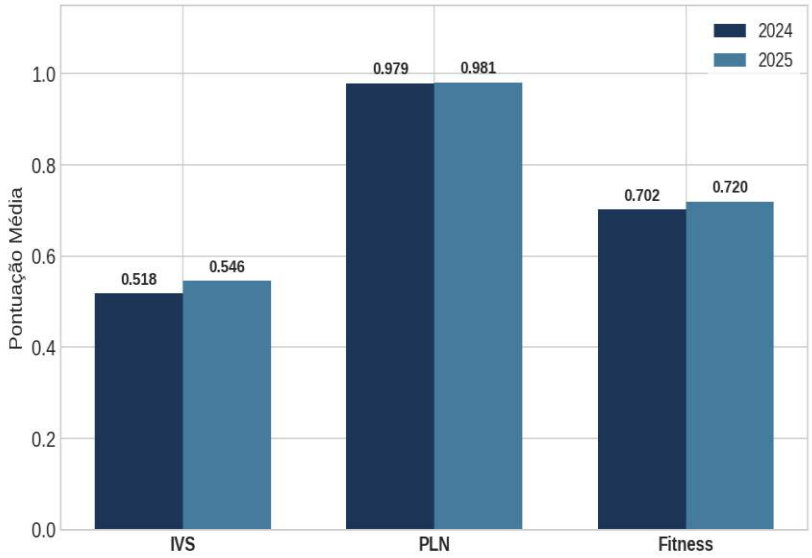
A arquitetura proposta foi avaliada utilizando uma base de dados proveniente de processos seletivos de cursos de qualificação profissional, composta por 227 candidatos inscritos no processo de 2024 e 160 em 2025. A avaliação teve como objetivo analisar o comportamento dos módulos integrados (IVS, PLN e AG), verificar a contribuição marginal de cada indicador na ordenação e discutir os limites do modelo diante das necessidades apontadas de calibração, análise de sensibilidade e governança algorítmica.

A Figura 2 apresenta a comparação das médias de IVS, PLN e *fitness* nos dois processos analisados. Observa-se um aumento discreto do IVS médio em 2025 em relação a 2024,

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

indicando uma maior concentração relativa de candidatos com os critérios de vulnerabilidade considerados pelo modelo. A Nota Semântica (PLN), por sua vez, apresentou valores médios elevados nos dois conjuntos, aproximadamente 0,88, acompanhados de baixa dispersão. Esse comportamento indica que as respostas discursivas apresentaram elevada similaridade em relação aos textos de referência utilizados pelo módulo de PLN. Como o *fitness* resulta da combinação entre IVS e PLN, a elevação observada no IVS também se refletiu no valor médio dessa função.

Figura 2. Comparação média de IVS, PLN e *Fitness* (2024 vs 2025)

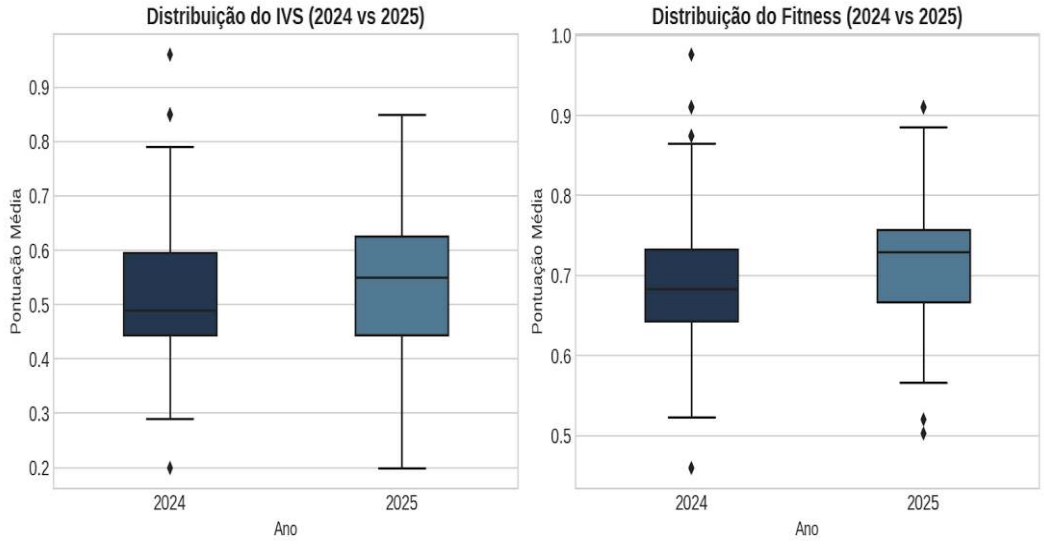


Fonte: Elaborado pelos autores (2026).

