



UNIVERSIDADE FEDERAL DO CEARÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
MESTRADO PROFISSIONAL EM POLÍTICAS PÚBLICAS E
GESTÃO DO ENSINO SUPERIOR

EDUARDO MELO ZINHANI

AVALIAÇÃO E GESTÃO DE CURSOS DE GRADUAÇÃO: DESENVOLVIMENTO
DE *FRAMEWORK* BASEADO EM INTELIGÊNCIA ARTIFICIAL

FORTALEZA

2026

EDUARDO MELO ZINHANI

**AVALIAÇÃO E GESTÃO DE CURSOS DE GRADUAÇÃO: DESENVOLVIMENTO DE
FRAMEWORK BASEADO EM INTELIGÊNCIA ARTIFICIAL**

Dissertação apresentada ao Programa de Pós-Graduação em Políticas Públicas e Gestão do Ensino Superior da Universidade Federal do Ceará (UFC), como requisito parcial à obtenção do título de Mestre em Políticas Públicas e Gestão do Ensino Superior. Área de concentração: Políticas Públicas e Gestão do Ensino Superior.

Orientador: Prof. Dr. Wagner Bandeira Andriola.

FORTALEZA

2026

EDUARDO MELO ZINHANI

AVALIAÇÃO E GESTÃO DE CURSOS DE GRADUAÇÃO: DESENVOLVIMENTO DE
FRAMEWORK BASEADO EM INTELIGÊNCIA ARTIFICIAL

Dissertação apresentada ao Programa de Pós-Graduação em Políticas Públicas e Gestão do Ensino Superior da Universidade Federal do Ceará (UFC), como requisito parcial à obtenção do título de Mestre em Políticas Públicas e Gestão do Ensino Superior. Área de concentração: Políticas Públicas e Gestão do Ensino Superior.

Aprovado em: 20/04/2026.

BANCA EXAMINADORA

Prof. Dr. Wagner Bandeira Andriola (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. João César Moura Mota
Universidade Federal do Ceará (UFC)

Prof. Dr. Alexandre Gonçalves Porto
Universidade do Estado de Mato Grosso (UNEMAT)

Prof. Dr. Benevid Felix da Silva
Universidade do Estado de Mato Grosso (UNEMAT)

Toda honra e toda glória sejam dadas ao senhor
nosso Deus.

AGRADECIMENTOS

A Deus, por me sustentar em cada etapa dessa jornada e por me dar forças nos momentos em que o cansaço quase venceu à vontade.

À minha esposa Flávia e aos meus filhos, pelo amor incondicional, pela paciência com as ausências e por serem o motivo de tudo isso fazer sentido.

Ao meu amigo Raphael Monteiro Pires, que me deu o empurrão que eu precisava e contribuiu diretamente para que este projeto se concretizasse. Sem essa força inicial, nada disso teria saído do papel.

À minha amiga Prof.^a Nilce Maria, companheira fiel de trabalho, sempre presente nos bastidores e que tanto contribuiu para o meu desenvolvimento profissional.

Ao amigo Prof. Benevid Felix da Silva, referência que carrego comigo na área da computação.

Ao amigo Prof. Alexandre Gonçalves Porto, pela parceria e amizade que o tempo só fortaleceu.

Ao meu orientador, Prof. Dr. Wagner Bandeira Andriola, pela condução segura e generosa desta pesquisa, pela confiança depositada no trabalho e pela disponibilidade constante.

Aos servidores da PROEG, em especial aos colegas Valdivina e Luizito da Diretoria de Gestão de Regulação do Ensino Superior, pelas valiosas contribuições e conhecimentos divididos, que viabilizaram a realização desta dissertação.

Aos colegas e professores do mestrado do POLEDUC, pelas trocas, reflexões e pelo companheirismo que tornaram essa caminhada mais leve.

À Universidade do Estado de Mato Grosso (UNEMAT), por viabilizar esta oportunidade de qualificação profissional, e à Universidade Federal do Ceará (UFC), pela excelência do programa que me acolheu.

A todos que, de alguma forma, contribuíram para a realização deste trabalho, meu sincero obrigado.

"O coração inteligente adquire o saber, o ouvido dos sábios procura a ciência."

(Provérbios 18,15)

RESUMO

A avaliação institucional de cursos de graduação no Brasil, regulamentada pelo Sistema Nacional de Avaliação da Educação Superior (SINAES), gera volumes significativos de dados avaliativos cuja tradução em planos de ação estratégicos frequentemente excede a capacidade analítica das Comissões Próprias de Avaliação (CPA) e dos coordenadores de curso, comprometendo a função formativa do sistema e reforçando uma cultura de conformidade em detrimento de uma cultura de qualidade. Esta dissertação tem como objetivo desenvolver e testar um framework baseado em agentes de Inteligência Artificial Generativa para apoiar os processos de avaliação institucional em cursos de graduação, transformando os critérios normativos do SINAES em diagnósticos estruturados e recomendações estratégicas acionáveis. A pesquisa adota o Design Science Research (DSR) como abordagem metodológica, operacionalizada pelo modelo DSRM, posicionando-se no quadrante de melhoria. Como contribuições metodológicas, propõem-se três artefatos analíticos construídos mediante Análise de Conteúdo aplicada ao Instrumento de Avaliação de Cursos de Graduação do INEP: a Matriz de Progressão Conceitual (MPC), que decompõe os cinco níveis conceituais de cada indicador em elementos constitutivos, atributos de progressão e evidências requeridas; a Matriz de Condições de Aplicabilidade (MCA), que codifica regras determinantes da aplicabilidade dos indicadores; e a Matriz de Compartilhamento de Evidências (MCE), que classifica o escopo de compartilhamento dos artefatos documentais. O framework foi materializado na plataforma Notacinco, aplicação web multi-tenant, com três agentes especializados de IA generativa (Diagnóstico, Recomendação e Síntese), e pipeline de Retrieval-Augmented Generation (RAG). A validação multimétodo, conduzida em sete cursos da Universidade do Estado de Mato Grosso (UNEMAT) com sessenta execuções, obteve *F1-Score* Geral de 0,9582, Acurácia do Diagnóstico de 1,0000, Coeficiente de Consistência Estrutural de 0,9759 e Índice de Consistência Semântica de 0,9509. Conclui-se que o framework contribui para transformar dados avaliativos em conhecimento acionável, apoiando coordenadores de curso e membros da CPA na formulação de planos de melhoria fundamentados em evidências e alinhados aos requisitos do SINAES.

Palavras-chave: avaliação institucional; SINAES; inteligência artificial generativa; *Design Science Research*; análise de conteúdo, interdisciplinar.

ABSTRACT

The institutional evaluation of undergraduate programs in Brazil, regulated by the National Higher Education Evaluation System (SINAES), generates significant volumes of assessment data whose translation into strategic action plans frequently exceeds the analytical capacity of Institutional Self-Assessment Committees (CPA) and course coordinators, compromising the formative function of the system and reinforcing a culture of compliance rather than a culture of quality. This dissertation aims to develop and test a framework based on Generative Artificial Intelligence agents to support the institutional evaluation processes of undergraduate programs, transforming the normative criteria of SINAES into structured diagnoses and actionable strategic recommendations. The research adopts Design Science Research (DSR) as its methodological approach, operationalized through the DSRM model, positioning the study within the improvement quadrant. As methodological contributions, three analytical artifacts are proposed, built through Content Analysis applied to INEP's Undergraduate Course Assessment Instrument: the Conceptual Progression Matrix (MPC), which decomposes the five conceptual levels of each indicator into constitutive elements, progression attributes, and required evidence; the Applicability Conditions Matrix (MCA), which codifies rules determining indicator applicability; and the Evidence Sharing Matrix (MCE), which classifies the scope of documentary artifact sharing. The framework was materialized in the Notacinco platform, a multi-tenant web application, featuring three specialized generative AI agents (Diagnosis, Recommendation, and Synthesis) and a Retrieval-Augmented Generation (RAG) pipeline. Multi-method validation, conducted across seven programs at the State University of Mato Grosso (UNEMAT) with sixty executions, yielded an Overall *FI-Score* of 0.9582, Diagnostic Accuracy of 1.0000, Structural Consistency Coefficient of 0.9759, and Semantic Consistency Index of 0.9509. It is concluded that the framework contributes to transforming evaluative data into actionable knowledge, supporting course coordinators and CPA members in formulating evidence-based improvement plans aligned with SINAES requirements.

Keywords: institutional evaluation; SINAES; generative artificial intelligence; Design Science Research; content analysis, interdisciplinary.

LISTA DE FIGURAS

Figura 1 – Proposições em diversos domínios (Propositions across various domains).	68
Figura 2 – Compartimentalização do genAI 1.0 e do genAI 2.0.....	72
Figura 3 – O ciclo de aprimoramento de habilidades do MAS para a educação.....	73
Figura 4 – Quadro de Contribuição de Conhecimento para a Pesquisa em Desenvolvimento	86
Figura 5 – Diagrama de fluxo: Integração da MPC com o <i>Framework</i> SINAES/IA.....	124
Figura 6 – Arquitetura em Camadas da Plataforma.....	146
Figura 7 – Diagrama de Sequência: Motor de Regras MCA (McaRuleEngine)	158
Figura 8 – Diagrama de Sequência: Serviço de Diagnóstico MPC.....	162
Figura 9 – Diagrama de fluxo de Dados do <i>Framework</i> : Os 4 Estágios.....	164
Figura 10 – Diagrama C4 (Nível Container): Visão Geral da Arquitetura	166
Figura 11 – Orquestração sequencial dos agentes de IA do <i>framework</i>	171
Figura 12 – Captura de tela das recomendações com código de cores para o gestor MCA	179
Figura 13 – Captura de tela do Relatório de Avaliação Integrado (Planos de Ação MPC).	180
Figura 14 – Captura de tela do cadastro/edição de curso e <i>upload</i> da DCN.	233
Figura 15 – Captura de tela da visualização de curso.....	234
Figura 16 – Captura de tela do <i>upload</i> do PPC do curso.....	235
Figura 17 – Captura de tela da configuração da Matriz de Condições de Aplicabilidade (MCA).....	236
Figura 18 – Captura de tela do cadastro de avaliação do curso.....	237
Figura 19 – Captura de tela de configuração da Matriz de Progressão Conceitual (MPC).	238
Figura 20 – Captura de tela do relatório de IA.	239
Figura 21 – Captura de tela da gestão de planos de ação e cadastro de evidências.	240
Figura 22 – Captura de tela da validação de artefatos pela Diretoria de Regulação.	241
Figura 23 – Captura de tela do relatório executivo.	242
Figura 24 – Captura de tela do Painel de Transparência de IA.	244

LISTA DE GRÁFICOS

Gráfico 1 – F1 Geral por cenário.....	187
Gráfico 2 – <i>Precision</i> vs. <i>Recall</i> por categoria e cenário.	191
Gráfico 3 – Métricas MCE por cenário.	193
Gráfico 4 – Distribuição dos valores de F1 Geral por cenário.	195
Gráfico 5 – Síntese visual dos resultados por cenário.	199
Gráfico 6 – F1 Geral por curso.	206
Gráfico 7 – <i>Precision</i> e <i>Recall</i> por categoria para cada curso.	209
Gráfico 8 – Métricas MCE por curso.	212
Gráfico 9 – Distribuição da classificação de relevância por sessão.	214
Gráfico 10 – Relação entre CC oficial e F1 Geral.....	218
Gráfico 11 – CCE e ICS por curso.	221
Gráfico 12 – Estabilidade por componente estrutural.	225
Gráfico 13 – Comparação entre CCE e ICS por curso.	228
Gráfico 14 – Visão integrada dos três procedimentos de validação.	230

LISTA DE TABELAS

Tabela 1 – Pesos dos componentes no conceito preliminar de curso	47
Tabela 2 – O Fluxo Avaliativo do SINAES por Faixa de CPC	48
Tabela 3 – Síntese dos procedimentos de validação do <i>framework</i>	98
Tabela 4 – Decomposição do Indicador 1.1 (Exemplo de Aplicação).....	116
Tabela 5 – Distribuição dos indicadores por dimensão e tipo de instrumento	118
Tabela 6 – Comparação semântica entre os instrumentos: exemplo do indicador 1.1	120
Tabela 7 – Decomposição do indicador 1.1 na MPC	121
Tabela 8 – Exemplo de regra de aplicabilidade (MCA) - Indicador 1.7	127
Tabela 9 – Domínios e variáveis determinantes da Matriz de Condições de Aplicabilidade (MCA).....	128
Tabela 10 – Catálogo completo de variáveis que compõem a MCA.....	130
Tabela 11 – Tipologia das regras de aplicabilidade da MCA	131
Tabela 12 – Padrões lógicos das regras de aplicabilidade da MCA	133
Tabela 13 – Exemplo de classificação de escopo de compartilhamento de evidências (MCE)	138
Tabela 14 – Distribuição dos indicadores por escopo de compartilhamento de evidências	140
Tabela 15 – Composição tecnológica adotada.....	145
Tabela 16 – Arquitetura com separação de interesses por Módulos.....	147
Tabela 17 – Distribuição quantitativa das entidades por agrupamento funcional	155
Tabela 18 – Pesos por dimensão e tipo de instrumento.	160
Tabela 19 – Critérios de Seleção	167
Tabela 20 – Análise comparativa dos provedores de IA generativa	168
Tabela 21 – Configuração dos agentes de IA do <i>framework</i>	171
Tabela 22 – Caracterização dos cenários controlados.	184
Tabela 23 – Desempenho geral do <i>framework</i> nos cenários controlados.....	185
Tabela 24 – F1 Geral e acerto de dimensão prioritária por cenário.....	185
Tabela 25 – Métricas de classificação por categoria de prioridade	187
Tabela 26 – Matriz de confusão do Cenário A (execução representativa, N = 43 indicadores).....	188
Tabela 27 – Síntese do atendimento aos limiares MCE	193
Tabela 28 – Variabilidade das métricas de classificação entre execuções.....	194

Tabela 29 – Cursos selecionados para a análise retrospectiva.....	200
Tabela 30 – Distribuição de indicadores por categoria no <i>ground truth</i>	201
Tabela 31 – Desempenho geral do <i>framework</i> na análise retrospectiva.....	201
Tabela 32 – Matriz de confusão: Curso de Tecnologia em Alimentos (execução com F1 Geral = 1,0)	202
Tabela 33 – Matriz de confusão: Educação Física (controle positivo, F1 Geral = 0,6111).	203
Tabela 34 – Métricas de classificação por curso (consolidado de todas as sessões).....	204
Tabela 35 – Acerto de dimensão prioritária por curso.....	206
Tabela 36 – <i>Precision</i> , <i>Recall</i> e F1 por categoria de prioridade e curso.....	208
Tabela 37 – Métricas MCE consolidadas (análise retrospectiva).....	210
Tabela 38 – Métricas MCE por curso (sessão de referência)	210
Tabela 39 – Resultados da classificação automática de relevância	212
Tabela 40 – Distribuição da classificação por nível de relevância (moda das rodadas).....	213
Tabela 41 – Variabilidade do F1 Geral por curso entre sessões	215
Tabela 42 – Métricas de consistência consolidadas.....	220
Tabela 43 – Métricas de consistência por curso	220
Tabela 44 – Estabilidade da classificação de prioridade por indicador (Educação Física, 6 execuções).....	222
Tabela 45 – Estabilidade por componente estrutural e curso	223
Tabela 46 – Médias consolidadas dos componentes de estabilidade.....	224
Tabela 47 – Variabilidade das métricas quantitativas por curso	226
Tabela 48 – Índice de Consistência Semântica (ICS) por curso.....	227
Tabela 49 – Síntese dos <i>outputs</i> gerados na prova de conceito	241
Tabela 50 – Registros de transparência de IA da prova de conceito	243
Tabela 51 – Síntese consolidada das métricas de validação	245

LISTA DE ABREVIATURAS E SIGLAS

ABMES	Associação Brasileira de Mantenedoras de Ensino Superior
AD	Acurácia do Diagnóstico
AGINOV	Agência de Inovação
API	<i>Application Programming Interface</i>
AUT	Autorização
AVA	Ambiente Virtual de Aprendizagem
AWS	<i>Amazon Web Services</i> (Serviços Web da Amazon)
BDI	<i>Belief-Desire-Intention</i> (Crença-Desejo-Intenção)
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CC	Conceito de Curso
CCE	Coefficiente de Consistência Estrutural
CEA	Comissão Especial de Avaliação
CEE/MT	Conselho Estadual de Educação de Mato Grosso
CEP	Comitê de Ética em Pesquisa
CEUA	Comitê de Ética na Utilização de Animais
CINE	Classificação Internacional Normalizada da Educação
CNPJ	Cadastro Nacional da Pessoa Jurídica
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
CONAES	Comissão Nacional de Avaliação da Educação Superior
CPA	Comissão Própria de Avaliação
CPC	Conceito Preliminar de Curso
CST	Curso Superior de Tecnologia
CSV	<i>Comma-Separated Values</i>
DCN	Diretrizes Curriculares Nacionais
DGRES	Diretoria de Regulação da Educação Superior
DIP	<i>Dependency Inversion Principle</i> (Princípio da Inversão de Dependência)
DOCX	<i>Microsoft Word Open XML Document</i> (formato de documento)
DOU	Diário Oficial da União

DP	Desvio Padrão
DSL	<i>Domain-Specific Language</i> (Linguagem de Domínio Específico)
DSR	<i>Design Science Research</i>
DSRM	<i>Design Science Research Methodology</i>
DTO	<i>Data Transfer Object</i>
EAD	Educação a Distância
EBIA	Estratégia Brasileira de Inteligência Artificial
e-MEC	Sistema e-MEC
ENADE	Exame Nacional de Desempenho dos Estudantes
ENC	Exame Nacional de Cursos
ETL	<i>Extract, Transform, Load</i>
FIES	Fundo de Financiamento Estudantil
GERES	Grupo Executivo para a Reformulação do Ensino Superior
GPT	<i>Generative Pre-trained Transformer</i>
HTTP	<i>Hypertext Transfer Protocol</i>
IA	Inteligência Artificial
IACG	Instrumentos de Avaliação de Cursos de Graduação
IAG	Inteligência Artificial Generativa
ICC-N2	Índice de Compartilhamento <i>Cross</i> -Curso (Nível 2)
ICS	Índice de Consistência Semântica
IDD	Indicador de Diferença entre os Desempenhos Observado e Esperado
IES	Instituição de Educação Superior
IGC	Índice Geral de Cursos
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
INPI	Instituto Nacional da Propriedade Industrial
JSON	JavaScript Object Notation
JSONB	JavaScript Object Notation Binary
KDD	<i>Knowledge Discovery in Databases</i>
KDT	<i>Knowledge Discovery from Text</i>
LDB	Lei de Diretrizes e Bases da Educação Nacional

LGPD	Lei Geral de Proteção de Dados Pessoais
LLM	<i>Large Language Model</i> (Modelo de Linguagem de Grande Escala)
MAS	<i>Multi-Agent System</i> (Sistema Multiagente)
MCA	Matriz de Condições de Aplicabilidade
MCE	Matriz de Classificação do Escopo de Compartilhamento de Evidências
MCP	<i>Model Context Protocol</i>
MCTI	Ministério da Ciência, Tecnologia e Inovação
MEC	Ministério da Educação
MPC	Matriz de Progressão Conceitual
MVC	<i>Model-View-Controller</i>
NA	Nota de Oportunidades de Ampliação da Formação
NC	Nota dos Concluintes no ENADE
ND	Nota de Proporção de Doutores
NDE	Núcleo Docente Estruturante
NF	Nota de Infraestrutura e Instalações Físicas
NIDD	Nota do Indicador de Diferença de Desempenho
NM	Nota de Proporção de Mestres
NO	Nota de Organização Didático-Pedagógica
NR	Nota de Regime de Trabalho
NSA	Não Se Aplica
ORM	<i>Object-Relational Mapping</i>
PAIUB	Programa de Avaliação Institucional das Universidades Brasileiras
PARU	Programa de Avaliação da Reforma Universitária
PDF	<i>Portable Document Format</i>
PDI	Plano de Desenvolvimento Institucional
PEAS	<i>Performance, Environment, Actuators, Sensors</i>
PHP	<i>Hypertext Preprocessor</i>
POLEDUC	Programa de Pós-Graduação em Políticas Públicas e Gestão do Ensino Superior
PPC	Projeto Pedagógico de Curso
ProUni	Programa Universidade para Todos

RAG	<i>Retrieval-Augmented Generation</i> (Geração Aumentada por Recuperação)
REC	Reconhecimento
REST	<i>Representational State Transfer</i>
SaaS	<i>Software as a Service</i>
SDK	<i>Software Development Kit</i>
SECITECI	Secretaria de Estado de Ciência, Tecnologia e Inovação
SES/MT	Sistema Estadual de Ensino de Mato Grosso
SESU	Secretaria de Educação Superior
SHA-512	<i>Secure Hash Algorithm 512 bits</i>
SINAES	Sistema Nacional de Avaliação da Educação Superior
SOLID	<i>Single Responsibility, Open-Closed, Liskov Substitution, Interface Segregation, Dependency Inversion</i>
SSE	<i>Server-Sent Events</i>
SUS	Sistema Único de Saúde
TAO	Taxa de Adequação de Orientação
TCC	Trabalho de Conclusão de Curso
TDE	Taxa de Diferenciação de Escopo
TDR	Taxa de Descoberta RAG
TFNM	Taxa de Falsos Negativos MCE
TRA	Taxa de Reconhecimento de Artefatos
UFC	Universidade Federal do Ceará
UNEMAT	Universidade do Estado de Mato Grosso
USD	<i>United States Dollar</i> (Dólar dos Estados Unidos)
UTF-8	<i>Unicode Transformation Format 8-bit</i>
UUID	<i>Universally Unique Identifier</i>

LISTA DE SÍMBOLOS

=	Igual a / atribuição de valor
≠	Diferente de (operador de comparação na DSL da MCA)
>	Maior que
<	Menor que
≥	Maior ou igual a
≤	Menor ou igual a
×	Multiplicação / cruzamento
/	Divisão (utilizado na fórmula do <i>FI-Score</i>)
+	Adição / operador de conjunção
→	Implica / fluxo de processamento
↔	Relação bidirecional / mapeamento mútuo
\$	Dólar (unidade monetária – referência a custos de API)
%	Porcentagem
±	Mais ou menos (indicador de margem de erro / intervalo de dispersão)
μ	Mu – média aritmética (notação estatística utilizada nos gráficos de resultados)
σ	Sigma – desvio padrão (notação estatística utilizada nos gráficos de resultados)
±1σ	Intervalo de um desvio padrão em torno da média (barras de erro nos gráficos)
κ	<i>Kappa de Fleiss</i> (coeficiente de concordância entre avaliadores)
F1	<i>FI-Score</i> (média harmônica entre <i>Precision</i> e <i>Recall</i>)
DP	Desvio Padrão (medida de dispersão estatística)
CV	Coefficiente de Variação (razão entre desvio padrão e média)
ms	Milissegundos (unidade de tempo de processamento)
C1	Conceito 1 do Instrumento de Avaliação de Cursos de Graduação do INEP
C2	Conceito 2 do Instrumento de Avaliação de Cursos de Graduação do INEP
C3	Conceito 3 do Instrumento de Avaliação de Cursos de Graduação do INEP
C4	Conceito 4 do Instrumento de Avaliação de Cursos de Graduação do INEP
C5	Conceito 5 do Instrumento de Avaliação de Cursos de Graduação do INEP
D1	Dimensão 1 – Organização Didático-Pedagógica
D2	Dimensão 2 – Corpo Docente e Tutorial
D3	Dimensão 3 – Infraestrutura

N1	Níveis de acesso a artefatos de evidência – direto
N2	Níveis de acesso a artefatos de evidência – cross-curso
N3	Níveis de acesso a artefatos de evidência – descoberta via RAG
M1 a M7	Codificação das métricas MCE (M1=TRA, M2=TAO, M3=TDE, M4=ICC-N2, M5=TFNM, M6=TDR, M7=Qualidade N3)
V01 a V20	Variáveis determinantes da Matriz de Condições de Aplicabilidade (MCA)
N+1	Conceito imediatamente superior ao atual na escala do instrumento INEP

SUMÁRIO

1	INTRODUÇÃO	21
1.1	Contextualização e Problemática	21
1.2	Delimitação do Problema de Pesquisa	26
1.3	Objetivos	27
1.3.1	Objetivo Geral.....	27
1.3.2	Objetivos Específicos.....	27
1.4	Justificativa e Relevância.....	28
1.4.1	Relevância Científica	28
1.4.2	Relevância para Gestores e Políticas Públicas	28
1.4.3	Relevância Institucional	29
1.5	Estrutura da Dissertação.....	29
2	FUNDAMENTAÇÃO TEÓRICA	32
2.1	Políticas Públicas de Avaliação na Educação Superior.....	32
2.1.1	Histórico da Avaliação no Ensino Superior Brasileiro.....	32
2.1.2	Marco Legal e Institucional	33
2.1.3	O Exame Nacional de Cursos e a Regulação Estatal	34
2.1.4	O SINAES: Síntese Evolutiva e Inovação Sistêmica.....	35
2.1.5	Dimensões e Instrumentos Avaliativos	36
2.1.6	Autoavaliação Institucional: Processos e Desafios	38
2.1.7	O Processo de Autoavaliação: O Papel e os Desafios da CPA	39
2.1.8	Os Documentos Fundamentais: A Centralidade do PDI e do PPC	42
2.1.9	O Motor Acadêmico: O Papel Crítico do NDE.....	43
2.1.10	Avaliação Externa e Integração Sistêmica	44
2.1.11	Os sistemas estaduais de ensino superior e sua articulação com o SINAES	45
2.1.12	Conceito de Curso e Conceito Preliminar de Curso: estrutura, ponderação e implicações regulatórias.....	46
2.1.13	Consequências regulatórias e estratégicas dos indicadores de qualidade: a lógica do incentivo institucional	49
2.1.14	Racionalidade institucional frente ao risco de rebaixamento: entre conformidade e qualidade	50
2.1.15	Desafios Contemporâneos e Perspectivas de Aprimoramento.....	51

2.2	Decisão Estratégica em Organizações Educacionais: Conceitos e Teorias.....	51
2.2.1	Teorias da Decisão Estratégica e seus Condicionantes nas Instituições de Educação Superior	54
2.2.2	A Teoria da Racionalidade Limitada e suas Implicações para a Gestão Educacional.....	54
2.2.3	A Perspectiva Comportamental: Heurísticas e Vieses no Processo Decisório....	55
2.3	Descoberta de Conhecimento.....	56
2.3.1	Knowledge Discovery in Databases (KDD).....	56
2.3.2	Knowledge Discovery from Text (KDT)	57
2.4	KDD e KDT como Referencial Contextualizador	58
2.5	Análise de Conteúdo: Fundamentos Teórico- Metodológicos.....	58
2.5.1	Conceituação e Evolução Histórica	58
2.5.2	Modalidades de Análise de Conteúdo.....	59
2.5.3	Fases da Análise de Conteúdo.....	59
2.5.4	Critérios de Qualidade	60
2.5.5	Aplicação ao Contexto do SINAES	61
2.6	Inteligência Artificial na Gestão Educacional.....	62
2.6.1	O Papel da Inteligência Artificial na Extração de Conhecimento.....	62
2.6.2	Inteligência Artificial Fraca e Forte	63
2.6.3	Redes Neurais Artificiais	64
2.6.4	IA Aplicada à Avaliação Institucional	64
2.6.5	Limitações das Abordagens Tradicionais para Processamento de Documentos Normativos.....	66
2.7	Inteligência Artificial Generativa: Aspectos Avançados para Aplicação em Sistemas de Apoio à Decisão	66
2.7.1	Agentes de Inteligência Artificial: Arquiteturas e Aplicações Educacionais.....	70
2.7.2	Agentes Inteligentes e Sistemas Multiagentes	74
2.7.2.1	Componentes funcionais de um agente inteligente.....	76
2.7.2.2	Padrões de orquestração e compartilhamento de contexto	78
2.7.3	<i>Prompt Engineering</i> : Fundamentos e Técnicas	79
2.7.4	<i>Retrieval-Augmented Generation (RAG)</i>	80
2.7.5	Limitações e Alucinações de LLMs.....	83
2.8	<i>Design Science Research</i> : Fundamentos e Justificativa de Escolha	84

2.8.1	Tipologia de Artefatos e Contribuições.....	85
2.8.2	Diretrizes para Rigor e Relevância	86
2.8.3	Avaliação de Artefatos em <i>Design Science Research</i>	87
3	METODOLOGIA	89
3.1	Classificação e Natureza da Pesquisa	89
3.2	Abordagem Metodológica: <i>Design Science Research</i>	89
3.2.1	Modelo Metodológico: DSRM	91
3.2.1.1	Identificação do Problema e Motivação.....	92
3.2.1.2	Definição de Objetivos da Solução	93
3.2.1.3	<i>Design</i> e Desenvolvimento	93
3.2.1.4	Demonstração.....	95
3.2.1.5	Avaliação.....	95
3.2.1.6	Comunicação.....	96
3.2.2	Procedimentos de Validação	97
3.2.2.1	Validação por Cenários Controlados.....	98
3.2.2.2	Validação por Análise Retrospectiva com Dados Oficiais.....	104
3.2.2.3	Validação por Consistência Interna.....	108
3.2.2.4	Prova de Conceito (Demonstração Integral).....	110
3.3	Ambiente de Estudo	111
3.4	Fontes de Dados	112
3.4.1	Dados Primários: Instrumentos do SINAES.....	113
3.4.2	Dados Secundários: Indicadores e Microdados Oficiais.....	113
3.4.3	Dados Institucionais.....	113
3.5	Aspectos Éticos e Conformidade Legal	113
3.5.1	Conformidade com LGPD	113
3.5.2	Uso Responsável de IA	114
3.5.3	Limitações Metodológicas	114
4	CONSTRUÇÃO DOS ARTEFATOS ANALÍTICOS DO FRAMEWORK	115
4.1	Fundamentação: Análise de Conteúdo Documental	115
4.1.1	Matriz de Progressão Conceitual (MPC)	115
4.1.2	Processo de Decomposição	116
4.1.3	Exemplo de Aplicação	116
4.1.4	Fundamentos e Processo de Construção	117

4.1.5	Corpus Analítico e Dimensões do Instrumento.....	117
4.1.6	Categorias Analíticas da Decomposição	118
4.1.7	Diferenças Semânticas entre os Instrumentos.....	119
4.1.8	Estrutura da Matriz Desenvolvida.....	120
4.1.9	Padrões Transversais Identificados	122
4.1.10	Integração da MPC com o <i>Framework</i>	123
4.1.11	Implicações para a Gestão Institucional.....	124
4.2	Matriz de Condições de Aplicabilidade (MCA)	125
4.2.1	Processo de Decomposição	126
4.2.2	Exemplo de Aplicação	127
4.2.3	Fundamentação Metodológica da MCA	127
4.2.4	Variáveis Determinantes: Estrutura e Domínios	128
4.2.5	Tipologia e Padrões Lógicos das Regras de Aplicabilidade	131
4.2.6	A MCA como Camada de Pré-Filtragem Contextual dos Agentes de IA	134
4.2.7	Implementação Técnica: Modelo de Dados e Motor de Regras	135
4.2.8	Contribuição da MCA para a Qualidade do Diagnóstico Institucional.....	135
4.3	Matriz de Compartilhamento de Evidências (MCE).....	136
4.3.1	Processo de Classificação	137
4.3.2	Exemplo de Aplicação	138
4.3.3	Taxonomia de Escopos.....	138
4.3.4	Resultados da Classificação	139
4.3.5	Análise por Dimensão Avaliativa.....	140
4.3.6	Exemplos de Aplicação Prática.....	141
4.3.7	Integração com a Arquitetura do <i>Framework</i>	142
4.3.8	Contribuição para a Gestão Institucional	143
5	DESENVOLVIMENTO DO <i>FRAMEWORK</i>	144
5.1	Arquitetura e Engenharia de <i>Software</i> da Aplicação.....	144
5.1.1	Pilha Tecnológica	144
5.1.2	Arquitetura em Camadas.....	146
5.1.3	Arquitetura Modular	147
5.1.4	Arquitetura <i>Multi-Tenant</i>	149
5.1.5	Modelo de Dados do Domínio SINAES	150
5.1.5.1	Instrumentos de Avaliação de Cursos de Graduação (IACG).....	150

5.1.5.2	Codificação da Matriz de Condições de Aplicabilidade (MCA)	151
5.1.5.3	Codificação da Matriz de Progressão Conceitual (MPC)	152
5.1.5.4	Codificação da Matriz de Compartilhamento de Evidências (MCE)	152
5.1.5.5	Avaliações e Planos de Ação (<i>Tenant-Scoped</i>)	154
5.1.6	Gestão de Dados Acadêmicos	155
5.1.7	ETL dos Dados do SINAES.....	156
5.1.8	Serviços Computacionais do <i>Framework</i>	157
5.1.8.1	Motor de Regras da MCA (McaRuleEngine)	158
5.1.8.2	Calculadora do Conceito de Curso (CcCalculatorService).....	159
5.1.8.3	Serviço de Diagnóstico MPC (MpcDiagnosticoTenantService).....	160
5.1.9	Fluxo de Dados do <i>Framework</i>	163
5.1.10	Controle de Acesso e Segurança	164
5.1.11	Síntese Arquitetural.....	165
5.2	Integração com Inteligência Artificial Generativa	166
5.2.1	Escolha do Provedor de IA	167
5.2.2	Módulo de IA: Estrutura e Organização	169
5.2.3	Arquitetura dos Agentes de IA.....	169
5.2.4	Implementação dos Agentes Especializados.....	171
5.2.5	Ferramentas (<i>Tools</i>) dos Agentes e Servidor MCP	172
5.2.6	Estratégia de Engenharia de <i>Prompt</i>	174
5.2.7	<i>Retrieval-Augmented Generation</i> (RAG) e Base de Conhecimento.....	176
5.2.8	<i>Pipeline</i> de Processamento Documental (RAG).....	177
5.2.9	Fluxos de Integração com a Aplicação	178
5.2.10	Base de Conhecimento Global e Aprendizado Organizacional	180
5.2.11	Transparência, Auditoria e Uso Responsável de IA.....	181
5.2.12	Síntese da Integração com IA	182
6	RESULTADOS E DISCUSSÃO.....	183
6.1	Validação por Cenários Controlados.....	183
6.1.1	Configuração dos Cenários	183
6.1.2	Resultados das métricas de classificação de prioridade.....	184
6.1.2.1	Desempenho geral.....	184
6.1.2.2	Desempenho por cenário.....	185
6.1.2.3	Métricas por categoria de prioridade	187

6.1.3	Resultados das métricas de reconhecimento de evidências (MCE).....	191
6.1.4	Estabilidade dos resultados entre execuções.....	194
6.1.5	Discussão dos resultados.....	196
6.2	Validação por Análise Retrospectiva com Dados Oficiais.....	199
6.2.1	Caracterização dos cursos selecionados.....	200
6.2.2	Resultados das métricas automáticas de classificação	201
6.2.2.1	Desempenho geral.....	201
6.2.2.2	Desempenho por curso.....	204
6.2.2.3	Análise da dimensão prioritária	206
6.2.2.4	Métricas detalhadas por categoria (sessão de referência).....	207
6.2.3	Resultados das métricas de reconhecimento de evidências (MCE).....	210
6.2.4	Resultados da classificação de relevância.....	212
6.2.5	Estabilidade dos resultados entre execuções.....	214
6.2.6	Discussão dos resultados.....	215
6.3	Validação por Consistência Interna.....	218
6.3.1	Métricas de consistência	219
6.3.2	Resultados do Coeficiente de Consistência Estrutural (CCE)	219
6.3.2.1	Resultado consolidado	219
6.3.2.2	Resultados por curso	220
6.3.3	Análise dos componentes de estabilidade.....	223
6.3.4	Variabilidade das métricas quantitativas	226
6.3.5	Consistência semântica das recomendações	226
6.3.6	Discussão dos resultados.....	228
6.4	Prova de Conceito (Demonstração Integral).....	230
6.4.1	Curso Selecionado e Dados de Entrada	231
6.4.2	Processo de Análise.....	232
6.4.2.1	Cadastro institucional do curso e inserção das DCN	232
6.4.2.2	Verificação dos dados cadastrais do curso	233
6.4.2.3	<i>Upload</i> do PPC e constituição da base de conhecimento RAG.....	234
6.4.2.4	Configuração do perfil de aplicabilidade com a Matriz MCA.....	235
6.4.2.5	Lançamento dos conceitos no Instrumento de Avaliação	236
6.4.2.6	Elaboração dos planejamentos com a Matriz MPC	237
6.4.2.7	Execução sequencial dos agentes de IA e geração do relatório	238

6.4.2.8	Preenchimento do plano de ação com evidências por indicador	239
6.4.2.9	Validação dos artefatos pela DGRES e operação da Matriz MCE	240
6.4.3	<i>Outputs</i> Gerados.....	241
6.4.4	Registros de Transparência de IA	242
6.5	Discussão Geral dos Resultados	244
6.5.1	Síntese das Métricas.....	245
6.5.2	Contribuições do <i>Framework</i>	246
6.5.3	Limitações Identificadas	247
6.5.4	Comparação com a Literatura	249
6.6	Implicações para a Gestão de IES	250
7	CONCLUSÕES	253
7.1	O Produto Tecnológico e a Contribuição para a Inovação.....	254
7.2	Limitações do Estudo.....	255
7.3	Recomendações para Trabalhos Futuros e Alinhamento Estratégico	257
	REFERÊNCIAS.....	260

1 INTRODUÇÃO

1.1 Contextualização e Problemática

A educação superior brasileira atravessa um período de profundas transformações, impulsionada por um ambiente globalizado caracterizado por rápidas mudanças tecnológicas, sociais e econômicas. Neste cenário volátil, incerto, complexo e ambíguo, as Instituições de Ensino Superior (IES) enfrentam desafios sem precedentes para manter-se relevantes e eficazes na formação de profissionais qualificados e na produção de conhecimento científico que impulse o desenvolvimento socioeconômico nacional.

As IES, enquanto centros estratégicos de conhecimento e inovação, desempenham um papel crucial na construção de uma sociedade mais competitiva e equitativa. Contudo, para que essa missão seja cumprida com excelência, torna-se imperativo que as políticas públicas, estratégias institucionais e práticas de gestão estejam alinhadas às demandas contemporâneas e futuras, sustentadas por processos decisórios fundamentados em evidências científicas e dados confiáveis. Conforme observam Carvalho *et al.* (2018, p. 240), "as mudanças e inovações fazem parte da sociedade e a universidade atua como indutora dessas inovações, por meio do ensino, da pesquisa, da extensão".

A Lei nº 10.861, de 14 de abril de 2004, que institui o Sistema Nacional de Avaliação da Educação Superior (SINAES), apresenta-se como principal instrumento regulador da qualidade educacional e excelência acadêmica no país. Esse sistema estabelece mecanismos de avaliação das instituições, dos cursos e do desempenho estudantil, promovendo a cultura da avaliação e o compromisso com a melhoria contínua da educação superior brasileira (SINAES, 2009). O SINAES é compreendido como política de Estado que visa subsidiar a qualidade educativa "ao regulamentar o sistema em todos os níveis de ensino", transformando a organização da educação por meio de medidas de controle e regulação (GRIBOSKI; FUNGHETTO, 2013).

O SINAES fundamenta-se em três pilares avaliativos complementares: avaliação institucional, avaliação de cursos de graduação e avaliação do desempenho dos estudantes por meio do Exame Nacional de Desempenho dos Estudantes (ENADE). Essa estrutura tridimensional visa proporcionar uma visão holística da qualidade educacional, fornecendo subsídios robustos para o planejamento estratégico institucional e a formulação de políticas

públicas educacionais mais efetivas (SINAES, 2009).

Cabe registrar, desde este ponto, que o SINAES convive no ordenamento brasileiro com os sistemas estaduais de ensino superior, responsáveis pela regulação das instituições estaduais de ensino superior. É o caso da Universidade do Estado de Mato Grosso (UNEMAT), instituição pública estadual vinculada ao Sistema Estadual de Ensino de Mato Grosso (SES/MT), cuja regulação compete ao Conselho Estadual de Educação (CEE/MT) e cuja avaliação externa é conduzida pela Secretaria de Estado de Ciência, Tecnologia e Inovação (SECITECI), em regime de alinhamento normativo ao SINAES, conforme detalhado na seção 2.1.11.

No âmbito específico da avaliação de cursos, dois indicadores assumem centralidade: o Conceito de Curso (CC), atribuído por comissões avaliadoras em visitas *in loco* com base no Instrumento de Avaliação de Cursos de Graduação do INEP, ou, no caso das instituições vinculadas a sistemas estaduais de ensino, por comissões designadas pelos respectivos órgãos reguladores, e o Conceito Preliminar de Curso (CPC), calculado a partir de componentes que incluem desempenho dos estudantes no ENADE, Indicador de Diferença entre os Desempenhos Observado e Esperado (IDD), características do corpo docente e percepção discente sobre o processo formativo. Esses indicadores orientam decisões regulatórias significativas, incluindo reconhecimento e renovação de reconhecimento de cursos.

Não obstante os avanços significativos proporcionados pelo SINAES, observa-se que os processos de avaliação institucional ainda enfrentam limitações importantes. A atual forma de divulgação dos resultados, muitas vezes desarticulada dos índices de qualidade (como o IGC) e o conceito institucional, acaba por reduzir a relevância do Plano de Desenvolvimento Institucional (PDI) tanto no processo avaliativo quanto na gestão prática das instituições (GRIBOSKI et al., 2018). Além disso, o desafio de equilibrar o crescimento quantitativo com a qualidade exige a definição de indicadores e de processos avaliativos consistentes que superem a mera conformidade burocrática (CARVALHO *et al.*, 2018). Esta realidade compromete a tempestividade e a precisão das análises, dificultando a identificação de padrões ocultos, tendências emergentes e correlações complexas que poderiam orientar decisões estratégicas mais assertivas.

A problemática central reside na gestão ineficiente da autoavaliação, que gera uma desorientação informacional, impedindo que gestores tomem decisões estratégicas fundamentadas em dados confiáveis. Esta incerteza epistemológica transforma a avaliação externa de uma oportunidade de crescimento em uma ameaça percebida, alimentando uma

cultura de conformidade em detrimento da cultura de qualidade. A autoavaliação institucional deve ser entendida como o ponto de partida para o planejamento estratégico e para a gestão universitária, pois é através dela que se obtém o diagnóstico necessário para orientar as decisões dos gestores (ANDRIOLA, 2009).

A análise dos instrumentos de avaliação do INEP revela uma estrutura textual complexa que, embora sistematizada em indicadores e conceitos, demanda interpretação especializada para tradução em ações concretas de melhoria. Cada indicador do Instrumento de Avaliação de Cursos de Graduação apresenta cinco critérios correspondentes aos conceitos 1 a 5, organizados em progressão qualitativa que incorpora elementos constitutivos, atributos de diferenciação e requisitos de evidenciação. A identificação precisa do *gap* entre o conceito atual de um curso e o conceito desejado, bem como a especificação das ações necessárias para progressão, constitui tarefa analítica que frequentemente excede a capacidade operacional das Comissões Próprias de Avaliação (CPA) e dos coordenadores de curso.

Simultaneamente, o paradigma da transformação digital tem revolucionado diversos setores da economia e da sociedade, oferecendo oportunidades sem precedentes para aprimorar processos organizacionais por meio da aplicação de tecnologias emergentes. A Inteligência Artificial (IA), particularmente em sua vertente generativa baseada em *Large Language Models* (LLMs), destaca-se como ferramenta transformadora capaz de processar grandes volumes de dados textuais, identificar padrões complexos e gerar sínteses contextualizadas que subsidiem a tomada de decisão estratégica.

Na educação superior, essas tecnologias apresentam potencial significativo para a gestão institucional. A IA pode subsidiar a análise de indicadores de qualidade, automatizar a interpretação de instrumentos normativos, gerar recomendações contextualizadas para processos de avaliação e apoiar coordenadores na formulação de planos de melhoria. A convergência entre as necessidades do SINAES e as possibilidades oferecidas pela IA configura oportunidade estratégica para revolucionar os processos de avaliação institucional. A aplicação sistemática dessas tecnologias aos dados avaliativos pode potencializar significativamente a descoberta de conhecimento, transformando informações brutas em conhecimento acionável que oriente gestores educacionais na tomada de decisões mais informadas e precisas.

Historicamente, os processos decisórios nas organizações dependiam quase exclusivamente da análise humana, inerentemente limitada por fatores como tempo, capacidade de processamento de grandes volumes de dados e vieses cognitivos. Segundo Leitner-Hanetseder e Lehner (2022), a introdução da Inteligência Artificial (IA) representou um marco

evolutivo nesse cenário ao automatizar tarefas analíticas complexas, viabilizando decisões mais ágeis, precisas e fundamentadas.

A principal virtude da IA no suporte à decisão estratégica reside em sua capacidade de operar com algoritmos que simulam cenários e avaliam múltiplas alternativas de forma simultânea. Radonjić, Duarte e Pereira (2022) observam que, embora os gestores sempre tenham tido abundância de dados à disposição, estes eram frequentemente utilizados apenas como mecanismos de controle, e não como fontes de conhecimento. Por meio da IA, essa massa de dados é transformada em inteligência acionável, livre de limitações humanas como a fadiga ou interferências emocionais, o que é fundamental em contextos de avaliação de alta complexidade.

Nesse panorama tecnológico, a Inteligência Artificial Generativa (IA Generativa) surge como subárea de destaque, voltada à criação de novos conteúdos a partir do aprendizado de padrões complexos. Dentre suas arquiteturas, os modelos baseados em *Transformers*, como a família GPT, destacam-se pela capacidade de gerar textos e análises técnicas contextualizadas (ALNAHHAS; YOUSEF; GAI, 2024). Esses modelos operam por meio de representações vetoriais densas que capturam relações semânticas entre elementos textuais, viabilizando tarefas de interpretação, síntese e geração de conteúdo a partir de documentos normativos e institucionais.

No âmbito da Universidade do Estado de Mato Grosso (UNEMAT), instituição pública estadual vinculada ao Sistema Estadual de Ensino de Mato Grosso (SES/MT) e regulada pelo Conselho Estadual de Educação (CEE/MT) em regime de alinhamento normativo ao SINAES, observa-se um cenário de transição institucional. Embora o Plano de Desenvolvimento Institucional (PDI) 2022-2028 estabeleça a modernização tecnológica como pilar estratégico, a realidade operacional ainda é marcada por processos manuais e não automatizados, tais como o registro de evidências em documentos não padronizados, a elaboração de relatórios avaliativos sem sistematização por indicadores e a ausência de ferramentas que relacionem os dados avaliativos a ações de melhoria. Esse hiato entre as diretrizes de inovação e a persistência de rotinas manuais constitui o cerne desta investigação, que busca estratégias para mitigar tal discrepância.

Considerando as potencialidades da IA e a necessidade de mitigar seus riscos, esta pesquisa propõe a aplicação dessas tecnologias no contexto da avaliação da educação superior brasileira. Ao alinhar a capacidade analítica da IA com as diretrizes do SINAES, busca-se transpor a barreira da análise meramente descritiva para uma descoberta de conhecimento que

subsidiar planos de ação acadêmicas e administrativas mais eficazes.

Neste contexto, a presente pesquisa propõe o desenvolvimento e a validação técnica de um *framework*¹ de apoio à avaliação institucional que integra: (a) três artefatos analíticos, a Matriz de Progressão Conceitual (MPC), a Matriz de Condições de Aplicabilidade (MCA) e a Matriz de Compartilhamento de Evidências (MCE), construídos mediante Análise de Conteúdo Categorical Temática aplicada ao Instrumento de Avaliação de Cursos de Graduação do INEP; (b) agentes de Inteligência Artificial Generativa para processamento contextual dos indicadores do SINAES e geração de recomendações estratégicas personalizadas, tendo as matrizes analíticas como base de conhecimento normativo; e (c) aplicação *web* que operacionaliza o *framework* com módulos de visualização, monitoramento de indicadores e interface conversacional para apoio à tomada de decisão. O projeto parte da realidade da UNEMAT, instituição pública estadual *multicampi* com base estruturada de dados históricos de avaliações. A avaliação de impacto do *framework* sobre as práticas de gestão e sobre a cultura de autoavaliação institucional constitui agenda de pesquisa futura, não contemplada nos objetivos deste trabalho.

Diferentemente de sistemas puramente estatísticos, este *framework* propõe uma abordagem integrada, dinâmica e interativa em que os dados educacionais são transformados em conhecimento acionável mediante a combinação de análise semântica dos instrumentos normativos e processamento contextual por agentes inteligentes. A proposição de três artefatos analíticos complementares, a MPC, que decompõe os critérios de avaliação em elementos constitutivos, atributos de progressão e evidências requeridas; a MCA, que codifica as condições de aplicabilidade dos indicadores a cada perfil de curso; e a MCE, que classifica o escopo de compartilhamento de evidências entre cursos, constitui contribuição metodológica central que viabiliza a integração entre o conhecimento normativo do SINAES e as capacidades de processamento dos agentes de IA generativa.

O *framework* contribui para a transformação de dados avaliativos em conhecimento acionável, subsidiando coordenadores de curso e membros da CPA na formulação de planos de melhoria fundamentados em evidências e alinhados aos requisitos do SINAES. Esta pesquisa insere-se no campo das políticas públicas e da gestão da educação superior, com ênfase no

¹ O termo *framework* (em tradução livre, "quadro de referência" ou "arcabouço") refere-se a um conjunto estruturado de conceitos, diretrizes e práticas que serve de base para a solução de problemas específicos ou para o desenvolvimento de sistemas, permitindo uma abordagem padronizada e organizada para lidar com processos complexos. No contexto da gestão e tecnologia, um *framework* é um modelo lógico que organiza informações, métricas e ferramentas de análise. Ele funciona como um "esqueleto" que suporta a integração de dados complexos, facilitando a interpretação e a tomada de decisão estratégica a partir de uma estrutura replicável e sistemática.

desenvolvimento e na validação técnica de artefato tecnológico destinado a apoiar processos avaliativos das IES brasileiras.

1.2 Delimitação do Problema de Pesquisa

Apesar do potencial transformador identificado, a integração entre IA e processos de avaliação institucional no contexto do SINAES ainda constitui área emergente na literatura científica e na prática organizacional das IES brasileiras. Observam-se lacunas significativas tanto no desenvolvimento de *frameworks* conceituais robustos quanto na implementação prática de soluções tecnológicas que efetivamente otimizem os processos decisórios na gestão educacional.

A problemática central desta pesquisa emerge da constatação de que as IES brasileiras dispõem de volumes consideráveis de dados avaliativos gerados pelos instrumentos do SINAES, mas carecem de ferramentas e metodologias adequadas para extrair o máximo valor dessas informações. A análise manual e fragmentada desses dados limita drasticamente o potencial de descoberta de conhecimento, perpetuando processos decisórios baseados em intuição ou evidências parciais.

Especificamente, identificam-se três lacunas inter-relacionadas:

Lacuna Analítica: Os critérios de atribuição de conceitos presentes no Instrumento de Avaliação de Cursos do INEP, embora textuais e estruturados, não estão decompostos em formato que permita identificação automática de lacunas e geração sistemática de recomendações. A passagem do conceito 3 para o conceito 4 em determinado indicador, por exemplo, requer a identificação precisa dos atributos adicionais que distinguem esses níveis, tarefa que demanda análise semântica especializada.

Lacuna Tecnológica: As ferramentas disponíveis para gestão de indicadores do SINAES limitam-se predominantemente a sistemas de visualização de dados que apresentam conceitos e notas sem oferecer análise contextual ou recomendações estratégicas. Não foram identificados, na literatura consultada, *frameworks* que integrem agentes de Inteligência Artificial Generativa ao processamento contextualizado dos instrumentos de avaliação do INEP para geração de recomendações personalizadas por curso. Tal ausência delimita o espaço de contribuição desta pesquisa, sem prejuízo de desenvolvimentos paralelos não acessados pelo levantamento bibliográfico realizado.

Lacuna Metodológica: A literatura sobre avaliação institucional carece de abordagens sistemáticas que articulem técnicas de Análise de Conteúdo para tratamento de documentos normativos com as capacidades de IA para geração de conhecimento aplicado. O *Design Science Research* (DSR) oferece um paradigma adequado para desenvolvimento de artefatos que preencham esta lacuna, porém sua aplicação ao domínio específico da avaliação de cursos de graduação no Brasil permanece incipiente.

A pesquisa adota o *Design Science Research* (DSR) como abordagem metodológica. Diferentemente de pesquisas de natureza puramente descritiva ou explicativa, que buscam compreender

fenômenos existentes, o DSR orienta a criação e a avaliação de artefatos tecnológicos concebidos para resolver problemas práticos, produzindo conhecimento simultaneamente sobre o artefato desenvolvido e sobre o contexto em que ele é aplicado (HEVNER *et al.*, 2004; PEFFERS *et al.*, 2007). Essa característica torna o DSR especialmente adequado ao desenvolvimento de um *framework* computacional destinado a apoiar processos avaliativos regulamentados, como os do SINAES. A operacionalização desta abordagem é detalhada no Capítulo 3.

