v7.n11.2026

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

USO DE INTELIGÊNCIA ARTIFICIAL (IA) COMO FERRAMENTA DE AUXÍLIO EM ANÁLISE E APERFEIÇOAMENTO DE PERFIL PROFISSIONAL

USE OF ARTIFICIAL INTELLIGENCE (AI) AS A SUPPORT TOOL IN THE ANALYSIS AND IMPROVEMENT OF PROFESSIONAL PROFILES

USO DE LA INTELIGENCIA ARTIFICIAL (IA) COMO HERRAMIENTA DE APOYO EN EL ANÁLISIS Y PERFECCIONAMIENTO DEL PERFIL PROFESIONAL

Vinicius Gussoni Vellosa[1], André Luiz da Silva[2], Fabiana Florian[3]

e7119641
https://doi.org/10.47820/recima21.v7i11.9641

PUBLICADO 11.2026

RESUMO

Este trabalho teve por objetivo desenvolver e avaliar uma plataforma apoiada em Modelos de Linguagem de Grande Escala que produza diagnóstico estruturado de perfis profissionais. O método foi experimental e controlado: doze perfis sintéticos, construídos para a pesquisa e distribuídos em quatro áreas de atuação e três níveis de senioridade, acrescidos de onze variantes demográficas, foram submetidos a três execuções independentes cada um, com instrumentação do servidor. A concordância do diagnóstico foi aferida por verificador independente, que leu os mesmos documentos sem consultar a plataforma e respondeu aos treze critérios factualmente verificáveis. As duas leituras concordaram em 92,1% das verificações determinadas, medida restrita a critérios factuais de perfis construídos; a pontuação manteve-se estável em onze dos doze perfis, com a variação concentrada nas categorias julgadas pelo modelo; e não se detectou efeito sistemático de gênero, ano de conclusão do curso ou prestígio da instituição sobre a nota. A ferramenta de edição do agente conversacional, porém, foi acionada em apenas catorze dos quarenta turnos que a solicitavam, e em 43% dos turnos que nada pediam. Conclui-se que o sistema identifica corretamente o que falta no perfil e não executa com confiabilidade as alterações que promete, distinção que a verificação externa tornou visível e que delimita o que ferramentas dessa natureza podem hoje oferecer.

PALAVRAS-CHAVE: Agentes conversacionais. Aperfeiçoamento de perfil profissional. Inteligência artificial. Modelos de linguagem de grande escala. Reprodutibilidade.

ABSTRACT

This study aimed to develop and evaluate a platform supported by Large Language Models that produces a structured diagnosis of professional profiles. The method was experimental and controlled: twelve synthetic profiles, built for this research and distributed across four fields and three seniority levels, plus eleven demographic variants, were each submitted to three independent runs with server instrumentation. The agreement of the diagnosis was assessed by an independent verifier, which read the same documents without consulting the platform and answered the thirteen factually verifiable criteria.

[1] Graduando em Engenharia de Computação pela Universidade de Araraquara (Uniara). Araraquara, São Paulo, Brasil.
[2] Doutor e mestre em Educação Escolar pela Universidade Estadual Paulista (Unesp), MBA em Gestão de Projetos, bacharel em Informática e engenheiro de computação; docente dos cursos de Sistemas de Informação e Engenharia de Computação e do Mestrado Profissional em Engenharia de Produção da Universidade de Araraquara (Uniara) e *Head* de Dados e Arquitetura na Serasa Experian, com atuação em engenharia de software, arquitetura de dados, inteligência artificial e gestão de tecnologia. Araraquara, São Paulo, Brasil.
[3] Doutora e graduada em Ciências Econômicas; docente da Universidade de Araraquara (Uniara) com atuação em disciplinas de trabalho de conclusão de curso e metodologia científica. Araraquara, São Paulo, Brasil.

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.

The two readings agreed on 92.1% of the determined checks, a measure restricted to factual criteria of constructed profiles; scoring remained stable in eleven of the twelve profiles, with variation concentrated in the categories judged by the model; and no systematic effect of gender, graduation year or institutional prestige on the score was detected. The editing tool of the conversational agent, however, was triggered in only fourteen of the forty turns that requested it, and in 43% of the turns that requested nothing. It is concluded that the system correctly identifies what is missing from a profile and does not reliably carry out the changes it promises, a distinction made visible by external verification and which delimits what tools of this kind can currently offer.

KEYWORDS: *Conversational agents. Professional profile improvement. Artificial intelligence. Large language models. Reproducibility.*

RESUMEN

Este trabajo tuvo por objetivo desarrollar y evaluar una plataforma apoyada en Modelos de Lenguaje de Gran Escala que produzca un diagnóstico estructurado de perfiles profesionales. El método fue experimental y controlado: doce perfiles sintéticos, construidos para la investigación y distribuidos en cuatro áreas de actuación y tres niveles de antigüedad, con once variantes demográficas adicionales, fueron sometidos a tres ejecuciones independientes cada uno, con instrumentación del servidor. La concordancia del diagnóstico se midió mediante un verificador independiente, que leyó los mismos documentos sin consultar la plataforma y respondió a los trece criterios fácticamente verificables. Las dos lecturas coincidieron en el 92,1% de las verificaciones determinadas, medida restringida a criterios fácticos de perfiles construidos; la puntuación se mantuvo estable en once de los doce perfiles, con la variación concentrada en las categorías juzgadas por el modelo; y no se detectó efecto sistemático del género, del año de finalización del curso ni del prestigio de la institución sobre la nota. La herramienta de edición del agente conversacional, sin embargo, se activó en apenas catorce de los cuarenta turnos que la solicitaban, y en el 43% de los turnos que nada pedían. Se concluye que el sistema identifica correctamente lo que falta en el perfil y no ejecuta con fiabilidad los cambios que promete, distinción que la verificación externa hizo visible y que delimita lo que herramientas de esta naturaleza pueden ofrecer hoy.

PALABRAS CLAVE: *Agentes conversacionales. Perfeccionamiento del perfil profesional. Inteligencia artificial. Modelos de lenguaje de gran escala. Reproducibilidad.*

1. INTRODUÇÃO

A constante evolução e a competitividade do mercado de trabalho exigem que os indivíduos destaquem suas qualificações de maneira estratégica, e nesse cenário a Inteligência Artificial (IA) desponta como instrumento capaz de realizar atividades complexas, como o processamento de dados e a resolução de problemas (Russell; Norvig, 2013). Este trabalho aborda o uso de ferramentas de IA, especificamente o Processamento de Linguagem Natural (PLN), para a análise e o aperfeiçoamento de perfis profissionais, delimitando-se ao uso da tecnologia pelo próprio profissional, sem intermediação de terceiros.

Apesar da importância da curadoria de perfis, como currículos e redes sociais profissionais, quem pretende revisar a própria apresentação enfrenta três obstáculos: desconhece os critérios pelos quais o perfil será lido, carece de distanciamento crítico sobre o

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.

próprio percurso e dispõe de tempo limitado para a tarefa. A questão de pesquisa é, pois: de que maneira uma plataforma apoiada em Modelos de Linguagem de Grande Escala pode diagnosticar o perfil profissional, e que evidências sustentam que esse diagnóstico é acurado e reprodutível? Dela desdobram-se três outras: em que medida o diagnóstico identifica corretamente o que falta no perfil, quando confrontado com verificação independente? Qual o grau de reprodutibilidade dos resultados? E com que confiabilidade o agente conversacional executa as alterações solicitadas?

A hipótese que sustenta esta investigação é a de que a implementação de uma plataforma baseada em Modelos de Linguagem de Grande Escala (LLMs) automatizará o diagnóstico estrutural do perfil, identificando corretamente os elementos ausentes ou mal apresentados quando confrontada com verificação independente. Formula-se, ainda, a hipótese secundária de que critérios vinculados exclusivamente a competências técnicas e experiências declaradas tornarão a pontuação insensível a marcadores demográficos — gênero, ano de conclusão do curso e prestígio da instituição —, de modo que perfis idênticos em conteúdo técnico, distintos apenas quanto a esses marcadores, não apresentarão variação sistemática de nota. Sistemas apoiados em LLMs apresentam comportamento não determinístico, razão pela qual a reprodutibilidade é tratada nesta pesquisa como objeto de mensuração, e não como pressuposto.

O objetivo geral é desenvolver e avaliar uma plataforma que utiliza IA para diagnosticar o perfil profissional e apoiar o seu aperfeiçoamento. Os objetivos específicos são estruturar um sistema de pontuação e diagnóstico automatizado; processar dados textuais propondo melhorias em resumos e experiências de trabalho; mensurar o desempenho operacional, a reprodutibilidade e a sensibilidade a marcadores demográficos; aferir a concordância do diagnóstico com verificação independente; e medir a confiabilidade do acionamento de ferramenta pelo agente conversacional.