A distribuição dos indicadores, apresentada na Figura 3, evidencia comportamento semelhante nos dois conjuntos de dados. Os valores de IVS e *fitness* apresentam dispersão moderada, sem ocorrência de valores extremos relevantes. Embora esse comportamento indique estabilidade descritiva dos indicadores nas bases analisadas, não permite afirmar, isoladamente, a robustez do algoritmo. Como a execução do AG foi realizada uma única vez e sem definição de semente aleatória, a avaliação não permite verificar a variabilidade da solução diante de diferentes inicializações. Dessa forma, os resultados devem ser interpretados como evidências referentes aos cenários experimentais analisados, sendo necessárias execuções repetidas, diferentes sementes, variações de hiperparâmetros e bases com maior heterogeneidade para uma avaliação mais ampla da estabilidade do método.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

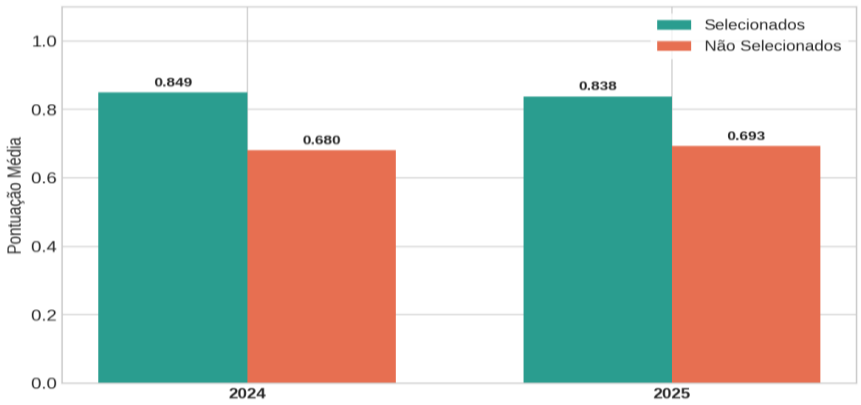
Figura 3. Distribuição (*Boxplot* IVS & *Fitness*)



Fonte: Elaborado pelos autores (2026).

Em relação à separação entre os grupos, a Figura 4 apresenta a comparação do *fitness* médio entre candidatos selecionados (30 primeiras colocações) e não selecionados. Os candidatos classificados apresentaram *fitness* médio numericamente superior ao dos não selecionados, com diferença média superior a 0,14 em ambos os anos. Esse resultado reflete diretamente o funcionamento do algoritmo, cuja função de aptidão foi concebida para maximizar a pontuação composta do grupo final. Trata-se, portanto, do comportamento esperado da função objetivo, e não de uma prova autônoma da qualidade ou equidade do método de seleção.

Figura 4. Comparação do *Fitness* Médio (Selecionados vs. Não Selecionados)