A UNEMAT, instituição *multicampi* que atende diversas regiões do estado, apresenta-se como contexto ideal para investigar essa problemática. A complexidade de sua estrutura organizacional, a diversidade de seus cursos de graduação e a disponibilidade de dados históricos de avaliação institucional configuram um ambiente propício para o desenvolvimento e validação de soluções inovadoras.

Diante desse cenário, formula-se o seguinte problema de pesquisa:

Como desenvolver um *framework* que integre Análise de Conteúdo e agentes de IA generativa para apoiar os processos de avaliação institucional de cursos de graduação, transformando os critérios normativos do SINAES em recomendações estratégicas acionáveis para gestores educacionais?

1.3 Objetivos

1.3.1 Objetivo Geral

Desenvolver e testar um *framework* baseado em agentes de Inteligência Artificial para apoiar os processos de avaliação de cursos de graduação e gerar um plano de ações acadêmicas e administrativas para aprimorar a qualidade educacional destes, conforme as diretrizes do SINAES.

1.3.2 Objetivos Específicos

- I. Realizar levantamento bibliográfico sobre *Design Science Research*, Análise de Conteúdo, Inteligência Artificial generativa e suas aplicações na gestão da educação superior, estabelecendo os fundamentos teórico-metodológicos da pesquisa;
- II. Construir, mediante aplicação sistemática de técnicas de Análise de Conteúdo ao Instrumento de Avaliação de Cursos de Graduação do INEP, as matrizes analíticas estruturantes do *framework*: Matriz de Progressão Conceitual (MPC), Matriz de

Condições de Aplicabilidade (MCA) e Matriz de Compartilhamento de Evidências (MCE);

- III. Integrar agentes de Inteligência Artificial Generativa ao *framework* para processamento contextual dos indicadores do SINAES e geração de recomendações estratégicas personalizadas, tendo as matrizes analíticas como base de conhecimento normativo;
- IV. Desenvolver aplicação *web* que operacionalize o *framework*, incorporando módulos de visualização, monitoramento de indicadores e interface conversacional para apoio à tomada de decisão de coordenadores e membros da CPA;
- V. Validar o *framework* por meio de estratégia multimétodo que combine verificação técnica por cenários controlados, análise retrospectiva com dados oficiais do Conceito de Curso, verificação de consistência interna e análise de aderência normativa aos instrumentos do SINAES.

1.4 Justificativa e Relevância

1.4.1 Relevância Científica

A pesquisa contribui para a área interdisciplinar emergente que articula políticas públicas de avaliação educacional, gestão universitária e inteligência artificial. Não foram localizados estudos que articulem sistematicamente a decomposição analítica dos instrumentos normativos do INEP com capacidades de processamento de linguagem de modelos generativos. A adoção do DSR como paradigma demonstra como artefatos tecnológicos podem constituir produtos científicos legítimos quando desenvolvidos com rigor metodológico (HEVNER *et al.*, 2004). A Matriz de Progressão Conceitual oferece contribuição replicável para análise de outros instrumentos regulatórios graduados, para além do caso SINAES.

1.4.2 Relevância para Gestores e Políticas Públicas

Os beneficiários diretos são coordenadores de curso, membros das CPAs e gestores institucionais das IES brasileiras que frequentemente operam sob condições de racionalidade limitada (SIMON, 1980), com acesso a grande volume de dados avaliativos, mas sem

ferramentas que transformem esses dados em planos de ação fundamentados. Em escala sistêmica, a solução pode informar o aprimoramento das políticas públicas de avaliação, contribuindo para que o SINAES opere de forma mais eficaz em sua função formativa, e não apenas regulatória e classificatória (BARREYRO, 2018).

1.4.3 Relevância Institucional

Para a UNEMAT, a solução oferece oportunidade concreta de modernizar seus processos de autoavaliação, alinhando-se ao PDI 2022-2028 (UNEMAT, 2022). A estrutura *multicampi*, com cursos de diferentes graus acadêmicos e modalidades, configura um contexto exigente para validação de um *framework* que deva operar com diferentes perfis de aplicabilidade dos indicadores, conferindo robustez adicional ao processo de teste.

1.5 Estrutura da Dissertação

Esta dissertação estrutura-se em sete capítulos, organizados de forma lógica e progressiva para apresentar o desenvolvimento completo da pesquisa, desde a fundamentação conceitual até a validação do artefato proposto.

O **Capítulo 1 – Introdução** apresenta a contextualização do problema, situando a avaliação institucional de cursos de graduação no contexto do SINAES e identificando as limitações operacionais enfrentadas por coordenadores de curso e membros das Comissões Próprias de Avaliação na transformação de dados avaliativos em planos de ação estratégicos. São formulados o problema de pesquisa, o objetivo geral e os objetivos específicos, seguidos da justificativa e relevância científica, institucional e para gestores e políticas públicas.

O **Capítulo 2 – Fundamentação Teórica** apresenta a revisão da literatura organizada em oito eixos temáticos. Inicialmente, discutem-se as políticas públicas de avaliação na educação superior brasileira, percorrendo o histórico dos programas avaliativos, o marco legal e institucional, a estrutura do SINAES, os processos de autoavaliação e o papel da CPA, os indicadores de qualidade (CC, CPC, IGC) e os desafios contemporâneos do sistema. Em seguida, abordam-se as teorias da decisão estratégica em organizações educacionais, com ênfase na racionalidade limitada e nas heurísticas e vieses do processo decisório. O capítulo prossegue com a apresentação dos conceitos de descoberta de conhecimento (KDD e KDT)

como referencial contextualizador, seguida dos fundamentos teórico-metodológicos da Análise de Conteúdo. Na sequência, examina-se a Inteligência Artificial na gestão educacional, incluindo abordagens tradicionais e suas limitações para o processamento de documentos normativos. A seção dedicada à Inteligência Artificial Generativa aprofunda os aspectos avançados para aplicação em sistemas de apoio à decisão, abrangendo agentes inteligentes e sistemas multiagentes, engenharia de *prompt*, *Retrieval-Augmented Generation* (RAG) e limitações dos modelos de linguagem. O capítulo encerra-se com os fundamentos do *Design Science Research*, incluindo tipologia de artefatos, diretrizes de rigor e relevância e estratégias de avaliação de artefatos.

O **Capítulo 3 – Metodologia** detalha os procedimentos metodológicos adotados para o desenvolvimento e a validação do *framework*. Apresenta a classificação e a natureza da pesquisa, a operacionalização do *Design Science Research* conforme o modelo DSRM de Peffers *et al.* (2007) e a estratégia de validação multimétodo empregada, que combina quatro abordagens complementares: validação por cenários controlados, análise retrospectiva com dados oficiais do Conceito de Curso, verificação de consistência interna e análise de aderência normativa aos instrumentos do SINAES. O capítulo descreve ainda o ambiente de estudo, a Universidade do Estado de Mato Grosso (UNEMAT), as fontes de dados primárias, secundárias e institucionais, e os aspectos éticos e de conformidade legal, incluindo observância à LGPD e diretrizes de uso responsável de IA.

O **Capítulo 4 – Construção dos Artefatos Analíticos do Framework** apresenta o desenvolvimento dos três artefatos conceituais que compõem a base de conhecimento normativo do *framework*. A Matriz de Progressão Conceitual (MPC) é descrita em detalhe, desde a fundamentação metodológica mediante Análise de Conteúdo Categórica Temática até a estrutura tridimensional resultante em elementos constitutivos, atributos de progressão e evidências requeridas, os padrões transversais identificados e as implicações para a gestão institucional. A Matriz de Condições de Aplicabilidade (MCA) é apresentada com suas variáveis determinantes, tipologia e padrões lógicos das regras de aplicabilidade, bem como sua função de pré-filtragem contextual para os agentes de IA. A Matriz de Compartilhamento de Evidências (MCE) é descrita com sua taxonomia de escopos (institucional, parcial e do curso), os resultados da classificação por dimensão avaliativa e exemplos de aplicação prática em contexto *multicampi*.

O **Capítulo 5 – Desenvolvimento do Framework** descreve a materialização dos artefatos analíticos em sistema computacional funcional. A primeira parte detalha a arquitetura

e engenharia de *software* da aplicação: pilha tecnológica, arquitetura em camadas e modular, modelo *multi-tenant*, modelo de dados do domínio SINAES, processos de ETL, serviços computacionais com o motor de regras da MCA, calculadora do Conceito de Curso e serviço de diagnóstico MPC, fluxo de dados e controle de acesso. A segunda parte apresenta a integração com Inteligência Artificial Generativa, abrangendo a seleção do provedor de IA, a arquitetura multiagente com três agentes especializados (Diagnóstico, Recomendação e Síntese), as ferramentas e o servidor MCP, a estratégia de engenharia de *prompt*, o *pipeline* RAG e a base de conhecimento, os fluxos de integração com a aplicação, a base de conhecimento global para aprendizado organizacional, as funcionalidades avançadas baseadas em agentes de IA e os mecanismos de transparência, auditoria e uso responsável de IA.

O **Capítulo 6 – Resultados e Discussão** documenta a aplicação do *framework* a cursos da UNEMAT por meio de prova de conceito, apresenta os resultados da validação multimétodo, incluindo análise retrospectiva com dados oficiais do CC, verificação de consistência e aderência normativa, e discute as contribuições, limitações e implicações dos resultados para a gestão de IES, em diálogo com a literatura pertinente.

O **Capítulo 7 – Conclusões** sintetiza os principais resultados alcançados em relação aos objetivos propostos, apresenta as contribuições científicas e tecnológicas da pesquisa, descreve o produto tecnológico desenvolvido, discute as limitações do estudo e propõe recomendações para trabalhos futuros.

A dissertação complementa-se com as referências bibliográficas.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Políticas Públicas de Avaliação na Educação Superior

A educação superior brasileira e seus mecanismos de avaliação constituem produto de um processo histórico complexo, marcado por transformações políticas, sociais e econômicas que moldaram progressivamente a configuração atual do sistema nacional de avaliação (SILVA; GHIRALDELLI, 2022). A compreensão desta trajetória é fundamental para contextualizar adequadamente as propostas contemporâneas de integração tecnológica aos processos avaliativos, uma vez que revela tanto os avanços conquistados quanto os desafios persistentes na busca pela qualidade educacional.

2.1.1 Histórico da Avaliação no Ensino Superior Brasileiro

A institucionalização da avaliação da educação superior no Brasil emergiu na década de 1970 com a criação da política de avaliação da pós-graduação pela CAPES, representando o primeiro esforço sistemático do Estado para implementar controle de qualidade no ensino superior (SOUZA; GUERRA, 2020). Este contexto caracterizava-se pela expansão acelerada do sistema, sustentada principalmente por instituições não universitárias, predominantemente privadas, suscitando preocupações quanto à manutenção de padrões mínimos de qualidade educacional (POLIDORI *et al.*, 2007).

O marco inaugural da avaliação sistemática da graduação ocorreu em 1983 com a criação do Programa de Avaliação da Reforma Universitária (PARU), primeira iniciativa governamental especificamente concebida para avaliar as Instituições de Ensino Superior brasileiras. Implementado pela CAPES em contexto político singular, caracterizado pelos movimentos pró-democratização, o programa estruturava-se em torno de dois eixos centrais: gestão das IES e produção/disseminação de conhecimentos, contemplando estudos sobre o impacto da Reforma Universitária (Lei 5.540/1968) em diversos aspectos institucionais (ANDRIOLA, 2008).

Uma das contribuições mais significativas do PARU foi a inauguração da concepção de avaliação formativa e emancipatória na educação superior brasileira, priorizando a "avaliação interna" como procedimento metodológico principal e propondo a participação

ativa da comunidade acadêmica (SOUZA; GUERRA, 2020). Apesar de suas inovações conceituais, o programa foi descontinuado em 1984, apenas um ano após sua implementação, privando o sistema educacional de análises que poderiam ter orientado políticas públicas mais efetivas (DOS-SANTOS-OLIVEIRA; ROTHEN, 2024).

Com o advento da Nova República em 1985, emergiram novas propostas por meio da Comissão Nacional para a Reformulação do Ensino Superior, que fundamentou a criação do Grupo Executivo para a Reformulação do Ensino Superior (GERES) em 1986. Embora o anteprojeto tenha sido rejeitado pela comunidade universitária, permaneceu como orientação política, estimulando iniciativas avaliativas em universidades como USP, UFMG, UnB e UFRJ.

O Programa de Avaliação Institucional das Universidades Brasileiras (PAIUB), criado em 1993, representou um marco fundamental na consolidação da cultura de avaliação na educação superior. Sustentado no princípio da adesão voluntária, o programa concebia a autoavaliação como etapa inicial de um processo que se estendia a toda a instituição e se completava com uma avaliação externa. Estabeleceu-se em sete princípios fundamentais: globalidade, comparabilidade, respeito à identidade institucional, não-premiação, adesão voluntária, legitimidade e continuidade (SINAES, 2009).

O PAIUB recuperou ideias fundamentais do PARU quanto ao foco na instituição e na participação da comunidade acadêmica, retomando a concepção de avaliação como instrumento de melhoria da qualidade educacional (SOUZA; GUERRA, 2019). Esta perspectiva contrastava com abordagens punitivas ou meramente regulatórias, enfatizando o caráter formativo e emancipatório da avaliação institucional. Apesar de sua duração relativamente curta e da interrupção do apoio ministerial, o programa recebeu ampla adesão das universidades, conferindo legitimidade à cultura de avaliação e tornando-se referência obrigatória para as reflexões que culminariam na concepção do Sistema Nacional de Avaliação da Educação Superior (SINAES, 2009).

2.1.2 Marco Legal e Institucional

A Lei de Diretrizes e Bases da Educação Nacional (Lei nº 9.394/1996) estabeleceu marcos normativos fundamentais para a avaliação da educação superior brasileira (HAAS; APARÍCIO, 2019; MARQUEZAN; SCREMIN, 2018). Em seu artigo 9º, a LDB determinou competências específicas da União, incluindo "assegurar um processo nacional de avaliação das instituições de educação superior, com a cooperação dos sistemas que tiverem

responsabilidade sobre este nível de ensino" e "autorizar, reconhecer, credenciar, supervisionar e avaliar, respectivamente, os cursos das instituições de educação superior e os estabelecimentos do seu sistema de ensino" (CARVALHO *et al.*, 2018; SOUZA; GUERRA, 2020).

Esta definição legal representou um avanço significativo ao institucionalizar a responsabilidade federal pela condução de processos avaliativos abrangentes e sistemáticos, estabelecendo bases jurídicas sólidas para a implementação de políticas nacionais de avaliação da qualidade educacional.

2.1.3 O Exame Nacional de Cursos e a Regulação Estatal

A Lei nº 9.131/1995 introduziu uma importante inovação no cenário avaliativo brasileiro mediante a criação do Exame Nacional de Cursos (ENC), posteriormente conhecido como "Provão". Esta iniciativa estabeleceu a aplicação anual de exames nacionais baseados nos conteúdos mínimos estabelecidos para cada curso e representando uma mudança paradigmática na abordagem estatal da avaliação educacional.

O ENC foi implementado sob a égide de um Estado regulador e avaliador, em substituição ao modelo tradicional de Estado executor direto. Esta transformação refletiu tendências internacionais de reforma do papel estatal, enfatizando funções de monitoramento e controle de qualidade em detrimento da provisão direta de serviços educacionais (SINAES, 2009).

A concepção original do SINAES estabeleceu um duplo mandato que constitui fonte permanente de tensão sistêmica: servir simultaneamente como instrumento de desenvolvimento institucional (função formativa) e como mecanismo de controle estatal (função regulatória). Esta dualidade cria o que pode ser denominado "um paradoxo avaliativo", a mesma ferramenta deve promover a autorreflexão institucional e fornecer subsídios para sanções administrativas (DOS-SANTOS-OLIVEIRA; ROTHEN, 2024; LORDÊLO; DAZZANI, 2009).

Dias Sobrinho (2008) argumenta que essa tensão é inerente a qualquer sistema nacional de avaliação que pretenda combinar responsabilização, controle e regulação (*accountability*) e melhoria da qualidade, formação e emancipação (*improvement*). No entanto, a análise do comportamento institucional das IES brasileiras revela que a dimensão punitiva tem consistentemente prevalecido sobre a dimensão formativa, resultando em estratégias

defensivas que privilegiam a conformidade sobre a qualidade.

O conceito de "*accountability* inteligente" (*smart accountability*), desenvolvido por O'Day (2002), oferece um arcabouço teórico para compreensão desta tensão. Segundo esta perspectiva, sistemas de avaliação eficazes devem equilibrar demandas por transparência e responsabilização com apoio adequado para melhoria institucional (AFONSO, 2012). A análise do SINAES revela um desequilíbrio significativo nesta equação, com ênfase desproporcional no controle em detrimento do suporte.

2.1.4 O SINAES: Síntese Evolutiva e Inovação Sistêmica

A criação do Sistema Nacional de Avaliação da Educação Superior (SINAES) pela Lei nº 10.861/2004, representou a síntese evolutiva das experiências anteriores, incorporando aprendizagens do PARU, PAIUB e ENC em uma proposta sistêmica e abrangente. O SINAES emergiu do trabalho da Comissão Especial de Avaliação (CEA), designada pela Portaria SESU em 28 de abril de 2003, que sistematizou décadas de experiência em avaliação educacional desde os anos 1980.

Conforme estabelecido em sua lei instituidora, o sistema visa: a melhoria da qualidade da educação superior; a orientação da expansão de sua oferta; o aumento permanente da eficácia institucional e efetividade acadêmica e social; e a promoção do aprofundamento dos compromissos e responsabilidades sociais das instituições de educação superior.

A nova proposta caracteriza-se pela integração de diversos instrumentos avaliativos em uma visão global do sistema educacional, buscando respeitar a identidade e diversidade institucionais enquanto possibilita a análise comparativa da realidade e missão de cada IES. O sistema fundamenta-se em três pilares complementares: avaliação institucional (interna e externa), avaliação de cursos de graduação e avaliação do desempenho estudantil através do Exame Nacional de Desempenho dos Estudantes (ENADE) (SINAES, 2009).

A filosofia avaliativa do SINAES fundamenta-se em princípios que transcendem a mera verificação de conformidade regulatória. Conforme estabelecido em seus documentos norteadores, o sistema busca promover a melhoria da qualidade da educação superior através de processos avaliativos que sejam formativos, sistêmicos, participativos e que respeitem a identidade e diversidade das instituições (SINAES, 2009).

Conforme observado por Dias Sobrinho (2003), o SINAES representa uma mudança paradigmática significativa em relação aos sistemas avaliativos anteriormente

vigentes, principalmente no que se refere aos conceitos de avaliação e educação superior. Esta transformação não constitui mera opção técnica, mas uma decisão fundamentalmente filosófica e política que reposiciona o papel da avaliação na educação superior brasileira.

A avaliação no SINAES caracteriza-se como formativa, voltada para processos em detrimento de produtos institucionais, prevendo a participação ativa dos agentes internos. Simultaneamente, o sistema incorpora o controle necessário que uma política pública requer do Estado, estabelecendo uma dupla perspectiva de limites e possibilidades. Ao impor limites através do controle, o SINAES abre possibilidades para o aperfeiçoamento da educação superior e o direcionamento de políticas públicas na área.

Entretanto, a operacionalização do SINAES ao longo de duas décadas revela uma tensão fundamental entre seus objetivos declarados e sua implementação prática. A análise crítica da literatura especializada (Barreyro, 2018; Polidori, 2009) indica que a dimensão regulatória e classificatória do sistema gradualmente se sobrepôs à sua dimensão formativa e emancipatória.

2.1.5 Dimensões e Instrumentos Avaliativos

O SINAES integra três modalidades de instrumentos avaliativos aplicados em momentos distintos, configurando abordagem sistemática e complementar que proporciona visão abrangente da qualidade educacional:

1. Avaliação das Instituições de Educação Superior: Dividida em duas etapas complementares: a autoavaliação coordenada pela Comissão Própria de Avaliação (CPA) de cada IES, e a avaliação externa realizada por comissões designadas pelo INEP para IES federais/privadas, seguindo diretrizes estabelecidas pela CONAES. Esta modalidade constitui o eixo central do sistema, organizando a coerência do conjunto de instrumentos avaliativos.

2. Avaliação dos Cursos de Graduação: Realizada através de instrumentos específicos e visitas *in loco* de comissões externas designadas pelo INEP para IES federais/privadas, seguindo diretrizes estabelecidas pela CONAES, esta modalidade foca na qualidade dos projetos pedagógicos, corpo docente, infraestrutura e resultados dos cursos ofertados pelas IES.

3. Avaliação do Desempenho dos Estudantes (ENADE): Avalia estudantes ingressantes e concluintes dos cursos de graduação através de exames aplicados periodicamente, permitindo análise da evolução do desempenho acadêmico e da efetividade dos processos

formativos. A seleção dos estudantes é realizada de forma amostral e as áreas participantes são definidas pela CONAES em ciclos trienais.

A Lei nº10.861, de 14 de abril de 2004, estabeleceu dez dimensões institucionais que constituem aspectos mínimos obrigatórios da autoavaliação, garantindo processo avaliativo abrangente e adequado às especificidades de cada IES:

Dimensão I - Missão e Plano de Desenvolvimento Institucional: Proporciona informações sobre finalidades, objetivos e compromissos da IES, práticas pedagógicas e administrativas, PDI e Projeto Pedagógico Institucional, especialmente quanto às atividades de ensino, pesquisa, extensão, gestão acadêmica e avaliação institucional.

Dimensão II - Política Institucional para Ensino, Pós-Graduação, Pesquisa e Extensão: Analisa concepção curricular e organização didático-pedagógica, práticas pedagógicas, pertinência dos currículos considerando objetivos institucionais e demandas sociais, científicas, econômicas e culturais.

Dimensão III - Responsabilidade Social da Instituição: Examina contribuições para inclusão social, desenvolvimento econômico e social, defesa do meio ambiente, preservação da memória cultural, produção artística e patrimônio cultural.

Dimensão IV - Comunicação com a Sociedade: Avalia estratégias de comunicação interna e externa, imagem pública institucional e divulgação de serviços, produtos e ações nos meios de comunicação social.

Dimensão V - Política Institucional de Gestão de Pessoal: Analisa políticas de pessoal, carreiras do corpo docente e técnico-administrativo, aperfeiçoamento, desenvolvimento profissional e condições de trabalho.

Dimensão VI - Organização e Gestão da Instituição: Examina funcionamento e representatividade dos colegiados, independência e autonomia na relação com a mantenedora, participação dos segmentos da comunidade universitária nos processos decisórios.

Dimensão VII - Infraestrutura Física: Avalia adequação da infraestrutura de ensino e pesquisa, bibliotecas, recursos de informação e comunicação à missão institucional.

Dimensão VIII - Planejamento e Avaliação: Analisa processos, resultados e eficácia da autoavaliação institucional, articulação entre planejamento estratégico e avaliação.

Dimensão IX - Política de Atendimento aos Estudantes: Examina políticas de acesso, permanência, apoio pedagógico, financeiro e social aos estudantes.

Dimensão X - Sustentabilidade Financeira: Avalia viabilidade financeira institucional considerando o significado social da continuidade dos compromissos na oferta da

educação superior.

Além disso, cada dimensão é desdobrada em indicadores específicos, avaliados por meio de uma escala de cinco níveis (LEITE, 2017). Esses conceitos não apenas refletem as expectativas do Estado, como também consolidam o modelo institucionalizado de avaliação da educação superior brasileira (LEITE, 2017).

2.1.6 Autoavaliação Institucional: Processos e Desafios

A autoavaliação constitui etapa fundamental do SINAES, representando um processo reflexivo através do qual as IES desenvolvem autoconhecimento sistemático sobre suas atividades, estruturas e resultados (DOS-SANTOS-OLIVEIRA; ROTHEN, 2024). Segundo o INEP, a autoavaliação objetiva produzir conhecimentos, refletir sobre atividades institucionais, identificar causas de problemas, aperfeiçoar a consciência pedagógica e capacidade profissional do corpo docente e técnico-administrativo, fortalecer relações de cooperação entre atores institucionais, tornar mais efetiva a vinculação com a comunidade e prestar contas à sociedade (ZIMMERMANN; ALVES, 2021).

Esta concepção transcende perspectivas meramente burocráticas ou regulatórias, posicionando a autoavaliação como instrumento de gestão estratégica capaz de promover um olhar reflexivo da instituição sobre ela mesma, nas diversas dimensões em que ela se constitui. O processo visa refletir sobre conteúdo e forma das ações administrativas, financeiras e pedagógicas, identificando potencialidades e fragilidades para projetar ações futuras aperfeiçoadas.

De acordo com Sanches e Raphael (2006), a avaliação institucional configura-se como um movimento abrangente de autoconhecimento, que permite a revisão das metas planejadas e a construção contínua da identidade universitária. Essa análise deve estar profundamente alinhada à missão, ao percurso histórico e ao contexto social da instituição. Embora a avaliação não solucione isoladamente todos os desafios institucionais, ela cumpre o papel fundamental de diagnosticar gargalos e estimular a reflexão crítica sobre as soluções mais adequadas.

Nesse sentido, quando o processo avaliativo é estruturado de maneira criteriosa, ele é capaz de evidenciar tanto as potencialidades quanto as fragilidades da instituição. Esse diagnóstico torna-se um subsídio indispensável para o planejamento estratégico e para a definição de novos rumos acadêmicos e administrativos (SANCHES, 2007).

Além disso, a autoavaliação responde a uma demanda por transparência e responsabilidade social. No caso das instituições públicas, o processo justifica o emprego dos recursos advindos da sociedade; já nas instituições privadas, atende ao crescente nível de exigência da comunidade escolar quanto ao retorno dos investimentos realizados. Assim, ao consolidar uma cultura de avaliação, as instituições fortalecem sua credibilidade e reconhecimento público (ANDRIOLA, 2004).

2.1.7 O Processo de Autoavaliação: O Papel e os Desafios da CPA

Toda IES, por força da Lei nº 10.861, de 14 de abril de 2004, deve constituir uma Comissão Própria de Avaliação (CPA). As Comissões Próprias de Avaliação (CPA) constituem instâncias estratégicas do SINAES, responsáveis pela coordenação dos processos internos de avaliação, sistematização e prestação de informações solicitadas pelo INEP. Conforme estabelecido pela Portaria MEC nº 2.051/2004, estas comissões devem ser constituídas no âmbito de cada IES, pública ou privada, observando diretrizes específicas que assegurem legitimidade e efetividade do processo.

A composição da CPA deve garantir participação de todos os segmentos da comunidade acadêmica (docente, discente e técnico-administrativo) e representantes da sociedade civil organizada, vedada a existência de maioria absoluta de qualquer segmento. Esta configuração plural visa assegurar diferentes perspectivas e interesses na condução dos processos avaliativos, conferindo legitimidade democrática às análises e proposições resultantes. A legitimidade do processo avaliativo está intrinsecamente ligada à sua natureza democrática, exigindo uma composição que neutralize hegemonias setoriais e incorpore o olhar da comunidade externa (DIAS SOBRINHO, 2008)

Para que a avaliação não se torne um instrumento de validação política da gestão, é indispensável que a CPA possua independência técnica para apontar fragilidades sem sofrer retaliações ou censuras institucionais (LEITE, 2017). A CPA deve atuar com autonomia em relação a conselhos e demais órgãos colegiados institucionais, condição fundamental para garantir independência técnica e política necessária ao desenvolvimento de avaliações críticas e propositivas. Esta autonomia não implica isolamento, mas capacidade de desenvolver trabalho técnico sem interferências que possam comprometer a objetividade e fidedignidade dos processos avaliativos.

A autoavaliação institucional requer o compromisso explícito dos dirigentes das

IES, que devem assegurar os recursos necessários e, principalmente, a utilização dos resultados para o planejamento (GRIBOSKI, 2012). A eficácia da autoavaliação depende da transição de uma cultura de “prestação de contas” para uma “cultura de autoconhecimento”, onde os dados servem de base para decisões estratégicas (POLIDORI, 2009).

O ciclo avaliativo proposto pelo SINAES compreende um percurso metodológico que vai além da coleta de dados, exigindo uma fase de sensibilização inicial e, crucialmente, um balanço crítico que retroalimente o planejamento institucional (INEP, 2017).

Para que a autoavaliação alcance resultados satisfatórios, é imprescindível o atendimento de condições fundamentais que assegurem credibilidade e utilidade do processo:

Equipe de Coordenação Representativa: Necessidade de constituição de equipe com participação de todos os atores institucionais, garantindo legitimidade e abrangência das análises desenvolvidas.

Compromisso Explícito dos Gestores: Fundamental que os gestores institucionais assumam compromisso explícito de utilizar os resultados da autoavaliação por meio de ações fundamentadas nos pontos fracos evidenciados no processo de avaliação. Sem esse compromisso, a autoavaliação pode tornar-se um exercício meramente formal.

Confiabilidade de Informações e Dados: Imprescindível que informações coletadas e os dados sejam confiáveis, sendo necessária a utilização de recursos tecnológicos adequados para conferir credibilidade ao processo avaliativo.

O não atendimento de qualquer um desses requisitos pode comprometer significativamente a eficácia da autoavaliação, resultando em processos que não contribuem efetivamente para a melhoria institucional e podem prejudicar a instituição nos momentos de avaliação externa.

A autoavaliação desenvolve-se por meio de etapas sistemáticas que asseguram uma abordagem metodológica rigorosa:

Preparação: Constituição da CPA conforme determinações legais, elaboração do projeto de autoavaliação, planejamento das atividades e mobilização da comunidade acadêmica.

Desenvolvimento: Execução das atividades de coleta de dados, aplicação de instrumentos, realização de consultas à comunidade acadêmica e análise das informações obtidas.

Consolidação: Sistematização dos resultados, elaboração de relatórios, proposição de ações de melhoria e divulgação dos resultados à comunidade.

Balanço Crítico: Reflexão sobre estratégias utilizadas, dificuldades e avanços

apresentados, permitindo planejamento de ações futuras e aprimoramento contínuo do processo.

O relatório produzido pela CPA é um documento central que subsidia tanto a gestão interna quanto a avaliação externa. No entanto, a efetividade das CPAs é frequentemente comprometida por desafios significativos, tais como: (a) baixa participação da comunidade acadêmica, que pode não se sentir engajada no processo; (b) A percepção da autoavaliação como um mero exercício burocrático para cumprir exigências legais; (c) A produção de relatórios que são superficiais, não identificam falhas críticas ou não geram ações de melhoria concretas.

A resistência ou apatia da comunidade acadêmica muitas vezes decorre da percepção de que os relatórios da CPA são peças documentais sem impacto prático na melhoria das condições de ensino e trabalho (FESTA, 2019). Agravando o cenário, as constantes mudanças nos instrumentos e sistemas do MEC/INEP podem frustrar as comissões e invalidar o trabalho já realizado, minando a credibilidade do processo internamente (CARVALHO; MALAVASI, 2020).

Uma CPA com atuação insuficiente gera incertezas para a gestão da IES. A finalidade da autoavaliação é construir um conhecimento profundo sobre a própria realidade institucional, identificando com clareza os pontos fortes e as fragilidades. Quando esse processo é falho, a liderança da instituição fica sem uma inteligência precisa sobre seu próprio estado. A visita *in-loco* passa a ser um "uma situação de incerteza avaliativa", baseada em informações incompletas ou imprecisas. O risco de rebaixamento conceitual é, portanto, uma consequência direta dessa incerteza.

Uma CPA robusta atua como um mecanismo de redução de riscos, permitindo que a instituição antecipe problemas que seriam detectados apenas na avaliação externa (BITTENCOURT, 2008). Fortalecer a CPA é, portanto, o primeiro e mais crucial passo para superar o dilema.

Desta forma, a autoavaliação institucional, operacionalizada através das Comissões Próprias de Avaliação (CPA), constitui o fundamento do SINAES e deveria funcionar como sistema de inteligência institucional para tomada de decisão estratégica. No entanto, múltiplos estudos (LEITE, 2017; ANDRIOLA, 2009; SORDI & LUDKE, 2009) identificam deficiências sistemáticas que comprometem sua eficácia.

As principais manifestações da gestão ineficiente incluem:

Baixa Participação Comunitária: *Surveys* realizados pelas CPA frequentemente apresentam taxas de resposta baixas, comprometendo a representatividade e validade dos

resultados.

Superficialidade Analítica: Relatórios de autoavaliação tendem a ser descritivos em vez de analíticos, limitando-se a apresentar dados sem identificar padrões, tendências ou relações causais.

Desconexão com a Gestão: Os resultados da autoavaliação raramente são integrados aos processos de planejamento estratégico e tomada de decisão institucional.

Periodicidade Inadequada: Muitas CPAs produzem relatórios apenas para cumprimento de exigências legais, sem estabelecer ciclos contínuos de monitoramento.

Limitações Metodológicas: Ausência de expertise em métodos quantitativos avançados, análise estatística e mineração de dados limita a capacidade de descoberta de conhecimento não trivial.

2.1.8 Os Documentos Fundamentais: A Centralidade do PDI e do PPC

No centro de qualquer processo avaliativo do SINAES estão dois documentos: o Plano de Desenvolvimento Institucional (PDI) e o Projeto Pedagógico de Curso (PPC). O PDI é o documento que expressa a missão e o planejamento estratégico da IES, enquanto o PPC define a identidade, os objetivos, a matriz curricular e as práticas pedagógicas de um curso específico. A comissão de avaliação externa tem como tarefa primordial verificar a coerência entre o que está prometido nesses documentos e a realidade operacional da instituição. O instrumento de avaliação de cursos é explicitamente projetado para auditar o PPC em suas três dimensões: Organização Didático-Pedagógica, Corpo Docente e Tutorial, e Infraestrutura.

Nesse contexto, o PPC deve ser entendido não como um documento estático e burocrático, mas como o "contrato legal" firmado entre a IES e o MEC. A avaliação *in-loco* é, em sua essência, uma auditoria que verifica o cumprimento desse contrato. Cada promessa feita no PPC, seja sobre a qualificação do corpo docente, os recursos da biblioteca, as metodologias de ensino ou as atividades de extensão, torna-se uma obrigação contratual. A melhoria da qualidade dos cursos e dos Projetos Pedagógicos (PPC) depende da articulação efetiva entre a Comissão Própria de Avaliação (CPA), o Núcleo Docente Estruturante (NDE) e as coordenações, permitindo uma "visão ampliada da realidade institucional" e direcionamentos precisos para a gestão (GRIBOSKI *et al.*, 2024).

Qualquer discrepância entre o prometido e o observado pode ser apontada como uma não conformidade, resultando em uma avaliação negativa. Portanto, um PPC desatualizado,

excessivamente ambicioso ou desalinhado com a prática cotidiana representa uma vulnerabilidade crítica para a instituição.

2.1.9 O Motor Acadêmico: O Papel Crítico do NDE

O Núcleo Docente Estruturante (NDE) é o grupo de professores com liderança acadêmica responsável pela concepção, implementação e atualização contínua do PPC. A existência e a atuação do NDE são, por si só, um indicador de qualidade explicitamente verificado nos instrumentos de avaliação do INEP. Seu papel não é meramente administrativo; é o NDE que garante a integração curricular, a relevância do conteúdo e a constante modernização do curso, sendo considerado a "alma" do projeto pedagógico.

O NDE transcende a função burocrática ao atuar como o núcleo intelectual do curso, sendo responsável por conferir identidade e coesão pedagógica à proposta formativa, assegurando que o currículo responda aos desafios contemporâneos da profissão. "O NDE deve ser o responsável pelo acompanhamento do curso, zelando pelo cumprimento das diretrizes curriculares nacionais e pela qualidade da formação oferecida" (INEP, 2017).

Dada sua responsabilidade direta sobre o "contrato" (o PPC), um NDE funcional e engajado constitui a mais importante linha de defesa da instituição contra um resultado avaliativo negativo. Inversamente, um NDE passivo, cujas reuniões são meramente protocolares, é um forte indicativo de risco. A eficácia de um curso de graduação está diretamente ligada à capacidade de seu NDE em transformar as diretrizes institucionais em ações pedagógicas concretas, sendo a qualidade de suas atas e relatórios um dos principais preditores de sucesso em avaliações *in loco* (BITTENCOURT, 2008).

A saúde do NDE funciona como um termômetro da saúde da gestão acadêmica do curso. Avaliadores experientes identificam rapidamente essa correlação. Um NDE forte, com atas de reunião detalhadas, discussões documentadas e evidências claras de sua influência sobre as melhorias do PPC, fornece uma narrativa poderosa de gestão da qualidade e compromisso com a melhoria contínua, que responde diretamente aos principais indicadores da avaliação. Para o gestor, analisar a funcionalidade do NDE é um passo indispensável na análise de risco. No âmbito do SINAES, o NDE é avaliado não apenas pela sua existência formal, mas pela sua capacidade de liderança e impacto real na gestão acadêmica, funcionando como uma garantia de que o projeto pedagógico não é um documento estático, mas um organismo vivo.

2.1.10 Avaliação Externa e Integração Sistêmica

A avaliação externa constitui a segunda etapa da avaliação institucional, realizada por comissões de pares especialmente designadas pelo INEP e constituídas a partir de um banco de avaliadores (MARQUEZAN; SCREMIN, 2018; NASCIMENTO; HETKOWSKI, 2009). Estas avaliações utilizam como referência os instrumentos estabelecidos pelo sistema e têm como objetivo verificar *in loco* as informações prestadas pela instituição na autoavaliação, conferindo credibilidade e objetividade ao processo (SINAES, 2009).

As visitas de avaliação externa aplicam conceitos ordenados em cinco níveis a cada uma das dimensões institucionais e ao seu conjunto, estabelecendo padrões comparativos de qualidade (LEITE, 2017). Para tanto, as instituições visitadas devem proporcionar condições necessárias à realização das visitas, incluindo disponibilização de sala privativa para os trabalhos da comissão, acesso a documentos complementares e acompanhamento aos locais programados (CONAES, 2004).

O ENADE desempenha papel fundamental na articulação sistêmica do SINAES, constituindo simultaneamente ponto de partida e chegada da avaliação (RIBEIRO; GUERRA, 2019). Conforme observado por Griboski (2012), "tudo se inicia com a avaliação do estudante (ENADE), que possibilita o cálculo dos insumos para compor o Conceito Preliminar de Curso (CPC). Com base na média dos CPCs, realiza-se o cálculo do indicador de qualidade da instituição (Índice Geral de Cursos) e, em seguida, realizam-se as avaliações *in loco*".

Esta arquitetura sistêmica permite que avaliadores e avaliados tracem rumos, metas e inovações na busca da qualidade, configurando um ciclo avaliativo que proporciona *feedback* contínuo para melhoria institucional e orientação de políticas públicas (LORDÊLO; DAZZANI, 2009).

A coordenação nacional do SINAES é exercida pela Comissão Nacional de Avaliação da Educação Superior (CONAES), órgão colegiado responsável pela formulação de diretrizes, critérios e estratégias para os processos avaliativos. A operacionalização técnica é responsabilidade do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), que desenvolve instrumentos, constitui comissões de avaliação e processa informações do sistema.

Esta estrutura de governança busca articular autonomia técnica com um controle democrático, assegurando que os processos avaliativos mantenham rigor metodológico enquanto respondem às necessidades e expectativas da comunidade acadêmica e da sociedade

(SINAES, 2009).

2.1.11 Os sistemas estaduais de ensino superior e sua articulação com o SINAES

A educação superior brasileira organiza-se em arquitetura regulatória dual, constitucionalmente assentada e consolidada pela Lei de Diretrizes e Bases da Educação Nacional (BRASIL, 1996). O sistema federal de ensino superior congrega as instituições federais, as privadas e as demais instituições integradas ao Sistema Federal de Ensino, cuja regulação compete ao Ministério da Educação (MEC) e cuja operacionalização técnica, no que se refere à avaliação, é atribuída ao Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) no âmbito do SINAES (SINAES, 2004). Os sistemas estaduais de ensino superior, conforme estabelece o artigo 17 da LDB, congregam as instituições mantidas pelo poder público estadual, sob a regulação dos respectivos Conselhos Estaduais de Educação. Coexistem, portanto, competências regulatórias distintas, ainda que orientadas por princípios convergentes.

A Universidade do Estado de Mato Grosso, instituição pública estadual *multicampi*, integra o Sistema Estadual de Ensino de Mato Grosso (SES/MT). Nesse arranjo, a autoavaliação institucional é conduzida pela Comissão Própria de Avaliação (CPA), estabelecida nos termos do artigo 11 da Lei nº 10.861/2004; a avaliação externa é operacionalizada pela Secretaria de Estado de Ciência, Tecnologia e Inovação (SECITECI), por meio de comissões verificadoras designadas a partir de banco de avaliadores; e a regulação, a supervisão e a emissão dos atos autorizativos, credenciamento, recredenciamento, reconhecimento e renovação de reconhecimento, competem ao Conselho Estadual de Educação (CEE/MT). As bases normativas desse fluxo estão consolidadas na Resolução nº 311/2008-CEE/MT (MATO GROSSO, 2008), que fixa o roteiro de verificação *in loco* com referência explícita ao SINAES, na Resolução Normativa nº 001/2011-CEE/MT (MATO GROSSO, 2011), que estabelece as diretrizes para a elaboração dos instrumentos de avaliação utilizados nos processos de regulação, e na Resolução nº 01/2017-CEE/MT (MATO GROSSO, 2017), que delimita as competências de cada instância no ciclo regulatório.

A articulação entre o sistema estadual matogrossense e o SINAES constitui alinhamento normativo explícito, e não mera convergência de fato. A Resolução Normativa nº 001/2011-CEE/MT invoca a Lei Federal nº 10.861/2004 como fundamento das diretrizes nela estabelecidas, de modo que os instrumentos de avaliação produzidos no âmbito do SES/MT

espelham a estrutura do Instrumento de Avaliação de Cursos de Graduação do INEP: organizam-se nas mesmas três dimensões, Organização Didático-Pedagógica, Corpo Docente e Tutorial e Infraestrutura, adotam a escala de cinco níveis conceituais para a atribuição do Conceito de Curso (CC) e observam os princípios avaliativos preconizados pelo sistema nacional, notadamente a integração entre autoavaliação e avaliação externa, a formalidade do processo e o uso dos resultados para a melhoria institucional. Essa homologia estrutural é reconhecida pela própria UNEMAT, cujo Projeto de Autoavaliação Institucional declara, como documento norteador, as orientações do SINAES e as Notas Técnicas do INEP/DAES/CONAES (UNEMAT, 2022).

A explicitação dessa dupla arquitetura regulatória é relevante para o presente trabalho por duas razões. Em primeiro lugar, sustenta a legitimidade metodológica de adotar o Instrumento de Avaliação de Cursos de Graduação do INEP como corpus para a construção das matrizes analíticas descritas no Capítulo 4, ainda que a validação do *framework* se opere em cursos da UNEMAT. Em segundo lugar, amplia o potencial de replicação do artefato proposto, cuja arquitetura, ancorada nos princípios comuns a ambos os subsistemas, admite aplicação tanto em instituições submetidas à regulação federal quanto em instituições vinculadas a sistemas estaduais de ensino que observem alinhamento normativo análogo ao do SES/MT.

2.1.12 Conceito de Curso e Conceito Preliminar de Curso: estrutura, ponderação e implicações regulatórias

O SINAES opera com uma "moeda" própria, materializada em conceitos que definem o status de um curso. Os dois principais são o Conceito de Curso (CC) e o Conceito Preliminar de Curso (CPC).

O **Conceito de Curso (CC)** e o **Conceito Preliminar de Curso (CPC)** constituem os principais indicadores sintéticos produzidos no âmbito do SINAES para mensuração da qualidade de cursos de graduação. O CC resulta da avaliação *in loco* (presencial ou virtual) conduzida por comissões externas designadas pelo INEP, com base no Instrumento de Avaliação de Cursos de Graduação, organizado em três dimensões: Organização Didático-Pedagógica, Corpo Docente e Tutorial, e Infraestrutura. O CPC, por sua vez, é calculado a partir de componentes objetivos, atribuindo pesos distintos a cada componente.

O Conceito Preliminar de Curso (CPC) exemplifica a transformação de um instrumento técnico em um ativo estratégico institucional. Originalmente concebido como um

indicador preliminar que funcionaria como "filtro" para avaliações *in loco*, o CPC adquiriu centralidade no sistema que transcende sua função original.

A fórmula do CPC, conforme detalhado na Tabela 1, combina de forma ponderada os eixos de Desempenho dos Estudantes (55%), Corpo Docente (30%) e Percepção Discente (15%). Esta composição reflete uma perspectiva de qualidade que equilibra resultados (*outcomes*) e insumos (*inputs*), seguindo modelos consolidados na literatura de avaliação educacional.

Tabela 1 – Pesos dos componentes no conceito preliminar de curso

Eixo de Avaliação	Indicador (Sigla)	Peso Individual	Total por Eixo
Desempenho dos Estudantes	Nota dos Concluintes no ENADE (NC)	20%	55%
	Nota do Indicador de Diferença (NIDD)	35%	
Corpo Docente	Nota de Proporção de Mestres (NM)	7,5%	30%
	Nota de Proporção de Doutores (ND)	15%	
	Nota de Regime de Trabalho (NR)	7,5%	
Percepção Discente	Organização Didático-Pedagógica (NO)	7,5%	15%
	Infraestrutura e Instalações Físicas (NF)	5%	
	Oportunidades de Ampliação da Formação (NA)	2,5%	

Fonte: SINAES (2017).

O Conceito de Curso (CC) é o conceito final, atribuído em uma escala de 1 a 5. Este conceito substitui o CPC e passa a ser a referência oficial do curso.

Entretanto, a literatura técnica aponta uma desconexão crítica entre esses dois indicadores. Estudos demonstram que "não há correlação lógica entre os dois conceitos (CPC e CC)", uma vez que as avaliações *in loco* (CC) resultam sistematicamente em notas superiores àquelas obtidas via CPC. Essa divergência gera incertezas regulatórias, pois um CPC insatisfatório (faixas 1 e 2) pode acarretar penalidades como a redução de vagas e a obrigatoriedade de visitas compulsórias, que muitas vezes atribuem conceitos satisfatórios (CC ≥ 3) ao curso. Essa realidade indica a necessidade de uma revisão metodológica nos critérios de atribuição de valores para assegurar a coerência do sistema de avaliação (TOZZI; TOZZI, 2017).

Embora projetado para ser um gatilho para avaliações, o CPC tornou-se, na prática, um fim em si mesmo. Para um curso com um CPC satisfatório (conceito 3 ou 4), este indicador funciona como um escudo e, ao mesmo tempo, um teto. É um escudo porque, conforme a regulamentação, protege o curso da obrigatoriedade de uma avaliação externa e de possíveis sanções, garantindo a renovação automática de seu reconhecimento. É um teto porque o risco

de perder essa proteção ao buscar um CC superior frequentemente supera a recompensa potencial. O CPC, portanto, deixa de ser "preliminar" e se transforma em um ativo estratégico que a IES busca proteger, evitando a incerteza de uma avaliação *in-loco*.

Se analisarmos através da lente da teoria da decisão sob incerteza (KNIGHT, 1921). As IES com CPC 3 ou 4 enfrentam uma situação de escolha em que os resultados possíveis (melhoria, manutenção ou piora do conceito) não podem ser determinados com certeza, mas apenas estimados probabilisticamente.

A teoria do prospecto (*Prospect Theory*) oferece um entendimento para a compreensão do dilema. Tversky e Kahneman (1992) demonstraram que indivíduos tendem a ser avessos à perda e preferem evitar perdas a obter ganhos equivalentes. No contexto do SINAES, o CPC satisfatório representa um ponto de referência a partir do qual perdas potenciais (rebaixamento) são percebidas como mais significativas que ganhos potenciais (melhoria do conceito).

Esta aversão à perda é amplificada pela "gestão ineficiente da autoavaliação", que cria incerteza epistemológica sobre a real condição institucional. Quando gestores não possuem informações confiáveis sobre sua prontidão avaliativa, o risco percebido aumenta exponencialmente, intensificando a preferência pelo status quo.

A regulamentação do SINAES estabelece caminhos distintos para os cursos, baseados em seu desempenho no CPC. Essa estrutura cria o ponto crucial da decisão para a IES. Cursos com CPC 1 ou 2 são compulsoriamente submetidos a uma visita de avaliação *in-loco* para verificação de suas condições de oferta e, caso o baixo desempenho seja confirmado, são levados a assinar um Protocolo de Compromisso com o MEC. Por outro lado, cursos com CPC 5 são dispensados da avaliação, tendo sua excelência reconhecida.

O dilema reside na faixa intermediária. Cursos com CPC 3 ou 4 têm a opção de solicitar uma avaliação *in-loco* para tentar elevar seu conceito para 4 ou 5. Contudo, não são obrigados a fazê-lo, podendo optar por transformar seu CPC em conceito permanente, o que garante a renovação do reconhecimento do curso, como pode ser observado na Tabela 2.

Tabela 2 – O Fluxo Avaliativo do SINAES por Faixa de CPC

Faixa de CPC	Status da Avaliação <i>In-Loco</i>	Consequência Regulatória Imediata	Implicação Estratégica para a IES
1 ou 2	Obrigatória	Abertura de processo de ofício pelo MEC; risco de sanções e celebração de Protocolo de	Ação Corretiva Urgente: A IES deve obrigatoriamente se preparar para a visita e para sanar as deficiências apontadas.

		Compromisso.	
3	Opcional	Renovação automática do reconhecimento do curso; elegibilidade para programas federais.	Zona de Conforto: A IES está em conformidade. A decisão de buscar um conceito maior envolve uma análise de risco versus recompensa.
4	Opcional	Renovação automática do reconhecimento do curso; elegibilidade para programas federais.	Posição Sólida: A IES possui um bom indicador. A busca pelo conceito 5 é uma questão de estratégia de mercado e busca pela excelência.
5	Dispensada	Renovação automática do reconhecimento do curso; conceito máximo atribuído.	Liderança de Mercado: A IES pode usar o conceito como um diferencial competitivo e focar na manutenção da excelência.

Fonte: Elaborado pelo autor (2026).

2.1.13 Consequências regulatórias e estratégicas dos indicadores de qualidade: a lógica do incentivo institucional

Os conceitos do SINAES transcendem a esfera acadêmica, possuindo um valor tangível e imediato no mercado. IES com conceitos elevados (4 e 5) utilizam essa chancela do MEC como uma poderosa ferramenta de marketing para atrair e reter alunos, construindo uma imagem de credibilidade e qualidade.

Mais criticamente, um conceito satisfatório (igual ou superior a 3) é um requisito fundamental para a participação da IES em programas de financiamento e bolsas do governo federal, como o Fundo de Financiamento Estudantil (FIES) ou o Programa Universidade para Todos (ProUni). Cursos com conceitos 1 ou 2 podem ser suspensos desses programas, o que representa um impacto financeiro e de captação devastador, especialmente para IES privadas (MORETTI, 2016).

Essa estrutura cria um forte incentivo para que as instituições busquem atingir e manter o conceito 3, a nota mínima para a conformidade e para o acesso aos programas federais. Essa busca pelo "suficiente" estabelece o que pode ser chamado de "Zona de Conforto". Embora seja uma estratégia racional do ponto de vista de uma IES individual, que busca minimizar riscos e garantir sua operação, essa tendência representa um risco sistêmico para a qualidade da educação superior no país. A acomodação com o conceito 3 pode levar a uma estagnação generalizada, onde a maioria dos cursos se contenta com a conformidade em vez de aspirar à excelência, minando o objetivo primordial do SINAES de promover a melhoria contínua da

qualidade.

A decisão de enfrentar uma avaliação *in-loco* voluntária não pode ser baseada apenas na análise do ambiente externo. Ela depende, fundamentalmente, de um diagnóstico honesto e rigoroso da prontidão interna da instituição. A "gestão ineficiente da autoavaliação", mencionada anteriormente, é o fator central que transforma a oportunidade de melhoria em uma ameaça de rebaixamento.

2.1.14 Racionalidade institucional frente ao risco de rebaixamento: entre conformidade e qualidade

O temor generalizado de um rebaixamento de conceito pode ser combatido e transformado em um plano de ação gerenciável. Isso requer a transição de uma ansiedade difusa para uma análise de risco específica, estruturada e baseada em evidências. As ferramentas para essa análise são os próprios instrumentos e diretrizes do SINAES.

O "Instrumento de Avaliação de Cursos de Graduação" (para Reconhecimento e Renovação de Reconhecimento) é o guia que a comissão do INEP utiliza para conduzir sua análise. Ele é estruturado em três dimensões principais:

- **Dimensão 1** - Organização Didático-Pedagógica;
- **Dimensão 2** - Corpo Docente e Tutorial; e
- **Dimensão 3** - Infraestrutura.

Uma análise detalhada deste instrumento permite identificar os indicadores de maior risco para um curso que se encontra em uma posição frágil.