A escolha do tema justifica-se pela convergência de três fatores. Sob a perspectiva social e econômica, a competitividade do mercado impõe ao profissional a tarefa de apresentar-se de forma estruturada, ao passo que o acesso a orientação especializada permanece restrito por custo. Sob a perspectiva tecnológica, o avanço dos LLMs tornou viável processar e avaliar documentos textuais não estruturados em escala. Sob a perspectiva acadêmica, identificou-se lacuna na literatura nacional: embora existam sistemas voltados ao recrutamento e à triagem de candidatos pelo lado das empresas, como o sistema Grace (Medeiros *et al.*, 2023), há escassez de ferramentas orientadas ao aperfeiçoamento do perfil sob a ótica do próprio profissional. Este

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.

trabalho busca preencher essa lacuna, com solução que amplia o acesso ao diagnóstico estruturado e submete os próprios resultados a verificação empírica.

2. REFERENCIAL TEÓRICO

2.1. Fundamentos de inteligência artificial e processamento de linguagem natural

A Inteligência Artificial (IA) é o ramo da Ciência da Computação dedicado a desenvolver sistemas capazes de se comportar de forma inteligente, resolvendo problemas complexos e reconhecendo padrões (Jackson, 1985). Ampara-se no estudo de agentes que percebem o ambiente e executam ações, com o duplo objetivo de compreender a inteligência humana e de construir entidades dotadas de tal capacidade (Russell; Norvig, 2013), e abrange raciocínio, generalização e aprendizado a partir de experiências prévias (Copeland, 2021).

Entre as suas áreas especializadas destaca-se o Processamento de Linguagem Natural (PLN), dedicado a dotar os sistemas da capacidade de interpretar textos de forma estruturada (Pereira, 2018). Para analisar documentos não estruturados, o PLN clássico encadeia etapas de pré-processamento: padronização de caracteres e remoção de *stopwords*; lematização, que unifica sinônimos e adequa verbos ao tempo presente (Akhmetov *et al.*, 2020; Barbosa *et al.*, 2017); e vetorização, que converte palavras em representações numéricas por técnicas como o *Term Frequency–Inverse Document Frequency* (TF-IDF), que pondera cada termo pela frequência e pela relevância (Rangel, 2019).

Esse encadeamento é anterior à consolidação dos Modelos de Linguagem de Grande Escala. Em arquiteturas contemporâneas, como a adotada neste trabalho, normalização, lematização e vetorização deixam de ser executadas pelo desenvolvedor: o próprio modelo recebe o documento não estruturado e devolve dados organizados. A substituição traz ganho de simplicidade, mas implica a perda de três garantias do encadeamento clássico — o determinismo do resultado, a auditabilidade das etapas intermediárias e a ausência de custo de inferência —, perdas examinadas nas subseções 2.4 e 3.3.

2.2. Modelos de linguagem e engenharia de prompt

Modelos de linguagem são, historicamente, modelos probabilísticos que preveem a distribuição de probabilidade de sequências de palavras em um idioma (Russell; Norvig, 2013). Os Modelos de Linguagem de Grande Escala operam sobre a mesma premissa, calculando o próximo *token* mais provável a partir de volume colossal de dados de treinamento (Carraro, 2025). O salto que viabilizou a IA generativa moderna veio da arquitetura *Transformer* (Vaswani *et al.*, 2017), cujo mecanismo de autoatenção processa os dados em paralelo e calcula o peso

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.

relativo de cada palavra frente às demais, conferindo ao modelo compreensão dinâmica do contexto. Seguem-se o ajuste fino e o aprendizado por reforço a partir de *feedback* humano, que alinham as respostas a instruções complexas (Carraro, 2025). Tunstall, von Werra e Wolf (2022) destacam a capacidade desses modelos de produzirem saídas em formatos estruturados a partir de entradas textuais livres, propriedade que fundamenta tecnicamente o processamento de currículos descrito na seção 3.

A interação com esses modelos é mediada por instruções em linguagem natural, os *prompts*, e a engenharia de *prompt* reúne as técnicas de formulação dessas instruções de modo a maximizar a qualidade e a consistência das respostas (Carraro, 2025). Diferentemente da programação convencional, ela opera por condicionamento estatístico: o texto da instrução desloca a distribuição de probabilidade das saídas, sem jamais fixá-la. Três técnicas são empregadas no sistema desenvolvido: o *prompt* de sistema, instrução persistente que define papel, tom e regras de conduta; o aprendizado por poucos exemplos, que fornece na própria instrução amostras do comportamento desejado, recurso descrito por Tunstall, von Werra e Wolf (2022) como forma de adaptação sem novo treinamento; e a especificação de esquema de saída. Soma-se o controle da temperatura, que regula a aleatoriedade na seleção do próximo *token*: valores mais altos ampliam a diversidade ao custo da previsibilidade (Carraro, 2025).

2.3. Agentes inteligentes e chamada de função

A noção de agente é anterior à dos Modelos de Linguagem de Grande Escala e constitui um dos conceitos estruturantes da Inteligência Artificial. Define-se agente como a entidade capaz de perceber o ambiente por meio de sensores e de atuar sobre ele por meio de atuadores, orientando as suas ações à consecução de um objetivo; agente racional é aquele que seleciona a ação que maximiza a medida de desempenho esperada (Russell; Norvig, 2013). A formalização importa porque um modelo de linguagem, isoladamente, não é um agente: produz texto, mas não atua sobre o ambiente.

A conversão ocorre quando se lhe concede a capacidade de invocar ferramentas externas. Schick *et al.* (2023) demonstraram que modelos de linguagem podem aprender, de forma autossupervisionada, a decidir quando chamar uma interface de programação, quais argumentos fornecer e como incorporar o resultado à continuação do texto. Yao *et al.* (2023) propuseram a alternância explícita entre etapas de raciocínio e etapas de ação, evidenciando que o entrelaçamento entre o que o modelo pensa e o que executa reduz a incidência de afirmações não fundamentadas. É esse mecanismo, designado na literatura como chamada de função (*function calling*), que sustenta o componente conversacional da plataforma. Há, todavia,

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.

aspecto que a literatura consultada trata de forma ainda incipiente e que se revelou determinante nesta pesquisa: a decisão de invocar uma ferramenta é, ela própria, uma saída probabilística do modelo, conforme a subseção 3.3 documenta empiricamente.

2.4. Avaliação, reprodutibilidade e viés algorítmico

Avaliar sistemas apoiados em Modelos de Linguagem de Grande Escala apresenta dificuldades que não se colocam no software convencional. Chang *et al.* (2024), em revisão sistemática da literatura, organizam o problema em três eixos: o que avaliar, onde avaliar e como avaliar. Registram que métricas automáticas isoladas são insuficientes para tarefas abertas, nas quais não existe resposta única correta, e recomendam a combinação de medição automatizada com avaliação humana, arranjo adotado nesta pesquisa.

Ao problema da avaliação soma-se o da reprodutibilidade. Ouyang *et al.* (2024) investigaram empiricamente o não determinismo de modelos de linguagem em tarefas de geração de código e constataram variação substancial entre execuções sucessivas de uma mesma requisição, mesmo sob parâmetros idênticos. Observam que grande parte da literatura da área reporta resultados obtidos em execução única, prática que compromete a validade das conclusões, e defendem que toda medição sobre sistemas dessa natureza seja repetida e acompanhada de medidas de dispersão. Duas consequências orientam o desenho descrito na metodologia: a variabilidade é resultado tão relevante quanto o valor médio, e o não determinismo alcança também as decisões funcionais do modelo, entre elas a de invocar ou não uma ferramenta.

Exigência de verificação análoga incide sobre o viés. A aplicação de IA no recrutamento e na seleção consolidou-se como resposta à necessidade de otimizar a análise manual de grandes volumes de dados (Medeiros *et al.*, 2023), e as organizações recorrem a aprendizado de máquina e a ferramentas de *People Analytics* para reduzir o tempo de triagem (Lima; Morais Júnior, 2022). O risco associado é a discriminação algorítmica: decisões preditivas apoiadas em padrões extraídos de bases históricas tendem a reproduzir e a amplificar os preconceitos nelas contidos (Pantoja; Leão, 2022). O caso emblemático da literatura relata o descarte de um sistema de recrutamento treinado com histórico de currículos predominantemente masculinos, que passou a penalizar candidatas ao identificar o termo “mulheres” nos perfis (Fidalgo, 2022; Pantoja; Leão, 2022). Soma-se a exclusão dos chamados “trabalhadores escondidos”, candidatos qualificados rejeitados por não atenderem a palavras-chave pré-estabelecidas no código (Pantoja; Leão, 2022).