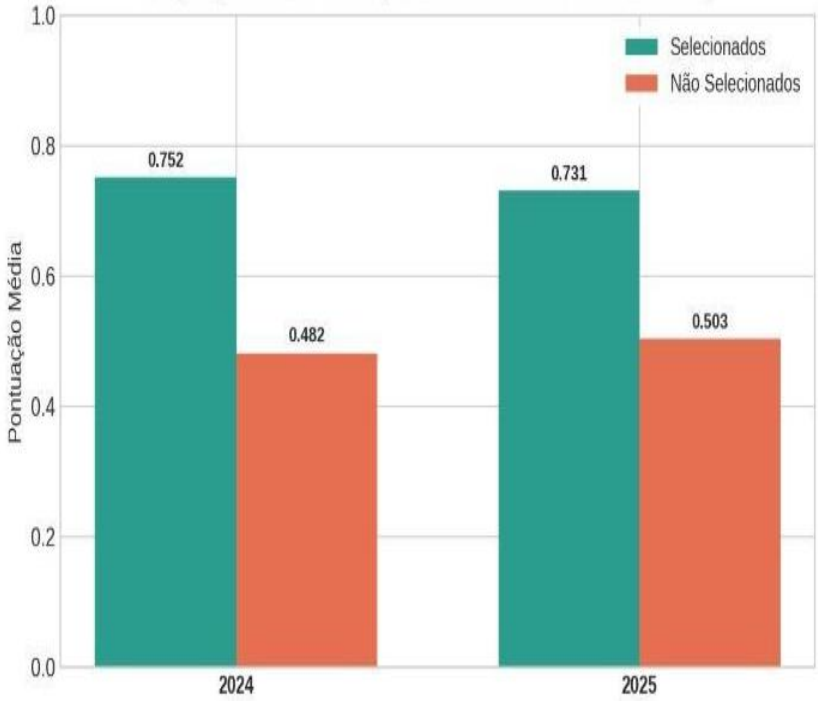


Fonte: Elaborado pelos autores (2026).

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

A análise do IVS (Figura 5) apresenta padrão semelhante, no qual os candidatos selecionados apresentam IVS médio superior ao dos não selecionados. Esse resultado é compatível com a composição da função de aptidão, que atribui ao IVS peso de 0,50, fazendo com que candidatos com maiores valores de vulnerabilidade social contribuam positivamente para a função objetivo.

Figura 5. Comparação do IVS Médio (Selecionados vs. Não Selecionados)

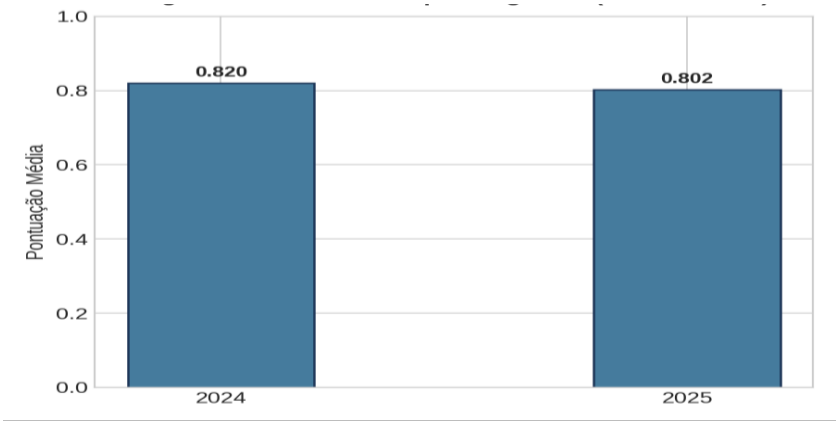


Fonte: Elaborado pelos autores (2026).

A Figura 6 apresenta a nota de corte correspondente ao 30º candidato da ordenação final. A variação observada entre 2024 e 2025 foi reduzida, indicando comportamento semelhante do limite de seleção nos dois conjuntos analisados. Entretanto, essa proximidade está restrita às bases utilizadas no experimento e não permite concluir que o modelo apresente previsibilidade geral diante de alterações mais significativas no perfil dos candidatos. A estabilidade observada deve, portanto, ser compreendida como uma característica descritiva dos dois processos avaliados, e não como uma propriedade generalizável do sistema.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

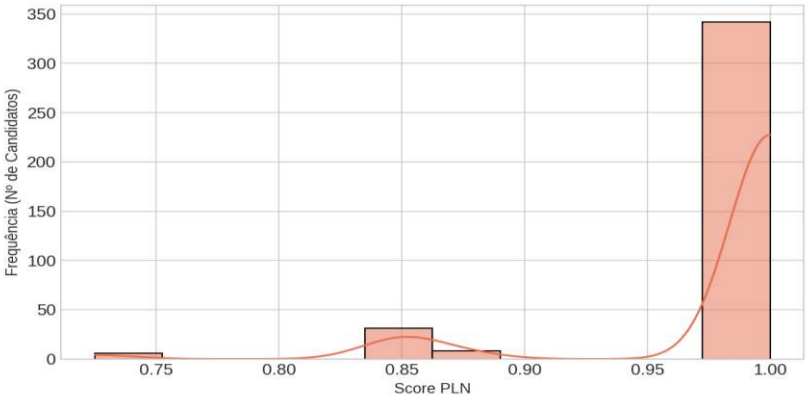
Figura 6. Nota de corte para ingresso (30° Colocado)



Fonte: Elaborado pelos autores (2026).