A avaliação *in-loco* não é uma busca subjetiva por "qualidade", mas sim uma auditoria objetiva de "evidências". A comissão constrói seu relatório com base nas provas documentais e nas verificações factuais que consegue coletar durante a visita. A ausência de evidência é, para todos os efeitos, tratada como ausência de qualidade. Isso significa que um curso pode ter práticas excelentes, mas receber um conceito baixo se falhar em documentar adequadamente seus processos, decisões e resultados. O ônus da prova recai inteiramente sobre a IES. O medo do rebaixamento pode, portanto, ser diretamente mitigado ao deslocar o foco institucional de "ser bom" para "provar que somos bons" por meio de uma documentação meticulosa e organizada, conforme detalhado em listas de verificação de instituições experientes.

2.1.15 Desafios Contemporâneos e Perspectivas de Aprimoramento

Apesar dos avanços significativos proporcionados pelo SINAES, persistem desafios importantes que limitam seu potencial transformador. Barreyro (2018) alertam que "na proposta do SINAES, a questão da qualidade foi mencionada como objetivo principal do sistema (melhoria da qualidade), mas não claramente definida, permanecendo como um conceito impreciso e, às vezes, contraditório".

Esta imprecisão conceitual pode gerar interpretações divergentes e comprometer a efetividade dos processos avaliativos. Entretanto, é questionável se uma definição rígida de qualidade seria adequada, considerando que o SINAES constitui um sistema de medição de desempenho que estabelece diversos critérios através de indicadores previamente formulados.

O volume e a complexidade das informações geradas pelo SINAES demandam instrumentos tecnológicos mais sofisticados para coleta, processamento e análise de dados. A dependência de métodos manuais e ferramentas básicas limita drasticamente o potencial de descoberta de conhecimento e compromete a tempestividade das análises necessárias para a tomada de decisões estratégicas.

O estabelecimento de uma cultura avaliativa sólida constitui desafio permanente que demanda tempo, continuidade, informação e reflexão. Como observado por Dias Sobrinho (2003), a aculturação de novos hábitos não ocorre de forma automática nem imediata, exigindo esforços sistemáticos para impulsionar a consciência institucional e possibilitar uma visão crítica e participativa.

A superação deste desafio requer estratégias inovadoras que democratizem o acesso às informações avaliativas, simplifiquem a interpretação de dados complexos e facilitem a participação ativa da comunidade acadêmica nos processos reflexivos e propositivos.

2.2 Decisão Estratégica em Organizações Educacionais: Conceitos e Teorias

A compreensão do processo decisório em organizações constitui elemento central para a gestão estratégica de instituições educacionais. Herbert Simon (1955), laureado com o Prêmio Nobel de Economia em 1978, desenvolveu contribuições seminais para este campo ao introduzir o conceito de racionalidade limitada (*bounded rationality*), que fundamenta grande

parte das teorias contemporâneas sobre decisão organizacional. Conforme argumenta Simon (1996), a racionalidade global, pressuposta pela teoria econômica neoclássica, assume que o tomador de decisão possui uma função de utilidade abrangente e consistente, conhece todas as alternativas disponíveis, pode computar o valor esperado de utilidade associado a cada alternativa e escolhe aquela que maximiza a utilidade esperada.

Em contraposição, a racionalidade limitada reconhece que os agentes decisórios enfrentam restrições cognitivas, informacionais e temporais que inviabilizam a otimização global. Simon propôs o conceito de *satisficing*, uma combinação dos termos *satisfy* (satisfazer) e *suffice* (bastar), para descrever o comportamento real dos decisores, que buscam alternativas satisfatórias em vez de ótimas (SIMON, 1956). Esta perspectiva possui relevância particular para a gestão de IES, nas quais os gestores frequentemente precisam tomar decisões complexas com informações incompletas sobre múltiplas dimensões da qualidade educacional, desde indicadores de desempenho acadêmico até aspectos de infraestrutura e corpo docente.

A teoria de Simon sobre racionalidade limitada estabelece que os seres humanos, embora orientados por objetivos, frequentemente falham em alcançá-los devido à interação entre aspectos de suas arquiteturas cognitivas e a complexidade essencial do ambiente que enfrentam (SIMON, 1980). No contexto do SINAES, essa complexidade manifesta-se na multiplicidade de indicadores avaliativos, na heterogeneidade dos cursos e na necessidade de articular dimensões qualitativas e quantitativas na interpretação dos resultados. A Comissão Própria de Avaliação (CPA), responsável pela autoavaliação institucional, opera tipicamente sob condições de racionalidade limitada, necessitando processar grandes volumes de dados com recursos analíticos restritos.

Planellas e Muni (2019) destacam que a tomada de decisões estratégicas representa um desafio significativo para gestores, dada sua complexidade inerente e o impacto duradouro que podem exercer sobre a organização. Diferentemente das decisões operacionais, voltadas à gestão cotidiana e às respostas imediatas a demandas rotineiras, as decisões estratégicas requerem planejamento criterioso e visão prospectiva. No âmbito das IES, essa distinção manifesta-se na diferença entre ações corretivas pontuais, como ajustes em disciplinas específicas, e decisões que redefinem o projeto pedagógico de um curso ou a política institucional de qualificação docente em resposta aos resultados do ENADE.

Henry Mintzberg (1976) ampliou a compreensão dos processos de decisão estratégica ao identificar diferentes modos pelos quais as organizações formulam suas estratégias. Conforme o autor, existem três modos principais de tomada de decisão estratégica:

o modo empreendedor, caracterizado por decisões proativas e arriscadas conduzidas por um líder visionário; o modo adaptativo (ou incremental), marcado por soluções reativas aos problemas existentes; e o modo de planejamento, que envolve abordagem sistemática e analítica para a formulação estratégica. Esta tipologia oferece lentes interpretativas relevantes para compreender como diferentes IES abordam os desafios impostos pelo sistema avaliativo brasileiro.

Conforme Mações (2018), as decisões estratégicas são projetadas com horizonte temporal de médio a longo prazo, tipicamente abrangendo períodos de dois a cinco anos, e impactam a organização em sua totalidade. Essas decisões envolvem escolhas fundamentais quanto a produtos a serem desenvolvidos, mercados a serem explorados e orientações prioritárias entre crescimento e rentabilidade. Aplicando essa perspectiva ao contexto educacional, observa-se que decisões sobre abertura ou fechamento de cursos, redefinição de matrizes curriculares e investimentos em infraestrutura laboratorial constituem exemplos de escolhas estratégicas que condicionam a trajetória institucional e sua capacidade de resposta aos indicadores avaliativos do SINAES.

O modo de planejamento, conforme descrito por Mintzberg (1973), pressupõe que a alta gestão desenvolve uma ideia razoavelmente clara da missão e objetivos da organização e, no desenvolvimento de suas estratégias, opta por utilizar um processo iterativo no qual a organização sonda o futuro, experimenta e aprende a partir de uma série de compromissos parciais (incrementais), em vez de formulações globais de estratégias totais. Esta perspectiva alinha-se com a necessidade das IES de desenvolver planos de ação acadêmicas e administrativas em resposta aos resultados do SINAES, ajustando continuamente suas estratégias com base em evidências geradas pelo sistema avaliativo.

No âmbito da gestão universitária, Simon (1980) sugeriu que as organizações estabeleçam racionalidade procedimental, assegurando que processos formais sejam seguidos para coletar, analisar e utilizar informações relevantes e garantir deliberação apropriada antes que uma decisão seja tomada. Esta recomendação fundamenta a proposta de *frameworks* tecnológicos que auxiliem os gestores educacionais a processarem sistematicamente os dados avaliativos disponíveis.

A aplicação dessas teorias ao contexto do SINAES revela um desafio central: como superar as limitações da racionalidade humana para aproveitar plenamente o potencial informacional dos indicadores avaliativos? Os dados gerados pelo Exame Nacional de Desempenho dos Estudantes (ENADE), pelos relatórios de autoavaliação institucional e pelas

avaliações *in loco* constituem um repositório valioso; porém, sua complexidade frequentemente excede a capacidade analítica dos gestores. É nesse contexto que emergem as tecnologias de inteligência artificial como instrumentos potencialmente capazes de ampliar a racionalidade procedimental nas IES, tema desenvolvido na seção seguinte.

A formulação de decisões estratégicas opera tipicamente sob condições de informação limitada e incerteza considerável, restringindo a aplicabilidade de experiências prévias dos gestores (PLANELLAS; MUNI, 2019). Essa característica introduz elevado nível de risco e dificulta a reversão das decisões uma vez implementadas. No contexto do SINAES, a incerteza manifesta-se na impossibilidade de prever com precisão os resultados das avaliações futuras (FELDMANN *et al.*, 2016), nas mudanças metodológicas nos instrumentos avaliativos (IKUTA, 2016) e nas dinâmicas do corpo discente que influenciam o desempenho no ENADE.

A utilização de sistemas de inteligência artificial para análise de dados históricos e identificação de padrões pode contribuir para reduzir, ainda que parcialmente, essa incerteza inerente ao processo decisório estratégico em IES.

2.2.1 Teorias da Decisão Estratégica e seus Condicionantes nas Instituições de Educação Superior

A compreensão dos fundamentos teóricos que sustentam a tomada de decisão estratégica constitui requisito essencial para o desenvolvimento de instrumentos de apoio à gestão educacional. No contexto das Instituições de Educação Superior (IES) brasileiras, submetidas às diretrizes avaliativas do SINAES, a aplicação dessas teorias adquire contornos específicos, dado que os gestores educacionais operam em um ambiente regulado, com múltiplas fontes de dados institucionais e pressões simultâneas de natureza acadêmica, administrativa e regulatória.

2.2.2 A Teoria da Racionalidade Limitada e suas Implicações para a Gestão Educacional

A Teoria da Decisão Racional, em sua formulação clássica, pressupõe que os tomadores de decisão possuem acesso completo às informações relevantes, capacidade ilimitada de processamento cognitivo e preferências estáveis e bem definidas. Nessa perspectiva idealizada, a decisão estratégica resultaria de um processo sistemático de

identificação de alternativas, avaliação de consequências e seleção da opção que maximiza os objetivos organizacionais.

Contudo, os trabalhos seminais de Herbert Simon (1955, 1980) demonstraram que a racionalidade humana opera sob restrições cognitivas significativas. O conceito de *bounded rationality* (racionalidade limitada) reconhece que os decisores não otimizam, mas satisfazem. Ou seja, selecionam alternativas que atendem a critérios mínimos aceitáveis, dadas as limitações de tempo, informação e capacidade de processamento. Essa constatação possui implicações diretas para a gestão de IES: coordenadores de curso e gestores acadêmicos frequentemente tomam decisões com base em informações fragmentadas, sob pressão temporal e com recursos analíticos limitados (PEREIRA; GASQUE, 2019).

No âmbito do SINAES, a racionalidade limitada manifesta-se quando gestores precisam interpretar indicadores complexos, como o Conceito Preliminar de Curso (CPC) ou os resultados do ENADE, sem dispor de ferramentas adequadas para análise multidimensional. Conforme argumenta Bazerman (2014), vieses como o excesso de confiança e a ilusão de controle podem levar decisores a subestimar riscos ou superestimar sua capacidade de prever resultados. No contexto educacional, isso pode traduzir-se em planos de melhoria baseados em diagnósticos superficiais ou em interpretações enviesadas dos dados avaliativos.

A proposição de um *framework* baseado em agentes inteligentes, conforme delineado nesta pesquisa, alinha-se precisamente à necessidade de compensar as limitações da racionalidade humana. Ao automatizar processos de descoberta de conhecimento em bases de dados institucionais, o *framework* pode ampliar a capacidade analítica dos gestores, oferecendo sínteses estruturadas que reduzem a carga cognitiva e minimizam a influência de vieses individuais na interpretação dos indicadores do SINAES.

2.2.3 A Perspectiva Comportamental: Heurísticas e Vieses no Processo Decisório

A Teoria da Decisão Comportamental, consolidada a partir dos trabalhos de Kahneman e Tversky (1974), evidenciou que o julgamento humano recorre sistematicamente a atalhos mentais, denominados heurísticas, que, embora funcionais em muitos contextos, podem produzir erros sistemáticos e previsíveis. Três heurísticas fundamentais foram identificadas: disponibilidade, representatividade e ancoragem.

A heurística da disponibilidade leva os decisores a superestimar a probabilidade de eventos que são facilmente recordados, seja por sua frequência recente, seja por seu impacto

emocional. Em uma IES, um coordenador de curso pode, por exemplo, atribuir peso desproporcional a reclamações recentes de estudantes, negligenciando tendências estruturais reveladas pelos dados históricos do ENADE. A heurística da representatividade, por sua vez, induz julgamentos baseados em estereótipos ou semelhanças superficiais, desconsiderando informações estatísticas relevantes. Já a ancoragem demonstra como estimativas iniciais, mesmo arbitrárias, influenciam julgamentos subsequentes, dificultando ajustes adequados diante de novas evidências.

Nesse sentido, o desenvolvimento de relatórios informativos inteligentes, um dos objetivos específicos do *framework* proposto, pode funcionar como um mecanismo para reduzir ou eliminar vieses cognitivos para melhorar a precisão e a racionalidade na tomada de decisões, apresentando informações de forma estruturada e contextualizada, de modo a facilitar o engajamento do pensamento analítico. Ao traduzir dados brutos do SINAES em visualizações interpretáveis e recomendações fundamentadas, os agentes inteligentes podem atenuar a influência de heurísticas disfuncionais no processo decisório.

2.3 Descoberta de Conhecimento

A literatura especializada distingue duas metodologias principais para a descoberta de conhecimento, que se diferenciam pela natureza dos dados processados: o *Knowledge Discovery in Databases* (KDD) e o *Knowledge Discovery from Text* (KDT). Ambas apresentam processos específicos de trabalho, dependentes do modo como os dados estão organizados (SILVA; GHIRALDELLI, 2022).

2.3.1 Knowledge Discovery in Databases (KDD)

O KDD refere-se à descoberta de conhecimento em dados estruturados, organizados em bancos de dados relacionais ou sistemas de informação que permitem a identificação de relacionamentos entre variáveis. Conforme Silva (2020, p. 164), "no KDD os dados estão organizados em um banco de dados, armazenados de modo a possibilitar a identificação de relacionamentos entre palavras ou conjunto de palavras". Os dados, nesse contexto, não se limitam a textos, podendo estar organizados em tabulações, listas ou tuplas.

É fundamental distinguir entre KDD e Mineração de Dados, conceitos

frequentemente confundidos como sinônimos. O KDD constitui um processo completo de descoberta de conhecimento, enquanto a Mineração de Dados representa sua etapa principal, na qual são aplicados algoritmos específicos para produzir padrões sobre bases de dados. O KDD descreve todo o processo de extração de conhecimento em bases de dados, sendo a Mineração de Dados a etapa central deste processo mais amplo.

No contexto da avaliação institucional, o KDD apresenta aplicabilidade direta na análise de dados estruturados provenientes de sistemas acadêmicos, como históricos de desempenho discente, taxas de evasão, indicadores de titulação docente e resultados do ENADE. A metodologia permite identificar correlações entre variáveis que podem orientar intervenções pedagógicas e administrativas.

2.3.2 Knowledge Discovery from Text (KDT)

O KDT destina-se à descoberta de conhecimento em dados não estruturados, predominantemente textuais. Silva (2020, p. 166) define essa metodologia como:

o processo de construir a informação e utilizá-la a partir de dados não estruturados em textos, considerando não somente o levantamento quantitativo de palavras, mas também associação de palavras, semântica, contexto dentro de uma base de linguagem natural.

Ambrósio e Moraes (2007, p. 6) destacam que as principais contribuições dessa área se relacionam à:

[...] busca de informações específicas em documentos, à análise qualitativa e quantitativa de grandes volumes de textos, e à melhor compreensão de textos disponíveis em documentos. Textos estes que podem estar representados das mais diversas formas, dentre elas: e-mails; arquivos em diferentes formatos (pdf, doc, txt, por exemplo); páginas Web; campos textuais em bancos de dados; textos eletrônicos digitalizados a partir de papéis.

Para a gestão educacional, o KDT revela-se particularmente útil na análise de relatórios de autoavaliação institucional, respostas abertas de questionários aplicados pela CPA, atas de reuniões de colegiados e documentos normativos. A metodologia possibilita a identificação de temas recorrentes, percepções predominantes e demandas não explicitadas em indicadores quantitativos.

2.4 KDD e KDT como Referencial Contextualizador

No contexto desta pesquisa, o KDD e o KDT oferecem referencial conceitual útil para situar a problemática da descoberta de conhecimento a partir de dados avaliativos das IES. Contudo, a abordagem adotada para construção das matrizes analíticas difere fundamentalmente do KDD clássico: trata-se de Análise de Conteúdo categorial temática (BARDIN, 2016), procedimento qualitativo e interpretativo que opera sobre documentos textuais normativos, não sobre bases de dados numéricas. Essa distinção é relevante para delimitar o escopo metodológico do trabalho, detalhado no Capítulo 4.

2.5 Análise de Conteúdo: Fundamentos Teórico-Metodológicos

2.5.1 Conceituação e Evolução Histórica

A Análise de Conteúdo constitui um conjunto de técnicas de análise das comunicações que visa obter, por procedimentos sistemáticos e objetivos de descrição do conteúdo das mensagens, indicadores, quantitativos ou qualitativos, que permitam a inferência de conhecimentos relativos às condições de produção e recepção dessas mensagens (BARDIN, 1977). Esta definição evidencia três características fundamentais da técnica: sua natureza sistemática (procedimentos explícitos e replicáveis), sua orientação objetiva (descrição baseada em critérios definidos) e sua finalidade inferencial (extração de conhecimentos que transcendem o conteúdo manifesto).

Krippendorff (2018) amplia a conceituação ao definir Análise de Conteúdo como técnica de pesquisa para fazer inferências replicáveis e válidas a partir de textos (ou outra matéria significativa) para os contextos de seu uso. Essa definição enfatiza dois critérios de qualidade, replicabilidade e validade, que distinguem a Análise de Conteúdo de interpretações impressionistas ou idiossincráticas de textos.

No contexto da gestão educacional, a Análise de Conteúdo apresenta aplicabilidade significativa para o tratamento de documentos normativos, relatórios institucionais e instrumentos de avaliação. Os critérios de atribuição de conceitos presentes no Instrumento de Avaliação de Cursos de Graduação do INEP, por exemplo, constituem corpus textual que demanda análise sistemática para extração de elementos operacionalizáveis em sistemas de

apoio à decisão.

2.5.2 Modalidades de Análise de Conteúdo

A literatura especializada distingue diferentes modalidades de Análise de Conteúdo, que variam conforme os objetivos da investigação e a natureza do material analisado. Bardin (2016) apresenta três modalidades principais:

Análise Categórica: Modalidade utilizada, baseada em operações de desmembramento do texto em unidades e posterior reagrupamento em categorias segundo critérios definidos. As categorias podem ser estabelecidas *a priori*, derivadas de teoria ou de hipóteses de pesquisa, ou emergir a posteriori, do próprio material analisado. No contexto da análise dos indicadores do SINAES, categorias como "elementos constitutivos", "atributos de progressão" e "evidências requeridas" exemplificam categorização *a priori* orientada pelos objetivos do *framework* proposto.

Análise de Avaliação: Foca nas atitudes e opiniões expressas no material, medindo a direção (positiva, negativa, neutra) e a intensidade dos juízos emitidos. Esta modalidade encontra aplicação na análise de respostas discursivas de questionários de autoavaliação institucional.

Análise de Enunciação: Considera a comunicação como processo dinâmico, analisando não apenas o conteúdo manifesto, mas também as condições de produção do discurso. Esta modalidade é particularmente útil para análise de documentos institucionais como Projetos Pedagógicos de Curso (PPC) e Planos de Desenvolvimento Institucional (PDI).

2.5.3 Fases da Análise de Conteúdo

Bardin (2016) organiza o processo de Análise de Conteúdo em três fases cronológicas que estruturam a aplicação da técnica:

Pré-análise: Esta fase corresponde ao período de organização, que tem por objetivo operacionalizar e sistematizar as ideias iniciais. Compreende atividades de leitura flutuante (primeiro contato com os documentos), escolha dos documentos (delimitação do corpus), formulação de hipóteses e objetivos, referenciação de índices e elaboração de indicadores, e preparação do material.

A leitura flutuante consiste em estabelecer contato com os documentos a analisar e conhecer o texto, deixando-se invadir por impressões e orientações. No caso da análise do Instrumento de Avaliação de Cursos do INEP, esta etapa envolve leitura integral do documento, identificando sua estrutura organizativa (dimensões, indicadores, conceitos) e as características estilísticas dos critérios de avaliação.

Exploração do Material: Fase de análise propriamente dita, consiste essencialmente em operações de codificação, decomposição ou enumeração, em função de regras previamente formuladas. A codificação corresponde a uma transformação dos dados brutos do texto, mediante recorte, agregação e enumeração, que permite atingir representação do conteúdo susceptível de esclarecer o analista acerca das características do texto.

O recorte define a unidade de registro, o segmento de conteúdo a considerar como unidade de base, e a unidade de contexto, o segmento mais amplo que fornece significado à unidade de registro. Na análise dos indicadores do SINAES, a unidade de registro pode ser definida como cada critério de conceito (1 a 5) de cada indicador, enquanto a unidade de contexto corresponde ao indicador completo com todos os seus critérios.

Tratamento dos Resultados, Inferência e Interpretação: Os resultados brutos são tratados de maneira a tornarem-se significativos e válidos. Operações estatísticas simples (frequências) ou mais complexas permitem estabelecer quadros de resultados, diagramas, figuras e modelos que condensam e põem em relevo as informações fornecidas pela análise. A partir dos resultados, o analista pode propor inferências e adiantar interpretações a propósito dos objetivos previstos ou relativas a outras descobertas inesperadas.

2.5.4 Critérios de Qualidade

A validade e confiabilidade da Análise de Conteúdo dependem da observância de critérios que assegurem rigor metodológico. Krippendorff (2018) sistematiza os principais critérios:

Exaustividade: O conjunto de categorias deve permitir a classificação de todo o material relevante, sem resíduos não categorizáveis. Este critério implica que a Matriz de Progressão Conceitual deve contemplar todos os elementos presentes nos critérios de avaliação do INEP.

Exclusividade: Cada unidade de registro deve ser classificável em uma única categoria, evitando ambiguidades. No contexto dos indicadores do SINAES, os elementos

constitutivos, atributos de progressão e evidências requeridas devem ser claramente distinguíveis.

Homogeneidade: As categorias devem ser construídas com base em princípio único de classificação, mantendo coerência interna. Categorias que misturam critérios distintos comprometem a validade da análise.

Pertinência: As categorias devem ser adequadas ao material analisado e aos objetivos da pesquisa. No *framework* proposto, as categorias derivam diretamente da estrutura dos indicadores do SINAES e da necessidade de operacionalização para geração de recomendações.

Objetividade: As regras de codificação devem ser suficientemente claras para permitir que diferentes codificadores, trabalhando independentemente, obtenham resultados consistentes. Este critério fundamenta a replicabilidade da análise.

2.5.5 Aplicação ao Contexto do SINAES

A aplicação da Análise de Conteúdo aos instrumentos avaliativos do SINAES demanda adaptações metodológicas que considerem as especificidades do corpus. Os critérios de atribuição de conceitos presentes no Instrumento de Avaliação de Cursos de Graduação apresentam uma estrutura textual padronizada que facilita a análise categorial:

Cada indicador é descrito por cinco critérios correspondentes aos conceitos 1 a 5, organizados em progressão qualitativa. O conceito 1 tipicamente descreve uma situação de ausência ou não atendimento do requisito avaliado. Os conceitos 2 a 5 descrevem graus progressivos de atendimento, incorporando elementos adicionais ou qualificadores que distinguem cada nível.

Esta estrutura sugere uma estratégia analítica baseada em três categorias principais:

Elementos Constitutivos: Identificação dos componentes obrigatórios mencionados em todos os níveis conceituais, que definem "o que" está sendo avaliado. Estes elementos permanecem constantes ao longo da progressão conceitual.

Atributos de Progressão: Extração dos qualificadores que diferenciam cada nível conceitual, representando "como" a qualidade é graduada. Atributos típicos incluem termos como "de maneira limitada", "claramente voltadas para", "práticas comprovadamente exitosas ou inovadoras".

Evidências Requeridas: Inferência da documentação ou demonstração necessária

para comprovar o atendimento ao critério em cada nível. Embora nem sempre explicitadas nos textos, as evidências podem ser deduzidas a partir dos elementos constitutivos e atributos de progressão.

A aplicação sistemática desta estratégia ao conjunto dos indicadores do instrumento de avaliação produz a Matriz de Progressão Conceitual, artefato central do *framework* proposto. Esta matriz estrutura o conhecimento normativo de forma que possibilita: (a) identificação automatizada do gap entre conceito atual e conceito desejado; (b) especificação dos requisitos adicionais para progressão; (c) geração de recomendações contextualizadas pelos agentes de IA.

2.6 Inteligência Artificial na Gestão Educacional

A revolução digital contemporânea tem transformado fundamentalmente os paradigmas organizacionais e gerenciais em diversos setores, sendo a educação superior uma das áreas que mais pode beneficiar-se da aplicação estratégica de tecnologias emergentes. A Inteligência Artificial (IA), em particular, representa oportunidade transformadora para otimizar processos decisórios, automatizar análises complexas e gerar insights valiosos que orientem gestores educacionais na condução mais eficaz de suas instituições.

2.6.1 O Papel da Inteligência Artificial na Extração de Conhecimento

A operacionalização das metodologias KDD e KDT depende fundamentalmente de técnicas de inteligência artificial (IA), especialmente no que concerne ao processamento de grandes volumes de dados e à identificação de padrões não triviais. "A Inteligência Artificial envolve utilizar métodos baseados no comportamento inteligente de humanos e outros animais para solucionar problemas complexos" (SILVA; GHIRALDELLI, 2022).

Artero (2009, p. 19) complementa essa definição ao explicar que:

[...] as definições de IA seguem na mesma linha que as definições para a IN [Inteligência Natural], simplesmente, adaptando-se para o computador e, em resumo, sugerem que a IA é o estudo de como fazer os computadores realizarem tarefas que, no momento, as pessoas fazem melhor, o que inclui a capacidade de adquirir e aplicar conhecimentos aprendidos.

Russell e Norvig (2013) propõem uma categorização das definições de inteligência artificial em quatro campos: sistemas que pensam como seres humanos; sistemas que pensam racionalmente; sistemas que atuam como seres humanos; e sistemas que atuam racionalmente. Essa multiplicidade de abordagens reflete a complexidade epistemológica do campo, que envolve contribuições de diversas ciências.

2.6.2 Inteligência Artificial Fraca e Forte

A literatura distingue dois paradigmas fundamentais no campo da inteligência artificial: a IA fraca e a IA forte. A IA fraca caracteriza-se pelo desenvolvimento de sistemas especialistas que utilizam teorias cognitivas e ciência da computação para prover automações específicas, em contextos em que algoritmos tradicionais seriam excessivamente complexos ou inviáveis. A IA forte, por sua vez, constitui uma concepção hipotética de inteligência artificial capaz de realizar todas as operações que um ser humano poderia executar (SILVA; GHIRALDELLI, 2022).

Conforme Artero (2009, p. 16):

No primeiro caso, acredita-se que as máquinas serão realmente capazes de pensar como uma pessoa, apresentando todas as características de consciência que os humanos possuem, enquanto o segundo grupo acredita apenas na possibilidade de as máquinas simularem o comportamento humano, mas sem a real consciência de suas ações.

O Teste de Turing, proposto por Alan Turing em 1950, constitui referência clássica para a distinção entre IA fraca e forte. Russell e Norvig (2013, p. 4) explicam que:

O teste de Turing [...] foi projetado para fornecer uma definição operacional satisfatória de inteligência. Em vez de propor uma lista longa e talvez controversa de qualificações exigidas para inteligência, ele sugeriu um teste baseado na impossibilidade de distinguir entre entidades inegavelmente inteligentes - os seres humanos.

Para os propósitos da gestão educacional e da avaliação institucional, as técnicas de IA fraca apresentam aplicabilidade imediata e resultados comprovados. Sistemas especialistas

podem processar dados do SINAES, identificar padrões de desempenho e sugerir correlações relevantes para a tomada de decisão, ainda que a interpretação final e a formulação de estratégias permaneçam sob responsabilidade dos gestores.

2.6.3 Redes Neurais Artificiais

Entre as técnicas de inteligência artificial aplicadas à descoberta de conhecimento, as redes neurais artificiais destacam-se pela capacidade de aprendizado e adaptação. Conforme Silva (2020), essas redes baseiam-se no modelo biológico neuronal desenvolvido por McCulloch e Pitts na década de 1940, constituindo modelos matemáticos que emulam o funcionamento do sistema nervoso.

Braga, Ludermir e Carvalho (2000, p. 16) descrevem o processo de treinamento de redes supervisionadas:

O objetivo é ajustar os parâmetros da rede, de forma a encontrar uma ligação entre os pares de entrada e saída fornecidos. [...] O professor indica explicitamente um comportamento bom ou ruim para a rede, visando a direcionar o processo de treinamento. A rede tem sua saída corrente (calculada) comparada com a saída desejada, recebendo informações do supervisor sobre o erro da resposta atual.

As redes neurais de aprendizagem por reforço, especificamente, apresentam processo de treinamento análogo à aprendizagem humana por repetição, onde a medida de desempenho baseia-se na verificação da correção das saídas mediante tentativa e erro (SILVA; GHIRALDELLI, 2022). Esse mecanismo possibilita que os sistemas se adaptem progressivamente aos padrões específicos dos dados institucionais, refinando sua capacidade de identificação de correlações relevantes.

2.6.4 IA Aplicada à Avaliação Institucional

A aplicação de Inteligência Artificial na gestão educacional representa evolução natural dos processos de descoberta de conhecimento discutidos anteriormente, oferecendo capacidades avançadas para processamento automático de informações, identificação de padrões complexos e geração de recomendações baseadas em evidências.

A integração de algoritmos de aprendizado de máquina, raciocínio baseado em casos e sistemas especialistas aos processos avaliativos permite desenvolver soluções que vão além da mera automatização de tarefas, proporcionando análise preditiva, identificação de tendências, detecção de anomalias e descoberta de oportunidades de melhoria que poderiam permanecer ocultas em análises convencionais.

Os algoritmos de *Machine Learning* constituem uma ferramenta fundamental para implementação de soluções inteligentes na gestão educacional. Conforme evidenciado por Mitchell (1997), o aprendizado de máquina permite que sistemas computacionais aprimorem automaticamente seu desempenho por meio da experiência, característica particularmente valiosa para análise de dados educacionais que apresentam padrões dinâmicos e evolutivos.

Algoritmos de Classificação Supervisionada: Técnicas como árvores de decisão (C4.5), redes neurais artificiais e máquinas de vetores de suporte podem ser empregadas para categorizar automaticamente estudantes, cursos ou instituições segundo critérios de qualidade predefinidos. Por exemplo, algoritmos de classificação podem identificar estudantes em risco de evasão com base em variáveis como frequência, desempenho acadêmico, situação socioeconômica e engajamento em atividades extracurriculares.

Algoritmos de Clustering: Técnicas de agrupamento não supervisionado permitem descobrir perfis naturais de discentes, docentes ou cursos com características similares, facilitando o desenvolvimento de estratégias diferenciadas de gestão. A identificação de clusters pode revelar segmentos específicos que demandam atenção particular ou que apresentam potencial de excelência a ser explorado.

Algoritmos de Associação: Regras de associação podem identificar relações não óbvias entre diferentes aspectos da vida acadêmica, como correlações entre métodos pedagógicos específicos e resultados de aprendizagem, ou entre infraestrutura disponível e satisfação estudantil.

Os sistemas especialistas constituem uma abordagem clássica da IA que pode ser particularmente útil para codificar conhecimento especializado sobre avaliação institucional e critérios de qualidade educacionais. Estes sistemas podem incorporar regras derivadas da experiência de avaliadores especialistas, diretrizes do SINAES e melhores práticas internacionais, proporcionando assistência inteligente para interpretação de resultados avaliativos e formulação de recomendações de melhoria.

A construção de bases de conhecimento especializadas permite sistematizar expertise acumulada por avaliadores experientes, democratizando acesso a análises sofisticadas

e garantindo consistência na interpretação de dados avaliativos. Além disso, sistemas especialistas podem ser atualizados continuamente com novo conhecimento, aprimorando progressivamente sua capacidade de fornecer orientações precisas e contextualizadas.

Uma das aplicações mais promissoras da IA na gestão educacional refere-se ao desenvolvimento de modelos preditivos capazes de antecipar problemas antes que se manifestem plenamente. Modelos que predizem desempenho de cursos em avaliações externas baseados em indicadores internos permitem ações proativas de melhoria, enquanto sistemas que identificam precocemente estudantes em risco de evasão possibilitam intervenções preventivas mais eficazes. Modelos que antecipem resultados prováveis baseados em dados internos contínuos permitem às instituições ajustarem estratégias e implementar melhorias antes das avaliações oficiais, otimizando resultados e reduzindo riscos de avaliações insatisfatórias.

2.6.5 Limitações das Abordagens Tradicionais para Processamento de Documentos Normativos

A aplicação de técnicas de aprendizado de máquina ao domínio educacional demonstra eficácia para predição de desempenho estudantil, identificação de perfis de evasão e análise de padrões em *logs* de sistemas de gestão de aprendizagem (BAKER; YACEF, 2009; ROMERO; VENTURA, 2020). Contudo, sua aplicação ao problema específico desta pesquisa apresenta limitações estruturais: os critérios do SINAES para distinção entre os conceitos 3 e 4 em determinado indicador não são reduzíveis a variáveis numéricas facilmente processáveis por modelos tradicionais de *machine learning*, demandando interpretação semântica contextualizada. Essa limitação justifica a opção por *Large Language Models* (LLMs) como núcleo do componente de processamento do *framework*.

2.7 Inteligência Artificial Generativa: Aspectos Avançados para Aplicação em Sistemas de Apoio à Decisão

A inteligência artificial generativa (IAG) representa uma evolução significativa no campo da computação, com implicações profundas para os processos decisórios organizacionais. Diferentemente dos sistemas tradicionais de IA, que operam primariamente em tarefas de classificação e predição, os modelos generativos são capazes de criar conteúdo,

textos, imagens, códigos e análises a partir de padrões aprendidos em grandes volumes de dados. O reconhecimento da relevância científica dessa tecnologia foi evidenciado pela concessão de dois Prêmios Nobel em 2024: em Física, para Geoffrey Hinton e John Hopfield, pelos fundamentos algorítmicos da IA, e em Química, para pesquisadores que aplicaram IA generativa na predição de estruturas proteicas (GENRE, 2025).

A definição precisa de inteligência artificial generativa permanece em construção na literatura, dada a complexidade e a rápida evolução do campo. De modo geral, pode ser descrita como uma subcategoria dentro do amplo campo da inteligência artificial, caracterizada pela capacidade de gerar novos conteúdos que não estavam explicitamente presentes nos dados de treinamento. Os grandes modelos de linguagem *Large Language Models* (LLMs), como GPT-4, Claude e Gemini, exemplificam essa categoria tecnológica, demonstrando capacidades avançadas de compreensão e geração de linguagem natural que expandem significativamente as possibilidades de interação humano-computador no contexto organizacional.

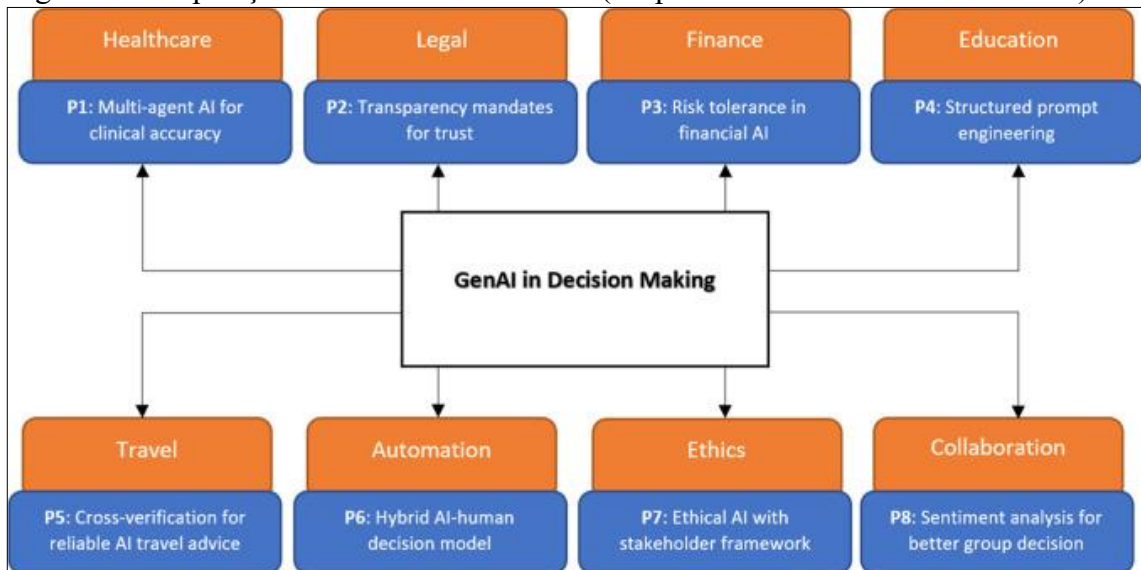
A adoção de IAG nas organizações tem apresentado crescimento acelerado. Conforme levantamento da McKinsey (2024), 65% dos respondentes em pesquisa global relataram que suas organizações utilizavam regularmente IA generativa em pelo menos uma função de negócio, representando quase o dobro do percentual registrado dez meses antes. No contexto educacional, esse movimento também é perceptível: pesquisa da Associação Brasileira de Mantenedoras de Ensino Superior (ABMES) indicou que 71% dos estudantes universitários brasileiros utilizam ferramentas de IA frequentemente em suas rotinas de estudo. Esses dados evidenciam uma transformação em curso que demanda resposta estratégica das instituições educacionais.

A aplicação da IAG ao processo decisório organizacional fundamenta-se em suas capacidades de processamento e síntese de informações complexas. Em ambiente de negócios caracterizado por crescente complexidade, a IAG possibilita a análise de grandes volumes de dados, a identificação de padrões complexos, a formulação de recomendações baseadas em evidências e a facilitação de decisões mais informadas e tempestivas. No contexto das IES, essas capacidades podem ser direcionadas para a análise de indicadores do SINAES, a identificação de pontos críticos nos relatórios de avaliação e a geração de recomendações para planos de melhoria.

A literatura recente sobre IAG destaca que, apesar das capacidades avançadas dessas tecnologias para analisar dados e gerar padrões, o julgamento humano permanece crítico em situações de alta incerteza (LÓPEZ *et al.*, 2025). Esta constatação alinha-se com o conceito

de racionalidade limitada de Simon: a IAG pode ampliar significativamente a capacidade de processamento informacional, mas não elimina a necessidade de julgamento contextualizado pelos gestores. No âmbito do SINAES, isso implica que os sistemas de IA devem ser concebidos como ferramentas de apoio à decisão, e não como substitutos do discernimento profissional dos avaliadores e gestores educacionais.

Figura 1 – Proposições em diversos domínios (Propositions across various domains).



Fonte: Adaptado de ALBASHRAWI (2025).

A Figura 1 sintetiza oito proposições derivadas da análise multidisciplinar conduzida por Albashrawi (2025), cada qual endereçando um domínio específico de aplicação da IAG ao processo decisório. No domínio da saúde (P1), propõe-se que a integração de *frameworks* multiagente pode aprimorar a confiabilidade diagnóstica ao mitigar vieses cognitivos inerentes a modelos individuais. No domínio jurídico (P2), sustenta-se que a implementação de mandatos de transparência, como a exigência de trilhas de auditoria e explicitação das fontes utilizadas pela IA, pode fortalecer a confiança pública em decisões judiciais assistidas por IAG. Para o domínio de negócios e finanças (P3), a proposição indica que sistemas de assessoria financeira baseados em IAG devem incorporar modelos de tolerância ao risco específicos para cada perfil de usuário, promovendo personalização e reduzindo inconsistências nas recomendações. No domínio do turismo (P5), recomenda-se que recomendações de viagem geradas por IAG sejam validadas mediante cruzamento com fontes verificadas, como órgãos oficiais de turismo e dados em tempo real, a fim de mitigar riscos de desinformação. A proposição relativa a sistemas tecnológicos e automação (P6) defende modelos híbridos de decisão que combinem a eficiência computacional da IAG com a

adaptabilidade do julgamento humano em ambientes dinâmicos. No domínio ético (P7), propõe-se a incorporação de *frameworks* alinhados a partes interessadas para assegurar equidade e prestação de contas nas decisões assistidas por IAG. Por fim, no contexto de decisões colaborativas e multidisciplinares (P8), a análise de sentimentos potencializada por IAG é apresentada como mecanismo para mitigar hostilidade e fomentar discussões construtivas em processos deliberativos coletivos (ALBASHRAWI, 2025).

Dentre as proposições formuladas, a P4, referente ao domínio educacional, apresenta particular relevância para o presente trabalho. Albashrawi (2025) argumenta que a eficácia da IAG na educação depende fundamentalmente da forma como os *prompts* são estruturados e como o conteúdo gerado é interpretado pelos atores educacionais. O autor observa que *prompts* inadequadamente formulados podem conduzir a respostas enviesadas, incompletas ou pedagogicamente imprecisas, comprometendo a qualidade do apoio decisório oferecido. A proposição sustenta, portanto, que a tomada de decisão educacional assistida por IAG deve ser orientada por *frameworks* estruturados de engenharia de *prompt*, os quais incluam formatos padronizados de consulta, verificação cruzada das respostas com fontes validadas e integração de mecanismos de *feedback* em sistemas de gestão da aprendizagem. Essa estruturação, segundo o autor, não apenas aprimora a acurácia e a equidade das saídas geradas, mas também possibilita a adaptação da IAG a necessidades de aprendizagem diversificadas. A revisão contínua do conteúdo gerado por educadores é destacada como condição necessária para garantir alinhamento com padrões curriculares e prevenir desinformação no contexto instrucional (ALBASHRAWI, 2025).

A proposição P4 encontra ressonância direta no escopo do presente trabalho. O *framework* proposto nesta dissertação situa-se precisamente na interseção entre IAG e tomada de decisão em contexto educacional regulado, ao buscar apoiar coordenadores de curso e membros de Comissões Próprias de Avaliação na análise de indicadores do SINAES. A diretriz de Albashrawi (2025), de que sistemas educacionais assistidos por IAG devem operar sob *frameworks* estruturados que garantam acurácia, verificação cruzada e supervisão humana, constitui princípio orientador da solução aqui desenvolvida. Assim como a proposição P4 advoga pela padronização das interações com modelos de linguagem no domínio instrucional, o *framework* proposto estrutura as consultas aos agentes de IA a partir de bases de conhecimento normativas, assegurando que as saídas produzidas estejam ancoradas nos critérios oficiais de avaliação e sujeitas à validação pelos gestores institucionais. Essa convergência reforça a pertinência da abordagem adotada e posiciona o artefato como contribuição aplicada à agenda

de pesquisa delineada pelo autor, estendendo seus princípios do contexto instrucional para o domínio da gestão da qualidade no ensino superior brasileiro.

No contexto específico da gestão de IES brasileiras, a IAG oferece possibilidades relevantes para enfrentar o que pode ser denominado “Dilema do Conceito Satisfatório”: situação na qual cursos que obtêm conceitos intermediários nas avaliações do SINAES enfrentam o desafio estratégico de mobilizar esforços de melhoria sem a pressão imediata de conceitos insatisfatórios. A capacidade da IAG de processar padrões em dados históricos de avaliações, identificar tendências e gerar análises comparativas pode subsidiar a identificação de oportunidades de aprimoramento mesmo em contextos de desempenho aparentemente estável.

A Estratégia Brasileira de Inteligência Artificial (EBIA), instituída pela Portaria MCTI nº 4.617 de 2021, reconhece a educação como um dos eixos prioritários para o desenvolvimento da IA no país. Conforme Rotta, Silva e Costa (2025), a EBIA desempenha um papel central na condução de ações governamentais favoráveis ao desenvolvimento de soluções em inteligência artificial, com implicações diretas para a gestão universitária. Sua aplicação no âmbito da avaliação institucional representa oportunidade de alinhar as práticas das IES com as diretrizes nacionais de inovação tecnológica, contribuindo simultaneamente para a melhoria da qualidade educacional e para o desenvolvimento do ecossistema de IA brasileiro.

Contudo, a implementação de IAG na gestão educacional demanda atenção a desafios específicos. A possibilidade de geração de informações imprecisas ou enviesadas, fenômeno denominado “alucinação” nos modelos generativos, requer mecanismos de validação e supervisão humana. Adicionalmente, questões de privacidade dos dados dos estudantes, transparência algorítmica e equidade nos resultados gerados constituem preocupações éticas que devem orientar o desenho de soluções baseadas em IAG para o contexto do SINAES. A seção seguinte apresenta a arquitetura de agentes inteligentes como abordagem para endereçar esses desafios mediante sistemas estruturados de processamento e validação.

2.7.1 Agentes de Inteligência Artificial: Arquiteturas e Aplicações Educacionais

Os agentes de inteligência artificial representam uma evolução arquitetural que transcende as capacidades dos modelos generativos tradicionais, configurando sistemas autônomos capazes de perceber seu ambiente, tomar decisões e executar ações para atingir objetivos específicos. De acordo com a definição da BELCIC e STRYKER (2025), um agente

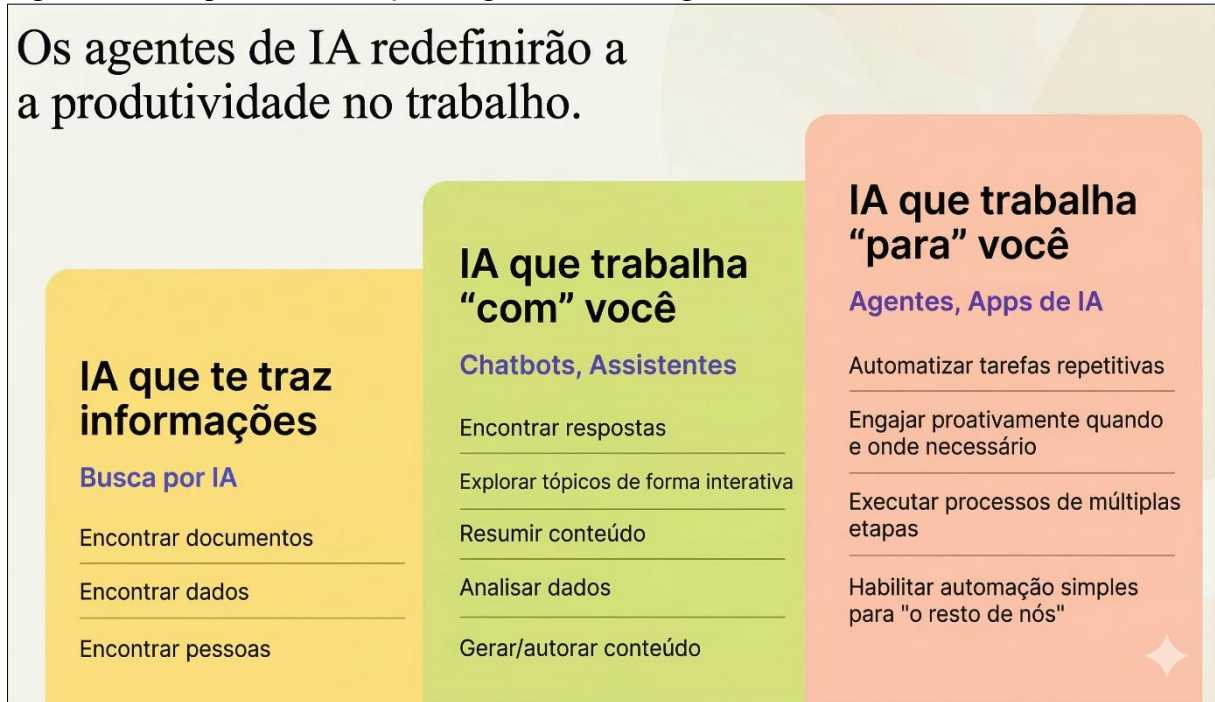
de IA é um sistema que executa tarefas autonomamente, projetando fluxos de trabalho com as ferramentas disponíveis. Franklin e Graesser (1997) oferecem uma definição mais abrangente: um agente autônomo é um sistema situado em um ambiente, que percebe esse ambiente e age sobre ele, ao longo do tempo, em busca de sua própria agenda e de modo a afetar o que percebe no futuro.

A distinção entre agentes de IA e sistemas de IA convencionais reside fundamentalmente na capacidade de agêntica, a habilidade de agir de forma autônoma e orientada a objetivos. Enquanto os LLMs tradicionais geram respostas com base em *prompts* específicos, os agentes podem decompor tarefas complexas em subtarefas, selecionar ferramentas apropriadas, executar ações sequenciais e adaptar seu comportamento com base nos resultados obtidos (AWS, 2025). Esta característica confere aos agentes um potencial diferenciado para aplicações em contextos de gestão, nos quais decisões frequentemente envolvem múltiplas etapas e fontes de informação.

A arquitetura de agentes inteligentes é composta por componentes funcionais interrelacionados que habilitam seu comportamento autônomo. De acordo com a síntese da literatura recente (GOOGLE CLOUD, 2025), os elementos centrais incluem: raciocínio (*reasoning*), que envolve o uso de lógica e informações disponíveis para tirar conclusões e resolver problemas; planejamento (*planning*), a capacidade de decompor objetivos complexos em subtarefas gerenciáveis; memória, que permite a retenção de informações de curto e longo prazo para um comportamento consistente; e uso de ferramentas (*tool use*), a habilidade de empregar recursos externos para estender funcionalidades (BANDI *et al.*, 2025). A integração desses componentes permite que agentes operem em cenários dinâmicos com níveis crescentes de autonomia.

O conceito de IA agêntica (*agentic AI*) emerge na literatura recente para descrever sistemas que combinam múltiplos tipos de inteligência artificial, habilitando capacidades de planejar, agir, aprender e melhorar (AEROSPIKE, 2025).

Figura 2 – Compartimentalização do genAI 1.0 e do genAI 2.0.



Fonte: Adaptado de AEROSPIKE (2025).

A Figura 2, demonstra que diferentemente da IA generativa, que produz conteúdo quando solicitada, a IA agêntica é proativa, tomando decisões e agindo em busca de objetivos. Esta evolução paradigmática representa uma transição de sistemas de resposta a *prompts* para sistemas de ação autônoma, com implicações significativas para a automação de processos analíticos e decisórios em organizações complexas como as IES.

Os sistemas multiagentes (*Multi-Agent Systems - MAS*) constituem arquitetura na qual múltiplos agentes interagem para resolver problemas ou alcançar objetivos compartilhados. De acordo com a AWS (2025), esses agentes podem ser homogêneos (similares em *design*) ou heterogêneos (diferentes em estrutura ou função) e podem colaborar, coordenar ou mesmo competir dependendo do contexto. A eficácia dos sistemas multiagentes em ambientes distribuídos e complexos os torna particularmente adequados para aplicações educacionais, nas quais diferentes aspectos da avaliação institucional podem ser endereçados por agentes especializados que compartilham informações e coordenam suas análises.

No domínio educacional, os agentes de IA têm sido aplicados em múltiplas frentes. Jiang *et al.* (2024) propuseram o *Von Neumann Multi-Agent System Framework for Education (AI Agent for Education)* Figura 3, uma arquitetura que integra múltiplos agentes baseados em LLM para apoiar processos de ensino-aprendizagem. Os autores argumentam que os sistemas multiagentes podem realizar tanto a circulação externa para aprendizes humanos, fomentando

a construção de conhecimento, quanto a circulação interna para aumentar a inteligência coletiva entre agentes baseados em LLM. Esta perspectiva sugere aplicações além da tutoria individualizada, abrangendo suporte a processos de gestão e avaliação institucional.

Figura 3 – O ciclo de aprimoramento de habilidades do MAS para a educação.



Fonte: The ability enhancement cycle of MAS for education (Jiang *et al.*, 2024, Adaptado/tradução nossa).

A aplicação de agentes inteligentes no contexto da educação superior tem sido objeto de crescente atenção na literatura. Uma revisão sistemática de Gu e Blackmore (2015) sobre modelagem baseada em agentes no domínio da educação superior identificou aplicações em simulação de dinâmicas institucionais, modelagem de comportamento de estudantes e análise de sistemas educacionais complexos. Mais recentemente, Alexandru *et al.* (2015) demonstraram a utilização de agentes inteligentes em sistemas adaptativos de e-learning, evidenciando sua capacidade de personalizar experiências educacionais com base em características individuais dos aprendizes.

A emergência da IA agêntica traz implicações específicas para a gestão de IES. Conforme ASC OHIO STATE (2025), os sistemas agênticos representam uma mudança de paradigma do modo 'Copiloto' (assistido) para o modo 'Auto piloto' (autônomo), com ênfase

em estruturas hierárquicas de agentes para coordenar trabalho complexo. No contexto da avaliação institucional, isso implica a possibilidade de sistemas que não apenas respondem a consultas específicas, mas monitoram proativamente indicadores, identificam desvios de desempenho e sugerem intervenções antes que problemas se consolidem.