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 mitigação exige diretrizes de conformidade, auditoria de dados e transparência, e o sistema deve ser desenhado para atuar de forma contramajoritária, focando estritamente nas competências técnicas e experiências do candidato (Medeiros *et al.*, 2023). Sob essa condição, o aperfeiçoamento automatizado de currículos pode atenuar interferências discriminatórias, sem eliminá-las. Vale aqui o que vale para a avaliação e para a reprodutibilidade: a ausência de viés não decorre da automação em si e só pode ser afirmada mediante verificação, exigência que orienta o teste de sensibilidade descrito na metodologia.

3. METODOLOGIA

Quanto à natureza, a pesquisa classifica-se como aplicada; quanto aos objetivos, como exploratória e descritiva, por investigar a aplicabilidade de LLMs em domínio ainda pouco sistematizado na literatura nacional e por registrar o comportamento do sistema desenvolvido. A abordagem é quantitativa: medem-se tempo de resposta, volume de requisições, dispersão das pontuações entre execuções repetidas, variação frente a marcadores demográficos, taxa de acionamento de ferramenta e coincidência entre o diagnóstico do sistema e a verificação independente do mesmo documento. Os procedimentos técnicos são dois: revisão bibliográfica sistematizada e pesquisa experimental, mediante o desenvolvimento de um protótipo funcional e a sua submissão a ensaio controlado em ambiente isolado. Descreve-se a seguir a plataforma avaliada no que é pertinente à avaliação, pois cada decisão de projeto aqui registrada delimita uma propriedade que a seção de resultados submete a medição, e as falhas documentadas ao longo do percurso constituem, elas próprias, resultado da investigação.

3.1. A plataforma avaliada: arquitetura e fluxo

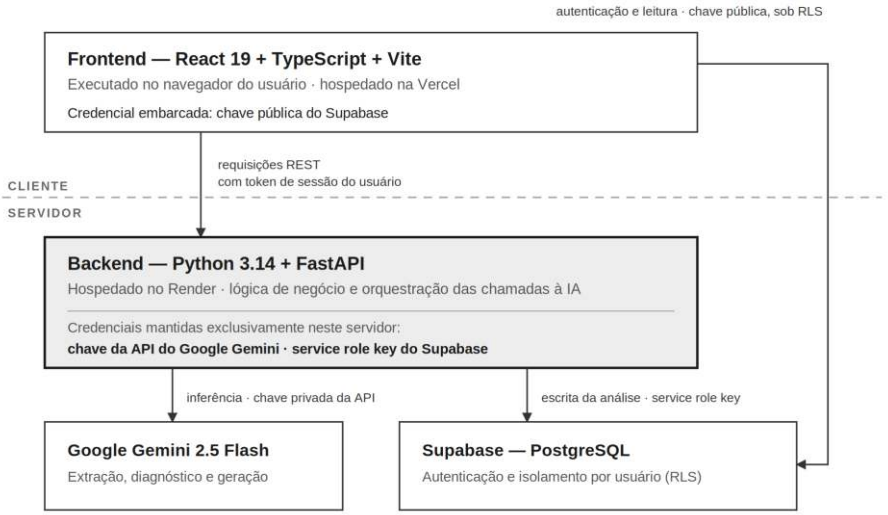
O Hudson.ai auxilia o profissional na análise e no aperfeiçoamento do próprio perfil, sem intermediação de terceiros, combinando diagnóstico automatizado, entrevista estratégica conduzida por agente conversacional e geração de materiais de posicionamento. O acesso se dá por interface web em três idiomas, e o idioma escolhido determina a língua de todo o material produzido pelo modelo. A versão aqui descrita corresponde ao estado do sistema em 21 de agosto de 2026.

A arquitetura é de três camadas desacopladas, no padrão cliente-servidor: o *frontend* em React 19 e TypeScript, hospedado na Vercel; o *backend* em Python com o *framework* FastAPI, hospedado no Render, responsável pela lógica de negócio, pela orquestração das chamadas ao Google Gemini e pela validação dos dados; e o banco gerenciado pelo Supabase, com PostgreSQL na nuvem, autenticação nativa e *Row Level Security*.

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 Figura 1 representa a organização das camadas e a distribuição das credenciais. O *frontend* comunica-se diretamente com o Supabase para autenticar e ler os dados do próprio usuário, mediante a chave pública do serviço e sob as políticas de *Row Level Security*; as credenciais privadas — a chave da API do Google Gemini e a *service role key* do Supabase — permanecem no servidor e jamais transitam pelo navegador.

Figura 1. Arquitetura em camadas e distribuição das credenciais entre cliente e servidor



Fonte: elaboração própria.

O percurso do usuário organiza-se em sete etapas sequenciais, do envio do perfil à entrega consolidada, com a progressão controlada por campo próprio na tabela de perfis, o que permite retomá-lo sem perda de contexto. O perfil é fornecido por arquivo PDF exportado do LinkedIn ou pela URL do perfil público; seguem-se o diagnóstico, a entrevista com o agente e a geração das campanhas de posicionamento e do *brief* estratégico, documento que consolida cargo-alvo, sinais de mercado e próximos passos. Concluído o *brief*, a análise é arquivada no histórico, e a falha de arquivamento passou a ser notificada, onde antes a perda do trabalho ocorria em silêncio.

A Figura 2 representa o percurso e indica o volume de requisições de cada etapa. O agente conversacional concentra a maior parte do processamento: das vinte a vinte e cinco requisições consumidas por uma sessão completa, de quinze a vinte originam-se dessa única etapa, porque o histórico do diálogo é reenviado a cada nova mensagem e a invocação de

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.

ferramenta demanda duas requisições, uma para decidir o acionamento e outra para produzir a resposta ao usuário.

Figura 2. Percurso do usuário na plataforma e distribuição das requisições ao modelo de linguagem



Fonte: elaboração própria.

A segurança apoia-se em três camadas: autenticação por *token* validado a cada requisição, isolamento por usuário mediante *Row Level Security* e credenciais sensíveis mantidas em variáveis de ambiente do servidor. As medidas guardam relação direta com o regime de proteção de dados examinado na subseção 2.4, pois a plataforma processa currículos, pretensões salariais e motivos de desligamento, informações que, no contexto de uma relação de trabalho, ensejam os cuidados discutidos por Lima e Morais Júnior (2022) a propósito do artigo 20 da Lei Geral de Proteção de Dados.

Uma restrição contratual do fornecedor de IA incide sobre a conformidade: no plano gratuito, os dados submetidos podem ser utilizados para o treinamento dos seus modelos, condição incompatível com a finalidade da plataforma. O plano pago foi contratado em 13 de outubro de 2026, antes da coleta aqui relatada, de modo que nenhum documento submetido durante os ensaios ficou sujeito a essa condição.

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.

3.2. Diagnóstico automatizado: o Hudson Score

O Hudson Score é o módulo de diagnóstico quantitativo da plataforma: pontuação na escala de 0 a 100 calculada pelo *backend* mediante a verificação de vinte e três critérios objetivos distribuídos em seis categorias. O serviço retorna, além da nota geral, o detalhamento por categoria, as lacunas críticas e as correções sugeridas. O Quadro 1 apresenta as categorias, os seus pesos brutos e os critérios considerados.

Quadro 1. Sistema de pontuação do Hudson Score

Categoria	Peso bruto	Critérios avaliados
Informações pessoais	**25 pts**	Nome, sobrenome, localização, correio eletrônico e telefone
Headline	**20 pts**	Aderência ao nível de senioridade, título corporativo, funções desempenhadas e especialidades
Resumo (About)	**10 pts**	Tom executivo, competências comportamentais e competências técnicas
Experiência profissional	**20 pts**	Clareza dos cargos, localização, datas e qualidade das descrições
Educação	**15 pts**	Instituição, grau obtido, área de formação, datas e descrição
Idiomas	**5 pts**	Quantidade de idiomas declarados e nível de proficiência
Total bruto	**95 pts**	Normalizado para a escala de 0 a 100

Fonte: elaboração própria.

