

**UNIVERSIDADE DE CAXIAS DO SUL
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA
PROGRAMA DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO – PPGA
MESTRADO EM ADMINISTRAÇÃO**

**ANÁLISE DE UM SISTEMA BASEADO EM INTELIGÊNCIA ARTIFICIAL PARA
IDENTIFICAÇÃO DE AÇÕES INSTITUCIONAIS**

LEONARDO ALVARENGA PEREIRA

**CAXIAS DO SUL
2026**

LEONARDO ALVARENGA PEREIRA

**ANÁLISE DE UM SISTEMA BASEADO EM INTELIGÊNCIA ARTIFICIAL PARA
IDENTIFICAÇÃO DE AÇÕES INSTITUCIONAIS**

Dissertação de Mestrado submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Administração da Universidade de Caxias do Sul, como parte dos requisitos necessários para obtenção do título de Mestre (a) em Administração. Área de Concentração: Estratégia e Inovação. Linha de Pesquisa: Estratégia e Operações.

Orientador: Prof. Dr. Roberto Birch Gonçalves.

**CAXIAS DO SUL
2026**

Dados Internacionais de Catalogação na Publicação (CIP)
Universidade de Caxias do Sul
Sistema de Bibliotecas UCS - Processamento Técnico

P436a Pereira, Leonardo Alvarenga

Análise de um sistema baseado em inteligência artificial para
identificação de ações institucionais [recurso eletrônico] / Leonardo
Alvarenga Pereira. – 2026.

Dados eletrônicos.

Dissertação (Mestrado) - Universidade de Caxias do Sul, Programa de
Pós-Graduação em Administração, 2026.

Orientação: Roberto Birch Gonçalves.

Modo de acesso: World Wide Web

Disponível em: <https://repositorio.ucs.br>

1. Inteligência artificial. 2. Administração pública. 3. Institutos federais
de educação, ciência e tecnologia. 4. Processamento de linguagem natural
(Computação). 5. Avaliação educacional. I. Gonçalves, Roberto Birch,
orient. II. Título.

CDU 2. ed.: 004.8

Catalogação na fonte elaborada pela(o) bibliotecária(o)
Ana Guimarães Pereira - CRB 10/1460

LEONARDO ALVARENGA PEREIRA

**ANÁLISE DE UM SISTEMA BASEADO EM INTELIGÊNCIA ARTIFICIAL PARA
IDENTIFICAÇÃO DE AÇÕES INSTITUCIONAIS**

Dissertação de Mestrado submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Administração da Universidade de Caxias do Sul, como parte dos requisitos necessários para obtenção do título de Mestre (a) em Administração. Área de Concentração: Estratégia e Inovação. Linha de Pesquisa: Estratégia e Operações.

Aprovado (a) em: 26/09/2026.

Banca Examinadora

Prof. Dr. Alex Eckert
Universidade de Caxias do Sul – UCS

Prof. Dr. Gabriel Vidor
Universidade de Caxias do Sul – UCS

Dr. Leonardo da Silva Cezarini
Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul - IFRS

DEDICATÓRIA

Dedico este trabalho à minha esposa Ingrid Viñas Pereira, pela compreensão nos momentos ausentes e pelo apoio contínuo à minha formação.

AGRADECIMENTOS

Agradeço, primeiramente, a Deus, por me conceder saúde, força e perseverança para concluir mais esta etapa da minha vida.

À minha esposa, pela compreensão nos momentos de ausência, pelo apoio incondicional e por estar sempre ao meu lado, incentivando-me a seguir em frente.

Aos meus pais, por todo o amor, pelos valores que me ensinaram e por serem a base sobre a qual construí tudo o que sou.

Ao IFRS – Campus Bento Gonçalves, por apoiar a minha formação e por proporcionar as condições necessárias para a realização deste trabalho.

Ao meu orientador, Prof. Dr. Roberto Birch Gonçalves, que teve paciência para me manter no rumo certo.

Aos mestres do curso, por fornecerem todo o conhecimento e a ajuda que procurava.

E, finalmente, aos colegas de curso, pela troca de ideias e pelo encorajamento que foram decisivos, muito em particular nos momentos mais difíceis deste percurso.

O homem não teria alcançado o possível se, repetidas vezes, não tivesse tentado o impossível.

Max Weber

RESUMO

Este trabalho aborda a aplicação da inteligência artificial (IA) no contexto das instituições públicas, especificamente no Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul – Campus Bento Gonçalves. O contexto da pesquisa decorre da necessidade de otimizar o processo de autoavaliação institucional, atualmente conduzido pela comissão própria de avaliação (CPA), tarefa trabalhosa que demanda um tempo considerável até sua conclusão. O objetivo geral da dissertação é a elaboração de um sistema baseado em inteligência artificial para identificar as ações institucionais apontadas nas pesquisas de autoavaliação do IFRS. Para alcançar esse objetivo, a pesquisa compreendeu os principais aspectos da inteligência artificial, identificando a utilização de tecnologia no setor público e analisar a relação entre a IA e a tecnologia nesse contexto. Em termos de metodologia, o trabalho tem um enfoque exploratório, analisando relatórios publicados pela instituição em anos anteriores, utilizando o Design Science Research (DSR) para a criação e avaliação de um artefato, denominado Auto Report, que contribuisse para a evolução do assunto de forma prática. O sistema desenvolvido converte a distribuição das respostas de cada questão em comandos estruturados, submetidos a quatro modelos de inteligência artificial generativa (ChatGPT, DeepSeek, Gemini e Claude), dos quais se obtêm proposições de ações para cada dimensão avaliada. Os resultados, obtidos a partir da análise das autoavaliações da comunidade interna dos ciclos de 2023, 2024 e 2025, demonstraram que, nas condições analisadas, o sistema alcançou convergência entre 90,9% e 100% com os relatórios manuais nas questões em que a comissão registrou ações, além de ofertar proposições pertinentes nas questões em que a comissão se limitou ao diagnóstico. A geração automatizada dos apontamentos pelo sistema ocorre em poucos minutos, reduzindo significativamente o tempo de trabalho da comissão, que permanece responsável pela validação crítica das ações antes de sua incorporação ao relatório final. Conclui-se que a inteligência artificial contribuiu para o processo de autoavaliação ao reproduzir, complementar e aprimorar o trabalho da comissão, promovendo uma gestão mais eficiente e responsiva às necessidades da comunidade acadêmica e da sociedade, e fornecendo subsídios sobre a integração da inteligência artificial nas práticas de gestão pública.

Palavras-chave: Inteligência Artificial. IFRS. Sistema. Autoavaliação Institucional. Tecnologia.

ABSTRACT

This study addresses the application of artificial intelligence (AI) within public institutions, specifically at the Federal Institute of Education, Science and Technology of Rio Grande do Sul (IFRS) – Bento Gonçalves Campus. The research context stems from the need to optimize the institutional self-assessment process—currently conducted by the Internal Evaluation Commission (CPA)—which is a laborious task requiring considerable time to complete. The dissertation's general objective is to develop an AI-based system to identify institutional actions highlighted in IFRS self-assessment surveys. To achieve this, the research examined key aspects of artificial intelligence, identified technology usage in the public sector, and analyzed the relationship between AI and technology in this context. Methodologically, the study adopts an exploratory approach, analyzing reports published by the institution in previous years and employing Design Science Research (DSR) to create and evaluate an artifact named "Auto Report," designed to advance the subject matter in a practical way. The developed system converts the response distribution for each question into structured commands submitted to four generative AI models (ChatGPT, DeepSeek, Gemini, and Claude), yielding proposed actions for each evaluated dimension. Results derived from analyzing internal community self-assessments from the 2023, 2024, and 2025 cycles demonstrated that, under the conditions analyzed, the system achieved 90.9% to 100% convergence with manual reports for questions where the commission had recorded actions, while also offering relevant proposals for questions where the commission had limited itself to diagnosis. The system generates these recommendations in a matter of minutes, significantly reducing the commission's workload, while the commission remains responsible for the critical validation of the actions before their incorporation into the final report. It is concluded that artificial intelligence contributed to the self-assessment process by replicating, complementing, and enhancing the commission's work, fostering management that is more efficient and responsive to the needs of the academic community and society, and providing insights into the integration of artificial intelligence into public management practices.

Keywords: Artificial Intelligence. IFRS. System. Institutional Self-Assessment. Technology.

LISTA DE FIGURAS

Figura 1 - Publicações de artigos com os termos "artificial intelligence" e "public service" na base de dados da Scopus	21
Figura 2 - Publicações de artigos com os termos "artificial intelligence" e "public service" na base de dados do Google Acadêmico	22
Figura 3 – “Árvore genealógica” das linguagens de programação	38
Figura 4 - Saídas da DSR proposta por Mason (2006)	56
Figura 5 - Processos da DSR em desenvolvimento de sistemas segundo Nunamaker, Chen e Purdin (1991)	57
Figura 6 - Saídas da DSR proposta na pesquisa	59
Figura 7 - Diagrama de Classes	62
Figura 8 - Tela inicial do sistema Auto Report	64
Figura 9 - Artefato proposto na pesquisa	74

LISTA DE QUADROS

Quadro 1 - Distribuição de perguntas por eixo	67
Quadro 2 - Decomposição do prompt de sugestões de ações	82
Quadro 3 - Características da integração de cada modelo de inteligência artificial ..	84
Quadro 4 - Evolução do prompt	88
Quadro 5 - Distribuição das classificações de correspondência por modelo e por ano	91
Quadro 6 - Convergência das ações do sistema conforme o recorte de análise	92
Quadro 7 - Convergência das ações por eixo do SINAES (2023-2025)	95
Quadro 8 - Correspondências por modelo no conjunto de três anos	99
Quadro 9 - Tempo de geração dos relatórios por modelo (em minutos)	100

LISTA DE ABREVIATURAS E SIGLAS

AGU	Advocacia-Geral da União
API	<i>Application Programming Interface</i>
CPA	Comissão Própria de Avaliação
DOM	<i>Document Object Model</i>
DSR	<i>Design Science Research</i>
GPS	<i>General Problem Solver</i>
GPT	<i>Generative Pre-trained Transformer</i>
HTML	<i>Hypertext Markup Language</i>
IA	Inteligência Artificial
IFRS	Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
PAIFRS	Programa de Autoavaliação do Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul
PHP	<i>Hypertext Preprocessor</i>
SINAES	Sistema Nacional de Avaliação da Educação Superior
SPA	Subcomissão Própria de Avaliação
STF	Supremo Tribunal Federal
TCU	Tribunal de Contas da União
TI	Tecnologia da Informação

SUMÁRIO

1	INTRODUÇÃO	14
1.1	TEMA E PROBLEMA	17
1.2	OBJETIVOS	18
1.2.1	Objetivo geral	18
1.2.2	Objetivos específicos.....	18
1.3	JUSTIFICATIVA DO ESTUDO	19
1.4	ADERÊNCIA À LINHA DE PESQUISA PPGA.....	23
2	REFERENCIAL TEÓRICO	25
2.1	INTELIGÊNCIA ARTIFICIAL (IA).....	25
2.1.1	História da Inteligência Artificial.....	28
2.1.2	Tipos de Inteligência Artificial	31
2.1.3	Linguagens de programação	33
2.1.4	Integração entre sistemas com IA generativa.....	39
2.1.5	ChatGPT.....	40
2.1.6	DeepSeek.....	42
2.1.7	Gemini	43
2.1.8	Claude	45
2.2	EFICIÊNCIA NO SETOR PÚBLICO	46
2.2.1	Tecnologias no setor público	49
2.3	RELAÇÃO ENTRE A IA E O SERVIÇO PÚBLICO.....	51
3	PROCEDIMENTOS METODOLÓGICOS.....	54
3.1	DELINEAMENTO DA PESQUISA	54
3.1.1	Sistema desenvolvido.....	59
3.2	CAMPO DE ESTUDO E AMOSTRA.....	65
3.3	PROCEDIMENTOS DE COLETA DE DADOS	69
3.4	PROCESSO DE ANÁLISE DE DADOS.....	71
3.5	ARTEFATO PROPOSTO	73
4	RESULTADOS E DISCUSSÃO	79
4.1	A CONSTRUÇÃO DOS PROMPTS UTILIZADOS NO SISTEMA DESENVOLVIDO	79

4.1.1	A representação dos dados quantitativos no comando	80
4.1.2	Os componentes do <i>prompt</i>	81
4.1.3	O nível de detalhe dos comandos: avaliação, instrumento e pergunta	83
4.1.4	Integração com as interfaces de programação dos modelos	84
4.1.5	Refinamento do <i>prompt</i>	86
4.2	A GERAÇÃO DOS RELATÓRIOS E A CARACTERIZAÇÃO DAS AÇÕES IDENTIFICADAS	89
4.3	ANÁLISE COMPARATIVA POR EIXO DOS SINAES	95
4.4	ANÁLISE COMPARATIVA ENTRE OS MODELOS DE INTELIGÊNCIA ARTIFICIAL	98
4.5	DISCUSSÃO DOS RESULTADOS E RESPOSTAS À QUESTÃO DE PESQUISA	102
5	CONSIDERAÇÕES FINAIS	106
5.1	IMPLICAÇÕES TEÓRICAS	109
5.2	IMPLICAÇÕES GERENCIAIS	110
5.3	LIMITAÇÕES DO ESTUDO E SUGESTÕES DE PESQUISAS FUTURAS	113
	REFERÊNCIAS.....	116
	APÊNDICE.....	122

1 INTRODUÇÃO

No dia 14 de maio de 2004 foi criada a Lei nº 10.861 que institui o Sistema Nacional de Avaliação da Educação Superior – SINAES, com o objetivo de avaliar a qualidade das instituições de educação superior, dos cursos e do desempenho dos estudantes no Brasil (BRASIL, 2004).

De acordo com esta Lei nº 10.861 a avaliação das instituições de educação superior visa identificar seu perfil e o significado de sua atuação através de suas atividades, cursos, programas, projetos e setores, considerando diversas dimensões institucionais, incluindo obrigatoriamente: a missão e o plano de desenvolvimento institucional; a política para o ensino, pesquisa, pós-graduação, extensão e suas formas de operacionalização, incluindo procedimentos de estímulo à produção acadêmica e concessão de bolsas de pesquisa e monitoria; a responsabilidade social da instituição, especialmente em relação à inclusão social, desenvolvimento econômico e social, defesa do meio ambiente, memória cultural, produção artística e patrimônio cultural; a comunicação com a sociedade; as políticas de pessoal, carreiras do corpo docente e técnico-administrativo, seu aperfeiçoamento, desenvolvimento profissional e condições de trabalho; a organização e gestão institucional, com ênfase no funcionamento e representatividade dos colegiados, independência e autonomia em relação à mantenedora, e participação da comunidade universitária nos processos decisórios; a infraestrutura física, especialmente para ensino e pesquisa, biblioteca, recursos de informação e comunicação; o planejamento e avaliação, com foco nos processos, resultados e eficácia da autoavaliação institucional; as políticas de atendimento aos estudantes; e a sustentabilidade financeira, considerando a importância social da continuidade dos compromissos na oferta da educação superior (BRASIL, 2004).

Para a avaliação das instituições de ensino superior, a lei de criação do SINAES estabeleceu o critério de que deverão ser empregados procedimentos e instrumentos diversos, incluindo a autoavaliação e a avaliação externa in loco, presencial ou virtual, e definiu que cada instituição de ensino superior, pública ou privada, deveria constituir uma Comissão Própria de Avaliação (CPA) no prazo de 60 (sessenta) dias a partir da publicação da lei, com a atribuição de conduzir os processos internos de avaliação da instituição, sistematizar e prestar as informações

solicitadas pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira - INEP (BRASIL, 2004).

No Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul – IFRS, criado em 29 de dezembro de 2008 através da Lei nº 11.892 (BRASIL, 2008), a primeira comissão própria de avaliação foi constituída em 15 de setembro de 2010, através da resolução nº 657, composta por uma comissão central na reitoria e uma subcomissão própria de avaliação (SPA) em cada campus (IFRS, 2010b).

De acordo com o Programa de Autoavaliação do Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul – PAIFRS, tem-se a seguinte definição sobre os instrumentos de autoavaliação:

Os instrumentos de autoavaliação que constituem o PAIFRS são disponibilizados no formato *online* para a comunidade interna, em programa desenvolvido pelo Departamento de TI da Instituição. Para a comunidade externa, o instrumento é disponibilizado no formato *off line* e enviado via correio eletrônico para as famílias dos alunos, bem como para instituições públicas e privadas parceiras ou mesmo em formato físico quando necessário (IFRS, 2012, p. 14).

Dentre os instrumentos que compõem a autoavaliação do IFRS tem-se: a Autoavaliação Institucional voltada para a participação de todos os segmentos da comunidade interna do IFRS (discentes, docentes e técnicos-administrativos), que envolve a avaliação das políticas de ensino, pesquisa e extensão, comunicação com a sociedade, infraestrutura, serviços, gestão e organização do IFRS; a Avaliação pela Comunidade Externa, que considera a visão de atores externos como famílias dos alunos e instituições parceiras; a Autoavaliação do Curso, que analisa a qualidade e relevância dos programas oferecidos; a Autoavaliação Discente, que obtém *feedback* dos alunos sobre suas experiências educacionais; a Avaliação Docente, que examina o desempenho e a prática pedagógica dos professores; e a Avaliação de Egressos, que investiga a trajetória profissional dos ex-alunos e sua percepção sobre a formação recebida (IFRS, 2010b).

Cada instrumento de autoavaliação aplicado de forma *online* possui um conjunto de questões objetivas com respostas dispostas em escala *Likert* de 5 pontos e uma questão aberta para comentários que o avaliador julgar importante (IFRS, 2010b).

A autoavaliação do IFRS contribui para o planejamento de gestão, contemplando indicadores que norteiam políticas para o ensino, a pesquisa, a

extensão e as respectivas normas de operacionalização, incluindo os procedimentos para estímulo à produção acadêmica, bolsas de pesquisa, monitoria e demais modalidades. Ela também auxilia na parte de infraestrutura física, como biblioteca, recursos de informação e comunicação. Em relação aos processos, o PAIFRS é fundamental para o apontamento de políticas de atendimento a estudantes e egressos e sustentabilidade financeira, considerando o significado social da continuidade dos compromissos na oferta da educação superior (IFRS, 2010b).

Os resultados da autoavaliação, produzidos anualmente, geram um relatório geral do IFRS e relatórios específicos para cada campus. As conclusões contidas nesses relatórios são discutidas com os responsáveis pela gestão do IFRS, servindo como base para o planejamento institucional do ano subsequente e apontamento de ações visando um aumento da qualidade da instituição. Além disso, esses resultados são apresentados e discutidos com toda a comunidade escolar e acadêmica (IFRS, 2010b).

A interpretação dos dados coletados através das pesquisas de autoavaliação institucional e a confecção dos relatórios, geral e específicos para cada campus, leva em média 4 meses atualmente. O processo é elaborado de forma manual, através do acesso ao sistema de autoavaliação disponibilizado pela reitoria do IFRS, que apresenta os resultados em forma de tabelas e gráficos. Tais relatórios são elaborados, em conjunto, pela CPA Central e CPA de cada campus seguindo um roteiro de elaboração definido pelo PAIFRS (IFRS, 2010b).

O roteiro de relatório foi construído de acordo com as 10 dimensões definidas pelo SINAES, orientando a construção dos relatórios com uma introdução sobre cada dimensão, uma explanação sobre o significado desta dimensão, a percepção dos avaliadores sobre a dimensão e um item de apontamento de ações de superação identificado a partir da interpretação dos dados pela CPA (IFRS, 2010b).

Diante disso, este estudo se propõe a desenvolver um sistema que utilize inteligência artificial com o objetivo de identificar as ações de superação apontadas nas pesquisas de autoavaliação institucional realizadas pela CPA do IFRS, diminuindo assim o tempo de elaboração dos relatórios apresentados pela comissão.

Esta dissertação está estruturada em cinco seções. A primeira seção, a introdução, abrange a delimitação do tema, a definição do problema de pesquisa, o objetivo geral, os objetivos específicos e a justificativa deste estudo. A segunda seção apresenta o referencial teórico relacionado aos temas abordados na dissertação,

fundamentando as proposições deste trabalho, com foco em Autoavaliação Institucional, Inteligência Artificial (IA) e a adoção de Sistemas de Informação com IA em instituições públicas. A terceira seção descreve os procedimentos metodológicos aplicados nesta pesquisa. Na quarta seção, os resultados são analisados e discutidos. A quinta e última seção contém as considerações finais, as limitações da pesquisa e sugestões para futuras pesquisas. Finalmente, são apresentadas as referências bibliográficas utilizadas na elaboração desta dissertação.

1.1 TEMA E PROBLEMA

Esta dissertação tem como tema o uso de Inteligência Artificial (IA) para auxiliar no apontamento de ações a partir de pesquisas de autoavaliação do IFRS – Campus Bento Gonçalves. De acordo com Russell e Norvig (2024) a Inteligência Artificial é definida de maneira diversa pelos pesquisadores, sendo caracterizada tanto pela fidelidade ao desempenho humano e pela capacidade de fazer a "coisa certa" (racionalidade), quanto pelo enfoque nos processos internos de pensamento e raciocínio ou no comportamento externo inteligente.

A evolução da IA ao longo da história pode ser sintetizada por alguns marcos como a definição dos fundamentos do campo com base na representação e raciocínio, o desenvolvimento de sistemas especialistas para resolver problemas do mundo real, a criação de técnicas de raciocínio probabilístico para lidar com a incerteza, e a integração do aprendizado profundo (redes neurais multicamadas) como um componente essencial da computação moderna (Russell; Norvig, 2024).

Apesar do ritmo mais lento do que o inicialmente previsto, a ascensão da inteligência artificial experimentou um renovado interesse a partir da década de 1990, impulsionado por investimentos de governos e empresas em seu desenvolvimento. Esse avanço foi favorecido pelo progresso tecnológico, pelo aprimoramento da eficiência da tecnologia da informação na captura e armazenamento de dados, pela maior acessibilidade a ferramentas tecnológicas e pela expansão dos serviços em nuvem (Rocha; Kissimoto, 2022).

A resolução de problemas utilizando sistemas de IA envolve um mecanismo de busca geral que reúne passos elementares de raciocínio para encontrar soluções completas, utilizando um conhecimento específico do domínio que permite passos de raciocínio mais amplos e facilita a resolução de casos típicos em áreas especializadas,

exigindo um conhecimento prévio da solução para problemas complexos (Russell; Norvig, 2024).

Utilizar o aprendizado de máquina pode ser uma das alternativas da utilização da IA para a resolução de problemas. Segundo Russell e Norvig (2024) o aprendizado de máquina funciona através da utilização de teorias existentes como base, fundamentando-se em teoremas rigorosos ou metodologia experimental consolidada, priorizando resultados experimentais e relevância para aplicações reais em vez de programação manual, afirmações filosóficas ou exemplos simplificados fabricados.

Segundo Li *et. al.* (2017), a inteligência artificial se destaca como um recurso fundamental no processo de tomada de decisões, em virtude de sua capacidade de aprendizado de máquina, favorecendo, assim, o desenvolvimento de novos modelos de negócios. Essa capacidade permite que a IA analise grandes volumes de dados e aprenda com eles, facilitando a tomada de decisões de forma mais precisa.

Dentro do IFRS – campus Bento Gonçalves, atualmente a comissão própria de avaliação (CPA) leva em torno de 4 meses para concluir os trabalhos de cada pesquisa de autoavaliação institucional. Desenvolver um sistema com IA utilizando aprendizado de máquina com base em relatórios produzidos em anos anteriores, pode agilizar o trabalho da comissão própria de avaliação na conclusão das ações apontadas nas pesquisas de autoavaliação institucional.

Assim, a partir do tema desta dissertação, a questão que norteia esta pesquisa é: Como a inteligência artificial contribui no apontamento de ações de superação identificadas na coleta de informações?

1.2 OBJETIVOS

1.2.1 Objetivo geral

O objetivo geral deste trabalho é elaborar um sistema baseado em inteligência artificial para identificar as ações institucionais apontadas nas pesquisas de autoavaliação do Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul – campus Bento Gonçalves.

1.2.2 Objetivos específicos

Os objetivos específicos deste trabalho, são:

- a) Compreender os principais aspectos da inteligência artificial;
- b) Identificar a utilização de tecnologia no setor público;
- c) Analisar a relação da inteligência artificial na utilização de tecnologia no setor público;
- d) Propor uma interface intuitiva para utilização do sistema desenvolvido.

1.3 JUSTIFICATIVA DO ESTUDO

Segundo Corrêa (1993), o papel social da organização pública é prestar serviços de qualidade, a responsabilidade de atender o público e economizar seus recursos sem esquecer de seus princípios fundamentais como legalidade, impessoalidade, moralidade, publicidade e eficiência.

Segundo Moraes (2024, p. 383), o princípio da eficiência tem a seguinte definição:

É aquele que impõe à Administração direta e indireta e a seus agentes a persecução do bem comum, por meio do exercício de suas competências de forma imparcial, neutra, transparente, participativa, eficaz, sem burocracia e sempre em busca da qualidade, primando pela adoção dos critérios legais e morais necessários para melhor utilização possível dos recursos públicos, de maneira a evitarem-se desperdícios e garantir-se uma maior rentabilidade social. Note-se que não se trata da consagração da tecnocracia, muito pelo contrário, o princípio da eficiência dirige-se para a razão e fim maior do Estado, a prestação dos serviços públicos essenciais à população, visando a adoção de todos os meios legais e morais possíveis para a satisfação do bem comum.

Da mesma forma, Matias-Pereira (2012, p. 17) define:

O princípio da eficiência que foi incluído pela emenda constitucional n° 19, de 1998, aos princípios constitucionais da administração pública (art. 37, caput, da constituição federal), impõe ao servidor público o encargo de realizar suas atribuições com maior agilidade, perfeição, qualidade e orientado pelas modernas técnicas administrativas.

De maneira geral, é compreendido que o estado, configurando-se como a esfera na qual o indivíduo exerce sua cidadania, tem o dever primordial de aprimorar a qualidade na oferta de seus serviços públicos. Tal atuação deve pautar-se pela eficiência, assegurando, assim, o cumprimento dos direitos dos cidadãos, conforme a Constituição Federal cita em seu artigo 37 (BRASIL, 1988):

Art. 37. A administração pública direta e indireta de qualquer dos Poderes da União, dos Estados, do Distrito Federal e dos Municípios obedecerá aos princípios de legalidade, impessoalidade, moralidade, publicidade e eficiência [...]

O Brasil caminhou no rumo da eficiência na administração pública quando, influenciado pelo gerencialismo britânico e pelos princípios do *New Public Management* (NPM) e da “*progressive governance*”, teve uma reforma gerencial formalizada pelo Plano Diretor de Reforma do Aparelho do Estado em 1995. Esse plano visava criar uma gestão pública chamada “administração gerencial”, baseada em flexibilidade, resultados, foco no cliente e controle social (BRASIL, 1995).

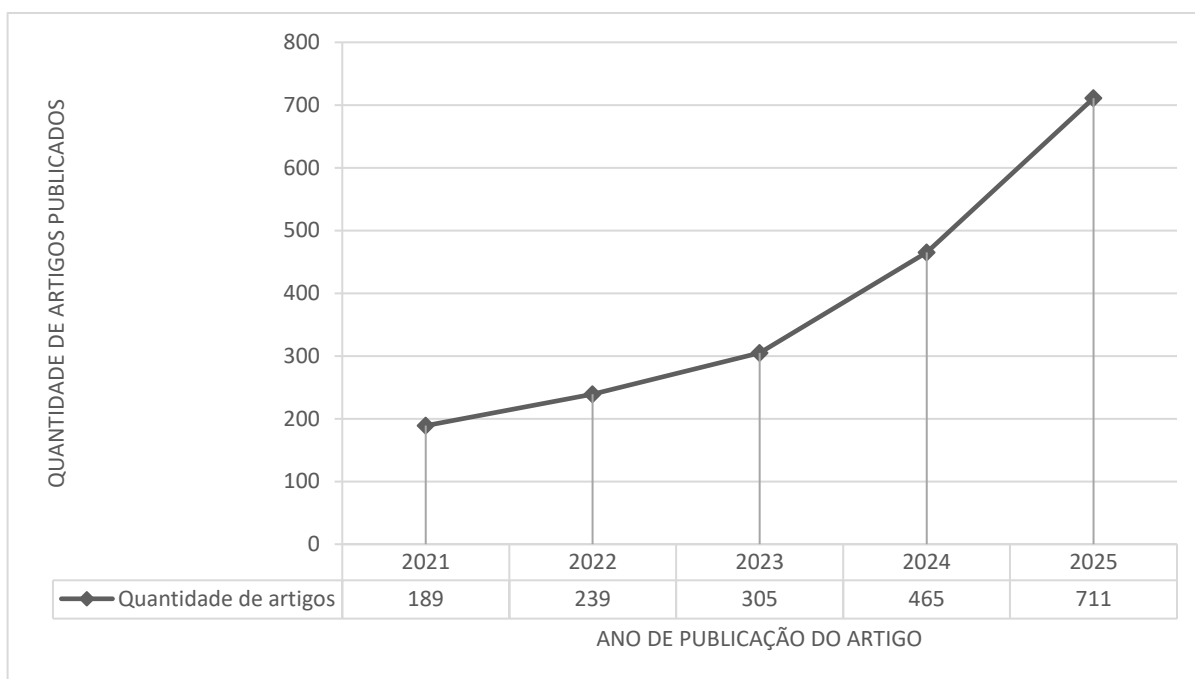
A administração pública brasileira, caracterizada por uma estrutura burocrática, pesada e centralizada, tem mostrado dificuldades para atender às demandas da modernidade. Isso resulta em incapacidade de responder de forma eficaz e rápida às necessidades dos cidadãos, funcionamento que afeta a competitividade internacional, incoerência no modelo organizacional, processos de decisão longos e complexos que impedem a solução eficiente de problemas, falta de transparência, desconfiança quanto à legalidade e desmotivação dos funcionários, além da desvalorização do conceito de serviço público (Matias-Pereira, 2012).

Desta forma, este estudo justifica-se pela busca de eficiência no serviço público, dado que, se a pesquisa confirmar que um sistema com IA agiliza a identificação de ações institucionais indicadas nas pesquisas de autoavaliação do campus Bento Gonçalves do IFRS, a comissão própria de avaliação da instituição poderá concluir seus resultados anuais com maior eficácia.

De forma bibliométrica, a pesquisa se justifica pelo crescente interesse acadêmico em inteligência artificial aplicada ao serviço público. Ao realizar uma pesquisa na base de dados da Scopus, utilizando a query de pesquisa: “(*TITLE-ABS-KEY ('artificial intelligence') AND TITLE-ABS-KEY ('public service')) AND PUBYEAR > 2020 AND PUBYEAR < 2026 AND (LIMIT-TO (DOCTYPE, 'ar'))*”, encontra-se como resultado 1909 artigos publicados. Da mesma forma, se for realizada uma pesquisa junto a base de dados do Google Acadêmico, utilizando os termos “*artificial intelligence*” e “*public service*”, limitando ao mesmo período, obtêm-se 64150 resultados de artigos. O crescimento na quantidade de artigos publicados na base de

dados da Scopus, pode ser melhor visualizado graficamente conforme apresentado na Figura 1.

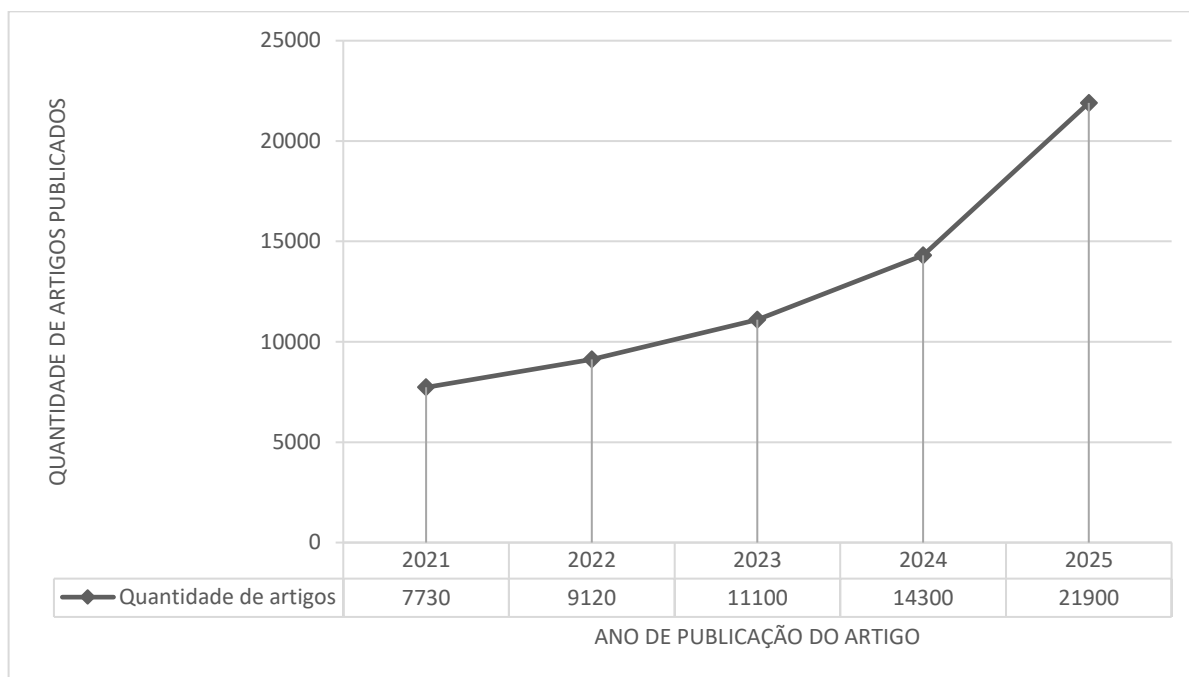
Figura 1 - Publicações de artigos com os termos "artificial intelligence" e "public service" na base de dados da Scopus



Fonte: autoria própria (2026).

Da mesma forma, fica evidente o crescimento na quantidade de artigos publicados na base de dados do Google Acadêmico, conforme apresentado na Figura 2.

Figura 2 - Publicações de artigos com os termos "artificial intelligence" e "public service" na base de dados do Google Acadêmico



Fonte: autoria própria (2026).

Segundo Alhosani e Alhashmi (2024), o setor público pode usar a IA para resolver vários problemas, incluindo, entre outros, a barreira linguística, atrasos na prestação de serviços, longos períodos de espera, falta de assistência massiva número de casos previsíveis e altas taxas de rotatividade. Ainda segundo os autores, o aumento da produtividade, a diminuição das cargas de trabalho e procedimentos simplificados são apenas algumas das maneiras pelas quais a IA ajudará o setor público, o governo e a sociedade.

Como exemplos de aplicação da IA no serviço público, Alhosani e Alhashmi (2024) citam o Departamento de Receita de Nova Gales do Sul, na Austrália, que utiliza IA desde 2018 para identificar e apoiar clientes vulneráveis que não conseguem pagar suas penalidades, prevendo a vulnerabilidade dos clientes e oferecendo opções de resolução alternativas, evitando que sejam forçados a pagar multas que não podem arcar, conectando-os a canais de resolução mais apropriados. Outro exemplo citado pelos autores é o dos Serviços de Saúde Nacional (NHS) do Reino Unido, onde a IA é utilizada para melhorar significativamente os cuidados de saúde, incluindo a detecção e prevenção precoce de doenças e o suporte à tomada de decisões clínicas, permitindo monitorar a saúde dos pacientes em tempo real e alertar os profissionais

sobre possíveis ameaças com base em dados como peso, níveis de açúcar no sangue e frequência cardíaca, como demonstrado durante a pandemia de COVID-19. Esses exemplos ilustram como a IA pode ser aplicada de maneira prática para melhorar a eficiência e a eficácia dos serviços públicos.