A aplicação de agentes ao contexto do SINAES possibilita endereçar limitações identificadas nos processos avaliativos tradicionais. Agentes especializados podem ser configurados para: processamento automatizado de relatórios de autoavaliação, identificando padrões e inconsistências; análise comparativa de indicadores do ENADE entre ciclos avaliativos e entre cursos similares; síntese de recomendações das comissões de avaliação externa; e geração de planos de ação baseados em evidências. A orquestração desses agentes em arquitetura multiagente permite que informações de múltiplas fontes sejam integradas para subsidiar decisões mais fundamentadas.

2.7.2 Agentes Inteligentes e Sistemas Multiagentes

A fundamentação teórica dos agentes de inteligência artificial empregados no *framework* proposto ancora-se em três contribuições seminais da literatura: a formalização do conceito de agente racional por Russell e Norvig (2013), a caracterização das propriedades constitutivas de agentes autônomos por Wooldridge (2009) e a sistematização dos princípios de sistemas multiagentes por Jennings *et al.* (1998). A articulação dessas bases teóricas permite compreender não apenas a natureza computacional dos agentes, mas também as implicações de sua aplicação em domínios organizacionais complexos, como a avaliação de cursos de graduação no âmbito do SINAES.

Russell e Norvig (2013) definem agente como qualquer entidade que percebe seu ambiente por meio de sensores e age sobre esse ambiente por meio de atuadores. Essa definição, aparentemente simples, encerra uma formalização que se aplica tanto a agentes físicos (robôs, por exemplo) quanto a agentes de *software*, nos quais os sensores correspondem a entradas de dados como consultas a bancos de dados, recebimento de documentos, leitura de indicadores e os atuadores correspondem às saídas produzidas como relatórios gerados, recomendações emitidas, alertas disparados. Os autores introduzem o conceito de agente racional como aquele que, para cada sequência de percepções possível, seleciona a ação que maximiza sua medida de desempenho, considerando as evidências fornecidas pela sequência perceptual e o conhecimento embutido no agente. Essa racionalidade não pressupõe onisciência; o agente

opera com informação limitada e deve aprender a partir da experiência para aprimorar sua eficácia ao longo do tempo.

Para estruturar o projeto de agentes, Russell e Norvig (2013) propõem o descritor PEAS (*Performance, Environment, Actuators, Sensors*), que especifica quatro dimensões constitutivas: a medida de desempenho que avalia o sucesso do agente, a natureza do ambiente no qual opera, os atuadores de que dispõe para agir e os sensores que utiliza para perceber.

Russell e Norvig (2013) classificam ainda os ambientes de tarefa segundo propriedades que determinam o grau de complexidade do projeto do agente: ambientes completamente ou parcialmente observáveis, determinísticos ou estocásticos, episódicos ou sequenciais, estáticos ou dinâmicos, discretos ou contínuos, e com agente único ou multiagente.

Complementarmente à formalização de Russell e Norvig, Wooldridge (2009) oferece uma caracterização das propriedades que distinguem agentes de programas convencionais. O autor identifica quatro atributos essenciais. A autonomia refere-se à capacidade do agente de operar sem intervenção direta de seres humanos ou de outros agentes, exercendo controle sobre seu próprio estado e comportamento. A reatividade denota a habilidade de perceber mudanças no ambiente e responder tempestivamente a elas. A proatividade (*pro-activeness*) descreve a capacidade de tomar iniciativa orientada a objetivos, em vez de apenas reagir a estímulos externos. Por fim, a habilidade social consiste na capacidade de interagir com outros agentes e, potencialmente, com seres humanos por meio de protocolos de comunicação para alcançar seus objetivos.

Wooldridge (2009) distingue ainda duas concepções de agêntica: a noção fraca (*weak notion*), que compreende os quatro atributos supracitados, e a noção forte (*strong notion*), que atribui aos agentes estados mentais intencionais como crenças, desejos e intenções, conforme o modelo BDI (*Belief-Desire-Intention*). Para os propósitos da presente pesquisa, adota-se a noção fraca de agêntica, uma vez que os agentes baseados em modelos de linguagem de grande escala (*Large Language Models - LLMs*) não possuem estados mentais intencionais em sentido estrito, mas emulam comportamento orientado a objetivos por meio de instruções (*prompts*) que definem seu papel, contexto e critérios de avaliação.

A dimensão dos sistemas multiagentes é formalizada por Jennings *et al.* (1998), que definem um sistema multiagente (*Multi-Agent System - MAS*) como aquele composto por múltiplos agentes que interagem entre si, frequentemente em nome de seus respectivos usuários, com diferentes objetivos e motivações. Os autores argumentam que o MAS são particularmente adequados para domínios nos quais: o problema é inerentemente distribuído, de modo que

diferentes aspectos requerem competências especializadas; a complexidade do problema excede a capacidade de um agente individual; ou a modularidade e escalabilidade do sistema são requisitos de projeto. Os três critérios aplicam-se ao contexto da avaliação institucional: os indicadores do SINAES abrangem dimensões heterogêneas com organização didático-pedagógica, corpo docente, infraestrutura, cada qual com critérios analíticos distintos; a análise integrada de dezenas de indicadores, cruzada com documentos institucionais e dados históricos, ultrapassa a capacidade de um único agente generalista; e a necessidade de evolução incremental do sistema demanda arquitetura que permita a adição de novos agentes sem reestruturação completa.

Jennings *et al.* (1998) identificam três padrões fundamentais de interação em sistemas multiagentes: cooperação, coordenação e negociação. Na cooperação, agentes trabalham conjuntamente para atingir um objetivo comum. Na coordenação, agentes ajustam suas atividades para evitar conflitos e garantir coerência. Na negociação, agentes com objetivos parcialmente conflitantes buscam acordos mutuamente aceitáveis. O *framework* proposto adota predominantemente o padrão de cooperação com coordenação sequencial: os agentes compartilham o objetivo comum de produzir análise integrada e acionável para o gestor, e suas atividades são coordenadas por meio de um fluxo determinístico que define a ordem de execução e as dependências entre etapas.

2.7.2.1 Componentes funcionais de um agente inteligente

A literatura recente sobre agentes baseados em LLMs identifica quatro componentes funcionais que habilitam o comportamento autônomo: raciocínio, planejamento, memória e uso de ferramentas (GOOGLE CLOUD, 2025; BANDI *et al.*, 2025). A compreensão desses componentes é necessária para fundamentar as decisões de projeto adotadas na arquitetura do *framework*.

O raciocínio (*reasoning*) constitui a capacidade do agente de utilizar informações disponíveis para derivar conclusões, estabelecer relações causais e formular julgamentos. Nos agentes baseados em LLMs, o raciocínio é potencializado por técnicas de engenharia de *prompt* como a cadeia de raciocínio (*chain-of-thought*), na qual o modelo é instruído a explicitar os passos intermediários de seu processo analítico (WEI *et al.*, 2022). No Agente de Diagnóstico do *framework*, por exemplo, o *prompt* instrui explicitamente a sequência analítica: identificar o conceito atual do indicador, comparar com os atributos dos conceitos superiores definidos na

MPC, listar os *gaps* específicos, classificar a prioridade e considerar o tipo de instrumento aplicável. Essa estruturação do raciocínio confere transparência ao processo decisório e facilita a auditoria das recomendações geradas.

O planejamento (*planning*) refere-se à capacidade de decompor objetivos complexos em subtarefas gerenciáveis e definir a sequência de ações necessárias para alcançá-los. Russell e Norvig (2013) dedicam parte substancial de sua obra à formalização do planejamento em agentes, distinguindo abordagens clássicas (planejamento como busca em espaço de estados) de abordagens situadas (planejamento reativo e hierárquico). Nos agentes contemporâneos baseados em LLMs, o planejamento opera de modo menos formal, porém análogo: diante de uma tarefa como "produzir recomendações para melhoria do indicador 1.5", o agente decompõe internamente a tarefa em etapas como consultar o diagnóstico do indicador, buscar evidências na base RAG, verificar artefatos existentes, formular a recomendação, e por fim decide autonomamente a ordem de execução.

A memória compreende os mecanismos que permitem ao agente reter e recuperar informações ao longo de sua operação. Distinguem-se dois tipos: a memória de curto prazo, que corresponde ao contexto mantido durante uma sessão de interação, e a memória de longo prazo, que persiste informações entre sessões. No *framework* proposto, a memória de curto prazo é implementada via *prompt* na janela de contexto do LLM. A memória de longo prazo, por sua vez, é operacionalizada pela base de conhecimento RAG, que armazena e indexa vetorialmente os documentos institucionais como o PPC, DCN, relatórios de autoavaliação e os artefatos produzidos em ciclos anteriores, constituindo o que se denomina memória institucional inteligente.

O uso de ferramentas (*tool use*) consiste na capacidade do agente de invocar recursos externos para complementar suas funcionalidades. Essa capacidade distingue os agentes contemporâneos dos modelos de linguagem convencionais, que operam exclusivamente sobre informações presentes no *prompt* e em seus parâmetros treinados. No *framework* proposto, cada agente dispõe de um conjunto curado de ferramentas como a consulta a indicadores, diagnóstico MPC, busca na base RAG, consulta a artefatos e a consulta ao plano de ação selecionadas conforme o princípio do menor privilégio: cada agente acessa apenas os dados necessários ao cumprimento de sua responsabilidade funcional. Essa delimitação é relevante tanto para a qualidade das saídas, evitando que o agente se disperse com informações irrelevantes, quanto para a governança do sistema.

2.7.2.2 Padrões de orquestração e compartilhamento de contexto

A operacionalização de sistemas multiagentes requer a definição de padrões de orquestração que governem a sequência de execução dos agentes e os mecanismos de compartilhamento de informações entre eles. A literatura identifica dois padrões fundamentais: a orquestração sequencial e a orquestração paralela (WANG *et al.*, 2024).

Na orquestração sequencial, os agentes são executados em cadeia, de modo que a saída de um agente constitui a entrada do agente subsequente. Esse padrão é adequado quando há dependência lógica entre as etapas do processamento: o diagnóstico precede a recomendação, que por sua vez precede a síntese. A vantagem reside na rastreabilidade em que cada saída intermediária pode ser inspecionada e auditada, e na previsibilidade do fluxo de processamento. A limitação consiste no tempo total de execução, que corresponde à soma dos tempos individuais de cada agente.

Na orquestração paralela, múltiplos agentes são executados simultaneamente sobre aspectos independentes do problema, e seus resultados são agregados por um agente consolidador. Esse padrão é eficiente quando as subtarefas não possuem dependência entre si, por exemplo, a análise simultânea de indicadores de dimensões distintas, e reduz o tempo total de processamento. A complexidade adicional reside na necessidade de mecanismos de sincronização e resolução de conflitos entre os resultados parciais.

O *framework* proposto adota a orquestração sequencial determinística com três estágios de processamento: o Agente de Diagnóstico produz o diagnóstico estruturado, que alimenta o Agente de Recomendação; a concatenação do diagnóstico e das recomendações serve como entrada para o Agente de Síntese. Essa linearidade foi adotada deliberadamente em detrimento de arquiteturas mais complexas de orquestração dinâmica, em razão de dois fatores: a rastreabilidade exigida pela validação científica da pesquisa, que requer que cada recomendação possa ser rastreada até o diagnóstico que a originou; e a previsibilidade operacional necessária ao contexto de gestão institucional, no qual os gestores precisam compreender a proveniência de cada orientação recebida.

O compartilhamento de contexto entre agentes ocorre por meio da transmissão explícita das saídas textuais: o texto integral produzido pelo Agente de Diagnóstico é incluído no *prompt* do Agente de Recomendação, e ambos os textos, diagnóstico e recomendações, compõem o *prompt* do Agente de Síntese. Adicionalmente, todos os agentes compartilham acesso à mesma base de conhecimento RAG e ao mesmo conjunto de dados estruturados do

curso, assegurando que operem sobre informações consistentes. Esse modelo de compartilhamento por passagem de mensagens textuais, em vez de memória compartilhada ou *blackboard*, alinha-se ao padrão de comunicação documentado na literatura de sistemas multiagentes (JENNINGS *et al.*, 1998) e garante que cada agente receba contexto completo para sua tarefa, sem ambiguidade sobre a procedência das informações.

Cabe registrar que a opção por implementação nativa dos agentes, sem utilização de *frameworks* de orquestração como *LangChain*, *CrewAI* ou *AutoGen*, fundamenta-se na necessidade de controle granular sobre parâmetros de geração (temperatura, limites de tokens, número máximo de passos), sobre o conjunto de ferramentas disponíveis a cada agente e sobre a integração estrita com a arquitetura *multi-tenant* da aplicação web. Essa decisão de projeto, embora demande maior esforço de desenvolvimento, assegura que a lógica de orquestração seja transparente, testável e completamente rastreável, requisitos indispensáveis para a validação científica do artefato produzido no âmbito da *Design Science Research* (DSR).

Em síntese, a fundamentação teórica em Russell e Norvig (2013), Wooldridge (2009) e Jennings *et al.* (1998) fornece os alicerces conceituais para o projeto da arquitetura multiagente do *framework*. A tradução dessas bases teóricas em decisões de projeto concretas é detalhada no Capítulo 5, que descreve a implementação técnica dos agentes, suas ferramentas e seus fluxos de integração com a aplicação.

2.7.3 Prompt Engineering: Fundamentos e Técnicas

O *prompt engineering* constitui disciplina emergente dedicada ao desenvolvimento de instruções eficazes para modelos de linguagem de grande escala (*Large Language Models - LLMs*) (AMATRIAIN, 2024). Conforme argumentam ZHAO *et al.* (2023), a qualidade das saídas produzidas por LLMs depende criticamente da formulação dos *prompts*, as instruções textuais que especificam a tarefa a ser executada, o contexto relevante e o formato esperado de resposta.

A relevância do *prompt engineering* para sistemas de apoio à decisão reside no fato de que LLMs não são programadas para tarefas específicas mediante código tradicional, mas instruídos por meio de linguagem natural. Esta característica oferece flexibilidade sem precedentes, permitindo que não programadores especifiquem comportamentos complexos, mas também introduz desafios de consistência e controle que demandam técnicas específicas de engenharia de *prompts*.

Wei *et al.* (2022) identificaram uma técnica particularmente relevante denominada *chain-of-thought prompting* (encadeamento de raciocínio). Esta técnica consiste em instruir o modelo a decompor problemas complexos em etapas intermediárias de raciocínio, explicitando o processo lógico que conduz à resposta final. Os autores demonstraram que o *chain-of-thought prompting* melhora significativamente o desempenho de LLMs em tarefas que requerem raciocínio matemático, lógico e de senso comum (WEI *et al.*, 2022).

No contexto do *framework* proposto para avaliação institucional, o *chain-of-thought prompting* apresenta aplicabilidade direta. A análise de indicadores do SINAES requer que o sistema: (a) identifique o conceito atual do curso em determinado indicador; (b) compare com os critérios dos conceitos superiores; (c) identifique as lacunas específicas; (d) formule recomendações que enderecem essas lacunas. Instruir o agente de IA a explicitar cada uma dessas etapas aumenta a transparência e a confiabilidade das recomendações geradas.

Outras técnicas de *prompt engineering* relevantes incluem:

Few-shot prompting: Fornecimento de exemplos de entrada-saída esperada antes da tarefa principal, demonstrando o padrão desejado de resposta. Para o *framework* proposto, exemplos de análises de indicadores com recomendações bem formuladas podem ser incluídos no *prompt* para calibrar o comportamento do modelo.

Role prompting: Especificação de um papel ou persona para o modelo assumir, contextualizando suas respostas. Instruir o agente a atuar como "especialista em avaliação institucional conforme diretrizes do SINAES" estabelece o registro discursivo e o nível de expertise esperados.

Structured output prompting: Solicitação de respostas em formato estruturado (JSON, tabelas, listas numeradas), facilitando o processamento computacional subsequente. Esta técnica é essencial para integração dos *outputs* da IA com a interface de visualização do *framework*.

2.7.4 *Retrieval-Augmented Generation (RAG)*

Os modelos de linguagem de grande escala, apesar de suas capacidades avançadas de compreensão e geração de texto, operam exclusivamente sobre dois conjuntos de informação: o conhecimento internalizado durante o treinamento que é estático e sujeito à defasagem temporal e o conteúdo fornecido na janela de contexto de cada requisição. Essa limitação estrutural implica que, quando aplicados a domínios especializados como a avaliação

institucional de cursos de graduação, os LLMs tendem a produzir respostas genéricas ou factualmente imprecisas, fenômeno amplamente documentado como alucinação (ZHAO *et al.*, 2023). A técnica de *Retrieval-Augmented Generation* (RAG) emerge como resposta arquitetural a essa limitação, ao integrar mecanismos de recuperação de informação com a capacidade generativa dos modelos de linguagem.

A concepção original do RAG foi formalizada por Lewis *et al.* (2020), que propuseram uma arquitetura composta por dois componentes complementares: um modelo recuperador (*retriever*), responsável por identificar documentos ou trechos relevantes em uma base de conhecimento externa, e um modelo gerador (*generator*), que condiciona a produção de texto aos documentos recuperados. Os autores demonstraram que essa combinação supera tanto modelos puramente paramétricos que dependem exclusivamente do conhecimento internalizado nos pesos da rede neural quanto abordagens puramente extrativas que se limitam a selecionar trechos sem capacidade de síntese ou reformulação. A contribuição central de Lewis *et al.* (2020) reside na demonstração de que o condicionamento da geração a evidências documentais externas melhora a factualidade das respostas e permite a atualização do conhecimento do modelo sem necessidade de retreinamento, mediante a simples atualização da base de documentos indexados.

O mecanismo de recuperação no RAG opera por meio de representações vetoriais densas (*embeddings*), nas quais tanto os documentos da base de conhecimento quanto as consultas dos usuários ou, no caso de sistemas agênticos, as consultas formuladas autonomamente pelos agentes são convertidas em vetores numéricos de alta dimensionalidade. A similaridade entre consulta e documentos é calculada por métricas como a similaridade cossenoidal, que mensura a proximidade semântica entre os vetores no espaço de representação. Essa abordagem supera as limitações das técnicas clássicas de recuperação baseadas em correspondência lexical (*keyword matching*), ao capturar relações semânticas entre termos que não compartilham superfície linguística como por exemplo, reconhecendo que uma consulta sobre "infraestrutura laboratorial" é semanticamente próxima de um trecho do PPC que descreve "equipamentos e espaços de prática de estágio".

Gao *et al.* (2023) oferecem uma taxonomia abrangente dos paradigmas de RAG, distinguindo três variantes: o RAG ingênuo (*naive RAG*), que implementa o fluxo linear de indexação, recuperação e geração; o RAG avançado (*advanced RAG*), que incorpora técnicas de pré-processamento e pós-processamento para refinar a qualidade da recuperação, como reranqueamento de resultados e segmentação semântica de documentos; e o RAG modular

(*modular RAG*), que decompõe o pipeline em módulos intercambiáveis, permitindo configurações flexíveis de busca, filtragem e geração. Os autores argumentam que a evolução do RAG ingênuo para as variantes avançada e modular responde a limitações práticas identificadas na aplicação da técnica, como a recuperação de trechos irrelevantes, a perda de contexto nas fronteiras de segmentação e a dificuldade de o modelo gerador integrar informações dispersas em múltiplos fragmentos recuperados.

No contexto do *framework* proposto, a adoção da técnica de RAG fundamenta-se em três requisitos do domínio de avaliação institucional. O primeiro diz respeito à especificidade documental: os critérios de atribuição de conceitos do SINAES exigem que as análises e recomendações sejam fundamentadas em documentos concretos do curso como o Projeto Pedagógico de Curso, a Diretrizes Curriculares Nacionais e relatórios de autoavaliação, cuja interpretação contextualizada não está presente nos parâmetros genéricos de treinamento do modelo. O segundo requisito refere-se à atualização contínua: documentos institucionais são revisados periodicamente pelos coordenadores e pelo NDE, de modo que o sistema precisa operar sobre a versão mais recente de cada documento sem depender de retreinamento do modelo. O terceiro requisito concerne à rastreabilidade: em um sistema de apoio à decisão voltado a gestores e membros da CPA, cada recomendação deve ser rastreável até a evidência documental que a fundamenta, o que só é possível quando o modelo gera suas respostas a partir de trechos identificados e citáveis.

Cabe distinguir, do ponto de vista epistemológico, a natureza do RAG em relação a outras abordagens de extração de conhecimento. Diferentemente do processo de Descoberta de Conhecimento em Bancos de Dados (KDD), que busca identificar padrões previamente desconhecidos em conjuntos de dados estruturados por meio de técnicas de mineração (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996), o RAG configura-se como mecanismo de fundamentação documental: sua função não é descobrir conhecimento novo, mas sim recuperar e articular conhecimento já existente em documentos normativos e institucionais para subsidiar a geração de análises contextualizadas. Essa distinção é relevante para a delimitação metodológica da presente pesquisa, que emprega Análise de Conteúdo (BARDIN, 2016) como técnica de decomposição dos documentos normativos e RAG como mecanismo de acesso contextual a esses documentos pelos agentes de IA, sem recorrer a técnicas de mineração de dados no sentido estrito do termo.

A eficácia do RAG como estratégia de mitigação de alucinações está condicionada à qualidade do pipeline de processamento documental com segmentação, geração de

embeddings e estratégia de recuperação, cuja implementação técnica no contexto do *framework* é detalhada no Capítulo 5.

2.7.5 Limitações e Alucinações de LLMs

A aplicação de LLMs em sistemas de apoio à decisão requer compreensão rigorosa de suas limitações inerentes. O fenômeno mais crítico documentado na literatura é a tendência dos modelos a produzirem "alucinações", *outputs* que parecem plausíveis e são apresentados com confiança, mas que contêm informações factuais incorretas, inconsistentes ou fabricadas.

As alucinações em LLMs derivam de características fundamentais de sua arquitetura e treinamento. Esses modelos são otimizados para produzir sequências de tokens estatisticamente prováveis dado o contexto, não para verificar a veracidade factual de suas afirmações. Consequentemente, quando o modelo não possui informação confiável sobre determinado tópico, ele pode gerar conteúdo plausível, mas incorreto, seguindo padrões linguísticos aprendidos durante o treinamento.

No contexto de sistemas de apoio à decisão para gestão educacional, as alucinações representam risco significativo. Um agente de IA que produza recomendações fundamentadas em interpretações incorretas dos critérios do SINAES, ou que fabrique evidências inexistentes sobre a situação de um curso, pode induzir gestores a decisões equivocadas com consequências potencialmente graves para a qualidade educacional e para os processos regulatórios.

Além das alucinações, outras limitações relevantes incluem:

Viés de confirmação: LLMs podem exibir tendência a concordar com premissas apresentadas pelo usuário, mesmo quando estas são incorretas, comprometendo a sua utilidade como ferramenta de análise crítica.

Inconsistência temporal: Modelos podem produzir respostas diferentes para *prompts* idênticos em momentos distintos, dificultando a replicabilidade das análises.

Limitações de contexto: Apesar de janelas de contexto expandidas (até 200.000 tokens em modelos recentes), LLMs ainda enfrentam dificuldades em manter coerência ao longo de documentos muito extensos ou em integrar informações dispersas no contexto.

Ausência de atualização: O conhecimento dos modelos é limitado aos dados de treinamento, não incorporando informações posteriores à data de corte. Alterações nos instrumentos de avaliação do INEP, por exemplo, podem não estar refletidas nas respostas do modelo.

2.8 *Design Science Research*: Fundamentos e Justificativa de Escolha

O *Design Science Research* (DSR) constitui paradigma metodológico consolidado nas ciências aplicadas, particularmente nos campos de sistemas de informação e engenharia, que se distingue das abordagens tradicionais de pesquisa por seu foco na construção e avaliação de artefatos destinados a resolver problemas práticos. Enquanto as ciências naturais e comportamentais buscam compreender fenômenos mediante observação e teorização, o DSR orienta-se para a criação de soluções que transformam estados existentes em estados desejados (HEVNER *et al.*, 2004).

A emergência do DSR como paradigma reconhecido remonta aos trabalhos seminiais de Herbert Simon sobre as "ciências do artificial", que estabeleceram as bases epistemológicas para considerar o *design* como atividade científica legítima. Simon (1996) argumentou que as ciências do artificial diferem das ciências naturais por tratarem não do que as coisas são, mas do que elas poderiam ser, uma orientação fundamentalmente prescritiva e transformadora. Esta perspectiva encontra ressonância direta no contexto da gestão educacional, onde gestores de Instituições de Ensino Superior (IES) necessitam não apenas compreender os fenômenos avaliativos, mas intervir sobre eles de forma estratégica.

Hevner *et al.* (2004) consolidaram o DSR como metodologia rigorosa ao proporem um arcabouço conceitual que articula relevância prática e rigor científico. Segundo os autores, a pesquisa em *Design Science* deve produzir artefatos viáveis na forma de constructos, modelos, métodos ou instâncias, que demonstrem utilidade para resolver problemas identificados no ambiente organizacional. Esta concepção alinha-se às necessidades contemporâneas das IES brasileiras, que enfrentam o desafio de transformar volumes crescentes de dados avaliativos do Sistema Nacional de Avaliação da Educação Superior (SINAES) em conhecimento acionável para a tomada de decisão.

Peppers *et al.* (2007) sistematizaram o processo DSR em seis etapas: identificação do problema; definição dos objetivos da solução; *design* e desenvolvimento; demonstração; avaliação; e comunicação. Hevner *et al.* (2004) estabelecem diretrizes de rigor metodológico, incluindo relevância para o problema prático, avaliação rigorosa do artefato e contribuição verificável ao corpo de conhecimento. A adoção do DSR nesta pesquisa justifica-se pela natureza do problema: o desenvolvimento de um artefato tecnológico para resolver problemas

práticos da gestão de avaliação de cursos no SINAES. A operacionalização de cada etapa e detalhada no Capítulo 3.

2.8.1 Tipologia de Artefatos e Contribuições

A literatura em *Design Science Research* estabelece uma tipologia de artefatos que orienta a natureza das contribuições científicas esperadas. Conforme sistematização complementada por Hevner *et al.* (2004), os artefatos produzidos em pesquisas de DSR podem ser categorizados em quatro tipos fundamentais:

Constructos constituem o vocabulário conceitual de um domínio, fornecendo a linguagem na qual problemas e soluções são definidos e comunicados.

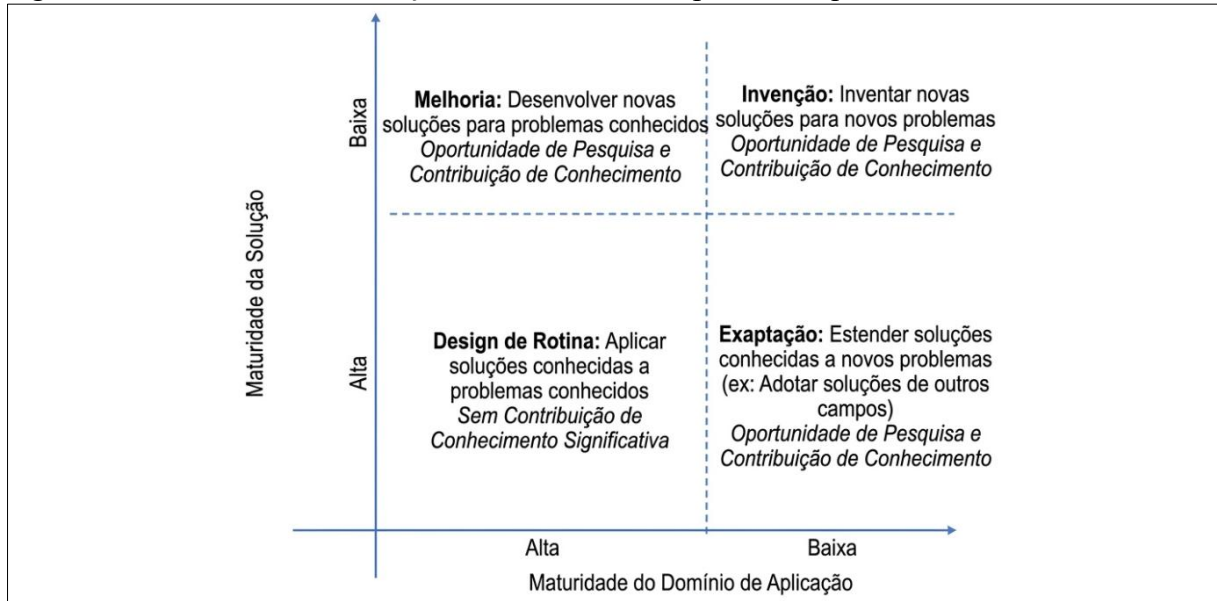
Modelos são representações abstratas que utilizam constructos para descrever tarefas, situações ou artefatos, capturando relações estruturais e funcionais entre elementos do domínio.

Métodos definem processos que orientam a resolução de problemas, estabelecendo sequências de passos baseados em constructos e modelos subjacentes.

Instâncias são realizações concretas de artefatos em ambientes específicos, demonstrando a viabilidade e utilidade prática dos constructos, modelos e métodos propostos.

Gregor e Hevner (2013) ampliaram esta tipologia ao proporem uma matriz de contribuição que considera o grau de maturidade do problema e da solução. Os autores identificam quatro quadrantes: **invenção** (problema novo, solução nova), **melhoria** (problema conhecido, solução nova), **exaptação** (problema novo, solução conhecida) e **rotina** (problema conhecido, solução conhecida). Esta categorização auxilia pesquisadores a posicionarem suas contribuições e a calibrarem expectativas quanto ao impacto científico esperado.

Figura 4 – Quadro de Contribuição de Conhecimento para a Pesquisa em Desenvolvimento



Fonte: Adaptado de *DSR Knowledge Contribution Framework* (Gregor e Hevner, 2013, tradução nossa).

No contexto da presente pesquisa, o *framework* proposto situa-se no quadrante de **melhoria**, uma vez que aborda problema conhecido, as dificuldades das IES em interpretar indicadores do SINAES e formular planos de ação estratégicos, com solução nova, a integração de agentes de IA generativa para descoberta de conhecimento e geração automatizada de recomendações. Esta classificação implica que a contribuição principal reside na demonstração de que tecnologias emergentes podem ser efetivamente aplicadas para resolver problemas persistentes na gestão educacional.

2.8.2 Diretrizes para Rigor e Relevância

O reconhecimento do DSR como paradigma científico legítimo depende da observância de critérios que assegurem simultaneamente relevância prática e rigor metodológico. Hevner *et al.* (2004) propuseram sete diretrizes que orientam a condução e avaliação de pesquisas em Design Science:

Diretriz 1 - Design como Artefato: A pesquisa deve produzir um artefato viável na forma de constructo, modelo, método ou instância. Esta diretriz estabelece o requisito fundamental de que pesquisas em DSR devem resultar em contribuições tangíveis, não apenas em descrições ou explicações de fenômenos.

Diretriz 2 - Relevância do Problema: O objetivo da pesquisa deve ser desenvolver soluções baseadas em tecnologia para problemas organizacionais importantes e relevantes. No

contexto do SINAES, a relevância manifesta-se na necessidade premente das IES de otimizarem seus processos de autoavaliação e gestão da qualidade educacional.

Diretriz 3 - Avaliação do *Design*: A utilidade, qualidade e eficácia do artefato devem ser rigorosamente demonstradas mediante métodos de avaliação bem executados. Esta diretriz fundamenta a necessidade de procedimentos de validação que transcendam a mera demonstração de funcionamento técnico.

Diretriz 4 - Contribuições da Pesquisa: A pesquisa deve prover contribuições claras e verificáveis nas áreas do artefato de *design*, fundamentos de *design* ou metodologias de *design*. As contribuições devem ser originais e relevantes para o avanço do conhecimento no domínio.

Diretriz 5 - Rigor da Pesquisa: A pesquisa deve aplicar métodos rigorosos tanto no desenvolvimento quanto na avaliação do artefato. O rigor manifesta-se na fundamentação teórica adequada, na aplicação de técnicas apropriadas e na documentação sistemática dos procedimentos.

Diretriz 6 - *Design* como Processo de Busca: O *design* de um artefato eficaz requer a utilização dos meios disponíveis para alcançar os fins desejados e satisfazer as leis do ambiente do problema. Esta diretriz reconhece a natureza iterativa do processo de *design*, que envolve ciclos de geração e avaliação de alternativas.

Diretriz 7 - Comunicação da Pesquisa: A pesquisa deve ser apresentada efetivamente tanto para audiências orientadas à tecnologia quanto para audiências orientadas à gestão. Esta diretriz é particularmente relevante para o mestrado profissional, que demanda comunicação acessível a gestores educacionais.

2.8.3 Avaliação de Artefatos em *Design Science Research*

A avaliação do artefato constitui etapa essencial do processo DSR. Hevner *et al.* (2004) estabelecem que a avaliação deve verificar em que medida o artefato proporciona utilidade, qualidade e eficácia na solução do problema que motivou sua construção, podendo assumir formas diversas dependendo do tipo de artefato e dos recursos disponíveis. Venable, Pries-Heje e Baskerville (2016) propõe tipologia organizada em dois eixos: natureza do ambiente de avaliação (natural vs. artificial) e momento da avaliação (ex ante vs. ex post). A avaliação em ambiente artificial - cenários controlados, simulações - e especialmente adequada nos estágios iniciais do desenvolvimento. Pries-Heje, Baskerville e Venable (2008) distinguem

estratégias formativas, orientadas a informar ajustes durante o desenvolvimento, de estratégias somativas, destinadas a verificar se os objetivos finais foram atingidos. Em pesquisas de mestrado com restrições temporais, estratégias formativas baseadas em cenários controlados e análise de consistência interna são frequentemente adotadas, com avaliações somativas em ambiente natural posicionadas como agenda futura (VENABLE *et al.*, 2016).

3 METODOLOGIA

O presente capítulo descreve os procedimentos metodológicos adotados para o desenvolvimento e a validação do *framework* proposto. A abordagem geral, o *Design Science Research* (DSR), e seus fundamentos conceituais foram apresentados na seção 2.7. Aqui, interessa detalhar como cada etapa do processo DSR (PEFFERS *et al.*, 2007) foi operacionalizada nesta pesquisa, quais instrumentos foram empregados e como os resultados de cada etapa informaram as subsequentes.

3.1 Classificação e Natureza da Pesquisa

Esta pesquisa caracteriza-se como aplicada quanto à sua natureza, pois visa gerar conhecimentos direcionados à solução de problemas específicos da gestão educacional, conforme conceituação de Gerhardt e Silveira (2009), que destaca que pesquisas aplicadas "objetivam gerar conhecimentos para aplicação prática, dirigidos à solução de problemas específicos". A investigação foca no desenvolvimento de um artefato tecnológico, o *framework* de apoio à avaliação institucional, que atenda necessidades concretas da Universidade do Estado de Mato Grosso e possa ser replicado em outras instituições de ensino superior.

Quanto aos objetivos, a pesquisa adota caráter exploratório e descritivo. O aspecto exploratório justifica-se pela natureza emergente da integração entre Inteligência Artificial e avaliação institucional no contexto da educação superior, área com lacunas significativas na literatura especializada. Como observado por Gil (2022), pesquisas exploratórias visam proporcionar maior familiaridade com problemas, tornando-os mais esclarecidos e facilitando a construção de hipóteses. O componente descritivo manifesta-se na caracterização detalhada dos processos de avaliação institucional, na especificação sistemática dos indicadores do SINAES e na descrição das funcionalidades do *framework* desenvolvido.

3.2 Abordagem Metodológica: *Design Science Research*

A presente investigação adota como paradigma metodológico o *Design Science Research* (DSR), abordagem consolidada para pesquisas que visam criar e avaliar artefatos tecnológicos destinados a resolver problemas práticos. Diferentemente das abordagens

comportamentais tradicionais, que buscam compreender fenômenos por meio de observação e experimentação, o DSR visa construir e avaliar artefatos que atendam necessidades identificadas (HEVNER *et al.*, 2004).

No âmbito do DSR, a natureza dos dados e procedimentos empregados mobiliza tanto elementos quantitativos quanto qualitativos, sem que isso configure uma abordagem de métodos mistos no sentido tradicional do termo. Trata-se, antes, de uma característica intrínseca ao desenvolvimento e à validação de artefatos tecnológicos aplicados a domínios complexos. Conforme observa Gil (2022), pesquisas exploratórias e descritivas frequentemente combinam diferentes técnicas de coleta e análise para proporcionar compreensão adequada dos fenômenos investigados. No mesmo sentido, Lakatos e Marconi (2017) reconhecem que a quantificação e a interpretação qualitativa constituem recursos complementares cuja articulação depende das exigências do objeto de estudo.

A dimensão quantitativa manifesta-se em dois momentos centrais. Primeiramente, nos dados que alimentam o *framework*: indicadores do SINAES, conceitos do CC oriundos das avaliações *in loco* do CEE/SECITECI-MT (no caso das IES do SES/MT), microdados do ENADE e informações do Censo da Educação Superior (bases nacionais do INEP das quais as IES estaduais também participam) compõem a base numérica processada pelos agentes de IA. Em segundo lugar, nos procedimentos de validação do artefato, que requerem métricas objetivas, taxa de correção lógica, acurácia do diagnóstico, coeficiente de consistência estrutural e taxa de aderência normativa, para aferir o desempenho do *framework* conforme os critérios de avaliação preconizados por Hevner *et al.* (2004).

A dimensão qualitativa, por sua vez, permeia a construção dos artefatos conceituais e a interpretação dos resultados. A elaboração das matrizes MPC, MCA e MCE empregou a Análise de Conteúdo Categórica Temática (BARDIN, 2016) como procedimento sistemático de decomposição dos critérios normativos do INEP, processo essencialmente interpretativo que exigiu identificação de categorias analíticas, codificação de elementos constitutivos e inferência de evidências requeridas. Ademais, a transformação dos padrões identificados pelos algoritmos em recomendações contextualizadas para gestores das IES requer interpretação que considere o contexto institucional e as especificidades de cada curso, dimensão que, como destaca Creswell (2010), demanda sensibilidade aos significados atribuídos pelos sujeitos aos fenômenos estudados.

Conforme argumentam Peffers *et al.* (2007), o *Design Science Research* constitui uma metodologia adequada para pesquisas em sistemas de informação que produzem artefatos

inovadores como contribuição científica. Os autores definem artefato como qualquer objeto projetado, constructos, modelos, métodos ou instâncias, que incorpore conhecimento para resolver problemas relevantes (BASKERVILLE *et al.*, 2018). No contexto desta dissertação, o *framework* proposto configura-se como artefato tecnológico que integra constructos (Matriz de Progressão Conceitual), modelos (arquitetura de agentes), métodos (procedimentos de análise) e uma instância implementada (aplicação web).

A escolha do DSR justifica-se por múltiplos fatores convergentes. Primeiramente, o problema de pesquisa demanda não apenas compreensão teórica, mas a produção de uma solução operacional que transforme dados avaliativos em conhecimento acionável para gestores. Em segundo lugar, a natureza do mestrado profissional orienta-se para a aplicação prática do conhecimento na solução de problemas concretos, alinhando-se à filosofia do DSR que valoriza tanto a relevância prática quanto o rigor científico (SANTANA *et al.*, 2023). Em terceiro lugar, o DSR oferece critérios claros de avaliação de artefatos que não dependem exclusivamente das percepções subjetivas dos usuários, permitindo validação técnica baseada em dados (SONNENBERG; BROCKE, 2012).

Gregor e Hevner (2013) identificam diferentes níveis de contribuição em pesquisas de *Design Science*: invenção (problema novo, solução nova), melhoria (problema conhecido, solução nova), exaptação (problema novo, solução conhecida) e rotina (problema conhecido, solução conhecida). Esta pesquisa posiciona-se no quadrante de melhoria, uma vez que aborda problema conhecido (dificuldades na gestão de indicadores do SINAES) com solução nova (*framework* baseado em agentes de IA generativa para descoberta de conhecimento e geração de recomendações estratégicas).

3.2.1 Modelo Metodológico: DSRM

O desenvolvimento desta pesquisa segue o modelo metodológico proposto por Peffers *et al.* (2007), denominado *Design Science Research Methodology* (DSRM), que estrutura o processo de investigação em seis atividades sequenciais e iterativas: (1) identificação do problema e motivação; (2) definição dos objetivos da solução; (3) *design* e desenvolvimento; (4) demonstração; (5) avaliação; e (6) comunicação. Cada atividade é descrita nas subseções seguintes, com indicação explícita de como foi operacionalizada no contexto desta dissertação e de como se conecta aos objetivos específicos e aos capítulos correspondentes.

Cabe ressaltar que o modelo DSRM não prescreve uma sequência rigidamente linear. Conforme esclarecem Peffers *et al.* (2007), pesquisadores podem iniciar a partir de qualquer uma das quatro primeiras atividades, e iterações entre etapas são esperadas à medida que o entendimento do problema e da solução se aprofunda. Na presente investigação, o ponto de entrada foi a identificação do problema, caracterizando uma abordagem centrada no problema, conforme tipologia dos próprios autores.

Nas subseções a seguir, cada atividade é descrita em detalhe, explicitando os procedimentos adotados, os dados mobilizados e as entregas produzidas em cada etapa do processo de pesquisa.

3.2.1.1 Identificação do Problema e Motivação

A primeira atividade do DSRM consiste em definir o problema de pesquisa e justificar o valor de uma solução. Peffers *et al.* (2007) recomendam que esta etapa atomize o problema para que a complexidade da solução proposta possa ser capturada de forma precisa. A identificação do problema foi conduzida por duas vias complementares.

A primeira via consistiu na revisão sistemática da literatura sobre políticas públicas de avaliação da educação superior no Brasil, com ênfase nos instrumentos normativos do SINAES e nos desafios enfrentados por Instituições de Ensino Superior (IES) nos processos de autoavaliação e avaliação externa. Essa revisão, apresentada no Capítulo 2, evidenciou três constatações centrais: (a) os instrumentos de avaliação de cursos do INEP possuem estrutura complexa com critérios graduados em cinco níveis conceituais, cuja interpretação contextualizada excede a capacidade de análise sistemática dos gestores e membros das Comissões Próprias de Avaliação (CPA); (b) a ausência de ferramentas que traduzam dados avaliativos em recomendações acionáveis compromete a função formativa do SINAES, reforçando seu caráter predominantemente regulatório; e (c) coordenadores de curso operam sob condições de racionalidade limitada (SIMON, 1955), dispondo de grande volume de dados sem instrumentos adequados para transformá-los em planos de ação fundamentados.

A segunda via envolveu a análise do contexto operacional da Universidade do Estado de Mato Grosso (UNEMAT), instituição pública estadual de ensino superior com estrutura *multicampi*, que concentra desafios representativos das IES brasileiras: diversidade de cursos de diferentes graus acadêmicos e modalidades, dispersão geográfica e necessidade de padronização de processos avaliativos. A conjugação dessas duas vias permitiu formular o

problema de pesquisa apresentado na seção 1.2, que articula a demanda prática por instrumentos de apoio à decisão com a lacuna identificada na literatura. Esta atividade relaciona-se diretamente ao Objetivo Específico I, que estabelece a realização de levantamento bibliográfico sobre os fundamentos teórico-metodológicos da pesquisa.

3.2.1.2 Definição de Objetivos da Solução

A segunda atividade do DSRM requer que os objetivos da solução sejam derivados racionalmente da definição do problema. Conforme Peffers *et al.* (2007), esses objetivos podem ser quantitativos, quando expressam em que termos a solução proposta deve ser superior às existentes, ou qualitativos, quando descrevem as funcionalidades esperadas do artefato. No presente caso, adotou-se formulação predominantemente qualitativa, dado o caráter exploratório da solução proposta.

A partir do problema identificado, foram definidos os seguintes objetivos para o artefato: (a) decompor os critérios de avaliação do SINAES em elementos analíticos estruturados mediante as matrizes MPC, MCA e MCE, de modo a tornar explícitos os requisitos de progressão conceitual, as condições de aplicabilidade e os escopos de compartilhamento de evidências entre indicadores; (b) integrar agentes de Inteligência Artificial generativa capazes de processar contextualmente os indicadores e gerar recomendações estratégicas personalizadas, tendo as matrizes como base de conhecimento normativo; (c) materializar o *framework* em aplicação *web* que disponibilize módulos de visualização, monitoramento de indicadores e interface conversacional para apoio à tomada de decisão de coordenadores de curso e membros da CPA.

Esses objetivos de solução correspondem, no plano da pesquisa, aos Objetivos Específicos II, III e IV, respectivamente. A vinculação entre os objetivos do artefato e os objetivos da pesquisa é intencional: no *Design Science Research*, a contribuição científica reside no próprio artefato e no conhecimento gerado durante seu desenvolvimento (HEVNER *et al.*, 2004), de modo que a especificação do artefato constitui, simultaneamente, a operacionalização dos objetivos acadêmicos.

3.2.1.3 Design e Desenvolvimento

A terceira atividade do DSRM compreende a criação do artefato, incluindo a determinação de suas funcionalidades e de sua arquitetura, seguida da implementação propriamente dita. Peffers *et al.* (2007) observam que esta atividade é inerentemente criativa e requer conhecimento tanto da teoria do domínio quanto das tecnologias disponíveis. No contexto desta pesquisa, o *design* e o desenvolvimento foram conduzidos em três frentes complementares, cada uma associada a um ou mais objetivos específicos.

A primeira frente consistiu na construção das matrizes analíticas da MPC, MCA e MCE, operacionalização do Objetivo Específico II. O procedimento empregou técnicas de Análise de Conteúdo Categórica Temática (BARDIN, 2016) aplicadas sistematicamente ao Instrumento de Avaliação de Cursos de Graduação do INEP, nas versões de autorização e reconhecimento/renovação de reconhecimento. A MPC decompõe os descritores dos cinco níveis conceituais de cada indicador, identificando atributos de progressão e evidências requeridas. A MCA codifica as condições institucionais e normativas que determinam a aplicabilidade de cada indicador a um dado curso. A MCE classifica o escopo de compartilhamento das evidências requeridas, distinguindo entre artefatos de responsabilidade do curso, do campus e da instituição. O processo de construção e os resultados dessas matrizes são apresentados no Capítulo 4.

A segunda frente envolveu a integração de agentes de Inteligência Artificial generativa ao *framework*, operacionalização do Objetivo Específico III. Foram projetados três agentes especializados: Diagnóstico, Recomendação e Síntese. Orquestrados de forma sequencial determinística, tendo as matrizes analíticas como base de conhecimento normativo. A arquitetura de agentes incorpora um *pipeline* de *Retrieval-Augmented Generation* (RAG) para processamento de documentos institucionais e técnicas estruturadas de *prompt engineering* para assegurar aderência aos critérios normativos do INEP.

A terceira frente compreendeu o desenvolvimento da aplicação *web* que operacionaliza o *framework*, correspondente ao Objetivo Específico IV. A aplicação implementa arquitetura *multi-tenant* em Laravel/PHP com PostgreSQL, incorporando módulos de cadastro de cursos e indicadores, motor de regras da MCA, calculadora de Conceito de Curso, visualização de dados, execução dos agentes de IA e interface conversacional para interação com os relatórios gerados. As decisões arquiteturais, o modelo de dados e os componentes de *software* são detalhados no Capítulo 5.

A natureza iterativa do processo de *design* manifestou-se ao longo do desenvolvimento: os resultados da construção das matrizes informaram requisitos da

arquitetura de agentes, e os testes iniciais do sistema realimentaram ajustes na estrutura das matrizes e nos *prompts* dos agentes. Essa dinâmica é prevista e encorajada pelo modelo DSRM (PEFFERS *et al.*, 2007), que reconhece ciclos de refinamento entre as atividades de *design*, demonstração e avaliação.

3.2.1.4 Demonstração

A quarta atividade do DSRM visa comprovar que o artefato funciona conforme projetado, mediante sua aplicação a uma ou mais instâncias do problema. Peffers *et al.* (2007) distinguem demonstração de avaliação: enquanto a demonstração evidencia a viabilidade técnica e funcional do artefato, a avaliação mensura em que medida ele efetivamente suporta a solução do problema. Esta distinção é fundamental para a estruturação dos procedimentos de validação adotados nesta pesquisa.

A demonstração do *framework* materializa-se como prova de conceito (descrita na seção 3.2.7), na qual o sistema é aplicado integralmente a um curso real da UNEMAT. O procedimento documenta todas as etapas do fluxo operacional, desde a configuração do perfil do curso e das variáveis da MCA na plataforma, passando pela inserção dos conceitos atribuídos pelo coordenador na autoavaliação, pelo *upload* de documentos institucionais para a base de conhecimento RAG, pela execução sequencial dos agentes de Diagnóstico, Recomendação e Síntese, até a geração do relatório final com diagnóstico, recomendações e plano de ação consolidado. A documentação *end-to-end* inclui capturas de tela, dados de entrada, *outputs* completos de cada agente e registros do painel de transparência de IA.

Esta atividade está associada ao Objetivo Específico V, especificamente na dimensão de verificação da viabilidade operacional do artefato. Os resultados da prova de conceito são apresentados no Capítulo 6.

3.2.1.5 Avaliação

A quinta atividade do DSRM requer observação e mensuração de quão bem o artefato suporta a solução do problema definido. Hevner *et al.* (2004) estabelecem que a avaliação deve verificar utilidade, qualidade e eficácia do artefato por meio de métodos bem executados. Venable, Pries-Heje e Baskerville (2016) contribuem com tipologia que organiza

as estratégias de avaliação em dois eixos, natureza do ambiente (natural *versus* artificial) e momento da avaliação (*ex ante versus ex post*), oferecendo referencial para a seleção de procedimentos compatíveis com as restrições do contexto investigado.

Nesta pesquisa, a avaliação do artefato é operacionalizada por meio de estratégia multimétodo que combina três procedimentos complementares, cada um mensurando uma dimensão específica da qualidade do *framework*: (a) validação por cenários controlados, que verifica a correção lógica dos *outputs* a partir de *inputs* predefinidos cujas respostas esperadas são conhecidas; (b) validação por análise retrospectiva com dados oficiais do Conceito de Curso (CC), que confronta os diagnósticos gerados pelo sistema com resultados de avaliações *in loco* já realizadas, utilizando dados oficiais de avaliações externas conduzidas pela SECITECI, homologadas pelo CEE/MT; e (c) validação por consistência interna, que mensura a estabilidade dos resultados entre execuções múltiplas com os mesmos dados de entrada.

A combinação desses três procedimentos assegura avaliação que captura as dimensões de correção funcional, relevância prática, confiabilidade operacional e validade normativa do artefato. Os procedimentos são detalhados na seção 3.2.2 deste capítulo. Os resultados da avaliação, juntamente com a discussão de suas implicações, são apresentados no Capítulo 6. Esta atividade operacionaliza integralmente o Objetivo Específico V da pesquisa.

3.2.1.6 Comunicação

A comunicação dos resultados ocorre por meio desta dissertação e do produto técnico associado, contribuindo tanto para a literatura acadêmica quanto para a prática profissional em gestão de IES.

A sexta e última atividade do DSRM consiste na comunicação do problema, do artefato, da utilidade e da novidade da solução para as audiências relevantes. Peffers *et al.* (2007) observam que a comunicação deve contemplar tanto audiências orientadas à tecnologia quanto audiências orientadas à gestão, adaptando linguagem e nível de detalhamento conforme o público-alvo.

No contexto do mestrado profissional, a comunicação ocorre por dois canais complementares. O primeiro é esta dissertação, que documenta integralmente o processo de pesquisa, desde os fundamentos teóricos até os resultados da validação, atendendo aos requisitos de rigor acadêmico e contribuição ao corpo de conhecimento. O segundo canal é o produto técnico associado ao programa, materializado na plataforma Notacinco

(notacinco.com.br), que disponibiliza o *framework* como aplicação *web* operacional, contribuindo diretamente para a prática profissional em gestão de IES.

A dupla natureza da comunicação, acadêmica e profissional, reflete a Diretriz 7 de Hevner *et al.* (2004), que preconiza a apresentação efetiva dos resultados para diferentes audiências, e alinha-se à vocação do mestrado profissional de produzir conhecimento com aplicabilidade direta. Dessa forma, a comunicação não se restringe à descrição do artefato, mas demonstra sua inserção no contexto das políticas públicas de avaliação da educação superior e seu potencial de contribuição para a melhoria dos processos de gestão institucional no âmbito do SINAES.

3.2.2 Procedimentos de Validação

A validação do *framework* segue abordagem multimétodo, conforme recomendações de Hevner *et al.* (2004) para avaliação de artefatos em *Design Science Research*. A literatura em DSR reconhece que artefatos computacionais podem ser avaliados por múltiplas estratégias, que variam desde demonstrações técnicas e análises formais até estudos empíricos com usuários (PEFFERS *et al.*, 2007). A seleção dos procedimentos de validação considera tanto o rigor científico exigido quanto as restrições práticas do contexto investigado, particularmente a indisponibilidade de dados históricos de autoavaliação institucional em nível de indicadores individuais e a natureza exploratória do artefato proposto.