A categoria de maior peso é Informações Pessoais, com 25 pontos, porque a ausência de dados de contato obsta a identificação do candidato por recrutadores; seguem-se *Headline* e Experiência Profissional, com 20 pontos cada. A soma dos pesos brutos é de 95 pontos, convertidos à escala de 0 a 100 pela razão entre a pontuação obtida e o máximo possível, normalização necessária porque a interface emprega o valor diretamente como percentual. Cabe declarar a origem desses pesos: foram fixados no desenho do sistema, a partir dos campos que a exportação do LinkedIn disponibiliza e da função de cada um na identificação do candidato, e não derivam de estudo de validação com recrutadores nem de calibração estatística. A escala mede, por conseguinte, a completude e a consistência formal do perfil frente a critérios declarados, e não a qualidade profissional de quem o assina; a sua validade de construto não foi submetida a verificação, limitação que se declara e que esta pesquisa não resolve.

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 escala vigente resulta de correção realizada em agosto de 2026. Até então o cálculo incluía um sétimo critério, a presença de fotografia profissional, com peso de 5 pontos; verificou-se que o arquivo PDF exportado pelo LinkedIn não contém a imagem do perfil e que, impossibilitado de verificar o critério, o modelo produzia respostas divergentes a cada execução. O critério foi removido do cálculo e convertido em recomendação textual. Dessa correção extrai-se princípio aplicável a todo sistema que combine código convencional e modelos de linguagem: não se deve solicitar ao modelo aquilo que o código pode verificar de forma determinística. Pontuações anteriores a 21 de agosto de 2026 empregavam escala distinta e não são comparáveis às posteriores.

3.3. O agente e o não determinismo da chamada de função

O agente Lander conduz a entrevista estratégica e responde pelas entregas qualitativas do sistema. É alimentado pelo modelo Google Gemini 2.5 Flash, configurado com *prompt* de sistema que define a sua persona e lhe impõe regras de conduta, com o tom fixado por exemplos incorporados à própria instrução, conforme a técnica descrita na subseção 2.2. A entrevista percorre oito etapas, do cargo-alvo e da urgência da busca à faixa salarial, à maior conquista profissional com métricas e às restrições inegociáveis. O agente dispõe de duas ferramentas: uma que altera o perfil em tempo real, com resultado refletido em painel lateral à conversa, e outra de pesquisa sobre empresas, acionada mediante solicitação explícita.

A edição do perfil pelo agente preserva a versão de partida. Até agosto de 2026 as alterações eram gravadas sobre os campos originais de título e resumo, de modo que o texto extraído se perdia; o sistema passou a manter duas versões, e as etapas subsequentes consomem a reposicionada. Sem o registro do ponto de partida, um sistema cuja finalidade é aperfeiçoar o perfil não teria como demonstrar a diferença que produziu. Ao término da entrevista, o Lander aciona serviços que geram quatro entregas consolidadas, entre elas o documento de posicionamento de mercado e a reescrita completa do perfil.

Durante os testes de agosto de 2026 observou-se comportamento anômalo no componente conversacional: o painel de exibição do perfil deixava de refletir as alterações, enquanto o agente afirmava textualmente tê-las realizado. Descartado o defeito de implementação — conferidos a instrução do agente, a declaração da ferramenta, a montagem do pacote de modificações no servidor e a aplicação pela interface —, reproduziu-se o mesmo diálogo em execuções sucessivas, e a causa evidenciou-se: em uma execução o modelo invocou a ferramenta de edição; em outra, limitou-se a descrever a alteração em texto, sem executá-la.

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.

O achado confirma, no domínio desta aplicação, a característica discutida na subseção 2.4: a chamada de função não é determinística. O modelo pode afirmar ter executado uma ação sem tê-la executado, possibilidade que deve ser assumida em todo ponto do sistema em que o efeito de uma ferramenta seja visível ao usuário. É propriedade dos modelos de linguagem, e não defeito de código. Três medidas foram implementadas: regra explícita na instrução do agente, vedando a afirmação de uma edição não acompanhada da efetiva invocação da ferramenta; redução da temperatura de 0,7 para 0,5; e registro de cada invocação no servidor, que permite distinguir a falha de execução da ausência de acionamento.

A taxa de acionamento anterior e posterior às medidas não foi mensurada no momento da correção, por esgotamento da cota diária de requisições. A eficácia da correção foi aferida nesta pesquisa, e os resultados da subseção 4.3 não a confirmam: a ferramenta foi acionada em 35% dos turnos que solicitam alteração, e um dos dois turnos de solicitação não a acionou em nenhuma das vinte repetições.

3.4. Propriedades avaliáveis do sistema

Esta subseção enuncia o que no artefato é passível de medição e contra que critério cada propriedade será julgada, de modo que a avaliação não se construa depois de conhecido o resultado. O Quadro 2 sintetiza o conjunto.

Quadro 2. Propriedades avaliáveis e respectivos critérios de julgamento

Propriedade	Como é medida	Critério de julgamento
Concordância do diagnóstico com verificação independente	Verificador independente aplica os treze critérios factuais aos doze perfis sintéticos	Coincidência com o que a plataforma apurou sobre o mesmo documento
Validade do verificador	Aplicação aos doze perfis sintéticos, de atributos conhecidos por construção	Coincidência com o gabarito e taxa de indeterminação
Reprodutibilidade do diagnóstico	Três execuções independentes de cada perfil da amostra sintética	Desvio-padrão inferior a cinco pontos na escala de zero a cem
Insensibilidade a marcadores demográficos	Pares de perfis distintos em um único marcador, três execuções cada	Diferença inferior à dispersão observada entre execuções repetidas
Confiabilidade do acionamento de ferramenta	Roteiro fixo de diálogo repetido para cada valor de temperatura	Taxa de acionamento nos turnos que solicitam edição
Custo temporal do ciclo	Instrumentação do servidor	Tempo medido por instrumentação sobre a amostra sintética

Fonte: elaboração própria.

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.

Duas distinções sustentam a leitura do quadro. A primeira separa o que o sistema mede de si mesmo daquilo que é conferido por fora dele: aferir o acerto da pontuação por ela mesma seria circular, razão pela qual a concordância do diagnóstico é medida por verificador independente, que lê o documento e responde aos treze critérios factuais sem consultar a plataforma. A segunda separa o que a instrumentação do servidor apura automaticamente, como tempos e volume de requisições, do que exige inspeção documento a documento, como a verificação dos critérios factuais.

À avaliação externa soma-se a verificação interna do código. A suíte automatizada reúne cinquenta e um testes implementados com Pytest, que cobrem o cálculo da pontuação, os modelos de dados e os montadores de instrução, sem efetuar chamadas à API do modelo de linguagem. A composição atual resulta da correção de um defeito de validade: a suíte anterior mantinha cópia da função de cálculo em seu próprio código e validava, portanto, uma réplica, e não o código executado em produção. Permanece limitação de cobertura, pois as rotinas que normalizam as saídas malformadas do modelo foram verificadas manualmente, sem integrar a suíte.

3.5. Amostra, instrumentos e análise

O universo compreende perfis profissionais em formato compatível com a exportação em PDF gerada pelo LinkedIn, e a pesquisa opera sobre doze perfis sintéticos construídos para este trabalho, distribuídos em quatro áreas de atuação e três níveis de senioridade, acrescidos de onze variantes derivadas de quatro deles. A opção por perfis construídos decorre de duas razões: o tratamento de currículos reais implicaria a coleta de dados pessoais, prática cuja problematização integra a fundamentação teórica deste trabalho, conforme a subseção 2.4; e somente perfis construídos admitem o controle isolado de variáveis, condição indispensável ao teste de sensibilidade, e possuem gabarito conhecido por construção, contra o qual se pode aferir a leitura do sistema.

Quanto aos aspectos éticos, o ensaio aqui relatado opera exclusivamente sobre perfis sintéticos construídos para esta pesquisa, com nomes, instituições e trajetórias fictícios. Não há, neste estudo, coleta de dados pessoais nem participação de pessoas, e o conjunto analisado reúne apenas medidas produzidas por instrumentação do servidor e pela inspeção dos documentos construídos.

Os instrumentos de coleta compreendem cinco conjuntos de registros, todos incidentes sobre a amostra sintética. O primeiro são registros automatizados de desempenho, obtidos por

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.

instrumentação do servidor: tempo de resposta por serviço, número de requisições ao modelo, volume de *tokens* e identificação do modelo acionado. O segundo mede a reprodutibilidade, submetendo cada perfil a três execuções independentes, com apuração de média, desvio-padrão e amplitude. O terceiro afere a taxa de acionamento de ferramenta pelo agente, por roteiro fixo de diálogo repetido cinco vezes para cada temperatura avaliada. O quarto é o teste de sensibilidade, em que quatro perfis-base e onze variantes, idênticos em conteúdo técnico e distintos apenas quanto a gênero, ano de conclusão do curso e prestígio da instituição, são submetidos a três execuções cada.