Apesar dos diversos exemplos de aplicação da IA no serviço público, ampliando o termo “institutional self-assessment surveys” na busca de publicações nas bases de dados da Scopus e do Google Acadêmico, não foram encontradas pesquisas que verifiquem a eficiência da utilização de um sistema com inteligência artificial especificamente no apontamento de ações provenientes de pesquisas de autoavaliação institucional. Como tal pesquisa será realizada no Brasil, também foi pesquisado o termo “inteligência artificial em pesquisas de autoavaliação” no site da Biblioteca Digital Brasileira de Teses e Dissertações, que resultou em um artigo de validação de um sistema de autoavaliação e orientação para a prática da educação física, sendo diferente da pesquisa realizada neste trabalho. Assim, a realização desta pesquisa justifica-se pela contribuição para a evolução do tema aplicado ao serviço público.

1.4 ADERÊNCIA À LINHA DE PESQUISA PPGA

A pesquisa proposta se encaixa na linha de pesquisa Estratégia e Operações por estar focada no desenvolvimento de um sistema de inteligência artificial para otimizar a análise das pesquisas de autoavaliação institucional. Esse sistema contribui diretamente para a eficácia organizacional ao automatizar a identificação das ações institucionais apontadas nessas pesquisas, facilitando a gestão e a tomada de decisão. A relação com a dimensão Sistemas se dá pelo fato de que a pesquisa envolve a criação de uma ferramenta tecnológica voltada à gestão de informações, promovendo maior precisão e agilidade na interpretação dos dados institucionais. Além disso, ao integrar inteligência artificial aos processos administrativos, a pesquisa favorece a melhoria contínua das operações internas do IFRS – campus Bento Gonçalves, tornando-as mais estratégicas e fundamentadas em evidências.

Na dimensão Mercado, a pesquisa fortalece o posicionamento institucional ao aprimorar a compreensão das necessidades e expectativas de seus públicos de interesse. A automação da análise das pesquisas permite que a instituição desenvolva estratégias mais eficazes para atender às demandas internas e externas, tornando a

gestão mais dinâmica e proativa. Dessa forma, a iniciativa está alinhada à busca por maior sustentabilidade organizacional, garantindo que as estratégias adotadas estejam embasadas em informações precisas e atualizadas. Ao integrar tecnologia e gestão estratégica, a pesquisa reafirma o papel da inteligência artificial como um elemento fundamental para o desempenho institucional em ambientes complexos e competitivos.

2 REFERENCIAL TEÓRICO

Neste capítulo, intitulado referencial teórico, será exposto o embasamento desta pesquisa, abrangendo abordagens teóricas específicas: Inteligência Artificial (2.1), Serviço Público (2.2) e a Relação entre a Inteligência Artificial e o Serviço Público (2.3). Cada uma dessas abordagens é discutida com uma contextualização histórica e a análise dos benefícios e desafios, conforme descrito na literatura e em artigos acadêmicos publicados na área.

2.1 INTELIGÊNCIA ARTIFICIAL (IA)

De acordo com Russell e Norvig (2024) a definição de Inteligência Artificial pode ser categorizada em diferentes abordagens, dependendo do enfoque: sistemas que pensam como humanos, sistemas que agem como humanos, sistemas que pensam racionalmente e sistemas que agem racionalmente. Os autores definem a IA de forma ampla como o estudo e o desenvolvimento de agentes inteligentes, que são sistemas capazes de perceber seu ambiente e tomar ações que maximizam suas chances de sucesso em um determinado objetivo. Essa definição destaca a IA como um campo que busca criar máquinas que simulam aspectos de comportamento e raciocínio humano, bem como agentes que atuam de forma racional.

Essa pluralidade de definições não é acidental, tampouco se restringe à obra desses autores. Haenlein e Kaplan (2019) observam que a delimitação do campo é permanentemente instável em razão do que denominam efeito IA. Sempre que uma técnica antes tida como inteligente passa ao uso corrente deixa de ser reconhecida como inteligência artificial e é reclassificada como simples automação. Para escapar dessa fronteira móvel, os autores propõem uma definição de caráter funcional, segundo a qual a IA é a capacidade de um sistema interpretar corretamente dados externos, aprender a partir desses dados e utilizar o aprendizado para atingir objetivos e tarefas específicos por meio de adaptação flexível. Trata-se de formulação próxima da noção de agente racional, à qual Russell e Norvig (2024) chegam ao final de seu percurso, com a diferença de que prescinde por completo da comparação com a mente humana.

Para afirmar que um programa pensa como um ser humano, é necessário compreender como os seres humanos pensam, o que envolve introspecção,

experimentos psicológicos e imagens cerebrais. Uma vez que uma teoria suficientemente precisa da mente humana seja desenvolvida, ela pode ser expressa em um programa de computador, e a correspondência entre os comportamentos do programa e os humanos pode indicar a presença de mecanismos semelhantes em ambos (Russell; Norvig, 2024). Os autores destacam que no desenvolvimento do *General Problem Solver* (GPS), o foco dos criadores estava em comparar os passos do programa com os de indivíduos humanos na resolução de problemas. A ciência cognitiva combina modelos computacionais de IA com técnicas experimentais da psicologia para criar teorias verificáveis sobre o funcionamento da mente. Este campo se baseia em investigações experimentais de humanos ou animais, diferenciando-se da IA, que evoluiu separando modelos de desempenho humano e eficácia algorítmica. Essa separação contribuiu para o rápido desenvolvimento de ambos os campos, que continuam a influenciar-se mutuamente, especialmente em áreas como a visão computacional.

Ainda segundo Russell e Norvig (2024) a proposta de Alan Turing em 1950 por meio do conhecido Teste de Turing, que oferece uma definição operacional de Inteligência Artificial é um exemplo da abordagem da IA com enfoque na ação de um sistema como um humano. Segundo essa proposta um computador pode ser considerado equivalente a um ser humano se, ao responder a perguntas feitas por um interlocutor, suas respostas não puderem ser distinguidas das de um humano. Para tal comportamento, um computador precisaria ter as seguintes capacidades: processamento de linguagem natural, representação de conhecimento, raciocínio automatizado, aprendizado de máquina, visão computacional e robótica.

Para definir sistemas que pensam racionalmente, Russell e Norvig (2024) explicam a “lei do pensamento” com base em Aristóteles, que foi pioneiro na codificação do “pensamento correto”, criando silogismos que estruturavam argumentos com conclusões logicamente válidas a partir de premissas corretas, iniciando assim o campo da lógica. No século XIX, lógicos desenvolveram uma notação precisa para expressar declarações e relações diversas, ultrapassando a aritmética básica. Por volta de 1965, surgiram programas capazes de resolver, em princípio, qualquer problema descrito em notação lógica, embora, em casos insolúveis, pudessem entrar em *loops* infinitos. A tradição logicista da IA busca criar sistemas inteligentes baseados nessa abordagem, mas enfrenta dois desafios

principais: a dificuldade de formalizar conhecimentos informais e a diferença entre resolver problemas teóricos e práticos.

A última abordagem que define o que é a IA para Russell e Norvig (2024) trata de sistemas que agem racionalmente ou segundo eles, definido como agentes racionais. Um agente é um sistema que age de forma autônoma, percebendo seu ambiente, adaptando-se a mudanças, persistindo ao longo do tempo e buscando metas. Um agente racional age para alcançar o melhor resultado ou, diante de incertezas, o resultado esperado mais favorável. A abordagem das "leis do pensamento" enfatiza inferências corretas como parte do comportamento racional, mas a racionalidade envolve mais do que inferências lógicas, incluindo ações instintivas que não requerem deliberação, como reflexos.

Russell e Norvig (2024) complementam o entendimento de agentes racionais dizendo que todas as habilidades necessárias para o Teste de Turing contribuem para que um agente aja racionalmente. Isso inclui representação de conhecimento, raciocínio e geração de linguagem natural, o que é essencial para interações sociais complexas e aprendizado eficiente para comportamento eficaz. A abordagem do agente racional é mais ampla que a das "leis do pensamento", pois inclui diferentes mecanismos de racionalidade e é mais adequada para desenvolvimento científico, com padrões de racionalidade bem definidos e modeláveis. No entanto, alcançar uma racionalidade perfeita é inviável em ambientes complexos devido às altas demandas computacionais.

É justamente dessa inviabilidade que parte a formulação de Poole e Mackworth (2023), para quem a Inteligência Artificial deve ser entendida menos como a busca de um ideal de racionalidade e mais como o estudo do projeto de agentes computacionais que se comportem de maneira adequada ao ambiente em que operam. Segundo os autores, cada sistema é caracterizado por dimensões como a modularidade, a forma de representação do conhecimento, a presença ou a ausência de aprendizado, o grau de incerteza envolvido, o horizonte de planejamento e o número de agentes que atuam simultaneamente. A pergunta deixa de ser se a máquina pensa e passa a ser qual combinação de capacidades um agente precisa reunir para dar conta de determinada tarefa. Essa combinação, contudo, não se estabeleceu de uma só vez, ela é produto de uma trajetória de quase oito décadas, feita de avanços, frustrações e mudanças de rumo.

2.1.1 História da Inteligência Artificial

O primeiro trabalho reconhecido como IA foi realizado por Warren McCulloch e Walter Pitts em 1943, que criaram um modelo de neurônios artificiais baseado em fisiologia cerebral, lógica proposicional e teoria da computação. Eles demonstraram que redes de neurônios artificiais podiam computar funções e implementar conectivos lógicos. Também sugeriram que redes poderiam aprender, conceito que foi expandido por Donald Hebb em 1949 com a regra de aprendizado de Hebb, ainda influente hoje (Russell; Norvig, 2024).

Em 1950, Marvin Minsky e Dean Edmonds construíram o SNARC, o primeiro computador de rede neural, com 3.000 válvulas eletrônicas para simular uma rede de 40 neurônios. Embora inicialmente recebida com ceticismo, Minsky provou teoremas que destacaram as limitações das redes neurais (Russell; Norvig, 2024).

A visão de Alan Turing foi especialmente influente. Em 1947, ele abordou a IA em palestras e, em 1950, publicou o artigo "*Computing Machinery and Intelligence*", introduzindo o Teste de Turing, aprendizado de máquina, algoritmos genéticos e aprendizagem por reforço. Turing propôs a ideia do "*Child Programme*", sugerindo que seria mais eficaz desenvolver um programa que imitasse a mente infantil, em vez da adulta (Russell; Norvig, 2024).

As ideias de Turing, entretanto, ainda não circulavam sob um nome comum. Haenlein e Kaplan (2019) situam a fundação institucional do campo em 1956, na conferência de verão realizada no Dartmouth College, ocasião em que a expressão inteligência artificial foi cunhada e em que se reuniram, em torno de um mesmo programa de pesquisa, os pesquisadores que conduziram as duas décadas seguintes. Segundo os autores, o encontro fixou não apenas o vocabulário do campo, mas também o otimismo que o caracterizaria: a suposição de que qualquer aspecto da inteligência humana poderia ser descrito com precisão suficiente para que uma máquina viesse a simulá-lo.

De acordo com Russell e Norvig (2024) os primeiros anos da IA foram marcados por sucessos limitados, mas impressionantes para a época. Com computadores primitivos e ferramentas de programação incipientes, os avanços eram notáveis e frequentemente desafiavam a visão predominante de que "uma máquina nunca poderá fazer X". Herbert Simon, em 1957, fez previsões otimistas sobre o futuro da IA acreditando que, em breve, máquinas poderiam pensar e resolver problemas

complexos como humanos. No entanto, essas expectativas foram frustradas pela incapacidade dos programas iniciais de lidar com problemas mais complexos.

As dificuldades enfrentadas incluíam a falta de conhecimento contextual nos programas, que revelaram a necessidade de compreensão profunda do conteúdo para evitar erros. A estratégia de tentar todas as combinações possíveis em busca de soluções mostrou-se insuficiente para problemas maiores, já que a complexidade aumentava exponencialmente. Esse desafio, conhecido como "explosão combinatória", foi uma crítica central que resultou na redução do apoio à pesquisa em IA no Reino Unido (Russell; Norvig, 2024).

Essa retração não foi pontual nem restrita ao Reino Unido. Haenlein e Kaplan (2019) descrevem o período como o primeiro de dois invernos da IA, ciclos em que a distância entre as promessas anunciadas e os resultados efetivamente entregues levou ao corte generalizado de financiamento e ao esvaziamento do interesse acadêmico e industrial pelo campo. O primeiro deles a partir de meados da década de 1970, o segundo ao final dos anos 1980, quando o mercado de sistemas especialistas entrou em colapso. Os autores ressaltam que, em ambos os casos, os fundamentos teóricos permaneceram válidos, o que faltava era capacidade computacional, volume de dados e maturidade de engenharia, condições que só viriam a reunir-se décadas mais tarde. Enquanto isso, a pesquisa buscou caminhos alternativos aos que haviam sido frustrados.

Na primeira década de pesquisa em IA, o principal enfoque foi em mecanismos de busca de uso geral para resolver problemas por meio de passos elementares de raciocínio. Essas abordagens, chamadas de métodos fracos, eram gerais, mas ineficazes para problemas grandes ou complexos. Como alternativa, surgiram métodos que incorporavam conhecimento específico de domínio para realizar raciocínios mais sofisticados e eficazes, essenciais para problemas especializados (Russell; Norvig, 2024).

Um exemplo dessa abordagem foi o programa DENDRAL, desenvolvido em Stanford por Ed Feigenbaum, Bruce Buchanan e Joshua Lederberg. O DENDRAL foi projetado para inferir a estrutura molecular a partir de informações fornecidas por um espectrômetro de massa. O DENDRAL foi significativo por ser o primeiro sistema de IA bem-sucedido com uso intensivo de conhecimento, cuja eficácia vinha de um grande número de regras específicas de domínio. Sistemas posteriores seguiram essa

abordagem: a separação entre o conhecimento (representado por regras) e o mecanismo de raciocínio (Russell; Norvig, 2024).

Nos anos 1980, o algoritmo de aprendizado por retropropagação, originalmente descoberto em 1969 por Bryson e Ho, foi redescoberto por diferentes grupos e aplicado a problemas de aprendizado em ciência da computação e psicologia. A pesquisa em redes neurais se dividiu em dois campos: um focado no desenvolvimento de algoritmos e arquiteturas de rede eficientes e na análise de suas propriedades matemáticas, e outro voltado para a modelagem das propriedades empíricas de neurônios reais e seus conjuntos. Essa divisão reflete a evolução da IA e da ciência cognitiva em caminhos distintos, mas interrelacionados (Russell; Norvig, 2024).

Encorajados pelos avanços na resolução de subproblemas da IA, os pesquisadores começaram a focar no desenvolvimento de agentes completos nos anos seguintes, como exemplificado pela arquitetura SOAR de Newell, Laird e Rosenbloom. Agentes inteligentes, como bots, tornaram-se comuns na Internet e são fundamentais em ferramentas como mecanismos de pesquisa e sistemas de recomendação. A construção de agentes completos mostrou a necessidade de reorganizar subcampos da IA, reconhecendo que sistemas sensoriais não fornecem informações totalmente confiáveis, exigindo que raciocínio e planejamento lidem com a incerteza (Russell; Norvig, 2024).

Nos últimos anos, as teorias existentes passaram a ser usadas como base, com ênfase em teoremas rigorosos e evidências experimentais, ao invés de intuição. A IA agora valoriza aplicações práticas e abandonou seu isolacionismo em relação a outras áreas da ciência da computação, integrando-se com teoria da informação, modelagem estocástica e controle. Exemplos incluem o domínio dos modelos ocultos de Markov (MOMs) no reconhecimento de fala e o retorno da tradução automática baseada em princípios da teoria da informação. As redes neurais, que inicialmente buscavam definir sua aplicabilidade, hoje são comparadas a técnicas de estatística e aprendizado de máquina, contribuindo para a mineração de dados e a indústria associada (Russell; Norvig, 2024).

A obra de Judea Pearl sobre raciocínio probabilístico popularizou as redes bayesianas, melhorando o raciocínio com conhecimento incerto e levando ao desenvolvimento de sistemas especialistas normativos, que seguem as leis da teoria de decisão. Revoluções similares ocorreram em robótica, visão computacional e

representação de conhecimento, que se tornaram mais formais e especializados (Russell; Norvig, 2024).

Nos últimos 60 anos, a ciência da computação focou principalmente nos algoritmos, mas estudos recentes em IA indicam que a ênfase nos dados pode ser mais relevante para muitos problemas. Isso se deve à crescente disponibilidade de grandes volumes de dados, como trilhões de palavras e bilhões de imagens. Observadores destacam que o “gargalo do conhecimento” da IA pode ser superado com métodos de aprendizado baseados em grandes quantidades de dados, em vez da engenharia de conhecimento manual. Isso está trazendo novas aplicações e revitalizando o campo da IA, com milhares de soluções já integradas em diversas indústrias (Russell; Norvig, 2024).

LeCun, Bengio e Hinton (2015) descrevem o aprendizado profundo como um conjunto de métodos capazes de aprender representações dos dados em múltiplos níveis de abstração, dispensando a engenharia manual de atributos que até então exigia a intervenção de especialistas de domínio a cada nova aplicação, exatamente o gargalo do conhecimento a que se referem os observadores mencionados por Russell e Norvig (2024). Dois anos depois, Vaswani *et al.* (2017) propuseram a arquitetura transformer, que substituiu as estruturas recorrentes por um mecanismo de atenção capaz de ponderar a relevância de cada elemento de uma sequência em relação a todos os demais. Ao permitir o processamento paralelo de sequências inteiras, essa arquitetura tornou economicamente viável o treinamento em escalas até então inatingíveis e converteu-se no fundamento técnico comum aos modelos existentes. É esse acúmulo que permite, hoje, distinguir o que a IA efetivamente faz daquilo que dela se espera.

2.1.2 Tipos de Inteligência Artificial

De acordo com Ludermitz (2021) a inteligência artificial pode ser classificada em três tipos principais: IA Fraca (ou IA Focada), IA Forte (ou IA Generalizada) e IA Superinteligente.

A IA Fraca consiste em algoritmos especializados que resolvem problemas em áreas ou tarefas específicas, como sistemas especialistas e sistemas de recomendação. Esses sistemas armazenam grandes quantidades de dados e realizam tarefas complexas, mas sempre dentro de um escopo limitado (Ludermitz,

2021). Russell e Norvig (2024) definem a IA Fraca como a ideia de que as máquinas talvez possam agir de maneira inteligente, ou agir como se fossem inteligentes.

Como alguns exemplos de IA Fraca, Mendonça *et. al.* (2024) apresenta assistentes virtuais (como Siri e Alexa), sistemas de recomendação (como os da Netflix e Amazon), algoritmos de busca (como o Google) e veículos autônomos. Embora a IA fraca seja altamente útil em setores como atendimento ao cliente, medicina, finanças e logística, suas capacidades são limitadas às funções para as quais foi programada.

A IA Forte, ou IA Generalizada, segundo Ludermir (2021) refere-se a algoritmos que conseguem executar múltiplas tarefas com competência comparável à dos humanos. Esses algoritmos frequentemente utilizam técnicas de Aprendizado de Máquina e atualmente representam o nível de IA em que nos encontramos, com aplicações como Visão Computacional. Complementando essa definição, Russell e Norvig (2024) definem a IA Forte como asserção de que as máquinas que o fazem estão realmente pensando, em vez de simularem o pensamento.

Searle (1980) formulou a oposição que os dois pares de definições reproduzem por meio do experimento mental que ficou conhecido como o Quarto Chinês, onde um indivíduo trancado em uma sala, sem qualquer domínio do idioma, manipula símbolos chineses seguindo um manual de regras escrito em sua própria língua e produz respostas indistinguíveis das de um falante nativo. Para o autor, a situação demonstra que a manipulação sintática de símbolos, por mais bem-sucedida que seja em seus resultados, não constitui compreensão semântica, razão pela qual a execução de um programa jamais bastaria para atribuir estados mentais a uma máquina. Recuperar essa origem esclarece a natureza da distinção, ela não versa sobre o que o sistema é capaz de fazer, mas sobre o estatuto que se deve conferir àquilo que ele faz. Trata-se, portanto, de uma questão filosófica, e não de uma fronteira técnica mensurável, o que explica por que as descrições da IA forte costumam formular-se menos em termos de desempenho e mais em termos de atributos internos.

Mendonça *et. al.* (2024) define que a IA forte é um conceito que se refere a sistemas de IA com capacidades intelectuais iguais ou superiores às dos humanos, capazes de entender, aprender e aplicar conhecimento de forma ampla em diferentes contextos. Essa IA seria capaz de raciocínio abstrato, solução de problemas complexos, compreensão emocional e criatividade. Entre suas características estão a

autonomia, o aprendizado contínuo, a consciência e autoconsciência, e a transferência de conhecimento entre domínios.

Segundo Ludermir (2021), IA Superinteligente seria aquela em que os algoritmos superam os humanos em praticamente todas as tarefas. No entanto, sistemas com IA Superinteligente ainda não existem, e não se sabe se um dia serão desenvolvidos com as técnicas atuais de IA.

Ainda que hipotético, esse último estágio é objeto de tratamento sistemático por Bostrom (2018), que define a superinteligência como qualquer intelecto capaz de superar amplamente o desempenho cognitivo humano em praticamente todos os domínios de interesse. O autor examina os caminhos pelos quais tal patamar poderia ser alcançado e sustenta que o problema decisivo não seria de ordem técnica, mas de controle, assegurar que um sistema com essa capacidade permanecesse alinhado a objetivos compatíveis com os humanos. É precisamente a distância entre essa especulação e o que se pode construir com as técnicas disponíveis que explica a posição predominante no campo.

Para Russell e Norvig (2024), a maioria dos pesquisadores de IA adota a hipótese de IA fraca, focando em desenvolver programas que funcionem, independentemente de serem considerados simulações de inteligência ou inteligência real.

Mendonça *et. al.* (2024) diz que o desenvolvimento da IA forte enfrenta desafios como a complexidade computacional, segurança e controle para garantir que ela opere de forma alinhada aos valores humanos, e questões éticas e regulatórias sobre decisões autônomas e os direitos de máquinas conscientes. Embora modelos de IA fraca como GPT-4 e Gemini demonstrem capacidades avançadas, eles ainda não alcançam o nível de IA forte.

Ainda segundo Mendonça *et. al.* (2024) para progredir em direção à IA forte, são necessárias inovações em modelos híbridos, aprendizado auto supervisionado, neuromorfismo e a integração de diversas disciplinas como neurociência e filosofia. A transição de IA fraca para IA forte é complexa, mas promete impactos profundos na sociedade com avanços em diversos setores.

2.1.3 Linguagens de programação

Entre 1936 e 1945, Konrad Zuse, um cientista alemão, construiu uma série de computadores eletromecânicos complexos, quase todos destruídos em 1945 após bombardeios aliados, exceto o modelo Z4. Depois de se separar de seu grupo de pesquisa, Zuse desenvolveu uma linguagem chamada Plankalkül, iniciada em 1943 como parte de sua tese de doutorado, para expressar computações no Z4. Em um manuscrito de 1945, publicado apenas em 1972, Zuse detalhou o Plankalkül e escreveu algoritmos para diversos problemas (Sebesta, 2018).

O Plankalkül era uma linguagem de programação avançada para sua época, especialmente em estruturas de dados. Utilizava bits como tipo de dado mais simples e construía tipos inteiros e de ponto flutuante com notação de complemento de dois e esquema de “bit oculto”. O manuscrito de Zuse incluía programas complexos como algoritmos de ordenação, análise de grafos, operações numéricas, análise de fórmulas lógicas e 49 páginas de algoritmos para xadrez. A notação de Plankalkül, que exigia múltiplas linhas por sentença, tornava sua leitura intimidadora e poderia dificultar a implementação na época (Sebesta, 2018).

No final dos anos 1940 e início dos anos 1950, as primeiras linguagens de programação, chamadas de pseudocódigos na época, surgiram em resposta à dificuldade de programar computadores primitivos. Esses computadores eram lentos, pouco confiáveis, caros e tinham memória limitada, além de não possuírem software de suporte. A programação era feita em código de máquina, o que era tedioso e sujeito a erros, pois usava códigos numéricos em vez de instruções textuais e endereçamento absoluto, complicando modificações e inserções de código. Problemas como esses incentivaram o desenvolvimento de montadores e linguagens de montagem para facilitar a programação (Sebesta, 2018).

O IBM 704, introduzido em 1954, representou um grande avanço na computação, principalmente por ter impulsionado o desenvolvimento da linguagem de programação Fortran. A lentidão dos sistemas de interpretação era tolerada no final da década de 1940 e até meados da década de 1950 devido à falta de hardware de ponto flutuante nos computadores, o que obrigava as operações de ponto flutuante a serem simuladas em software, um processo demorado. Com grande parte do tempo do processador dedicada a isso, a sobrecarga da interpretação era vista como aceitável. No entanto, muitos programadores preferiam usar código de máquina ou de montagem para maior eficiência. O lançamento do IBM 704, com suporte a instruções

de ponto flutuante e indexação em hardware, marcou o fim da era da interpretação na computação científica, expondo o custo da interpretação (Sebesta, 2018).

A primeira linguagem de programação funcional surgiu para atender à necessidade de processamento de listas, impulsionada pelas primeiras aplicações em Inteligência Artificial (IA), cujo interesse cresceu na década de 1950. Esse interesse vinha de áreas como a linguística, que buscava processar a linguagem natural; a psicologia, que queria modelar o armazenamento e recuperação de informações humanas; e a matemática, que desejava mecanizar processos como a prova de teoremas. Todas essas áreas concluíram que era necessário desenvolver métodos para que computadores processassem dados simbólicos em listas encadeadas, contrastando com a predominância do uso de dados numéricos em vetores na época (Sebesta, 2018).

O Lisp dominou as aplicações de IA por muitos anos e superou sua má reputação inicial com implementações compiladas que tornaram o código mais rápido do que quando executado em interpretadores. Além de seu sucesso em IA, o Lisp foi pioneiro na programação funcional, uma área muito ativa em pesquisa de linguagens de programação. Muitos especialistas acreditam que a programação funcional oferece uma estratégia superior ao desenvolvimento de software em comparação com a programação procedural em linguagens imperativas (Sebesta, 2018).

O primeiro passo em direção à sofisticação das linguagens de programação foi o ALGOL 60, que surgiu em 1958 como resultado de esforços de grupos de usuários de computadores para criar uma linguagem de programação universal para aplicações científicas. Estes grupos foram motivados pelo desenvolvimento de várias linguagens de alto nível, o que dificultava o compartilhamento de programas devido à dependência de arquiteturas específicas (Sebesta, 2018).

A ambição de universalidade do ALGOL 60 expõe aquilo que Scott (2015) identifica como o núcleo do projeto de qualquer linguagem de programação: a necessidade de conciliar critérios que competem entre si. Segundo o autor, uma linguagem é sempre um compromisso entre expressividade, legibilidade, confiabilidade, portabilidade e custo de implementação, de modo que o ganho em um desses eixos costuma ser obtido às custas de outro. Scott (2015) sustenta ainda que a história do campo pode ser lida como uma sucessão de elevações no nível de abstração, da manipulação direta de endereços de memória ao encapsulamento de dados e comportamento, e deste aos modelos declarativos, nos quais se especifica o

que deve ser computado sem descrever exaustivamente como. Vista sob essa ótica, a trajetória posterior das linguagens deixa de configurar uma sequência de substituições e revela-se como a ocupação de nichos distintos, cada qual privilegiando um conjunto próprio de critérios.

Nos anos 1970 e início dos anos 1980, o Basic foi uma linguagem de programação amplamente usada, mas pouco reconhecida, por ser ignorada por cientistas da computação e por ter, em suas primeiras versões, uma estrutura deselegante e poucas sentenças de controle. O design do Basic foi influenciado principalmente pelo Fortran, com elementos da sintaxe do ALGOL 60, e evoluiu de forma despadronizada. O Basic se tornou muito popular em microcomputadores por ser fácil de aprender, especialmente para iniciantes, e por ter dialetos que podiam ser implementados em máquinas com pouca memória. Com o aumento da capacidade dos microcomputadores e a chegada de novas linguagens, o uso do Basic diminuiu, mas foi revitalizado com o lançamento do Visual Basic em 1991 (Sebesta, 2018).

As linguagens de programação continuaram evoluindo ao longo do tempo e tiveram uma nova “revolução” com o uso da *internet*, que cresceu significativamente nos anos 1990 com os navegadores gráficos, criando a necessidade de computação dinâmica em documentos HTML, inicialmente estáticos. A computação no lado do servidor era feita com CGI, enquanto no navegador surgiram os applets Java. Com o tempo, essas soluções foram superadas por linguagens de scripting, como JavaScript, criada por Brendan Eich na Netscape e inicialmente chamada Mocha e depois LiveScript. Em 1995, tornou-se JavaScript em colaboração com a Sun Microsystems (Sebesta, 2018).

O JavaScript é mais comumente usado em navegadores, embutido em documentos HTML e interpretado pelo navegador, para validação de formulários e criação de conteúdo dinâmico. Embora seu nome sugira relação com Java, as semelhanças se limitam à sintaxe. Java é fortemente tipada e orientada a objetos, enquanto JavaScript é dinamicamente tipada e não tem suporte completo para herança e vinculação dinâmica de métodos. Um de seus principais usos é a modificação dinâmica de documentos HTML por meio do *Document Object Model* (DOM), que permite o controle interativo dos elementos de uma página (Sebesta, 2018).

Em 1994, foi criada a linguagem de programação PHP, por Rasmus Lerdorf. Originalmente, "PHP" significava "Personal Home Page", mas posteriormente passou

a significar "PHP: Hypertext Preprocessor". O PHP é uma linguagem de *scripting* do lado servidor, embutida em HTML e projetada para aplicações *Web*. Seu código é interpretado no servidor e gera saída em HTML, invisível para os navegadores. A linguagem tem semelhanças com JavaScript em sua sintaxe e tipagem dinâmica, e seus vetores combinam características de JavaScript e Perl (Sebesta, 2018).

Desenvolvida inicialmente por Guido van Rossum nos anos 1990, surgiu o Python, uma linguagem de *scripting* orientada a objetos. É uma linguagem de código aberto disponível para várias plataformas e usada para tarefas como administração de sistemas, programação CGI e outras aplicações menores. A sintaxe de Python é original, com tipagem dinâmica e estruturas de dados como listas, tuplas (listas imutáveis) e dicionários, permitindo que os usuários criem módulos em linguagens compiladas para adicionar funcionalidades ao interpretador (Sebesta, 2018).

O traço comum a JavaScript, PHP e Python não é apenas cronológico. Ousterhout (1998) identificava já à época em que essas linguagens se consolidavam a divisão de trabalho que passaria a caracterizar o desenvolvimento de aplicações: linguagens de sistema, fortemente tipadas e compiladas, destinam-se à construção de componentes de alto desempenho a partir do zero, ao passo que linguagens de *scripting*, dinamicamente tipadas e interpretadas, destinam-se a conectar componentes já existentes. O autor argumenta que a tipagem dinâmica, frequentemente tratada como fragilidade, é justamente o que confere a essas linguagens sua função própria, pois permite manipular dados de origens heterogêneas sem a declaração prévia de estruturas rígidas, condição para que a linguagem opere como elemento de ligação entre sistemas que não foram concebidos para conversar entre si.

Na figura 3 é apresentada uma "árvore genealógica" das linguagens de programação, que ajuda a entender a relação entre as primeiras linguagens criadas e as linguagens mais modernas utilizadas atualmente. No diagrama, o eixo vertical corresponde à linha do tempo: os anos dispostos à esquerda, do topo (1957) para a base (2004), indicam a ordem cronológica de surgimento de cada linguagem. As ligações entre as linguagens, por sua vez, representam relações de influência, isto é, o modo como uma linguagem deu origem a outra ou inspirou sua concepção. As ligações predominantemente verticais acompanham essa linha do tempo e conectam uma linguagem às suas sucessoras diretas dentro de uma mesma linhagem. Já as ligações horizontais conectam linguagens situadas em um mesmo período, porém

atualizações em estruturas de dados dependem de procedimentos específicos, que são detalhados pelo programador com base em seu conhecimento do domínio. Nas linguagens de programação, as estruturas de dados podem representar fatos, mas a inferência é dependente de procedimentos implementados pelo programador (Russell; Norvig, 2024).

2.1.4 Integração entre sistemas com IA generativa

De acordo com Russell e Norvig (2024) existem três tipos principais de aprendizagem de máquinas: o aprendizado supervisionado, onde o agente recebe exemplos de pares de entrada e saída (rótulos) e aprende uma função de mapeamento entre eles. O aprendizado não supervisionado, onde o agente identifica padrões nas entradas sem feedback explícito. E o aprendizado por reforço, onde o agente aprende com base em reforços, recebendo recompensas e punições.

Entre os modelos de redes neurais existentes, temos o GPT (*Generative Pre-trained Transformer*), que é um modelo de rede neural treinado em grandes quantidades de texto para gerar textos novos e de alta qualidade a partir de uma entrada pré-definida e contexto. Utiliza a técnica de "*transformer*", que considera o contexto das palavras ao gerar texto. O GPT é treinado por meio de aprendizado não supervisionado, onde aprende a prever a próxima palavra em uma sequência e ajusta seus pesos para minimizar o erro entre a previsão e a palavra real. O ChatGPT é uma versão ajustada de um modelo GPT de grande escala (Ramos, 2023).

A técnica de transformer a que Ramos (2023) se refere foi proposta por Vaswani *et al.* (2017) e repousa em um mecanismo de atenção, que atribui a cada elemento da sequência de entrada um peso proporcional à sua relevância em relação a todos os demais, permitindo ao modelo capturar dependências de longo alcance sem percorrer o texto sequencialmente. O que essa arquitetura tornou possível foi demonstrado por Brown *et al.* (2020), ao ampliarem a escala do modelo e do *corpus* de treinamento, os autores constataram a emergência do aprendizado no contexto, isto é, a capacidade de executar tarefas inteiramente novas a partir apenas de uma instrução ou de poucos exemplos apresentados na própria entrada, sem qualquer atualização dos parâmetros do modelo. O achado tem consequência prática direta, se a tarefa passa a ser especificada no texto de entrada, a formulação desse texto deixa de ser detalhe de uso e converte-se em variável de projeto.

Modelos de linguagem de IA generativa, como o ChatGPT, podem ser uma ferramenta valiosa para pesquisadores em diversos campos, pois utilizam textos públicos da internet como base para buscar informações, analisá-las e gerar respostas a perguntas e solicitações através de comandos chamados "prompts". A criação de diálogos eficientes e personalizados melhora a interação, tornando as respostas mais precisas e a experiência mais satisfatória. Essa prática é chamada de Engenharia de Prompt e busca tornar a IA mais acessível e fácil de usar (Ramos, 2023).

O amadurecimento dessa prática conduziu à sua sistematização. White *et al.* (2023) propõem organizá-la sob a forma de um catálogo de padrões de *prompt*, à semelhança dos padrões de projeto consagrados na engenharia de software, documentando construções reutilizáveis e demonstrando como podem ser combinadas em fluxos de trabalho. A contribuição é menos técnica do que metodológica, ao nomear e descrever essas construções, os autores tornam possível documentar e reproduzir um *prompt*, requisito indispensável quando ele deixa de ser digitado por uma pessoa e passa a ser emitido por um sistema.