A estratégia de validação fundamenta-se na distinção conceitual entre **demonstração** e **avaliação** estabelecida por Peffers *et al.* (2007) no modelo DSRM. A demonstração comprova que o artefato funciona conforme projetado, evidenciando sua viabilidade técnica. A avaliação, por sua vez, observa quão bem o artefato suporta a solução do problema definido, requerendo procedimentos que mensurem propriedades específicas como correção, relevância, estabilidade e conformidade normativa. Os procedimentos descritos a seguir operacionalizam a fase de avaliação do DSRM, enquanto a prova de conceito materializa a demonstração.

A combinação de quatro métodos complementares assegura uma avaliação que captura múltiplas dimensões da qualidade do artefato, conforme sintetizado na Tabela 3.

Tabela 3 – Síntese dos procedimentos de validação do *framework*

Procedimento	Dimensão Avaliada	Dados Utilizados	Métricas Principais
Cenários Controlados	Correção funcional	Cenários sintéticos com perfis contrastantes de cursos (conceitos predefinidos, variáveis MCA e artefatos de evidência), cujo <i>ground truth</i> é gerado deterministicamente pelo motor de regras MPC/MCA	<i>Precision</i> , <i>Recall</i> e <i>F1-Score</i> por categoria de prioridade (crítico, médio, baixo, NSA); F1 Geral (<i>macro-average</i>); verificação de dimensão prioritária; métricas MCE de reconhecimento de evidências (TRA, TAO, TDE, ICC-N2, TFSM, TDR)
Análise Retrospectiva	Relevância prática	Conceitos oficiais do CC atribuídos por comissões do CEE/SECITECI-MT em avaliações <i>in loco</i> de cursos da UNEMAT, realizadas no âmbito do Sistema Estadual de Ensino de Mato Grosso, inseridos como autoavaliação no <i>framework</i> ; perfis MCA e artefatos de evidência dos cursos selecionados	Métricas automáticas de classificação (<i>Precision</i> , <i>Recall</i> , F1 por categoria); Acurácia do Diagnóstico (AD); AD Expandida; <i>Precision</i> de Críticos; métricas MCE; classificação de relevância das recomendações por agente auxiliar (escala de 3 pontos, com análise de concordância via <i>Kappa de Fleiss</i>)
Consistência Interna	Confiabilidade operacional	Execuções múltiplas do <i>framework</i> com os mesmos dados de entrada e configurações idênticas dos agentes	Coefficiente de Consistência Estrutural (proporção de elementos diagnósticos estáveis entre execuções); análise de consistência semântica das recomendações textuais
Prova de Conceito	Viabilidade operacional	Dados reais de um curso da UNEMAT (fluxo integral do <i>framework</i>)	Documentação <i>end-to-end</i> do processo operacional

Fonte: Elaborado pelo autor (2026), com base em Hevner *et al.* (2004) e Peffers *et al.* (2007).

3.2.2.1 Validação por Cenários Controlados

A validação por cenários controlados constitui procedimento analítico que verifica a correção lógica do *framework* mediante a submissão de *inputs* predefinidos cujos *outputs* esperados são conhecidos pelo pesquisador. Este método é reconhecido na literatura de DSR

como avaliação baseada em cenários (*scenario-based evaluation*), na qual o artefato é exercitado em situações representativas que permitem verificar se seu comportamento corresponde ao previsto pela teoria subjacente (HEVNER *et al.*, 2004).

Definição dos cenários: São construídos três cenários sintéticos que representam perfis contrastantes de cursos de graduação em diferentes situações avaliativas. Cada cenário define: (a) o perfil do curso (grau acadêmico, modalidade, área de conhecimento); (b) os conceitos atribuídos a cada indicador aplicável segundo a MCA; (c) as variáveis contextuais do perfil MCA (existência de hospital-escola, laboratórios especializados, programas de extensão, entre outras); e (d) artefatos de evidência sintéticos em diferentes níveis de acesso para validação do reconhecimento de evidências pela MCE.

Os cenários são concebidos de modo a exercitar diferentes comportamentos do sistema, contemplando: (i) cenário com predominância de conceitos baixos (1 ou 2) em uma dimensão específica, gerando alta concentração de gaps críticos e dimensão prioritária claramente identificável; (ii) cenário com conceitos médios uniformes (todos conceito 3), configurando um caso extremo sem gaps críticos e sem dimensão prioritária destacada, de modo a verificar a capacidade discriminativa do sistema na ausência de fragilidades severas; e (iii) cenário com contraste extremo entre dimensões uma em patamar máximo e outra com alta concentração de conceitos 1 e 2 utilizando instrumento de avaliação distinto (Autorização, com pesos 40/35/25, em contraste com Reconhecimento, com pesos 35/20/45) para verificar a adaptação do *framework* a diferentes configurações de pesos dimensionais.

A diversidade de cenários é planejada para abranger dois tipos de instrumento (Reconhecimento e Autorização), cursos de diferentes áreas de conhecimento e graus acadêmicos, e perfis MCA variados desde configurações mínimas (todas as variáveis desativadas) até configurações máximas (com hospital-escola, laboratórios especializados e demais contextos habilitados). Cada cenário utiliza dados de cursos reais da UNEMAT no que se refere à estrutura institucional (identificadores, instrumentos, indicadores), porém os conceitos atribuídos são controlados pelo pesquisador para garantir previsibilidade dos resultados esperados.

Geração do *ground truth*: Para cada cenário, a resposta esperada (*ground truth*) é gerada deterministicamente pelo motor de regras MPC/MCA, sem envolvimento de inteligência artificial. O procedimento aplica as regras de progressão conceitual da MPC a cada indicador: indicadores com conceitos 1 ou 2 são classificados como críticos, com conceito 3 como médios e com conceito 4 como baixos; indicadores com conceito 5 são tratados como em patamar

máximo (sem *gap*) e indicadores marcados como não se aplica pela MCA são excluídos da análise. A dimensão prioritária é determinada pela maior concentração de *gaps* críticos; em caso de empate, prevalece a dimensão com maior volume total de *gaps*.

É necessário explicitar que o *ground truth* assim gerado compartilha as mesmas regras de classificação que foram codificadas nos *prompts* dos agentes de IA. Essa identidade entre regras do *ground truth* e instruções dos agentes configura verificação de conformidade funcional: os procedimentos de validação mensuram se o agente segue as regras internas do sistema com fidelidade, e não se essas regras produzem classificações úteis ou corretas do ponto de vista da gestão educacional. Conforme discutido por Hevner *et al.* (2004, Diretriz 3), a avaliação rigorosa de artefatos em *Design Science Research* requer, idealmente, critérios externos ao artefato. No presente trabalho, a validação por cenários controlados e por análise retrospectiva atende à dimensão de conformidade funcional e estabilidade operacional; a dimensão de eficácia prática, que requereria avaliação por especialistas humanos, constitui limitação metodológica central, tratada na seção 7.2.

A natureza determinística do *ground truth* assegura que existe uma única resposta correta para cada cenário segundo os critérios do próprio *framework*, eliminando a subjetividade na avaliação dos outputs dos agentes e permitindo cálculo automatizado das métricas de validação. Essa propriedade é adequada para verificar a conformidade funcional, ainda que insuficiente, por si só, para demonstrar a relevância prática dos diagnósticos produzidos.

Validação do reconhecimento de evidências (MCE): Complementarmente à validação da classificação de prioridade dos indicadores, os cenários controlados incorporam a verificação do reconhecimento de evidências institucionais pela Matriz de Compartilhamento de Evidências (MCE). Para cada cenário, são construídos artefatos de evidência sintéticos com documentos institucionais realistas distribuídos em três níveis de acesso que testam capacidades distintas dos agentes:

Nível 1 (N1): artefatos vinculados diretamente ao plano MPC do curso em avaliação, acessíveis por consulta direta à base de dados;

Nível 2 (N2): artefatos compartilhados entre cursos, vinculados a planos de outros cursos da mesma instituição, verificando a capacidade de identificação de evidências *cross-course*;

Nível 3 (N3): artefatos indexados exclusivamente na base de conhecimento RAG (*Retrieval-Augmented Generation*), sem vínculo direto ao plano do curso, verificando a

capacidade de descoberta semântica por meio de busca vetorial.

Adicionalmente, indicadores MCE sem artefato associado compõem um grupo de controle: para estes, espera-se que o agente recomende a produção de novas evidências, e não a vinculação a artefatos inexistentes.

Métricas de classificação de prioridade: A avaliação da correção funcional do *framework* adota abordagem de classificação multiclasse com métricas de *Information Retrieval*, computadas de forma automatizada. Cada indicador do instrumento de avaliação é classificado em uma categoria de prioridade segundo regra determinística derivada da MPC: *crítico* (conceitos 1 ou 2), *médio* (conceito 3), *baixo* (conceito 4) e *NSA* (não se aplica, conforme determinado pela MCA). Para cada categoria de prioridade, são calculadas três métricas complementares:

- I. ***Precision***: razão entre o número de indicadores corretamente classificados naquela categoria pelo *framework* (interseção entre o output do agente e o *ground truth*, considerando tanto o código do indicador quanto a categoria de prioridade atribuída) e o número total de indicadores que o agente classificou naquela categoria. A *Precision* mensura a proporção de falsos positivos indicadores que o agente identificou como pertencentes a determinada categoria, mas que, segundo o *ground truth*, pertencem a outra categoria ou não deveriam ter sido incluídos.
- II. ***Recall***: razão entre o número de indicadores corretamente classificados e o número total de indicadores que efetivamente pertencem àquela categoria no *ground truth*. O *Recall* mensura a proporção de falsos negativos com indicadores que o agente deixou de identificar ou classificou em categoria distinta da esperada.
- III. ***F1-Score***: média harmônica entre *Precision* e *Recall*, calculada pela fórmula $F1 = 2 \times (Precision \times Recall) / (Precision + Recall)$, que sintetiza ambas as dimensões de erro em uma métrica única.

A métrica-resumo adotada é o F1 Geral (*macro-average*), obtido pela média aritmética dos *F1-Scores* de todas as categorias (crítico, médio, baixo e NSA). A opção pelo *macro-average* justifica-se por tratar todas as categorias com peso igual, independentemente do número de indicadores em cada uma, evitando viés para categorias com maior representatividade nos cenários. O limiar de aceitação estabelecido para o F1 Geral é $\geq 0,70$,

definido por *design* heurístico com base em dois critérios: (i) na literatura de classificação de textos, valores de F1 entre 0,70 e 0,80 são tipicamente considerados indicadores de desempenho satisfatório para tarefas de complexidade moderada (SOKOLOVA; LAPALME, 2009), e o patamar de 0,70 representa o limite inferior dessa faixa; (ii) no contexto de uso do *framework*, um F1 de 0,70 implica que, na média entre *Precision* e *Recall*, aproximadamente 70% das classificações por categoria são corretas, nível que, embora insuficiente para operação autônoma, é adequado para um sistema de apoio à decisão cujos *outputs* serão revisados pelo coordenador de curso. Não foram identificados, na revisão de literatura, benchmarks específicos para classificação de indicadores de avaliação educacional que permitissem fundamentar o limiar em comparação direta com sistemas similares.

Complementarmente, são realizadas duas verificações adicionais:

- I. **Verificação de Dimensão Prioritária:** verifica se a dimensão apontada pelo agente de IA como mais frágil corresponde à dimensão com maior concentração de gaps críticos no *ground truth*. Quando não há indicadores críticos no cenário (por exemplo, quando todos possuem conceito 3), o *ground truth* considera que nenhuma dimensão se destaca, e qualquer resposta do agente é aceita.
- II. **Verificação de NSA:** avalia se os indicadores marcados como "não se aplica" pela MCA foram corretamente omitidos pelo agente (isto é, não apareceram como gap no diagnóstico), também mediante cômputo de *Precision*, *Recall* e *F1-Score*.

A adoção de *Precision*, *Recall* e *F1-Score* por categoria, em substituição a uma proporção simples de acerto global, permitem distinguir entre erros de omissão (o agente não identificou um gap real com falso negativo, capturado pelo *Recall*) e erros de excesso (o agente identificou gaps inexistentes com falso positivo, capturado pela *Precision*). A granularidade por categoria (crítico, médio, baixo, NSA) possibilita identificar em quais faixas de conceito o *framework* apresenta melhor ou pior desempenho, orientando ajustes específicos na engenharia de *prompt* ou na lógica de processamento dos agentes.

Métricas de reconhecimento de evidências (MCE): A validação do reconhecimento de evidências emprega sete métricas complementares, organizadas por dimensão avaliada:

- I. **Taxa de Reconhecimento de Artefatos (TRA):** proporção dos artefatos existentes que o agente reconhece em suas recomendações, verificando se os agentes consultam a base de evidências antes de recomendar. Limiar de aceitação: $\geq 80\%$.
- II. **Taxa de Adequação de Orientação (TAO):** proporção das recomendações cuja classificação MCE (vincular, produzir, vincular com complemento, revisar) é compatível com o esperado no *ground truth*. O procedimento aceita flexibilidades em classificações semanticamente equivalentes como por exemplo, quando o *ground truth* espera vinculação e o agente recomenda revisão de artefato existente, a orientação é considerada adequada. Limiar de aceitação: $\geq 70\%$.
- III. **Taxa de Diferenciação de Escopo (TDE):** capacidade do agente de distinguir entre evidências de diferentes escopos: institucional (menciona compartilhamento), parcial (menciona complemento específico do curso) e de curso (não menciona compartilhamento). Limiar de aceitação: $\geq 60\%$.
- IV. **Índice de Compartilhamento Cross-Curso (ICC-N2):** proporção dos artefatos compartilhados entre cursos (Nível 2) que o agente identifica, verificando a capacidade de navegação em relações *cross-curso*. Limiar de aceitação: $\geq 70\%$.
- V. **Taxa de Falsos Negativos MCE (TFNM):** proporção dos artefatos existentes para os quais o agente recomenda produção de novo artefato ou ignora completamente, quando deveria recomendar vinculação ao artefato disponível. Limiar de aceitação: $\leq 20\%$.
- VI. **Taxa de Descoberta RAG (TDR):** proporção dos artefatos indexados exclusivamente na base de conhecimento RAG (Nível 3) que o agente consegue descobrir por meio de busca semântica. Esta métrica é de caráter exploratório, sem limiar fixo, dado o caráter investigativo da busca semântica em bases de conhecimento institucional.
- VII. **Qualidade da Descoberta N3:** quando o agente descobre um artefato via busca RAG, esta métrica verifica se a orientação produzida é adequada (vinculação ou vinculação com complemento). Também de caráter exploratório.

Extração e comparação automatizada: Os *outputs* textuais dos agentes de IA são processados por agentes auxiliares de *parsing* com output estruturado (*JSON schema*), que extraem os elementos estruturais, indicadores identificados com suas prioridades, dimensão apontada como prioritária, indicadores omitidos como NSA e classificações MCE das recomendações. Os elementos extraídos são então comparados automaticamente com o *ground truth* para cômputo das métricas, eliminando subjetividade na avaliação e assegurando reprodutibilidade do procedimento.

3.2.2.2 Validação por Análise Retrospectiva com Dados Oficiais

A validação por análise retrospectiva constitui procedimento que utiliza dados oficiais de avaliações *in loco* do CEE/SECITECI-MT (no caso das IES do SES/MT), para verificar a relevância prática dos diagnósticos e recomendações produzidos pelo *framework*. Diferentemente dos cenários controlados, que empregam dados sintéticos com respostas esperadas predefinidas, este procedimento opera com dados reais de avaliações *in loco*, confrontando os *outputs* do *framework* com o histórico avaliativo efetivo dos cursos.

Fundamentação: As avaliações *in loco* realizadas por comissões designadas produzem o Conceito de Curso (CC), que registra o conceito atribuído pela comissão avaliadora a cada indicador do instrumento de avaliação. O procedimento utiliza os conceitos do CC como *input* para o *framework*, tratando-os como se fossem uma autoavaliação realizada pelo coordenador. Esta decisão metodológica fundamenta-se no seguinte raciocínio: se o *framework* é capaz de gerar diagnósticos e recomendações pertinentes a partir dos conceitos oficiais atribuídos pela avaliação, demonstra-se que o sistema produz *outputs* relevantes para situações avaliativas reais, independentemente de quem tenha originado os dados de entrada.

Seleção dos cursos: São selecionados cinco cursos da Universidade do Estado de Mato Grosso (UNEMAT) que tenham passado por avaliação *in loco* com Conceito de Curso (CC) atribuído nos ciclos avaliativos recentes. A seleção observa critérios de diversidade para maximizar a representatividade dos perfis avaliados:

- I. **Grau acadêmico:** inclusão de licenciaturas, bacharelados e cursos de tecnologia;
- II. **Área de conhecimento:** cobertura de diferentes áreas (ciências humanas, saúde, exatas, sociais aplicadas);

- III. **Faixa de CC:** distribuição entre cursos com CC baixo, médio e elevado, incluindo ao menos um curso com $CC \geq 4$ para funcionar como controle positivo na ausência de indicadores com conceitos 1 ou 2, o *framework* não deve gerar alertas de criticidade, verificando-se assim a capacidade discriminativa do sistema;
- IV. **Distribuição de indicadores:** cursos com alta concentração de conceitos 1 e 2 contrapostos a cursos com maioria de conceitos 4 e 5.

Os dados dos cursos são importados a partir de registros oficiais da UNEMAT, incluindo os conceitos atribuídos a cada indicador na avaliação *in loco* e os perfis de aplicabilidade da MCA. A importação é realizada por meio de *seeders* automatizados que garantem a integridade e a rastreabilidade dos dados.

Procedimento: Para cada curso selecionado, executam-se as seguintes etapas:

- (a) Cadastro do perfil do curso no *framework*, incluindo grau acadêmico, modalidade, área de conhecimento e demais variáveis da Matriz de Condições de Aplicabilidade (MCA);
- (b) Inserção dos conceitos atribuídos pela comissão avaliadora do INEP a cada indicador, conforme dados públicos do CC, como autoavaliação vigente no sistema;
- (c) Geração do *ground truth* determinístico pelo motor de regras MPC/MCA, que classifica cada indicador segundo o conceito oficial: *crítico* (conceitos 1 ou 2), *médio* (conceito 3), *baixo* (conceito 4) e *sem gap* (conceito 5); indicadores marcados como NSA pela MCA são excluídos da análise;
- (d) Execução integral da cadeia de agentes (Diagnóstico, Recomendação e Síntese);
- (e) Extração dos elementos estruturais do diagnóstico gerado e cálculo automático das métricas de classificação, incluindo *Precision*, *Recall* e *F1-Score* por categoria, F1 Geral, verificação de dimensão prioritária e métricas MCE de reconhecimento de evidências;
- (f) Exportação das recomendações geradas em formato estruturado para classificação de relevância por agente auxiliar, com registro integral dos

parâmetros operacionais de inferência (modelo utilizado, temperatura, consumo de tokens e custo de processamento).

Validação do reconhecimento de evidências (MCE) no contexto retrospectivo:

A validação MCE na análise retrospectiva emprega a mesma estrutura de três níveis de acesso descrita na seção 3.2.2.1, com uma diferenciação relevante: os artefatos de evidência são distribuídos entre os cursos segundo um gradiente progressivo de complexidade. Os cursos são configurados para exercitar individualmente cada nível desde cursos com artefatos exclusivamente diretos (N1) até cursos cujos artefatos são descobríveis apenas por busca semântica na base de conhecimento RAG (N3). Esta distribuição graduada permite avaliar a capacidade de descoberta e orientação de evidências em contextos de crescente complexidade informacional.

Adicionalmente, artefatos de um curso externo, não incluído entre os cinco cursos avaliados, são indexados na base de conhecimento institucional sem vínculo a curso específico, tornando-os acessíveis a qualquer curso do *tenant* por meio de busca semântica. Este arranjo verifica a capacidade do pipeline RAG de navegar por evidências institucionais transversais em contextos avaliativos distintos.

Métricas automáticas de classificação: As métricas automáticas reutilizam o mesmo arcabouço de *Precision*, *Recall* e *F1-Score* por categoria descrito na seção 3.2.2.1, aplicado agora a dados reais. Adicionalmente, são calculadas três métricas específicas deste procedimento:

- I. **Acurácia do Diagnóstico (AD):** corresponde ao *Recall* da categoria *crítico*, sendo a razão entre os indicadores com conceitos 1 ou 2 corretamente identificados pelo *framework* como críticos e o total de indicadores com conceitos 1 ou 2 presentes no CC oficial. Esta métrica mensura a capacidade do *framework* de detectar as fragilidades mais severas registradas na avaliação oficial. Quando o curso não possui indicadores com conceitos 1 ou 2, a métrica é reportada como não aplicável. Limiar de aceitação: $\geq 0,80$.
- II. **Acurácia do Diagnóstico Expandida (AD Expandida):** extensão da AD que inclui, além dos indicadores críticos, os indicadores com conceito 3 priorizados pelo *framework* para melhoria. A AD Expandida avalia a capacidade do *framework* de identificar o conjunto completo de indicadores

que demandam atenção (conceitos 1, 2 e 3), excluindo apenas os indicadores em patamar satisfatório (conceitos 4 e 5). Limiar de aceitação: $\geq 0,70$.

- III. **Precision de Críticos:** razão entre os indicadores corretamente classificados como críticos pelo *framework* e o total de indicadores que o *framework* classificou nessa categoria. Esta métrica mensura a taxa de falsos alarmes com indicadores sinalizados como críticos que, na avaliação oficial, possuem conceitos iguais ou superiores a 3.

Complementarmente, são identificados e listados os falsos negativos (indicadores C1-C2 não sinalizados pelo *framework*) e os falsos positivos (indicadores sinalizados como críticos que possuem conceitos iguais ou superiores a 3 no CC oficial), permitindo análise qualitativa dos padrões de erro.

Métricas de relevância com classificação por agente auxiliar: O segundo grupo compreende métricas de relevância obtidas por classificação automática realizada por agente de IA auxiliar. Para cada recomendação produzida pelo *framework*, um agente classificador independente atribui classificação em escala ordinal de três pontos:

- I. **(3) Plenamente relevante:** a recomendação endereça uma fragilidade real do curso e a ação proposta é condizente com o tipo de melhoria necessária;
- II. **(2) Parcialmente relevante:** a recomendação endereça uma fragilidade real, porém a ação proposta é genérica ou incompleta;
- III. **(1) Não relevante:** a recomendação não endereça uma fragilidade efetivamente existente ou a ação proposta é inadequada.

As recomendações são exportadas em formato estruturado com metadados de rastreabilidade (curso, código INEP, indicador). O agente classificador recebe as recomendações e os textos oficiais dos indicadores IACG, e executa múltiplas rodadas independentes de classificação na mesma escala de três pontos. Os resultados das múltiplas rodadas são consolidados por meio de análise de concordância (moda por recomendação, concordância intra-agente e *Kappa de Fleiss*), permitindo avaliar a estabilidade interna das classificações do agente. A partir das classificações consolidadas (moda), são calculadas: a média de relevância por curso, a média geral consolidada e a distribuição percentual por nível de relevância.

A opção pela classificação exclusiva por agente auxiliar, em detrimento da

avaliação humana direta, fundamenta-se no escopo desta pesquisa, que prioriza a verificação da conformidade funcional e da estabilidade operacional do *framework*. A classificação por agente permite operacionalizar a avaliação de relevância de forma sistemática e reproduzível, com métricas de concordância que atestam a estabilidade interna do instrumento classificador. É necessário explicitar, contudo, que essa abordagem constitui *proxy* experimental: as classificações produzidas refletem o julgamento de um agente de IA sobre os *outputs* de outro agente da mesma família de modelos (Claude Sonnet 4.6), o que introduz potencial viés compartilhado de treinamento. A classificação por agente é, portanto, indicador de coerência interna do sistema, e não evidência de relevância prática das recomendações para a gestão educacional. A validação por especialistas humanos, coordenadores de curso, membros de CPA e avaliadores institucionais, que permitiria aferir a utilidade percebida e a concordância agente-humano, constitui limitação metodológica central desta pesquisa e trabalho futuro prioritário, conforme detalhado na seção 7.2.

A combinação de métricas automáticas de classificação de prioridade com métricas de relevância obtidas pelo agente auxiliar permite avaliar tanto a correção lógica do diagnóstico (se o *framework* identifica corretamente os indicadores problemáticos) quanto a pertinência prática das recomendações (se as ações propostas são coerentes com o contexto avaliativo do curso).

3.2.2.3 Validação por Consistência Interna

A validação por consistência interna verifica a confiabilidade operacional do *framework*, mensurando a estabilidade dos resultados produzidos quando os mesmos dados de entrada são submetidos ao sistema em diferentes momentos. Este procedimento é particularmente relevante para artefatos que integram modelos de Inteligência Artificial generativa, cuja natureza estocástica introduz variabilidade textual nas respostas mesmo quando os parâmetros de geração são controlados (AMATRIAIN, 2024).

Procedimento: A análise de consistência é operacionalizada como pós-processamento das execuções da análise retrospectiva (seção 3.2.2.2), utilizando os mesmos cursos e dados de entrada. Para cada curso, os elementos estruturais dos diagnósticos e recomendações gerados em cada execução são comparados automaticamente, sem necessidade de execuções adicionais dos agentes de IA. A configuração dos agentes é mantida constante em todas as execuções: modelo de IA, temperatura, limite de tokens e *prompts*. As saídas de cada

execução são registradas integralmente, incluindo os textos dos agentes, os elementos estruturais extraídos e as métricas calculadas, permitindo análise comparativa completa. Para os cenários controlados, a estabilidade entre execuções é reportada como elemento complementar na seção 6.1.4 (variabilidade das métricas de classificação entre execuções), porém sem o cálculo formal do CCE e do ICS, que requerem a comparação detalhada dos elementos estruturais e textos de recomendação, análise conduzida exclusivamente para as execuções da análise retrospectiva.

A configuração dos agentes adota temperaturas diferenciadas por função: o agente de Diagnóstico opera com temperatura baixa (0,1), priorizando a estabilidade e a reprodutibilidade dos elementos estruturais do diagnóstico; o agente de Recomendação opera com temperatura ligeiramente mais elevada (0,3), permitindo variação na formulação textual das orientações sem comprometer a consistência das categorias de ação recomendadas. Esta diferenciação é intencional e visa maximizar a estabilidade dos *outputs* diagnósticos, que devem ser determinísticos, sem eliminar completamente a variabilidade na geração de recomendações textuais, onde certa diversidade de formulação é aceitável e até desejável.

A configuração dos agentes adota temperaturas diferenciadas por função, definidas por *design* heurístico: o agente de Diagnóstico opera com temperatura baixa (0,1), priorizando a estabilidade e a reprodutibilidade dos elementos estruturais do diagnóstico; o agente de Recomendação opera com temperatura ligeiramente mais elevada (0,3), permitindo variação na formulação textual das orientações sem comprometer a consistência das categorias de ação recomendadas. É necessário registrar que essas temperaturas foram definidas por raciocínio teórico, maior determinismo para tarefas de classificação, maior diversidade para geração textual, sem ablação experimental que compare o desempenho com configurações alternativas (por exemplo, temperatura 0,0 ou 0,3 para o agente de Diagnóstico). A ausência de otimização experimental dos hiperparâmetros constitui limitação metodológica, e a ablação de temperaturas e outros parâmetros de geração é recomendada como trabalho futuro para identificar a configuração que maximiza a relação entre estabilidade e qualidade dos *outputs*.

Níveis de consistência: A análise distingue dois níveis de consistência. O primeiro nível, denominado consistência estrutural, refere-se à estabilidade dos elementos objetivos do diagnóstico: quais indicadores são identificados como críticos, médios ou de baixa prioridade; qual a classificação de prioridade atribuída a cada indicador; quais dimensões são apontadas como prioritárias; e quais indicadores são corretamente omitidos como não aplicáveis (NSA). Esses elementos podem ser comparados de forma automática e objetiva entre execuções, por

serem extraídos e representados em formato estruturado.

O segundo nível, denominado consistência semântica, refere-se à equivalência de sentido das recomendações textuais geradas. Reconhece-se que variações na redação são esperadas e aceitáveis, dado o caráter generativo dos modelos de IA, desde que o conteúdo prescritivo seja equivalente, isto é, que as mesmas categorias de ação (curto, médio ou longo prazo), os mesmos tipos de intervenção e as mesmas evidências requeridas sejam mantidos entre execuções.

Métrica: Coeficiente de Consistência Estrutural: proporção dos elementos estruturais do diagnóstico que permanecem idênticos entre as execuções. Consideram-se elementos estruturais: (a) o conjunto de indicadores identificados como críticos ou prioritários; (b) a classificação de prioridade (crítico, médio, baixo) atribuída a cada indicador; (c) a identificação das dimensões com maior concentração de fragilidades; e (d) os indicadores corretamente omitidos como não aplicáveis. O coeficiente é calculado pela razão entre a quantidade de elementos que permaneceram estáveis em todas as execuções e o total de elementos avaliados.

Adicionalmente, a variabilidade das métricas quantitativas entre execuções (F1 Geral, F1 por categoria, métricas AD e métricas MCE) é mensurada pelo desvio-padrão, permitindo quantificar a dispersão dos resultados em torno da média. O consolidado de cada sessão de validação registra média, desvio-padrão, valor mínimo e valor máximo de cada métrica ao longo das repetições.

É importante registrar que a expectativa metodológica não é de consistência absoluta (coeficiente igual a 1,0), o que seria incompatível com a natureza generativa dos modelos de IA utilizados. Espera-se consistência elevada nos elementos estruturais (superior a 0,80), com variação aceitável na formulação textual das recomendações. A combinação de temperaturas baixas nos agentes de diagnóstico e síntese, juntamente com o *ground truth* determinístico gerado pelo motor de regras MPC/MCA, constitui mecanismo de controle que visa maximizar a estabilidade dos *outputs* sem eliminar completamente a variabilidade inerente ao modelo.

3.2.2.4 Prova de Conceito (Demonstração Integral)

A prova de conceito materializa a atividade de **Demonstração** do modelo DSRM (PEFFERS *et al.*, 2007), comprovando que o artefato funciona conforme projetado em ambiente

operacional real. Diferentemente dos procedimentos de avaliação descritos nas subseções anteriores, que mensuram propriedades específicas do artefato, a prova de conceito documenta a execução integral do *framework* desde a entrada de dados até a geração do relatório final com plano de ação, evidenciando a viabilidade prática da solução proposta.

Procedimento:

Seleciona-se um curso específico da UNEMAT para aplicação integral do *framework*. O processo é documentado em todas as suas etapas: (a) configuração do perfil do curso e das variáveis da MCA na plataforma; (b) inserção dos conceitos atribuídos pelo coordenador na autoavaliação vigente; (c) *upload* de documentos institucionais (PPC, DCN, relatórios anteriores) para a base de conhecimento RAG; (d) execução sequencial dos agentes de Diagnóstico, Recomendação e Síntese; e (e) geração do relatório de avaliação com diagnóstico, recomendações e plano de ação consolidado.

A documentação inclui capturas de tela da interface da aplicação, dados de entrada utilizados, *outputs* completos de cada agente, registros do painel de transparência de IA (tokens consumidos, tempo de resposta, modelo utilizado) e o relatório final gerado. Esta documentação *end-to-end* permite que outros pesquisadores e gestores compreendam o funcionamento operacional do sistema e avaliem sua adequação a seus próprios contextos institucionais.

A prova de conceito distingue-se dos procedimentos de validação por não empregar métricas quantitativas de avaliação. Seu propósito é demonstrar que o *framework* é tecnicamente viável e operacionalmente funcional, constituindo evidência complementar que, em conjunto com os resultados dos procedimentos de validação, fundamenta a avaliação global do artefato nos termos exigidos pelo *Design Science Research*.

3.3 Ambiente de Estudo

A presente investigação delimita-se ao contexto institucional da Universidade do Estado de Mato Grosso (UNEMAT), instituição pública estadual de ensino superior vinculada ao Sistema Estadual de Ensino de Mato Grosso (SES/MT). Nessa condição, a UNEMAT opera sob fluxo regulatório distinto daquele das IES federais e privadas: a autoavaliação é conduzida pela Comissão Própria de Avaliação (CPA) da instituição, a avaliação externa é realizada por comissões verificadoras designadas pela Secretaria de Estado de Ciência, Tecnologia e Inovação (SECITECI), e a regulação e a emissão de atos autorizativos competem ao Conselho Estadual de Educação (CEE/MT), conforme estabelecem a Resolução Normativa nº 001/2011-

CEE/MT e a Resolução nº 01/2017-CEE/MT. Esse fluxo opera em alinhamento normativo ao SINAES (Lei nº 10.861/2004), conforme detalhado na seção 2.1.11, razão pela qual os instrumentos estaduais de avaliação espelham, em dimensões, indicadores e escala conceitual, o Instrumento de Avaliação de Cursos de Graduação do INEP.

A escolha da UNEMAT como *locus* de pesquisa justifica-se por três razões convergentes: (a) sua estrutura *multicampi*, que reproduz, em escala institucional, a diversidade de perfis de curso típica do sistema brasileiro de educação superior; (b) a disponibilidade de dados históricos de avaliações externas conduzidas pelo CEE/SECITECI-MT, com Conceitos de Curso atribuídos por comissões de pares segundo instrumento alinhado ao SINAES; e (c) a existência de processo institucional de autoavaliação consolidado, coordenado pela CPA em articulação com a Diretoria de Regulação da Educação Superior (DGRES/UNEMAT). O alinhamento normativo entre SES/MT e SINAES garante, portanto, que a validação do *framework* na UNEMAT preserve a consistência metodológica com o instrumento federal que serviu de *corpus* para a construção das matrizes MPC, MCA e MCE.

Conforme estabelecido em seu Plano de Desenvolvimento Institucional (PDI) 2022-2028 (UNEMAT, 2022), a UNEMAT organiza-se em estrutura acadêmico-administrativa distribuída geograficamente pelo território mato-grossense, configurando um desafio significativo para os processos de gestão e avaliação. Esta complexidade institucional justifica a necessidade de ferramentas tecnológicas que integrem informações de múltiplas fontes e contextos.

O recorte temporal da pesquisa compreende os dados de avaliação institucional correspondentes ao período de 2018 a 2025, abrangendo ciclos avaliativos dos cursos de graduação para diversas áreas de conhecimento representadas na instituição. O *framework* será aplicado em sete cursos presenciais da instituição: Curso de Bacharelado em Agronomia do Campus Universitário de Pontes e Lacerda; Curso de Bacharelado em Medicina do Campus Universitário de Cáceres; Curso de Licenciatura e Bacharelado em Educação Física do Campus Universitário de Cáceres; Curso de Licenciatura em Pedagogia do Campus Universitário de Juara; Curso de Licenciatura em Pedagogia do Campus Universitário de Pontes e Lacerda; Curso de Tecnologia em Alimentos do Campus Universitário de Barra do Bugres; Curso de Tecnologia em Gestão Turismo do Campus Universitário de Nova Xavantina.

3.4 Fontes de Dados

3.4.1 Dados Primários: Instrumentos do SINAES

A fonte primária de dados para o desenvolvimento do *framework* é constituída pelos instrumentos oficiais de avaliação do SINAES, a saber:

- **Instrumento de Avaliação de Cursos de Graduação (INEP):** Documento normativo que estabelece os indicadores, dimensões e critérios de atribuição de conceitos para avaliação de cursos presenciais e a distância. Este instrumento constitui base para a construção da Matriz de Progressão Conceitual.

3.4.2 Dados Secundários: Indicadores e Microdados Oficiais

A pesquisa utiliza bases de dados públicas disponibilizadas pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP):

- **Indicadores de Qualidade:** Indicadores derivados do ENADE - CPC, IDD, IGC e o próprio Conceito ENADE - que são nacionais e, portanto, também se aplicam aos cursos da UNEMAT que participam do Exame. Conceito de Curso (CC) da UNEMAT, oriundo dos relatórios de verificação *in loco* do CEE/SECITECI-MT (que não são publicados pelo INEP).

- **Censo da Educação Superior:** Conceito de Curso (CC) da UNEMAT, oriundo dos relatórios de verificação *in loco* do CEE/SECITECI-MT (que não são publicados pelo INEP).

3.4.3 Dados Institucionais

Complementarmente, a pesquisa utiliza dados provenientes dos sistemas acadêmicos e administrativos da UNEMAT, incluindo informações sobre a estrutura *multicampi* dos cursos, histórico de avaliações *in loco* e documentos institucionais. (PDI, PPC).

3.5 Aspectos Éticos e Conformidade Legal

3.5.1 Conformidade com LGPD

O *framework* foi desenvolvido em conformidade com a Lei Geral de Proteção de Dados Pessoais (Lei nº 13.709/2018). Os dados utilizados na pesquisa são:

- **Dados públicos:** Indicadores do INEP, disponíveis em bases de dados abertas
- **Dados institucionais:** Utilizados mediante autorização formal da administração da UNEMAT
- **Dados agregados:** Sem identificação individual de estudantes ou docentes

3.5.2 Uso Responsável de IA

A integração com IA generativa observa princípios de uso responsável:

- Transparência sobre o uso de IA no processo de análise
- Supervisão humana sobre as recomendações geradas
- Não utilização de dados sensíveis nas requisições à API externa
- Documentação dos *prompts* utilizados para fins de replicabilidade

3.5.3 Limitações Metodológicas

Reconhecem-se as seguintes limitações metodológicas:

Generalização: O *framework* foi desenvolvido e validado no contexto específico da UNEMAT, podendo requerer adaptações para aplicação em outras IES com características distintas.

Dependência de Dados de Qualidade: A eficácia das análises depende da qualidade e completude dos dados de entrada cadastrados no sistema.

Variabilidade da IA: Apesar das estratégias de controle, modelos de IA generativa podem apresentar variações em suas respostas, mitigadas parcialmente pelos testes de consistência.

Escopo Temporal: A validação concentra-se em dados de ciclos avaliativos recentes, não contemplando análises longitudinais de longo prazo.

4 CONSTRUÇÃO DOS ARTEFATOS ANALÍTICOS DO FRAMEWORK

4.1 Fundamentação: Análise de Conteúdo Documental

O desenvolvimento do *framework* requer a decomposição sistemática dos critérios de avaliação do SINAES em elementos acionáveis que possam ser processados computacionalmente e utilizados para geração de recomendações. Para tanto, adota-se a técnica de **Análise de Conteúdo**, conforme sistematização de Bardin (2016), especificamente a variante de análise categorial temática aplicada a documentos normativos.

A Análise de Conteúdo constitui um conjunto de técnicas de análise das comunicações que visa obter, por procedimentos sistemáticos e objetivos de descrição do conteúdo das mensagens, indicadores que permitam a inferência de conhecimentos relativos às condições de produção/recepção dessas mensagens (BARDIN, 1977). No contexto desta pesquisa, a "mensagem" analisada corresponde aos critérios textuais de atribuição de conceitos presentes no Instrumento de Avaliação de Cursos do INEP.

A contribuição específica da Análise de Conteúdo reside na garantia de rigor e sistematicidade na fase de estruturação do conhecimento normativo. Diferentemente de abordagens puramente computacionais que podem capturar padrões estatísticos sem compreensão semântica profunda, a Análise de Conteúdo assegura que a decomposição dos indicadores preserve o sentido normativo original e as distinções qualitativas entre níveis conceituais. Essa preservação é crítica para que as recomendações geradas pelo sistema sejam não apenas tecnicamente corretas, mas também normativamente válidas à luz das diretrizes do SINAES.

4.1.1 Matriz de Progressão Conceitual (MPC)

A principal contribuição metodológica desta pesquisa reside na proposição da **Matriz de Progressão Conceitual (MPC)**, estrutura analítica que decompõe cada indicador do SINAES em três dimensões: (a) Elementos Constitutivos; (b) Atributos de Progressão; e (c) Evidências Requeridas.

4.1.2 Processo de Decomposição

O procedimento de construção da MPC segue as etapas da Análise de Conteúdo:

Pré-análise: Leitura flutuante do Instrumento de Avaliação de Cursos, identificação do *corpus* (totalidade dos indicadores das três dimensões) e formulação das regras de codificação.

Exploração do Material: Para cada indicador, executa-se:

- I. Identificação dos elementos constitutivos presentes em todos os níveis conceituais.
- II. Extração dos atributos de progressão que diferenciam os conceitos 1-5.
- III. Inferência das evidências requeridas para comprovação.

Tratamento dos Resultados: Sistematização em formato tabular padronizado, validação cruzada e ajustes iterativos.

4.1.3 Exemplo de Aplicação

A título ilustrativo, apresenta-se na Tabela 4, a decomposição do Indicador 1.1 (Políticas institucionais no âmbito do curso):

Tabela 4 – Decomposição do Indicador 1.1 (Exemplo de Aplicação)

Conceito	Elementos Constitutivos	Atributo de Progressão	Evidências Requeridas
1	Políticas de ensino, extensão e pesquisa no PDI	Não implantação	Ausência de evidências
2	Políticas de ensino, extensão e pesquisa no PDI	Implantação limitada	Documentação parcial de implantação
3	Políticas de ensino, extensão e pesquisa no PDI	Implantação completa	Documentação completa de implantação
4	Políticas de ensino, extensão e pesquisa no PDI + Alinhamento ao perfil do egresso	Implantação com foco formativo	Documentação + Relação com perfil do egresso
5	Políticas de ensino, extensão e pesquisa no PDI + Alinhamento ao perfil do egresso + Práticas inovadoras	Excelência comprovada	Documentação + Alinhamento + Evidências de inovação

Fonte: Elaborado pelo autor (2026), com base na Análise de Conteúdo do Instrumento de Avaliação de Cursos de Graduação do INEP (2017).

Esta estrutura permite que o sistema identifique automaticamente o *gap* entre o conceito atual e o conceito desejado, fornecendo subsídios específicos para os agentes de IA gerarem recomendações contextualizadas.

4.1.4 Fundamentos e Processo de Construção

A Matriz de Progressão Conceitual (MPC) constitui o artefato central do *framework* proposto, operando como camada intermediária entre o conhecimento normativo do SINAES e as capacidades de processamento dos agentes de Inteligência Artificial generativa. Conforme delineado nas seções anteriores, a MPC foi construída mediante aplicação sistemática da técnica de Análise de Conteúdo, na modalidade categorial temática, aos critérios de atribuição de conceitos presentes no Instrumento de Avaliação de Cursos de Graduação do INEP, publicado em outubro de 2017.

A construção da MPC fundamenta-se na premissa de que os critérios textuais de avaliação do INEP, embora padronizados e sistematizados em progressão qualitativa (conceitos 1 a 5), não se encontram decompostos em formato que permita identificação automatizada de *gaps* e geração sistemática de recomendações.

Cabe ressaltar que o INEP disponibiliza dois instrumentos distintos de Avaliação de Cursos de Graduação (Presencial e a Distância): o instrumento de **Autorização**, aplicado a cursos novos que ainda não iniciaram suas atividades, e o instrumento de **Reconhecimento e Renovação de Reconhecimento**, aplicado a cursos já em funcionamento. Esta distinção é normativamente significativa, pois implica diferenças semânticas nos critérios que refletem o estágio de maturidade do curso avaliado. Em decorrência, a MPC foi construída em duas versões complementares, uma para cada tipo de instrumento, preservando as especificidades normativas de cada contexto avaliativo.

4.1.5 Corpus Analítico e Dimensões do Instrumento

O corpus submetido à Análise de Conteúdo compreende a totalidade dos indicadores distribuídos nas três dimensões avaliativas estabelecidas pelo INEP. O instrumento de Autorização contempla 55 indicadores, enquanto o instrumento de Reconhecimento e Renovação de Reconhecimento contempla 58 indicadores, organizados conforme a estrutura

apresentada no Tabela 5.

Tabela 5 – Distribuição dos indicadores por dimensão e tipo de instrumento

Dimensão	Autorização	Reconhecimento
1 - Organização Didático-Pedagógica	24	24
2 - Corpo Docente e Tutorial	15	16
3 - Infraestrutura	16	18
Total de indicadores	55	58

Fonte: Elaborado pelo autor (2026), a partir dos Instrumentos de Avaliação de Cursos do INEP (2017).

A diferença quantitativa entre os dois instrumentos decorre da presença de três indicadores exclusivos no instrumento de Reconhecimento: o indicador 2.3 (Atuação do coordenador), que avalia a gestão efetiva da coordenação em curso já em funcionamento; o indicador 3.16 (Comitê de Ética em Pesquisa - CEP); e o indicador 3.17 (Comitê de Ética na Utilização de Animais - CEUA), ambos vinculados a requisitos que somente podem ser verificados em cursos com atividades de pesquisa já implantadas.

Adicionalmente, a análise identificou que 32 indicadores no instrumento de Autorização e 34 no instrumento de Reconhecimento possuem observações que estabelecem condições de aplicabilidade, tornando-os obrigatórios apenas para determinados tipos de curso (por exemplo, licenciaturas, cursos de Direito, cursos da área de saúde) ou modalidades específicas (presencial ou a distância). Essa constatação reforça a necessidade de um mecanismo complementar de filtragem, a Matriz de Condições de Aplicabilidade (MCA), que determine, para cada curso concreto, quais indicadores são efetivamente pertinentes à avaliação.

4.1.6 Categorias Analíticas da Decomposição

Seguindo o procedimento descrito na seção 4.1.2, cada indicador foi submetido a processo de decomposição em três categorias analíticas definidas *a priori*, derivadas da estrutura textual dos próprios critérios de avaliação. Essas categorias atendem aos critérios de qualidade propostos por Krippendorff (2018) para a Análise de Conteúdo: exaustividade, exclusividade, homogeneidade, pertinência e objetividade. As três categorias são descritas a seguir.

a) Elementos Constitutivos. Compreendem os componentes obrigatórios mencionados nos critérios de avaliação, independentemente do conceito atribuído.

Representam "o quê" está sendo avaliado e permanecem constantes ao longo da progressão conceitual. No indicador 1.1 (Políticas institucionais no âmbito do curso), por exemplo, os elementos constitutivos identificados foram: políticas institucionais de ensino, extensão e pesquisa; o PDI como documento-base; e o âmbito do curso como contexto de aplicação. Esses elementos aparecem em todos os cinco critérios, variando apenas os qualificadores que os acompanham.

b) Atributos de Progressão. Constituem os qualificadores que diferenciam cada nível conceitual (1 a 5), representando o gradiente de qualidade esperado pelo INEP. A extração desses atributos revela o padrão semântico pelo qual o instrumento gradua a qualidade: o conceito 1 tipicamente descreve ausência ou não atendimento; o conceito 2 indica atendimento limitado ou parcial; o conceito 3 estabelece o patamar de suficiência; o conceito 4 incorpora elementos de diferenciação qualitativa (como alinhamento ao perfil do egresso ou consideração de características locais e regionais); e o conceito 5 exige evidências de excelência, frequentemente expressas por termos como "práticas comprovadamente exitosas ou inovadoras".

c) Evidências Requeridas. Representam a documentação ou demonstração necessária para comprovar o atendimento ao critério em cada nível. Embora nem sempre explicitadas nos textos originais do INEP, as evidências podem ser inferidas a partir dos elementos constitutivos e dos atributos de progressão. Esta categoria é particularmente relevante para a gestão institucional, pois fornece aos coordenadores de curso e membros da CPA orientações concretas sobre quais documentos, registros e práticas devem ser preparados ou demonstrados durante avaliações *in loco*.

A articulação dessas três categorias configura uma estrutura tridimensional que permite ao *framework*: (a) identificar automaticamente o *gap* entre o conceito atual e o conceito desejado para cada indicador; (b) especificar os requisitos adicionais necessários para a progressão conceitual; e (c) fornecer subsídios documentais para que os agentes de IA gerem recomendações contextualizadas e acionáveis.

4.1.7 Diferenças Semânticas entre os Instrumentos

A análise comparativa dos dois instrumentos revelou uma diferença semântica estruturante que perpassa a totalidade dos indicadores compartilhados e que possui implicações diretas para a operacionalização do *framework*. No instrumento de Autorização, os critérios são

formulados predominantemente com verbos no campo da **previsão** como "previsto", "prevê", "prevista", "planejado", refletindo que o curso ainda não iniciou suas atividades e, portanto, a avaliação incide sobre o que está documentado e projetado. Já no instrumento de Reconhecimento e Renovação de Reconhecimento, os critérios utilizam verbos no campo da **implantação** "implantado", "implementado", "constante", "adotando-se", indicando que a avaliação verifica a efetiva operacionalização das práticas.

Essa distinção não é meramente estilística. Ela determina a natureza da evidência exigida em cada contexto avaliativo. Ao passo que um curso em processo de autorização pode demonstrar conformidade por meio de documentação prospectiva (como o PPC, convênios firmados e regulamentos aprovados), um curso em reconhecimento necessita comprovar resultados concretos de suas atividades (como relatórios de implantação, registros de práticas realizadas e indicadores de desempenho).

Tabela 6 – Comparação semântica entre os instrumentos: exemplo do indicador 1.1

Conceito	Autorização	Reconhecimento
1	...não estão previstas no âmbito do curso.	...não estão implantadas no âmbito do curso.
3	...estão previstas no âmbito do curso.	...estão implantadas no âmbito do curso.
5	...pressupondo-se práticas exitosas ou inovadoras para a sua revisão.	...adotando-se práticas comprovadamente exitosas ou inovadoras para a sua revisão.

Fonte: Elaborado pelo autor (2026), a partir dos Instrumentos de Avaliação de Cursos do INEP (2017).

Observa-se, na Tabela 6, que a diferença entre "pressupondo-se" (Autorização) e "adotando-se práticas *comprovadamente* exitosas" (Reconhecimento) no conceito 5 é particularmente significativa: no primeiro caso, basta a previsão de práticas exitosas; no segundo, exige-se comprovação de que tais práticas foram efetivamente adotadas e produziram resultados. Essa nuance foi preservada na MPC e incorporada à lógica dos agentes de IA, de modo que as recomendações geradas pelo *framework* sejam semanticamente aderentes ao tipo de avaliação enfrentada pelo curso.

4.1.8 Estrutura da Matriz Desenvolvida

A aplicação do procedimento de decomposição ao *corpus* completo resultou em duas matrizes: a MPC para Autorização, com 275 registros de decomposição (55 indicadores × 5 conceitos), e a MPC para Reconhecimento, com 290 registros (58 indicadores × 5 conceitos). Cada registro contém os elementos constitutivos, o atributo de progressão específico daquele nível conceitual, a evidência requerida inferida e o critério original do INEP, mantido para fins de rastreabilidade e validação cruzada.

A título ilustrativo, o Tabela 7 apresenta a decomposição completa do indicador 1.1 (Políticas institucionais no âmbito do curso) na versão de Reconhecimento, demonstrando como as três categorias analíticas se articulam para cada nível conceitual.

Tabela 7 – Decomposição do indicador 1.1 na MPC

Conceito	Atributo de Progressão	Evidências Requeridas
1	Não implantação no âmbito do curso	Ausência de evidências de implantação das políticas no curso
2	Implantação limitada	Documentação parcial de implantação das políticas
3	Implantação plena	Documentação completa de implantação das políticas de ensino, extensão e pesquisa
4	Implantação com alinhamento ao perfil do egresso e promoção de oportunidades de aprendizagem	Documentação de implantação + evidências de alinhamento com perfil do egresso + relatórios de aprendizagem
5	Implantação com alinhamento + práticas comprovadamente exitosas ou inovadoras para revisão	Documentação completa + alinhamento + registros comprovados de práticas exitosas/inovadoras + processos de revisão

Fonte: Elaborado pelo autor (2026).

A Tabela 7 explicita o padrão de progressão identificado pela Análise de Conteúdo: os elementos constitutivos permanecem constantes (políticas de ensino, extensão e pesquisa no PDI), enquanto os atributos de progressão graduem a qualidade desde a ausência de implantação (conceito 1), passando pela implantação limitada (conceito 2) e plena (conceito 3), até a incorporação de elementos diferenciadores como o alinhamento ao perfil do egresso (conceito 4) e a comprovação de práticas exitosas ou inovadoras (conceito 5). As evidências requeridas, por sua vez, especificam a materialidade documental necessária em cada nível, fornecendo orientação concreta para a gestão.

4.1.9 Padrões Transversais Identificados

A fase de tratamento dos resultados, conforme proposta por Bardin (2016), permitiu identificar padrões transversais que permeiam o conjunto dos indicadores e que fundamentam a lógica algorítmica dos agentes de IA do *framework*. Dentre os padrões mais relevantes, destacam-se:

Padrão de progressão cumulativa. Na maioria dos indicadores, os conceitos superiores incorporam integralmente os requisitos dos conceitos inferiores, acrescidos de elementos adicionais. Assim, o conceito 4 tipicamente engloba todos os elementos do conceito 3, adicionando um diferenciador qualitativo (como contextualização local e regional, articulação com o perfil do egresso ou avaliação periódica). Esse padrão cumulativo permite que o sistema identifique, de forma determinística, quais requisitos adicionais separam o conceito atual do conceito desejado.