Em relação ao componente textual, a Figura 7 evidencia forte concentração das Notas Semânticas em valores superiores a 0,85. A baixa variabilidade observada reduz a capacidade do componente de diferenciar candidatos dentro das bases analisadas. Consequentemente, embora o PLN contribua para a composição do *fitness*, a maior parte da variação utilizada na ordenação dos candidatos decorreu das diferenças observadas no IVS. Nesse cenário experimental, o componente textual funcionou principalmente como indicador de proximidade entre as respostas dos candidatos e o perfil de referência definido para o processo. Esse resultado também evidencia uma limitação do modelo, pois a concentração das notas pode restringir a contribuição efetiva do PLN para a diferenciação da lista final. Estudos posteriores podem avaliar diferentes modelos, estratégias de treinamento ou procedimentos de calibração para verificar se é possível obter maior capacidade de diferenciação sem comprometer a interpretação dos resultados.

Figura 7. Distribuição do *score* PLN evidenciando baixa variabilidade entre candidatos



Fonte: Elaborado pelos autores (2026).

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Para avaliar o impacto da distribuição de pesos da função objetivo ($\alpha * IVS + \beta * NS$), realizou-se uma análise de sensibilidade variando as combinações de α e β. Na simulação com α=0,80 e β=0,20, a taxa de concordância na lista dos 30 selecionados em relação ao cenário padrão (α=0,50; β=0,50) foi de 93,3%, alterando a condição de apenas 2 candidatos. Por outro lado, na configuração invertida (α=0,30; β=0,70), a alteração atingiu 23,3% das vagas (7 candidatos). Os resultados indicam que a definição dos pesos influencia a composição da lista de selecionados, especialmente quando se atribui maior contribuição ao componente textual. Essa sensibilidade reforça que os valores adotados constituem escolhas de modelagem e devem ser definidos e avaliados em conjunto com os objetivos institucionais do processo seletivo, não sendo possível considerá-los como pesos universalmente válidos.

Quanto ao desempenho computacional, o tempo médio de execução do pipeline (sequência de etapas) completo, incluindo cálculo dos *embeddings*, processamento do IVS e otimização pelo AG, foi de aproximadamente 4,2 segundos por processo seletivo. Além disso, a comparação entre a solução obtida pelo AG e a ordenação determinística dos candidatos pelo *fitness* apresentou 100% de equivalência no cenário experimental sem sub-restrições combinatórias. Esse resultado é particularmente relevante porque demonstra que, para a formulação atualmente utilizada, a aplicação do AG não produziu uma solução diferente daquela obtida pela ordenação direta dos candidatos segundo a função objetivo. Assim, não foi observada vantagem do método evolutivo quanto à composição final da seleção nesse cenário. A utilização do AG permanece tecnicamente pertinente como estrutura extensível para situações futuras nas quais sejam incorporadas restrições combinatórias simultâneas, modalidades de cotas ou outras condições que tornem a solução não trivial por ordenação direta.

De maneira geral, os resultados demonstram a viabilidade técnica da integração entre IVS, PLN e AG para processamento e ordenação de candidatos nos cenários analisados. Entretanto, os experimentos também evidenciaram limitações relevantes: a baixa variabilidade das notas produzidas pelo PLN, a influência dos pesos definidos a priori, a equivalência entre AG e ordenação determinística no cenário sem restrições combinatórias e a ausência de avaliações formais de *fairness* e de múltiplas execuções do AG. Dessa forma, os resultados sustentam o uso da arquitetura como proposta experimental de apoio à decisão, mas não permitem afirmar superioridade, robustez geral, equidade ou justiça distributiva do método. A adoção institucional dependeria de validação adicional, incluindo testes em diferentes bases e instituições, análises sistemáticas de sensibilidade, avaliações formais de possíveis disparidades entre grupos e definição de mecanismos de governança e supervisão humana.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

4. CONSIDERAÇÕES FINAIS

Este estudo apresentou o desenvolvimento de uma arquitetura híbrida de inteligência artificial para apoiar a seleção e a ordenação de candidatos a partir da integração de critérios socioeconômicos e informações discursivas. A proposta combinou um Índice de Vulnerabilidade Social (IVS), representação semântica das respostas dos candidatos por meio de modelo *Transformer* e um algoritmo genético para a composição do conjunto de candidatos selecionados.