Persiste, ainda assim, uma limitação relevante: o modelo responde a partir do conhecimento incorporado durante o treinamento, o que compromete tarefas ancoradas em documentos determinados. Para contorná-la, Lewis *et al.* (2020) propõem a geração aumentada por recuperação (Retrieval-Augmented Generation), arquitetura que acopla ao modelo generativo um mecanismo de busca sobre uma base documental externa, de modo que cada resposta seja condicionada aos documentos efetivamente recuperados. A estratégia é especialmente pertinente ao setor público, no qual a resposta deve poder ser remetida ao documento que a fundamenta. Sua adoção, contudo, pressupõe que o modelo possa ser acionado a partir do sistema que detém esses documentos.

É possível integrar diversas linguagens de programação às APIs dos GPTs existentes, desde que sejam feitas as devidas configurações e utilizadas ferramentas apropriadas (Gallardo; Machuca; Semblantes, 2023).

2.1.5 ChatGPT

O ChatGPT é um modelo generativo de linguagem desenvolvido pela empresa OpenAI e tornou-se uma das ferramentas de inteligência artificial mais difundidas para tarefas de processamento de linguagem natural e desenvolvimento

de software. Sua arquitetura baseia-se em *transformers* otimizados por aprendizado por reforço a partir de *feedback* humano (*Reinforcement Learning from Human Feedback – RLHF*), abordagem projetada para priorizar a aplicabilidade a um amplo conjunto de tarefas e o suporte a diversas linguagens de programação, como Python, Ruby e Java (Subahi, 2026).

O procedimento de otimização a que Subahi (2026) se refere foi descrito por Ouyang *et al.* (2022) e desenvolve-se em três etapas: o ajuste fino supervisionado do modelo pré-treinado sobre demonstrações redigidas por anotadores humanos, o treinamento de um modelo de recompensa a partir de comparações em que esses anotadores ordenam por preferência múltiplas saídas geradas para um mesmo comando e, por fim, a otimização da política do modelo por aprendizado por reforço contra esse modelo de recompensa. O resultado central do estudo esclarece por que a técnica se generalizou, um modelo de 1,3 bilhão de parâmetros assim alinhado foi preferido pelos avaliadores humanos a outro de 175 bilhões que não passara pelo procedimento, o que evidencia que a utilidade percebida de um modelo de linguagem depende menos de sua escala do que de seu alinhamento à intenção de quem o utiliza. Os autores registram, contudo, que o alinhamento não elimina a produção de conteúdo factualmente incorreto.

O ChatGPT reúne diversas gerações de modelos de inteligência artificial desenvolvidos pela OpenAI ao longo do tempo. De acordo com Shean *et al.* (2026), sua geração mais recente é disponibilizada em múltiplas configurações e em diferentes níveis de raciocínio (baixo, médio e alto), que determinam a quantidade de recursos de cadeia de pensamento alocados internamente pelo modelo durante a resolução de uma tarefa.

Em avaliação comparativa conduzida por Shean *et al.* (2026) com questões de múltipla escolha no estilo de provas de residência em oftalmologia, os modelos da família GPT-5 alcançaram desempenho elevado, respondendo corretamente cerca de noventa por cento das questões, e todas as variantes superaram com folga o desempenho de candidatos humanos. Os autores observaram, contudo, que as configurações otimizadas para raciocínio obtiveram resultados superiores aos das versões básicas e que o modelo básico mais recente apresentou a menor acurácia do estudo, o que sugere um possível compromisso em que os modelos mais novos priorizam eficiência computacional, latência e escalabilidade de implantação em detrimento da acurácia bruta em tarefas de raciocínio complexo. O estudo também

evidenciou limitações persistentes diante de questões que dependem da interpretação de imagens e de questões de ordem superior, voltadas à tomada de decisão e ao raciocínio em múltiplas etapas, nas quais a acurácia de todos os modelos avaliados se reduziu de forma consistente (Shean *et al.*, 2026).

Em avaliação voltada ao desenvolvimento de linguagens específicas de domínio, Subahi (2026) observou que o ChatGPT entregou alta clareza e precisão sintática, organizando suas respostas em estruturas bem definidas e em gramática no formato *Extended Backus–Naur Form (EBNF)*, o que tornou as saídas mais legíveis e prontas para processamento automatizado. Em consequência desse equilíbrio entre estrutura, correção e usabilidade, o ChatGPT obteve as médias gerais mais elevadas na avaliação realizada por especialistas, além de figurar entre os modelos com maior taxa de conformidade sintática.

2.1.6 DeepSeek

Concebido pela DeepSeek AI, o DeepSeek ganhou projeção acelerada em razão de seu forte desempenho em tarefas de linguagem natural. Sua concepção combina a estratégia de mistura de especialistas (*Mixture-of-Experts – MoE*) com uma estrutura de *transformer*, arranjo que processa e classifica grandes volumes textuais de modo econômico e atinge alta exatidão com custo de treinamento reduzido (Subahi, 2026). Conforme o autor, o sistema revela-se particularmente eficaz ao extrair respostas confiáveis a partir de comandos curtos e mostra-se sólido na geração de código e na dedução lógica, ainda que, em alguns testes, registre latência superior à de parte de seus concorrentes.

Subahi (2026) descreve o DeepSeek Coder, voltado à geração de código e à engenharia de software, com dimensões que variam de 1,3 a 33 bilhões de parâmetros e amplas janelas de contexto, e o DeepSeek R1, dedicado à resolução de problemas entre domínios distintos e provido de raciocínio sequencial baseado em cadeia de pensamento, recurso que amplia o entendimento de suas deduções. Dentre essas opções, o presente estudo recorre ao deepseek-chat, configuração de propósito geral orientada ao diálogo, que herda a economia computacional proporcionada pela arquitetura de mistura de especialistas e a aptidão para produzir previsões consistentes a partir de instruções sucintas, conforme apontado pelo mesmo autor.

O raciocínio em cadeia de pensamento que distingue o R1 foi obtido por um procedimento incomum, relatado pelos próprios desenvolvedores do modelo. Guo *et al.* (2025) descrevem que as capacidades de raciocínio emergiram de aprendizado por reforço puro, em que a recompensa era atribuída exclusivamente pela verificação da resposta final, sem que trajetórias de raciocínio anotadas por humanos fossem fornecidas em qualquer etapa. Desse processo emergiram, sem programação explícita, comportamentos como a autorreflexão, a verificação de passos intermediários e a adaptação dinâmica de estratégia. Os autores registram, todavia, que a versão treinada apenas por reforço apresentava baixa legibilidade e mistura de idiomas no encadeamento do raciocínio, limitação apenas parcialmente mitigada, na versão final. O relato ajuda a compreender tanto as virtudes quanto as fragilidades que o modelo revela quando posto lado a lado com seus concorrentes.

Quando confrontado com outros sistemas, o DeepSeek apresenta um perfil de contrastes. Na verificação objetiva conduzida por Subahi (2026), foi o que exibiu a maior validade semântica entre as construções aprovadas, sinal de sólida aderência às restrições impostas. Em contrapartida, obteve uma das mais baixas taxas de aprovação sintática, por mostrar-se suscetível a variações de formatação e de vocabulário, como unidades de medida e palavras-chave estruturais. Em outras comparações, e a depender da versão avaliada, o modelo foi posicionado entre os de resposta mais célere, embora com exatidão apenas intermediária (Liu *et al.*, 2026), tendo ainda sua legibilidade apontada como inferior à dos pares, por recorrer a uma redação de maior densidade (Hang *et al.*, 2026).

Já no estudo dedicado à criação de linguagens específicas de domínio, o DeepSeek sobressaiu-se pelo rigor sintático e pelo minucioso detalhamento em todas as fases, incluindo a validação lógica e o tratamento de casos-limite, como a rejeição de intervalos inválidos para preservar a integridade dos tipos (Subahi, 2026). Estruturadas em blocos rotulados e em notação gramatical formal, suas saídas mostraram-se precisas e imediatamente aproveitáveis para o processamento automático.

2.1.7 Gemini

Criado pelo Google, o Gemini distingue-se por sua natureza multimodal, sendo capaz de interpretar e combinar conteúdos em texto, imagem, áudio e vídeo

dentro de uma mesma rede neural do tipo *transformer*. Essa característica o habilita à análise de dados em alto nível e à produção de conteúdo em formatos variados (Subahi, 2026). Conforme o referido estudo, trata-se de um sistema com desempenho regular nas atividades convencionais dos modelos de linguagem, frequentemente entre os mais ágeis em tempo de execução, cuja engenharia de processamento favorece a rapidez sem comprometer de modo relevante a precisão dos resultados, embora o tratamento simultâneo de múltiplas mídias demande recursos computacionais mais robustos.

A natureza dessa multimodalidade é detalhada por Anil *et al.* (2023), onde os autores esclarecem que o modelo não acopla codificadores independentes por modalidade a um núcleo textual, mas é treinado desde o início sobre sequências que intercalam texto, imagem, áudio e vídeo, de forma que todas as modalidades compartilhem uma mesma representação interna. Essa opção de projeto tem duas consequências que se manifestam nos resultados, de um lado, sustenta a robustez do modelo em tarefas que dependem da leitura de imagens, uma vez que o conteúdo visual não sofre a perda de informação inerente à sua tradução prévia para texto. De outro, esclarece a exigência computacional apontada por Subahi (2026), pois o custo de processar múltiplas mídias é assumido desde a origem, e não delegado a componentes acessórios.

Segundo Shean *et al.* (2026) a versão mais atual contempla diferentes perfis de operação e graus de raciocínio (alto e baixo), parâmetro que regula o volume de processamento dedicado às cadeias de pensamento acionadas para solucionar cada problema. Nesse conjunto inserem-se ainda configurações orientadas à eficiência, categoria à qual pertence o gemini-3.5-flash, coerente com os atributos de baixa latência e ampla escalabilidade de implantação que, segundo os mesmos autores, costumam ser priorizados pelas variantes voltadas ao desempenho computacional.

No experimento comparativo relatado por Shean *et al.* (2026), o Gemini 3 Pro obteve o índice de acerto mais alto entre os sistemas testados, próximo de noventa e quatro por cento, posicionando-se à frente de toda a família GPT-5 nos diversos domínios e patamares de dificuldade. Verificou-se que o perfil ajustado para raciocínio rendeu os melhores resultados e que esse modelo sofreu a menor perda de acerto nos itens que exigiam leitura de imagens, fato atribuído ao amplo acervo de treinamento visual mantido pela Google. Por outro lado, constatou-se uma degradação generalizada de desempenho, comum a todos os sistemas, nas tarefas

de maior complexidade cognitiva, que demandam decisão e encadeamento de várias etapas de raciocínio.

No campo da construção de linguagens específicas de domínio, Subahi (2026) apontou que o Gemini produziu material pertinente ao contexto, embora com clareza apenas mediana, valendo-se de descrições em linguagem quase natural e de um repertório limitado de construções, sem incorporar recursos mais elaborados de escalonamento ou de execução de protocolos. Mesmo diante dessa expressividade restrita, o modelo destacou-se pela aderência às regras formais, com a maioria de suas saídas em conformidade com as especificações sintáticas examinadas.

2.1.8 Claude

Idealizado pela Anthropic, o Claude foi construído sobre uma arquitetura *transformer* ajustada por realimentação humana. O modelo tem como traço mais singular, segundo Subahi (2026), a ênfase de seu projeto em segurança e em princípios éticos, orientação que se traduz em maior exatidão e nas mais baixas taxas de erro em problemas de classificação, nos quais costuma superar os concorrentes em precisão. Tal qualidade, no entanto, vem acompanhada de elevada demanda computacional, cujo custo operacional pode reduzir sua conveniência em cenários de grande escala ou com infraestrutura restrita.

A orientação a que Subahi (2026) atribui esse desempenho não decorre apenas de uma escolha editorial da empresa, mas de um método de alinhamento específico, descrito por Bai *et al.* (2022) sob a denominação de IA Constitucional. Diferentemente do procedimento adotado no ChatGPT, em que o modelo de recompensa se constrói a partir de comparações realizadas por anotadores humanos, a abordagem constitucional substitui grande parte dessa realimentação pela autocrítica do próprio modelo, orientada por um conjunto explícito e público de princípios, a constituição. O modelo gera uma resposta, avalia à luz desses princípios, revisa, e o conjunto das revisões passa a alimentar a etapa de aprendizado por reforço. Segundo os autores, o procedimento reduz a dependência da rotulagem humana de conteúdo nocivo e, sobretudo, torna auditáveis os critérios de comportamento, que passam a estar registrados em texto em vez de permanecerem implícitos nas preferências dos anotadores. Trata-se de atributo relevante em

aplicações do setor público, nas quais a motivação dos atos administrativos é exigência de princípio.

A oferta do Claude organiza-se em camadas sucessivas de capacidade e em gerações que se renovam com frequência. Um dos modelos destacados é o claude-sonnet-4-6, com configuração que procura conciliar a alta qualidade de saída característica do modelo com uma demanda de recursos mais comedida do que a de suas variantes mais robustas, atenuando justamente a sobrecarga operacional assinalada por Subahi (2026).

Segundo Liu *et al.* (2026) identificaram uma de suas versões como a mais exata dentre sete sistemas confrontados, com índices superiores a oitenta por cento em bases de questões técnicas e desempenho notável em coerência semântica e em raciocínio fundamentado em evidências, embora ao preço de um tempo de resposta mais alongado. Hang *et al.* (2026), por sua vez, observaram qualidade informacional equiparável à dos pares e a capacidade de calibrar o nível de linguagem segundo o destinatário.

Referente à elaboração de linguagens específicas de domínio, o Claude evidenciou notável expressividade, introduzindo construções avançadas que viabilizam semânticas mais ricas, ramificações complexas de fluxo de trabalho e respostas guiadas por protocolos e lógicas refinadas de escalonamento (Subahi, 2026). A isso soma-se elevada legibilidade, obtida por meio de uma estrutura hierárquica e de tom narrativo, próxima da linguagem corrente e, por isso, acessível a públicos sem domínio técnico. Tais virtudes, porém, têm como contrapartida menor precisão formal e cobertura incompleta: o modelo tendeu a negligenciar situações de exceção e cenários de erro, além de adotar uma sintaxe de inspiração natural carente do detalhamento formal necessário a uma interpretação inequívoca.

2.2 EFICIÊNCIA NO SETOR PÚBLICO

A partir da década de 1980, houve uma transição da administração pública clássica para uma nova abordagem, impulsionada pela busca de soluções para questões como eficiência, eficácia, efetividade, legitimidade democrática e o impacto das tecnologias da informação. Essa mudança também foi influenciada por avanços em disciplinas como teoria organizacional, ciência política e economia. Com essas novas perspectivas, houve um esforço para deixar de lado as generalizações e

aproveitar a ampla gama de informações sobre a administração pública em diferentes países (Matias-Pereira, 2012).

Esse movimento recebeu, na literatura internacional, a denominação de New Public Management (NPM) e, na literatura brasileira, ficou conhecido como Nova Gestão Pública (NGP), tendo sido sistematizado por Hood (1991) a partir da identificação de sete doutrinas recorrentes nas reformas administrativas do período: a gestão profissional ativa, os padrões e as medidas explícitas de desempenho, a ênfase no controle de resultados em detrimento do controle de procedimentos, a desagregação de grandes unidades, a introdução de competição, a adoção de práticas gerenciais próprias do setor privado e a disciplina no uso dos recursos. Hood (1991) adverte que o movimento articula valores administrativos que nem sempre convivem sem tensão. A busca por economia e parcimônia, não se harmoniza automaticamente com as exigências de honestidade e equidade, nem com as de robustez e resiliência institucional. A ressalva é relevante porque situa a eficiência não como um fim absoluto, mas como um valor entre outros, cuja prevalência precisa ser justificada em cada caso.

Nessa linha tem-se o conceito de Administração Pública Gerencial, que surgiu na segunda metade do século XX como resposta à expansão das funções econômicas e sociais do Estado, além do avanço tecnológico e da globalização. A busca por eficiência na administração pública, visando a redução de custos e aumento da qualidade dos serviços para beneficiar o cidadão, tornou-se essencial. Assim, a reforma do Estado passou a ser guiada por valores de eficiência e qualidade na prestação de serviços, promovendo o desenvolvimento de uma cultura gerencial nas organizações públicas (Matias-Pereira, 2012).

Essa abordagem da administração, amplamente aceita no setor privado, é igualmente aplicável ao setor público, onde sua implementação pode ser ainda mais crucial. No entanto, destaca-se que, na Administração Pública, o objetivo não se limita à redução de insumos ou custos, mas envolve a alocação eficiente desses recursos para maximizar a satisfação do interesse público (Coelho, 2011).

De acordo com o art. 37 da Constituição Federal, a Administração Pública é orientada pelos princípios constitucionais da legalidade, impessoalidade, moralidade e publicidade, que funcionam como diretrizes fundamentais e norteiam a atividade administrativa (Camargo; Guimarães, 2013).

Em 1998, a Emenda Constitucional nº 19, destacou-se por reformar profundamente a administração, redefinindo a relação entre servidores e Estado e introduzindo um pensamento gerencial que se afastava dos valores tradicionais. Entre essas mudanças, o conceito de eficiência foi elevado a um princípio constitucional, determinando o modo e a intensidade que a administração pública deve alcançar seus objetivos, com uma relação adequada entre as medidas adotadas e os fins almejados (Camargo; Guimarães, 2013).

O princípio da eficiência é definido, conforme Moraes (2024, p. 383):

É aquele que impõe à Administração direta e indireta e a seus agentes a persecução do bem comum, por meio do exercício de suas competências de forma imparcial, neutra, transparente, participativa, eficaz, sem burocracia e sempre em busca da qualidade, primando pela adoção dos critérios legais e morais necessários para melhor utilização possível dos recursos públicos, de maneira a evitarem-se desperdícios e garantir-se uma maior rentabilidade social. Note-se que não se trata da consagração da tecnocracia, muito pelo contrário, o princípio da eficiência dirige-se para a razão e fim maior do Estado, a prestação dos serviços públicos essenciais à população, visando a adoção de todos os meios legais e morais possíveis para a satisfação do bem comum.

O Brasil avançou em direção à eficiência na administração pública quando, sob a influência do gerencialismo britânico e dos princípios do New Public Management (NPM) e da "*progressive governance*", instituiu a reforma gerencial por meio do Plano Diretor de Reforma do Aparelho do Estado em 1995. Esse plano tinha como objetivo estabelecer uma forma de gestão pública denominada "administração gerencial", pautada na flexibilidade, resultados, foco no cliente e no controle social. (BRASIL, 1995).

A formulação que sustentou esse plano é apresentada por Bresser-Pereira (1996), responsável por sua condução. O autor distingue três formas históricas de administração (a patrimonialista, a burocrática e a gerencial) e sustenta que a última não rompe integralmente com a anterior. Dela conserva princípios como a admissão por mérito, a existência de um sistema estruturado de carreiras e a avaliação de desempenho, afastando-se, porém, no ponto decisivo do controle, que deixa de incidir a priori sobre os procedimentos e passa a incidir a posteriori sobre os resultados. Bresser-Pereira (1996) argumenta que a administração burocrática, ao presumir a má-fé do administrador e cercar cada ato de formalidades destinadas a preveni-la, revelou-se lenta e cara justamente onde deveria ser confiável, tornando-se inadequada a um Estado cujas funções sociais e econômicas se haviam ampliado. É

desse deslocamento do procedimento para o resultado que decorre o sentido operacional atribuído à eficiência.

A eficiência implica a execução das atribuições com rapidez, qualidade e proficiência, sendo avaliada tanto de forma qualitativa quanto quantitativa. Embora possua diversas definições, a eficiência é essencialmente um conceito econômico, não jurídico, pois orienta a atividade administrativa a buscar os melhores resultados com o menor custo, utilizando os recursos disponíveis (Camargo; Guimarães, 2013).

No campo específico da educação superior pública, a busca por eficiência e por responsabilização materializou-se na institucionalização de processos sistemáticos de avaliação. Dias Sobrinho (2003) situa a avaliação institucional entre a regulação, voltada ao controle e à prestação de contas, e a emancipação, orientada à formação e ao aperfeiçoamento coletivo da instituição. Sob essa perspectiva, o Sistema Nacional de Avaliação da Educação Superior (SINAES) não se reduz a um dispositivo legal, constituindo uma política pública que articula, em um mesmo sistema, a avaliação externa, a avaliação do desempenho dos estudantes e a autoavaliação institucional (Ristoff; Giolo, 2006). Nesse arranjo, a Comissão Própria de Avaliação (CPA) é a instância responsável por conduzir a autoavaliação e por converter seus resultados em subsídios para a gestão.

Quanto aos métodos de autoavaliação, a literatura acadêmica tem-se dedicado tanto aos desafios operacionais das CPAs quanto às técnicas empregadas no tratamento dos dados coletados. Ao analisarem a autoavaliação de uma instituição federal de ensino superior, Reis, Silveira e Ferreira (2010) demonstram que o emprego de técnicas estatísticas, como a análise de agrupamentos, qualifica a interpretação dos resultados e amplia seu uso efetivo na gestão, superando a simples apresentação descritiva de tabelas e gráficos. Andriola (2005), por sua vez, discute os desafios e as necessidades que se apresentam às CPAs para a implementação do SINAES e evidencia que a consolidação de uma cultura avaliativa depende menos do cumprimento formal da norma do que da apropriação dos resultados pela comunidade acadêmica. Tais contribuições reforçam a pertinência de aprimorar os instrumentos de análise dos dados de autoavaliação, questão retomada nas proposições metodológicas desta pesquisa.

2.2.1 Tecnologias no setor público

As tecnologias de informação e comunicação no setor governamental promovem maior eficiência na gestão pública, permitindo acesso a informações em tempo real, integração de sistemas e maior participação cidadã. Isso resulta em um desempenho melhor do governo, tanto em desenvolvimento macroeconômico quanto em um enfoque mais local (Araújo *et. al.*, 2012).

A Tecnologia da Informação (TI) é altamente valorizada nas organizações por oferecer suporte ao planejamento, auxiliar decisões estratégicas e manter a competitividade. No entanto, por exigir especialização e representar custos significativos, é necessário um gerenciamento eficaz para não comprometer o desempenho institucional (Gonçalves *et. al.*, 2020).

Iniciativas de uso de tecnologias da informação e comunicação no setor governamental, conhecidas como tecnologias de governo eletrônico (e-gov), garantem maior eficiência nas atividades públicas e significativa redução de custos (Gonçalves *et. al.*, 2020).

As questões burocráticas poderiam ser totalmente automatizadas e digitalizadas com o uso de tecnologias, permitindo que documentos governamentais sejam processados em tempo real por meio de certificados digitais em portais de governo eletrônico. Essa estrutura melhora a eficiência do governo, que deve focar em prover serviços essenciais, gerir a administração pública e criar condições para o desenvolvimento econômico pelo setor privado. Embora a transformação atual caminhe para a automação do que já existe, as tecnologias atuais têm potencial para sustentar uma nova forma de governo moderno (Gonçalves *et. al.*, 2020).

Segundo Mergel, Edelman e Haug (2019), a distância entre automatizar o que já existe e sustentar uma nova forma de governo pode ser explicada a partir de três fenômenos habitualmente tratados como sinônimos: a digitização, que consiste na mera conversão de conteúdo analógico em formato digital. A digitalização, que reorganiza processos já existentes em torno de meios digitais, sem alterar a lógica. E, por fim, a transformação digital, que envolve a organização como um todo, modificando sua cultura, suas relações com os interessados e a própria concepção dos serviços prestados. Os autores observam que o resultado de curto prazo dessas iniciativas costuma ser medido pelo número de serviços digitais disponibilizados, ao passo que o impacto de longo prazo se expressa em ganhos de efetividade e na satisfação do cidadão, o que explica por que iniciativas tecnicamente bem-sucedidas podem, ainda assim, não produzir a transformação anunciada.

No caso brasileiro, Ribeiro e Segatto (2025), analisando os microdados da Pesquisa TIC Governo Eletrônico 2021, constataam que, embora organizações governamentais brasileiras venham adotando tecnologias e técnicas de inteligência artificial, o movimento permanece restrito a um grupo reduzido de entidades e ocorre de maneira heterogênea entre os âmbitos federal e estadual. As autoras associam essa heterogeneidade às diferentes capacidades em tecnologia da informação instaladas em cada organização, sustentando que a existência de recursos tecnológicos prévios opera como condição para a incorporação da IA ao cotidiano administrativo.

Dentro do IFRS, a estrutura organizacional conta com uma reitoria sediada em Bento Gonçalves e 17 *campi* distribuídos pelo estado do Rio Grande do Sul (IFRS, 2023). A reitoria possui uma equipe de TI e centraliza alguns sistemas de uso comum entre todos os *campi* da instituição, porém, cada campus possui uma equipe de TI própria, que tem autonomia para desenvolvimento e investimento próprio em tecnologia.

Em virtude desta autonomia entre as TIs dos *campi* do IFRS, a instituição possui vários sistemas que não interagem entre si, tornando alguns processos ineficientes e custosos, uma vez que exigem retrabalho por parte dos servidores que executam as tarefas, tendo que informar manualmente alguns dados que poderiam ser obtidos automaticamente de outros sistemas.

2.3 RELAÇÃO ENTRE A IA E O SERVIÇO PÚBLICO

A transformação digital na administração pública, impulsionada pela crescente multiplicação de processos tecnológicos e pela transferência massiva de dados, impõe uma nova configuração à ação pública. Essa mudança envolve novos princípios para a administração, um papel renovado para os agentes públicos e a transição para instituições cada vez mais digitais. Nesse contexto, a inclusão de ferramentas como a inteligência artificial traz vantagens e desafios que devem ser considerados na elaboração e execução de políticas públicas (Valle; Gallo, 2020).

A Administração Pública muitas vezes enfrenta dificuldades para atingir a eficiência desejada devido à falta de pessoal e recursos financeiros, resultando em insatisfação dos cidadãos e dificuldades no controle de fraudes e corrupção. Para superar esses desafios, o uso da Inteligência Artificial é analisado como uma forma

de promover a eficiência administrativa, dado que essa tecnologia oferece benefícios significativos, possibilitando a melhoria das suas atividades por meio de soluções tecnológicas disruptivas (Desordi; Bona, 2020).

O que essas soluções efetivamente compreendem foi sistematizado por Wirtz, Weyerer e Geyer (2019), que identificam dez áreas de aplicação da inteligência artificial no setor público, descrevendo, para cada uma, seu funcionamento e a forma pela qual gera valor público. Os autores agrupam, em paralelo, os obstáculos à sua adoção em quatro conjuntos: os desafios tecnológicos, ligados à qualidade e à interoperabilidade dos dados, os legais e regulatórios. Os éticos, relativos a vieses e à responsabilização por decisões automatizadas, e os sociais, referentes à aceitação pelos cidadãos e aos efeitos sobre o trabalho dos servidores. Esse arcabouço permite ler as iniciativas já em curso no país não como casos isolados, mas como instâncias de categorias determinadas.

Diversos órgãos públicos no Brasil já utilizam sistemas de Inteligência Artificial para melhorar a eficiência operacional. O Tribunal de Contas da União (TCU), por exemplo, emprega os robôs Alice, Sofia e Monica para detectar irregularidades em contratações públicas com recursos federais. A robô Alice já contribuiu para a suspensão de licitações irregulares, evidenciando como a IA pode otimizar, agilizar e tornar mais eficiente o serviço público (Desordi; Bona, 2020).

Outros exemplos que podem ser citados na utilização de inteligência artificial no serviço público são os casos da Advocacia-Geral da União (AGU) com o sistema Sapiens, que emprega inteligência artificial e automação para acelerar o trabalho jurídico dos procuradores, e o programa Victor, no Supremo Tribunal Federal (STF), focado em aumentar a eficiência na análise de processos, economizando tempo e recursos humanos (Valle; Gallo, 2020).

Zuiderwijk, Chen e Salem (2021), em revisão sistemática da produção científica sobre o tema, constata o predomínio de estudos exploratórios, conceituais e qualitativos e a escassez de pesquisas que examinem empiricamente os efeitos concretos da adoção de IA sobre o desempenho das organizações públicas. Os autores propõem, em consequência, uma agenda de pesquisa que enfatiza a necessidade de avaliar os resultados efetivamente obtidos e de considerar as especificidades do contexto institucional em que a tecnologia é implantada.

Reunindo análise bibliográfica, documental e questionário aplicado a servidores e empregados públicos, Toledo e Mendonça (2023) identificam percepção

crescente de ganhos de eficiência e agilidade decorrentes da IA, acompanhada, porém, de despreparo dos próprios servidores para lidar com essas mudanças digitais. Os autores ressaltam que o número reduzido de respondentes limita a validade externa das conclusões, que devem ser lidas como retrato pontual de percepção, e não como generalização ao serviço público. Ainda assim, o achado sugere que a adoção da tecnologia é menos uma questão de disponibilidade de ferramentas do que de preparo institucional.

Atualmente, o IFRS ainda não conta com inteligência artificial integrada aos sistemas desenvolvidos pela instituição. No entanto, existem diversas oportunidades de aplicação dessa tecnologia para aprimorar suas operações e serviços. Por exemplo, a IA poderia ser utilizada para otimizar o atendimento aos alunos por meio de chatbots inteligentes, automatizar processos administrativos complexos, ou analisar grandes volumes de dados para apoiar a tomada de decisões estratégicas. A implementação de soluções baseadas em IA representa uma oportunidade significativa para aumentar a eficiência e a qualidade dos serviços oferecidos pelo IFRS, alinhando a instituição às inovações tecnológicas que têm transformado a administração pública.

3 PROCEDIMENTOS METODOLÓGICOS

Nesse capítulo serão apresentados os procedimentos que norteiam a metodologia deste trabalho. Primeiramente será apresentado o delineamento da pesquisa, seguido pelo campo de pesquisa, coleta de dados, proposta de solução, critérios de avaliação e análise dos resultados.

De acordo com Matias-Pereira (2019, p. 11), tem-se como definição de metodologia científica:

Metodologia científica pode ser definida como o estudo dos métodos ou dos instrumentos necessários para elaborar um estudo científico. A utilização adequada desse elenco de técnicas e processos que integram a metodologia científica permite ao pesquisador obter melhor desempenho e qualidade na produção científica.

A metodologia também pode ser entendida como a lógica do procedimento científico que possui caráter normativo, distinto da teoria substantiva. É difícil separar teoria e método, pois uma investigação sem premissas teóricas não só é infrutífera, mas impossível (Matias-Pereira, 2019).

3.1 DELINEAMENTO DA PESQUISA

De acordo com Sampieri, Collado e Lucio (2013), estudos exploratórios são realizados para investigar temas ou problemas pouco estudados, com muitas incertezas ou que ainda não foram abordados. Esses estudos são indicados quando a literatura disponível apresenta apenas direções não exploradas e ideias vagas, ou quando se busca examinar temas e áreas sob novas perspectivas. Esse é o caso deste estudo, pois a aplicação de inteligência artificial à análise das pesquisas de autoavaliação institucional e à identificação das ações de superação apontadas pelas comissões próprias de avaliação constitui um tema ainda pouco explorado na literatura, aqui examinado sob uma nova perspectiva.

Assim sendo, o presente estudo pode ser classificado como exploratório já que se propõe a explorar os relatórios elaborados manualmente pelas comissões próprias de avaliação e compará-los com os resultados obtidos em nossa pesquisa, com o objetivo de avaliar a convergência entre as ações institucionais identificadas por ambos os métodos.

Já que o objetivo geral desta pesquisa é a elaboração de um sistema que utilize inteligência artificial para identificar as ações institucionais apontadas nas pesquisas de autoavaliação do Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul – campus Bento Gonçalves, optou-se por utilizar a metodologia do Design Science Research (DSR).

De acordo com Dresh, Lacerda e Antunes (2015), a Design Science Research é um método de pesquisa voltado para a solução de problemas, utilizado quando o objetivo é desenvolver um artefato ou prescrição. Esse método busca compreender o problema e construir e avaliar artefatos que promovam melhorias nas condições existentes. É uma abordagem que contribui para reduzir a distância entre teoria e prática nas pesquisas.

A pesquisa baseada na Design Science Research é caracterizada pela orientação à solução de problemas específicos, visando uma solução satisfatória, e não necessariamente a ótima. As soluções devem ser generalizáveis para uma classe de problemas, permitindo que outros pesquisadores e profissionais em diferentes contextos possam utilizar o conhecimento gerado (Dresh; Lacerda; Antunes, 2015).

Manson (2006), propôs um modelo que aponta as saídas que podem ser geradas a partir da realização de cada uma das etapas da DSR. Nesse modelo, o autor explana que após a conscientização do problema, o pesquisador deve apresentar uma proposta contendo evidências, a caracterização do ambiente, métricas de aceitação, e a definição dos atores e classes de problemas relacionados. Na etapa de sugestão, são feitas tentativas para resolver o problema, registrando premissas e escolhas. O desenvolvimento resulta em um ou mais artefatos, com justificativas para as ferramentas e métodos usados, além de formas de validação. Na etapa de avaliação, os artefatos são testados, comparando suas medidas de performance aos requisitos anteriores e detalhando o processo para evitar vieses. Por fim, a conclusão reúne os resultados, sintetiza as aprendizagens e justifica a contribuição do estudo para a classe de problemas definida inicialmente (Dresh; Lacerda; Antunes, 2015).

O modelo de saídas da DSR proposto por Mason (2006), conforme relatado por Dresh, Lacerda e Antunes (2015) pode ser observado na Figura 4:

Figura 4 - Saídas da DSR proposta por Mason (2006)



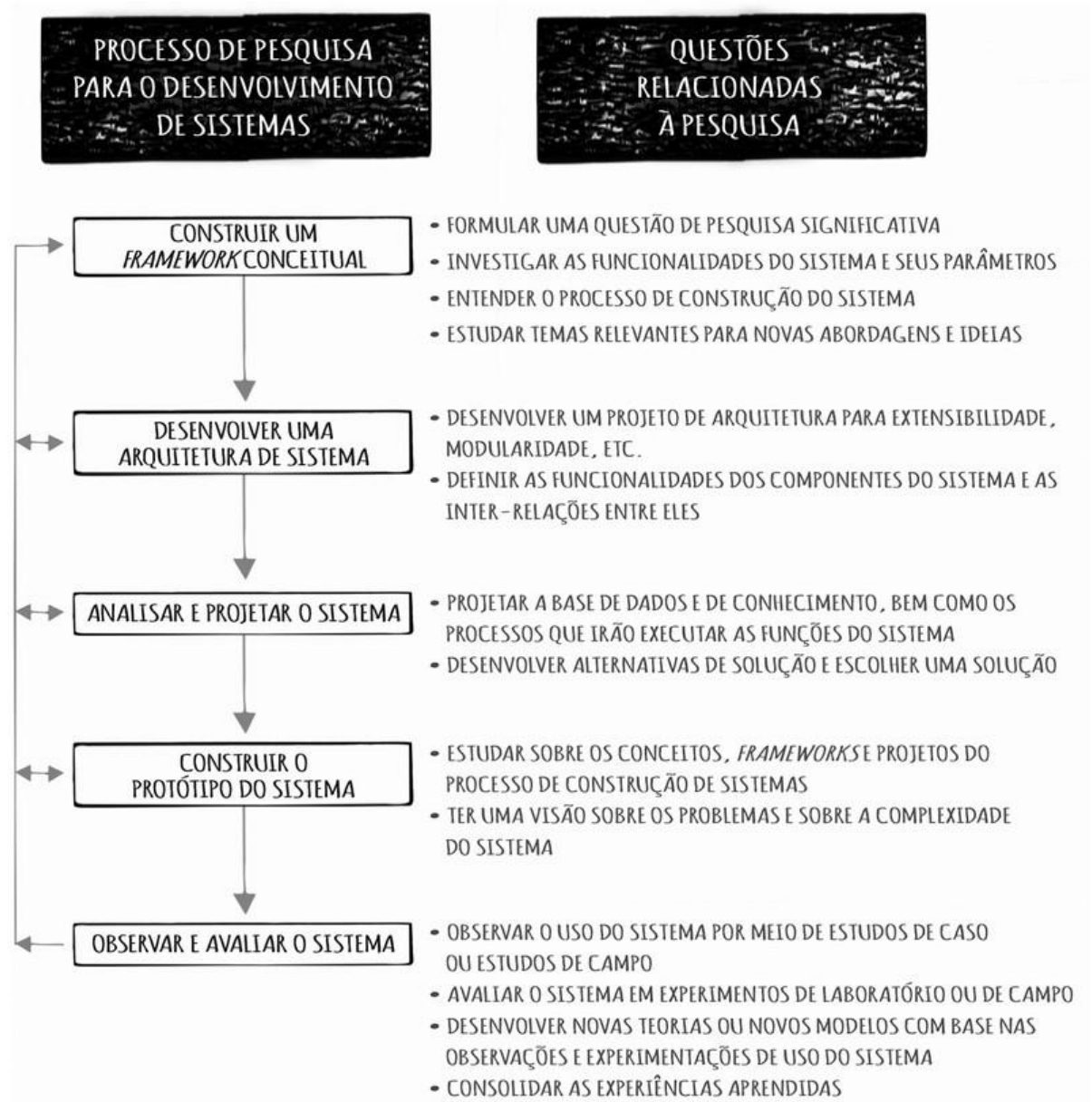
Fonte: Dresh, Lacerda e Antunes (2015, p. 83).