Padrão de excelência no conceito 5. A análise revelou que 15 indicadores em cada instrumento utilizam, no conceito 5, expressões que remetem a práticas "exitosas ou inovadoras" (no instrumento de Autorização) ou "comprovadamente exitosas ou inovadoras" (no instrumento de Reconhecimento). Esse padrão indica que a excelência, na concepção do INEP, não se limita ao cumprimento integral dos requisitos, mas demanda evidências de diferenciação institucional por meio de iniciativas que transcendam o atendimento normativo. Para o *framework*, esse padrão operacionaliza-se como um indicador qualitativo que exige dos agentes de IA a geração de recomendações que orientem a IES na documentação e sistematização de suas práticas inovadoras.

Padrão de condicionalidade. Cerca de 58% dos indicadores no instrumento de Autorização (32 de 55) e 59% no instrumento de Reconhecimento (34 de 58) possuem condições de aplicabilidade que os tornam obrigatórios apenas para subconjuntos específicos de cursos. Indicadores como o 1.8 (Estágio curricular supervisionado - relação com a rede de escolas da Educação Básica), obrigatório apenas para licenciaturas, ou o 3.15 (Núcleo de Práticas Jurídicas), exclusivo para cursos de Direito, exemplificam esse padrão. A identificação sistemática dessas condições na MPC fundamentou o desenvolvimento da Matriz de Condições de Aplicabilidade (MCA), artefato complementar que implementa a lógica de filtragem no sistema.

4.1.10 Integração da MPC com o *Framework*

No contexto da arquitetura do *framework*, a MPC opera como base de conhecimento estruturado que alimenta os três agentes especializados de IA generativa. A integração ocorre em três níveis funcionais:

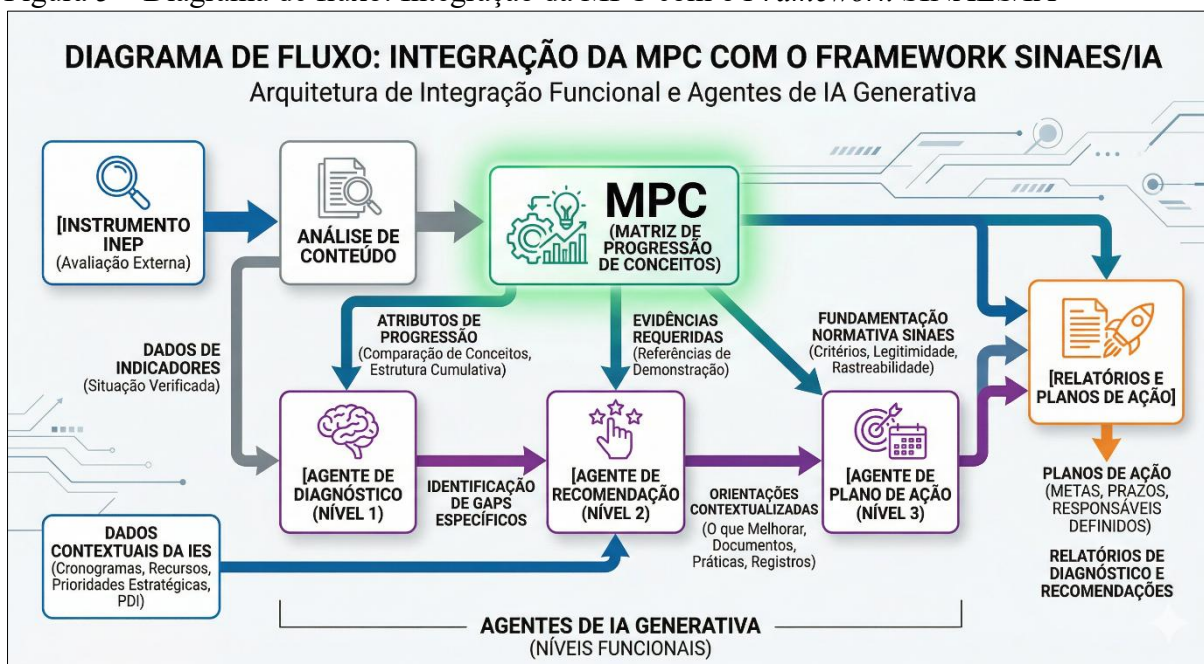
No primeiro nível, o **Agente de Diagnóstico** utiliza os atributos de progressão da MPC para comparar o conceito atual de cada indicador com os requisitos dos conceitos superiores, identificando de forma automatizada os *gaps* específicos que separam a situação verificada da situação desejada. Essa comparação é viabilizada pela estrutura cumulativa da MPC, que explicita quais elementos devem ser acrescentados para a progressão de um conceito a outro.

No segundo nível, o **Agente de Recomendação** utiliza as evidências requeridas da MPC como referência para gerar orientações contextualizadas. Ao combinar a informação sobre o que falta (proveniente do diagnóstico) com a especificação do que deve ser demonstrado (proveniente das evidências requeridas), o agente produz recomendações que indicam não apenas o que melhorar, mas quais documentos, registros e práticas devem ser preparados ou reestruturados.

No terceiro nível, o **Agente de Plano de Ação** articula os elementos da MPC com dados contextuais da IES, como cronogramas institucionais, recursos disponíveis e prioridades estratégicas definidas no PDI, para transformar as recomendações em planos de ação com metas, prazos e responsáveis definidos. Neste nível, a MPC assegura que cada ação proposta esteja normativamente fundamentada nos critérios do SINAES, conferindo legitimidade e rastreabilidade ao processo de melhoria contínua.

A Figura 5 sintetiza o fluxo de integração entre a MPC e os demais componentes do *framework*.

Figura 5 – Diagrama de fluxo: Integração da MPC com o *Framework* SINAES/IA



Fonte: Elaborado pelo autor (2026).

4.1.11 Implicações para a Gestão Institucional

A MPC desenvolvida oferece contribuições diretas para a prática da gestão de cursos de graduação em pelo menos três dimensões. Em primeiro lugar, ao decompor critérios textuais complexos em elementos discretos e operacionalizáveis, a matriz reduz a ambiguidade interpretativa que historicamente dificulta o trabalho de coordenadores de curso e membros da CPA na leitura dos instrumentos do INEP. Conforme aponta a literatura, a gestão ineficiente da autoavaliação gera desorientação informacional que impede decisões estratégicas fundamentadas (ANDRIOLA, 2009).

Em segundo lugar, a estrutura tridimensional da MPC viabiliza o que se pode denominar de **diagnóstico diferencial de indicadores**: para cada indicador, o gestor pode identificar precisamente em qual nível conceitual o curso se encontra e quais requisitos específicos separam esse nível do conceito imediatamente superior. Esse diagnóstico diferencial constitui informação de elevado valor estratégico, pois permite a priorização de ações de melhoria com base no esforço relativo necessário para a progressão conceitual.

Em terceiro lugar, a categoria de evidências requeridas funciona como *checklist* normativo para preparação de avaliações *in loco*. Ao especificar quais documentos, registros e demonstrações são necessários para cada nível conceitual, a MPC orienta a IES na organização de seu acervo documental e na sistematização de suas práticas, contribuindo para a transição de

uma cultura de conformidade reativa para uma cultura de qualidade proativa, objetivo central das políticas de avaliação do SINAES.

Dessa forma, a MPC não se configura apenas como artefato técnico de suporte computacional, mas como contribuição metodológica com aplicabilidade direta na gestão educacional, replicável a outros contextos institucionais que enfrentem desafios similares na interpretação e operacionalização dos critérios avaliativos do SINAES.

4.2 Matriz de Condições de Aplicabilidade (MCA)

A Matriz de Progressão Conceitual (MPC), descrita na seção anterior, responde ao desafio de operacionalizar computacionalmente os critérios qualitativos de atribuição de conceitos presentes no Instrumento de Avaliação de Cursos de Graduação do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Contudo, a decomposição do "como" cada indicador é avaliado constitui apenas uma das dimensões necessárias para que o *framework* produza análises contextualmente precisas. Anterior a essa etapa, coloca-se uma questão de natureza lógica igualmente crítica: determinado indicador "se aplica" ao curso analisado? A ausência de uma resposta sistemática a esse questionamento exporia o sistema ao risco de gerar recomendações baseadas em indicadores inaplicáveis à realidade específica do curso, comprometendo, em consequência, a credibilidade e a utilidade gerencial do diagnóstico produzido.

É nesse contexto que se insere a Matriz de Condições de Aplicabilidade (MCA), artefato complementar à MPC que formaliza, de maneira computacionalmente operável, as condições sob as quais cada indicador do Instrumento se aplica, ou não se aplica a um curso específico. A MCA não opera de forma autônoma: juntas, MPC e MCA formam o sistema de conhecimento normativo estruturado que alimenta os agentes de inteligência artificial do *framework* proposto. Enquanto a MPC alimenta o raciocínio avaliativo dos agentes, a MCA assegura que esse raciocínio incida exclusivamente sobre os indicadores pertinentes ao perfil do curso em análise.

A MPC operacionaliza o "como" cada indicador é avaliado, porém nem todos os indicadores do Instrumento de Avaliação de Cursos se aplicam indistintamente a qualquer curso de graduação. Variáveis como o grau acadêmico, a modalidade de ensino e disposições do PPC e das DCN condicionam a aplicabilidade de diversos indicadores, determinando situações de "Não Se Aplica" (NSA) ou ajustes no escopo da avaliação. A Matriz de Condições de

Aplicabilidade (MCA) formaliza, de maneira computacionalmente operável, as condições sob as quais cada indicador se aplica, ou não se aplica, a um curso específico. Enquanto a MPC responde ao "como", a MCA responde ao "se", constituindo uma camada de pré-filtragem que assegura que os agentes de IA processem exclusivamente os indicadores pertinentes ao perfil do curso em análise.

A construção da MCA adota a mesma Análise de Conteúdo Categórica Temática (BARDIN, 2016) empregada na MPC, preservando a coerência metodológica do sistema de conhecimento normativo. Foram definidas três categorias analíticas *a priori*:

- a) **Variáveis Determinantes:** Os atributos do curso testados pela condição de aplicabilidade, grau acadêmico, modalidade de ensino, previsões do PPC e das DCN, características de infraestrutura. Representam o "o que" condiciona a aplicabilidade.
- b) **Lógica de Avaliação:** A expressão formal da condição em proposições booleanas estruturadas no formato *IF-THEN-ELSE*. Representa o "como" a condição é computacionalmente avaliada.
- c) **Efeito sobre o Indicador:** O resultado da avaliação da regra, NSA, Aplicável ou Aplicável com Recorte de Escopo (temporal ou espacial). Representa a "consequência" para o diagnóstico.

4.2.1 Processo de Decomposição

O procedimento de construção da MCA segue as etapas da Análise de Conteúdo:

Pré-análise: Leitura flutuante das condições de aplicabilidade presentes no Instrumento de Avaliação de Cursos, identificação do corpus (totalidade das marcações NSA e ajustes de escopo) e formulação das regras de codificação.

Exploração do Material: Para cada condição de aplicabilidade, executa-se:

- a) Identificação das variáveis determinantes que condicionam a aplicação do indicador.
- b) Formalização da lógica de avaliação em estrutura IF-THEN-ELSE
- c) Classificação do efeito sobre o indicador (NSA, Aplicável, Recorte Temporal)

Tratamento dos Resultados: Sistematização em formato tabular padronizado, classificação tipológica das regras e validação cruzada com o texto normativo original.

4.2.2 Exemplo de Aplicação

A título ilustrativo, a Tabela 8 apresenta a decomposição da condição de aplicabilidade do Indicador 1.7 (Estágio curricular supervisionado):

Tabela 8 – Exemplo de regra de aplicabilidade (MCA) - Indicador 1.7

Variável Determinante	Lógica de Avaliação	Efeito
V06 – DCN prevê estágio obrigatório	IF V06 = Sim THEN Aplicável	Aplicável (obrigatório por DCN)
V05 – PPC prevê estágio	IF V06 = Não AND V05 = Sim THEN Aplicável	Aplicável (previsto no PPC)
V05 e V06 – Nenhum prevê	IF V06 = Não AND V05 = Não THEN NSA	Não Se Aplica

Fonte: Elaborado pelo autor (2026), com base na Análise de Conteúdo do Instrumento de Avaliação de Cursos de Graduação do INEP (2017) e na metodologia de Bardin (2016).

Neste exemplo, observa-se o Padrão de Hierarquia Normativa: a previsão nas DCN prevalece sobre o PPC. O motor de regras avalia primeiro a fonte normativa hierarquicamente superior, reproduzindo no plano computacional a lógica jurídica da relação entre norma geral e norma específica. Essa estrutura permite que o *framework* filtre automaticamente os indicadores inaplicáveis antes de acionar os agentes de IA evitando diagnósticos e recomendações descontextualizados.

4.2.3 Fundamentação Metodológica da MCA

A construção da MCA adotou a mesma orientação metodológica empregada na elaboração da MPC: a Análise de Conteúdo Categórica Temática, conforme sistematizada por Bardin (2016). A escolha pela continuidade metodológica não representa mera conveniência procedimental, mas atende a três exigências convergentes de rigor científico.

A primeira exigência refere-se à consistência interna do *framework*. Empregar o mesmo *framework* metodológico na construção de ambas as matrizes preserva a coerência epistemológica do sistema de conhecimento normativo, reforçando a validade interna do artefato produzido. A segunda exigência concerne ao rigor na extração das informações: os procedimentos sistemáticos de codificação propostos por Bardin (2016) garantem que todas as

condições de aplicabilidade presentes no Instrumento sejam identificadas e classificadas sem ambiguidade, evitando omissões ou interpretações idiossincráticas. A terceira, não menos relevante para um trabalho de mestrado profissional, é a replicabilidade: as regras de codificação explícitas permitem que outros pesquisadores e gestores institucionais reproduzam o processo de análise e verifiquem os resultados obtidos.

No plano operacional, a Análise de Conteúdo foi aplicada ao *corpus* normativo formado pelo conjunto das condições de aplicabilidade presentes no Instrumento de Avaliação de Cursos com os textos que determinam, para cada indicador, as situações nas quais é atribuída a marcação "Não Se Aplica" (NSA) ou realizados ajustes no escopo da avaliação. Conforme a conceituação clássica, a Análise de Conteúdo consiste em um *"conjunto de técnicas de análise das comunicações que visa obter, por procedimentos sistemáticos e objetivos de descrição do conteúdo das mensagens, indicadores [...] que permitam a inferência de conhecimentos relativos às condições de produção/recepção dessas mensagens"* (BARDIN, 2016, p. 48). No contexto da MCA, a "mensagem" analisada são os textos normativos das condições de aplicabilidade, e o "conhecimento inferido" são as regras lógicas que governam a aplicabilidade de cada indicador.

Foram definidas três categorias analíticas *a priori*, derivadas da leitura flutuante do corpus: (a) as Variáveis Determinantes, correspondentes aos atributos do curso testados pela condição como grau acadêmico, modalidade de ensino, conteúdo do PPC e das DCN, características de infraestrutura, entre outros; (b) a Lógica de Avaliação, expressa na forma de proposições booleanas estruturadas no formato *IF-THEN-ELSE*; e (c) o Efeito sobre o Indicador, que registra o resultado da avaliação da regra: NSA, Aplicável, ou Aplicável com Recorte de Escopo (temporal ou espacial).

4.2.4 Variáveis Determinantes: Estrutura e Domínios

A etapa de exploração do material, segunda fase da Análise de Conteúdo segundo Bardin (2016), permitiu identificar e catalogar as variáveis sobre as quais as condições de aplicabilidade do Instrumento operam. O processo de codificação revelou 20 variáveis determinantes, organizadas em cinco domínios conforme a natureza da fonte de informação de cada uma. A Tabela 9 a seguir apresenta a síntese dessa estrutura.

Tabela 9 – Domínios e variáveis determinantes da Matriz de Condições de Aplicabilidade

(MCA)

Domínio	Variáveis	Descrição	Qtd.	Fonte
Cadastral	V01, V02, V03	Características formais do curso: grau acadêmico, modalidade de ensino e área de conhecimento.	3	e-MEC
PPC (Projeto Pedagógico de Curso)	V04, V05, V07, V09, V11–V15, V18–V20	Elementos previstos no PPC: oferta EaD, TCC, estágio, atividades complementares, laboratórios, material didático, ambientes profissionais, integração com redes públicas, pesquisa com seres humanos e animais.	13	PPC do Curso
Normativo (DCN)	V06, V08, V10	Exigências estabelecidas nas Diretrizes Curriculares Nacionais: obrigatoriedade de TCC, estágio e atividades complementares.	3	DCN do Curso
PPC/DCN Combinado	V17	Integração com o SUS, prevista nas DCN, no PPC ou em ambos.	1	PPC e DCN
Infraestrutura	V16	Condições físicas da IES: existência de espaço individual para todos os docentes do curso.	1	Infraestrutura da IES

Fonte: Elaborado pelo autor (2026), com base na Análise de Conteúdo do Instrumento de Avaliação de Cursos de Graduação do INEP (2017) e na metodologia de Bardin (2016).

A distribuição das variáveis entre os cinco domínios reflete a própria lógica do Instrumento de Avaliação de Cursos: as condições de aplicabilidade ancoram-se, predominantemente, nas características pedagógicas declaradas no Projeto Pedagógico do Curso (PPC), responsável por 13 das 20 variáveis identificadas. Essa concentração é metodologicamente significativa, pois reforça a centralidade do PPC como documento-referência na avaliação do SINAES, na medida em que suas disposições determinam não apenas o que é avaliado (MPC), mas também quais indicadores são pertinentes à avaliação (MCA).

O domínio Cadastral, composto pelas variáveis V01 (Grau Acadêmico), V02 (Modalidade de Ensino) e V03 (Área de Conhecimento), opera como filtro estrutural primário: são as variáveis mais universalmente verificadas, pois características como o grau acadêmico e a modalidade do curso condicionam diretamente um conjunto expressivo de regras de aplicabilidade. O domínio Normativo (DCN) introduz uma camada de hierarquia regulatória: em determinadas regras, a previsão nas Diretrizes Curriculares Nacionais prevalece sobre a disposição do PPC, o que exige do sistema a capacidade de distinguir a fonte normativa de cada

obrigação. Por sua vez, a variável V16 (Espaço Individual para Todos os Docentes), único representante do domínio Infraestrutura, ilustra como condições de natureza estrutural da IES também incidem sobre a aplicabilidade de indicadores específicos.

A Tabela 10 a seguir relaciona o catálogo completo de variáveis que compõem a Matriz de Condições de Aplicabilidade (MCA).

Tabela 10 – Catálogo completo de variáveis que compõem a MCA

Cód.	Variável	Descrição	Tipo	Fonte
V01	Grau Acadêmico	Tipo de diploma conferido pelo curso	Cadastral	e-MEC / Cadastro do Curso
V02	Modalidade de Ensino	Forma predominante de oferta do curso	Cadastral	e-MEC / Cadastro do Curso
V03	Área de Conhecimento	Área específica do curso para regras setoriais	Cadastral	e-MEC / Tabela CAPES
V04	Oferta de Disciplinas EaD	Curso presencial que oferta disciplinas na modalidade EaD (Portaria 1.134/2016)	PPC	PPC do Curso
V05	Previsão de TCC no PPC	O PPC contempla Trabalho de Conclusão de Curso	PPC	PPC do Curso
V06	Previsão de TCC nas DCN	As DCN do curso prevêem obrigatoriamente TCC	Normativo	DCN do Curso
V07	Previsão de Estágio no PPC	O PPC contempla Estágio Curricular Supervisionado	PPC	PPC do Curso
V08	Previsão de Estágio nas DCN	As DCN do curso prevêem obrigatoriamente estágio	Normativo	DCN do Curso
V09	Previsão de Ativ. Complementares no PPC	O PPC contempla Atividades Complementares	PPC	PPC do Curso
V10	Previsão de Ativ. Complementares nas DCN	As DCN prevêem obrigatoriamente atividades complementares	Normativo	DCN do Curso
V11	Utilização de Laboratórios de Form. Básica	O PPC prevê uso de laboratórios didáticos de formação básica	PPC	PPC do Curso
V12	Utilização de Laboratórios de Form. Específica	O PPC prevê uso de laboratórios didáticos de formação específica	PPC	PPC do Curso
V13	Previsão de	O PPC contempla material didático	PPC	PPC do

Cód.	Variável	Descrição	Tipo	Fonte
	Material Didático no PPC	institucional (para presenciais)		Curso
V14	Previsão de Ambientes Profissionais (EaD)	O PPC de curso EaD prevê utilização de ambientes profissionais	PPC	PPC do Curso
V15	Atividades Presenciais na Sede (EaD)	Curso EaD prevê atividades presenciais na sede da IES	PPC	PPC do Curso
V16	Espaço Individual para Todos os Docentes	A IES possui espaço de trabalho individual para todos os docentes do curso	Infraestrutura	Infraestrutura da IES
V17	Integração com SUS/Sistema de Saúde	Curso da área de saúde com integração prevista nas DCN/PPC com o SUS	PPC/DCN	PPC e DCN do Curso
V18	Integração com Redes Públicas de Ensino	O PPC prevê integração com as redes públicas de ensino da educação básica	PPC	PPC do Curso
V19	Pesquisa com Seres Humanos	O PPC contempla realização de pesquisa envolvendo seres humanos	PPC	PPC do Curso
V20	Pesquisa com Animais	O PPC contempla utilização de animais em pesquisas	PPC	PPC do Curso

Fonte: Elaborado pelo autor (2026), com base na Análise de Conteúdo do Instrumento de Avaliação de Cursos de Graduação do INEP (2017).

4.2.5 Tipologia e Padrões Lógicos das Regras de Aplicabilidade

O processo de codificação e classificação, etapa analítica central da Análise de Conteúdo categorial, resultou na identificação de 25 regras de aplicabilidade decompostas e na elaboração de uma tipologia com 9 tipos distintos de regras. Essa tipologia não é uma classificação meramente descritiva: ela constitui o vocabulário de padrões computacionais que o motor de regras da MCA (detalhado na seção 4.2.5) precisa ser capaz de interpretar e avaliar. A Tabela 11 a seguir apresenta a tipologia completa.

Tabela 11 – Tipologia das regras de aplicabilidade da MCA

Tipo de Regra	Descrição	Qtd.
RECORTE_TEMPORAL	Ajusta o período avaliado (acervo bibliográfico, laboratórios, espaços docentes) conforme o grau acadêmico:	5

Tipo de Regra	Descrição	Qtd.
	1.º ano para CST; dois primeiros anos para bacharelados e licenciaturas.	
NSA_SIMPLES	Não Se Aplica de forma direta, baseado em uma única variável determinante.	2
NSA_CONDICIONAL	Não Se Aplica mediante combinação de duas ou mais variáveis avaliadas conjuntamente.	6
OBRIGATORIEDADE_HIERARQUICA	As Diretrizes Curriculares Nacionais (DCN) prevalecem sobre o PPC na determinação da obrigatoriedade do indicador (TCC, estágio, atividades complementares).	4
OBRIGATORIEDADE_SETORIAL	Obrigatório apenas para cursos de áreas de conhecimento específicas (ex.: saúde, Direito).	2
OBRIGATORIEDADE_CONDICIONAL	Obrigatoriedade dependente da combinação de condições cadastrais e normativas.	2
OBRIGATORIEDADE_EXCLUSIVA	Indicador exclusivo para um único tipo de curso (ex.: licenciatura), com atribuição automática de NSA para todos os demais.	1
OBRIGATORIEDADE_PPC	Obrigatório quando previsto no PPC, independentemente da área de conhecimento.	2
EXCLUSIVIDADE	Indicador exclusivo para modalidade ou configuração específica (ex.: EaD com ambientes profissionais previstos).	2

Fonte: Elaborado pelo autor (2026), com base na Análise de Conteúdo do Instrumento de Avaliação de Cursos de Graduação do INEP (2017).

Da análise da tipologia, depreende-se que a maior complexidade lógica se concentra nas regras do tipo NSA_CONDICIONAL (6 ocorrências) e OBRIGATORIEDADE_HIERARQUICA (4 ocorrências). As primeiras exigem a avaliação conjunta de duas ou mais variáveis, o que requer operadores lógicos de conjunção (AND) ou disjunção (OR) nas expressões computacionais. As segundas introduzem um mecanismo de precedência normativa: quando as Diretrizes Curriculares Nacionais estabelecem a obrigatoriedade de determinada atividade como TCC, estágio supervisionado ou atividades complementares, essa determinação prevalece sobre qualquer disposição contrária ou omissão no PPC do curso. Tal hierarquia reproduz, no plano computacional, a lógica jurídica da relação entre norma geral (DCN) e norma específica (PPC).

A categoria `RECORTE_TEMPORAL` merece destaque analítico particular. Diferentemente dos demais tipos, que produzem decisões binárias (aplica ou não aplica), o `RECORTE_TEMPORAL` não elimina o indicador da análise, mas ajusta o escopo do que deve ser avaliado: para Cursos Superiores de Tecnologia (CST), considera-se apenas o primeiro ano do curso; para bacharelados e licenciaturas, os dois primeiros anos. Essa distinção tem implicações práticas diretas para o *framework*, pois o contexto fornecido ao agente de inteligência artificial precisa incorporar esse recorte sob pena de que as recomendações geradas se refiram a um conjunto de evidências temporalmente inadequado ao perfil do curso.

Do ponto de vista da estrutura lógica subjacente às regras, a Análise de Conteúdo permitiu identificar cinco padrões formais distintos, apresentados na Tabela 12 a seguir.

Tabela 12 – Padrões lógicos das regras de aplicabilidade da MCA

Padrão Lógico	Estrutura IF-THEN-ELSE	Exemplo de Regra
Padrão 1 – Teste Simples	IF [variável] = [valor] THEN [efeito]	R10: IF V01 = Licenciatura THEN NSA
Padrão 2 – Teste Composto (AND)	IF [var_A] = [valor] AND [var_B] = [valor] THEN [efeito]	R06: IF V02 = EaD AND V14 = Sim THEN aplica
Padrão 3 – Teste Composto (OR)	IF [var_A] = [valor] OR [var_B] = [valor] THEN [efeito]	R07: IF V02 = EaD OR (V02 = Presencial AND V04 = Sim) THEN aplica
Padrão 4 – Hierarquia Normativa	IF DCN prevê THEN obrigatório; ELSE IF PPC não prevê THEN NSA; ELSE aplica	R16: DCN prevalece sobre PPC na obrigatoriedade do TCC
Padrão 5 – Recorte Temporal Condicional	IF V01 = CST THEN recorte = 1.º ano; ELSE recorte = 2 primeiros anos	R01: Ajusta escopo do acervo bibliográfico básico

Fonte: Elaborado pelo autor (2026).

Os Padrões 1 a 3 correspondem a estruturas lógicas elementares, tratáveis por operadores booleanos simples. O Padrão 4 - Hierarquia Normativa é o de maior relevância para a fidelidade do sistema ao ordenamento jurídico-normativo da avaliação: a estrutura em cascata (IF DCN... ELSE IF PPC... ELSE...) espelha a lógica de prevalência das DCN sobre o PPC, garantindo que o sistema reproduza corretamente a relação de subordinação entre os documentos normativos. O Padrão 5 - Recorte Temporal Condicional, por sua vez, opera não sobre a aplicabilidade do indicador em si, mas sobre o escopo temporal da evidência a ser analisada, o que o distingue fundamentalmente dos demais padrões e demanda tratamento

diferenciado no motor de regras.

4.2.6 A MCA como Camada de Pré-Filtragem Contextual dos Agentes de IA

A integração entre a MCA e os agentes de inteligência artificial do *framework* opera em dois planos distintos e complementares, ambos orientados a maximizar a precisão e a relevância das análises geradas.

O primeiro plano é a **filtragem prévia de indicadores**. Antes de qualquer processo de geração de recomendações, o agente responsável pelo diagnóstico recebe não a lista integral dos indicadores do Instrumento de Avaliação, mas apenas o subconjunto daqueles cuja aplicabilidade ao curso analisado foi confirmada pelo motor de regras da MCA. Essa filtragem produz benefícios operacionais e qualitativos: do ponto de vista operacional, reduz o volume de tokens processados em cada chamada aos modelos de linguagem de grande escala (LLM), otimizando o custo computacional da análise; do ponto de vista qualitativo, elimina o risco de que o agente gere recomendações referentes a dimensões irrelevantes para o curso, por exemplo, sugerir melhorias em laboratórios de formação básica a um curso que, por característica pedagógica declarada no PPC, não os utiliza.

O segundo plano é a **contextualização enriquecida do *prompt***. As condições específicas identificadas pela MCA, em especial os recortes temporais e as regras de hierarquia normativa, são incorporadas ao contexto fornecido ao agente. Assim, quando um CST é submetido à análise, o agente recebe a instrução explícita de considerar apenas o acervo bibliográfico relativo ao primeiro ano do curso; quando um indicador de estágio supervisionado é obrigatório por força de DCN, essa informação é destacada no contexto, permitindo que as recomendações reflitam a natureza regulatória da exigência. Essa contextualização reproduz, no ambiente dos agentes de IA, o discernimento que um avaliador institucional experiente exerceria ao interpretar as condições do Instrumento.

Essa arquitetura de pré-filtragem contextual representa uma contribuição técnica relevante do *framework* para a área de aplicação de inteligência artificial à gestão educacional. A MCA, ao introduzir uma camada de raciocínio condicional anterior à geração de análises, eleva o nível de especificidade e de acurácia contextual das recomendações produzidas aproximando-as da qualidade de um parecer especializado e distanciando-as das limitações dos sistemas de recomendação genérica.

4.2.7 Implementação Técnica: Modelo de Dados e Motor de Regras

A operacionalização computacional da MCA no sistema *framework*, requer três componentes estruturais interdependentes: o modelo de dados relacional, o motor de regras (*Rule Engine*) e a interface de cadastro do perfil de aplicabilidade do curso.

O modelo de dados é composto por três tabelas fundamentais. A tabela MCA_VARIAVEIS armazena o catálogo completo das 20 variáveis determinantes, com seus códigos, domínios, tipos de dado (enumerado ou booleano) e fontes de informação. A tabela MCA_REGRAS registra as 25 regras decompostas, cada qual associada aos indicadores correspondentes e contendo a expressão lógica em formato JSON estruturado, o tipo da regra conforme a tipologia definida e o efeito esperado (NSA, Aplicável ou Recorte Temporal com especificação do período). A tabela MCA_CURSO_PERFIL armazena, para cada curso cadastrado no *tenant*, os valores atribuídos às variáveis determinantes, alimentados pelo coordenador do curso por meio de um formulário de cadastro guiado pelo sistema.

O motor de regras é implementado como uma *Service Class*, responsável por: (1) receber o identificador do curso e carregar os valores das variáveis a partir de MCA_CURSO_PERFIL; (2) iterar sobre todos os indicadores do Instrumento de Avaliação; (3) avaliar, para cada indicador, as regras de aplicabilidade vinculadas conforme as expressões lógicas armazenadas; e (4) retornar o perfil completo de aplicabilidade do curso, discriminando, para cada indicador, o resultado: APLICA, NSA ou APLICA_COM_RECORTE (com a especificação do período pertinente). Esse perfil é disponibilizado como insumo tanto para o módulo de diagnóstico quanto para o contexto de *prompt* dos agentes de IA.

O formulário de cadastro do perfil de aplicabilidade é organizado por domínio de variáveis, com lógica condicional que exhibe apenas as perguntas pertinentes ao perfil do curso sendo cadastrado. Por exemplo, se o curso é identificado como presencial (V02 = Presencial), as perguntas relativas a atividades presenciais na sede (V15), variável relevante apenas para cursos EaD, são automaticamente ocultadas. Essa abordagem de *interface adaptativa* reduz a carga cognitiva sobre o coordenador e minimiza o risco de preenchimento inconsistente, que comprometeria a precisão das regras de aplicabilidade.

4.2.8 Contribuição da MCA para a Qualidade do Diagnóstico Institucional

A Matriz de Condições de Aplicabilidade representa, na arquitetura do *framework*,

a camada lógica que antecede a avaliação propriamente dita. Sua ausência implicaria que o sistema trataria todos os indicadores do SINAES como universalmente aplicáveis a qualquer curso, situação que, na prática avaliativa real, não corresponde à operação do Instrumento de Avaliação do INEP. A imprecisão decorrente dessa generalização indevida produziria diagnósticos que incluiriam análises de dimensões inexistentes na realidade do curso, recomendações descontextualizadas e, conseqüentemente, planos de ação com baixo grau de aderência à situação institucional específica.

Do ponto de vista da gestão educacional, a MCA contribui para o que Simon (1976) denomina racionalidade procedimental: ao estruturar os procedimentos de análise segundo regras explícitas e sistematicamente aplicadas, o *framework* reduz a dependência de julgamentos individuais dos gestores sobre quais indicadores são pertinentes ao curso, substituindo interpretações *ad hoc* por critérios formalizados e consistentes com a normativa do SINAES. Essa formalização é especialmente relevante em contextos institucionais nos quais a rotatividade de coordenadores de curso implica frequente perda de memória organizacional sobre as especificidades do Instrumento de Avaliação.

A decomposição realizada nesta pesquisa identificou 20 variáveis determinantes, 25 regras de aplicabilidade e 9 tipos distintos de padrões lógicos. Esse conjunto, combinado com a MPC, que decompõe os critérios qualitativos de atribuição de conceitos, forma a base de conhecimento normativo completa que confere ao *framework* capacidade de gerar análises e recomendações verdadeiramente contextualizadas para cada curso específico. Conforme a fundamentação metodológica adotada, a utilização da Análise de Conteúdo Categórica Temática (BARDIN, 2016) como metodologia de decomposição garante o rigor científico necessário ao trabalho acadêmico, ao mesmo tempo em que produz um artefato computacionalmente operável, atendendo, assim, ao duplo critério de qualidade preconizado pelo *Design Science Research*: relevância para a prática e rigor para a ciência (HEVNER *et al.*, 2004).

4.3 Matriz de Compartilhamento de Evidências (MCE)

A construção da Matriz de Progressão Conceitual (MPC), produziu a categoria de **Evidências Requeridas** para cada nível conceitual de cada indicador do instrumento de avaliação. Ao operacionalizar essa categoria no contexto de uma IES *multicampi* que oferta dezenas de cursos de graduação, identificou-se uma questão estruturante para a arquitetura do *framework*: nem todas as evidências exigidas pelo INEP são exclusivas de um curso. Parcela

significativa dos documentos, registros e demonstrações que comprovam o atendimento a determinados indicadores são, por natureza, institucionais, ou seja, um único documento serve como evidência para múltiplos cursos avaliados simultaneamente.

Essa constatação possui implicações diretas para a gestão. Quando a IES não identifica sistematicamente quais evidências são compartilháveis, três problemas recorrentes se manifestam: (a) **retrabalho**, pois diferentes coordenadores produzem documentos redundantes para comprovar o mesmo fato institucional; (b) **inconsistência**, pois versões divergentes de um mesmo documento podem ser apresentadas em avaliações de cursos distintos; e (c) **subaproveitamento de ativos documentais**, pois a IES pode já dispor de evidências robustas em nível institucional sem que coordenadores de curso tenham conhecimento ou acesso a elas.

Conforme observam Griboski, Peixoto e Hora (2018), a desarticulação entre os instrumentos de gestão institucional compromete a eficácia dos processos avaliativos e reduz a relevância do PDI como eixo estruturante da qualidade. No mesmo sentido, a literatura sobre autoavaliação institucional (ANDRIOLA, 2009; BITTENCOURT *et al.*, 2008) destaca que a fragmentação da informação entre setores e cursos constitui um dos principais obstáculos à construção de uma cultura de qualidade nas IES brasileiras.

Diante dessa problemática, desenvolveu-se um artefato complementar à MPC: a **Matriz de Compartilhamento de Evidências (MCE)**, que categoriza sistematicamente cada indicador do instrumento de avaliação segundo a natureza institucional, parcial ou específica do curso das evidências que o sustentam.

4.3.1 Processo de Classificação

O procedimento de construção da MCE segue análise semântica dos indicadores:

Pré-análise: Leitura flutuante dos indicadores dos dois instrumentos de avaliação (Autorização e Reconhecimento/Renovação), identificação do corpus (113 indicadores no total) e definição dos critérios de classificação segundo a taxonomia de escopos.

Exploração do Material: Para cada indicador, executa-se:

- a) Análise da natureza dos documentos comprobatórios tipicamente exigidos em avaliações *in loco*.
- b) Verificação se a evidência é de produção e manutenção institucional, parcial ou exclusiva do curso.
- c) Atribuição do escopo correspondente segundo a taxonomia definida

Tratamento dos Resultados: Sistematização em formato tabular com distribuição por escopo e por dimensão avaliativa, validação cruzada entre os dois instrumentos.

4.3.2 Exemplo de Aplicação

A título ilustrativo, a tabela 13 apresenta a classificação de três indicadores representativos de cada escopo:

Tabela 13 – Exemplo de classificação de escopo de compartilhamento de evidências (MCE)

Indicador	Escopo	Evidência Típica	Implicação para a Gestão
1.12 - Apoio ao discente	Institucional	Programa institucional de acolhimento, Núcleo de Acessibilidade, apoio psicopedagógico	Cadastro único pela IES; vinculação automática a todos os cursos
3.6 - Bibliografia básica	Parcial	Acervo da biblioteca e assinaturas digitais (institucional) + relação título-componente curricular (do curso)	Camada institucional herdada; coordenação complementa a relação título-UC
1.3 - Perfil do egresso	Curso	Descrição no PPC conforme DCN específicas da área	Produção exclusiva pelo NDE e coordenação do curso

Fonte: Elaborado pelo autor (2026), com base na análise dos Instrumentos de Avaliação de Cursos de Graduação do INEP (2017).

Essa classificação permite que o *framework* diferencie, nos relatórios e planos de ação, entre recomendações de vinculação (quando a evidência institucional já existe e precisa apenas ser associada ao acervo do curso) e recomendações de produção (quando a evidência efetivamente não existe e precisa ser criada), discriminando entre lacunas reais de qualidade e lacunas meramente documentais.

4.3.3 Taxonomia de Escopos

A classificação adota uma taxonomia de três níveis, definida a partir da análise semântica dos critérios de avaliação e da natureza dos documentos que tipicamente compõem o acervo comprobatório das IES em avaliações *in loco*. Os três escopos são:

a) Escopo institucional. A evidência é produzida, mantida e válida em nível

institucional, servindo integralmente para todos os cursos da IES sem necessidade de adaptação ou complemento. O documento comprobatório é único e sua apresentação em uma avaliação de curso específico dispensa qualquer contextualização adicional. Exemplos típicos incluem: salas de aula e seus laudos de acessibilidade (indicador 3.4), laboratórios de informática (indicador 3.5), o Ambiente Virtual de Aprendizagem (indicador 1.17), o programa institucional de apoio ao discente (indicador 1.12), o biotério (indicador 3.13), o Comitê de Ética em Pesquisa (indicador 3.16 do instrumento de Reconhecimento) e os espaços de trabalho para docentes em tempo integral (indicador 3.1).

b) Escopo parcial. A evidência possui uma camada institucional compartilhável e uma camada específica do curso. O documento-base é institucional, mas requer complemento ou contextualização que demonstre sua aplicação no âmbito do curso avaliado. Este é o escopo que demanda maior atenção da gestão, pois exige articulação entre setores institucionais e coordenações de curso. Exemplos incluem: políticas de ensino, extensão e pesquisa (indicador 1.1), cujo PDI é institucional, mas a materialização no PPC é do curso; os convênios de estágio (indicador 1.7), que podem ser compartilhados entre cursos, mas cuja distribuição de vagas e orientação são específicas; e a bibliografia básica (indicador 3.6), cujo acervo e assinaturas digitais são institucionais, mas a relação título-componente curricular é específica de cada curso.

c) Escopo do curso. A evidência é inteiramente produzida e válida apenas no contexto de um curso específico, sem possibilidade de reaproveitamento por outros cursos. São documentos cuja natureza está intrinsecamente vinculada ao PPC, às DCN e às particularidades da formação oferecida. Exemplos incluem: objetivos do curso (indicador 1.2), perfil profissional do egresso (indicador 1.3), estrutura curricular (indicador 1.4), conteúdos curriculares (indicador 1.5), metodologia (indicador 1.6), o NDE (indicador 2.1), o colegiado de curso (indicador 2.11) e os laboratórios de formação específica (indicador 3.9).

4.3.4 Resultados da Classificação

A aplicação da taxonomia ao conjunto completo de indicadores dos dois instrumentos de avaliação do INEP produziu a distribuição apresentada na Tabela 14. A classificação foi realizada para ambos os instrumentos de Autorização (55 indicadores) e Reconhecimento e Renovação de Reconhecimento (58 indicadores) e considerou as observações de aplicabilidade presentes em cada indicador.

Tabela 14 – Distribuição dos indicadores por escopo de compartilhamento de evidências

Escopo da Evidência	AUT (n)	AUT (%)	REC (n)	REC (%)	Implicação
Institucional	15	27%	17	29%	Reuso total
Parcial	21	38%	21	36%	Reuso + complemento
Curso	19	35%	20	34%	Produção exclusiva
Total	55	100%	58	100%	

Fonte: Elaborado pelo autor (2026), a partir da análise dos Instrumentos de Avaliação de Cursos do INEP (2017).

Os dados revelam que aproximadamente dois terços dos indicadores (65% no instrumento de Autorização e 66% no de Reconhecimento) possuem evidências total ou parcialmente compartilháveis entre cursos. Essa constatação é estrategicamente relevante: significa que uma IES que sistematize a gestão de seu acervo documental institucional pode reduzir significativamente o esforço de preparação para avaliações *in loco*, ao mesmo tempo em que garante consistência na documentação apresentada.

4.3.5 Análise por Dimensão Avaliativa

A distribuição dos escopos de compartilhamento não é uniforme entre as três dimensões avaliativas do INEP. A análise dimensional revela padrões distintos que refletem a natureza dos objetos avaliados em cada dimensão.

A **Dimensão 1 - Organização Didático-Pedagógica** concentra a maior proporção de indicadores com escopo de curso (como objetivos, perfil do egresso, estrutura curricular, conteúdos e metodologia), o que é coerente com o fato de que essa dimensão avalia elementos constitutivos do PPC, documento inerentemente específico de cada curso. Não obstante, indicadores como o apoio ao discente (1.12) e o AVA (1.17) são de natureza institucional, e vários outros (políticas institucionais, convênios de estágio, atividades complementares) possuem camada institucional relevante.

A **Dimensão 2 - Corpo Docente e Tutorial** apresenta uma particularidade significativa: o currículo Lattes de cada docente é um documento único, independentemente de quantos cursos o professor integre. Assim, indicadores relativos à experiência profissional (2.6), à experiência em docência superior (2.8), na educação básica (2.7) e em EaD (2.9) possuem escopo institucional, o mesmo currículo comprova o requisito para todos os cursos. Contudo, o

percentual de titulação (2.4) e a produção científica (2.15) são de escopo parcial, pois, embora os diplomas e publicações sejam os mesmos, o cálculo do percentual é realizado sobre o corpo docente específico de cada curso.

A **Dimensão 3 - Infraestrutura** é a que apresenta maior concentração de indicadores institucionais. Espaços físicos como salas de aula (3.4), sala de professores (3.3), laboratórios de informática (3.5) e biotérios (3.13) são, por definição, recursos compartilhados entre múltiplos cursos de um mesmo *campus*. Os laudos de acessibilidade, os inventários de equipamentos e os planos de manutenção dessas instalações são documentos únicos que servem como evidência para todos os cursos que utilizam aqueles espaços. Essa característica é particularmente relevante para universidades *multicampi* como a UNEMAT, em que a mesma infraestrutura pode sustentar avaliações de diversos cursos em um mesmo período.

4.3.6 Exemplos de Aplicação Prática

Para ilustrar a aplicabilidade concreta da classificação, consideram-se três cenários representativos dos diferentes escopos.

Cenário 1 - Escopo institucional: Apoio ao discente (indicador 1.12). Programas de acolhimento e permanência, o Núcleo de Acessibilidade, o serviço de apoio psicopedagógico e os programas de monitoria e nivelamento são, na maioria das IES, estruturas institucionais que atendem indistintamente a todos os cursos. Quando o *framework* identifica que um determinado curso necessita comprovar ações de apoio ao discente para atingir o conceito 3, o Agente de Recomendação pode orientar: "a IES já dispõe de Programa Institucional de Acolhimento e Permanência, vincule este documento ao acervo do curso e solicite relatório de atendimento setorial".

Cenário 2 - Escopo parcial: Bibliografia básica (indicador 3.6). O acervo da biblioteca, as assinaturas de bases digitais e o programa de aquisição são institucionais e servem para todos os cursos. Entretanto, a relação entre títulos e componentes curriculares é específica de cada curso. Nesse caso, o *framework* orienta a coordenação a produzir apenas o complemento necessário, a planilha de correspondência título-UC, sem necessidade de recompilar informações sobre o acervo geral que a biblioteca já possui.

Cenário 3 - Escopo do curso: Perfil profissional do egresso (indicador 1.3). Não há possibilidade de compartilhamento. O perfil é definido no PPC em função das DCN específicas da área, das competências próprias da formação e das demandas regionais. Nesse

caso, o *framework* direciona integralmente a responsabilidade de produção da evidência ao NDE e à coordenação do curso, sem sugerir qualquer reaproveitamento de documentos de outros cursos.

4.3.7 Integração com a Arquitetura do *Framework*

A classificação de escopo de compartilhamento integra-se à arquitetura do *framework* em três pontos críticos de processamento:

No nível da **base de dados**, o sistema implementa arquitetura *multi-tenant* que distingue entre dados institucionais (compartilhados entre todos os cursos da IES assinante) e dados do curso (específicos de cada coordenação). O campo *escopo_evidencia* na tabela de MCE_INDICADORES determina em qual camada cada evidência deve ser armazenada. Documentos com escopo institucional são cadastrados uma única vez pela administração da IES e automaticamente vinculados a todos os cursos. Documentos com escopo de curso são cadastrados exclusivamente pela coordenação responsável. Documentos com escopo parcial possuem registro duplo: a camada institucional é herdada e a camada do curso é complementada localmente.

No nível dos **agentes de IA**, a classificação de escopo é incorporada ao *prompt* contextual dos agentes de Diagnóstico e Recomendação. Quando o Agente de Diagnóstico identifica um *gap* em indicador com escopo institucional, o Agente de Recomendação é instruído a verificar primeiro se a evidência já existe na camada institucional antes de sugerir sua produção ao coordenador. Essa lógica evita recomendações redundantes e direciona o esforço do gestor exclusivamente para as evidências que dependem de sua atuação.

No nível dos **relatórios e planos de ação**, a classificação permite ao *framework* gerar dois tipos distintos de recomendação: (a) recomendações de *vinculação*, quando a evidência institucional existe, mas não está formalmente associada ao acervo do curso, nesse caso, a ação requerida é administrativa, não substantiva; e (b) recomendações de *produção*, quando a evidência efetivamente não existe e precisa ser criada, nesse caso, o plano de ação especifica responsáveis, prazos e entregas. Essa diferenciação permite que os relatórios gerados pelo *framework* discriminem entre lacunas reais de qualidade e lacunas meramente documentais, informação de elevado valor para a gestão estratégica da IES.

4.3.8 Contribuição para a Gestão Institucional

A Matriz de Compartilhamento de Evidências, embora conceitualmente simples, endereça uma fragilidade estrutural recorrente na gestão de IES brasileiras. Conforme documentado por Bittencourt *et al.* (2008), a articulação entre setores institucionais e coordenações de curso constitui fator crítico para o sucesso nas avaliações *in loco*. A ausência de mecanismos que explicitem quais documentos são compartilhados e quais são exclusivos resulta em esforços duplicados, lacunas não identificadas e perda de oportunidades de demonstração de qualidade institucional.

No contexto da UNEMAT, a aplicabilidade dessa classificação é amplificada pela natureza *multicampi* da instituição. Cursos oferecidos em diferentes *campi* compartilham infraestrutura localizada (como laboratórios e bibliotecas), políticas definidas no PDI e programas institucionais de apoio ao discente, ao mesmo tempo em que mantêm PPC, NDE e colegiados próprios. O *framework*, ao incorporar a lógica de compartilhamento, permite que a gestão institucional identifique, para cada período avaliativo, quais documentos já estão disponíveis e precisam apenas ser vinculados aos cursos em avaliação, e quais documentos demandam produção por parte de cada coordenação.

Essa capacidade de discriminação entre vinculação e produção de evidências transforma a preparação para avaliações *in loco* de um processo reativo e fragmentado em um processo planejado e coordenado, objetivo alinhado à proposta de Andriola (2009) de que a autoavaliação institucional deve constituir o ponto de partida para o planejamento estratégico e para a gestão universitária. O *framework* operacionaliza esse princípio ao traduzir a classificação de escopo em ações concretas, rastreáveis e atribuíveis, contribuindo para a consolidação de uma cultura de qualidade baseada em evidências nas IES atendidas.

5 DESENVOLVIMENTO DO *FRAMEWORK*

5.1 Arquitetura e Engenharia de *Software* da Aplicação

O capítulo anterior apresentou os três artefatos analíticos que fundamentam o *framework*: a Matriz de Progressão Conceitual (MPC), a Matriz de Condições de Aplicabilidade (MCA) e a Matriz de Compartilhamento de Evidências (MCE). Este capítulo descreve a materialização desses artefatos em um sistema computacional funcional, detalhando as decisões arquiteturais, os componentes de *software*, o modelo de dados e os mecanismos de processamento que viabilizam a operacionalização do *framework* proposto.

A construção da aplicação orientou-se por dois princípios fundamentais derivados do contexto da pesquisa. O primeiro consiste na aderência ao modelo *multi-tenant*, necessário para que múltiplas IES possam utilizar o *framework* de forma isolada e segura em uma única instância da plataforma. O segundo consiste na modularidade arquitetural, que permite a evolução independente dos componentes, instrumentos normativos do SINAES, dados acadêmicos institucionais e camada de inteligência artificial, sem comprometer a integridade do sistema como um todo.

Conforme argumentam Hevner *et al.* (2004), o rigor na construção de artefatos tecnológicos em pesquisas de *Design Science Research* demanda não apenas a descrição funcional do sistema, mas também a explicitação das decisões de *design* que fundamentam sua arquitetura. Esta seção atende a esse requisito ao apresentar, de forma sistemática, a pilha tecnológica selecionada, a organização modular do código-fonte, a estratégia de isolamento de dados por *tenant*, o modelo relacional do domínio SINAES e os serviços computacionais que implementam a lógica de negócio do *framework*.

5.1.1 Pilha Tecnológica

A seleção das tecnologias para implementação do *framework* considerou critérios de maturidade, escalabilidade, adequação ao domínio educacional e disponibilidade de ecossistema de bibliotecas e ferramentas. A Tabela 15 apresenta a composição tecnológica adotada.

Tabela 15 – Composição tecnológica adotada

Backend	PHP 8.4 / Laravel 12	<i>Framework</i> web maduro com ecossistema robusto para aplicações <i>SaaS</i> , suporte nativo a filas, cache, broadcasting e ORM <i>Eloquent</i> .
Banco de Dados	PostgreSQL 15	Suporte a tipos complexos (JSONB, <i>arrays</i>), <i>schemas</i> múltiplos, <i>full-text search</i> via <i>tvector</i> e índices GIN para consultas em campos semiestruturados.
Cache e Filas	Redis	Armazenamento em memória para cache de sessões, filas de processamento assíncrono e comunicação em tempo real via <i>pub/sub</i> .
Processamento Assíncrono	Laravel Horizon	Painel de monitoramento e gerenciamento de filas Redis, com controle de concorrência, políticas de <i>retry</i> e métricas de desempenho
Multi-tenancy	Stancl Tenancy 3.x	Biblioteca especializada para <i>multi-tenancy</i> em Laravel, com identificação por domínio e isolamento configurável de <i>cache</i> , <i>filesystem</i> e filas.
Controle de Acesso	Spatie Permission	Gerenciamento de <i>roles</i> e permissões com granularidade por recurso e ação, integrado ao <i>middleware</i> do <i>framework</i> .
Frontend	Blade / JavaScript / Bootstrap	Templates <i>server-side</i> com tema administrativo profissional (Bootstrap 5), tabelas <i>server-side</i> via Yajra DataTables e interatividade via Axios.
Integração com IA	API Claude (Anthropic)	Provedor de IA generativa selecionado conforme critérios descritos na seção 5.2.1, com janela de contexto expandida (200K tokens)
Containerização	Docker / Docker Swarm	Orquestração de containers para aplicação, banco de dados e serviços auxiliares, garantindo paridade entre ambientes de desenvolvimento e produção

Fonte: Elaborado pelo autor (2026).

A escolha do PostgreSQL como sistema gerenciador de banco de dados relacional merece destaque pela sua relevância para o domínio do *framework*. Os instrumentos de avaliação do SINAES possuem estrutura hierárquica (instrumento → dimensão → indicador → conceito) que se beneficia do suporte a *schemas* múltiplos do PostgreSQL. A aplicação utiliza *schemas* dedicados como `sinaes`, `universidade`, `pessoa`, `comum`, entre outros, como mecanismo de organização lógica que reflete a separação de domínios do negócio no próprio banco de dados. Essa abordagem facilita a manutenção, melhora a legibilidade das consultas e reforça o isolamento conceitual entre módulos distintos.