Os resultados demonstraram a viabilidade técnica da abordagem para integrar diferentes dimensões em um fluxo único de análise. Entretanto, observou-se que os escores provenientes do módulo de PLN apresentaram baixa variabilidade entre os candidatos, limitando sua capacidade de diferenciação no cenário avaliado. A análise de sensibilidade também indicou que alterações nos pesos atribuídos ao IVS e ao componente semântico podem modificar a composição da seleção, evidenciando que esses parâmetros constituem escolhas de modelagem e não configurações universalmente adequadas.

Quanto ao algoritmo genético, sua solução apresentou equivalência com a ordenação determinística no cenário experimental sem restrições combinatórias adicionais. Esse resultado indica que não foi demonstrada vantagem do algoritmo genético sobre a solução determinística nas condições avaliadas, uma vez que a seleção dos 30 candidatos com maiores scores individuais pode ser obtida diretamente por ordenação. Dessa forma, a utilização do algoritmo genético mostra-se mais justificável em cenários futuros nos quais sejam incorporadas restrições de composição, cotas ou outras condições que transformem o problema em uma otimização efetivamente combinatória.

Diante desses resultados, a arquitetura não deve ser interpretada como mecanismo autônomo de decisão ou como instrumento capaz de assegurar, por si só, equidade ou ausência de vieses no processo seletivo. Sua utilização deve permanecer subordinada à avaliação humana e a critérios institucionais previamente estabelecidos. Entre as principais limitações do estudo estão a baixa capacidade discriminativa observada no componente de PLN, a dependência dos resultados em relação aos pesos definidos no modelo, a ausência de múltiplas execuções do algoritmo genético com diferentes sementes e a não realização de métricas formais de disparidade entre grupos.

Como trabalhos futuros, recomenda-se ampliar a avaliação do componente semântico, testar diferentes modelos e estratégias de representação textual, realizar múltiplas execuções do algoritmo genético e avaliar sistematicamente a sensibilidade de seus parâmetros. Também se mostra relevante incorporar restrições combinatórias ao problema de seleção, comparar

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

diferentes estratégias de otimização e empregar métricas formais de equidade e disparidade entre grupos. Essas extensões podem contribuir para uma avaliação mais abrangente da arquitetura e para sua eventual utilização como ferramenta de apoio à decisão em contextos institucionais reais.

REFERÊNCIAS

BASSO-POLETTO, Daniela; EFROM, Cora; RODRIGUES, Maria Beatriz. Ações afirmativas no ensino superior: revisão quantitativa e qualitativa da literatura. *Revista Electrónica Educare*, Heredia, v. 24, n. 1, p. 1-18, 2020. Disponível em: https://www.scielo.sa.cr/scielo.php?script=sci_arttext&pid=S1409-42582020000100292. Acesso em: 31 jul. 2026.

BRASIL. Lei nº 12.711, de 29 de agosto de 2012. Dispõe sobre o ingresso nas universidades federais e nas instituições federais de ensino técnico de nível médio e dá outras providências. Brasília, DF: Presidência da República, 2012.

BRASIL. Lei nº 13.409, de 28 de dezembro de 2016. Altera a Lei nº 12.711, de 29 de agosto de 2012, para dispor sobre a reserva de vagas para pessoas com deficiência nos cursos técnico de nível médio e superior das instituições federais de ensino. Brasília, DF: Presidência da República, 2016.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília, DF: Presidência da República, 2018.

BRASIL. Lei nº 14.723, de 13 de novembro de 2023. Altera a Lei nº 12.711, de 29 de agosto de 2012, para atualizar a política de cotas no âmbito das instituições federais de educação superior e de ensino técnico de nível médio. Brasília, DF: Presidência da República, 2023.

COSTA JÚNIOR, João Fernando et al. A inteligência artificial como ferramenta de apoio no ensino superior. *REBENA – Revista Brasileira de Ensino e Aprendizagem*, [S. l.], v. 6, p. 246–269, 2023. Disponível em: https://rebena.emnuvens.com.br/revista/article/view/111. Acesso em: 29 jul. 2026.

CUNHA, Mário de Jesus et al. Transformações no ensino superior mediadas pela inteligência artificial: uma revisão de estudos sobre ensino, aprendizagem e avaliação. *Interference: A Journal of Audio Culture*, [S. l.], v. 12, n. 1, p. 483–497, 2026. Disponível em: https://interferencejournal.emnuvens.com.br/revista/article/view/747. Acesso em: 29 jul. 2026.