Em 1991, Nunamaker, Chen e Purdin publicaram um texto influente que introduziu a DSR na área de sistemas de informação. Eles propuseram a integração entre o processo de pesquisa tradicional e o desenvolvimento de sistemas, sugerindo uma abordagem multimetodológica que abrange desde a construção de teorias até o desenvolvimento prático de sistemas por meio de experimentação e observação (Dresh; Lacerda; Antunes, 2015).

O processo de pesquisa apresentado começa com a construção de um framework conceitual para justificar a pesquisa e estudar temas relevantes que fomentem novas ideias e abordagens. A segunda etapa, desenvolver a arquitetura do sistema, envolve a apresentação dos componentes do artefato, suas funcionalidades e a definição dos requisitos para avaliação posterior. Na terceira etapa, analisar e projetar o sistema, o pesquisador aplica conhecimento científico para criar e selecionar alternativas de solução. A quarta etapa, construir o protótipo do sistema, permite testar o artefato em condições reais, verificando sua viabilidade e possíveis problemas. Por fim, a etapa de observar e avaliar o sistema visa comprovar sua performance e aplicabilidade, possibilitando a proposição de novas teorias e modelos para apoiar pesquisas futuras.

O modelo do processo da DSR para uma pesquisa em desenvolvimento de sistemas proposto por Nunamaker, Chen e Purdin (1991), de acordo com Dresh, Lacerda e Antunes (2015) pode ser observado na Figura 5:

Figura 5 - Processos da DSR em desenvolvimento de sistemas segundo Nunamaker, Chen e Purdin (1991)



Fonte: Dresh, Lacerda e Antunes (2015, p. 76).

Neste trabalho, foram integradas as propostas de processos da DSR apresentadas por Dresh, Lacerda e Antunes (2015) utilizando o modelo de saídas

proposto por Mason (2006) e a proposta de processos de desenvolvimento de software proposta por Nunamaker, Chen e Purdin (1991). Na primeira etapa, que compreende a proposta, foi delineado o objetivo geral deste estudo, que consiste na elaboração de um sistema baseado em inteligência artificial para identificar ações institucionais apontadas em pesquisas de autoavaliação. A etapa seguinte, denominada tentativa, utilizou a abordagem proposta por Nunamaker, Chen e Purdin (1991), com o desenvolvimento de um protótipo funcional do sistema e a realização de sua avaliação funcional.

Na terceira etapa, de desenvolvimento, foi finalizada a criação do sistema, considerando as aprovações e refinamentos obtidos na fase anterior, resultando no artefato concluído. A quarta etapa, chamada de avaliação, teve como foco a comparação das ações identificadas pelo sistema desenvolvido com as ações registradas nos relatórios da comissão própria de avaliação, elaborados manualmente em seus relatórios anuais. Por fim, na etapa de conclusão, foram apresentados os resultados obtidos tanto no desenvolvimento quanto na avaliação do artefato, respondendo à questão central da pesquisa: "Como a inteligência artificial contribui para o apontamento de ações de superação identificadas na coleta de informações?".

Para ilustrar a unificação dos passos deste trabalho de acordo com os processos da DSR apresentados por Dresh, Lacerda e Antunes (2015), foi criada a Figura 6. Nela pode-se notar que no processo de desenvolvimento do protótipo do sistema proposto por Mason (2006), utilizando os processos de desenvolvimento de software propostos por Nunamaker, Chen e Purdin (1991).

Figura 6 - Saídas da DSR proposta na pesquisa

Fonte: autoria própria (2026).

Desta forma, de acordo com o exposto este trabalho tem como delineamento uma pesquisa exploratória com a utilização do Design Science Research (DSR) como metodologia.

3.1.1 Sistema desenvolvido

Para operacionalizar o artefato apresentado, foi desenvolvido um sistema de informação denominado Auto Report, concebido especificamente para receber os dados das pesquisas de autoavaliação institucional e, a partir deles, gerar de forma automatizada o apontamento das ações institucionais. A construção desse sistema corresponde à materialização do artefato previsto pelo Design Science Research, etapa em que o conhecimento teórico discutido no referencial é convertido em uma solução prática e passível de avaliação. O sistema foi projetado como uma aplicação web, opção que se justifica pela necessidade de acesso multiplataforma pelos membros das comissões próprias e subcomissões de avaliação, dispensando a instalação de softwares locais e permitindo a utilização a partir de qualquer dispositivo conectado à rede da instituição.

A opção por uma arquitetura web também se mostra coerente com o cenário tecnológico do IFRS, no qual a reitoria centraliza determinados sistemas de uso comum, enquanto cada campus mantém autonomia de desenvolvimento. Por ser independente do sistema de coleta atualmente em operação, o Auto Report pôde ser construído sem demandar alterações na infraestrutura existente, recebendo os dados por meio de importação de arquivos e preservando, assim, a separação entre o ambiente de coleta e o ambiente de análise. Essa decisão de projeto reduz riscos de interferência nos sistemas oficiais da instituição e favorece a reprodutibilidade da solução em outros *campi*.

No que se refere às tecnologias empregadas, o núcleo do sistema foi desenvolvido na linguagem PHP, em sua versão 8.2. A escolha do PHP fundamenta-se em sua ampla utilização no desenvolvimento de aplicações web e em sua natureza de linguagem de script do lado servidor, interpretada no servidor e geradora de saída em HTML, conforme caracterizado por Sebesta (2018). Além dos aspectos técnicos, a escolha considerou o contexto institucional em que o sistema será mantido, uma vez que a equipe de Tecnologia da Informação do campus Bento Gonçalves do IFRS possui ampla familiaridade com o PHP, linguagem empregada na maioria dos sistemas desenvolvidos internamente na instituição. Essa convergência tecnológica favorece a manutenção, a continuidade e a incorporação do artefato à infraestrutura já existente, reduzindo a curva de aprendizado e a dependência de competências externas. A versão 8.2 foi adotada por incorporar recursos modernos de tipagem, melhorias de desempenho e suporte ativo, aspectos relevantes para a manutenibilidade e a segurança de um sistema destinado ao tratamento de dados institucionais. Soma-se a isso a maturidade do ecossistema da linguagem, que oferece bibliotecas consolidadas para integração com serviços externos por meio de interfaces de programação de aplicações (APIs), recurso essencial para a comunicação com os modelos de inteligência artificial utilizados na pesquisa.

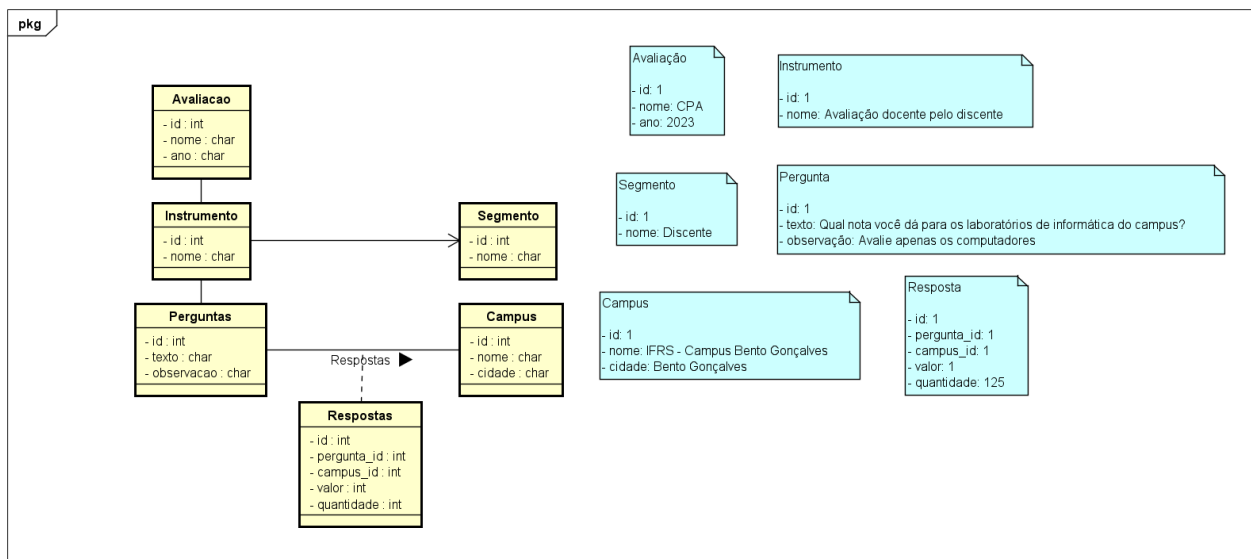
Sobre a linguagem PHP, foi adotado o *framework* Laravel, em sua versão 8. Um *framework* consiste em um conjunto estruturado de bibliotecas e convenções que padroniza o desenvolvimento e acelera a construção de aplicações, reduzindo a quantidade de código escrito manualmente pelo programador. O Laravel organiza a aplicação segundo o padrão arquitetural *Model-View-Controller* (MVC), que separa a camada de dados (*models*), a camada de apresentação (*views*) e a camada de controle de fluxo (*controllers*). Essa separação de responsabilidades favorece a

organização do código, a extensibilidade e a modularidade do sistema, atributos destacados por Nunamaker, Chen e Purdin (1991) como desejáveis no desenvolvimento de artefatos computacionais. Na prática, a camada de modelos do Laravel foi utilizada para representar as entidades do domínio da autoavaliação, enquanto os controladores concentraram a lógica de importação de dados, de comunicação com os modelos de inteligência artificial e de geração dos documentos de saída.

Para a persistência das informações, o sistema utiliza o sistema gerenciador de banco de dados MySQL, de natureza relacional. O MySQL foi selecionado por sua ampla adoção, gratuidade, confiabilidade e pela integração nativa com o *framework* Laravel, que dispõe de uma camada de mapeamento objeto-relacional capaz de traduzir as operações sobre os objetos do sistema em comandos de banco de dados. O modelo relacional mostrou-se adequado ao problema investigado, uma vez que os dados da autoavaliação apresentam estrutura bem definida e relacionamentos claros entre avaliações, instrumentos, perguntas, segmentos, *campi* e respostas, o que permite a organização das informações em tabelas interligadas por chaves primárias e estrangeiras.

A modelagem do sistema foi representada por meio de um diagrama de classes, apresentado na Figura 7. O diagrama foi elaborado durante a etapa de análise e projeto do sistema e orientou tanto a estruturação do banco de dados quanto a definição das entidades manipuladas pela aplicação. Junto a cada classe, o diagrama apresenta um objeto de exemplo, ilustrando como os atributos são efetivamente preenchidos com dados reais da autoavaliação conduzida pela CPA do campus Bento Gonçalves.

Figura 7 – Diagrama de classes



Fonte: autoria própria (2026).

Conforme se observa na Figura 7, a classe “Avaliacao” representa cada ciclo avaliativo conduzido pela comissão e é composta pelos atributos *id*, *nome* e *ano*, como ilustrado pelo objeto de exemplo referente à avaliação da CPA no ano de 2023. Cada avaliação agrega um ou mais instrumentos, modelados pela classe “Instrumento”, formada pelos atributos *id* e *nome*. O instrumento corresponde ao questionário aplicado a determinado público, como, por exemplo, o instrumento de avaliação docente respondido pelos discentes. Cada instrumento, por sua vez, está associado a um “Segmento”, classe composta pelos atributos *id* e *nome*, que identifica o público respondente (discente, docente ou técnico-administrativo) e que materializa, no modelo de dados, a subdivisão da comunidade interna descrita na delimitação da amostra.

A classe “Perguntas” reúne as questões que compõem cada instrumento e é definida pelos atributos *id*, *texto* e *observacao*, sendo o campo de observação destinado a orientações complementares de preenchimento, como no exemplo em que se solicita ao respondente avaliar apenas os computadores dos laboratórios de informática. As respostas obtidas são representadas pela classe “Respostas”, vinculada às perguntas e composta pelos atributos *id*, *pergunta_id*, *campus_id*, *valor* e *quantidade*. Cabe destacar que essa classe armazena os dados de forma agregada. Em vez de registrar cada resposta individual, o sistema contabiliza, para cada

pergunta e cada campus, quantas vezes determinado valor da escala Likert foi assinalado, conforme indicado pelo objeto de exemplo, em que o valor 1 foi escolhido 125 vezes. Essa opção de modelagem é coerente com a natureza dos dados utilizados, já que, conforme descrito nos procedimentos de coleta, as informações são extraídas dos relatórios já resumidos e anonimizados, dispensando o tratamento de respostas brutas e individualizadas.

Por fim, a classe “Campus”, formada pelos atributos *id*, nome e cidade, identifica a unidade institucional à qual as respostas se referem, como o IFRS – Campus Bento Gonçalves. Os relacionamentos entre as classes refletem a lógica do instrumento de autoavaliação: uma avaliação contém instrumentos, cada instrumento dirige-se a um segmento e agrega perguntas, e cada pergunta acumula respostas associadas a um campus. Essa estrutura confere ao sistema a capacidade de organizar e recuperar os dados por eixo, por segmento e por período, o que é fundamental para a posterior comparação entre as ações apontadas pela inteligência artificial e aquelas registradas manualmente pela comissão.

O diferencial do sistema reside em sua integração com modelos de inteligência artificial generativa, responsáveis pela proposição das ações institucionais a partir dos dados armazenados. Conforme apontado por Gallardo, Machuca e Semblantes (2023), é possível integrar diferentes linguagens de programação às APIs dos modelos generativos existentes, desde que realizadas as configurações adequadas e utilizadas as ferramentas apropriadas. No Auto Report, essa integração foi implementada por meio de requisições às APIs de cada provedor, nas quais o sistema envia um comando estruturado, contendo os dados quantitativos de cada pergunta, e recebe como retorno o texto com as ações sugeridas. A construção desses comandos apoia-se na prática de engenharia de prompt descrita por Ramos (2023), que busca tornar as respostas mais precisas e a interação mais eficiente por meio da elaboração cuidadosa das instruções enviadas ao modelo.

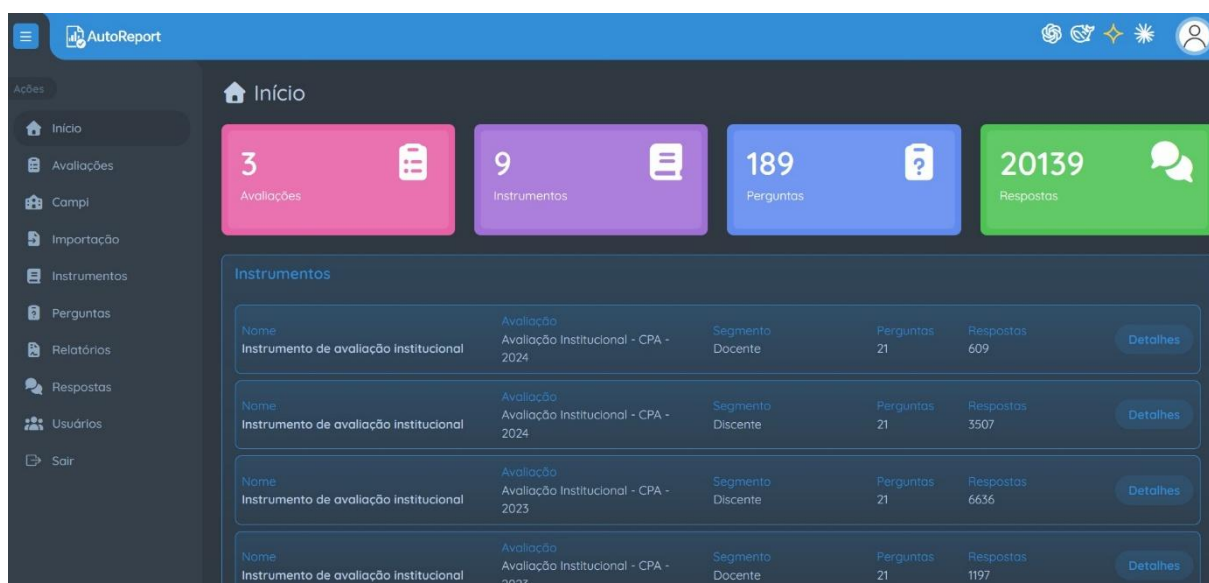
Com o objetivo de comparar o desempenho de diferentes soluções e não restringir os resultados a um único fornecedor, o sistema foi integrado a quatro modelos distintos de inteligência artificial. O primeiro deles é o ChatGPT, desenvolvido pela OpenAI, por meio do modelo *gpt-5.5*, reconhecido por sua ampla aplicabilidade a tarefas de processamento de linguagem natural. O segundo é o DeepSeek, da DeepSeek AI, na configuração *deepseek-chat*, voltada ao diálogo de propósito geral e caracterizada pela economia computacional proporcionada por sua arquitetura de

mistura de especialistas, conforme descrito por Subahi (2026). O terceiro modelo é o Gemini, do Google, na variante *gemini-3.5-flash*, orientada à eficiência, com baixa latência e ampla escalabilidade de implantação. O quarto modelo é o Claude, da Anthropic, na configuração *claude-sonnet-4-6*, que procura conciliar a alta qualidade de saída característica do modelo com uma demanda de recursos mais comedida do que a de suas variantes mais robustas.

A opção por integrar simultaneamente os quatro modelos atende a um propósito metodológico central da pesquisa, submeter o mesmo conjunto de dados e o mesmo comando a sistemas distintos, permitindo avaliar comparativamente a aderência das ações sugeridas às ações efetivamente apontadas pela comissão, evidenciando eventuais diferenças de precisão, de cobertura e de pertinência entre as soluções. Dessa forma, o sistema não se limita a empregar a inteligência artificial como ferramenta, mas constitui-se também em um ambiente de experimentação que viabiliza a análise comparativa prevista na etapa de avaliação do artefato.

Quanto ao comportamento do sistema, o Auto Report organiza-se em torno de um painel principal que centraliza o acesso às suas funcionalidades, apresentado na Figura 8. A interface foi concebida de modo a oferecer uma visão consolidada dos dados carregados e atende ao objetivo específico de propor uma interface intuitiva para a utilização do sistema, facilitando o trabalho dos membros da comissão que não possuem, necessariamente, formação técnica em tecnologia da informação.

Figura 8 – Tela inicial do sistema Auto Report



Fonte: autoria própria (2026).

Como se verifica na Figura 8, a tela inicial exibe indicadores sintéticos do volume de dados sob análise. No exemplo apresentado, três avaliações, quatro instrumentos, setenta e cinco perguntas e mais de dezessete mil respostas, além de uma listagem dos instrumentos cadastrados, com a identificação da avaliação, do segmento, do número de perguntas e do número de respostas associadas a cada um. O menu lateral organiza o acesso às entidades do sistema: Avaliações, Campi, Instrumentos, Perguntas, Respostas e Relatórios; refletindo, na navegação, a mesma estrutura definida no diagrama de classes. Essa correspondência direta entre o modelo de dados e a interface contribui para a coerência do sistema e para a previsibilidade de seu uso.

O fluxo de utilização do sistema inicia-se com a exportação dos dados a partir do sistema de autoavaliação em operação no IFRS, no formato CSV, seguida de sua importação para o Auto Report. Como o sistema oficial não permite conexões externas diretas ao seu banco de dados, a transferência por arquivo mostra-se necessária e, ao mesmo tempo, possibilita a padronização dos dados conforme o formato esperado pela aplicação. Uma vez importadas as informações, o usuário aciona, pela interface, a funcionalidade de geração do apontamento de ações, ocasião em que o sistema percorre as perguntas, monta os comandos correspondentes e os submete ao modelo de inteligência artificial configurado pelo usuário. Ao final do processamento, é disponibilizado para *download* um documento contendo as ações sugeridas, com a identificação do perfil dos avaliadores e a apresentação gráfica dos resultados.

Esse documento gerado pelo sistema constitui o insumo a ser confrontado com os relatórios elaborados manualmente pela comissão, conforme descrito nos procedimentos de análise de dados. Desse modo, o comportamento do Auto Report encerra o ciclo previsto no artefato proposto, ao receber os dados da autoavaliação, processar essas informações por meio de inteligência artificial e devolver, de forma estruturada e em tempo substancialmente inferior aos quatro meses atualmente demandados, um conjunto de ações institucionais que pode subsidiar o trabalho das comissões próprias de avaliação dos campi do IFRS.

3.2 CAMPO DE ESTUDO E AMOSTRA

O objetivo geral desta dissertação é desenvolver um sistema baseado em inteligência artificial capaz de identificar ações institucionais destacadas nas pesquisas de autoavaliação realizadas pelo Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul (IFRS) – campus Bento Gonçalves. A escolha do IFRS - campus Bento Gonçalves como campo de estudo, apesar da autoavaliação institucional ser aplicada anualmente em todos os campi do IFRS por suas respectivas comissões próprias de avaliação, justifica-se pelo contato mais próximo do pesquisador com a comissão avaliadora do referido campus. Essa proximidade possibilita uma análise mais detalhada e criteriosa das ações identificadas pelo sistema, favorecendo a compreensão e avaliação dos resultados obtidos.

Para compor a amostra utilizada na avaliação dos resultados da pesquisa e comparação das ações sugeridas pelo sistema, foram analisadas as avaliações institucionais realizadas pela CPA, correspondentes aos anos de 2023, 2024 e 2025.

A autoavaliação institucional conduzida pela Comissão Própria de Avaliação (CPA) do IFRS não se restringe a um único instrumento de coleta. De maneira a contemplar os diferentes segmentos e perspectivas que compõem a realidade do campus, a CPA emprega, em periodicidade anual, seis instrumentos avaliativos distintos: a Autoavaliação da Comunidade Interna; a Autoavaliação do Curso; a Autoavaliação Discente; a Avaliação pela Comunidade Externa; a Avaliação Docente; e a Avaliação dos Egressos. Cada instrumento dirige-se a um público específico e capta dimensões particulares da atuação institucional, de modo que a triangulação dessas fontes permite traçar um retrato mais abrangente do campus em determinado momento.

No que se refere às dimensões avaliadas, os instrumentos estruturam-se a partir das exigências previstas na Lei nº 10.861/2004, que institui o Sistema Nacional de Avaliação da Educação Superior (SINAES). Tais dimensões são organizadas em cinco eixos: o Eixo 1, de Planejamento e Avaliação Institucional; o Eixo 2, de Desenvolvimento Institucional; o Eixo 3, de Políticas Acadêmicas; o Eixo 4, de Políticas de Gestão; e o Eixo 5, de Infraestrutura Física. Esses eixos orientam tanto a formulação das questões respondidas pela comunidade quanto a posterior organização e análise dos resultados no relatório produzido pela CPA.

Para os fins desta dissertação, e considerando a natureza do sistema a ser desenvolvido, optou-se por delimitar a análise exclusivamente às avaliações da comunidade interna. Essa delimitação justifica-se por dois motivos: primeiro, porque

a comunidade interna constitui o público com maior representatividade e regularidade de participação no processo avaliativo; segundo, porque suas respostas concentram-se em um conjunto comum e padronizado de questões, o que favorece o tratamento automatizado e a comparação dos resultados ao longo dos anos de 2023, 2024 e 2025.

A comunidade interna, por sua vez, subdivide-se em três segmentos: o segmento docente, o segmento discente e o segmento dos técnicos administrativos. A esses três segmentos é submetido um conjunto de 21 questões, formuladas como afirmativas e respondidas segundo uma escala do tipo Likert de cinco pontos (concordo totalmente, concordo, não concordo nem discordo, discordo e discordo totalmente). As questões distribuem-se entre os cinco eixos do SINAES, conforme apresentado no Quadro 1.

Quadro 1 – Distribuição de perguntas por eixo

Eixo	Pergunta
Eixo 1 – Planejamento e Avaliação Institucional	Os resultados do processo de Avaliação Institucional são divulgados à comunidade acadêmica?
	A instituição utiliza os resultados do processo de Avaliação Institucional para realizar ajustes e melhorias em seu planejamento?
Eixo 2 – Desenvolvimento Institucional	A missão, os valores e a visão da Instituição são conhecidos e aplicados nas atividades cotidianas?
	A Instituição garante a inclusão social das pessoas com necessidades específicas em todos os níveis?
Eixo 3 – Políticas Acadêmicas	A Instituição oferece a possibilidade de participar dos processos de discussão para construção e/ou reformulação de propostas de cursos?
	A Instituição oferece a possibilidade de participação em projetos (ensino, pesquisa, extensão) que integre docentes, discentes e técnicos-administrativos?
	O Portal do IFRS (ifrs.edu.br) fornece, com clareza e agilidade, informações sobre o Instituto e o funcionamento da instituição?

	O site do campus apresenta informações sobre as atividades de ensino, pesquisa e extensão do IFRS à comunidade externa?
	Os meios de comunicação utilizados pelo IFRS são eficazes para divulgar as atividades da instituição?
	Os meios de comunicação utilizados pelo campus são eficazes para divulgar as atividades da instituição?
	A Instituição possui políticas bem definidas para ingresso de estudantes?
	A Instituição possui políticas bem definidas para a permanência e êxito dos estudantes?
Eixo 4 – Políticas de Gestão	A Instituição fomenta a qualificação dos servidores, visando o aprimoramento de suas atividades?
	A Instituição oferece a possibilidade de participar de Conselhos, Comissões, Colegiados e/ou Grupos de Trabalho no IFRS?
	A Instituição divulga seu regimento, portarias, resoluções, ordens de serviço e demais regulamentações do IFRS?
Eixo 5 – Infraestrutura Física	A biblioteca possui acervo virtual e/ou plataformas de pesquisas adequadas de acordo com as necessidades dos cursos?
	As salas de aula apresentam infraestrutura física e tecnológica adequada ao número de estudantes?
	Os serviços de manutenção (higienização, segurança, etc.) atendem às necessidades do Campus?
	Os servidores e estudantes possuem infraestrutura e local adequado para a realização de suas atividades?
	Os docentes possuem local adequado para a realização de atendimentos aos discentes?
	O campus oferece acesso satisfatório à internet?

Fonte: autoria própria (2026).

Cabe ressaltar que, embora respondidas pelos três segmentos, algumas questões possuem maior aderência a um público específico, como aquelas relativas ao atendimento docente ou à participação em projetos, o que será considerado no momento da análise e do confronto entre as ações sugeridas pelo sistema baseado

em inteligência artificial e aquelas identificadas nos relatórios elaborados manualmente pela CPA.

3.3 PROCEDIMENTOS DE COLETA DE DADOS

As técnicas de coleta e análise de dados são essenciais para a implementação dos métodos de pesquisa e do plano de trabalho do pesquisador. A escolha da técnica requer uma reflexão sobre os dados desejados, sua obtenção e fontes. Além disso, o pesquisador deve considerar a comunidade científica com a qual pretende dialogar, respeitando seus critérios e procedimentos para coleta e análise de dados (Dresh; Lacerda; Antunes, 2015).

Em estudos que envolvem a análise de dados e informações, o investigador deve escolher uma técnica de coleta de dados adequada para o desenvolvimento e conclusão da pesquisa (Martins, 2008).

Entre as várias técnicas de coleta de dados, uma delas se encaixa perfeitamente no estudo proposto, a análise de conteúdo. É uma técnica objetiva, sistemática e quantitativa para estudar e analisar a comunicação, permitindo inferências confiáveis a partir de discursos escritos ou orais em diversos contextos, como programas de mídia, artigos, livros, discursos e regulamentos. Ela pode ser usada tanto para fins exploratórios quanto de verificação.

No caso desta pesquisa, foi utilizada a técnica de análise de conteúdo com base nos relatórios anuais publicados anualmente pela comissão própria de avaliação (CPA) do IFRS – campus Bento Gonçalves. Nestes relatórios, são apresentados alguns resumos dos dados obtidos nas avaliações e são propostas ações institucionais identificadas nas respostas obtidas.

A análise de conteúdo, conforme proposta por Bardin (2016), organiza-se em três etapas fundamentais: a pré-análise, na qual se realiza a organização do material e a leitura inicial do corpus; a exploração do material, momento em que os dados são codificados e categorizados; e o tratamento dos resultados, etapa em que se procede à inferência e à interpretação. Esta pesquisa adota essas etapas como referência tanto para o exame dos relatórios elaborados manualmente quanto para a avaliação dos resultados produzidos pelo sistema, garantindo que ambos os conjuntos de dados sejam submetidos a um mesmo enquadramento analítico.

O corpus de análise é composto pelos relatórios anuais de autoavaliação institucional publicados pela CPA do IFRS – campus Bento Gonçalves, referentes aos anos de 2023, 2024 e 2025. A delimitação recai exclusivamente sobre as avaliações da comunidade interna, abrangendo os três segmentos que a compõem (docente, discente e técnico-administrativo) e o conjunto de 21 questões respondidas, distribuídas entre os cinco eixos do SINAES: Planejamento e Avaliação Institucional; Desenvolvimento Institucional; Políticas Acadêmicas; Políticas de Gestão; e Infraestrutura Física. Esses eixos funcionam, na presente pesquisa, como categorias de análise previamente definidas, o que permite classificar e organizar de maneira sistemática as ações institucionais identificadas por cada método.

No que se refere aos dados gerados manualmente, a análise parte das ações institucionais já destacadas nos relatórios da CPA. Nesses documentos, a comissão sintetiza os resultados das respostas obtidas, em geral apresentados de forma quantitativa, por meio dos percentuais de concordância e discordância de cada questão e, a partir dessa leitura, propõe ações institucionais consideradas relevantes, frequentemente reunidas nas seções de superação e de avaliação qualitativa. A análise de conteúdo, nesse caso, busca extrair e categorizar essas ações segundo os eixos correspondentes, registrando quais aspectos foram efetivamente apontados pela comissão e a que segmento e dimensão se vinculam.

Quanto aos dados gerados automaticamente, o sistema baseado em inteligência artificial é alimentado com as mesmas respostas da comunidade interna que originaram os relatórios manuais, produzindo, de modo autônomo, a identificação de ações institucionais destacadas. Aplica-se, sobre esse produto, o mesmo procedimento de codificação e categorização por eixo, de forma que as ações sugeridas pelo sistema sejam organizadas em uma estrutura comparável àquela utilizada na análise dos relatórios da CPA.

A análise final consiste no confronto entre os dois conjuntos de resultados. Essa comparação permite verificar o grau de convergência entre as ações identificadas manualmente e aquelas sugeridas pelo sistema, bem como evidenciar eventuais divergências, isto é, ações destacadas exclusivamente por um dos métodos. A comparação é conduzida em cada eixo, observando-se aspectos como a quantidade de ações identificadas, a aderência das ações às questões e aos segmentos avaliados, e a pertinência das sugestões frente ao contexto institucional. Nesse ponto, a proximidade do pesquisador com a comissão avaliadora do campus,

mencionada anteriormente, assume papel central, ao possibilitar uma avaliação qualitativa e criteriosa da relevância das ações apontadas pelo sistema, contribuindo para verificar em que medida a abordagem automatizada é capaz de reproduzir, complementar ou aprimorar o trabalho atualmente realizado de forma manual.

3.4 PROCESSO DE ANÁLISE DE DADOS

A análise e interpretação dos dados dependem da natureza dos documentos. No caso de documentos de segunda mão, como relatórios de empresas e órgãos governamentais que já passaram por tratamento analítico, os procedimentos são semelhantes aos utilizados em pesquisas bibliográficas (Gil, 2022).

Diferentes modelos de análise de conteúdo têm sido propostos e devem abordar seis questões essenciais: os dados que serão analisados, sua definição, a população de origem, o contexto de análise, os limites da análise e o objetivo das inferências (Gil, 2022).

Nesta pesquisa, utilizou-se os dados coletados nas pesquisas de autoavaliação realizadas previamente no IFRS – campus Bento Gonçalves. Estes dados foram retirados dos relatórios elaborados pela comissão própria de avaliação da instituição e em seguida, serão armazenados em um banco de dados montado para o desenvolvimento do sistema com inteligência artificial proposto.

Não foi necessária uma limpeza dos dados, uma vez que os mesmos por serem removidos diretamente dos relatórios já estão resumidos e anônimos, sem dados com “lixo”, caso fossem coletados diretamente através de uma aplicação diretamente com a população do campus.

Da mesma forma, os dados utilizados como apontamento de ações pela CPA do IFRS – campus Bento Gonçalves foram retirados dos relatórios publicados anualmente pela comissão, sem a necessidade de tratamento dos mesmos para a utilização nesta pesquisa.

Nessa fase de comparação das ações indicadas pelo sistema com aquelas apontadas pelos relatórios da CPA utilizou o método comparativo para a conclusão dos resultados. O método comparativo permite a análise de dados concretos e a identificação de elementos constantes e gerais, funcionando como uma experimentação indireta. O método é útil tanto para estudos abrangentes quanto para análises específicas, e pode ser empregado em todas as fases de uma investigação,

desde a descrição e classificação até a explicação de fenômenos e a identificação de vínculos causais entre fatores (Marconi; Lakatos, 2022).

Para viabilizar a comparação entre os dois conjuntos de resultados, adotou-se como unidade de análise a "ação institucional destacada", entendida como cada apontamento de melhoria, fragilidade ou potencialidade extraído das respostas da comunidade interna. Tanto as ações registradas manualmente nos relatórios da CPA quanto aquelas sugeridas pelo sistema baseado em inteligência artificial são tratadas a partir dessa mesma unidade, o que assegura que a comparação ocorra sobre elementos equivalentes e comparáveis.

Cada ação institucional, independentemente de sua origem, é classificada segundo três descritores comuns: o eixo do SINAES a que se vincula (Planejamento e Avaliação Institucional; Desenvolvimento Institucional; Políticas Acadêmicas; Políticas de Gestão; e Infraestrutura Física); a questão ou questões do instrumento às quais se relaciona, dentre as 21 que compõem a avaliação da comunidade interna; e o segmento de origem das respostas (docente, discente ou técnico-administrativo). Essa estrutura de categorização, definida previamente, organiza os dados de ambas as fontes em um mesmo padrão e permite que a comparação seja realizada de maneira sistemática, eixo a eixo.

A comparação propriamente dita é conduzida com o apoio de uma matriz comparativa, na qual as ações apontadas pela CPA e as sugeridas pelo sistema são dispostas lado a lado dentro de cada categoria. Sobre essa matriz, aplica-se um critério de correspondência que classifica cada ação em três níveis: correspondência total, quando a ação identificada pelo sistema coincide com aquela registrada pela comissão; correspondência parcial, quando há aproximação temática, mas com diferenças de abrangência, foco ou formulação; e ausência de correspondência, quando a ação é destacada por apenas um dos métodos. Esse critério confere objetividade ao confronto e torna o procedimento replicável.