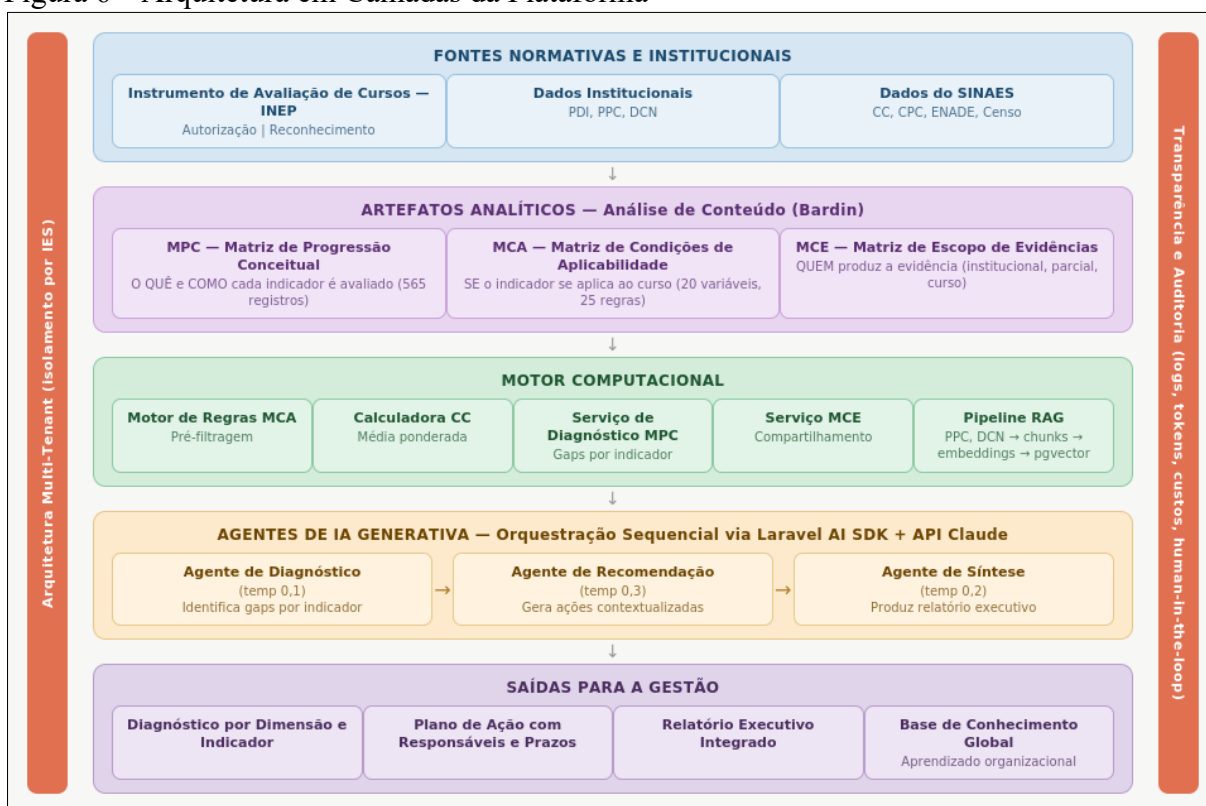
O tipo JSONB do PostgreSQL é utilizado extensivamente para armazenar dados semiestruturados, como as variáveis manuais do perfil MCA de cada curso (`variaveis_manuais`), os documentos compartilháveis e específicos da MCE (`documentos_compartilháveis`, `documentos_especificos`) e metadados de configuração dos *tenants*. Esta flexibilidade permite que o *schema* evolua sem necessidade de migrações

estruturais para cada novo atributo, característica relevante em um domínio sujeito a atualizações periódicas dos instrumentos normativos do INEP.

5.1.2 Arquitetura em Camadas

A estrutura arquitetural da aplicação *web*, detalhada na Figura 6, adota um modelo de arquitetura em camadas que isola responsabilidades para garantir escalabilidade e manutenções eficientes. O fluxo de dados inicia-se na Camada de Apresentação, composta por interfaces em Blade e JavaScript, e transita pela Camada de Aplicação, onde *middlewares* e *controllers* gerenciam as requisições, até atingir a lógica de negócio na Camada de Domínio. A persistência e a integração com serviços externos, como o PostgreSQL e a API generativa do Claude, são intermediadas pela Camada de Infraestrutura, seguindo rigorosamente o princípio de Inversão de Dependência (DIP) para assegurar que as abstrações de alto nível não sejam afetadas pelas implementações técnicas subjacentes.

Figura 6 – Arquitetura em Camadas da Plataforma



Fonte: Elaborado pelo autor (2026).

5.1.3 Arquitetura Modular

A aplicação adota arquitetura modular orientada a domínios, na qual o código-fonte é organizado em módulos independentes dentro do diretório `core/`, em contraste com a estrutura monolítica padrão do Laravel (diretório `app/`). Cada módulo encapsula um domínio de negócio específico e possui estrutura interna padronizada que separa responsabilidades em camadas bem definidas.

Esta decisão arquitetural fundamenta-se no princípio da separação de interesses (*separation of concerns*), conforme proposto por Martin (2017) no contexto dos princípios SOLID de engenharia de *software*. A modularização permite que alterações nos instrumentos do SINAES, como a publicação de novos critérios de avaliação pelo INEP, sejam absorvidas no módulo SINAES sem impactar os módulos de gestão universitária ou de controle de acesso.

A Tabela 16 apresenta os módulos implementados e suas responsabilidades no contexto do *framework*.

Tabela 16 – Arquitetura com separação de interesses por Módulos.

Módulo	Responsabilidade
Sinaes	Domínio central: instrumentos IACG, dimensões, indicadores, conceitos, MPC (atributos e evidências), MCA (variáveis e motor de regras), MCE (escopos de compartilhamento), avaliações de curso e planos de ação
Universidade	Gestão acadêmica: instituições de ensino superior, campi, unidades organizacionais, cursos de graduação (com atributos INEP completos), gestores de curso e eventos de domínio
Tenant	Lógica de <i>multi-tenancy</i> : modelo <i>Tenant</i> , identificação por domínio, <i>bootstrappers</i> de isolamento (cache, filesystem, filas, Redis), ciclo de vida de <i>tenants</i>
User	Gestão de usuários: autenticação, perfil, segurança, <i>impersonation</i> , relação User ↔ <i>Tenant</i> (pivot UserTenant)
Acl	Controle de acesso: <i>roles</i> e permissões via Spatie Permission, enums de recursos e ações, middleware de autorização
Pessoa	Dados cadastrais: pessoas físicas, endereços, telefones, e-mails, documentos, deficiências, etnias - shared entre contextos SaaS e <i>Tenant</i>
Comum	Utilitários compartilhados: dados geográficos (países, estados, cidades),

	áreas de conhecimento CNPq, classificação CINE, áreas de avaliação INEP, IES cadastradas, helpers e classes base
Planos	Planos de assinatura: <i>plans</i> , preços, features e subscriptions para o modelo SaaS
Auditoria	Rastreabilidade: <i>logs</i> de atividade, <i>logs</i> de autenticação e <i>logs</i> de sistema
Modulos	Gerenciamento dinâmico dos módulos da plataforma: ativação, configuração e navegação
IA	Camada de integração com os Agentes de Inteligência Artificial.

Fonte: Elaborado pelo autor (2026).

Internamente, cada módulo segue uma estrutura em camadas que implementa o padrão *Repository*, conforme recomendado por Fowler (2002) para mediação entre a camada de domínio e a camada de mapeamento de dados. A organização interna de um módulo típico compreende:

- **Controllers**: recebem requisições HTTP, delegam ao *Service* e retornam respostas; não contêm lógica de negócio;
- **Services**: concentram a lógica de negócio; dependem exclusivamente de *interfaces* de *Repository* injetadas via construtor;
- **Repositories** (Interface + Implementações): encapsulam o acesso a dados; toda entidade possui uma interface contratual e pelo menos uma implementação concreta (Eloquent ORM);
- **Form Requests**: validação e autorização de dados de entrada;
- **DTOs** (*Data Transfer Objects*): estruturas tipadas para transferência de dados entre camadas;
- **Models**: representação das entidades do domínio via Eloquent ORM;
- **Enums**: enumerações tipadas que codificam valores de domínio;
- **Jobs**: unidades de trabalho assíncrono processadas via filas Redis/Horizon.

Essa estratificação garante que a lógica de cálculo do Conceito de Curso, por exemplo, resida exclusivamente no `CcCalculatorService` do módulo Sinaes, enquanto o acesso aos dados de avaliação seja mediado pelo `CursoAvaliacaoRepositoryInterface`. A substituição da implementação do repositório, de Eloquent ORM para *Query Builder* ou cache, não requer alteração nos serviços consumidores, preservando o princípio da inversão de

dependência.

5.1.4 Arquitetura *Multi-Tenant*

O *framework* foi concebido como plataforma SaaS (*Software as a Service*) capaz de atender múltiplas Instituições de Ensino Superior de forma simultânea e isolada. Essa decisão arquitetural reflete a premissa de que o artefato tecnológico desenvolvido não se destina exclusivamente à UNEMAT, mas possui potencial de replicação para outras IES brasileiras, conforme discutido na seção 1.4 da justificativa.

A implementação do *multi-tenancy* utiliza a biblioteca *Stancl Tenancy*, configurada com as seguintes características:

Identificação por domínio: Cada IES acessa a plataforma por meio de um domínio exclusivo (ex.: ``unemat.notacinco.com.br``). O middleware ``InitializeTenancyByDomain`` identifica o *tenant* correspondente e inicializa o contexto isolado antes que qualquer lógica de aplicação seja executada.

Banco de dados compartilhado: Diferentemente de abordagens que criam um banco de dados separado por *tenant*, a aplicação utiliza banco único com isolamento lógico. As tabelas *tenant-scoped*, como avaliações de cursos, planos de ação e perfis MCA, possuem coluna ``tenant_id`` com escopo global automático (*Global Scope*) que filtra os registros de acordo com o *tenant* ativo. Essa estratégia reduz a complexidade operacional de gerenciamento de múltiplos bancos de dados, mantendo o isolamento dos dados.

Isolamento de serviços auxiliares: Além do isolamento no banco de dados, a plataforma isola automaticamente por *tenant* os seguintes serviços:

- Cache Redis: cada *tenant* possui prefixo exclusivo nas chaves de cache, evitando colisão entre dados cacheados de diferentes IES;
- *Filesystem*: arquivos enviados (artefatos, anexos, documentos) são armazenados em diretórios segregados por *tenant*;
- Filas de processamento: *jobs* assíncronos preservam o contexto do *tenant* que os originou, garantindo que o processamento ocorra no escopo correto;
- Permissões Spatie: a cache de permissões é automaticamente prefixada por *tenant*, permitindo que cada IES mantenha configuração independente de perfis e permissões.

Dois contextos de operação: A plataforma opera em dois contextos distintos e complementares:

- **Contexto Central (SaaS):** domínio principal da plataforma, onde são gerenciados *tenants*, planos de assinatura, configurações globais e dados normativos do SINAES (instrumentos, indicadores, MPC). Esses dados são universais e, portanto, compartilhados entre todas as IES assinantes;
- **Contexto *Tenant*:** domínio específico de cada IES, onde operam as funcionalidades de gestão acadêmica (cursos, campi, unidades), avaliações de cursos, planos de ação MPC e os agentes de IA.

Essa separação é relevante para o *framework* na medida em que os dados normativos do SINAES, instrumentos de avaliação, indicadores, critérios de conceito, atributos da MPC e regras da MCA, são cadastrados uma única vez no contexto central e compartilhados por todos os *tenants*. Quando o INEP publica uma atualização nos instrumentos de avaliação, a manutenção é realizada centralmente e propagada instantaneamente para todas as IES da plataforma. Os dados de cada IES, avaliações históricas, conceitos atribuídos, planos de ação, artefatos documentais, permanecem estritamente isolados no escopo do respectivo *tenant*.

5.1.5 Modelo de Dados do Domínio SINAES

O modelo de dados constitui o alicerce sobre o qual o *framework* operacionaliza as matrizes analíticas descritas nas seções anteriores. A modelagem foi estruturada no *schema* PostgreSQL `sinaes`, segregando logicamente as entidades do domínio regulatório das demais entidades da aplicação.

O modelo organiza-se em quatro agrupamentos funcionais: (a) a representação dos instrumentos normativos do INEP; (b) a codificação da Matriz de Condições de Aplicabilidade; (c) a estrutura da Matriz de Progressão Conceitual; e (d) os registros de avaliação e planos de ação *tenant-scoped*.

5.1.5.1 Instrumentos de Avaliação de Cursos de Graduação (IACG)

A estrutura hierárquica dos instrumentos do INEP é representada por quatro entidades principais, cujos relacionamentos refletem fielmente a organização normativa:

`iacg_instrumento`: Registra os tipos de instrumento (Autorização, Reconhecimento, Renovação de Reconhecimento), cada um identificado por UUID. A distinção entre instrumentos é semanticamente significativa, conforme demonstrado na seção 4.1.4, pois os critérios de avaliação diferem entre cursos em processo de autorização e cursos em reconhecimento.

`iacg_dimensao`: Representa as três dimensões avaliativas estabelecidas pelo INEP: Dimensão 1 (Organização Didático-Pedagógica), Dimensão 2 (Corpo Docente e Tutorial) e Dimensão 3 (Infraestrutura). Cada dimensão está vinculada a um instrumento específico.

`iacg_indicadores`: Entidade central do modelo, registra os indicadores de avaliação com seus respectivos códigos, nomes, observações de aplicabilidade e elementos constitutivos. Conforme apresentado na seção 4.1.5, o instrumento de Autorização contempla 55 indicadores e o de Reconhecimento, 58 indicadores. Cada indicador pertence a uma dimensão e pode estar associado a condições de aplicabilidade.

`iacg_conceitos`: Armazena os cinco critérios de atribuição de conceito (escala 1 a 5) para cada indicador. Cada registro contém o texto normativo completo do critério, preservando a rastreabilidade ao instrumento original do INEP. Essa entidade é fundamental para o funcionamento da MPC, pois os atributos de progressão e evidências requeridas são derivados da análise semântica desses critérios.

`iacg_aplicabilidades`: Registra as condições de aplicabilidade dos indicadores, distinguindo entre modalidades presencial e a distância. Esta entidade complementa a MCA ao fornecer as regras de base para determinação da aplicabilidade de cada indicador.

Essas entidades são globais (não possuem `tenant_id`), pois representam a normativa do SINAES, compartilhada entre todas as IES da plataforma.

5.1.5.2 Codificação da Matriz de Condições de Aplicabilidade (MCA)

A MCA é codificada no modelo de dados por meio de três entidades que implementam o sistema de regras descrito na seção 4.2:

`mca_variaveis`: Define as 20 variáveis determinantes da aplicabilidade (V01 a V20, acrescidas de CATEGORIA_ADM). Cada variável possui nome, descrição, tipo de dado e domínio de valores válidos. As variáveis V01 (grau do curso), V02 (modalidade) e V03 (área de avaliação) são derivadas automaticamente do cadastro do curso, enquanto as variáveis V04 a V20 são coletadas manualmente via formulário do perfil MCA.

`mca_regras`: Armazena as regras de aplicabilidade em formato de DSL (*Domain-Specific Language*) JSON. Cada regra contém uma expressão lógica composta por operadores (*AND*, *OR*, *NOT*, *IF/THEN/ELSE*) e operadores de comparação (*=*, *!=*, *contains*, *not_contains*) sobre as variáveis definidas. O efeito de cada regra é codificado como um dos três resultados possíveis: *APLICA*, *APLICA_COM_RECORTE* ou *NSA* (Não Se Aplica).

`mca_regra_indicador`: Tabela *pivot* que associa regras a indicadores, com campo de prioridade que determina a ordem de avaliação. Um mesmo indicador pode estar associado a múltiplas regras, e a resolução segue a prioridade: *NSA* > *APLICA_COM_RECORTE* > *APLICA*.

`mca_curso_perfil` (*tenant-scoped*): Registra o perfil de variáveis de cada curso para fins de avaliação MCA. Contém as variáveis manuais (V04–V20) em campo JSONB, permitindo flexibilidade na evolução do conjunto de variáveis sem necessidade de migrações estruturais.

5.1.5.3 Codificação da Matriz de Progressão Conceitual (MPC)

A MPC é materializada no banco de dados por meio de duas entidades que representam a decomposição analítica descrita na seção 4.1:

`mpc_atributos`: Registra os atributos de progressão extraídos da análise de conteúdo dos critérios do INEP. Cada registro está vinculado a um indicador (*id_iacg*), a um conceito-alvo (2 a 5) e a uma ordem dentro daquele nível conceitual. A restrição de unicidade *(id_iacg, conceito_alvo, ordem)* garante a integridade referencial da decomposição.

`mpc_evidencias`: Detalha as evidências requeridas para comprovação de cada atributo de progressão. Cada evidência está vinculada a um atributo (*id_mpc_atributo*) e possui ordem sequencial, permitindo a apresentação estruturada dos requisitos documentais para cada nível conceitual.

Essas entidades, assim como as do IACG, são globais e compartilhadas entre todos os *tenants*, pois representam o conhecimento normativo derivado da análise de conteúdo dos instrumentos do INEP.

5.1.5.4 Codificação da Matriz de Compartilhamento de Evidências (MCE)

A Matriz de Compartilhamento de Evidências é codificada no modelo de dados por meio de duas estruturas complementares: uma entidade de referência normativa e uma extensão semântica aplicada à entidade de artefatos da MPC. Enquanto as entidades descritas nas seções 5.1.5.1 a 5.1.5.3 formalizam, respectivamente, os instrumentos do INEP, as condições de aplicabilidade e os atributos de progressão conceitual, a MCE opera em uma dimensão transversal a todas essas estruturas, ela determina, para cada indicador, o grau em que as evidências documentais produzidas por um curso podem ser compartilhadas com outros cursos da mesma instituição dentro da plataforma.

O modelo de dados materializa a taxonomia adotada, conforme o modelo conceitual descrito no Capítulo 4, por meio das seguintes estruturas:

`mce_indicadores`: Entidade global de referência que registra a classificação MCE de cada indicador do INEP. Cada registro está vinculado a um indicador (`id_iacg`) por relação de unicidade e cada indicador possui exatamente uma classificação MCE. O campo ``escopo_evidencia`` armazena o escopo atribuído ao indicador, restrito aos três valores canônicos do domínio: institucional, parcial e curso. Dois campos JSONB complementam o registro: ``documentos_compartilhaveis``, que lista os tipos documentais da camada institucional do indicador, e ``documentos_especificos``, que enumera os documentos da camada exclusiva do curso. Esses campos alimentam as orientações exibidas ao usuário na interface de gestão de artefatos. O campo ``justificativa`` registra o fundamento analítico da classificação, preservando a rastreabilidade da decisão de curadoria. Assim como as entidades do IACG e da MPC, ``mce_indicadores`` é uma entidade global, sem ``tenant_id``, pois representa conhecimento normativo derivado da análise de conteúdo dos instrumentos do INEP e é compartilhada entre todos os *tenants* da plataforma.

`natureza` (extensão em ``mpc_artefatos``): Para suportar a distinção entre as duas camadas do escopo parcial, a entidade de artefatos da MPC é estendida com a coluna `natureza`, restrita aos valores institucional e específico. Um artefato de natureza institucional documenta a parte compartilhável de um indicador, válida para todos os cursos, enquanto um artefato de natureza específica documenta o complemento particular de um determinado curso. Para indicadores de escopo curso, todos os artefatos são classificados como institucional, pois não há camada compartilhável separada; o artefato é o único produzido para aquele indicador naquele curso. Essa distinção é relevante para o funcionamento do mecanismo de sugestões e para a classificação de lacunas no relatório avaliativo.

A integração entre a MCE e o subsistema de artefatos opera por meio de dois

mecanismos computacionais de relevância arquitetural. O primeiro é o serviço de sugestões, que adapta seu comportamento ao escopo MCE do indicador associado ao item de plano em análise: para o escopo institucional, o sistema realiza uma busca transversal por artefatos aprovados vinculados ao mesmo indicador em qualquer curso do *tenant*; para o escopo parcial, o sistema executa duas buscas independentes, uma para a camada institucional e outra para complementos de outros cursos; para o escopo curso, as sugestões são desabilitadas, e o sistema orienta o usuário à criação exclusiva. O segundo mecanismo é a classificação de lacunas, que discrimina, para fins de relatório, cinco estados distintos de evidenciação: ausência completa de artefato (*sem_artefato*), artefato em elaboração ou revisão (*artefato_pendente*), lacuna documental onde é a situação em que o artefato institucional existe no *tenant* mas não foi vinculado ao curso em questão (*lacuna_documental*), evidenciação parcial para ser aplicável ao escopo parcial quando apenas uma das duas camadas está documentada (*parcialmente_evidenciado*), e evidenciação completa (*evidenciado*). Essa taxonomia de lacunas permite ao *framework* distinguir ações de natureza substantiva, que requerem a produção de novos documentos, de ações de natureza administrativa, que requerem apenas a vinculação de artefatos já existentes.

5.1.5.5 Avaliações e Planos de Ação (*Tenant-Scoped*)

O quarto agrupamento compreende as entidades que registram os dados específicos de cada IES:

`curso_avaliacoes` (*tenant-scoped*): Registra as avaliações realizadas para cada curso, vinculadas ao tipo de instrumento utilizado. Armazena o Conceito de Curso (CC) ponderado calculado pelo sistema, bem como as médias por dimensão. Cada avaliação pode conter múltiplos registros de conceitos por indicador na entidade associada **`curso_avaliacao_indicadores`**.

`curso_avaliacao_indicadores` (*tenant-scoped*): Registra o conceito atribuído a cada indicador em uma avaliação específica, incluindo observações e justificativas. Esta entidade viabiliza o cálculo do CC e a identificação de *gaps* pela MPC.

`mpc_planos` (*tenant-scoped*): Representa os planos de ação de melhoria contínua, vinculados a um curso e opcionalmente a uma avaliação. Cada plano possui título, status (rascunho, em andamento, concluído), datas de início e término previsto, e pode conter múltiplos itens de ação.

`mpc_plano_itens` (*tenant-scoped*): Detalha os itens de cada plano de ação, vinculando indicadores, atributos da MPC e evidências requeridas a ações concretas com responsáveis, prazos e status de acompanhamento. O conceito atual e o conceito-alvo são registrados para cada item, permitindo o monitoramento da progressão.

`mpc_artefatos` (*tenant-scoped*): Registra os artefatos (documentos, registros, evidências) produzidos no âmbito dos planos de ação. Cada artefato possui natureza classificada conforme o escopo MCE (institucional, parcial ou curso), status de validação (rascunho, submetido, em revisão, aprovado, rejeitado) e pode conter anexos (`mpc_artefato_anexos`), textos ricos (`mpc_artefato_textos`) e histórico de alterações de status (`mpc_artefato_historico`).

A Tabela 17 sintetiza a distribuição quantitativa das entidades por agrupamento funcional.

Tabela 17 – Distribuição quantitativa das entidades por agrupamento funcional

Agrupamento	Entidades	Escopo	Schema
Instrumentos IACG	5	Global	sinaes
Matriz de Aplicabilidade (MCA)	4	Global + <i>Tenant</i>	sinaes
Matriz de Progressão (MPC)	2	Global	sinaes
Avaliações e Planos de Ação	8	<i>Tenant</i>	sinaes
Total	19	-	-

Fonte: Elaborado pelo autor (2026).

A separação entre entidades globais e *tenant-scoped* operacionaliza um dos princípios centrais do *framework*: a normativa do SINAES é universal, enquanto sua aplicação é contextual e particular a cada IES. Essa distinção, materializada na presença ou ausência da coluna `tenant_id` e do escopo global correspondente nos modelos *Eloquent*, assegura que as atualizações nos instrumentos de avaliação sejam automaticamente refletidas em todas as IES assinantes, ao passo que os dados institucionais permanecem hermeticamente isolados.

5.1.6 Gestão de Dados Acadêmicos

Paralelamente ao domínio SINAES, a aplicação mantém modelo de dados dedicado à representação da estrutura acadêmico-administrativa das IES, organizado no *schema*

PostgreSQL `universidade`. Esse modelo é essencial para o funcionamento do *framework*, pois os dados acadêmicos fornecem o contexto institucional necessário para a aplicação das matrizes analíticas.

A entidade central é o Curso (`universidade.cursos`), que registra informações completas em conformidade com os atributos utilizados pelo INEP: grau acadêmico (Bacharelado, Licenciatura, Tecnológico), modalidade (Presencial, EAD), turnos de funcionamento (armazenados em JSONB para suportar múltiplos turnos), código e-MEC, situação no MEC, área de conhecimento CNPq, classificação CINE, dados de autorização e reconhecimento (número de portaria DOU, data de publicação, prazo de validade), entre outros.

Esses atributos são diretamente utilizados pelas variáveis automáticas da MCA (V01 - grau, V02 - modalidade, V03 - área de avaliação) para determinação da aplicabilidade dos indicadores.

O modelo hierárquico Instituição → Campus → Unidade → Curso reflete a organização administrativa típica de universidades brasileiras, especialmente de instituições *multicampi* como a UNEMAT. Cada campus possui endereço, contatos e pode abrigar múltiplas unidades organizacionais. A entidade Curso vincula-se a um campus e pode possuir gestores designados (`curso_gestores`), viabilizando a atribuição de responsabilidades nos planos de ação da MPC.

5.1.7 ETL dos Dados do SINAES

A alimentação do sistema com os dados normativos do SINAES segue processo de ETL (*Extract, Transform, Load*) implementado por meio de comandos especializados. Essa abordagem foi adotada em razão da natureza dos dados de origem, instrumentos de avaliação disponibilizados pelo INEP em formato textual que demandam tratamento e normalização antes da inserção no modelo relacional.

O principal mecanismo de ETL é o comando `sinaes:mpc:importar`, que realiza a importação da Matriz de Progressão Conceitual a partir de arquivos CSV estruturados. O processo compreende as seguintes etapas:

Extração: O arquivo CSV contém dez colunas que representam a decomposição analítica de cada indicador: tipo de instrumento, código do indicador, nome, observações, elementos constitutivos, conceito-alvo, atributo de progressão, ordem do atributo, evidência requerida e ordem da evidência. Os dados são provenientes da análise de conteúdo descrita na

seção 4.1.1, realizada manualmente a partir dos instrumentos oficiais do INEP (2017).

Transformação: O processo de transformação inclui: (a) detecção e remoção de BOM (*Byte Order Mark*) UTF-8; (b) conversão de *encoding* Latin-1/Windows-1252 para UTF-8, necessária para preservar caracteres acentuados do português; (c) mapeamento de tipos de instrumento, a *string* "Autorização" é mapeada para o tipo `AUTORIZACAO`, enquanto "Reconhecimento" é mapeado simultaneamente para `RECONHECIMENTO` e `RENOVACAO_RECONHECIMENTO`, refletindo a equivalência normativa entre esses dois instrumentos; (d) resolução de referências cruzadas entre instrumentos e indicadores, com cache local para evitar consultas redundantes ao banco de dados.

Carga: A inserção utiliza operações idempotentes (``updateOrCreate``) com critérios de unicidade compostos: ``(id_iacg, conceito_alvo, ordem)`` para atributos e ``(id_mpc_tributo, ordem)`` para evidências. Essa idempotência garante que execuções repetidas do processo, necessárias após revisões na análise de conteúdo, não gerem duplicações nem perda de dados. Adicionalmente, o campo ``elementos`` do indicador é atualizado durante a carga, sincronizando os elementos constitutivos identificados na MPC com a entidade do indicador.

O comando suporta a opção (`--dry-run`), que executa todo o processo dentro de uma transação deliberadamente revertida (*rollback*), permitindo validação prévia sem efeito colateral no banco de dados. Em execução normal, o processo é envolto em transação atômica para garantir consistência total: ou todos os registros são importados com sucesso, ou nenhuma alteração é persistida.

A MCE é alimentada por processo análogo, utilizando arquivo CSV específico (``MPC_compartilhamento_evidencias.csv``) que classifica cada indicador quanto ao seu escopo de compartilhamento de evidências (institucional, parcial ou curso), conforme descrito na seção 4.3.

A estratégia de ETL garante consistência na classificação entre todas as instâncias da plataforma e permite a atualização da curadoria por meio da reexecução do processo de importação, sem necessidade de intervenção direta no banco de dados.

5.1.8 Serviços Computacionais do *Framework*

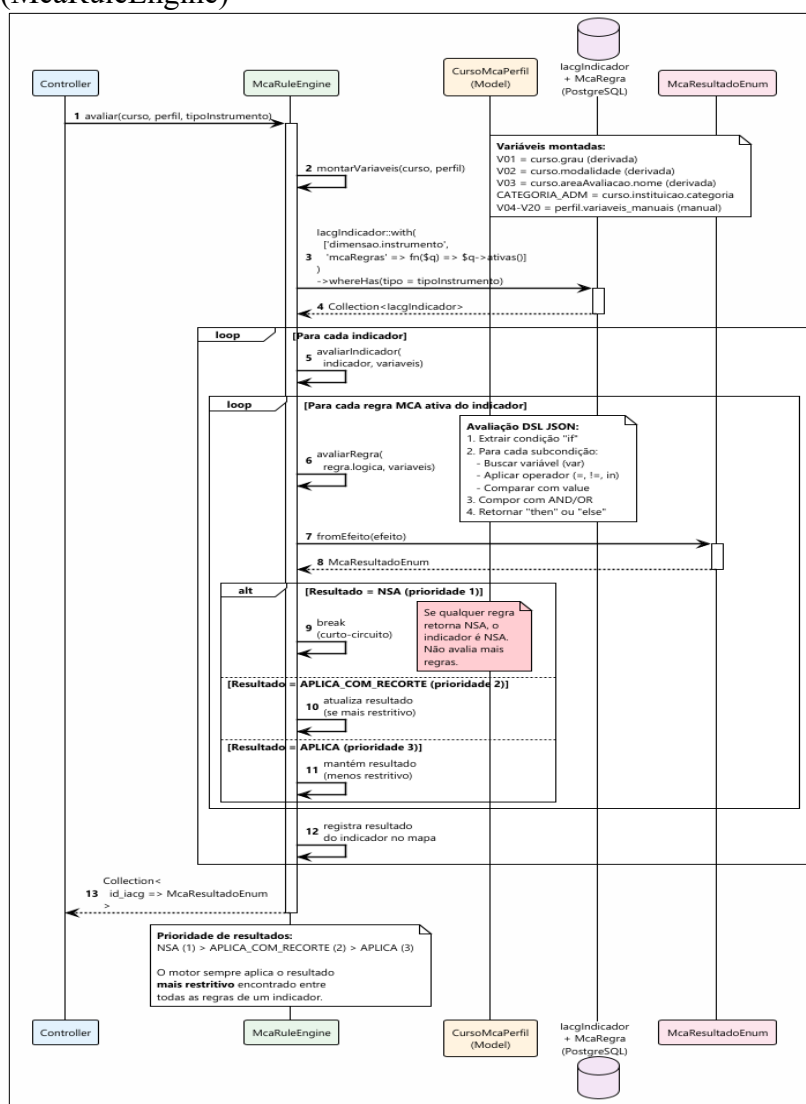
A lógica de negócio do *framework* é implementada por serviços especializados que operam sobre os dados normativos e institucionais descritos nas seções anteriores. Esses serviços constituem o núcleo computacional da aplicação, materializando em código as

operações analíticas que o *framework* propõe. Destacam-se três serviços fundamentais.

5.1.8.1 Motor de Regras da MCA (McaRuleEngine)

O `McaRuleEngine` implementa o sistema especialista que determina a aplicabilidade de cada indicador do SINAES para um curso específico, conforme descrito conceitualmente na seção 4.2. Conforme a Figura 7, trata-se de um serviço puro, sem estado persistente e sem dependência de repositórios, que recebe como entrada os dados do curso, seu perfil MCA e o tipo de instrumento, retornando uma coleção indexada por indicador com o respectivo resultado: `APLICA`, `APLICA_COM_RECORTE` ou `NSA`.

Figura 7 – Diagrama de Sequência: Motor de Regras MCA (McaRuleEngine)



Fonte: Elaborado pelo autor (2026).

O processo de avaliação opera em três níveis:

Montagem do contexto: As 20 variáveis determinantes são montadas a partir de duas fontes: variáveis automáticas derivadas do cadastro do curso (V01 - grau acadêmico, V02 - modalidade, V03 - área de avaliação CNPq) e variáveis manuais fornecidas pelo perfil MCA do curso (V04 a V20, como existência de estágio obrigatório, TCC, laboratórios especializados, entre outras). A categoria administrativa da IES completa o conjunto.

Avaliação por indicador: Para cada indicador ativo, o motor recupera as regras associadas (via tabela *pivot* `mca_regra_indicador`) e as avalia em ordem de prioridade. A hierarquia de prioridade segue a lógica: `NSA` (prioridade 1) > `APLICA_COM_RECORTE` (prioridade 2) > `APLICA` (prioridade 3). Se uma regra retorna `NSA`, a avaliação é interrompida imediatamente (*short-circuit*), pois nenhuma outra regra pode sobrepor a não aplicabilidade.

Interpretação da DSL: Cada regra é expressa em DSL JSON com suporte a: (a) resultado direto (`{"resultado": "nsa"}`); (b) estruturas condicionais (`IF/THEN/ELSE`); (c) operadores lógicos compostos (`AND`, `OR`, `NOT`); e (d) operadores de comparação sobre variáveis (`=`, `!=`, `contains`, `not_contains`). Essa arquitetura de DSL declarativa permite que novas regras de aplicabilidade sejam cadastradas sem alteração no código-fonte do motor, característica relevante para a manutenibilidade do sistema diante de atualizações nos instrumentos do INEP.

A adoção de uma abordagem baseada em sistema especialista, em detrimento de algoritmos de aprendizado de máquina, justifica-se pela natureza determinística das regras de aplicabilidade do SINAES. As condições que determinam se um indicador se aplica ou não a um curso específico são normativamente definidas e não admitem margem de interpretação probabilística. Conforme observa Russell e Norvig (2013), sistemas especialistas constituem abordagem adequada quando o domínio é bem delimitado e o conhecimento pode ser representado por regras explícitas.

5.1.8.2 Calculadora do Conceito de Curso (CcCalculatorService)

O `CcCalculatorService` implementa o cálculo do Conceito de Curso (CC) segundo a fórmula oficial de média ponderada por dimensão estabelecida pelo INEP. O serviço recebe uma avaliação de curso, o instrumento utilizado e, opcionalmente, o resultado da avaliação

MCA, retornando o CC em escala contínua de 1,00 a 5,00.

A Tabela 18 descreve como os pesos por dimensão variam conforme o tipo de instrumento:

Tabela 18 – Pesos por dimensão e tipo de instrumento.

Dimensão	Autorização	Reconhecimento / Renovação	Customizado
1 - Organização Didático-Pedagógica	0,4	0,3	De 0,0 a 1
2 - Corpo Docente e Tutorial	0,2	0,4	De 0,0 a 1
3 - Infraestrutura	0,4	0,3	De 0,0 a 1

Fonte: INEP (2017).

Quando integrado ao resultado da MCA, o serviço exclui automaticamente do cálculo os indicadores classificados como `NSA`, exceto aqueles com nota forçada pelo avaliador (`nsa_forcado`). Para cada dimensão, a média é calculada como a razão entre a soma dos conceitos dos indicadores aplicáveis e o total de indicadores aplicáveis naquela dimensão. O CC final é a soma ponderada das médias dimensionais.

O serviço também disponibiliza método para cálculo das médias por dimensão de forma desagregada, retornando para cada dimensão: a média, o total de indicadores e o número de indicadores com conceito atribuído. Essa informação alimenta dashboards e relatórios que permitem à gestão institucional identificar dimensões com desempenho inferior.

5.1.8.3 Serviço de Diagnóstico MPC (MpcDiagnosticoTenantService)

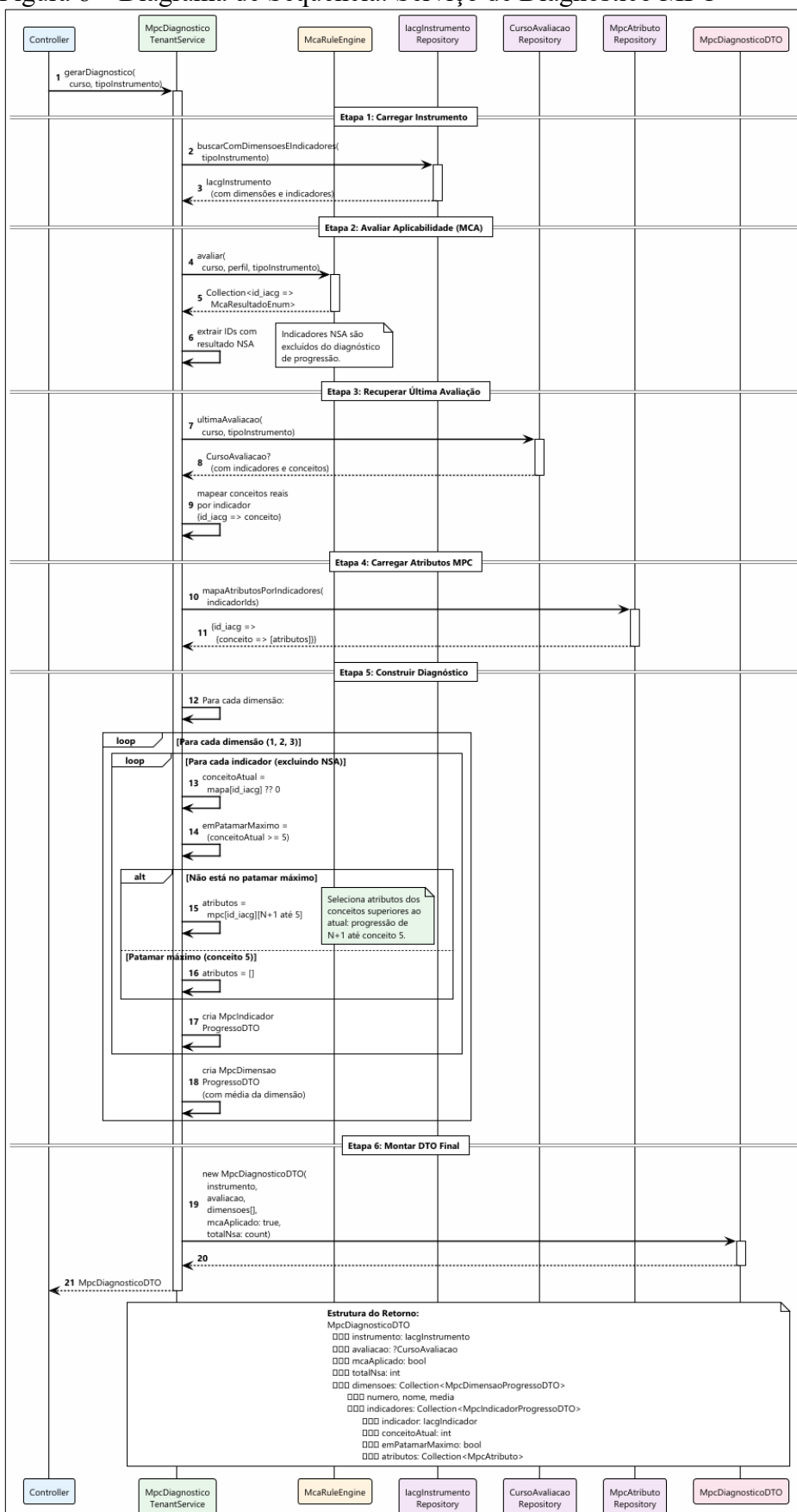
O `MpcDiagnosticoTenantService` constitui o ponto de integração central entre as três matrizes analíticas e os dados institucionais de cada curso. Este serviço gera o diagnóstico de progressão conceitual que fundamenta as recomendações dos agentes de IA e os planos de ação de melhoria.

O fluxo de geração do diagnóstico compreende as seguintes etapas:

1. **Carregamento do instrumento:** recupera o instrumento de avaliação com todas as dimensões, indicadores e conceitos associados, utilizando *eager loading* para evitar o problema N+1 de consultas;
2. **Avaliação MCA:** se o curso possui perfil MCA processado, o motor de regras é acionado para determinar quais indicadores são efetivamente aplicáveis. Os indicadores classificados como `NSA` são excluídos do diagnóstico;

3. **Recuperação da avaliação:** localiza a avaliação mais recente compatível com o tipo de instrumento, obtendo o conceito atribuído a cada indicador;
4. **Montagem do mapa de conceitos:** para cada indicador com conceito registrado, cria a correspondência entre indicador e conceito atual;
5. **Construção do diagnóstico por dimensão:** para cada indicador aplicável, o serviço recupera os atributos da MPC correspondentes aos conceitos superiores ao atual (conceito N+1 até conceito 5). Indicadores no conceito máximo (5) recebem a marcação `emPatamarMaximo`. O resultado é estruturado em DTOs hierárquicos: `MpcDiagnosticoDTO` → `MpcDimensaoProgressoDTO` → `MpcIndicadorProgressoDTO`.
6. **Diagnóstico prospectivo:** para cursos que ainda não passaram por avaliação (típico de cursos em processo de autorização), o serviço opera em modo prospectivo, assumindo conceito inicial zero para todos os indicadores. Essa funcionalidade permite que coordenadores de cursos novos utilizem a MPC para planejar a preparação para a primeira avaliação *in loco*.

Figura 8 – Diagrama de Sequência: Serviço de Diagnóstico MPC



Fonte: Elaborado pelo autor (2026).

Conforme demonstrado na Figura 8, o diagnóstico resultante fornece, para cada indicador de cada dimensão: o conceito atual, os atributos de progressão necessários para atingir cada nível superior e as evidências requeridas para comprovação. Essa informação estruturada constitui o insumo primário para os agentes de IA, conforme descrito na seção 4.1.10, e para os planos de ação gerenciados pela coordenação de curso.

5.1.9 Fluxo de Dados do *Framework*

A operacionalização do *framework* pode ser compreendida como um fluxo de dados que atravessa quatro estágios sequenciais, desde a fonte normativa até a ação institucional:

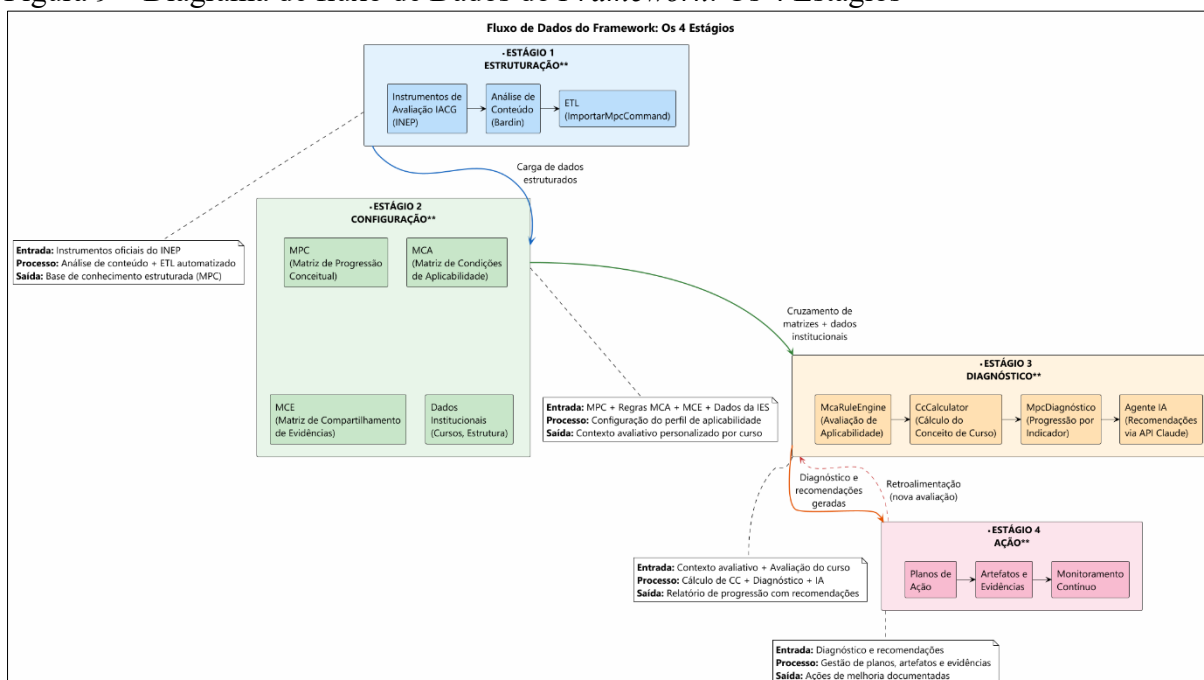
Estágio 1 - Ingestão normativa: Os instrumentos de avaliação do INEP são submetidos à análise de conteúdo (processo manual descrito na seção 3.6) e os resultados são importados para o banco de dados via processos ETL. Neste estágio, são populadas as entidades do IACG, os atributos e evidências da MPC, as regras da MCA e as classificações da MCE.

Estágio 2 - Contextualização institucional: Os dados acadêmicos da IES são cadastrados no sistema: instituição, campi, unidades, cursos com seus atributos INEP, gestores e perfil MCA de cada curso. Os conceitos de avaliações anteriores são registrados por indicador. Neste estágio, a MCA é processada para determinar, para cada curso, quais indicadores são efetivamente aplicáveis.

Estágio 3 - Diagnóstico analítico: O `MpcDiagnosticoTenantService` cruza os dados normativos (MPC) com os dados institucionais (avaliações) e os resultados da MCA para produzir o diagnóstico de progressão conceitual. Para cada indicador aplicável, o sistema identifica o conceito atual, os requisitos para progressão e as evidências necessárias. O `CcCalculatorService` calcula o CC ponderado e as médias dimensionais.

Estágio 4 - Ação institucional: Com base no diagnóstico, os gestores (coordenadores de curso, membros da CPA) elaboram planos de ação vinculados a indicadores específicos, com atributos da MPC como referência normativa. Os artefatos documentais são produzidos, classificados conforme o escopo MCE e submetidos a fluxo de validação. Os agentes de IA, quando acionados, processam o diagnóstico para gerar recomendações contextualizadas e sínteses executivas.

Figura 9 – Diagrama de fluxo de Dados do *Framework*: Os 4 Estágios



Fonte: Elaborado pelo autor (2026).

Esse fluxo materializa a proposta central da dissertação: transformar dados avaliativos brutos em conhecimento acionável para a gestão de cursos de graduação. A MPC fornece a estrutura analítica, a MCA garante a pertinência contextual, a MCE otimiza o esforço documental e os serviços computacionais automatizam o cruzamento e a síntese das informações.

5.1.10 Controle de Acesso e Segurança

O *framework* implementa modelo de controle de acesso baseado em permissões granulares, alinhado ao princípio do menor privilégio. As permissões são compostas pela concatenação de dois enumeradores: o recurso (ex.: `tenant-sinaes-mca`, `tenant-universidade-curso`) e a ação (`read`, `create`, `update`, `delete`, `export`, `import`, `enable`, `disable`). A string resultante (ex.: `tenant-sinaes-mca-read`) é verificada em três pontos:

Middleware de rota: cada controlador declara as permissões requeridas por método, impedindo o acesso antes da execução da lógica de negócio;

Diretiva Blade: os elementos de interface (botões, links, menus) são condicionalmente renderizados via diretiva `@can`, garantindo que os usuários visualizem apenas as ações que estão autorizados a executar;

Políticas (Policies): a autorização é verificada no nível do modelo para operações que dependem de relações específicas entre o usuário e a entidade.

No contexto *multi-tenant*, as permissões são isoladas por *tenant*, de modo que o administrador de uma IES não possui qualquer visibilidade ou acesso aos dados de outra IES. O cache de permissões do Spatie é automaticamente prefixada com o identificador do *tenant* ativo, eliminando riscos de vazamento entre contextos.

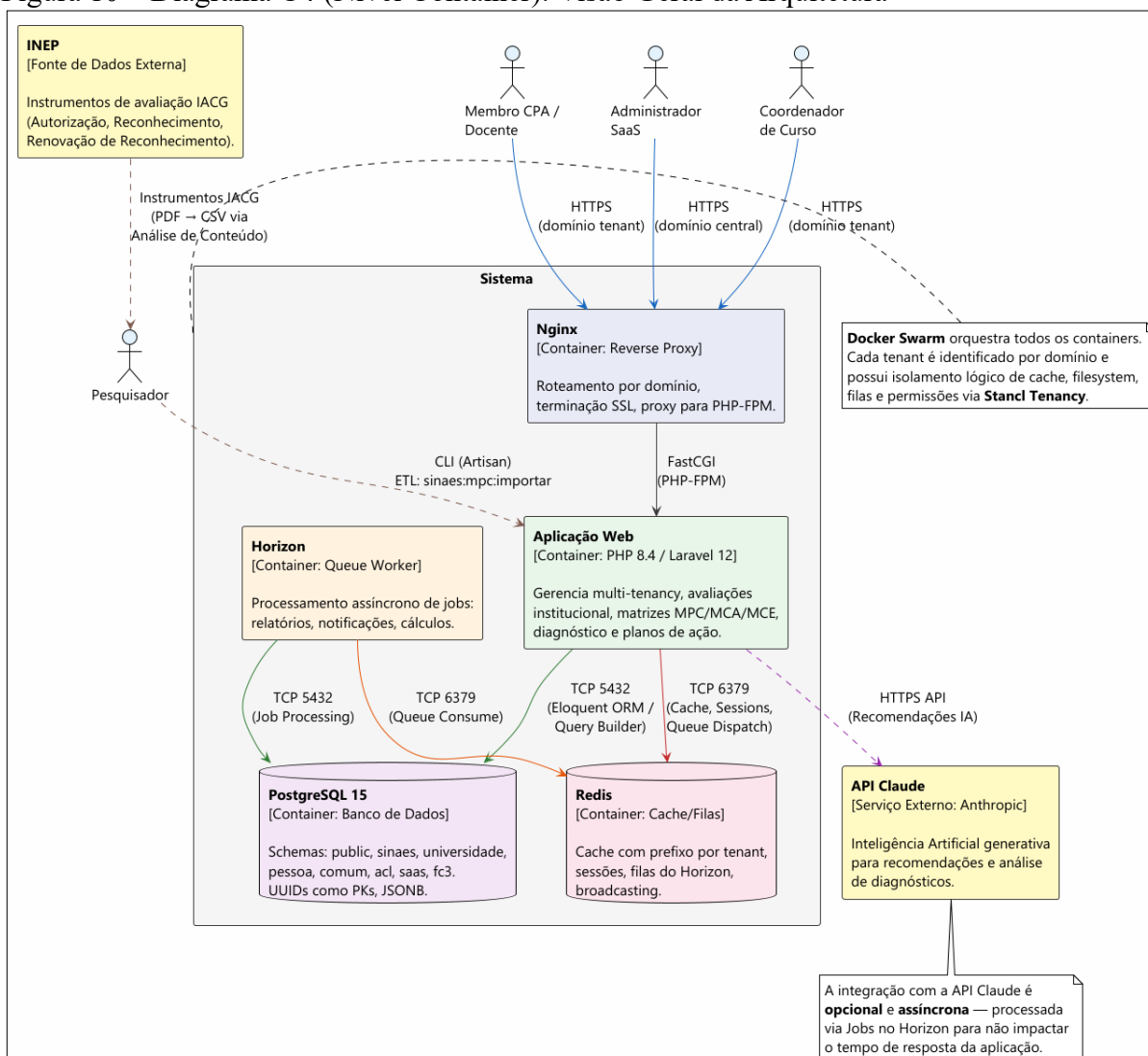
A conformidade com a Lei Geral de Proteção de Dados Pessoais (LGPD – Lei nº 13.709/2018) é assegurada por medidas complementares: (a) os dados enviados à API de IA generativa são agregados e desprovidos de identificação individual de usuários, de estudantes ou docentes; (b) o acesso ao sistema requer autenticação em todos os contextos; (c) os *logs* de autenticação registram IP, *user-agent* e geolocalização para fins de auditoria; (d) toda a comunicação utiliza protocolo HTTPS.

5.1.11 Síntese Arquitetural

A arquitetura descrita nesta seção materializa, no plano tecnológico, as contribuições analíticas apresentadas nas seções anteriores do capítulo. A modularidade do sistema permite que cada componente do *framework*, instrumentos normativos, matrizes analíticas, dados institucionais e camada de inteligência, evolua de forma independente. O modelo *multi-tenant* viabiliza a aplicação do *framework* por múltiplas IES sem comprometimento do isolamento de dados. O modelo relacional do *schema* `sinaes` codifica com fidelidade a estrutura hierárquica dos instrumentos do INEP e as decomposições analíticas da MPC, MCA e MCE. Os serviços computacionais, motor de regras, calculadora de CC e serviço de diagnóstico, automatizam as operações analíticas que, realizadas manualmente, constituiriam tarefa exaustiva e propensa a inconsistências.