DEVLIN, J. et al. BERT: pre-*training* of deep bidirectional *transformers* for language understanding. arXiv preprint, 2018. Disponível em: https://arxiv.org/abs/1810.04805. Acesso em: 30 jul. 2026.

FERNANDES, Taynara Roberta; PASSADOR, Claudia Souza. Contexto socioeconômico e infraestrutura escolar no desempenho acadêmico: revisão sistemática da literatura. *Regae: Revista de Gestão e Avaliação Educacional*, [S. l.], v. 12, n. 21, e84346, 2023. DOI: 10.5902/2318133884346. Disponível em: https://periodicos.ufsm.br/regae/article/view/84346. Acesso em: 17 ago. 2026.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

HEGGLER, J. M.; SZMOSKI, R. M.; MIQUELIN, A. F. As dualidades entre o uso da inteligência artificial na educação e os riscos de vieses algorítmicos. *Educação & Sociedade*, v. 46, e289323, 2025. Disponível em: https://doi.org/10.1590/ES.289323. Acesso em: 30 jul. 2026.

HOLLAND, John H. *Adaptation in natural and artificial systems*. Cambridge, MA: MIT Press, 1992. DOI: https://doi.org/10.7551/mitpress/1090.001.0001. Acesso em: 30 jul. 2026.

HOLMES, Wayne et al. *Artificial intelligence in education: promises and implications for teaching and learning.* London: Routledge, 2019. Disponível em: https://www.researchgate.net/publication/332180327_Artificial_Intelligence_in_Education_Promise_and_Implications_for_Teaching_and_Learning. Acesso em: 31 jul. 2026.

LIPSKY, Michael. *Burocracia de nível de rua: dilemas do indivíduo nos serviços públicos.* Tradução de Arthur Eduardo Moura da Cunha. Brasília, DF: Escola Nacional de Administração Pública (Enap), 2019. 430 p. Disponível em: https://bibliotecadigital.enap.gov.br/bitstream/1/4158/1/Burocracia%20de%20n%C3%ADvel%20de%20rua_Michael%20Lipsky.pdf. Acesso em: 30 jul. 2026.

MACHADO, Jessica Arnaut Prado. *Inteligência artificial aplicada ao apoio do ensino-aprendizagem na educação profissional e tecnológica: uma reflexão crítica sobre a prática docente em cursos técnicos.* 2026. 29 f. Trabalho de Conclusão de Curso (Especialização em Docência na Educação Profissional e Tecnológica) – Instituto Federal de Educação, Ciência e Tecnologia da Bahia, Camaçari, 2026. Disponível em: https://repositorio.ifba.edu.br/jspui/bitstream/123456789/1187/1/TCC_Jessica%20Arnaut%20Prado%20Machado.pdf. Acesso em: 17 ago. 2026.

MATOS, M. de Q. S.; VIANA, H. B.; NASCIMENTO, E. R.; TAVARES, C. Z. Inteligência artificial na educação: como algoritmos adaptativos podem elevar o desempenho acadêmico. *Revista on line de Política e Gestão Educacional,* Araraquara, v. 30, n. 00, p. e026056, 2026. DOI: 10.22633/rpge.v30i00.21254. Disponível em: https://periodicos.fclar.unesp.br/rpge/article/view/21254. Acesso em: 17 ago. 2026.

MITCHELL, Melanie. An introduction to *genetic algorithms*. Cambridge: MIT Press, 1998.

MOREIRA, Aline Leitão; OLIVEIRA, Meirivâni Meneses de; OLIVEIRA, Cintya Kelly Barroso; ARAÚJO, Francisco Adeil Gomes de; FERNANDES, Maria Jucineide da Costa. Inteligência artificial na escola: um relato de formação continuada para professores da rede pública estadual de ensino do Ceará. *Revista Docentes*, [S. l.], v. 10, n. 39, p. 48–54, 2025. Disponível em: https://periodicos.seduc.ce.gov.br/revistadocentes/article/view/1869. Acesso em: 17 ago. 2026.