A partir dessa classificação, a análise assume uma dimensão quantitativa e outra qualitativa, que se complementam. Na dimensão quantitativa, contabiliza-se, por eixo, o número de ações identificadas por cada método e o número de correspondências totais, parciais e ausentes, o que permite mensurar o grau de convergência entre as abordagens e evidenciar em quais eixos ela é maior ou menor. Na dimensão qualitativa, examinam-se as divergências, sobretudo as ações destacadas exclusivamente por um dos métodos, buscando compreender se

representam lacunas, redundâncias ou contribuições efetivamente novas trazidas pela abordagem automatizada.

É nesse exame qualitativo que se evidencia a pertinência do método comparativo descrito por Marconi e Lakatos (2022), uma vez que a identificação de elementos constantes e gerais entre as duas fontes funciona como uma forma de experimentação indireta, permitindo inferir em que medida o sistema reproduz a lógica de análise da comissão. A proximidade do pesquisador com a CPA do campus, mencionada na delimitação do estudo, cumpre aqui papel determinante, ao possibilitar a avaliação criteriosa da relevância das ações sugeridas pelo sistema e a verificação de sua aderência ao contexto institucional.

Por fim, a aplicação desse procedimento aos três anos que compõem a amostra permite não apenas uma comparação pontual, mas também a observação da consistência dos resultados ao longo do tempo. Dessa forma, torna-se possível concluir se o sistema baseado em inteligência artificial é capaz de reproduzir, complementar ou aprimorar o trabalho de identificação de ações atualmente realizado de forma manual pela comissão própria de avaliação.

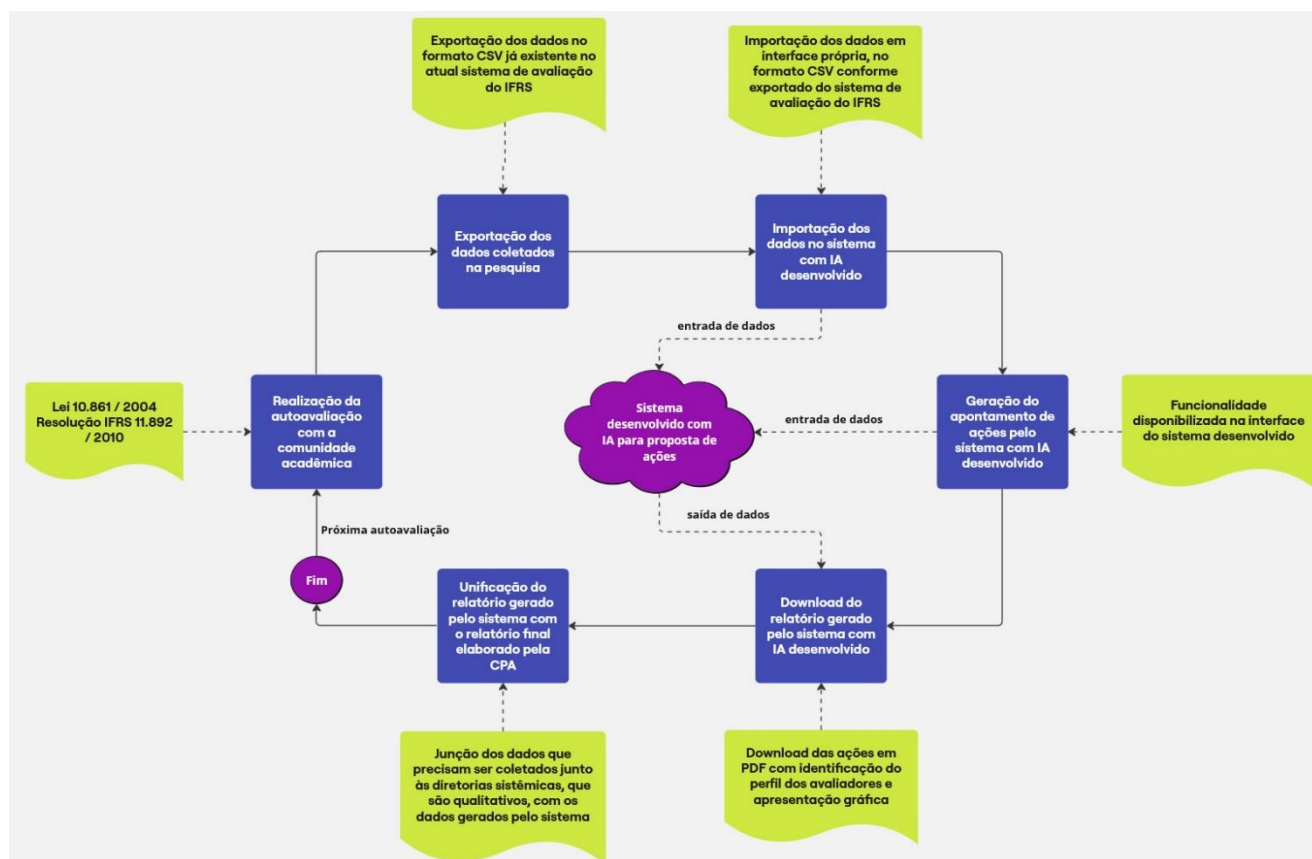
3.5 ARTEFATO PROPOSTO

Um artefato, é um objeto artificial concebido pelo homem, que, apesar de sua natureza projetada, está sujeito às leis naturais das ciências tradicionais. Ele pode ser analisado sob dois aspectos: descritivo, relacionado à comunicação e detalhamento de seus componentes, e imperativo, focado nas questões normativas de sua construção e aplicação (Dresh; Lacerda; Antunes, 2015).

A eficácia de um artefato depende da relação entre seu propósito, suas características internas e o ambiente externo em que opera, configurando-o como uma interface entre esses dois ambientes, com seus componentes organizados para alcançar objetivos específicos (Dresh; Lacerda; Antunes, 2015).

Nesta pesquisa é proposto o modelo de artefato apresentado na Figura 9, que representa a relação entre as etapas da conclusão de uma autoavaliação institucional, onde são apontadas ações institucionais com o apoio do sistema desenvolvido com a utilização de inteligência artificial. Neste artefato, também são apresentadas as relações entre o sistema utilizado atualmente para coleta das avaliações da comunidade do IFRS e o sistema com IA proposto nessa pesquisa.

Figura 9 - Artefato proposto na pesquisa



Fonte: autoria própria (2026).

De forma detalhada o funcionamento do artefato proposto nesta pesquisa pode ser explicado inicialmente pela montagem do banco de dados utilizado para as propostas de ações institucionais. Após a realização da pesquisa de avaliação institucional utilizando o sistema já em operação do IFRS, é necessário exportar as informações através de consultas personalizadas diretas ao banco de dados utilizado.

Depois da exportação destes dados, é necessária a realização de uma importação ao banco de dados utilizado pelo sistema desenvolvido com inteligência artificial, denominado nesta pesquisa de Auto Report, já que o sistema utilizado para autoavaliações no IFRS não permite conexões de sistemas externos ao banco de dados. Com isso, foi possível corrigir alguns padrões esperados pela IA para o funcionamento correto no apontamento das ações.

Assim que os dados coletados estiveram disponíveis para a utilização pelo sistema desenvolvido com inteligência artificial, foi disparada uma ação dentro do

sistema que permitiu a geração de um documento com as ações baseadas nestes dados.

A ideia central do artefato proposto reside na simulação, conduzida por modelos de inteligência artificial generativa, de questionamentos a respeito das ações capazes de aprimorar os resultados obtidos em cada questão das pesquisas de autoavaliação institucional. Diferentemente de uma simples consolidação estatística dos dados, o artefato organiza, para cada pergunta, o seu enunciado, o eixo do SINAES a que se vincula, o segmento respondente e a distribuição das respostas na escala Likert de cinco pontos, convertendo esse conjunto em um comando estruturado que orienta o modelo a propor ações de superação pertinentes àquela questão específica. Com isso, busca-se reproduzir, de maneira automatizada, o raciocínio que a comissão própria de avaliação realiza ao interpretar os percentuais de concordância e discordância e, a partir deles, sugerir medidas de melhoria para a instituição.

Essa abordagem apoia-se na capacidade dos modelos generativos de analisar grandes volumes de informação e produzir respostas contextualizadas a partir de um comando previamente definido. Conforme discutido no referencial teórico, modelos dessa natureza são treinados para gerar texto coerente e de alta qualidade a partir de uma entrada e de um contexto, o que os torna adequados à tarefa de transformar dados quantitativos em proposições qualitativas de ação. No artefato proposto, cada questão avaliada é tratada como uma unidade de análise, o modelo recebe o panorama das respostas e é instigado a indicar quais providências poderiam elevar o nível de concordância em avaliações futuras. A lógica do artefato, portanto, não se limita a descrever o resultado obtido, mas avança no sentido prescritivo, alinhando-se ao propósito do Design Science Research (DSR) de produzir soluções voltadas à melhoria das condições existentes.

Definidos esses fundamentos conceituais, optou-se por estruturar a solução de modo a operar com quatro modelos de inteligência artificial generativa, submetidos ao mesmo conjunto de dados e ao mesmo comando. São eles o ChatGPT, da OpenAI, o DeepSeek, da DeepSeek AI, o Gemini, do Google e o Claude, da Anthropic. A opção por múltiplos modelos, em vez de um único fornecedor, integra a própria concepção do artefato, pois permite confrontar as ações sugeridas por cada modelo entre si e, sobretudo, com as ações efetivamente registradas pela comissão, conferindo maior robustez à avaliação do artefato e reduzindo a dependência em relação ao comportamento de uma única solução.

A qualidade das proposições geradas está diretamente relacionada à forma como os comandos são elaborados, razão pela qual o artefato incorpora, em sua concepção, a prática de engenharia de prompt. Por meio dessa prática, busca-se construir instruções claras e personalizadas que tornem as respostas mais precisas e úteis, delimitando o papel que o modelo deve assumir, o formato esperado da saída e o escopo da análise. No artefato, o comando enviado a cada modelo contextualiza o cenário da autoavaliação institucional, apresenta o enunciado da questão e a distribuição das respostas e solicita, de modo explícito, a proposição de ações de superação vinculadas àquele resultado. A possibilidade de integrar diferentes modelos a um sistema desenvolvido em PHP decorre, ainda, da disponibilidade de interfaces de programação de aplicações oferecidas pelos próprios provedores, que viabilizam o envio dos comandos e o recebimento das respostas desde que realizadas as configurações adequadas.

As ações retornadas pelos modelos são organizadas segundo os cinco eixos do SINAES, Planejamento e Avaliação Institucional, Desenvolvimento Institucional, Políticas Acadêmicas, Políticas de Gestão e Infraestrutura Física, e relacionadas ao segmento respondente correspondente. Essa estruturação não é acessória, mas parte essencial do artefato, uma vez que assegura que o produto gerado pela inteligência artificial seja diretamente comparável ao formato dos relatórios elaborados manualmente pela comissão, os quais também se organizam a partir das dimensões definidas pelo SINAES. Dessa forma, o artefato foi concebido de modo a produzir uma saída já categorizada, o que viabiliza a etapa posterior de confronto entre as ações sugeridas automaticamente e aquelas apontadas pela CPA, conforme detalhado nos procedimentos de análise de dados.

A operação do artefato pressupõe, ainda, a articulação entre o sistema de coleta atualmente em uso no IFRS e o sistema desenvolvido nesta pesquisa, representada na Figura 9. Como o sistema oficial de autoavaliação não permite conexões externas diretas ao seu banco de dados, os dados são exportados em formato CSV e, em seguida, importados para o ambiente do sistema desenvolvido, ocasião em que se padronizam as informações conforme o formato esperado pela inteligência artificial. Essa separação entre o ambiente de coleta e o ambiente de análise preserva a integridade dos sistemas oficiais da instituição e mantém o artefato independente, característica que favorece sua eventual aplicação em outros campi do IFRS. Disponibilizados os dados, o usuário aciona, pela interface do sistema, a

funcionalidade responsável por percorrer as questões, montar os comandos e submetê-los ao modelo selecionado, dando origem ao documento com o apontamento das ações.

Cabe detalhar o sentido atribuído à expressão simulação de perguntas, que constitui o núcleo do artefato. Para cada questão da autoavaliação, o modelo é orientado a comportar-se como um analista de dados educacionais que, diante do resultado obtido, pergunta a si mesmo quais ações poderiam elevar a percepção da comunidade em relação àquele aspecto avaliado. Uma questão que apresente elevação nos índices de discordância sinaliza uma fragilidade institucional e, portanto, demanda a indicação de medidas corretivas, já uma questão com predomínio de concordância tende a ser tratada como uma potencialidade a ser preservada. Ao percorrer todas as questões dessa maneira, o artefato reproduz, de forma sistemática, o exercício interpretativo que fundamenta a elaboração dos apontamentos de superação registrados manualmente pela comissão, conferindo regularidade e abrangência a um procedimento que, no formato atual, depende integralmente da disponibilidade e da leitura individual dos avaliadores.

Considera-se, ainda, que algumas questões possuem maior aderência a determinado público, como aquelas relativas ao atendimento docente aos discentes ou à participação em projetos de ensino, pesquisa e extensão. Por essa razão, o segmento respondente é incluído no comando enviado ao modelo, de modo que as ações sugeridas guardem coerência com a origem das respostas e com o contexto a que se referem. Essa preocupação reflete a estruturação dos instrumentos descrita na delimitação da amostra, na qual a comunidade interna se subdivide nos segmentos docente, discente e técnico-administrativo. A inclusão do segmento como parâmetro do comando contribui para que o artefato não produza ações genéricas, mas proposições ajustadas à realidade de cada público, aproximando-se da sensibilidade contextual que caracteriza o trabalho da comissão.

Por fim, a concepção do artefato está diretamente vinculada à questão que norteia esta pesquisa, relativa ao modo como a inteligência artificial contribui no apontamento de ações de superação identificadas na coleta de informações. Ao gerar, de maneira autônoma, um conjunto de ações organizadas por eixo e por segmento, o artefato cria as condições necessárias para que se verifique em que medida tais ações convergem com aquelas elaboradas manualmente, permitindo avaliar se a abordagem automatizada é capaz de reproduzir, complementar ou aprimorar o trabalho da

comissão. Trata-se, portanto, de um artefato projetado não apenas para entregar um resultado prático à instituição, mas também para servir de instrumento de investigação do próprio fenômeno estudado, em conformidade com os pressupostos do Design Science Research.

Em síntese, o artefato proposto articula três elementos complementares: a representação estruturada dos dados da autoavaliação, a simulação de ações de superação por meio de modelos generativos e a organização das saídas em um formato comparável ao dos relatórios da comissão. Esses elementos, em conjunto, materializam a proposta de uma solução que pretende reduzir o tempo necessário à elaboração dos relatórios, hoje estimado em cerca de quatro meses, sem prescindir da análise crítica humana, uma vez que as ações indicadas pela inteligência artificial são submetidas à apreciação do usuário antes de sua incorporação ao relatório final.

4 RESULTADOS E DISCUSSÃO

Este capítulo apresenta os resultados obtidos a partir do desenvolvimento e da operação do sistema Auto Report, descrito nos procedimentos metodológicos. A exposição inicia-se pela apresentação detalhada da construção dos comandos enviados aos modelos de inteligência artificial, etapa que constitui o cerne do funcionamento do artefato e que condiciona, de forma direta, a qualidade das ações institucionais sugeridas. Em seguida, são discutidos a integração técnica entre o sistema e as interfaces de programação de aplicações dos provedores, o processo iterativo de refinamento dos comandos e os resultados da comparação entre as ações geradas automaticamente e aquelas registradas manualmente pela comissão própria de avaliação.

4.1 A CONSTRUÇÃO DOS PROMPTS UTILIZADOS NO SISTEMA DESENVOLVIDO

A eficácia de um sistema baseado em inteligência artificial generativa está diretamente associada à forma como as instruções são transmitidas ao modelo. Conforme discutido no referencial teórico, modelos de linguagem como o ChatGPT, o DeepSeek, o Gemini e o Claude são treinados para gerar texto coerente e contextualizado a partir de uma entrada previamente definida, denominada prompt ou comando. A prática de elaborar cuidadosamente esses comandos, conhecida como engenharia de *prompt*, busca tornar as respostas mais precisas e a interação mais eficiente (Ramos, 2023). No contexto do sistema Auto Report, essa prática assumiu papel central, uma vez que o objetivo não era apenas obter respostas plausíveis, mas sim respostas comparáveis, padronizadas e diretamente úteis ao trabalho da comissão própria de avaliação.

Toda a lógica de construção dos comandos foi concentrada em uma única classe da aplicação, responsável por reunir os dados quantitativos da autoavaliação, montar as instruções textuais e submetê-las ao modelo de inteligência artificial selecionado pelo usuário. Essa centralização atende ao atributo de modularidade destacado por Nunamaker, Chen e Purdin (1991) como desejável no desenvolvimento de artefatos computacionais, pois isola a responsabilidade de comunicação com os modelos generativos em um único ponto do sistema, facilitando manutenção, auditoria

e evolução dos comandos. A presente seção descreve, primeiramente, a anatomia do prompt definitivo adotado pelo sistema; em seguida, a forma como esse comando é integrado às diferentes APIs; e, por fim, o percurso iterativo que conduziu da formulação inicial, ainda incipiente, até a versão final empregada na geração dos relatórios.

4.1.1 A representação dos dados quantitativos no comando

Antes de qualquer instrução ser enviada ao modelo, o sistema necessita converter os dados armazenados em uma representação textual legível para a inteligência artificial. Conforme descrito na modelagem do sistema, as respostas são persistidas de forma agregada na classe Resposta, que registra, para cada pergunta e cada campus, quantas vezes determinado valor da escala Likert foi assinalado. A classe de controle do prompt percorre as perguntas de um instrumento e, para cada uma, recupera as respostas correspondentes ao campus em análise, concatenando-as em uma cadeia de texto estruturada.

A montagem dessa cadeia segue um padrão fixo. Para cada pergunta, o sistema imprime o enunciado e, na sequência, a distribuição quantitativa das respostas, no formato exemplificado a seguir, em que se reproduz a estrutura efetivamente gerada pelo código do sistema:

```
O campus oferece acesso satisfatório à internet?  
Respostas: 125 respostas com nota 1; 84 respostas com  
nota 2;  
60 respostas com nota 3; 95 respostas com nota 4;  
110 respostas com nota 5.
```

Essa formatação não é arbitrária. Ao apresentar o enunciado integral da pergunta seguido da contagem absoluta de respostas por ponto da escala, o sistema fornece ao modelo o panorama completo de cada questão, preservando tanto a polaridade das respostas quanto a intensidade da concordância ou discordância. A opção por enviar as quantidades absolutas, em vez de percentuais previamente calculados, foi deliberada, pois transfere ao próprio modelo a tarefa de ponderar a distribuição, ao mesmo tempo em que mantém a fidelidade aos dados originais extraídos dos relatórios da comissão. A repetição desse padrão para todas as

perguntas de um instrumento gera um bloco de texto, internamente designado questionamento, que constitui a base factual sobre a qual o modelo é instado a raciocinar.

Um cuidado adicional incorporado ao comando diz respeito à explicitação da escala. Como os dados são transmitidos apenas como notas numéricas de um (1) a cinco (5), o sistema antecede o bloco de dados com a informação de que a nota 1 significa discordar totalmente e a nota 5 significa concordar totalmente. Sem essa explicação, o modelo não teria como saber se os valores mais altos representam avaliações positivas ou negativas, o que comprometeria toda a interpretação subsequente. Trata-se de um exemplo concreto de como a engenharia de prompt precisa fornecer ao modelo não apenas os dados, mas também a chave de leitura desses dados.

4.1.2 Os componentes do *prompt*

O comando definitivo adotado pelo sistema é composto por três elementos complementares, que atuam em conjunto para orientar o comportamento do modelo: a definição do papel a ser assumido, a instrução de tarefa propriamente dita e as regras de formatação da resposta. Cada um desses elementos foi incorporado ao *prompt* como resposta a uma limitação identificada nas versões iniciais, refinando a instrução passada para atender o tipo de retorno esperado.

O primeiro componente é a atribuição de papel (*system role*). Em todas as chamadas aos modelos, o sistema informa que a inteligência artificial deve comportar-se como um analista de dados educacionais. Essa instrução é transmitida no campo reservado às mensagens de sistema, separado da mensagem do usuário nos modelos que oferecem essa distinção. A definição de um papel especializado condiciona o modelo a adotar um vocabulário, um enfoque e um nível de profundidade compatíveis com o domínio da avaliação institucional, em vez de produzir respostas genéricas. Ao se posicionar como analista de dados educacionais, o modelo tende a interpretar a distribuição das respostas sob a ótica da gestão acadêmica, propondo medidas pertinentes ao contexto de uma instituição de ensino.

O segundo componente é a instrução de tarefa. O sistema emprega dois comandos distintos, conforme o objetivo da requisição. No primeiro, voltado à caracterização do público respondente, a instrução pergunta ao modelo, com base

nos dados fornecidos, como ele define os respondentes. No segundo, voltado à proposição de melhorias, a instrução solicita quais ações o modelo sugere para melhorar as notas obtidas pelos respondentes. Essa dupla formulação reflete a própria estrutura dos relatórios da comissão, que combinam uma leitura interpretativa do perfil das respostas com o apontamento de ações de superação. A separação entre os dois tipos de comando permite que o sistema gere, para cada instrumento e para cada pergunta, tanto uma análise descritiva quanto uma recomendação prescritiva, alinhando-se ao caráter normativo do artefato preconizado pelo *Design Science Research*.

O terceiro componente é o conjunto de regras de formatação da resposta. Ao final da instrução, o comando determina que a resposta seja produzida em HTML válido, sem as marcações de abertura e fechamento de documento, estruturando o conteúdo por meio de elementos específicos para títulos, parágrafos, listas e tabelas, e proíbe expressamente o uso de marcação *markdown*. Essa exigência decorre diretamente do modo como o relatório final é construído, conforme se detalha adiante, e foi a solução encontrada para o problema de formatação que comprometia a apresentação do documento em PDF nas versões iniciais do sistema.

A combinação desses três componentes com o bloco de dados quantitativos resulta no comando completo enviado ao modelo. A título de ilustração, o comando de sugestão de ações, em sua forma definitiva, apresenta a estrutura reproduzida no Quadro 2, que decompõe o prompt em suas partes constitutivas e indica a função desempenhada por cada uma.

Quadro 2 – Decomposição do *prompt* de sugestões de ações

Componente	Trecho do comando	Função
Papel (system)	“Você é um analista de dados educacionais.”	Especializa o comportamento do modelo no domínio educacional.
Chave de leitura	“...onde a nota 1 significa discordo totalmente e a nota 5 significa concordo totalmente...”	Define a polaridade da escala Likert para o modelo.
Instrução de tarefa	“...quais ações você sugere para melhorar as notas obtidas pelos	Direciona o modelo à proposição prescritiva de ações.

	respondentes? Defina a resposta em um único parágrafo.”	
Regras de formato	“Responda utilizando HTML válido, sem <html> ou <body>... Não utilize markdown.”	Garante a renderização correta no documento PDF.
Dados (questionamento)	Enunciado de cada pergunta somada à distribuição das respostas por nota.	Fornece a base para a análise.

Fonte: autoria própria (2026).

4.1.3 O nível de detalhe dos comandos: avaliação, instrumento e pergunta

Um aspecto relevante da construção dos *prompts* diz respeito ao nível de detalhe com que os comandos são aplicados. O sistema não se limita a submeter um único bloco de dados ao modelo, mas organiza a comunicação em três níveis distintos, refletindo a própria hierarquia das entidades da autoavaliação. No nível mais amplo, o da avaliação, o sistema percorre todos os instrumentos de um ciclo avaliativo e gera, para cada um deles, uma caracterização dos respondentes e um conjunto de ações sugeridas. No nível intermediário, o do instrumento, os comandos consideram apenas as perguntas de um questionário específico, como o instrumento de avaliação docente respondido pelos docentes. No nível mais fino, o da pergunta, o comando contempla uma única questão e sua respectiva distribuição de respostas.

Essa estrutura de três níveis confere flexibilidade ao artefato e atende a diferentes necessidades de análise. A geração no nível da pergunta permite que o relatório apresente, logo após o gráfico de cada questão, uma interpretação localizada da distribuição das respostas e uma sugestão de melhoria pontual. Já a geração no nível do instrumento sintetiza, ao final de cada questionário, um conjunto de ações de superação que considera o conjunto das perguntas em sua totalidade, aproximando-se da lógica com que a comissão própria de avaliação encerra cada seção de seus relatórios. A coexistência dos dois níveis assegura que o documento final ofereça tanto o detalhamento questão a questão quanto a visão consolidada por instrumento, característica que será retomada na análise da estrutura do relatório gerado.

Para cada um desses níveis, a classe de controle de execução do *prompt* disponibiliza um par de métodos, um destinado a definir os dados (caracterização dos

respondentes) e outro destinado a sugerir as ações (proposição de melhorias). Há, portanto, seis pontos de entrada distintos, que compartilham o mesmo núcleo de construção do comando e diferenciam-se apenas pelo escopo dos dados considerados e pela instrução de tarefa empregada. Essa simetria garante consistência entre os diferentes níveis de análise e evita divergências de comportamento entre as partes do relatório.

4.1.4 Integração com as interfaces de programação dos modelos

Definida a estrutura do comando, o sistema precisa transmiti-lo ao modelo de inteligência artificial escolhido pelo usuário e recuperar a resposta gerada. Essa comunicação ocorre por meio das interfaces de programação de aplicações disponibilizadas por cada provedor. Conforme apontado por Gallardo, Machuca e Semblantes (2023), é possível integrar diferentes linguagens de programação às APIs dos modelos generativos existentes, desde que realizadas as configurações adequadas e utilizadas as ferramentas apropriadas. No Auto Report, desenvolvido em PHP com o framework Laravel, essa integração foi implementada de modo que o mesmo comando textual pudesse ser encaminhado, conforme a seleção do usuário, a um entre quatro modelos distintos.

Embora o conteúdo do comando seja idêntico para todos os modelos, a forma de invocação varia conforme as particularidades de cada *interface*, o que exigiu um tratamento condicional no código. A seleção do modelo é determinada por uma preferência associada ao usuário autenticado, e o sistema dirige a requisição ao provedor correspondente. O Quadro 3 sintetiza as diferenças técnicas observadas na integração com cada um dos quatro modelos.

Quadro 3 – Características da integração de cada modelo de inteligência artificial

Sistema de IA	Modelo utilizado	Método da API	Particularidade
ChatGPT (OpenAI)	gpt-5.5	Método <i>chat</i>	Separa papel (<i>system</i>) e mensagem (<i>user</i>). Retorno no formato JSON a ser decodificado.

Deepseek (Deepseek AI)	deepseek-chat	Método <i>chat</i>	Estrutura de mensagens equivalente à OpenAI. Resposta acessada por arranjo associativo.
Gemini (Google)	gemini-3.5-flash	Método <i>generateText</i>	Não distingue papel de mensagem: papel e instrução são concatenados em texto único.
Claude (Anthropic)	claude-sonnet-4.6	Método <i>chat</i>	Recebe o papel em campo próprio (<i>system</i>), separado da lista de mensagens.

Fonte: autoria própria (2026).

Como se observa no Quadro 3, três dos quatro modelos (ChatGPT, DeepSeek e Claude) adotam uma estrutura de comunicação semelhante, na qual o papel atribuído ao modelo é transmitido em um campo dedicado, distinto da mensagem do usuário. Essa separação é benéfica do ponto de vista da engenharia de *prompt*, pois evidencia ao modelo a fronteira entre a definição de seu comportamento e a tarefa concreta a ser executada. O Gemini, por sua vez, não oferecia essa distinção na biblioteca empregada, de modo que foi necessário concatenar a definição do papel e a instrução em uma única cadeia de texto antes do envio. Apesar dessa diferença estrutural, o conteúdo semântico transmitido ao modelo permaneceu equivalente, preservando a comparabilidade entre os resultados dos quatro provedores.

A integração com o Gemini exigiu ainda um cuidado adicional relacionado à estabilidade do serviço. Durante os testes, observou-se a ocorrência intermitente de respostas de erro com código 503, indicativas de indisponibilidade temporária do serviço. Para mitigar esse comportamento e evitar a interrupção da geração de um relatório inteiro em razão de uma falha pontual, o sistema foi dotado de um mecanismo de nova tentativa: ao identificar o erro 503, a aplicação aguarda um breve intervalo e reenvia o comando antes de, em caso de persistência da falha, propagar a exceção. Esse tratamento confere robustez ao processo de geração e reflete uma preocupação prática com a confiabilidade do artefato em condições reais de uso, aspecto valorizado na avaliação de soluções no âmbito do *Design Science Research*.

A decisão de integrar simultaneamente os quatro modelos, em vez de restringir a solução a um único fornecedor, integra a própria concepção metodológica da pesquisa. Ao submeter o mesmo conjunto de dados e o mesmo comando a sistemas distintos, o Auto Report constitui-se em um ambiente de experimentação que viabiliza a comparação entre as ações sugeridas por cada modelo e, sobretudo, entre essas sugestões e as ações efetivamente apontadas pela comissão. A uniformização do comando entre os provedores é, portanto, condição necessária para a validade dessa comparação, pois assegura que eventuais diferenças nos resultados decorram das características intrínsecas de cada modelo, e não de variações na formulação das instruções.

4.1.5 Refinamento do *prompt*

A versão do comando descrita nas subseções anteriores não foi obtida de imediato, mas resultou de um processo iterativo de refinamento, coerente com o *Design Science Research*, em que o artefato é sucessivamente construído, avaliado e aprimorado. As primeiras formulações dos *prompts*, empregadas nas etapas iniciais de desenvolvimento, apresentavam limitações que comprometiam a qualidade e a usabilidade das respostas, e foi a identificação dessas limitações que motivou a incorporação progressiva dos elementos que compõem o comando final. Esta subseção reconstitui esse percurso, organizando-o em três deficiências principais e nas soluções a elas associadas.

As primeiras versões do comando consistiam essencialmente na apresentação dos dados acompanhada de uma pergunta direta, sem qualquer instrução que orientasse o modelo quanto à perspectiva a ser adotada. Na prática, o sistema enviava ao modelo apenas a distribuição das respostas e a solicitação de uma análise, sem informar de que ponto de vista essa análise deveria ser conduzida. O resultado eram respostas excessivamente genéricas, que ora se limitavam a reformular os números apresentados, ora propunham considerações pouco aderentes ao contexto da gestão acadêmica. Para corrigir essa deficiência, incorporou-se a atribuição explícita do papel de analista de dados educacionais, transmitida no campo de sistema das requisições. Essa mudança alinhou o comportamento do modelo ao domínio da avaliação institucional, fazendo com que as respostas passassem a refletir

o enfoque, o vocabulário e o nível de profundidade esperados de um profissional dessa área.

Nas formulações iniciais, a pergunta dirigida ao modelo era vaga e não delimitava com precisão o que se esperava como resposta. Pedidos genéricos de análise geravam saídas de extensão e foco imprevisíveis, dificultando tanto a leitura quanto a posterior comparação com os relatórios da comissão. O refinamento consistiu em reformular a instrução de tarefa de modo a explicitar o objetivo pretendido. Para a caracterização do público, passou-se a perguntar, de forma direta, como o modelo define os respondentes a partir dos dados; para a proposição de melhorias, passou-se a solicitar quais ações o modelo sugere para elevar as notas obtidas, acrescentando, no nível da pergunta, a determinação de que a resposta fosse apresentada em um único parágrafo. Essa delimitação do formato textual reduziu a variabilidade das respostas e conferiu regularidade ao conjunto de ações geradas, tornando-as efetivamente comparáveis entre si e em relação ao trabalho manual da comissão. A apresentação explícita da chave de leitura da escala Likert, mencionada anteriormente, integra igualmente esse esforço de tornar a pergunta mais precisa, ao eliminar a ambiguidade quanto ao significado das notas.

A terceira e mais visível limitação das versões iniciais manifestava-se apenas no produto final do sistema. Como os comandos não estabeleciam qualquer regra sobre o formato da resposta, os modelos retornavam o texto com marcação *markdown*, recurso utilizado por padrão para estruturar listas, títulos e ênfases. Essa marcação, contudo, é incompatível com o mecanismo de geração do documento final, que converte conteúdo HTML em PDF. Em consequência, os relatórios produzidos exibiam símbolos de marcação em meio ao texto, como asteriscos e cardinais, em vez de listas e títulos devidamente formatados, comprometendo a apresentação e a legibilidade do documento. A solução adotada foi acrescentar ao comando um conjunto explícito de regras de formatação, determinando que a resposta fosse produzida em HTML válido, sem as marcações de documento, estruturada por meio de elementos de título, parágrafo, lista e tabela, e proibindo expressamente o uso de *markdown*. Com essa instrução, os modelos passaram a retornar conteúdo diretamente compatível com o método gerador de arquivos em PDF, eliminando o problema de formatação e produzindo um documento de aparência profissional.

A relevância dessa última correção só se compreende plenamente ao se considerar o modo como o relatório é montado. A geração do documento é

coordenada por um segundo componente do sistema, responsável por orquestrar as chamadas ao construtor de comandos e por organizar as respostas em uma estrutura sequencial. Esse componente percorre os instrumentos e as perguntas de uma avaliação e, para cada elemento, registra um bloco identificado por sua natureza, título, introdução, pergunta, gráfico, análise ou conclusão, compondo um arranjo ordenado que reproduz a lógica dos relatórios da comissão. Esse arranjo é então convertido em um arquivo PDF por meio de uma biblioteca que interpreta marcação HTML. É justamente por essa razão que o conteúdo retornado pelos modelos precisa estar em HTML válido, já que o texto gerado pela inteligência artificial é inserido diretamente no corpo do documento e renderizado como parte integrante do relatório. Qualquer marcação estranha ao HTML, como a sintaxe *markdown*, seria exibida literalmente, descaracterizando o resultado.

O Quadro 4 sistematiza o percurso de refinamento, contrastando a formulação inicial com a versão definitiva e indicando o efeito de cada aprimoramento sobre o resultado obtido.

Quadro 4 – Evolução do *prompt*

Aspecto	Formulação inicial	Versão final
Papel do modelo	Não definido, com respostas genéricas e pouco aderentes ao contexto educacional.	Definido como “analista de dados educacionais”, especializando a análise.
Instrução de tarefa	Pergunta vaga, sem delimitação de objetivo ou de extensão da resposta.	Pergunta precisa, com objetivo explícito e delimitação do formato em parágrafo único
Chave da escala	Notas enviadas sem explicação da polaridade da escala <i>Likert</i> .	Explicação de que a nota 1 é discordância total e a nota 5, concordância total.
Formato da resposta	Sem regras, obtendo retornos em <i>markdown</i> , que geravam símbolos soltos no PDF.	HTML válido sem <i>markdown</i> , renderizando corretamente no arquivo em formato PDF.
Resultado no relatório	Texto sem formatação, de foco variável e de difícil comparação.	Texto formatado, padronizado e diretamente comparável aos relatórios da CPA.

Fonte: autoria própria (2026).

O percurso descrito evidencia que a construção do *prompt* definitivo foi, ela própria, um resultado da pesquisa, e não apenas um pressuposto técnico. Cada um dos três aprimoramentos, a atribuição de papel, o refinamento da pergunta e a imposição de regras de formato, respondeu a uma limitação concreta observada na operação do artefato e contribuiu para aproximar as saídas geradas automaticamente do padrão esperado nos relatórios institucionais. Essa progressão ilustra de maneira concreta o caráter iterativo do *Design Science Research*, no qual o conhecimento sobre a classe de problemas é gerado e consolidado ao longo dos sucessivos ciclos de construção e avaliação do artefato.