O quinto afere a concordância do diagnóstico com verificação independente. Treze dos vinte e três critérios do sistema de pontuação são factualmente verificáveis por inspeção do documento: nome, sobrenome, localização, endereço eletrônico e telefone; localização e datas das experiências; instituição, grau, área e datas da formação; e quantidade e proficiência dos idiomas. Um verificador desenvolvido para esta pesquisa lê o documento e responde a esses critérios sem consultar a plataforma, admitindo três estados: atende, não atende e não determinado, este reservado aos casos em que o padrão não permite decisão segura. O procedimento incide apenas sobre os perfis sintéticos, porque só neles existe gabarito, e o próprio verificador foi validado antes do uso, tendo coincidido com o gabarito em 152 das 152 verificações determinadas, com taxa de indeterminação de 2,6%. A coincidência entre essa verificação e o que a plataforma apurou mede, portanto, a concordância entre duas leituras do mesmo documento sobre critérios factuais, e não a acurácia global do diagnóstico, que nenhum instrumento aqui empregado afere.

A análise obedece a encadeamento fixo: aplicação da plataforma sobre a amostra sintética, execução dos ensaios, apuração das métricas de cada instrumento e resposta às questões formuladas na introdução. Apuram-se a coincidência entre a verificação independente e o diagnóstico, as medidas de dispersão entre execuções repetidas (adotando-se como critério de estabilidade o desvio-padrão inferior a cinco pontos na escala de zero a cem), a variação frente a marcadores demográficos, a taxa de acionamento de ferramenta e os tempos de processamento.

Delimita-se, desde já, o alcance das afirmações que esses dados autorizam. A verificação independente sustenta que o diagnóstico identifica corretamente o que falta no documento, mas incide sobre perfis construídos, únicos para os quais existe gabarito, e não autoriza estender a mesma concordância a documentos reais, cuja diversidade de formatação é maior. Os dados não sustentam que o perfil de um profissional tenha melhorado após o uso da plataforma, nem admitem comparação com o que ele obteria por outro meio, pois o desenho não

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.

comporta grupo de comparação nem medição do perfil em segundo momento: as propriedades do Quadro 2 respondem se o sistema faz aquilo a que se propõe, e com que estabilidade, não se o perfil de quem o utiliza melhora. O desfecho de mercado — entrevistas obtidas, recolocação, tempo de transição — e a utilidade percebida por quem utiliza a ferramenta permanecem fora do escopo.

4. RESULTADOS E DISCUSSÃO

Esta seção apresenta os resultados da aplicação da plataforma sobre a amostra sintética, na ordem dos instrumentos declarados. Os dados das duas primeiras subseções provêm de trinta e seis execuções do ciclo de extração e diagnóstico, conduzidas em 17 de outubro de 2026 sobre os doze perfis sintéticos.

4.1. Desempenho operacional

As trinta e seis execuções consumiram setenta e duas requisições ao modelo, duas por execução, e todas concluíram sem falha. Os tempos apurados pela instrumentação do servidor constam da Tabela 1.

Tabela 1. Tempos de processamento por etapa, em trinta e seis execuções

Etapa	Média (s)	Desvio-padrão (s)	Mínimo (s)	Máximo (s)
Extração do documento	13,18	4,24	7,91	24,84
Diagnóstico	11,22	2,70	6,69	17,32
Ciclo completo	24,40	6,36	14,72	39,75

Fonte: dados da pesquisa.

A extração é a etapa mais lenta e a mais dispersa, com desvio-padrão de 4,24 segundos contra 2,70 do diagnóstico. A dispersão não é aleatória: a duração da extração acompanha o volume de texto do documento, com coeficiente de correlação de 0,75 entre o número de caracteres do título e do resumo e o tempo medido. Os perfis mais extensos demandaram entre 16 e 20 segundos em média, e os mais enxutos ficaram abaixo de 11. O ciclo completo manteve-se inferior a quarenta segundos em todas as execuções.

O custo apurado pela diferença no painel de faturamento do fornecedor situou-se em torno de três centavos de real por requisição, cerca de dois reais para o conjunto aqui relatado. O valor é aferido por diferença de saldo, e não por contagem de *tokens*, razão pela qual se reporta como ordem de grandeza, sem constituir restrição ao desenho.

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.

4.2. Reprodutibilidade do diagnóstico

O critério de estabilidade fixado na subseção 3.4 é o desvio-padrão da pontuação inferior a cinco pontos, na escala de zero a cem, entre execuções repetidas de um mesmo documento. O limiar é decisão de projeto, e não derivação de distribuição amostral: corresponde a 5% da escala e à ordem de grandeza do peso de um critério isolado, de que é exemplo a localização da experiência, com cinco pontos, de modo que oscilações abaixo dele não chegam a inverter a leitura de um critério. A Tabela 2 apresenta, para cada perfil, a média, o desvio-padrão, a amplitude e as três pontuações obtidas.

Tabela 2. Pontuação por perfil em três execuções independentes

Perfil	Média	Desvio-padrão	Amplitude	Execuções (pontos)
P01	89,4	4,56	7,9	94,7 · 86,8 · 86,8
P02	69,3	3,06	5,3	71,1 · 65,8 · 71,1
P03	52,6	0,00	0,0	52,6 · 52,6 · 52,6
P04	72,8	3,00	5,2	76,3 · 71,1 · 71,1
P05	52,6	0,00	0,0	52,6 · 52,6 · 52,6
P06	89,5	0,00	0,0	89,5 · 89,5 · 89,5
P07	50,9	3,00	5,2	52,6 · 47,4 · 52,6
P08	83,3	1,50	2,6	84,2 · 81,6 · 84,2
P09	79,8	3,06	5,3	81,6 · 81,6 · 76,3
P10	90,3	4,02	7,9	86,8 · 94,7 · 89,5
P11	47,4	0,00	0,0	47,4 · 47,4 · 47,4
P12	79,0	6,94	13,1	81,6 · 71,1 · 84,2

Fonte: dados da pesquisa.

Onze dos doze perfis atendem ao critério, e o desvio-padrão médio da amostra é de 2,43 pontos. Quatro perfis (P03, P05, P06 e P11) reproduziram pontuação idêntica nas três execuções. O P12 não atende: desvio-padrão de 6,94 pontos e amplitude de 13,1, oscilando entre 71,1 e 84,2. O resultado é relatado como obtido, sem exclusão do caso discrepante, porque a estabilidade é propriedade que se afirma do sistema. Três execuções por perfil, contudo, produzem estimativa de dispersão de baixa precisão: com n igual a três, o desvio-padrão amostral tem intervalo de confiança largo, de modo que os valores aqui reportados delimitam ordens de grandeza e a classificação de um perfil isolado deve ser lida como indicativa. A tabela mostra ainda que a pontuação cresce de 47,4 pontos no perfil mais lacunar a 90,3 no mais completo, o que indica que a escala discrimina documentos de qualidade distinta.

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 variação tampouco se distribui uniformemente pelas seis categorias, e o modo como se distribui reproduz a distinção adotada no desenho da pesquisa. As duas categorias integralmente verificáveis por inspeção, dados pessoais e idiomas, não variaram em nenhum dos doze perfis; as duas de natureza mista, experiência e formação, variaram em um perfil cada; e as duas que dependem de juízo do modelo concentram a instabilidade, com o título profissional variando em dois perfis e o resumo em seis, metade da amostra. A separação entre critérios factuais e critérios julgados, proposta na seção 3 como recurso de método, encontra correspondência empírica.

A variabilidade não se origina exclusivamente na etapa de diagnóstico. Dois perfis divergiram na própria extração: no P01, o comprimento do resumo extraído foi de 1.112 caracteres em duas execuções e de 1.346 na terceira; no P09, o número de formações identificadas foi de uma em duas execuções e de duas na terceira, erro de leitura conferido manualmente, uma vez que o documento declara uma única formação.

Um caso particular delimita com precisão a origem do fenômeno. Na primeira e na terceira execução do P01, a impressão digital da extração foi idêntica em todos os campos registrados, e ainda assim a categoria resumo recebeu 10,0 pontos na primeira e 2,5 na terceira. Ressalve-se que a impressão digital é representação reduzida do texto extraído e não assegura identidade byte a byte; ainda assim, a coincidência de todos os campos aferidos, combinada à divergência da nota, é compatível com o não determinismo relatado por Ouyang *et al.* (2024) e constitui a evidência mais direta, entre as aqui obtidas, de que a instabilidade reside no julgamento do modelo, e não apenas na leitura do documento.