NOGUEIRA, Luiz Guilherme Curcino. *Apoiando a gestão de projetos de software em educação: uma abordagem com inteligência artificial.* 2025. Monografia (Bacharelado em Ciência da Computação) – Instituto Federal Goiano, Campus Morrinhos, Morrinhos, 2025. Disponível em: https://repositorio.ifgoiano.edu.br/bitstream/prefix/5502/1/tcc_Luiz%20Guilherme.pdf. Acesso em: 17 ago. 2026.

RUSSELL, Stuart; NORVIG, Peter. *Artificial intelligence: a modern approach.* 4. ed. Hoboken: Pearson, 2022.

SANTOS, Sanval Ebert de Freitas. *Modelo conceitual para ambientes virtuais de ensino e aprendizagem com integração de técnicas de inteligência artificial*. 2025. 128 f. Tese (Doutorado em Difusão do Conhecimento) – Instituto Federal de Educação, Ciência e Tecnologia da Bahia, Salvador, 2025. Disponível em:

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

https://repositorio.ifba.edu.br/jspui/bitstream/123456789/966/1/Tese_Sanval%20Ebert%20de%20Freitas%20Santos_PPGC.pdf. Acesso em: 17 ago. 2026.

SILVA, Guilherme Henrique Gomes da. *Equidade no acesso e permanência no ensino superior: o papel da educação matemática frente às políticas de ações afirmativas*. 2016. Trabalho acadêmico – Universidade Estadual Paulista (UNESP), 2016. Disponível em: https://repositorio.unesp.br/entities/publication/554428ba-3368-450a-8615-56e15e0b5e67. Acesso em: 30 jul. 2026.

SOUSA, Andréia da Silva Quintanilha; NASCIMENTO, Emanuelle Lourenço do. Ações afirmativas na educação superior: redistribuição, reconhecimento e representação em disputa. *PerCursos*, Florianópolis, v. 27, e0203, 2026. DOI: https://doi.org/10.5965/19847246272026e0203. Acesso em: 30 jul. 2026.

TAVARES SANCHES, Tarcisio; ÂNGELO RODRIGUES MOCBEL, Miguel; CALDAS DE CASTRO, Vicente; AUGUSTO MACHADO DOS SANTOS, Carlos; GONÇALVES MARTINS, Leonardo; MAÍRA MENEZES DE OLIVEIRA, Waldma; JÚNIOR MELO E SILVA, Álvaro; DE SOUZA FARIAS, Fabricio. Avaliação da influência de fatores sociais e econômicos no uso da metodologia PBL + M-Learning para o ensino da matemática. *RENOTE*, Porto Alegre, v. 17, n. 1, p. 326–335, 2019. DOI: 10.22456/1679-1916.95800. Disponível em: https://seer.ufrgs.br/index.php/renote/article/view/95800. Acesso em: 21 ago. 2026.

VASWANI, Ashish et al. *Attention is all you need.* In: Advances in Neural Information Processing Systems. 2017. Disponível em: https://arxiv.org/abs/1706.03762. Acesso em: 31 jul. 2026.

VIEIRA, Enilton Ferreira; VIDEIRA, Piedade Lino; CUSTÓDIO, Elivaldo Serrão. As políticas de ações afirmativas na graduação e pós-graduação da UNIFAP: entre acesso, permanência e equidade. *Linguagens, Educação e Sociedade*, [S. l.], v. 29, n. 61, p. 1–30, 2025. Disponível em: https://www.periodicos.ufpi.br/index.php/lingedusoc/article/view/7088. Acesso em: 30 jul. 2026.

WAINER, Jacques; MELGUIZO, Tatiana. Políticas de inclusão no ensino superior: avaliação do desempenho dos alunos baseado no Enade de 2012 a 2014. *Educação e Pesquisa*, São Paulo, v. 44, e162807, 2018. DOI: https://doi.org/10.1590/S1517-9702201612162807. Disponível em: http://educa.fcc.org.br/scielo.php?pid=S1517-97022018000100305&script=sci_arttext. Acesso em: 30 jul. 2026.

ZAWACKI-RICHTER, Olaf et al. Systematic review of research on artificial intelligence applications in higher education – where are the educators?. *International Journal of Educational Technology in Higher Education*, v. 16, n. 39, 2019. Disponível em: https://www.researchgate.net/publication/336846972_Systematic_review_of_research_on_artificial_intelligence_applications_in_higher_education_-_where_are_the_educators. Acesso em: 31 jul. 2026.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.