Cabe destacar, por fim, que a uniformidade do comando entre os quatro modelos foi mantida durante todo o processo de refinamento. Cada ajuste introduzido foi aplicado simultaneamente às chamadas de todos os provedores, de modo que, ao final, ChatGPT, DeepSeek, Gemini e Claude recebessem instruções equivalentes em conteúdo, ainda que adaptadas às particularidades sintáticas de cada *interface*. Essa decisão preserva a validade da comparação entre os modelos, assegurando que as diferenças observadas em seus resultados sejam atribuíveis às características de cada solução, e não a discrepâncias na formulação dos comandos. Definida e estabilizada a construção dos *prompts*, tornou-se possível proceder à geração sistemática dos relatórios e à comparação de suas saídas com os apontamentos da comissão própria de avaliação, objeto das seções seguintes.

4.2 A GERAÇÃO DOS RELATÓRIOS E A CARACTERIZAÇÃO DAS AÇÕES IDENTIFICADAS

Estabilizada a construção dos prompts descrita na seção anterior, o sistema Auto Report foi empregado na geração dos relatórios correspondentes às avaliações da comunidade interna do IFRS – campus Bento Gonçalves, referentes aos anos de 2023, 2024 e 2025. Esta seção apresenta o produto efetivamente gerado pelo artefato, caracterizando a amostra analisada, descrevendo a estrutura dos documentos produzidos e quantificando as ações institucionais identificadas por cada um dos quatro modelos de inteligência artificial integrados à solução. Cabe antecipar que o principal resultado desta seção é a elevada convergência entre as ações propostas pelo sistema e aquelas registradas manualmente pela comissão. Considerando

apenas as questões em que a comissão efetivamente apontou ações, a taxa de convergência situou-se entre 90,9% e 100% (92,9% em 2023, 90,9% em 2024 e 100% em 2025), resultado que sustenta a viabilidade do artefato como apoio à elaboração dos relatórios de autoavaliação institucional. Ressalva-se, contudo, que esses percentuais refletem o cenário específico avaliado e não devem ser tomados como valores fixos a serem reproduzidos em outros contextos

A amostra submetida ao sistema abrange os três segmentos da comunidade interna (docente, discente e técnico-administrativo) e o conjunto de vinte e uma questões distribuídas entre os cinco eixos do SINAES. Para cada um dos três anos, o mesmo conjunto de dados e o mesmo comando foram submetidos aos quatro modelos, o que gerou, por ano, quatro relatórios completos, cada qual contendo a caracterização dos respondentes, a apresentação gráfica das respostas de cada pergunta, a análise localizada de cada questão e a síntese das ações de superação ao final de cada instrumento. A operação central realizada pelo artefato, receber a distribuição das respostas na escala Likert e devolver, para cada questão, proposições qualitativas de ação, materializa o que Ramos (2023) descreve a respeito dos modelos de linguagem generativa, treinados para produzir respostas coerentes e contextualizadas a partir de uma entrada e de um comando previamente definido. Da mesma forma, o resultado observado confirma, na prática, a orientação de Li *et al.* (2017) e de Russell e Norvig (2024), para quem o traço distintivo da inteligência artificial reside justamente na capacidade de analisar grandes volumes de dados e, a partir deles, subsidiar a tomada de decisão de forma mais precisa. O documento produzido pelo sistema reproduz, em sua sequência interna, a lógica dos relatórios institucionais elaborados pela comissão, o que assegura sua comparabilidade direta com o material manual.

Para caracterizar as ações identificadas, cada apontamento gerado pelos modelos foi classificado quanto à sua correspondência com as ações registradas manualmente pela comissão, segundo o critério definido nos procedimentos de análise de dados, que distingue correspondência total, correspondência parcial e ausência de correspondência. Essa classificação, aplicada questão a questão nos três anos, permite quantificar o volume de ações efetivamente convergentes com o trabalho da comissão e constitui a base sobre a qual se assenta a discussão desenvolvida a seguir. O Quadro 5 sintetiza a distribuição dessas classificações por modelo e por ano.

Quadro 5 – Distribuição das classificações de correspondência por modelo e por ano

Ano	Modelo de IA	Correspondência total	Corresp. parcial	Ausência de corresp.	Algum grau de corresp.
2023	ChatGPT	7	6	8	13
2023	Deepseek	8	5	8	13
2023	Gemini	7	6	8	13
2023	Claude	9	4	8	13
2024	ChatGPT	3	7	11	10
2024	Deepseek	3	7	11	10
2024	Gemini	3	7	11	10
2024	Claude	3	7	11	10
2025	ChatGPT	3	10	8	13
2025	Deepseek	3	10	8	13
2025	Gemini	1	12	8	13
2025	Claude	3	10	8	13

Fonte: autoria própria (2026). Nota: valores expressos em número de questões sobre um total de 21 questões avaliadas por ano.

Como se observa no Quadro 5, a distribuição das classificações é homogênea entre os quatro modelos dentro de cada ano, o que decorre da uniformização do comando descrita na seção anterior e reforça a validade da comparação. As diferenças entre modelos, quando existem, restringem-se a deslocamentos pontuais entre a correspondência total e a parcial. Em 2023, o Claude concentrou o maior número de correspondências totais (nove), enquanto, em 2025, o Gemini foi o único a registrar apenas uma correspondência total, migrando as demais para a faixa parcial. Em 2024, os quatro modelos apresentaram distribuição idêntica. O exame do desempenho relativo de cada modelo, contudo, é retomado de forma dedicada na seção 4.4, cabendo aqui apenas registrar o volume e a natureza das ações produzidas.

É possível observar, no Quadro 5, o comportamento do total de ausências de correspondência, que permanece elevado e estável (oito em 2023 e 2025, e onze em 2024). Esse resultado, todavia, não decorre de falha do sistema em identificar ações pertinentes, mas da própria natureza dos relatórios manuais tomados como

referência. Em um número expressivo de questões, a comissão própria de avaliação não registra qualquer ação de superação específica, limitando-se a apresentar o diagnóstico quantitativo das respostas. Nessas questões, não há ação manual com a qual confrontar a sugestão do sistema, de modo que a classificação recai necessariamente sobre a ausência de correspondência, ainda que o modelo tenha proposto ações consistentes. Essa constatação torna indispensável distinguir dois recortes de leitura dos resultados, um que considera a totalidade das vinte e uma questões e outro que exclui aquelas em que a comissão não apontou ações.

O Quadro 6 apresenta, para os três anos, o número de questões em que a comissão registrou ações de superação e o número daquelas em que apenas o diagnóstico foi apresentado, sem qualquer apontamento de ação. Como a distribuição das classificações é idêntica entre os modelos referente às ausências, os percentuais de convergência calculados a partir desse recorte aplicam-se igualmente aos quatro modelos.

Quadro 6 – Convergência das ações do sistema conforme o recorte de análise

Indicador	2023	2024	2025
Questões avaliadas	21	21	21
Questões com ação registrada pela comissão	12	11	13
Questões sem ação registrada pela comissão	9	10	8
Convergência considerando as 21 questões	61,9%	47,6%	61,9%
Convergência excluindo as questões sem ação	92,9%	90,9%	100%

Fonte: autoria própria (2026). Nota: os percentuais de convergência correspondem à soma das correspondências total e parcial e são iguais para os quatro modelos de IA utilizados, dado que as ausências se concentram nas mesmas questões sem ação da comissão.

Quando se consideram todas as vinte e uma questões, a convergência do sistema com o relatório manual situa-se em 61,9% em 2023 e 2025 e em 47,6% em 2024, patamar que, tomado isoladamente, poderia sugerir aderência moderada. Contudo, ao se excluírem as questões em que a comissão não apontou ações, a convergência se eleva de maneira acentuada, alcançando 92,9% em 2023, 90,9% em 2024 e 100% em 2025. Ou seja, quando há uma ação manual a ser reproduzida, o

sistema quase sempre identifica uma ação correspondente, seja de forma total ou parcial.

Essa distinção é decisiva para a correta interpretação dos resultados. O elevado número de ausências no recorte completo não reflete incapacidade do artefato de sugerir ações relevantes, mas sim a assimetria entre os dois métodos. O sistema propõe ações de melhoria para todas as questões, ao passo que a comissão o faz de forma seletiva, concentrando seus apontamentos em parte das dimensões avaliadas. Esse comportamento encontra respaldo no referencial teórico, onde Li *et al.* (2017) orientam que a inteligência artificial se destaca como recurso fundamental na tomada de decisões precisamente por sua capacidade de processar grandes volumes de dados e aprender com eles. O resultado obtido confirma essa orientação, já que o sistema percorreu, todas as questões dos três ciclos avaliativos, dimensão que a análise manual, condicionada pela leitura individual dos avaliadores, não cobre integralmente. Do ponto de vista da questão de pesquisa, esse padrão sinaliza que a inteligência artificial não apenas reproduz o trabalho manual, mas tende a ampliá-lo, cobrindo dimensões que a comissão, por limitação de tempo ou por opção metodológica, deixou sem apontamento.

A elevada convergência observada no recorte que exclui as questões sem ação dialoga com a caracterização dos modelos atuais proposta por Ludermit (2021). Segundo a autora, o estágio atual da tecnologia corresponde a algoritmos capazes de executar tarefas com competência comparável à humana quando aplicados a um escopo delimitado, entendimento reforçado por Mendonça *et al.* (2024), onde modelos como os empregados nesta pesquisa demonstram capacidades avançadas dentro de tarefas específicas, ainda que não configurem uma inteligência de propósito geral. O apontamento de ações de superação a partir da distribuição de respostas constitui exatamente uma tarefa dessa natureza, o que ajuda a explicar por que o sistema reproduziu a leitura da comissão, sempre que havia uma ação manual a ser identificada.

Cabe registrar, por fim, o desempenho do artefato quanto ao tempo de elaboração dos relatórios, aspecto diretamente vinculado ao problema que motivou a pesquisa. Enquanto a comissão própria de avaliação despende, em média, cerca de quatro meses para concluir a interpretação dos dados e a redação dos relatórios, o sistema Auto Report gerou o conjunto completo de apontamentos, para as vinte e uma questões de cada ano, em intervalos que variaram de aproximadamente dezoito

minutos a pouco mais de uma hora, conforme o modelo empregado. Para os fins desta seção, importa destacar a ordem de grandeza do ganho obtido, independentemente do modelo, a geração dos apontamentos, que manualmente consumia meses, passou a ser concluída pelo sistema em questão de minutos, reduzindo significativamente o tempo de trabalho da comissão, sem prescindir da posterior apreciação crítica das ações por seus membros.

A significativa redução do tempo de trabalho da comissão, materializa o princípio da eficiência tal como definido pelos autores estudados. Matias-Pereira (2012) orienta que esse princípio impõe ao servidor público realizar suas atribuições com maior agilidade, perfeição e qualidade, apoiando-se nas modernas técnicas administrativas. Camargo e Guimarães (2013), no mesmo sentido, definem a eficiência como a execução das atribuições com rapidez e proficiência, buscando os melhores resultados com o menor custo a partir dos recursos disponíveis. Coelho (2011) acrescenta que, na Administração Pública, não se trata apenas de reduzir custos, mas de alocar os recursos de modo a maximizar a satisfação do interesse público. O artefato desenvolvido responde a essas orientações ao preservar a qualidade do apontamento, enquanto reduz drasticamente o tempo e o esforço humano empregados, liberando os avaliadores para atividades de maior valor analítico.

Da mesma forma, Alhosani e Alhashmi (2024) orientam que, entre as principais contribuições da IA ao setor público, figuram o aumento da produtividade, a diminuição das cargas de trabalho e a simplificação de procedimentos, exatamente os efeitos observados na substituição de um processo manual pelo processo automatizado de geração de relatórios apresentado neste trabalho. Gonçalves *et al.* (2020) destacam que as tecnologias de governo eletrônico garantem maior eficiência nas atividades públicas e significativa redução de custos, enquanto Desordi e Bona (2020) e Valle e Gallo (2020) apresentam casos brasileiros em que a inteligência artificial otimiza, agiliza e torna mais eficiente o serviço público. O Auto Report insere-se nessa mesma classe de aplicações, transpondo para o contexto da autoavaliação institucional do IFRS a lógica de eficiência que esses autores identificam em outros órgãos, o que confere aderência às orientações do referencial e reforça a pertinência do artefato como resposta ao problema investigado.

4.3 ANÁLISE COMPARATIVA POR EIXO DOS SINAES

Caracterizado o volume geral das ações no item anterior, esta seção confronta as ações produzidas pelos modelos de inteligência artificial e aquelas registradas manualmente pela comissão própria de avaliação, organizando o exame a partir dos cinco eixos definidos pelo Sistema Nacional de Avaliação da Educação Superior (SINAES). Como estabelecido nos procedimentos de análise de dados, cada ação institucional é tratada como uma unidade de análise e classificada segundo o eixo a que se vincula, a questão a que se relaciona e o segmento de origem das respostas, o que assegura que a comparação ocorra sobre elementos equivalentes.

A análise apoia-se em uma constatação decisiva, já antecipada na seção anterior e reforçada pelos dados, a classificação de correspondência é praticamente idêntica entre os quatro modelos. As diferenças entre eles restringem-se ao grau da correspondência (total ou parcial) em poucas questões, mas a distribuição das ausências é a mesma para ChatGPT, DeepSeek, Gemini e Claude, uma vez que as ausências se concentram nas questões em que a comissão não apontou ação. Por essa razão, a leitura por eixo apresentada a seguir vale para os quatro modelos, ficando o exame das diferenças de precisão entre eles reservado à seção 4.4.

O Quadro 7 consolida, para os três anos reunidos, a distribuição das questões por eixo, discriminando aquelas em que a comissão registrou ações de superação e aquelas em que apenas apresentou o diagnóstico, bem como o grau de convergência das ações do sistema nas questões apontadas pela comissão. Os valores referem-se ao conjunto dos três ciclos avaliativos (2023, 2024 e 2025), onde cada questão é contabilizada uma vez por ano.

Quadro 7 – Convergência das ações por eixo do SINAES (2023-2025)

Eixo	Questões (3 anos)	Com ação	Sem ação	Convergência	Total / Parcial
Eixo 1 – Planejamento e Avaliação Institucional	6	3	3	100%	2 / 1
Eixo 2 – Desenvolvimento Institucional	6	3	3	100%	0 / 3
Eixo 3 – Políticas Acadêmicas	24	9	15	89%	1 / 7
Eixo 4 – Políticas de Gestão	9	4	5	100%	1 / 3

Eixo 5 – Infraestrutura Física	18	17	1	100%	10 / 7
--------------------------------	----	----	---	------	--------

Fonte: autoria própria (2026). Nota: a coluna “Convergência” expressa a proporção de questões com ação da comissão em que o sistema apresentou correspondência total ou parcial. A coluna “Total / Parcial” discrimina, entre essas convergências, o número de correspondências totais e parciais.

O Quadro 7 revela um padrão consistente, sempre que a comissão registra uma ação de superação, o sistema tende a identificar uma ação correspondente. Nos Eixos 1, 2, 4 e 5, a convergência é integral, e mesmo no Eixo 3, o único a apresentar uma ausência entre as questões com ação, a taxa permanece elevada, em 89%. A leitura eixo a eixo, esclarece as particularidades de cada dimensão e o significado dos casos de correspondência apenas parcial e de ausência.

No Eixo 1 – Planejamento e Avaliação Institucional, composto pelas questões relativas à divulgação e ao uso dos resultados da avaliação, a comissão apontou ações em três das seis oportunidades ao longo dos três anos, e em todas elas o sistema convergiu. As ações da comissão concentraram-se na necessidade de ampliar a divulgação dos resultados e de evidenciar como a avaliação se converte em planejamento, e os modelos propuseram a diversificação de canais, a criação de painéis de transparência e campanhas de devolutiva, o que resultou em correspondências totais. Trata-se de um eixo em que o raciocínio prescritivo do sistema reproduziu com fidelidade a leitura da comissão, comportamento coerente com a orientação de Ramos (2023), para quem os modelos generativos, a partir de um comando bem definido, geram respostas contextualizadas e pertinentes ao problema apresentado.

O Eixo 2 – Desenvolvimento Institucional, que reúne as questões sobre missão, valores e visão e sobre inclusão social, apresentou convergência integral nas três questões com ação da comissão, embora na forma de correspondência parcial. Essa característica decorre da natureza das ações registradas manualmente neste eixo, nos casos de inclusão, a comissão descreveu ações concretas e específicas já executadas, como a atuação do Núcleo de Atendimento às Pessoas com Necessidades Específicas (NAPNE), horas de assessoria e projetos de acessibilidade. Como os modelos não tinham acesso a esse relato de execução, sugeriram medidas de aprimoramento pertinentes, porém de caráter mais genérico e prospectivo. Essa diferença é uma decorrência do modo de operação descrito por

Ramos (2023), onde o modelo generativo produz suas respostas exclusivamente a partir da entrada e do contexto que lhe são fornecidos, de modo que, recebendo apenas a distribuição das respostas e não o histórico das iniciativas institucionais, tende a formular proposições prospectivas. A aproximação temática é evidente, mas a diferença de foco entre o realizado e o proposto explica a predominância da correspondência parcial.

O Eixo 3 – Políticas Acadêmicas, o mais extenso, com oito questões e, portanto, vinte e quatro observações no triênio, é também o mais heterogêneo. Nele se concentra a maior parte das questões sem ação da comissão (quinze das vinte e quatro), sobretudo nas dimensões de comunicação institucional e de participação, em que a comissão frequentemente se limitou ao diagnóstico. Nas nove questões com ação, a convergência foi de 89%, com uma única ausência. Esse caso isolado, a questão sobre a participação em projetos de ensino, pesquisa e extensão (em 2024), é elucidativo, pois o registro da comissão consistiu na enumeração dos projetos efetivamente executados no ano, um relato de resultados, e não em uma ação de superação prospectiva. Os modelos, por sua vez, sugeriram ampliar a divulgação e o acesso às oportunidades, de modo que a divergência reflete mais a diferença de propósito entre um relato de execução e uma proposição de melhoria do que uma falha do sistema.

No Eixo 4 – Políticas de Gestão, relativo à qualificação dos servidores, à participação em colegiados e à divulgação das regulamentações, a comissão apontou ações em quatro das nove observações, e em todas houve convergência. À semelhança do Eixo 2, parte das ações da comissão descreveu iniciativas concretas de qualificação como bolsas, afastamentos e valores investidos, enquanto os modelos propuseram ampliar a divulgação e o alcance dessas iniciativas, o que produziu correspondências parciais na maioria dos casos e uma correspondência total, reafirmando o padrão prospectivo já discutido.

O Eixo 5 – Infraestrutura Física, que abrange biblioteca, salas de aula, serviços de manutenção, espaços de trabalho e estudo, atendimento docente e acesso à internet, é o eixo de maior aderência entre os métodos. Das dezoito observações do triênio, dezessete tiveram ações registradas pela comissão, e todas convergiram com o sistema, com predominância de correspondências totais. Esse resultado é coerente com o perfil das respostas, onde os itens de infraestrutura foram os mais críticos da avaliação. Internet, salas de aula e manutenção reuniram os

maiores índices de insatisfação, o que levou tanto a comissão quanto os modelos a apontarem, de forma convergente, ações objetivas de diagnóstico técnico, ampliação de cobertura de rede, modernização de equipamentos e melhoria dos espaços. Essa aderência quase perfeita ilustra a orientação de Li *et al.* (2017) e de Russell e Norvig (2024), diante de dados quantitativos concretos e de sinal inequívoco, a inteligência artificial mostra-se especialmente eficaz em analisá-los e converter esse diagnóstico em decisões precisas, aqui na forma de ações de superação praticamente idênticas às da comissão.

Em síntese, a análise por eixo demonstra que a convergência entre o sistema e a comissão é robusta e distribuída, não se restringindo a dimensões específicas. Onde a comissão registrou ações de superação, o sistema quase sempre as reproduziu. As correspondências parciais concentraram-se justamente nos eixos em que a comissão descreveu iniciativas já executadas, e a única ausência entre as questões com ação resultou de um relato de resultados, e não de uma proposição de melhoria. Essa competência distribuída é coerente com a caracterização de Ludermir (2021) e de Mendonça *et al.* (2024) sobre o estágio atual da inteligência artificial, capaz de desempenho comparável ao humano quando aplicada a um escopo delimitado. As numerosas ausências observadas no Eixo 3 e, em menor grau, nos demais eixos referem-se a questões em que a comissão não apontou ações, nas quais o sistema, ainda assim, ofertou proposições pertinentes.

4.4 ANÁLISE COMPARATIVA ENTRE OS MODELOS DE INTELIGÊNCIA ARTIFICIAL

A integração simultânea de quatro modelos de inteligência artificial generativa ao sistema Auto Report atende a um propósito metodológico que ultrapassa a simples geração de ações, constituindo um ambiente de experimentação que viabiliza o confronto entre as soluções. Esta seção dedica-se a essa comparação, examinando o desempenho relativo de ChatGPT, DeepSeek, Gemini e Claude tanto na aderência às ações institucionais apontadas pela comissão quanto no tempo de processamento demandado. Mais do que descrever os resultados, busca-se aqui verificar em que medida o comportamento de cada modelo corresponde às características que o referencial teórico lhes atribui, encontrando as equivalências entre o que os autores orientam e o que a pesquisa observou.

Tomando como referência o mesmo conjunto de dados e o mesmo comando submetidos a todos os provedores, o Quadro 8 consolida, para os três anos reunidos, o número total de correspondências de cada modelo em relação ao relatório manual, sobre o universo de sessenta e três observações (vinte e uma questões em cada um dos três anos).

Quadro 8 – Correspondências por modelo no conjunto de três anos

Modelo	Correspondência total	Correspondência parcial	Ausência de corresp.	Algum grau de corresp.
ChatGPT (gpt-5.5)	13	23	27	36 (57,1%)
Deepseek (deepseek-chat)	14	22	27	36 (57,1%)
Gemini (gemini-3.5-flash)	11	25	27	36 (57,1%)
Claude (claude-sonnet-4.6)	15	21	27	36 (57,1%)

Fonte: autoria própria (2026).

O dado mais expressivo do Quadro 8 é a igualdade da taxa de convergência global entre os quatro modelos, já que todos alcançaram 36 correspondências de algum grau em 63 observações, o equivalente a 57,1%. Essa coincidência não é fortuita, decorre do fato de que as ausências se concentram nas mesmas questões, de modo que nenhum modelo consegue convergir onde não há ação manual a reproduzir. Em outras palavras, quando avaliados pelo critério de haver ou não correspondência, os quatro modelos são equivalentes, o que reforça que a uniformização do comando cumpriu seu papel metodológico de isolar o efeito do modelo. Essa convergência entre soluções de arquiteturas distintas, quando confrontadas com uma tarefa bem delimitada, é compatível com a leitura de Ludermir (2021), que situa o estágio atual da inteligência artificial na execução competente de tarefas específicas.

A diferenciação entre os modelos manifesta-se não na quantidade de convergências, mas na sua qualidade, isto é, na proporção de correspondências totais frente às parciais. Sob esse critério, o modelo de IA Claude destacou-se, com quinze correspondências totais, seguido pelo DeepSeek, com quatorze, pelo ChatGPT, com

treze, e pelo Gemini, com onze. Esse ordenamento encontra respaldo direto no referencial teórico. Subahi (2026) orienta que o Claude, por sua ênfase em segurança e precisão, alcança maior exatidão e as mais baixas taxas de erro em problemas de classificação, superando os concorrentes em precisão. Liu *et al.* (2026), no mesmo sentido, identificaram uma de suas versões como a mais exata dentre sete sistemas confrontados, com desempenho notável em coerência semântica e raciocínio fundamentado em evidências.

Na outra extremidade, o Gemini apresentou a menor incidência de correspondências totais, migrando com mais frequência para a faixa parcial. Também esse comportamento dialoga com o referencial conforme Subahi (2026), que descreve o Gemini como um sistema de desempenho regular nas atividades convencionais dos modelos de linguagem, orientado à eficiência, e Shean *et al.* (2026) observam que as configurações voltadas à eficiência tendem a priorizar latência e escalabilidade em detrimento da acurácia bruta em tarefas de raciocínio mais complexo. A variante *gemini-3.5-flash* empregada nesta pesquisa insere-se exatamente nessa categoria, o que ajuda a compreender por que suas respostas, embora pertinentes, alcançaram com menor frequência a correspondência total. O ChatGPT, por fim, posicionou-se em patamar intermediário e regular, condizente com a caracterização de Subahi (2026), que lhe atribui alta clareza e as médias gerais mais elevadas em conformidade sintática, atributos associados a um desempenho equilibrado.

O segundo eixo de comparação diz respeito ao tempo de geração dos relatórios, detalhado aqui por permitir o confronto direto entre as soluções. O Quadro 9 apresenta o tempo despendido por cada modelo na geração do conjunto completo de apontamentos, para as vinte e uma questões, em cada um dos três anos, bem como a média do período.

Quadro 9 – Tempo de geração dos relatórios por modelo (em minutos)

Modelo	2023	2024	2025	Média
Deepseek (deepseek-chat)	18,26	19,26	18,56	18,7
ChatGPT (gpt-5.5)	28,00	26,57	26,00	26,9
Claude (claude-sonnet-4.6)	43,49	42,34	45,08	43,6

Gemini (gemini-3.5-flash)	56,45	60,81	66,78	61,3
---------------------------	-------	-------	-------	------

Fonte: autoria própria (2026).

Os tempos revelam uma hierarquia estável ao longo dos três anos. O DeepSeek foi consistentemente o mais rápido, com média de 18,7 minutos, resultado condizente com a orientação de Subahi (2026), que atribui a esse modelo a economia computacional proporcionada por sua arquitetura de mistura de especialistas e a aptidão para produzir respostas confiáveis a partir de instruções sucintas. O ChatGPT posicionou-se como o segundo mais ágil, com média de 26,9 minutos. O Claude demandou tempo intermediário-alto, com média de 43,6 minutos, o que também encontra amparo no referencial de Subahi (2026), que adverte que a elevada exatidão do Claude vem acompanhada de alta demanda computacional, e Liu *et al.* (2026) que registram que o tempo de resposta mais alongado é a contrapartida de sua acurácia superior. A equivalência entre o observado e o orientado é notável, pois o modelo mais preciso foi também um dos mais lentos, exatamente o compromisso que a literatura antecipa.

O Gemini, por fim, foi o mais lento, com média de 61,3 minutos e tendência de crescimento ao longo dos anos. Esse resultado configura uma divergência aparente em relação ao referencial, que atribui ao Gemini um perfil de baixa latência e ampla escalabilidade (Subahi, 2026; Shean *et al.*, 2026). A divergência, contudo, não invalida a orientação teórica, mas evidencia a influência do contexto de integração. Conforme descrito nos procedimentos, a comunicação com o Gemini registrou indisponibilidades temporárias do serviço (erros de código 503), o que exigiu a implementação de um mecanismo de nova tentativa que, ao reenviar comandos após falhas, eleva o tempo total de processamento. Trata-se, portanto, de um caso em que as condições operacionais de uma aplicação real modularam o comportamento previsto pela literatura, distinção relevante para a interpretação dos resultados e coerente com a natureza avaliativa do Design Science Research.

Convém detalhar a natureza dessas indisponibilidades, por terem constituído o principal obstáculo operacional enfrentado na integração. De maneira recorrente, ao longo das tentativas de geração dos relatórios, o Gemini retornou uma mensagem de indisponibilidade temporária, com o seguinte teor: *“This model is currently experiencing high demand. Spikes in demand are usually temporary. Please try again*

later.”. Diferentemente de uma falha pontual, esse retorno manifestou-se de forma reiterada, interrompendo o processamento antes da conclusão do relatório e obrigando o sistema a reenviar os comandos. Foi precisamente essa recorrência que motivou a implementação do mecanismo de nova tentativa descrito nos procedimentos metodológicos, o qual, ao aguardar um breve intervalo e reencaminhar a requisição diante de cada ocorrência, garantiu a integralidade da geração, mas à custa de um tempo total substancialmente mais elevado. Esse comportamento explica, tanto a posição do Gemini como o modelo mais lento no Quadro 9 quanto a tendência de crescimento de seu tempo ao longo dos anos, e reforça que a divergência em relação ao perfil de baixa latência descrito na literatura decorre menos das características intrínsecas do modelo do que das condições de disponibilidade do serviço durante a realização da pesquisa.

Confrontados os dois critérios, delineia-se um quadro de compromissos entre exatidão e velocidade que reproduz, em boa medida, o retratado no referencial teórico. O Claude oferece a maior precisão, ao preço de um tempo de resposta mais elevado, enquanto o DeepSeek combina boa aderência com o menor tempo de execução, mostrando-se a alternativa mais equilibrada em termos de custo-benefício computacional. O ChatGPT apresenta desempenho intermediário e regular em ambas as dimensões, já o Gemini, na configuração adotada, reuniu a menor incidência de correspondências totais com o maior tempo de processamento. É importante ressaltar que essas diferenças não comprometem a validade global da solução, visto que todos os modelos reduziram o tempo de elaboração de meses para minutos e alcançaram idêntica taxa de convergência com o trabalho da comissão. A escolha entre eles configura-se, assim, menos como uma questão de eficácia e mais como uma decisão de conveniência operacional, a ser ponderada conforme a infraestrutura disponível e a prioridade atribuída à precisão ou à rapidez.

4.5 DISCUSSÃO DOS RESULTADOS E RESPOSTAS À QUESTÃO DE PESQUISA

As seções anteriores percorreram a dimensão quantitativa da comparação, mensurando o grau de convergência entre as ações geradas pelo sistema e aquelas registradas pela comissão própria de avaliação. Esta seção avança para a dimensão qualitativa, na qual o exame das divergências assume papel central. Mais do que aferir coincidências, interessa compreender o significado das ações destacadas

exclusivamente por um dos métodos, investigando se representam lacunas do sistema, redundâncias ou contribuições efetivamente novas trazidas pela abordagem automatizada.

O primeiro achado a ser discutido é a assimetria entre os dois recortes de análise. Considerando a totalidade das vinte e uma questões, a convergência situou-se entre 47,6% e 61,9%, excluídas as questões sem ação da comissão, ela se elevou para patamares entre 90,9% e 100%. Essa diferença, longe de indicar inconsistência, revela a distinção fundamental de propósito entre os dois métodos. A comissão, condicionada pelo tempo e pela leitura individual dos avaliadores, aponta ações de forma seletiva, concentrando-se nas dimensões que julga mais críticas e, por vezes, limitando-se ao diagnóstico. O sistema, por concepção, percorre todas as questões e propõe ações para cada uma, comportando-se de maneira exaustiva.

O exame qualitativo das ausências é, nesse sentido, particularmente revelador. Nas questões em que a comissão não apontou ações, majoritariamente nas dimensões de comunicação institucional e de participação do Eixo 3, o sistema não permaneceu omissor. Ofertou proposições consistentes, como a diversificação de canais de comunicação, a criação de painéis de transparência, a segmentação de mensagens por público e a institucionalização de espaços de escuta. A proximidade do pesquisador com a comissão avaliadora do campus, mencionada na delimitação do estudo, permitiu avaliar que tais proposições mostram-se pertinentes ao contexto institucional e aderentes aos problemas identificados nos próprios dados. Esse comportamento confirma a orientação de Li *et al.* (2017) quanto à capacidade da inteligência artificial de analisar grandes volumes de dados e deles extrair subsídios para a decisão. Ao cobrir as dimensões que a comissão havia deixado sem apontamento, o sistema não apenas reproduz, mas amplia o alcance da análise. Assim, boa parte das ausências não representa lacuna do sistema, mas contribuição complementar.

Um segundo achado qualitativo diz respeito à natureza das correspondências parciais, predominantes nos Eixos 2 e 4. Nesses eixos, as ações da comissão frequentemente assumiram a forma de relatos de iniciativas já executadas, enquanto o sistema desprovido de acesso a esse histórico de execução e alimentado apenas pela distribuição das respostas, propôs medidas prospectivas de aprimoramento. Essa diferença de perspectiva temporal é uma decorrência direta do modo de operação descrito por Ramos (2023), que descreve que o modelo generativo produz

suas respostas a partir da entrada e do contexto fornecidos, e não de um conhecimento situado sobre as ações em curso. A correspondência parcial expressa que a comissão relata o que fez, enquanto o sistema sugere o que fazer. Reconhecer essa distinção é relevante, pois indica que a integração entre os dois métodos é complementar, e não substitutiva, uma vez que o sistema não elimina o conhecimento da comissão, mas o enriquece com um leque sistemático de proposições de melhoria.

A consistência desses padrões ao longo dos três anos analisados confere robustez às conclusões. A elevada convergência nas questões com ação, a concentração das ausências nas questões sem ação e a predominância de correspondências parciais nos eixos de relato de execução repetiram-se em 2023, 2024 e 2025, evidenciando que o comportamento do sistema é estável e consistente ao longo dos três ciclos analisados nesta unidade. Essa regularidade dialoga com Russell e Norvig (2024), para quem o deslocamento do foco dos algoritmos para os dados é um dos traços da inteligência artificial contemporânea, cujo desempenho se sustenta na análise de grandes quantidades de informação. Alimentado a cada ciclo por um novo conjunto de respostas, o artefato manteve comportamento coerente, exatamente a operação que Ramos (2023) descreve ao tratar da transformação de dados em proposições por meio de comandos estruturados. Da mesma forma, a capacidade dos modelos de ajustar o nível de linguagem ao destinatário, apontada por Hang *et al.* (2026), favoreceu a produção de ações redigidas em termos diretamente aproveitáveis pela comissão, o que contribui para a utilidade prática do documento gerado.

Retomando a questão que norteia esta pesquisa “*de que modo a inteligência artificial contribui no apontamento de ações de superação identificadas na coleta de informações?*”, os resultados permitem uma resposta em três níveis, cada qual ancorado no referencial teórico. Em primeiro lugar, a inteligência artificial reproduz o trabalho da comissão, pois sempre que há uma ação manual a ser identificada, o sistema apresenta uma ação correspondente. Esse desempenho é compatível com a competência que Ludermitz (2021) e Mendonça *et al.* (2024) atribuem à inteligência artificial em tarefas delimitadas. Em segundo lugar, a inteligência artificial complementa esse trabalho, pois nas questões em que a comissão se limitou ao diagnóstico, o sistema ofertou proposições pertinentes, ampliando a cobertura das ações a dimensões antes negligenciadas. Esta constatação concretiza a capacidade analítica descrita por Li *et al.* (2017). Em terceiro lugar, a inteligência artificial aprimora

o processo em sua dimensão temporal, já que a geração dos apontamentos, antes conduzida manualmente ao longo de cerca de quatro meses, passa a ser produzida pelo sistema em poucos minutos, o que reduz significativamente o tempo de trabalho da comissão e materializa o princípio da eficiência tal como orientado por Matias-Pereira (2012), Camargo e Guimarães (2013) e Coelho (2011), sem prescindir da apreciação crítica humana, uma vez que as ações permanecem submetidas à validação da comissão antes de sua incorporação ao relatório final.