4.3. Confiabilidade do acionamento de ferramenta

O instrumento submete o agente conversacional a roteiro fixo de cinco turnos, repetido dez vezes para cada um de dois valores de temperatura, 0,5 e 0,7, totalizando cem requisições, concluídas sem falha. Dois dos cinco turnos solicitam explicitamente alteração do perfil; os demais declaram intenção ou encerram a conversa. O acionamento da ferramenta é observável de forma direta na resposta da interface: a presença de alterações no campo de retorno destinado ao perfil indica que a função foi invocada, e a sua ausência, que o agente respondeu apenas em texto.

Considerados os quarenta turnos que solicitam alteração, a ferramenta foi acionada em catorze, taxa de 35%: 30% a 0,5 e 40% a 0,7. Em cerca de dois terços das solicitações, portanto, o agente respondeu que faria a alteração sem efetuá-la. O valor agregado resulta da média de dois comportamentos opostos, e a Tabela 3 os separa.

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.

Tabela 3. Acionamento da ferramenta de edição por turno do roteiro, em dez diálogos por temperatura

Turno	Conteúdo do turno	Solicita edição	T 0,5	T 0,7
1	Declara intenção de reposicionamento	Não	**8/10**	**5/10**
2	Declara objetivo de carreira	Não	**7/10**	**5/10**
3	Pede alteração do título, fornecendo o texto	Sim	**0/10**	**0/10**
4	Pede reescrita do resumo e gravação no perfil	Sim	**6/10**	**8/10**
5	Encerra a conversa	Não	**1/10**	**0/10**

Fonte: dados da pesquisa.

O terceiro turno não acionou a ferramenta em nenhuma das vinte repetições, nas duas temperaturas. A regularidade afasta a hipótese de variabilidade do modelo e sugere comportamento determinado por característica do próprio turno. A explicação mais imediata, a de que o agente consideraria a tarefa cumprida por ter alterado o título antes sem que isso lhe fosse pedido, não se sustenta: em cinco diálogos o título havia sido alterado antes do terceiro turno e em quinze não, e em nenhum dos dois grupos houve acionamento.

A comparação entre os dois turnos que solicitam alteração revela que diferem em dois aspectos ao mesmo tempo: o terceiro fornece o texto pronto e instrui a atualização do título, ao passo que o quarto pede conteúdo gerado pelo próprio agente e acrescenta ordem explícita de gravação. Os dados não permitem atribuir a diferença a um dos dois fatores, limitação de desenho sanável pela inclusão de dois turnos que variem um fator de cada vez.

Observou-se, em sentido inverso, acionamento em turnos que não solicitam alteração alguma. Dos sessenta turnos dessa natureza, vinte e seis produziram edição no perfil, taxa de 43% (53% a 0,5 e 33% a 0,7): a simples declaração de intenção, no primeiro turno, bastou para que o agente reescrevesse os campos de cargo pretendido. O comportamento é o inverso do desejável nas duas direções: a ferramenta é acionada em 43% das vezes em que não deveria e deixa de ser acionada em 65% das vezes em que deveria.

Quanto à temperatura, o experimento não detectou efeito. A diferença entre as duas condições corresponde a dois acionamentos nos turnos que solicitam alteração e a seis nos demais, sobre bases de vinte e de trinta observações; cada acionamento isolado desloca a taxa em cinco e em 3,3 pontos percentuais, de modo que a resolução da medida é da ordem do efeito observado. Os tempos foram equivalentes: 3,22 segundos a 0,5 e 3,17 a 0,7. Afirma-se, pois, que não se identificou efeito da temperatura, e não que tal efeito inexista.

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.

4.4. Sensibilidade a marcadores demográficos

O instrumento compara a pontuação atribuída a pares de perfis idênticos em conteúdo técnico e distintos apenas quanto a um marcador: o ano de conclusão do curso, o gênero sugerido pelo prenome ou o prestígio da instituição. Quatro perfis-base (P02, P04, P09 e P12) deram origem a onze variantes, submetidas, com as bases, a três execuções cada, no mesmo dia e em sessão contínua, de modo que eventual alteração no modelo do fornecedor não se confunda com efeito de marcador. O P09 não recebeu variante de ano de conclusão, por ser perfil da área médica. Foram realizadas quarenta e cinco execuções válidas.

A subseção 3.4 enuncia o critério: diferença inferior à dispersão observada entre execuções repetidas. Ele exigiu operacionalização, feita após a coleta e aqui declarada como tal, porque três células produziram desvio-padrão igual a zero e, tomada a dispersão par a par, qualquer diferença contra elas excederia o critério, artefato da medida e não propriedade do sistema. Adotou-se referência única: o desvio-padrão agrupado das quinze células, de 4,36 pontos. A Tabela 4 apresenta as comparações.

Tabela 4. Diferença de pontuação entre perfis-base e variantes demográficas

Variante	Marcador alterado	Base	Variante	Diferença
P02F	Ano de conclusão adiantado	71,10	65,80	-5,30
P02G	Prenome de outro gênero	71,10	67,53	-3,57
P02I	Instituição de menor prestígio	71,10	66,70	-4,40
P04F	Ano de conclusão adiantado	72,83	72,83	0,00
P04G	Prenome de outro gênero	72,83	72,83	0,00
P04I	Instituição de menor prestígio	72,83	76,33	+3,50
P09G	Prenome de outro gênero	79,83	83,33	+3,50
P09I	Instituição de menor prestígio	79,83	79,83	0,00
P12F	Ano de conclusão adiantado	81,60	81,57	-0,03
P12G	Prenome de outro gênero	81,60	85,07	+3,47
P12I	Instituição de menor prestígio	81,60	82,47	+0,87

Fonte: dados da pesquisa.

A média das onze diferenças é de -0,18 ponto e a média dos valores absolutos, de 2,24 pontos, ambas inferiores à dispersão de referência, e apenas duas comparações excedem o critério, ambas no mesmo perfil-base. Agrupadas por marcador, as diferenças médias são de -1,78 ponto para o ano de conclusão, +0,85 para o prenome e -0,01 para a instituição; dentro de

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.

cada marcador os sinais se contradizem, sem concentração em um dos lados, de modo que não se identifica direção sistemática. Os dados não sustentam a existência de penalização ou de favorecimento associado aos três marcadores examinados.

O P02 constitui a exceção e merece registro. Suas três variantes pontuaram abaixo da base, em -5,30, -3,57 e -4,40 pontos, e é o único perfil-base em que a direção se repete nos três marcadores. O caso mais nítido é o do ano de conclusão: a base obteve 71,1 pontos nas três execuções e a variante, 65,8 nas três, com dispersão nula de ambos os lados. O achado é relatado sem que dele se extraia generalização: o mesmo marcador produziu diferença de -0,03 ponto no P12 e nenhuma no P04.

Três execuções por célula delimitam o que este instrumento consegue detectar. O desenho identifica diferenças da ordem de cinco pontos; um viés sistemático de dois pontos permaneceria invisível. A instabilidade da própria estimativa de dispersão ilustra o limite: o P02 apresentou desvio-padrão de 3,06 pontos no lote da manhã e de 0,00 no da tarde, no mesmo dia e sobre o mesmo documento. Registre-se ainda que, em duas das três execuções da variante P09l, o extrator identificou duas formações acadêmicas onde o documento declara uma, repetindo comportamento já observado no perfil-base. Afirma-se, pois, que não foi detectado efeito sistemático dos marcadores examinados, e não que tal efeito inexista.

4.5. Concordância com a verificação independente

O verificador descrito entre os instrumentos de coleta foi aplicado aos doze perfis sintéticos e respondeu, por inspeção direta do documento e sem consultar a plataforma, aos treze critérios factualmente verificáveis; a plataforma processou os mesmos documentos e produziu as suas próprias respostas. A comparação das duas leituras, par a par, mede a concordância entre elas, com denominador explícito: das cento e cinquenta e seis verificações possíveis, treze critérios em doze perfis, quatro foram declaradas indeterminadas pelo verificador, taxa de 2,6%, restando cento e cinquenta e duas determinadas; destas, cento e quarenta coincidem, o que corresponde a 92,1% de concordância. A distribuição importa mais que o valor agregado: onze dos treze critérios apresentaram concordância integral e toda a discordância concentra-se em dois, de naturezas distintas, como mostra a Tabela 5.