Cabe registrar, por fim, que a contribuição da inteligência artificial não se apresentou como um substituto do julgamento humano, mas como um instrumento de apoio à decisão. As ações geradas automaticamente, ainda que pertinentes, tendem a assumir caráter mais genérico e prospectivo, carecendo do conhecimento situado que a comissão detém sobre as iniciativas em andamento e sobre as particularidades do campus. Essa constatação está em harmonia com Coelho (2011), para quem a eficiência na Administração Pública não se resume à redução de custos, mas visa à melhor alocação dos recursos para maximizar a satisfação do interesse público, finalidade que permanece sob responsabilidade dos avaliadores. A articulação entre a profundidade e a rapidez do sistema, de um lado, e a sensibilidade contextual da comissão, de outro, é o que confere ao artefato seu valor efetivo: a inteligência artificial contribui ao antecipar, sistematizar e ampliar o conjunto de ações de superação, oferecendo à comissão um ponto de partida qualificado que abrevia substancialmente o esforço de interpretação e redação, preservando a decisão final sob responsabilidade humana.

5 CONSIDERAÇÕES FINAIS

Esta dissertação teve origem na constatação de um gargalo operacional concreto no processo de autoavaliação institucional do Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul – Campus Bento Gonçalves. Instituído pela Lei nº 10.861/2004, o Sistema Nacional de Avaliação da Educação Superior (SINAES) atribui às comissões próprias de avaliação a responsabilidade de conduzir os processos internos de avaliação, sistematizar as informações coletadas e apontar ações de superação capazes de elevar a qualidade da instituição. No caso examinado, a interpretação dos dados e a redação dos relatórios, geral e específicos por campus, consumiam, em média, cerca de quatro meses de trabalho manual da comissão própria de avaliação (CPA) e das comissões próprias de avaliação locais, condição que retardava a devolutiva à comunidade acadêmica e comprometia a tempestividade das decisões de gestão apoiadas nesses resultados. Diante desse cenário, formulou-se a questão que norteou toda a investigação: como a inteligência artificial contribui no apontamento de ações de superação identificadas na coleta de informações?

Para responder a essa pergunta, o objetivo geral do trabalho consistiu em elaborar um sistema baseado em inteligência artificial capaz de identificar as ações institucionais apontadas nas pesquisas de autoavaliação do IFRS – Campus Bento Gonçalves. Esse objetivo foi alcançado por meio do desenvolvimento e da avaliação do artefato denominado Auto Report, concebido segundo a lógica da Design Science Research (DSR), método orientado à solução de problemas e à construção de artefatos generalizáveis a uma classe de problemas (Dresh; Lacerda; Antunes, 2015). Em essência, o sistema converte a distribuição das respostas da escala *Likert* de cada questão em comandos estruturados, que são submetidos a quatro modelos de inteligência artificial generativa (ChatGPT, Deepseek, Gemini e Claude), dos quais se obtêm proposições qualitativas de ação para cada dimensão avaliada. Mais do que consolidar estatisticamente os dados, o artefato reproduz de maneira automatizada, o raciocínio prescritivo que a comissão realiza ao interpretar os percentuais de concordância e discordância e, a partir deles, sugerir medidas de melhoria para a instituição.

Os objetivos específicos, por sua vez, foram integralmente contemplados ao longo do percurso da pesquisa. O primeiro deles, compreender os principais aspectos

da inteligência artificial, foi atendido no referencial teórico, que percorreu a história, os tipos e as linguagens da área, bem como as características dos modelos generativos contemporâneos e das interfaces de programação que viabilizam sua integração a sistemas de informação. O segundo, identificar a utilização de tecnologia no setor público, desenvolveu-se a partir da discussão sobre a eficiência administrativa, evidenciando o papel das tecnologias de informação e comunicação na modernização da Administração Pública. O terceiro, analisar a relação da inteligência artificial com a utilização de tecnologia no setor público, articulou esses dois campos, situando a IA na agenda de eficiência do serviço público. O quarto e último objetivo específico, propor uma interface intuitiva para utilização do sistema desenvolvido, materializou-se na concepção do Auto Report, cuja tela permite ao usuário selecionar o instrumento, o ciclo avaliativo e o modelo de inteligência artificial desejado, disparando, com poucos comandos, a geração automatizada do documento completo de apontamentos.

A avaliação do artefato, conduzida sobre as avaliações da comunidade interna dos ciclos de 2023, 2024 e 2025, produziu achados consistentes ao longo dos três ciclos analisados. Quando confrontadas com os relatórios elaborados manualmente pela comissão, as ações geradas pelo sistema apresentaram convergência entre 90,9% e 100% nas questões em que havia ação efetivamente registrada pela CPA, alcançando 92,9% em 2023, 90,9% em 2024 e 100% em 2025. Tais percentuais devem ser interpretados como evidência de viabilidade em um cenário específico e controlado, e não como um patamar necessariamente reproduzível em outras unidades ou instituições. Considerada a totalidade das vinte e uma questões avaliadas por ano, a convergência situou-se entre 47,6% e 61,9%, patamar que não decorre de falha do artefato, mas da assimetria de propósito entre os dois métodos. Enquanto a comissão aponta ações de forma seletiva, concentrando-se nas dimensões que julga mais críticas e, por vezes, limitando-se ao diagnóstico, o sistema percorre exaustivamente todas as questões e propõe uma ação para cada uma. O exame qualitativo revelou que boa parte das ausências de correspondência não representa lacuna do sistema, mas contribuição complementar, uma vez que, nas questões em que a comissão se absteve de apontar ações, o artefato ofertou proposições pertinentes e aderentes aos problemas identificados nos próprios dados.

Esses resultados permitiram responder à questão de pesquisa em três níveis, cada qual ancorado no referencial teórico. Em primeiro lugar, a inteligência artificial

reproduz o trabalho da comissão, pois, sempre que existe uma ação manual a ser identificada, o sistema apresenta uma ação correspondente. O desempenho é compatível com a competência que Ludermit (2021) e Mendonça *et al.* (2024) atribuem à IA na execução de tarefas delimitadas. Em segundo lugar, a inteligência artificial complementa esse trabalho, pois, nas questões em que a comissão se limitou ao diagnóstico, o sistema ampliou a cobertura da análise a dimensões antes negligenciadas, concretizando a capacidade analítica descrita por Li *et al.* (2017). Em terceiro lugar, a inteligência artificial aprimora o processo em sua dimensão temporal. A geração dos apontamentos, antes conduzida manualmente ao longo de cerca de quatro meses, passou a ser produzida pelo sistema em intervalos que variaram de aproximadamente dezoito minutos a pouco mais de uma hora, conforme o modelo empregado, o que reduz significativamente o tempo de trabalho da comissão e materializa o princípio da eficiência tal como orientado por Matias-Pereira (2012), Camargo e Guimarães (2013) e Coelho (2011), sem prescindir da posterior apreciação crítica das ações pela comissão.

A comparação entre os quatro modelos, viabilizada pela uniformização dos comandos, revelou equivalência na taxa global de convergência, de 57,1% para todos, diferenciando-os não na quantidade, mas na qualidade das correspondências e no tempo de processamento. O modelo Claude destacou-se pela maior precisão, com o maior número de correspondências totais, enquanto o Deepseek revelou-se o mais rápido. O Gemini apresentou a menor incidência de correspondências totais e o maior tempo de geração, condicionado por indisponibilidades temporárias do serviço que exigiram mecanismos de nova tentativa, e o ChatGPT posicionou-se em patamar intermediário e regular. Esses comportamentos confirmaram as caracterizações da literatura recente sobre os modelos (Subahi, 2026; Liu *et al.*, 2026; Shean *et al.*, 2026), reforçando a validade da avaliação conduzida.

Cabe destacar, que a contribuição da inteligência artificial não se apresentou como um substituto do julgamento humano, mas como um instrumento de apoio à decisão. As ações geradas automaticamente, ainda que pertinentes, tendem a assumir caráter mais genérico e prospectivo, carecendo do conhecimento situado que a comissão detém sobre as iniciativas em andamento e sobre as particularidades do campus. É justamente a articulação entre a profundidade e a rapidez do sistema, de um lado, e a sensibilidade contextual dos avaliadores, de outro, que confere ao artefato seu valor efetivo.

5.1 IMPLICAÇÕES TEÓRICAS

Do ponto de vista teórico, este estudo contribui para a literatura de Design Science Research ao apresentar um artefato de natureza prescritiva, concebido para a produção de apontamentos de ações a partir de pesquisas de autoavaliação institucional em instituições públicas de ensino. Conforme orientam Dresh, Lacerda e Antunes (2015), a DSR busca reduzir a distância entre teoria e prática, produzindo soluções que, embora satisfatórias e não necessariamente ótimas, sejam generalizáveis a outros contextos. Ao demonstrar a viabilidade e a utilidade do Auto Report em um caso concreto, a pesquisa oferece um conhecimento que pode ser compartilhado a outras comissões de avaliação e a outras instituições submetidas ao SINAES, ampliando o repertório de artefatos disponíveis para o enfrentamento desse tipo de problema.

A elevada convergência observada nas questões com ação registrada confirma a caracterização de Ludermitz (2021) e de Mendonça *et al.* (2024), para quem o estágio atual da tecnologia corresponde a algoritmos capazes de executar tarefas com competência comparável à humana quando aplicados a um escopo delimitado. O apontamento de ações de superação a partir da distribuição de respostas constitui precisamente uma tarefa dessa natureza, o que ajuda a explicar por que o sistema reproduziu com fidelidade a leitura da comissão sempre que havia uma ação manual a ser identificada. Da mesma forma, a capacidade do artefato de percorrer a totalidade das questões e de propor ações mesmo onde a análise manual se absteve confirma a orientação de Li *et al.* (2017), segundo a qual a inteligência artificial se destaca na tomada de decisão pela sua aptidão para processar grandes volumes de dados e deles extrair subsídios pertinentes.

Uma contribuição teórica particularmente relevante emergiu do exame das correspondências apenas parciais e da distinção entre o realizado e o proposto. Nas dimensões em que a comissão relatou iniciativas já executadas, o sistema, alimentado exclusivamente pela distribuição das respostas, formulou proposições prospectivas de aprimoramento. Esse fenômeno ilustra o modo de operação descrito por Ramos (2023), para quem o modelo generativo produz suas respostas a partir da entrada e do contexto que lhe são fornecidos, e não de um conhecimento situado sobre as ações em curso. A pesquisa evidencia uma fronteira concreta da aplicação desses modelos,

a qualidade das proposições é diretamente dependente da riqueza do contexto disponibilizado no comando, o que oferece subsídios para futuras discussões sobre engenharia de *prompt* em contextos organizacionais. A consistência do comportamento do artefato ao longo dos três ciclos, por fim, dialoga com Russell e Norvig (2024), para quem o deslocamento do foco dos algoritmos para os dados é um dos traços definidores da inteligência artificial contemporânea.

O estudo contribui, ainda, para a literatura comparativa entre modelos de linguagem. Ao submeter ChatGPT, Deepseek, Gemini e Claude a uma mesma tarefa, com dados reais, em língua portuguesa e em um ambiente de integração de produção, a pesquisa ofereceu corroboração às caracterizações de precisão, clareza e eficiência descritas por Subahi (2026), Liu *et al.* (2026) e Shean *et al.* (2026), habitualmente derivadas de cenários controlados. Merece registro a divergência observada no tempo de resposta do Gemini que, embora descrito na literatura como um modelo de baixa latência, apresentou o maior tempo de geração em virtude de indisponibilidades operacionais do serviço. Esse achado enriquece a discussão teórica ao evidenciar a distância entre o perfil de desempenho atribuído a um modelo em condições ideais e o seu comportamento efetivo em uma aplicação real, distinção coerente com a natureza avaliativa da Design Science Research.

Por fim, ao utilizar no contexto da autoavaliação institucional a lógica de eficiência que a literatura identifica em outros órgãos públicos, a pesquisa articula os campos da inteligência artificial e da administração pública. Os resultados dialogam diretamente com Alhosani e Alhashmi (2024), que apontam o aumento da produtividade, a diminuição das cargas de trabalho e a simplificação de procedimentos entre as principais contribuições da IA ao setor público, e com os casos brasileiros descritos por Desordi e Bona (2020) e por Gonçalves *et al.* (2020), estendendo essa agenda a um domínio ainda pouco explorado empiricamente.

5.2 IMPLICAÇÕES GERENCIAIS

No plano gerencial, a principal implicação do estudo reside na disponibilização de um instrumento capaz de reduzir significativamente o tempo dedicado à elaboração dos relatórios de autoavaliação. A geração dos apontamentos, antes conduzida manualmente ao longo de cerca de quatro meses, passa a ser produzida de forma automatizada em poucos minutos, diminuindo de maneira expressiva o tempo de

trabalho da comissão. Essa redução libera os membros da comissão própria de avaliação e das subcomissões para atividades de maior valor analítico, como a validação crítica das ações, a articulação com a gestão e o acompanhamento da implementação das medidas propostas. Ao preservar a qualidade dos apontamentos enquanto diminui o esforço humano empregado, o artefato responde diretamente ao princípio da eficiência que a legislação impõe à Administração Pública, entendido não apenas como redução de custos, mas como a melhor alocação dos recursos para maximizar a satisfação do interesse público (Coelho, 2011).

É fundamental, contudo, que os gestores compreendam o sistema como uma ferramenta de apoio à decisão, e não como um substituto do julgamento humano. As ações geradas automaticamente devem ser tomadas como um ponto de partida qualificado, que abrevia substancialmente o esforço de interpretação e redação, mas que permanece submetido à apreciação da comissão antes de sua incorporação ao relatório final. Essa articulação preserva o conhecimento situado dos avaliadores sobre as iniciativas em andamento e sobre as particularidades do campus, elementos que o modelo, alimentado apenas pela distribuição das respostas, não é capaz de captar. O uso responsável do artefato pressupõe, portanto, a manutenção do avaliador humano no centro do processo decisório.

Como o sistema percorre exaustivamente todas as questões, ele cobre dimensões que a comissão, por limitação de tempo ou por opção metodológica, deixou sem apontamento. Ao ofertar proposições pertinentes nessas dimensões, o artefato amplia o alcance da análise e reduz o risco de que aspectos relevantes da avaliação passem despercebidos, oferecendo à gestão um panorama mais completo de oportunidades de melhoria. Trata-se de um ganho qualitativo que se soma à economia de tempo e reforça a utilidade prática da solução.

Diante da equivalência na taxa global de convergência, a decisão passa a orientar-se por critérios como precisão, rapidez, estabilidade e custo. Quando a exatidão dos apontamentos for prioritária, modelos de maior precisão, como o Claude, mostram-se indicados, ainda que ao custo de maior tempo de processamento. Quando a agilidade for determinante, alternativas como o Deepseek podem ser preferíveis. Convém também frisar, que a implementação adote a repetição automática das requisições incompletas, evitando tanto a interrupção do processo por falhas pontuais quanto a sobrecarga do serviço, para lidar com as indisponibilidades

observadas em serviços de terceiros, a exemplo das que afetaram o desempenho do Gemini durante a pesquisa.

Do ponto de vista da governança da solução, a experiência revelou a necessidade de procedimentos de exportação e importação de dados, uma vez que o sistema de autoavaliação atualmente em operação no IFRS não permite a conexão direta de aplicações externas ao seu banco de dados. A concentração da lógica de construção dos comandos em um único ponto da aplicação, atributo de modularidade destacado por Nunamaker, Chen e Purdin (1991), facilita a manutenção, a auditoria e a evolução dos comandos, aspecto relevante para a sustentabilidade da ferramenta ao longo do tempo. Por fim, a natureza replicável do artefato sugere que os ganhos observados no Campus Bento Gonçalves podem ser estendidos aos demais campi do IFRS e a outras instituições submetidas ao SINAES, contribuindo para uma devolutiva mais rápida à comunidade e para um planejamento institucional mais responsivo às suas necessidades.

Por ser concebido a partir de um instrumento comum a todas as instituições submetidas ao SINAES, o artefato não está preso às particularidades do IFRS – campus Bento Gonçalves. Os demais campi do IFRS, os outros Institutos Federais e, de modo geral, as instituições de ensino superior (IES) que operam com comissões próprias de avaliação podem valer-se do mesmo desenho para converter a distribuição das respostas em apontamentos de ações, adaptando os comandos ao seu próprio dicionário de dimensões e à redação de seus instrumentos. A principal contribuição prática do estudo reside, portanto, em sua aplicabilidade, trata-se de um método replicável, de baixo custo e independente de fornecedores específicos, uma vez que opera sobre modelos de inteligência artificial acessíveis e sobre dados que a própria instituição já coleta.

Para as IES, os ganhos potenciais vão além da economia de tempo. A geração automatizada de apontamentos permite antecipar a devolutiva à comunidade acadêmica, encurtando o intervalo entre a coleta das percepções e a resposta institucional, o que tende a fortalecer a adesão dos respondentes e a própria cultura de avaliação. Ao cobrir exaustivamente todas as questões, o sistema oferece à gestão um panorama mais completo de oportunidades de melhoria, útil para instruir a elaboração e a revisão de instrumentos de planejamento, como o Plano de Desenvolvimento Institucional (PDI) e os planos de ação das unidades. A padronização dos apontamentos também viabiliza comparações longitudinais entre

ciclos e comparações entre unidades de uma mesma rede, criando condições para o compartilhamento de boas práticas e para um benchmarking interno mais consistente.

A adoção por outras instituições, contudo, pressupõe algumas condições que a experiência do estudo tornou evidentes. É necessário dispor de mecanismos de exportação e importação de dados, já que nem todos os sistemas de autoavaliação permitem a conexão direta de aplicações externas. Também torna-se indispensável preservar a validação crítica dos apontamentos pela comissão, mantendo o avaliador humano no centro da decisão. Observadas essas condições, o estudo oferece às IES não apenas um resultado, mas um caminho prático e auditável para incorporar a inteligência artificial à gestão da autoavaliação institucional.

5.3 LIMITAÇÕES DO ESTUDO E SUGESTÕES DE PESQUISAS FUTURAS

Como toda investigação científica, este estudo apresenta limitações que devem ser explicitadas para a correta interpretação de seus resultados e para o balizamento de pesquisas futuras. A primeira delas decorre do recorte adotado, a avaliação do artefato restringiu-se às pesquisas da comunidade interna do IFRS – Campus Bento Gonçalves, o que, embora coerente com a orientação da Design Science Research a uma classe de problemas, limita a generalização direta dos achados a outros instrumentos, segmentos e instituições. Os resultados obtidos devem, portanto, ser lidos como evidências consistentes de viabilidade e utilidade em um contexto específico, e não como conclusões universalmente extrapoláveis.

Cabe enfatizar que os percentuais de convergência relatados (entre 90,9% e 100% nas questões com ação) foram obtidos em um cenário controlado, com uma única unidade, um único instrumento e um mesmo conjunto de avaliadores. Eventuais estudos de replicação em outros campi, instituições ou instrumentos podem encontrar taxas distintas, sem que isso, por si só, invalide a abordagem proposta, cuja contribuição central reside no método e na viabilidade demonstrada, e não na magnitude exata dos índices.

Uma segunda limitação relaciona-se ao padrão de referência empregado na avaliação. A convergência das ações geradas pelo sistema foi mensurada em relação aos relatórios elaborados manualmente pela comissão, os quais são, eles próprios, produto de um processo seletivo e sujeito à leitura individual dos avaliadores. Assim, o parâmetro de comparação não constitui um padrão isento, mas um referencial

condicionado pelas escolhas e pelas limitações de tempo da própria comissão. Além disso, a avaliação da pertinência das ações propostas nas questões sem correspondência manual apoiou-se, em parte, na proximidade do pesquisador com a comissão avaliadora do campus, o que confere caráter qualitativo a esse julgamento e o torna passível de viés.

Uma terceira limitação diz respeito aos insumos fornecidos aos modelos. O sistema foi alimentado exclusivamente pela distribuição quantitativa das respostas na escala *Likert*, sem incorporar as respostas às questões abertas dos instrumentos nem o histórico de iniciativas já executadas pela instituição. Essa ausência de contexto explica a predominância de correspondências parciais nas dimensões marcadas por relatos de execução e sinaliza que o desempenho do artefato poderia ser aprimorado com a ampliação das informações submetidas. Por fim, os resultados encontram-se vinculados a versões específicas dos modelos utilizados, avaliadas em um determinado momento. Dada a rápida evolução da tecnologia, o desempenho relativo entre eles pode alterar-se ao longo do tempo, e as indisponibilidades operacionais observadas em um dos serviços podem não se repetir em outras condições.

A partir dessas limitações, apresentam-se diversas oportunidades de investigação futura. Uma primeira direção consiste em aplicar o artefato aos demais campi do IFRS e a outras instituições submetidas ao SINAES, bem como estendê-lo aos demais instrumentos de autoavaliação, ampliando a base empírica e testando a generalização da solução. Uma segunda direção envolve o enriquecimento dos comandos com o processamento das respostas às questões abertas, por meio de técnicas de processamento de linguagem natural e de análise de sentimento, e com a incorporação do histórico de ações já implementadas pela instituição, o que tende a reduzir as correspondências apenas parciais e a aproximar as proposições do sistema da realidade situada da comissão.

Uma terceira frente promissora é a condução de uma validação formal do artefato junto aos membros da comissão própria de avaliação, mediante instrumentos estruturados de concordância, comparações entre relatórios manuais e automatizados e medidas de confiabilidade entre avaliadores, o que conferiria maior robustez à aferição da qualidade das ações geradas. Por fim, sugere-se a realização de estudos que avaliem o impacto do uso continuado do sistema sobre a qualidade do planejamento institucional, bem como análises comparativas de custo entre os

diferentes modelos e provedores, contribuindo para decisões de adoção mais bem fundamentadas por parte das instituições públicas.

REFERÊNCIAS

- ALAWADHI, S. *et al.* **Building Understanding of Smart City Initiatives.** H.J. Scholl et al. (Eds.): EGOV 2012, LNCS 7443, p. 40–53, 2012.
- ALHOSANI, K.; ALHASHMI, S.M. **Opportunities, challenges, and benefits of AI innovation in government services: a review.** Discover Artificial Intelligence, 4. ed., art. n. 18, 2024.
- ANDRIOLA, W. B. **Desafios e necessidades que se apresentam às Comissões Próprias de Avaliação (CPAs) das Instituições de Educação Superior (IES) visando à implementação do SINAES.** In: RISTOFF, D.; ALMEIDA JÚNIOR, V. (org.). Avaliação participativa: perspectivas e desafios. Brasília: INEP, 2005. p. 57-70.
- ANIL, R. *et al.* **Gemini: a family of highly capable multimodal models.** arXiv:2312.11805, 2023. Disponível em: <https://arxiv.org/abs/2312.11805>. Acesso em: 10 de jul. de 2026.
- ARAGÃO, A. **O princípio da eficiência.** Revista de Direito Administrativo, Rio de Janeiro, 237: 1-6, 2004.
- ARAÚJO, T. *et al.* **Tecnologias de informação e comunicação aplicadas à gestão no setor público: o modelo e-república.** Rio de Janeiro, 2012.
- BAHIENSE, G. **Alinhamento estratégico de tecnologia da informação no setor público: A oferta de serviços eletrônicos em secretarias de fazenda no Brasil.** Fundação Getúlio Vargas, Escola de administração de empresas de São Paulo, São Paulo, 2002.
- BARDIN, L. **Análise de conteúdo.** Edição revista e ampliada. São Paulo: Edições 70, 2016.
- BAI, Y. *et al.* **Constitutional AI: harmlessness from AI feedback.** arXiv:2212.08073, 2022. Disponível em: <https://arxiv.org/abs/2212.08073>. Acesso em: 14 de jul. de 2026.
- BOSTROM, N. **Superinteligência: caminhos, perigos, estratégias.** Rio de Janeiro: DarkSide Books, 2018.
- BRASIL. **Constituição da República Federativa do Brasil, de 05 de outubro de 1988.** Brasília, DF: Diário Oficial da União, 1988.
- BRASIL. **Plano Diretor da Reforma do Aparelho do Estado.** Presidência da República, Câmara da Reforma do Estado, Ministério da Administração Federal e Reforma do Estado. Brasília, 1995.
- BRASIL. **Lei Nº 10.861, de 14 de abril de 2004.** Institui o Sistema Nacional de Avaliação da Educação Superior - SINAES e dá outras providências. Brasília, DF: Diário Oficial da União, 2004.

BRASIL. **Lei Nº 11.892, de 29 de dezembro de 2008.** Institui a Rede Federal de Educação Profissional, Científica e Tecnológica, cria os Institutos Federais de Educação, Ciência e Tecnologia, e dá outras providências. Brasília, DF: Diário Oficial da União, 2008.

BRESSER-PEREIRA, L. C. **Da administração pública burocrática à gerencial.** Revista do Serviço Público, Brasília, v. 47, n. 1, p. 7-40, 1996.

BROWN, T. B. *et al.* **Language models are few-shot learners.** NeurIPS, 2020. Curran Associates, 2020. p. 1877-1901.

CAMARGO, F.; GUIMARÃES, K. **O princípio da eficiência na gestão pública.** Revista CEPPG - CESUC - Centro de Ensino Superior de Catalão, Ano XVI nº 28, 2013.

CHIAVENATO, I. **Introdução à teoria geral da administração: uma visão abrangente da moderna administração das organizações.** 10. ed. São Paulo: Atlas, 2024.

COELHO, J. **Princípio de Eficiência: parâmetro para uma nova gestão.** Revista do TCU, v. 122, p. 56-63, 2011.

CORRÊA, R. A. A. **Qualidade de vida, qualidade no trabalho, qualidade do trabalho, qualidade do atendimento público e competitividade.** Revista de Administração pública, Rio de Janeiro, v. 27, n. 1, p. 113-23, 1993.

DIAS SOBRINHO, J. **Avaliação: políticas educacionais e reformas da educação superior.** São Paulo: Cortez, 2003.

DESORDI, D.; BONA, C. **A inteligência artificial e a eficiência na administração pública.** Revista de Direito, Viçosa, ISSN 2527-0389, v.12, n. 2, 2020.

DRESH, A.; LACERDA, D.; ANTUNES, J. **Design Science Research: método de pesquisa para avanço da ciência e tecnologia.** 1. ed. Porto Alegre, Bookman: 2015.

GALLARDO, C.; MACHUCA, C.; SEMBLANTES, Y. **ChatGPT API: Brief overview and integration in Software Development.** International Journal of Engineering Insights: v. 1, n.1, Regular Paper, 2023.

GIL, A. **Como elaborar projetos de pesquisa.** 7. ed. Barueri: Atlas, 2022.

GONÇALVES, W. *et al.* **Terceirização de serviços de tecnologia da informação no setor público: uma abordagem usando o AHP.** Gestão Contemporânea, Vila Velha, v.10, n.1, p. 54-76, 2020.

GOODFELLOW, I. *et al.* **Generative Adversarial Networks.** Advances in Neural Information Processing Systems, Nova Iorque, 2014.

GUEDES, I; CÂNDIDO, G. **Modernização Administrativa no Setor Público: utilização de tecnologias de gestão e de informação para o gerenciamento do conhecimento.** Revista Gestão.Org, v. 6, n. 2, p. 158-173, 2008.

GUO, D. *et al.* **DeepSeek-R1 incentivizes reasoning in LLMs through reinforcement learning.** Nature, v. 645, p. 633-638, 2025.

HAENLEIN, M.; KAPLAN, A. **A brief history of artificial intelligence: on the past, present, and future of artificial intelligence.** California Management Review, v. 61, n. 4, p. 5-14, 2019.

HANG, Y. *et al.* **Evaluation of large Language models on pediatric asthma: a comparative study of Claude3-Opus, Gemini 2.0, ChatGPT-4o, and DeepSeek - a cross-sectional questionnaire study.** BMC Medical Informatics and Decision Making, v. 26, art. n. 77. Shangai: 2026.

HOOD, C. **A public management for all seasons?** Public Administration, v. 69, n. 1, p. 3-19, 1991.

IFRS. **Resolução Nº 068, de 28 de julho de 2010.** Bento Gonçalves, RS: IFRS, 2010a.

IFRS. **Resolução Nº 657, de 15 de setembro de 2010.** Bento Gonçalves, RS: IFRS, 2010b.

IFRS. **Programa de autoavaliação do Instituto Federal de Educação Ciência e Tecnologia do Rio Grande do Sul – PAIFRS.** Bento Gonçalves, RS: IFRS, 2012.

IFRS. **Plano de desenvolvimento institucional 2024-2028.** Aprovado pelo Conselho Superior do IFRS, conforma a resolução Nº 54 de 12 de dezembro de 2023. Bento Gonçalves, RS: IFRS, 2023.

LACERDA, D.; DRESH, A.; PROENÇA, A.; ANTUNES, J. **Design Science Research: método de pesquisa para a engenharia de produção.** Revista Gestão e Produção, v. 20, n. 4, p. 741-761, São Carlos, 2013.

LECUN, Y.; BENGIO, Y.; HINTON, G. **Deep learning.** Nature, v. 521, n. 7553, p. 436-444, 2015.

LEWIS, P. *et al.* **Retrieval-augmented generation for knowledge-intensive NLP tasks.** NeurIPS, 2020. Curran Associates, 2020. p. 9459-9474.

LI, B.; Hou, B.; Yu, W.; Lu, X.; Yang, C; **Applications of artificial intelligence in intelligent manufacturing: A review.** Frontiers of Information Technology & Electronic Engineering. 18(1), p. 86–96., Zhejiang: 2017.

LIU, H. *et al.* **LLM-driven collaborative framework for knowledge-enhanced cancer pain assessment and management.** NPJ Digital Medicine, v. 9, art. n. 180, 2026.

LUDERMIR, T. **Inteligência Artificial e Aprendizado de Máquina: estado atual e tendências.** ESTUDOS AVANÇADOS (ONLINE), v. 35, p. 85-94, 2021.

MAKSYM, C. **Inteligência artificial aplicada nos serviços públicos rumo ao desenvolvimento sustentável.** International Journal of Digital Law | IJDL, Belo Horizonte, v. 2, n. 1, edição especial suplementar, 2021.

MARCONI, M.; LAKATOS, E. **Metodologia científica.** 8. ed. Barueri: Atlas, 2022.

MARTINS, G. **Estudo de caso: uma estratégia de pesquisa.** 2. ed. São Paulo: Atlas, 2008.

MATIAS-PEREIRA, J. **Curso de Gestão Estratégica na Administração Pública.** 1. ed. São Paulo: Atlas, 2012.

MATIAS-PEREIRA, J. **Manual de metodologia da pesquisa científica.** 4. ed. São Paulo: Atlas, 2019.

MENDONÇA, M. *et. al.* **Evolução da Inteligência Artificial: da IA fraca à IA forte na era dos *chatbots*, com instâncias na educação.** Ciência e tecnologia: Catalisadores da inovação / Organizador Ernane Rosa Martin. Ponta Grossa: Atena, 2024.

MERGEL, I.; EDELMANN, N.; HAUG, N. **Defining digital transformation: results from expert interviews.** Government Information Quarterly, v. 36, n. 4, art. n. 101385, 2019.

MORAES, A. **Direito Constitucional.** 40. ed. São Paulo: Atlas, 2024.

OLSEN, W. **Coleta de dados: debates e métodos fundamentais em pesquisa social.** 1. ed. Porto Alegre: Penso, 2015.

OUSTERHOUT, J. K. **Scripting: higher-level programming for the 21st century.** Computer, v. 31, n. 3, p. 23-30, 1998.

OUYANG, L. *et al.* **Training language models to follow instructions with human feedback.** NeurIPS, 2022. Curran Associates, 2022. p. 27730-27744.

PECI, A.; PIERANTI, O.; RODRIGUES, S. **Governança e New Public Management: convergências e contradições no contexto brasileiro.** O&S, v. 15, n. 46, p. 39-55, 2008.

POOLE, D. L.; MACKWORTH, A. K. **Artificial intelligence: foundations of computational agents.** 3. ed. Cambridge: Cambridge University Press, 2023.

RAMOS, A. **Inteligência Artificial Generativa baseada em grandes modelos de linguagem-ferramentas de uso na pesquisa acadêmica.** Universidade Federal do Rio Grande do Norte, Natal: 2023.

REIS, C.; SILVEIRA, S.; FERREIRA, M. **Autoavaliação em uma instituição federal de ensino superior: resultados e implicações**. Avaliação (Campinas), Sorocaba, v. 15, n. 3, p. 109-129, nov. 2010.

RIBEIRO, M. M.; SEGATTO, C. I. **Inteligência artificial nas organizações públicas brasileiras: heterogeneidades e capacidades em tecnologia da informação**. Revista de Administração Pública, Rio de Janeiro, v. 59, n. 1, art. n. e2024-0066, 2025.

RISTOFF, D.; GIOLO, J. O **SINAES como sistema**. Revista Brasileira de Pós-Graduação, Brasília, v. 3, n. 6, p. 193-213, 2006.

ROCHA, I.; KISSIMOTO, K. **Barreiras e benefícios na adoção de inteligência artificial e IoT na gestão de operação**. Revista de Administração Mackenzie - RAM, 23(4), São Paulo: 2022.

RUSSELL, S.; NORVIG, P. **Inteligência artificial: uma abordagem moderna**. 4. ed. Rio de Janeiro: Grupo Editorial Nacional S.A Publicado pelo selo LTC | Livros Técnicos e Científicos Ltda., 2024.

SAMPIERI, R.; COLLADO, C.; LUCIO, M. **Metodologia de Pesquisa**. 5. ed. Porto Alegre, Penso, 2013.

SCOTT, M. L. **Programming language pragmatics**. 4. ed. Waltham: Morgan Kaufmann, 2015.

SEARLE, J. R. **Minds, brains, and programs**. Behavioral and Brain Sciences, v. 3, n. 3, p. 417-424, 1980.

SEBESTA, R. **Conceitos de linguagens de programação**. 11. ed. Porto Alegre: Bookman, 2018.

SHEAN, R. S. *et. al.* **Comparative performance of Gemini 3 Pro and GPT-5 family models on ophthalmology board-style questions**. Ophthalmology Science. v. 6, n. 5, art. n. 101145, 2026.

SUBAHI, A. **Prompt Engineering for Advancing Software Engineering: Using Generative Language Models in the Development of Domain-Specific Languages**. IEEE Access, v. 14, p. 22834-22850. Al Jamoum: 2026.

TOLEDO, A. T.; MENDONÇA, M. **A aplicação da inteligência artificial na busca de eficiência pela administração pública**. Revista do Serviço Público, Brasília, v. 74, n. 2, p. 410-438, 2023.

VALLE, V.; GALLO, W. **Inteligência artificial e capacidades regulatórias do Estado no ambiente da administração pública digital**. Revista de Direito Administrativo e Constitucional, ano 20, n. 82, p. 67-86, Belo Horizonte, 2020.

VAN ZIJL, W.; BURNHAM, K.; ECIM, D. **How close is AI to replacing accounting consultants? Insights from a comparative study of multiple AI models and exit-**

level accounting students. *Meditari Accountancy Research*, v. 34, n. 7, p. 53-78, 2026.

VASWANI, A. *et al.* **Attention is all you need.** NIPS, 2017. Curran Associates, 2017. p. 5998-6008.

WHITE, J. *et al.* **A prompt pattern catalog to enhance prompt engineering with ChatGPT.** arXiv:2302.11382, 2023. Disponível em: <https://arxiv.org/abs/2302.11382>. Acesso em: 11 de jul. de 2026.

WIRTZ, B. W.; WEYERER, J. C.; GEYER, C. **Artificial intelligence and the public sector: applications and challenges.** *International Journal of Public Administration*, v. 42, n. 7, p. 596-615, 2019.

ZUIDERWIJK, A.; CHEN, Y.-C.; SALEM, F. **Implications of the use of artificial intelligence in public governance: a systematic literature review and a research agenda.** *Government Information Quarterly*, v. 38, n. 3, art. n. 101577, 2021.