A primeira concentração de divergências está na localização da experiência profissional, em que a plataforma respondeu negativamente nos doze perfis, sem exceção, enquanto o verificador identificou a informação em oito deles. A inspeção do código esclarece o comportamento: a estrutura de dados que recebe cada experiência, definida no modelo WorkExperience, comporta quatro campos (empresa, cargo, período e descrição) e não dispõe

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.

de campo para localização. O documento traz a informação, a extração não tem onde armazená-la, e o modelo, ao julgar a estrutura extraída, responde com exatidão que a localização não consta: não há erro de julgamento, mas informação descartada antes da etapa de julgamento. A consequência é visível na tela de diagnóstico, pois o sistema recomenda que o usuário informe a localização de cada experiência, informação que o documento já continha, e a categoria de experiência profissional aparece pontuada em quinze de vinte. O defeito subtrai pontos e ainda produz recomendação improcedente.

Tabela 5. Concordância entre a verificação independente e o diagnóstico, por critério

Critério	Concordância	Observação
Nome	**12/12**	—
Sobrenome	**12/12**	—
Localização	**12/12**	—
Endereço eletrônico	**12/12**	—
Telefone	**12/12**	—
Localização da experiência	**4/12**	Critério inatingível
Datas da experiência	**12/12**	—
Instituição de formação	**12/12**	—
Grau obtido	**8/12**	Ambiguidade no documento
Área de formação	**8/8**	**4 indeterminados**
Datas da formação	**12/12**	—
Quantidade de idiomas	**12/12**	—
Proficiência em idiomas	**12/12**	—
Total	**140/152**	**4 indeterminados; 92,1% de concordância**

Fonte: dados da pesquisa.

A consequência extrapola o critério. Os cinco pontos atribuídos à localização da experiência são inalcançáveis para qualquer perfil, de modo que o máximo bruto efetivamente obtenível é de noventa pontos, e não os noventa e cinco adotados como divisor na normalização descrita na subseção 3.2. O teto real da escala é, por conseguinte, de 94,7 pontos, valor que os perfis P01 e P10 alcançaram em suas melhores execuções e que nenhum perfil superou. Nenhum usuário da plataforma, na versão avaliada, pode obter pontuação igual a cem.

A segunda concentração está no grau obtido, com quatro divergências, nos perfis P03, P05, P07 e P11. Nesses documentos, a seção de formação registra apenas a denominação do

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.

curso, como em “Engenharia de Computação”, sem as expressões bacharelado, licenciatura ou tecnólogo. O verificador aplica regra estrita e exige a expressão do grau; a plataforma adota leitura permissiva e toma a denominação do curso como grau declarado. Os dois comportamentos são defensáveis diante de documento ambíguo, e são exatamente esses os quatro perfis em que o próprio verificador se declarou indeterminado no critério vizinho, o de área de formação. Registra-se a divergência como ambiguidade do documento.

Separadas as duas concentrações, o quadro se esclarece. Excluída a localização da experiência, por inatingível, a concordância sobe para 97,1%; excluídos também os quatro casos ambíguos de grau, as cento e trinta e seis verificações restantes coincidem integralmente. A leitura que se sustenta é que o diagnóstico identifica corretamente os atributos factuais do documento sempre que esses atributos chegam até ele e estão declarados sem ambiguidade: as falhas situam-se na extração e na especificação dos critérios, não no julgamento do modelo. Reporta-se como resultado do instrumento a concordância de 92,1% apurada sobre o conjunto completo, com a ressalva de que o padrão de referência é a leitura do verificador, ela própria sujeita a erro, ainda que validada contra o gabarito nos termos da metodologia.

Frente ao que a literatura consultada registra, o trabalho apresenta vantagens e desvantagens. Os sistemas ali descritos orientam-se ao recrutamento e à triagem pelo lado das empresas, a exemplo do sistema Grace (Medeiros *et al.*, 2023), e a plataforma aqui avaliada inverte a orientação, servindo ao profissional que examina o próprio documento. No plano do método, as medições foram repetidas e acompanhadas de medidas de dispersão, procedimento que Ouyang *et al.* (2024) reclamam de uma área que reporta execução única, e a avaliação do sistema foi feita por instrumento externo a ele, arranjo que Chang *et al.* (2024) recomendam para tarefas sem resposta única correta. As desvantagens são duas: o sistema depende de modelo de terceiro, cujo não determinismo apenas administra, e a concordância foi aferida sobre perfis construídos, não sobre documentos reais.

5. CONSIDERAÇÕES FINAIS

A questão que abriu esta pesquisa indagava de que maneira uma plataforma apoiada em Modelos de Linguagem de Grande Escala pode diagnosticar o perfil profissional e que evidências sustentam que esse diagnóstico é acurado e reprodutível. Esta seção responde a ela, às três subquestões dela desdobradas e às hipóteses enunciadas na introdução. A resposta que os dados autorizam é mais estreita do que a pergunta: a plataforma diagnostica com acerto e falha na execução.

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 primeira subquestão perguntava em que medida o diagnóstico identifica corretamente o que falta no perfil quando confrontado com verificação independente. O verificador aplicado aos doze perfis sintéticos coincidiu com a plataforma em cento e quarenta das cento e cinquenta e duas verificações determinadas, 92,1% de concordância sobre critérios factuais de perfis construídos: confirma-se a primeira cláusula da hipótese principal nesse alcance, sem autorizar afirmação sobre a qualidade global do diagnóstico nem sobre documentos reais. A distribuição importa mais que o número, pois onze dos treze critérios coincidiram integralmente e toda a divergência concentrou-se em dois: a ambiguidade do grau declarado e o defeito de modelagem da localização da experiência, que torna cinco pontos inalcançáveis e reduz o teto efetivo da escala a 94,7 pontos, conforme a subseção 4.5. O achado só apareceu porque a verificação foi feita por fora do sistema.

A segunda subquestão tratava do grau de reprodutibilidade. Onze dos doze perfis atenderam ao critério de desvio-padrão inferior a cinco pontos, com média de 2,43, e a variação não se distribuiu ao acaso: as categorias verificáveis por inspeção não variaram em nenhuma das trinta e seis execuções, ao passo que o resumo variou em metade dos perfis, de modo que a divisão entre critério factual e critério julgado encontrou correspondência empírica. Um perfil ficou fora do critério, com amplitude de 13,1 pontos. Daí a recomendação de projeto: a pontuação discrimina documentos de qualidade distinta, mas não deve ser apresentada como medida exata, convindo exibi-la acompanhada da sua dispersão.

A terceira subquestão perguntava com que confiabilidade o agente executa as alterações solicitadas, e a resposta constitui o achado mais consequente deste trabalho: a hipótese falava em automatizar o diagnóstico, e é a automação da execução que não se sustenta. O agente declara que fará a alteração e não a faz na maior parte das vezes em que ela lhe é pedida, e altera o perfil em parte das vezes em que nada lhe foi pedido, nas proporções apuradas na subseção 4.3. A literatura descreve o mecanismo pelo qual um modelo decide invocar uma ferramenta externa (Schick *et al.*, 2023; Yao *et al.*, 2023); o que esta pesquisa acrescenta é a medida da sua confiabilidade em operação: a decisão de agir é saída probabilística do modelo, e não garantia de execução.

A hipótese secundária sustentava que vincular os critérios a competências técnicas e experiências declaradas tornaria a pontuação insensível a marcadores demográficos. O teste não a refutou: a média das onze diferenças ficou 0,18 ponto abaixo de zero, os sinais se contradizem dentro de cada marcador e apenas duas comparações excedem a referência de dispersão. Tampouco a confirmou em sentido forte, pelo limite de resolução declarado na

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.

subseção 4.4. A formulação que os dados autorizam é a de que não foi detectado efeito sistemático dos três marcadores, e não a de que tal efeito inexista.

Do trabalho decorrem aplicações práticas de duas naturezas. A primeira é a própria plataforma, em operação e acessível pela internet, que produz diagnóstico estruturado de um perfil em menos de quarenta segundos e a custo da ordem de três centavos de real por requisição. A segunda é de método: verificar por fora do sistema os critérios que ele afirma aplicar é procedimento reaproveitável por qualquer sistema de pontuação apoiado em modelos de linguagem, e foi o que revelou aqui um defeito que nenhuma medida interna teria exposto. Acrescente-se a recomendação de que todo agente que se proponha a alterar dados do usuário confirme a alteração efetivada, e não a prometida.