APÊNDICE

AutoReport — Manual do Usuário

AutoReport

Sistema de Geração de Relatórios de Avaliação com Inteligência Artificial

MANUAL DO USUÁRIO

Guia completo de operação do sistema

© AutoReport — 2026

Sumário

- 1. Introdução
 - 1.1. Para que serve o sistema
 - 1.2. Principais conceitos
- 2. Modelo de dados
 - 2.1. Entidades e atributos
 - 2.2. Como as entidades se relacionam
- 3. Acesso ao sistema
 - 3.1. Tela de login
- 4. Visão geral da interface
 - 4.1. Elementos comuns a todas as telas
 - 4.2. Painel inicial (Início)
- 5. Modelos de IA e troca rápida
- 6. Menu do usuário e perfil
 - 6.1. Menu do usuário
 - 6.2. Atualização de perfil
- 7. Módulos de cadastro e listagem
 - 7.1. Padrão comum de listagem
 - 7.2. Avaliações
 - 7.3. Campi
 - 7.4. Instrumentos
 - 7.5. Perguntas
 - 7.6. Respostas
 - 7.7. Usuários
- 8. Importação de respostas (CSV)
- 9. Geração de relatórios com IA
- 10. Configurações
- 11. Encerramento de sessão
- 12. Fluxo de uso resumido

1. Introdução

O AutoReport é um sistema web voltado à análise de dados de avaliações institucionais. Ele centraliza todo o ciclo de trabalho de uma avaliação — desde o cadastro das perguntas e dos participantes até o armazenamento das respostas — e entrega, ao final, relatórios analíticos gerados automaticamente com o apoio de modelos de Inteligência Artificial (IA).

Este manual descreve, tela a tela, todas as funcionalidades disponíveis. Ele foi elaborado para que qualquer usuário consiga operar o sistema com autonomia, independentemente de conhecimento técnico prévio.

1.1. Para que serve o sistema

Instituições de ensino realizam periodicamente avaliações institucionais, como as conduzidas pela Comissão Própria de Avaliação (CPA). Essas avaliações envolvem diferentes públicos (discentes, docentes e técnicos-administrativos), diversos campi e um grande volume de respostas. O AutoReport organiza esse conjunto de informações e transforma dados brutos em relatórios prontos para leitura e tomada de decisão.

Entre os principais benefícios do sistema estão:

- Centralização do cadastro de avaliações, instrumentos, perguntas, campi e respostas em um único lugar.
- Importação em massa de respostas a partir de arquivos CSV, dispensando o cadastro manual item a item.
- Geração automática de relatórios analíticos com apoio de IA, disponibilizados para download.
- Troca rápida entre diferentes modelos de IA integrados, conforme a preferência do usuário.

1.2. Principais conceitos

Antes de utilizar o sistema, é útil compreender os elementos que o compõem. A seção seguinte, sobre o modelo de dados, detalha como esses conceitos se relacionam.

- **Avaliação:** Uma campanha de avaliação identificada por um nome e um ano — por exemplo, “Avaliação Institucional - CPA”, referente a 2023.
- **Instrumento:** Um questionário aplicado dentro de uma avaliação, associado a um segmento (público-alvo). Uma mesma avaliação pode ter vários instrumentos.
- **Segmento:** O público a que o instrumento se destina, como discente, docente ou técnico-administrativo.
- **Pergunta:** Cada questão que compõe um instrumento, com o texto da pergunta e observações opcionais.
- **Campus:** Unidade física da instituição, identificada por nome e cidade (o plural utilizado no sistema é “campi”).
- **Resposta:** O registro consolidado que cruza uma pergunta com um campus, guardando um valor (por exemplo, uma nota) e a quantidade de respondentes que o assinalaram.
- **Relatório:** Documento analítico produzido a partir de uma avaliação ou instrumento, gerado com o modelo de IA configurado e disponibilizado para download.

2. Modelo de dados

O diagrama de classes a seguir representa as entidades do sistema e como elas se relacionam. Compreender essa estrutura ajuda a entender por que alguns cadastros exigem que outros já existam previamente.

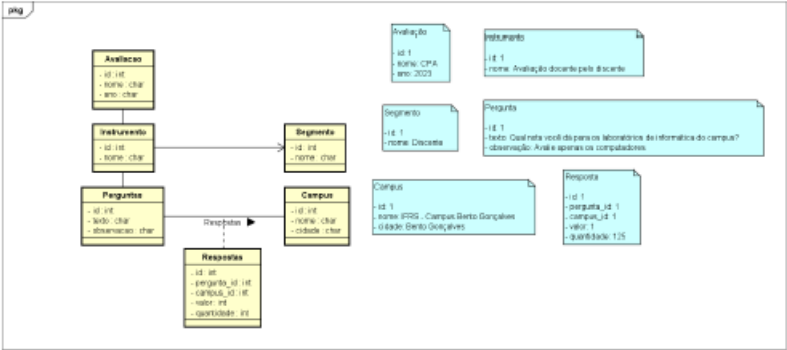


Diagrama de classes do AutoReport.

2.1. Entidades e atributos

As entidades do sistema, com seus respectivos atributos, são:

Campo	Obrigatório	Descrição
Avaliação	—	id, nome e ano. Representa a campanha de avaliação.
Instrumento	—	id e nome. Vincula-se a uma Avaliação e a um Segmento.
Segmento	—	id e nome. Define o público-alvo do instrumento.
Pergunta	—	id, texto e observação. Pertence a um Instrumento.
Campus	—	id, nome e cidade. Unidade em que as respostas são coletadas.
Resposta	—	id, pergunta_id, campus_id, valor e quantidade. Consolida as respostas.

2.2. Como as entidades se relacionam

- Uma Avaliação reúne um ou mais Instrumentos.
- Cada Instrumento está associado a um Segmento (o público que responde) e é composto por várias Perguntas.
- As Respostas ligam uma Pergunta a um Campus. Cada registro de resposta guarda um valor (por exemplo, “Nota 1”, “Nota 2” etc.) e a quantidade de pessoas que escolheram aquele valor.

Exemplo prático

Para a pergunta “Qual nota você dá para os laboratórios de informática do campus?”, aplicada no campus IFRS - Campus Bento Gonçalves, uma resposta pode registrar o valor 1 com quantidade 125 — ou seja, 125 respondentes atribuíram nota 1. Assim, as respostas são armazenadas de forma consolidada, e não individualmente.

Ordem recomendada de cadastro

Como as entidades dependem umas das outras, cadastre-as nesta ordem: 1) Avaliação; 2) Campi; 3) Instrumento (que exige avaliação e segmento); 4) Perguntas (que exigem instrumento); 5) Respostas (que exigem pergunta e campus). Depois desses cadastros, os relatórios já podem ser gerados.

3. Acesso ao sistema

3.1. Tela de login

Ao abrir o AutoReport, é exibida a tela de acesso. Nela, o usuário informa suas credenciais para entrar no sistema.



Tela de login — “Acesse sua conta”.

Para entrar, siga os passos:

1. Informe o e-mail cadastrado no campo “E-mail”.
2. Informe a senha no campo “Senha”. O ícone de olho, à direita do campo, permite exibir ou ocultar a senha digitada.
3. Clique em “Entrar”.

A tela oferece ainda dois recursos complementares:

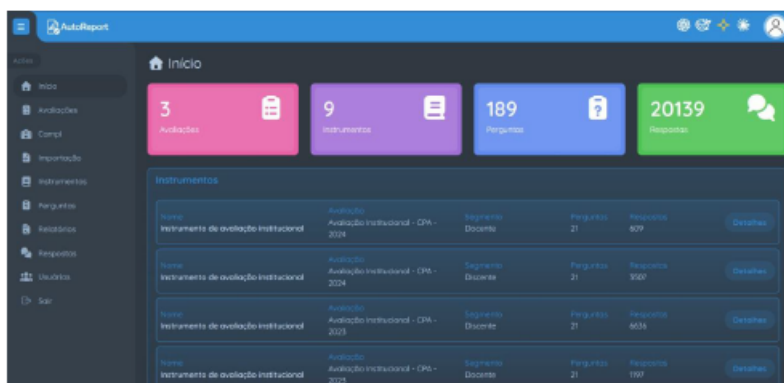
- **Esqueceu sua senha?:** link para iniciar a recuperação de acesso, caso o usuário não se lembre da senha.
- **Ainda não cadastrado? Registre-se:** botão que direciona ao cadastro de um novo usuário.

Dica

Os campos marcados com asterisco vermelho (*) são de preenchimento obrigatório. Esse padrão vale para todos os formulários do sistema.

4. Visão geral da interface

Após o login, o usuário é levado ao painel inicial. Todas as telas internas compartilham a mesma estrutura, o que torna a navegação previsível.



Tela inicial (Início), com os indicadores e a lista de instrumentos.

4.1. Elementos comuns a todas as telas

- **Barra superior:** exibe o logotipo do AutoReport à esquerda, os ícones dos modelos de IA integrados ao centro-direita e o acesso ao menu do usuário na extremidade direita.
- **Menu lateral (“Ações”):** concentra os atalhos para todos os módulos: Início, Avaliações, Campi, Importação, Instrumentos, Perguntas, Relatórios, Respostas, Usuários e Sair.
- **Área de conteúdo:** ocupa o centro da tela e apresenta o módulo selecionado — uma listagem, um formulário ou um painel.
- **Rodapé:** traz a identificação “© AutoReport - 2026”.

4.2. Painel inicial (Início)

O painel inicial reúne, na parte superior, cartões com os totais gerais do sistema — por exemplo, a quantidade de avaliações, de instrumentos, de perguntas e de respostas cadastradas. Logo abaixo, é apresentada uma lista de instrumentos com as informações de nome, avaliação vinculada, segmento, número de perguntas e número de respostas. O botão “Detalhes”, ao lado de cada instrumento, abre informações mais completas sobre ele.

5. Modelos de IA e troca rápida

Um dos diferenciais do AutoReport é a integração com quatro modelos de Inteligência Artificial. Em todas as telas internas, os quatro logotipos aparecem na barra superior, próximos ao menu do usuário.

O comportamento dos ícones é o seguinte:

- **Ícone em amarelo:** indica o modelo de IA atualmente configurado, ou seja, o que será utilizado na geração de relatórios.
- **Ícones em branco:** representam os demais modelos disponíveis, que ainda não estão em uso.

A troca é imediata: basta clicar sobre o ícone do modelo desejado para que ele passe a ser o modelo ativo. Não é necessário abrir telas adicionais para essa alteração rápida — embora o mesmo ajuste também esteja disponível na tela de Configurações (ver a seção 11).

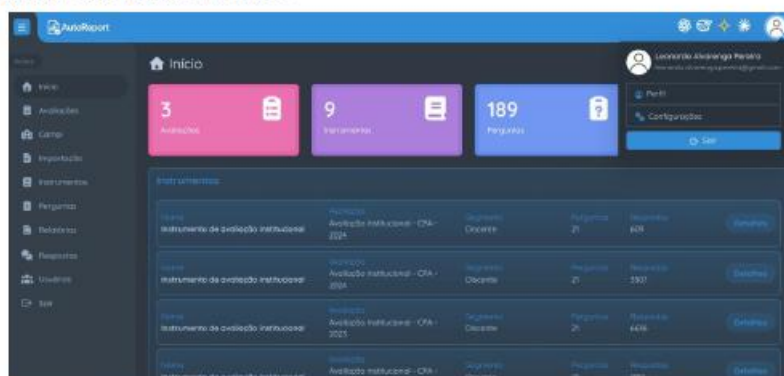
Por que trocar de modelo?

Cada modelo de IA tem características próprias de estilo, custo e disponibilidade. A troca rápida permite escolher, a cada geração de relatório, aquele que melhor atende à necessidade do momento, sem interromper o fluxo de trabalho.

6. Menu do usuário e perfil

6.1. Menu do usuário

Ao clicar no ícone de usuário, no canto superior direito, abre-se um menu suspenso com o nome e o e-mail do usuário logado, além de três opções.

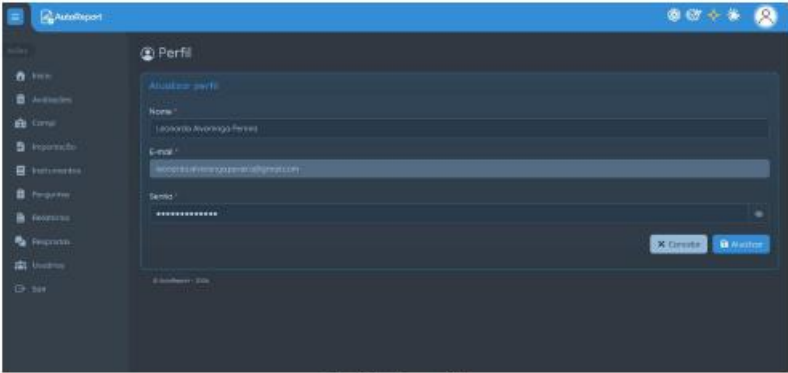


Menu do usuário aberto, com as opções Perfil, Configurações e Sair.

- **Perfil:** abre a tela de atualização dos dados pessoais.
- **Configurações:** abre a tela de configuração do modelo de IA.
- **Sair:** encerra a sessão e retorna à tela de login (mesma ação do item “Sair” do menu lateral).

6.2. Atualização de perfil

Na tela de perfil, o usuário pode revisar e atualizar seus próprios dados.



Tela "Atualizar perfil".

Campo	Obrigatório	Descrição
Nome	Sim	Nome completo do usuário.
E-mail	Sim	Endereço de e-mail associado à conta.
Senha	Sim	Senha de acesso; o ícone de olho permite exibir ou ocultar o valor digitado.

Após revisar os dados, clique em "Atualizar" para salvar as alterações ou em "Cancelar" para descartá-las.

7. Módulos de cadastro e listagem

Os módulos de dados do AutoReport seguem um padrão comum de operação. Compreendê-lo uma vez facilita o uso de todos eles.

7.1. Padrão comum de listagem

As telas de listagem (avaliações, campi, instrumentos, perguntas, respostas e usuários) apresentam:

- **Campo de busca:** para localizar registros por palavra-chave.
- **Botão “Adicionar”:** que abre o formulário de cadastro de um novo registro.
- **Tabela de registros:** com as colunas de dados e uma coluna “Ações”.
- **Rodapé de contagem:** que informa quantos registros estão sendo exibidos (por exemplo, “Mostrando 3 de 3 registros”).

Na coluna “Ações”, cada registro costuma oferecer três operações:

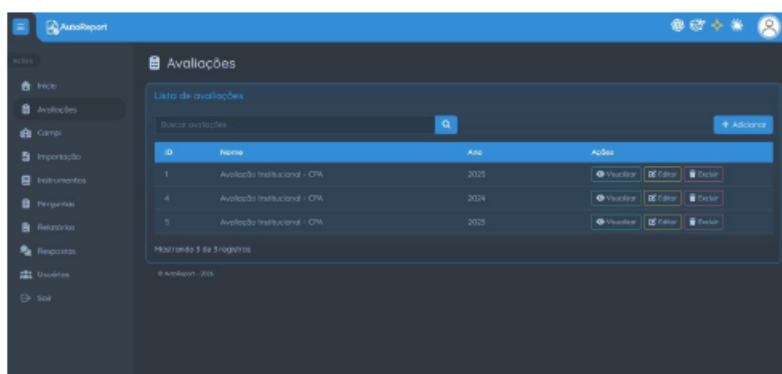
- **Visualizar:** abre os detalhes do registro somente para leitura.
- **Editar:** abre o formulário preenchido para alteração dos dados.
- **Excluir:** remove o registro do sistema (ação normalmente definitiva).

Atenção com a exclusão

A exclusão de um registro pode afetar outros que dependem dele. Por exemplo, excluir uma avaliação impacta os instrumentos, as perguntas e as respostas a ela vinculados. Utilize essa ação com cautela.

7.2. Avaliações

O módulo de avaliações lista todas as campanhas de avaliação cadastradas, exibindo o identificador, o nome e o ano de cada uma.



Listagem de avaliações.

Cadastro de avaliação

Ao clicar em “Adicionar”, é exibido o formulário de nova avaliação.

Formulário “Adicionar avaliação”.

Campo	Obrigatório	Descrição
Nome	Sim	Nome da avaliação — por exemplo, “Avaliação Institucional - CPA”.
Ano	Sim	Ano de referência da avaliação.

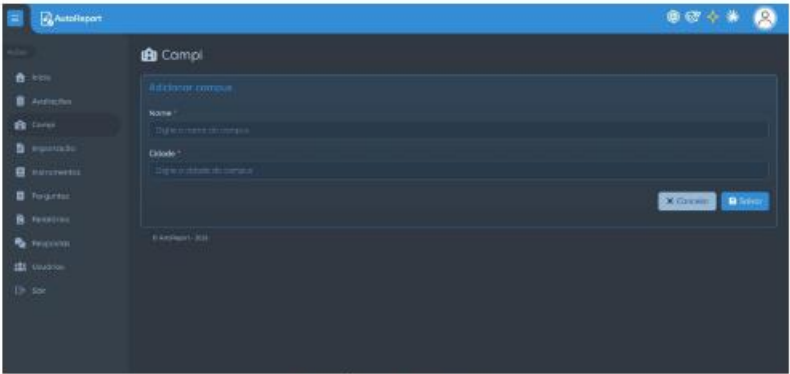
Clique em “Salvar” para gravar a avaliação ou em “Cancelar” para desistir.

7.3. Campi

O módulo de campi lista as unidades da instituição, com identificador, nome e cidade. “Campi” é o plural de “campus”.

Listagem de campi.

Cadastro de campus

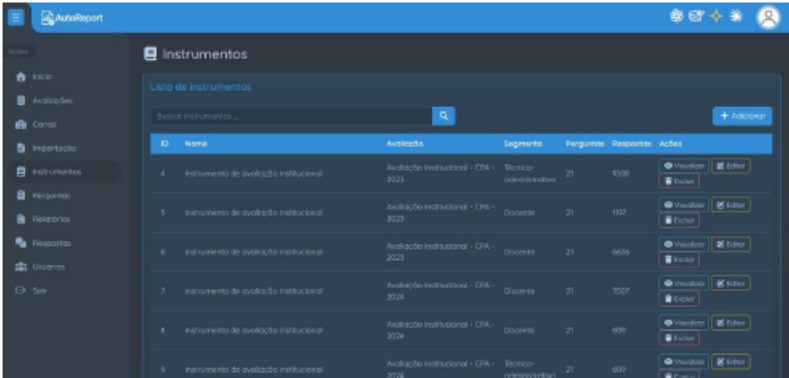


Formulário "Adicionar campus".

Campo	Obrigatório	Descrição
Nome	Sim	Nome do campus — por exemplo, "IFRS - campus Bento Gonçalves".
Cidade	Sim	Cidade onde o campus está localizado.

7.4. Instrumentos

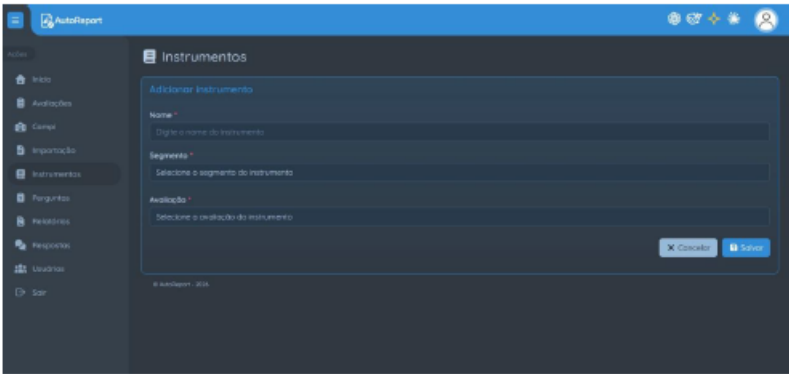
A listagem de instrumentos apresenta, além do identificador e do nome, a avaliação vinculada, o segmento, a quantidade de perguntas e a quantidade de respostas de cada instrumento. Como um instrumento pode reunir muitos registros, a lista pode se estender por várias linhas.



Listagem de instrumentos.

Cadastro de instrumento

O formulário de instrumento exige que a avaliação e o segmento correspondentes já existam, pois ambos são selecionados a partir de listas.

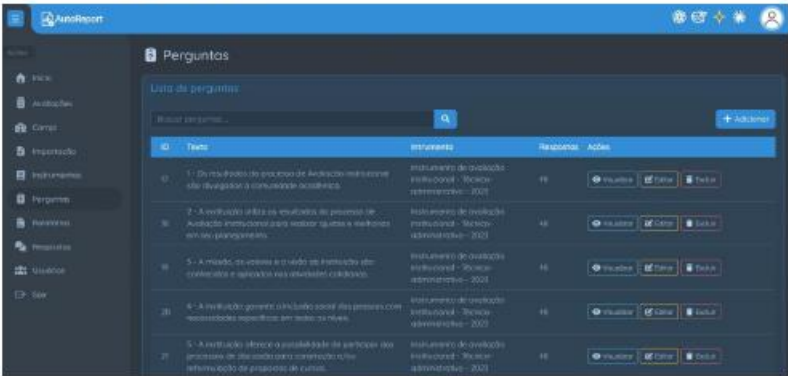


Formulário “Adicionar instrumento”.

Campo	Obrigatório	Descrição
Nome	Sim	Nome do instrumento.
Segmento	Sim	Público-alvo do instrumento (por exemplo, docente, discente ou técnico-administrativo), selecionado em uma lista.
Avaliação	Sim	Avaliação à qual o instrumento pertence, selecionada em uma lista.

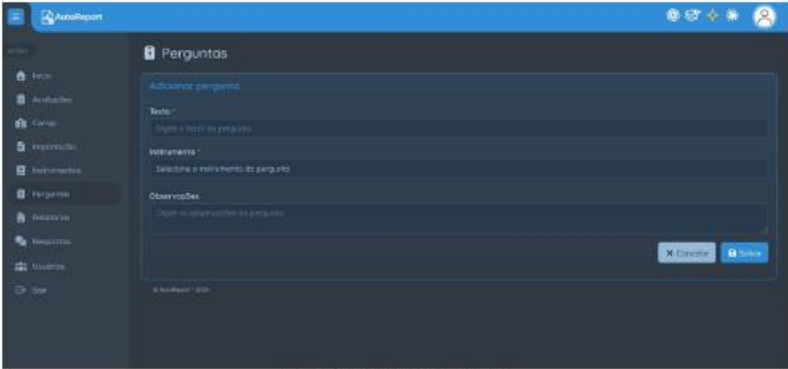
7.5. Perguntas

A listagem de perguntas mostra o identificador, o texto da pergunta, o instrumento a que ela pertence e a quantidade de respostas associadas.



Listagem de perguntas.

Cadastro de pergunta

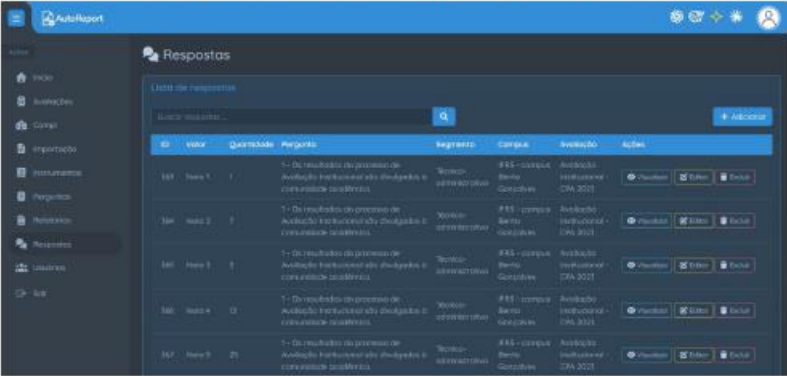


Formulário "Adicionar pergunta".

Campo	Obrigatório	Descrição
Texto	Sim	Enunciado da pergunta.
Instrumento	Sim	Instrumento ao qual a pergunta pertence, selecionado em uma lista.
Observações	Não	Anotações complementares sobre a pergunta (campo opcional).

7.6. Respostas

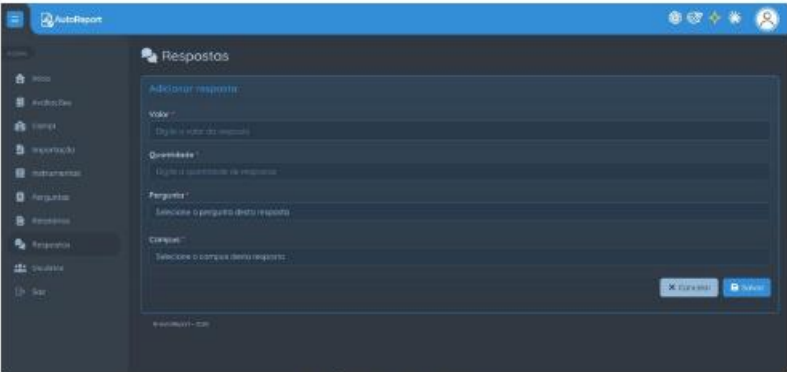
A listagem de respostas é a mais detalhada do sistema. Cada linha mostra o identificador, o valor, a quantidade, a pergunta, o segmento, o campus e a avaliação correspondentes. As respostas são armazenadas de forma consolidada: um mesmo enunciado gera várias linhas, uma para cada valor possível (por exemplo, “Nota 1”, “Nota 2” etc.), acompanhado da quantidade de respondentes.



Listagem de respostas.

Cadastro de resposta

O cadastro manual é útil para ajustes pontuais. Para grandes volumes, utilize a importação por CSV (seção 8).

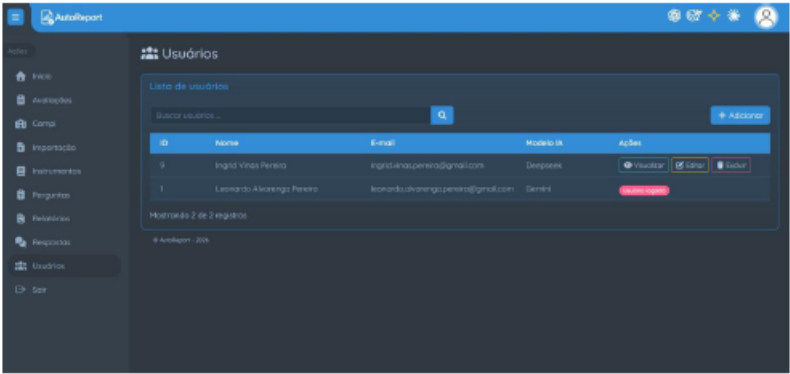


Formulário “Adicionar resposta”.

Campo	Obrigatório	Descrição
Valor	Sim	Valor da resposta (por exemplo, a nota atribuída).
Quantidade	Sim	Número de respondentes que assinalaram esse valor.
Pergunta	Sim	Pergunta a que a resposta se refere, selecionada em uma lista.
Campus	Sim	Campus em que a resposta foi coletada, selecionado em uma lista.

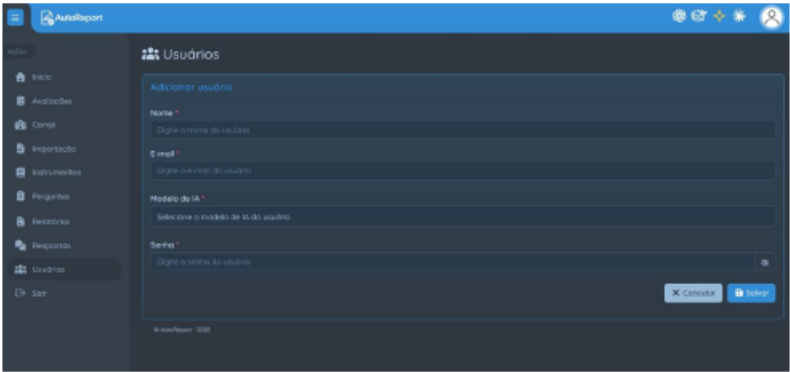
7.7. Usuários

O módulo de usuários lista as pessoas com acesso ao sistema, exibindo identificador, nome, e-mail e o modelo de IA associado a cada uma. O usuário atualmente autenticado é destacado com a marcação “Usuário logado”, e não apresenta as ações de edição e exclusão em sua própria linha da lista (a alteração dos próprios dados é feita pela tela de Perfil).



Listagem de usuários.

Cadastro de usuário

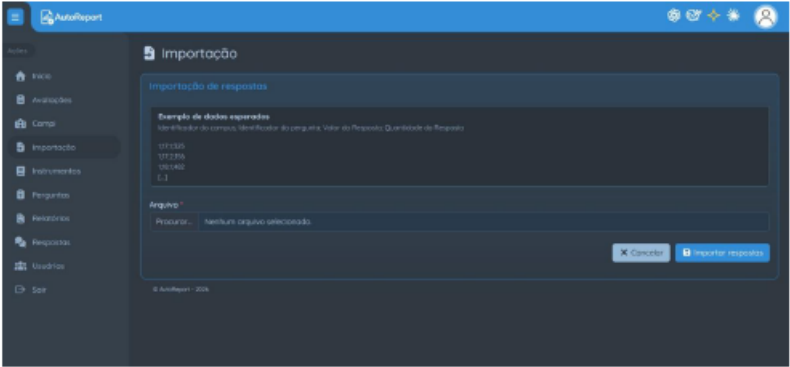


Formulário “Adicionar usuário”.

Campo	Obrigatório	Descrição
Nome	Sim	Nome completo do usuário.
E-mail	Sim	E-mail que servirá de login.
Modelo de IA	Sim	Modelo de IA preferido do usuário, selecionado em uma lista.
Senha	Sim	Senha de acesso; o ícone de olho permite exibir ou ocultar o valor.

8. Importação de respostas (CSV)

Cadastrar respostas uma a uma seria inviável em avaliações com milhares de registros. Por isso, o AutoReport oferece a importação de respostas a partir de um arquivo no formato CSV, acelerando enormemente esse trabalho.



Tela de importação de respostas.

8.1. Formato do arquivo

A própria tela exibe um exemplo dos dados esperados. Cada linha do arquivo deve conter, separados por ponto e vírgula, os seguintes campos, nesta ordem:

- Identificador do campus;
- Identificador da pergunta;
- Valor da resposta;
- Quantidade da resposta.

Um exemplo de conteúdo válido é:

1;17;1;325
1;17;2;156
1;18;1;482

No primeiro exemplo, para o campus de identificador 1 e a pergunta de identificador 17, foram registradas 325 respostas com o valor 1.

8.2. Como importar

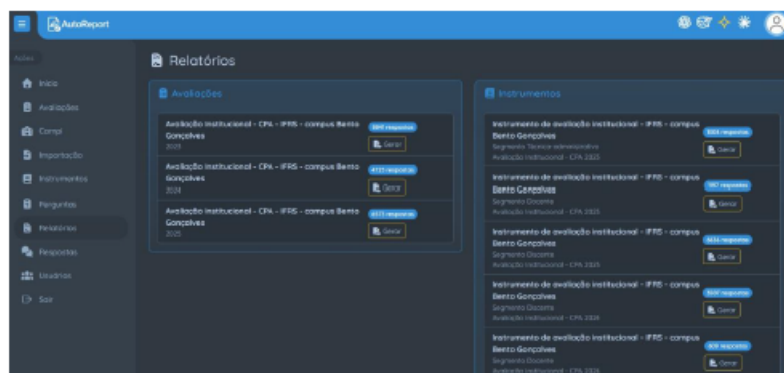
4. Acesse “Importação” no menu lateral.
5. Confira o exemplo de dados esperados exibido na tela e prepare seu arquivo CSV nesse mesmo formato.
6. Clique em “Procurar...” e selecione o arquivo no seu computador.
7. Clique em “Importar respostas” para processar o arquivo. Para desistir, clique em “Cancelar”.

Boas práticas

Confira se os identificadores de campus e de pergunta usados no arquivo correspondem exatamente aos cadastrados no sistema. Identificadores incorretos podem gerar respostas associadas ao registro errado.

9. Geração de relatórios com IA

A geração de relatórios é a funcionalidade central do AutoReport. Com base nas respostas coletadas, o sistema produz relatórios analíticos utilizando o modelo de IA configurado e os disponibiliza para download ao término do processamento.



Tela de relatórios, dividida em "Avaliações" e "Instrumentos".

9.1. Organização da tela

A tela de relatórios é dividida em duas colunas, que representam os dois níveis de agregação possíveis:

- **Avaliações:** gera um relatório consolidado de toda uma avaliação, considerando o conjunto de suas respostas. Cada item mostra o nome da avaliação, o ano e o total de respostas.
- **Instrumentos:** gera um relatório específico de um instrumento. Cada item mostra o nome do instrumento, o segmento, a avaliação vinculada e o total de respostas.

9.2. Como gerar um relatório

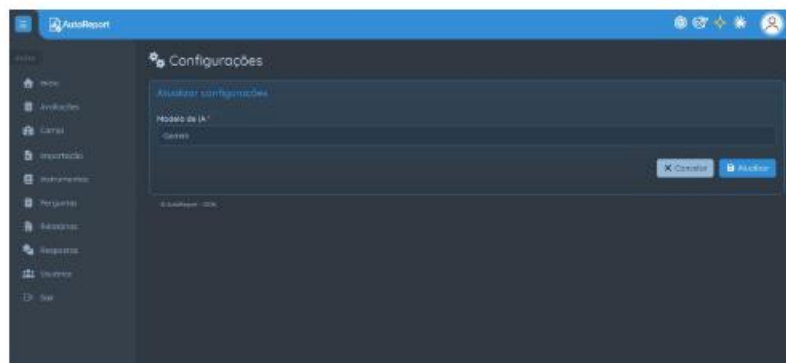
- Verifique, na barra superior, qual modelo de IA está ativo (ícone em amarelo). Se necessário, troque o modelo antes de gerar (ver seção 5).
- Localize, na coluna adequada, a avaliação ou o instrumento desejado.
- Clique no botão "Gerar" (identificado pelo ícone de PDF) correspondente àquele item.
- Aguarde o processamento. Ao final, o relatório é ofertado para download.

O modelo de IA importa

O relatório é produzido pelo modelo de IA que estiver configurado no momento da geração. Como a barra superior mostra o modelo ativo em amarelo, confirme essa escolha antes de clicar em "Gerar" para obter o resultado com o modelo desejado.

10. Configurações

A tela de configurações, acessível pelo menu do usuário, permite definir o modelo de IA de forma persistente — alternativa à troca rápida pelos ícones da barra superior.



Tela de configurações, com o modelo de IA selecionado (Gemini).

Campo	Obrigatório	Descrição
Modelo de IA	Sim	Modelo de Inteligência Artificial a ser utilizado nas gerações de relatório, selecionado em uma lista.

Após escolher o modelo, clique em "Atualizar" para gravar a preferência ou em "Cancelar" para manter a configuração anterior.

11. Encerramento de sessão

Ao terminar o uso do sistema, encerre a sessão para proteger seus dados. Há duas formas de fazê-lo:

- Clicar em "Sair", no final do menu lateral "Ações".
- Abrir o menu do usuário, no canto superior direito, e clicar em "Sair".

Em ambos os casos, o sistema encerra a sessão e retorna à tela de login.

12. Fluxo de uso resumido

Para uma primeira operação completa do AutoReport — do zero até o relatório — sugere-se o seguinte roteiro:

12. Acessar o sistema pela tela de login.
13. Cadastrar as avaliações (nome e ano).
14. Cadastrar os campi (nome e cidade).
15. Cadastrar os instrumentos, associando-os às avaliações e aos segmentos.
16. Cadastrar as perguntas de cada instrumento.
17. Registrar as respostas — manualmente ou, preferencialmente, por importação de arquivo CSV.
18. Confirmar o modelo de IA desejado na barra superior ou em Configurações.
19. Gerar o relatório da avaliação ou do instrumento e baixá-lo ao final do processamento.

Este manual descreve as funcionalidades observadas nas telas do sistema. Detalhes de comportamento podem variar conforme atualizações futuras do AutoReport.