As limitações encontradas desenham a agenda de pesquisas futuras. Cabe corrigir a estrutura de dados da experiência profissional, de modo a acomodar a localização, e refazer a aferição de concordância sobre a escala assim restabelecida. O roteiro que mediu o acionamento de ferramenta carece de dois turnos que variem um fator de cada vez. A concordância aferida sobre perfis construídos pede replicação sobre documentos reais, com tratamento adequado de dados pessoais e apreciação ética prévia. As rotinas de normalização, hoje verificadas de forma manual, devem integrar a suíte automatizada. Resta, ainda, o que este desenho não comportou: grupo de comparação, medição do perfil em segundo momento e aferição, junto a profissionais, da utilidade percebida das sugestões produzidas.

O problema que motivou a investigação era de acesso. Quem pretende revisar a própria apresentação desconhece os critérios pelos quais será lido, carece de distanciamento sobre o próprio percurso e dispõe de pouco tempo, ao passo que a orientação especializada é cara. Um diagnóstico automatizado, acurado quanto ao que aponta e obtido em dezenas de segundos por alguns centavos, desloca essa barreira na medida em que os resultados permitem: o sistema diz corretamente o que falta e ainda não executa com confiabilidade o que promete alterar. Reconhecer essa distância, e medi-la, é a contribuição que este trabalho oferece a quem venha a construir ferramentas da mesma espécie.

DECLARAÇÕES

Conflito de interesses. O desenvolvimento da plataforma e a escolha do tema decorrem de oportunidade oferecida ao primeiro autor por empresa de consultoria em carreira profissional, que detém participação societária no sistema aqui avaliado e pretende utilizá-lo e comercializá-lo. A empresa não participou do desenho dos instrumentos, da coleta dos dados nem da análise dos resultados. O primeiro autor declara interesse econômico indireto no objeto avaliado,

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.

circunstância que orientou duas decisões de método: a concordância do diagnóstico é aferida por verificador independente, e não pelo próprio sistema, e o alcance das afirmações é delimitado ao final da seção de metodologia. Os achados desfavoráveis são relatados como obtidos, entre eles o defeito de modelagem da subseção 4.5 e a baixa confiabilidade do acionamento de ferramenta da subseção 4.3.

Financiamento. A pesquisa não recebeu financiamento de agência de fomento. Os custos de operação durante a coleta, entre eles o plano pago do fornecedor do modelo de linguagem e a hospedagem da aplicação, foram arcados pelo primeiro autor.

Uso de inteligência artificial generativa. Declara-se o emprego do assistente Claude, da Anthropic, em duas frentes: na construção do protótipo, participou da escrita e da revisão do código-fonte; na elaboração deste artigo, participou da redação de seções do texto, da construção das figuras, da escrita dos programas de coleta e do verificador, e do tratamento dos dados. Não foi empregado na concepção da pesquisa, nas decisões de método nem na execução da coleta, e toda saída produzida com o seu apoio foi conferida pelos autores contra os dados brutos e a bibliografia. A ferramenta não figura como autora, e a responsabilidade por todo o conteúdo é dos autores.

REFERÊNCIAS

AKHMETOV, I. *et al*. Highly language-independent word lemmatization using a machine-learning classifier. **Computación y Sistemas**, v. 24, n. 3, p. 1353-1364, 2020. DOI: 10.13053/CyS-24-3-3775. Disponível em: https://www.cys.cic.ipn.mx/index.php/CyS/article/view/3775. Acesso em: 4 nov. 2026.

BARBOSA, J. *et al*. Introdução ao processamento de linguagem natural usando python. *In*: ESCOLA REGIONAL DE INFORMÁTICA DO PIAUÍ, 3., 2017, Teresina. **Anais** [...]. Teresina: [s. n.], 2017. v. 1, p. 336-360.

CARRARO, Fabrício. **Inteligência artificial e ChatGPT**: da revolução dos modelos de IA generativa à engenharia de prompt. Edição ampliada e atualizada. São Paulo: Casa do Código, 2025.

CHANG, Yupeng *et al*. A survey on evaluation of large language models. **ACM Transactions on Intelligent Systems and Technology**, [s. l.], v. 15, n. 3, p. 1-45, 2024. DOI: 10.1145/3641289. Disponível em: https://dl.acm.org/doi/10.1145/3641289. Acesso em: 23 ago. 2026.

COPELAND, B. J. Artificial intelligence. **Encyclopedia Britannica**, 2021. Disponível em: https://www.britannica.com/technology/artificial-intelligence. Acesso em: 6 abr. 2026.

FIDALGO, Luiza Barreto Braga. Game over para os trabalhadores em plataformas digitais: o adoecimento na lógica do jogo. *In*: ARAUJO, Adriane Reis de *et al*. (coord.). **O uso de dados pessoais e inteligência artificial na relação de trabalho**: proteção, discriminação, violência e assédio digital. Brasília: Ministério Público do Trabalho, 2022. p. 353-376.

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.

JACKSON, P. C. **Introduction to artificial intelligence**. New York: Dover Publications, 1985.

LIMA, Francisco Gérson Marques de; MORAIS JÚNIOR, Ricardo Antônio Maia de. Decisões automatizadas nas relações de emprego: art. 20 da LGPD ante o silêncio da CLT. *In*: ARAUJO, Adriane Reis de *et al*. (coord.). **O uso de dados pessoais e inteligência artificial na relação de trabalho**: proteção, discriminação, violência e assédio digital. Brasília: Ministério Público do Trabalho, 2022.

MEDEIROS, Ana Clara B. *et al*. Grace: sistema de recomendação de currículos com inteligência artificial. *In*: SIMPÓSIO BRASILEIRO DE BANCO DE DADOS, 38., 2023, Belo Horizonte. **Anais Estendidos** [...]. Porto Alegre: SBC, 2023. p. 8-14. DOI: 10.5753/sbbd_estendido.2023.232712. Disponível em: https://sol.sbc.org.br/index.php/sbbd_estendido/article/view/25607. Acesso em: 4 nov. 2026.

OUYANG, Shuyin *et al*. An empirical study of the non-determinism of ChatGPT in code generation. **ACM Transactions on Software Engineering and Methodology**, [s. l.], 2024. DOI: 10.1145/3697010. Disponível em: https://dl.acm.org/doi/10.1145/3697010. Acesso em: 23 ago. 2026.

PANTOJA, Adilson Carvalho; LEÃO, Semírames de Cássia Lopes. Discriminação algorítmica no mercado de trabalho: o viés preditivo na interpretação de dados. *In*: ARAUJO, Adriane Reis de *et al*. (coord.). **O uso de dados pessoais e inteligência artificial na relação de trabalho**: proteção, discriminação, violência e assédio digital. Brasília: Ministério Público do Trabalho, 2022. p. 33-60.

PEREIRA, Silvio. **Processamento de linguagem natural**. São Paulo: IME-USP, 2018. Disponível em: https://www.ime.usp.br/~slago/IA-pln.pdf. Acesso em: 6 abr. 2026.

RANGEL, M. M. **Categorização automática de conjuntos de dados de portais de dados abertos utilizando aprendizado de máquina supervisionado**. 2019. Trabalho de Conclusão de Curso (Bacharelado em Sistemas de Informação) - Instituto de Computação, Universidade Federal Fluminense, Niterói, 2019. Disponível em: https://app.uff.br/riuff/handle/1/13072. Acesso em: 4 nov. 2026.

RUSSELL, Stuart; NORVIG, Peter. **Inteligência artificial**. Tradução da 3. ed. Rio de Janeiro: Elsevier, 2013.

SCHICK, Timo *et al*. Toolformer: language models can teach themselves to use tools. *In*: ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS, 36., 2023, New Orleans. **Proceedings** [...]. [S. l.]: Curran Associates, Inc., 2023. Disponível em: https://papers.nips.cc/paper_files/paper/2023/hash/d842425e4bf79ba039352da0f658a906-Abstract-Conference.html. Acesso em: 4 nov. 2026.

TUNSTALL, Lewis; VON WERRA, Leandro; WOLF, Thomas. **Natural language processing with transformers**. Revised edition. Sebastopol: O'Reilly Media, 2022.

VASWANI, Ashish *et al*. Attention is all you need. *In*: ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS, 30., 2017, Long Beach. **Proceedings** [...]. [S. l.]: Curran Associates, Inc., 2017. p. 5998-6008. 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://papers.neurips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html. Acesso em: 4 nov. 2026.

YAO, Shunyu *et al*. ReAct: synergizing reasoning and acting in language models. *In*: INTERNATIONAL CONFERENCE ON LEARNING REPRESENTATIONS, 11., 2023, Kigali. **Proceedings** [...]. [S. l.: s. n.], 2023. Disponível em: https://iclr.cc/virtual/2023/poster/11003. Acesso em: 4 nov. 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.