Conforme proposto por Peffers *et al.* (2007) no contexto do DSRM, a descrição detalhada do artefato tecnológico constitui etapa indispensável para assegurar a reprodutibilidade da pesquisa e a avaliação objetiva da solução. As decisões de *design* aqui documentadas, desde a escolha do PostgreSQL com *schemas* múltiplos até a implementação do motor de regras em DSL declarativa, refletem a busca por uma arquitetura que conjugue rigor técnico com flexibilidade para absorver futuras atualizações nos instrumentos do SINAES.

Figura 10 – Diagrama C4 (Nível Container): Visão Geral da Arquitetura



Fonte: Elaborado pelo autor (2026).

5.2 Integração com Inteligência Artificial Generativa

As seções precedentes deste capítulo descreveram os artefatos analíticos do *framework* e a arquitetura de *software* que os materializa em aplicação funcional. A presente seção detalha a implementação da camada de inteligência artificial generativa, componente que transforma os artefatos analíticos de referências normativas estáticas em instrumentos de análise dinâmica e geração automatizada de conhecimento. Enquanto as matrizes fornecem a estrutura e o conhecimento do domínio, os agentes de IA operam como o motor analítico que processa dados contextuais, identifica padrões e produz recomendações acionáveis para a gestão de cursos de graduação.

A integração da IA ao *framework* não constitui mera adição de funcionalidade

tecnológica, mas operacionaliza o objetivo central da dissertação: automatizar a descoberta de conhecimento nos dados avaliativos do SINAES. Conforme argumentado na fundamentação teórica, a racionalidade limitada dos gestores (SIMON, 1955) e a complexidade dos instrumentos de avaliação do INEP justificam o desenvolvimento de artefatos tecnológicos que expandam a capacidade analítica das Comissões Próprias de Avaliação e dos coordenadores de curso. A implementação aqui descrita detalha como essa expansão é concretizada no plano técnico.

5.2.1 Escolha do Provedor de IA

A seleção do provedor de IA generativa para integração ao *framework* considerou cinco critérios técnicos derivados das demandas funcionais do sistema: (a) capacidade de processamento de textos longos, necessária para análise de documentos institucionais extensos como PDI, PPC e DCN; (b) consistência e replicabilidade das respostas, fundamental para validação científica das análises; (c) qualidade da API e disponibilidade de SDK com suporte a padrões contemporâneos de integração; (d) custo por token processado, determinante para viabilidade financeira da solução SaaS; e (e) conformidade com requisitos de privacidade de dados, em atendimento à LGPD.

Conforme os critérios de seleção apresentados na Tabela 19 e após análise comparativa entre os três principais provedores disponíveis no mercado, como OpenAI (GPT-4 Turbo/GPT-4o), Anthropic (Claude Sonnet/Opus) e Google (Gemini 1.5 Pro), optou-se pela utilização da API Claude da Anthropic como provedor primário. A decisão fundamentou-se, preponderantemente, em dois fatores: a janela de contexto expandida de 200.000 tokens, que permite processar documentos extensos como o PPC e as DCN em uma única requisição sem necessidade de fragmentação; e a consistência superior observada em tarefas de análise estruturada, característica verificada empiricamente durante os ciclos de prototipação da arquitetura de *prompts*.

Tabela 19 – Critérios de Seleção

Critério	Peso	Descrição
Capacidade de Processamento de Texto Longo	Alto	Análise de documentos extensos (PDI, PPC)
Consistência nas Respostas	Alto	Replicabilidade das análises
API Estável e Documentada	Alto	Integração com Laravel

Custo por Token	Médio	Viabilidade financeira
Compliance com LGPD	Alto	Proteção de dados institucionais

Fonte: Elaborado pelo autor (2026).

A comparação entre provedores de IA generativa apresentada na Tabela 20 foi conduzida em janeiro/2026. O mercado de modelos de linguagem de grande escala está em acelerada evolução, com lançamentos frequentes de novas versões e alterações de desempenho, custo e disponibilidade. Os critérios utilizados na análise comparativa foram definidos *a priori* com base nas necessidades funcionais do *framework*; a escolha do provedor adotado reflete o estado do mercado na data da coleta e pode ser revisada em versões futuras sem alteração da arquitetura conceitual do *framework*.

Tabela 20 – Análise comparativa dos provedores de IA generativa

Provedor	Modelo	Vantagens	Desvantagens	Custo Aproximado
Anthropic	Claude Sonnet 4 / Opus 4	Janela de 200K tokens; consistência em análise estruturada; SDK Laravel nativo	Menor penetração de mercado no Brasil	US\$ 3-15/M tokens
OpenAI	GPT-4 Turbo / GPT-4o	Ampla documentação; ecossistema maduro	Janela menor (128K); inconsistências em tarefas iterativas	US\$ 5-30/M tokens
Google	Gemini 1.5 Pro	Janela de 1M tokens; custo competitivo	API menos madura; menor consistência em saídas estruturadas	US\$ 1,25-5/M tokens

Fonte: Elaborado pelo autor (2026), com base em dados públicos dos provedores (2025-2026)

A integração técnica é viabilizada pelo Laravel AI SDK (v0.4.3), pacote oficial do ecossistema Laravel que oferece abstração unificada para comunicação com provedores de IA generativa como OpenAI, Anthropic, Gemini e muitos outros. Esse SDK permite a definição de agentes de IA como classes PHP anotadas com atributos declarativos de provedor, modelo, temperatura e limites de tokens, mantendo a coerência com o padrão arquitetural MVC adotado pela aplicação usando uma interface consistente e amigável ao Laravel. A utilização de SDK nativo, em vez de chamadas REST diretas, reduz a complexidade de integração, padroniza o tratamento de erros e viabiliza funcionalidades avançadas como *streaming* de respostas, gerenciamento de conversações e invocação de ferramentas (*tool use*) pelos agentes.

Para a geração de *embeddings* vetoriais, necessários ao mecanismo de *Retrieval-Augmented Generation* (RAG), optou-se pelo modelo text-embedding-3-small da OpenAI, cuja relação custo-eficiência é reconhecida na literatura de sistemas RAG. Essa decisão de utilizar provedores distintos para geração de texto e para geração de *embeddings* exemplifica a

flexibilidade proporcionada pela abstração do Laravel AI SDK, que permite selecionar o modelo mais adequado para cada tarefa sem acoplamento entre componentes.

5.2.2 Módulo de IA: Estrutura e Organização

A camada de inteligência artificial foi implementada como módulo independente na arquitetura modular descrita na seção 5.1.3, seguindo o princípio de separação de interesses que orienta toda a organização do código-fonte. Essa modularização permite que a camada de IA evolua, caso necessário com a troca de provedor, adição de novos agentes, alteração de estratégias de *prompt*, sem impactar os módulos SINAES, Universidade ou *Tenant*.

O módulo compreende cinco componentes estruturais: (a) Agentes, classes que encapsulam a lógica de cada agente especializado; (b) Ferramentas (*Tools*), que expõem dados da aplicação aos agentes durante o processamento; (c) Serviços, que orquestram a execução dos agentes e o pipeline de processamento; (d) Modelos de dados, que persistem os resultados das análises; e (e) Jobs assíncronos, que executam as operações de maior custo computacional em segundo plano via filas Redis gerenciadas pelo Horizon.

A infraestrutura de pacotes utilizada na integração compreende: o Laravel AI SDK, que fornece a abstração unificada para comunicação com provedores de IA e implementa os mecanismos de agentes, ferramentas, *embeddings* e conversações; a extensão *pgvector* do PostgreSQL, que habilita a busca por similaridade vetorial necessária ao RAG; e bibliotecas de extração de texto (PDF e DOCX) para alimentação do pipeline de processamento documental.

5.2.3 Arquitetura dos Agentes de IA

A camada de inteligência artificial do *framework* adota arquitetura multiagente composta por três agentes especializados, cada qual responsável por uma etapa distinta do ciclo analítico. Essa especialização segue o princípio de separação de responsabilidades aplicado ao domínio de sistemas multiagentes: ao invés de concentrar toda a lógica analítica em um único agente generalista, a decomposição em agentes com papéis definidos permite controle granular sobre parâmetros de geração (temperatura, limites de tokens), conjunto de ferramentas disponíveis e critérios de validação específicos de cada etapa.

Agente de Diagnóstico. Constitui o primeiro componente da cadeia analítica.

Recebe como entrada o perfil do curso (grau acadêmico, modalidade, área de conhecimento), o subconjunto de indicadores filtrados pela MCA e os conceitos atribuídos pelo coordenador na avaliação vigente. Sua função é comparar, para cada indicador, o conceito atual com os atributos de progressão dos conceitos superiores definidos na MPC, identificando os gaps específicos que separam a situação verificada da situação desejada. O agente opera com temperatura baixa (0,1) para maximizar a consistência entre execuções, conforme recomendado para tarefas analíticas determinísticas (AMATRIAIN, 2024). Possui acesso a ferramentas de consulta aos indicadores do curso, ao diagnóstico MPC e à base de conhecimento RAG.

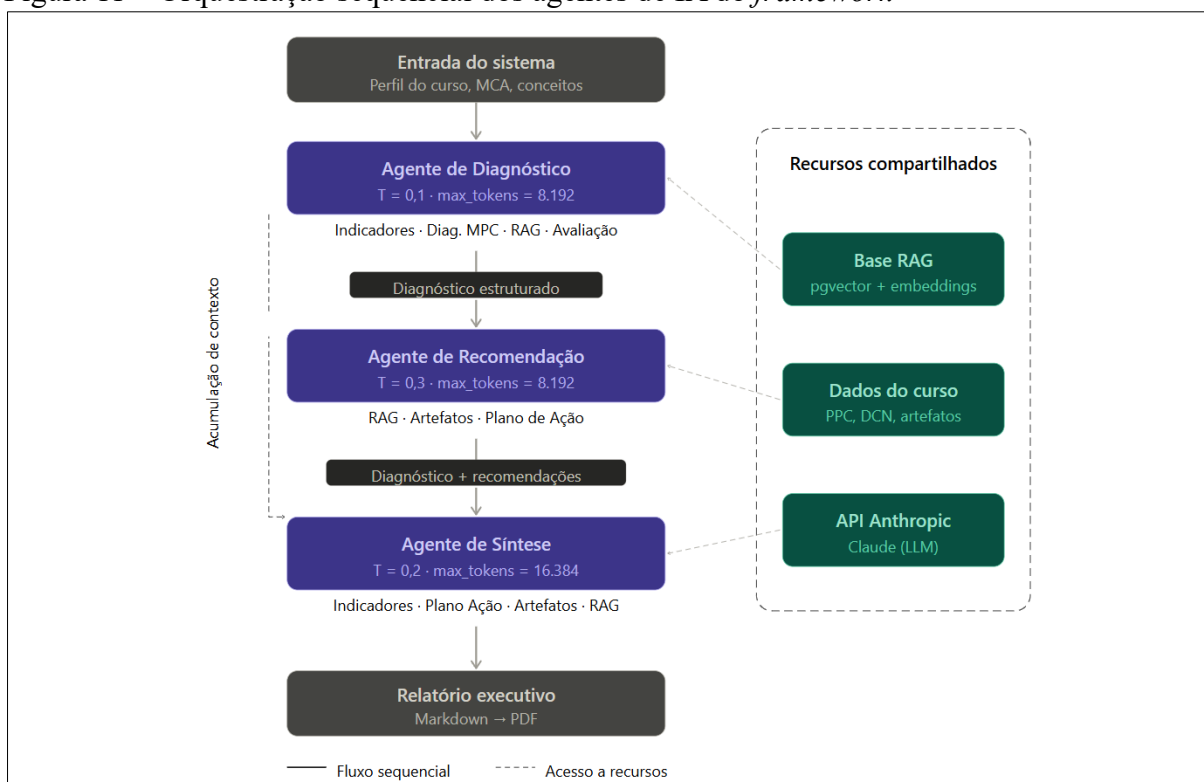
Agente de Recomendação. Opera a partir do diagnóstico produzido pelo agente anterior e consulta ativamente a base de conhecimento do curso, PPC, DCN e planejamentos anteriores, para fundamentar suas recomendações. Para cada *gap* identificado, o agente gera orientações que especificam: a ação concreta a ser implementada, o responsável sugerido (coordenador, NDE, CPA ou docente), o prazo estimado (curto, médio ou longo prazo), as evidências documentais a serem produzidas e o nível de esforço requerido. A temperatura é configurada em 0,3, permitindo maior variabilidade criativa na formulação de recomendações enquanto preserva a aderência normativa. Possui ferramentas de busca na base de conhecimento, consulta a artefatos existentes e consulta ao plano de ação.

Agente de Síntese. Consolida os *outputs* dos dois agentes anteriores em relatório executivo estruturado, organizado em seis seções padronizadas: sumário executivo, diagnóstico por dimensão, plano de ação consolidado, análise de artefatos e evidências, recomendações estratégicas e considerações finais. O agente utiliza temperatura de 0,2 e limite expandido de tokens (16.384) para produzir documento extenso com coerência textual. A saída é formatada em *Markdown* para renderização na interface da aplicação e posterior exportação em PDF.

A orquestração entre agentes opera em orquestração sequencial, com comunicação unidirecional entre estágios: o Agente de Diagnóstico produz o diagnóstico estruturado, que alimenta o Agente de Recomendação; a concatenação de diagnóstico e recomendações serve de entrada para o Agente de Síntese. A Figura 11 sintetiza essa arquitetura e a Tabela 21 define os parâmetros de configuração dos agentes de IA. A implementação dos agentes foi realizada de forma nativa via API da Anthropic, sem utilização de *frameworks* de orquestração como LangChain, CrewAI ou AutoGen, por razões de controle granular sobre o fluxo de processamento e integração estrita com a aplicação web. Essa decisão foi adotada deliberadamente em detrimento de arquiteturas mais complexas de orquestração dinâmica, em razão de dois fatores: a rastreabilidade exigida pela validação científica da pesquisa e a

previsibilidade operacional necessária ao contexto de gestão institucional, no qual os gestores precisam compreender a proveniência de cada recomendação.

Figura 11 – Orquestração sequencial dos agentes de IA do *framework*



Fonte: Elaborado pelo autor (2026).

Tabela 21 – Configuração dos agentes de IA do *framework*

Agente	Temperatura	Tokens Máximo	Passos Máximo	Ferramentas
Diagnóstico	0,1	8.192	10	Indicadores, Diagnóstico MPC, Base RAG, Avaliação
Recomendação	0,3	8.192	15	Base RAG, Artefatos, Plano de Ação
Síntese	0,2	16.384	15	Indicadores, Plano de Ação, Artefatos, Base RAG

Fonte: Elaborado pelo autor (2026).

5.2.4 Implementação dos Agentes Especializados

Cada agente é implementado como classe PHP que implementa a interface *Agent* do Laravel AI SDK, anotada com atributos declarativos que configuram o provedor, modelo, temperatura, limite de tokens e número máximo de passos de raciocínio. Essa abordagem

declarativa, além de seguir convenções idiomáticas do ecossistema Laravel, facilita a manutenção e a documentação dos parâmetros de cada agente.

Agente de Diagnóstico (*DiagnosticAgent*): Recebe como dependências injetadas no construtor o modelo do Curso, o DTO (*Data Transfer Object*) de diagnóstico MPC e, opcionalmente, os resultados da MCA. A instrução do agente (método *instructions()*) é construída dinamicamente, incorporando dados parametrizados do curso como nome, grau, modalidade e tipo de instrumento, ao contexto normativo. A cadeia de raciocínio instrui o agente a processar cada indicador aplicável em cinco etapas sequenciais: identificação do conceito atual, comparação com atributos dos conceitos superiores, listagem de *gaps*, classificação de prioridade e consideração do tipo de instrumento. A saída é estruturada por dimensão avaliativa, produzindo um diagnóstico hierarquizado que alimenta o agente seguinte.

Agente de Recomendação (*RecomendacaoAgent*): Recebe o curso e o texto do diagnóstico produzido pelo agente anterior. Suas instruções enfatizam a consulta ativa à base de conhecimento: para cada *gap*, o agente deve (1) buscar experiência prévia na base RAG, (2) verificar artefatos reutilizáveis de outros cursos, e (3) gerar recomendação com cinco componentes obrigatórios. As regras incluem a priorização de indicadores das dimensões com maior peso no CC, a consideração do perfil do curso e a preferência por recomendações específicas em detrimento de orientações genéricas. A utilização do escopo MCE é explicitamente instruída: quando há evidências institucionais compartilháveis, o agente deve recomendar a vinculação em vez da produção redundante de documentos.

Agente de Síntese (*SinteseAgent*): Recebe o curso, o plano de ação, o texto do diagnóstico e o texto das recomendações. Produz relatório executivo em *Markdown*, seguindo estrutura de seis seções definidas no *prompt*: sumário executivo com limite de 300 palavras; diagnóstico por dimensão com médias e indicadores prioritários; plano de ação consolidado com cronograma e responsáveis; análise de artefatos com status e lacunas; recomendações estratégicas com identificação de impacto imediato; e considerações finais com projeção de melhoria no CC. As regras instruem o agente a usar linguagem formal adequada a documento institucional, fundamentar afirmações com dados concretos, não repetir o diagnóstico na íntegra e incluir tabelas quando necessário.

5.2.5 Ferramentas (*Tools*) dos Agentes e Servidor MCP

Os agentes de IA do *framework* não operam exclusivamente a partir do conteúdo

textual do *prompt* e não possuem acesso irrestrito aos dados da aplicação. Cada agente dispõe de um subconjunto curado de ferramentas, selecionado conforme o princípio do menor privilégio adaptado ao contexto de agentes de IA, os agentes dispõem de um conjunto de ferramentas (*tools*) que lhes permitem consultar dados estruturados da aplicação durante o processo de geração de respostas. Essa capacidade de invocação de ferramentas (*tool use*) constitui um dos avanços arquiteturais dos modelos de linguagem contemporâneos, ao permitir que o modelo alterne entre raciocínio textual e acesso a informações externas de forma autônoma e contextualizada.

O Agente de Diagnóstico acessa indicadores, diagnóstico MPC, base RAG e avaliações históricas, porém não consulta artefatos ou planos de ação, dados que pertencem ao escopo funcional de outros agentes. Cada ferramenta é implementada como classe que define três elementos: uma descrição textual que o modelo utiliza para decidir quando invocá-la; um *schema* JSON que especifica os parâmetros de entrada esperados; e um método *handler* que executa a consulta e retorna os resultados formatados como texto. A formatação textual dos resultados é deliberada: em vez de retornar estruturas JSON brutas, cada ferramenta produz representação textual legível que otimiza a compreensão pelo modelo de linguagem.

A ferramenta de busca na base de conhecimento (BuscarBaseConhecimentoTool) merece descrição pormenorizada por sua relevância para a qualidade das recomendações. Ao receber uma consulta textual do agente, a ferramenta executa busca vetorial bidirecional: primeiro nos documentos específicos do curso (filtrados por *tenant* e curso), depois na base global de conhecimento acumulado. Os resultados incluem o tipo de documento, o *score* de similaridade, a fonte de referência e o conteúdo do *chunk*, permitindo que o agente avalie a relevância e a proveniência de cada fragmento recuperado. Essa transparência na apresentação dos resultados de busca é relevante para a rastreabilidade das recomendações geradas.

Cada agente possui acesso a um subconjunto específico de ferramentas, selecionado conforme sua responsabilidade funcional. As ferramentas implementadas compreendem: ConsultarIndicadoresCursoTool, que retorna os indicadores do curso com conceitos atribuídos, médias por dimensão e CC calculado; ConsultarMpcDiagnosticoTool, que fornece o diagnóstico de progressão conceitual com gaps identificados por indicador; ConsultarMcaPerfilTool, que informa o perfil de aplicabilidade do curso; ConsultarArtefatosTool, que lista artefatos produzidos com seus respectivos status; ConsultarPlanoAcaoTool, que retorna os itens do plano de ação vigente; ConsultarAvaliacaoCursoTool, que fornece dados de avaliações históricas; e

BuscarBaseConhecimentoTool, que executa a busca semântica na base RAG descrita na seção anterior.

Essa arquitetura de ferramentas confere aos agentes a capacidade de acessar informações atualizadas e específicas do curso sem que seja necessário incluir todos os dados no *prompt* inicial. O modelo decide autonomamente, durante o processamento, quais ferramentas invocar e em que sequência, com base no raciocínio que está desenvolvendo. Essa dinâmica reduz o volume de tokens na requisição inicial, melhora a precisão das análises e permite que os agentes operem sobre dados que evoluem ao longo do tempo, como conceitos recém-atribuídos ou artefatos recém-produzidos, sem necessidade de reconfiguração dos *prompts*.

Complementarmente, o *framework* disponibiliza um servidor MCP (*Model Context Protocol*) que expõe as funcionalidades do domínio como protocolo padronizado de integração. O MCP é um protocolo aberto que permite a comunicação entre aplicações e modelos de IA de forma interoperável, possibilitando que ferramentas externas de desenvolvimento assistido por IA consultem os dados do *framework*. Essa exposição padronizada viabiliza cenários de extensibilidade nos quais agentes de IA externos possam acessar dados de cursos, indicadores e diagnósticos para análises complementares, ampliando o ecossistema de inteligência artificial disponível para a gestão institucional.

5.2.6 Estratégia de Engenharia de *Prompt*

As instruções fornecidas a cada agente de IA foram desenvolvidas seguindo princípios consolidados de engenharia de *prompt* documentados na literatura (WEI *et al.*, 2022; AMATRIAIN, 2024). A estratégia adotada combina três técnicas complementares, articuladas para maximizar a aderência das saídas ao domínio normativo do SINAES.

A primeira técnica consiste na **definição de papel e contexto normativo**. Cada *prompt* inicia com a atribuição de um papel especializado ao agente "especialista em avaliação de cursos de graduação conforme o SINAES", "consultor em melhoria contínua" ou "redator de relatórios institucionais", seguida da descrição do contexto normativo no qual o agente opera. Esse contexto inclui a referência explícita à MPC como base de conhecimento estruturado, a distinção entre instrumentos de Autorização e de Reconhecimento e as regras específicas derivadas da MCA (indicadores filtrados, recortes temporais). A contextualização normativa é parametrizada dinamicamente com os dados do curso em análise como nome, grau acadêmico,

modalidade e tipo de instrumento, assegurando que cada execução do agente receba instruções ajustadas ao perfil específico do curso.

A segunda técnica é a **cadeia de raciocínio estruturada (*chain-of-thought*)**. Para o Agente de Diagnóstico, o *prompt* instrui explicitamente a sequência analítica: (1) identificar o conceito atual; (2) comparar com os atributos dos conceitos superiores (N+1 até 5); (3) listar os *gaps* específicos; (4) classificar a prioridade; (5) considerar o tipo de instrumento. Essa decomposição reduz a incidência de omissões e promove análises sistematicamente completas, conforme demonstrado por Wei *et al.* (2022) em tarefas de raciocínio com modelos de linguagem. Para o Agente de Recomendação, a cadeia instrui a consulta prévia à base de conhecimento, a verificação de artefatos reutilizáveis e a geração de recomendações com cinco componentes obrigatórios: ação, responsável, prazo, evidências e nível de esforço.

A terceira técnica refere-se à **especificação do formato de saída**. Cada agente recebe instruções sobre a estrutura esperada de sua resposta. O Agente de Diagnóstico deve organizar a análise por dimensão e, dentro de cada dimensão, por indicador. O Agente de Síntese recebe a definição das seis seções obrigatórias do relatório e suas respectivas diretrizes de conteúdo. Essa especificação garante a padronização das saídas, requisito indispensável tanto para o processamento *downstream* na cadeia multiagente quanto para a apresentação consistente dos resultados na interface da aplicação.

Complementarmente, todos os *prompts* incorporam regras negativas explícitas como instruções que delimitam o que o agente não deve fazer. Dentre as mais relevantes: "não invente informações e baseie-se apenas nos dados fornecidos"; "se o indicador foi marcado NSA pela MCA, ignore-o completamente"; "diferencie entre previsto (Autorização) e implantado (Reconhecimento)". Essas restrições visam mitigar riscos conhecidos de modelos generativos, como alucinação (geração de informações plausíveis, porém incorretas) e generalização indevida (produção de recomendações genéricas desconectadas do contexto específico do curso).

A calibração iterativa dos *prompts* constitui etapa metodológica relevante. Os *prompts* não foram formulados em versão única e definitiva, mas submetidos a ciclos sucessivos de teste, análise de saídas e refinamento. Cada ciclo envolveu a execução do agente com dados de cursos da UNEMAT, a análise qualitativa das saídas geradas e o ajuste das instruções para corrigir deficiências identificadas como diagnósticos que omitiam indicadores condicionais, recomendações excessivamente genéricas ou relatórios que repetiam o diagnóstico sem sintetizá-lo. Esse processo iterativo está em consonância com a abordagem de *Design Science*

Research adotada na pesquisa, que preconiza ciclos de *design* - avaliação - refinamento até que o artefato atinja nível satisfatório de desempenho (HEVNER *et al.*, 2004).

5.2.7 *Retrieval-Augmented Generation (RAG) e Base de Conhecimento*

A integração entre os agentes de IA e os documentos institucionais do curso como Projeto Pedagógico de Curso (PPC), Diretrizes Curriculares Nacionais (DCN) e planejamentos anteriores, emprega a técnica de *Retrieval-Augmented Generation (RAG)*, que combina recuperação de informações com geração de texto para fundamentar as saídas dos modelos em evidências documentais concretas.

A decisão arquitetural pela adoção de RAG, em detrimento do envio integral dos documentos na janela de contexto do modelo, fundamenta-se em três considerações técnicas. Embora a janela de 200.000 tokens do Claude comporte documentos extensos, o envio integral a cada requisição implicaria custo elevado por chamada, inviabilizando a operação em escala SaaS. Além disso, o envio integral não permite o reaproveitamento de processamento entre chamadas consecutivas. Por fim, a abordagem RAG possibilita a construção de base de conhecimento acumulativa, na qual documentos de múltiplos cursos e ciclos avaliativos são indexados e disponibilizados para consulta contextual pelos agentes, viabilizando a funcionalidade de aprendizado organizacional discutida na seção de desenvolvimento.

O pipeline de processamento de documentos para RAG compreende as seguintes etapas: (a) extração de texto dos documentos enviados em formato PDF ou DOCX, por meio de bibliotecas especializadas; (b) normalização e limpeza textual; (c) segmentação do texto em *chunks* (fragmentos) de aproximadamente 500 a 800 *tokens*, com sobreposição de 100 *tokens* entre *chunks* consecutivos para preservar a continuidade semântica nas fronteiras de segmentação; (d) geração de *embeddings* vetoriais para cada *chunk*, utilizando o modelo *text-embedding-3-small* via Laravel AI SDK; e (e) armazenamento dos *chunks* e respectivos *embeddings* em tabela dedicada no PostgreSQL com suporte à extensão *pgvector*, que habilita busca por similaridade vetorial diretamente no banco de dados relacional.

Quando um agente necessita consultar a base de conhecimento, invoca a ferramenta de busca semântica (BuscarBaseConhecimentoTool), que traduz a consulta do agente em vetor de *embedding*, executa busca por similaridade cossenoidal na tabela de *chunks* e retorna os fragmentos mais relevantes, juntamente com metadados de origem (documento, página, seção). Esse mecanismo permite que os agentes fundamentem suas análises em trechos específicos do

PPC ou das DCN, citando a fonte documental e a página de referência nas recomendações geradas. O limiar mínimo de similaridade é configurado em 0,6, valor empiricamente calibrado para equilibrar precisão (evitar retorno de conteúdo irrelevante) e cobertura (não excluir conteúdo marginalmente relevante).

A base de conhecimento opera em dois escopos complementares. O escopo específico do curso compreende os documentos (PPC, DCN) enviados pelo coordenador e os artefatos produzidos nos planos de ação do próprio curso. O escopo global, por sua vez, acumula conhecimento derivado de planos de ação concluídos em quaisquer cursos da plataforma, após processo de sanitização que remove dados identificadores do *tenant* de origem. Essa camada global implementa, no plano técnico, o conceito de memória institucional inteligente: cursos que iniciam seu processo de planejamento podem se beneficiar de experiências anteriores de outros cursos que enfrentaram gaps semelhantes, sem que haja exposição de dados sensíveis entre as instituições.

5.2.8 Pipeline de Processamento Documental (RAG)

A implementação do *pipeline* RAG compreende serviço especializado (DocumentoProcessorService) que executa a cadeia completa de processamento: extração textual, normalização, segmentação em chunks e geração de embeddings. O processamento é acionado automaticamente ao *upload* de documentos PPC ou DCN pelo coordenador do curso, por meio de *job* assíncrono que opera na fila dedicada *ia-processamento*.

A estratégia de segmentação (*chunking*) adota abordagem hierárquica que preserva a coerência semântica dos fragmentos: o texto é primeiro dividido por seções ou capítulos quando a estrutura do documento permite, depois por parágrafos dentro de cada seção, e finalmente os parágrafos são agrupados até atingir o tamanho-alvo de aproximadamente 700 tokens. A sobreposição de 100 tokens entre *chunks* consecutivos assegura que informações situadas nas fronteiras de segmentação não sejam perdidas. *Chunks* com menos de 100 tokens são descartados ou incorporados ao *chunk* precedente para evitar fragmentos excessivamente curtos que produziriam *embeddings* de baixa qualidade representacional.

Cada *chunk* é armazenado com metadados estruturados em campo JSONB: página de origem, seção, título da seção e informações de contexto documental. Esses metadados são retornados junto com o conteúdo nas buscas por similaridade, permitindo que os agentes citem a fonte com precisão: "conforme o PPC, seção 4.2, página 38". Adicionalmente, cada *chunk*

possui *hash* de conteúdo que permite detectar alterações quando o documento é reenviado, viabilizando a invalidação seletiva de *embeddings* desatualizados sem necessidade de reprocessar toda a base.

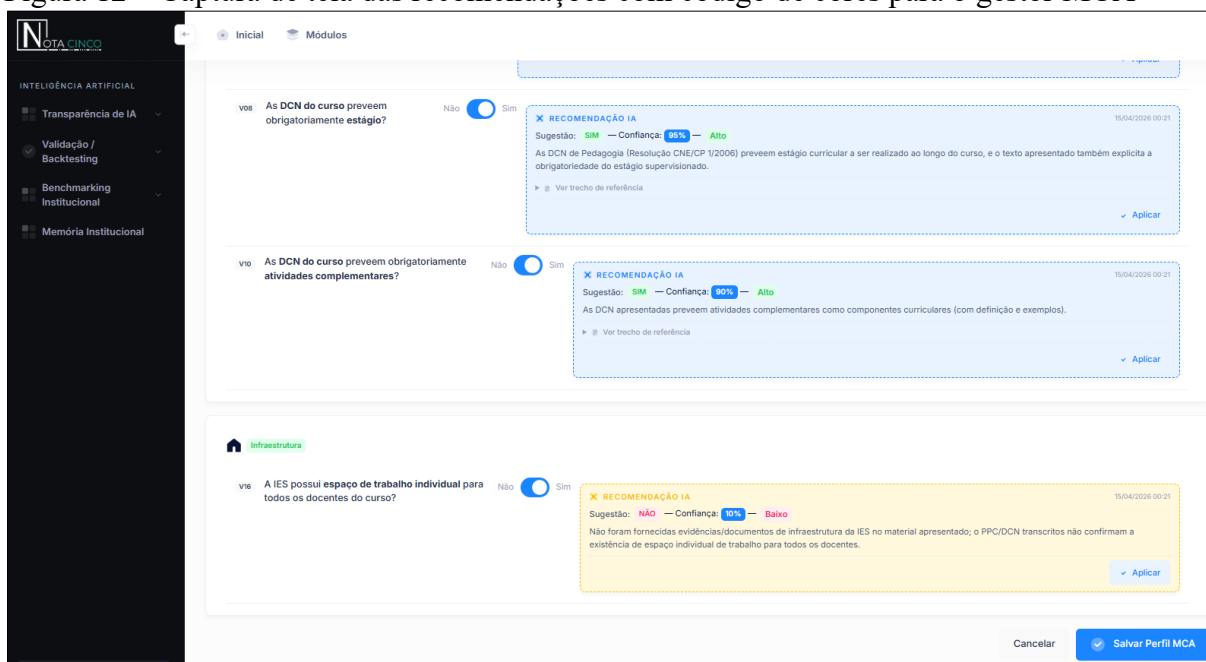
A extensão *pgvector* do PostgreSQL armazena os vetores de *embedding* (1.536 dimensões) em coluna tipada, com índice *IVFFlat* que acelera as buscas por similaridade. Essa decisão de manter os *embeddings* no mesmo banco de dados relacional da aplicação, em vez de utilizar serviço vetorial dedicado (como Pinecone ou Weaviate), fundamenta-se em dois critérios: redução da complexidade operacional, ao eliminar a necessidade de gerenciar um serviço adicional; e viabilidade técnica, dado que o volume de documentos por *tenant*, tipicamente dezenas de PDFs, não milhares, está dentro da capacidade de processamento eficiente do *pgvector*.

5.2.9 Fluxos de Integração com a Aplicação

A integração dos agentes de IA com os fluxos operacionais da aplicação ocorre em três pontos de contato principais, cada qual correspondendo a uma funcionalidade distinta oferecida ao gestor.

Recomendações IA na Matriz de Condições de Aplicabilidade: O primeiro ponto de integração ocorre na interface de configuração do perfil MCA do curso. Quando o coordenador acessa a aba MCA e o curso possui PPC e DCN carregados no sistema, o *framework* oferece a funcionalidade de geração de recomendações assistidas por IA. O Agente de Recomendação analisa os documentos do curso via base RAG e sugere, para cada variável MCA (V04 a V20), o valor recomendado (verdadeiro ou falso) acompanhado de grau de confiança percentual, justificativa textual e referência ao trecho do documento que fundamenta a sugestão. Conforme a Figura 12, as recomendações são apresentadas como cards visuais posicionados abaixo de cada variável, com código de cores (verde para confiança alta, amarelo para média, vermelho para baixa) que orienta o gestor sobre a segurança da sugestão. O gestor pode aceitar, rejeitar ou ignorar cada recomendação; a decisão final permanece sob controle humano. As recomendações são persistidas e invalidadas automaticamente quando os documentos de origem são atualizados.

Figura 12 – Captura de tela das recomendações com código de cores para o gestor MCA



Fonte: Elaborado pelo autor (2026).

Relatório de Avaliação Integrado: O segundo ponto de integração ocorre no contexto dos planos de ação MPC. Após o coordenador configurar um plano de ação com indicadores selecionados, conceitos atuais e conceitos-alvo, o *framework* oferece a geração de relatório executivo por meio da cadeia completa dos três agentes (Diagnóstico → Recomendação → Síntese). A execução é processada assincronamente, com *streaming* da resposta em tempo real na interface por meio de *Server-Sent Events* (SSE), permitindo que o gestor acompanhe a geração do relatório progressivamente. O relatório produzido é armazenado com versionamento: cada alteração solicitada gera nova versão com registro do *prompt* que a originou, assegurando rastreabilidade total do processo de elaboração.

Figura 13 – Captura de tela do Relatório de Avaliação Integrado (Planos de Ação MPC)

Fonte: Elaborado pelo autor (2026).

5.2.10 Base de Conhecimento Global e Aprendizado Organizacional

Uma das contribuições funcionais do *framework* que decorre diretamente da arquitetura RAG é a construção progressiva de base de conhecimento global. Quando um plano de ação é concluído em qualquer *tenant* da plataforma, um *job* assíncrono (IndexarPlanejamentoJob) extrai dados relevantes como indicadores trabalhados, ações implementadas, resultados obtidos e os indexa na base de conhecimento com escopo global (*tenant_id* nulo). Antes da indexação, o *job* executa processo de sanitização que remove nomes de instituição, campus e pessoas, preservando apenas o conteúdo genérico de valor analítico.

Essa base global materializa, no plano computacional, um mecanismo de aprendizado organizacional: as experiências acumuladas de cursos que já enfrentaram e superaram *gaps* avaliativos específicos ficam disponíveis para consulta pelos agentes de quaisquer cursos da plataforma. Um coordenador que inicia o planejamento para elevar o conceito do indicador 1.3 (Perfil do Egresso) pode se beneficiar, via recomendações do agente, de estratégias que se mostraram eficazes em outros cursos que trabalharam o mesmo indicador sem ter conhecimento direto dessa experiência e sem que dados sensíveis sejam expostos.

Esse mecanismo responde a uma das fragilidades recorrentes identificadas na literatura sobre avaliação institucional: a perda de memória organizacional decorrente da rotatividade de gestores e da descontinuidade dos processos de autoavaliação (ANDRIOLA, 2009). Ao codificar experiências bem-sucedidas em base de conhecimento consultável por agentes inteligentes, o *framework* transforma conhecimento tácito institucional em

conhecimento explícito reutilizável, contribuindo para a construção do que a literatura de gestão denomina organizações que aprendem (*learning organizations*).

5.2.11 Transparência, Auditoria e Uso Responsável de IA

A integração de IA generativa em processos de gestão institucional exige mecanismos de transparência e auditabilidade que assegurem a confiabilidade dos resultados e a conformidade ética do sistema. O *framework* implementa essas exigências por meio de três camadas complementares.

A primeira camada é o **painel de transparência de IA**, que registra automaticamente por meio de middleware interceptador integrado ao Laravel AI SDK cada interação dos agentes com o provedor de IA. Os registros incluem: o agente executado, o modelo e provedor utilizados, o resumo do *prompt* enviado, os tokens de entrada e saída consumidos, o custo estimado da requisição em dólares, o tempo de resposta em milissegundos, a quantidade de ferramentas invocadas pelo agente e o status da operação (sucesso, erro ou timeout). Esses registros são persistidos em tabela dedicada (*ia.logs_agentes*) e disponibilizados aos gestores como painel de monitoramento, permitindo acompanhar o uso de IA e os custos associados.

A segunda camada compreende os **procedimentos de validação** descritos na seção 3.2.2, cuja implementação no *framework* está integrada ao módulo de IA. O *backtesting* compara as recomendações geradas pela IA com os resultados de avaliações *in loco* já realizadas; os testes de consistência verificam se a mesma entrada produz saídas equivalentes em execuções distintas; e a validação por *benchmark* confronta os *outputs* com os critérios dos instrumentos oficiais do INEP.

A terceira camada refere-se ao **princípio de supervisão humana** que permeia toda a arquitetura de integração. Em nenhum ponto do fluxo operacional as análises ou recomendações da IA são aplicadas automaticamente sem intervenção do gestor. As recomendações MCA requerem aceitação individual; o relatório de avaliação é apresentado como proposta editável. Essa arquitetura de "*human-in-the-loop*" alinha-se aos princípios de uso responsável de IA preconizados pela Estratégia Brasileira de Inteligência Artificial (BRASIL, 2021) e assegura que o *framework* funcione como instrumento de apoio à decisão e não como substituto do julgamento humano na gestão educacional.

5.2.12 Síntese da Integração com IA

A integração da inteligência artificial generativa ao *framework* proposto configura-se como camada computacional que articula os artefatos analíticos (MPC, MCA, MCE) com as capacidades de processamento de linguagem natural dos modelos de IA contemporâneos. A arquitetura multiagente, com três agentes especializados, ferramentas de acesso a dados, pipeline RAG e base de conhecimento global, materializa a proposta de automatizar a descoberta de conhecimento nos dados avaliativos do SINAES.

Do ponto de vista da gestão institucional, a integração transforma o *framework* de um sistema de informação para registro e cálculo de indicadores em uma plataforma de inteligência institucional capaz de produzir diagnósticos fundamentados, recomendações contextualizadas e relatórios executivos que sintetizam informações de múltiplas fontes. A base de conhecimento global, que acumula experiências de cursos da plataforma, agrega a dimensão de aprendizado organizacional ao sistema, respondendo à fragmentação de conhecimento que a literatura identifica como um dos principais obstáculos à eficácia da autoavaliação institucional (LEITE, 2017; ANDRIOLA, 2009).

As decisões arquiteturais documentadas nesta seção como modularização do código, orquestração sequencial dos agentes, calibração iterativa dos *prompts*, mecanismos de transparência e supervisão humana refletem a busca por uma solução que conjugue inovação tecnológica com rigor metodológico e responsabilidade ética, em conformidade com os princípios do *Design Science Research* e com as diretrizes da EBIA para aplicação de IA no setor educacional.

6 RESULTADOS E DISCUSSÃO

Este capítulo apresenta os resultados obtidos na etapa de avaliação do artefato conforme o modelo DSRM (PEFFERS *et al.*, 2007).

6.1 Validação por Cenários Controlados

A validação por cenários controlados constitui o primeiro procedimento de avaliação do framework, conforme definido na seção 3.2.2.1. Este procedimento submete o *framework* a situações avaliativas predefinidas cujos resultados esperados (*ground truth*) são conhecidos pelo pesquisador, permitindo mensurar a correção lógica do sistema de forma objetiva e reproduzível, seguindo a abordagem de avaliação baseada em cenários (*scenario-based evaluation*) preconizada por Hevner *et al.* (2004).

Foram executadas quatro sessões de validação com o modelo Claude Sonnet 4.6 (Anthropic), totalizando 21 execuções independentes. Cada sessão processou os três cenários controlados definidos na metodologia, com repetições que permitiram avaliar tanto a correção funcional quanto a consistência dos resultados. O custo total de processamento foi de USD 21,93 (média de USD 1,04 por execução), com duração média de inferência de 483.958 milissegundos (8,05 minutos) por execução.

6.1.1 Configuração dos Cenários

Foram construídos três cenários sintéticos representando perfis distintos de cursos de graduação da Universidade do Estado de Mato Grosso (UNEMAT). Todos os cursos estão cadastrados no sistema com dados reais (grau acadêmico, modalidade, área de conhecimento), porém os conceitos atribuídos a cada indicador e os perfis da Matriz de Condições de Aplicabilidade (MCA) são sintéticos, criados especificamente para exercitar diferentes comportamentos do *framework*, conforme descrito na seção 3.2.2.1. A Tabela 22 sintetiza a configuração dos cenários.

Tabela 22 – Caracterização dos cenários controlados.

Cenário	Curso	Instrumento	Perfil de conceitos	Dimensão prioritária esperada	Propósito
A	Tecnologia em Gestão Turismo (Nova Xavantina)	Reconhecimento (D1=35%, D2=20%, D3=45%)	18 indicadores C2 em D1; D2 predominantemente C4	D1	Verificar detecção de alta concentração de gaps críticos em dimensão específica
B	Pedagogia (Juara)	Reconhecimento (D1=35%, D2=20%, D3=45%)	Todos os conceitos 3; 14 indicadores NSA	Nenhuma (empate)	Edge-case: capacidade discriminativa sem gaps críticos; tratamento de indicadores NSA
C	Medicina (Cáceres)	Autorização (D1=40%, D2=35%, D3=25%)	D3 com 4 C1 + 3 C2; D2 com 11 C5	D3	Contraste extremo entre dimensões; instrumento com pesos distintos

Fonte: Elaborado pelo autor (2026).

O Cenário A apresenta predominância de conceitos baixos (C2) na Dimensão 1, com 18 dos 24 indicadores nesse patamar, gerando alta concentração de gaps críticos e uma dimensão prioritária claramente identificável. O Cenário B constitui *edge-case* (caso limite) em que todos os indicadores aplicáveis possuem conceito 3, configurando cenário sem gaps críticos e sem dimensão prioritária destacada, e verificando adicionalmente o tratamento de 14 indicadores marcados como "não se aplica" (NSA) pela MCA. O Cenário C utiliza instrumento de Autorização (pesos 40/35/25, distintos dos pesos 35/20/45 do Reconhecimento), com contraste extremo entre uma dimensão em patamar máximo (D2, com 11 indicadores C5) e outra em situação de alta fragilidade (D3, com 4 indicadores C1 e 3 C2).

6.1.2 Resultados das métricas de classificação de prioridade

6.1.2.1 Desempenho geral

A Tabela 23 apresenta o desempenho consolidado do *framework* considerando todas as 21 execuções e os três cenários.

Tabela 23 – Desempenho geral do *framework* nos cenários controlados

Métrica	Valor
Total de execuções	21
F1 Geral (média)	0,8546
F1 Geral (DP)	0,1981
F1 Geral (mínimo)	0
F1 Geral (máximo)	0,933
Limiar de aceitação (F1 \geq 0,70)	Atendido
Custo total (USD)	21,93
Custo por execução (USD)	1,04
Duração média por execução (ms)	483.958

Fonte: Elaborado pelo autor (2026).

O *framework* obteve F1 Geral médio de 0,8546 (DP = 0,1981) considerando todas as 21 execuções, superando o limiar de aceitação de F1 \geq 0,70. A dispersão observada (DP = 0,1981) é influenciada por uma execução anômala com F1 = 0 no Cenário C, cuja análise detalhada é apresentada na seção 6.1.4. Todas as métricas consolidadas nesta seção e nas subsequentes incluem a execução anômala, adotando como princípio de que todos os dados coletados devem ser reportados. A título de contextualização, o F1 médio das 20 execuções restantes é de 0,8973 (DP = 0,0335), valor que permite dimensionar o impacto do outlier sobre a métrica consolidada, mas que não substitui o F1 de 0,8546 como resultado oficial do procedimento.

6.1.2.2 Desempenho por cenário

A análise por cenário revela padrões distintos de desempenho associados às diferentes configurações de perfis avaliativos, conforme pode ser observado na Tabela 24.

Tabela 24 – F1 Geral e acerto de dimensão prioritária por cenário

Cenário	Execuções	F1 Geral (média)	F1 (DP)	F1 Min.	F1 Max.	Dimensão prioritária (acertos/total)	Taxa de acerto
A	7	0,8674	0	0,8674	0,8674	7/7	100%
B	7	0,8968	0,0247	0,8698	0,9219	4/7	57,14%

Cenário	Execuções	F1 Geral (média)	F1 (DP)	F1 Min.	F1 Max.	Dimensão prioritária (acertos/total)	Taxa de acerto
C	7	0,7997	0,3526	0	0,933	6/7	85,71%

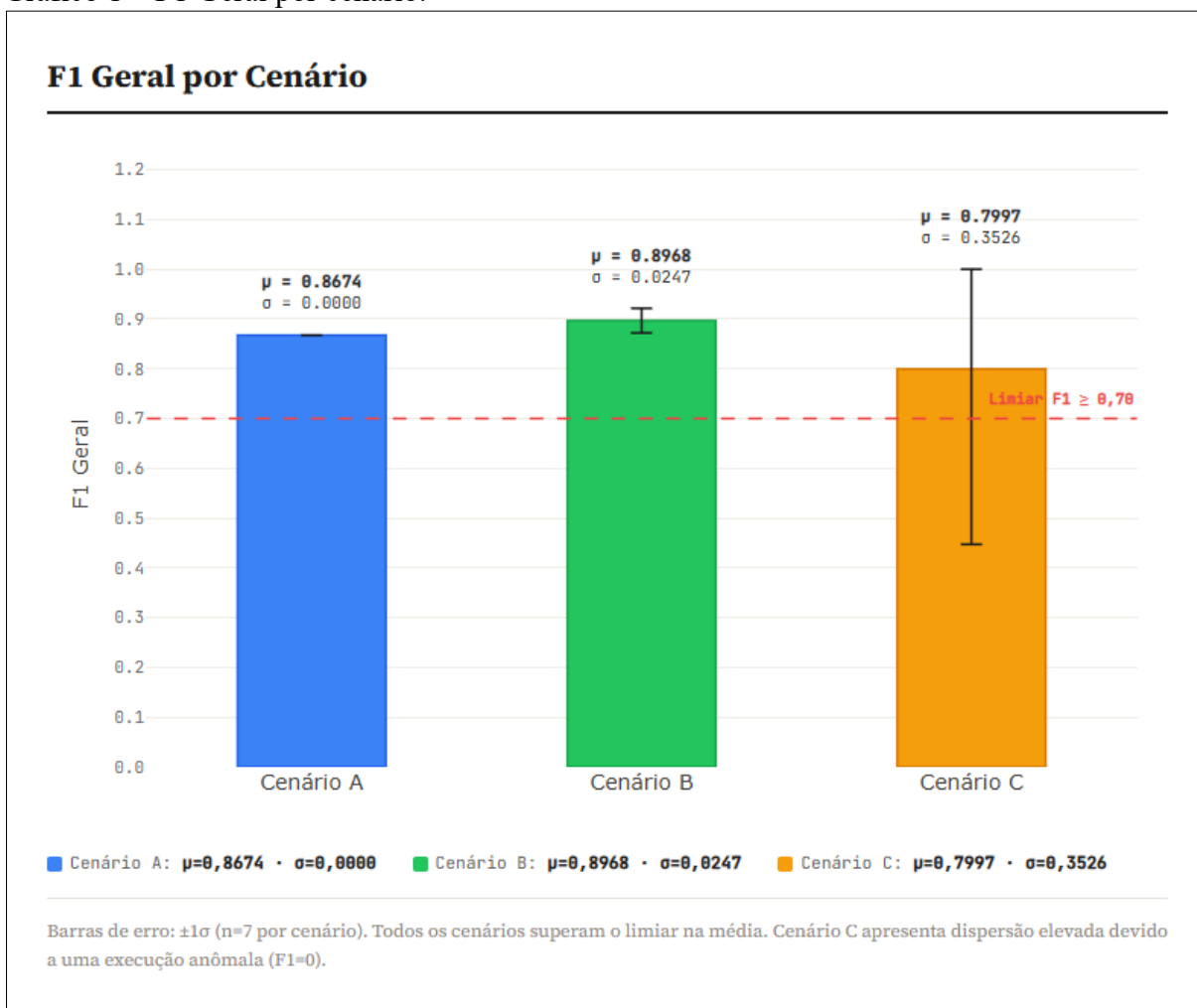
Fonte: Elaborado pelo autor (2026).

Cenário A (predominância de *gaps* críticos em D1). O *framework* apresentou desempenho com variabilidade nula, com F1 Geral de 0,8674 e desvio-padrão zero em todas as sete execuções, indicando reprodutibilidade perfeita dos elementos estruturais do diagnóstico. A identificação da dimensão prioritária (D1) foi correta em 100% das execuções. Este resultado demonstra que, quando o cenário apresenta perfil de fragilidade inequívoco, com alta concentração de *gaps* críticos em dimensão específica, o *framework* produz diagnósticos determinísticos apesar da natureza estocástica do modelo de linguagem.

Cenário B (*edge-case* sem *gaps* críticos). O F1 Geral médio de 0,8968 (DP = 0,0247) indica desempenho elevado e consistente. Contudo, um achado relevante refere-se à identificação da dimensão prioritária: o *framework* acertou em apenas 4 das 7 execuções (57,14%). Este cenário, por definição, não possui dimensão prioritária destacada (todos os indicadores possuem conceito 3), e o *ground truth* aceita qualquer resposta como válida. A taxa de 57,14% sinaliza que, em algumas execuções, o agente apontou uma dimensão específica como mais frágil quando nenhuma se destacava, revelando tendência em identificar padrões prioritários mesmo em cenários homogêneos. Embora essa tendência não constitua erro segundo o critério de aceitação adotado, ela indica que o agente opera com viés de ação, preferindo emitir orientação específica a declarar ausência de prioridade.

Cenário C (contraste extremo entre dimensões). O F1 Geral médio de 0,7997 apresentou alta dispersão (DP = 0,3526), atribuível a uma execução anômala com F1 = 0. As demais seis execuções produziram F1 = 0,933 de forma consistente (DP = 0), com identificação correta da dimensão prioritária (D3) em 6 das 7 execuções (85,71%). Este cenário, que combina instrumento de Autorização (pesos distintos) com contraste extremo entre dimensões, demonstra que o *framework* se adapta a diferentes configurações de pesos dimensionais e identifica corretamente a dimensão com maior concentração de fragilidades.

Gráfico 1 – F1 Geral por cenário.



Fonte: Elaborado pelo autor (2026).

6.1.2.3 Métricas por categoria de prioridade

A Tabela 25 detalha as métricas de *Precision*, *Recall* e *F1-Score* por categoria de prioridade para cada cenário.

Tabela 25 – Métricas de classificação por categoria de prioridade

Cenário	Categoria	<i>Precision</i> (média)	<i>Recall</i> (média)	F1 (média)	F1 (DP)
A	Crítico	0,7000	1,0000	0,8235	0
A	Médio	0,7333	1,0000	0,8462	0
A	Baixo	0,6667	1,0000	0,8000	0
A	NSA	1,0000	1,0000	1,0000	0

Cenário	Categoria	<i>Precision</i> (média)	<i>Recall</i> (média)	F1 (média)	F1 (DP)
B	Crítico	1,0000	1,0000	1,0000	0
B	Médio	0,6330	0,9805	0,7579	0,1051
B	Baixo	1,0000	1,0000	1,0000	0
B	NSA	0,7515	1,0000	0,8294	0,2026
C	Crítico	0,6429	0,8571	0,7347	0,324
C	Médio	0,6667	0,8571	0,7500	0,3307
C	Baixo	0,8571	0,8571	0,8571	0,378
C	NSA	0,8571	0,8571	0,8571	0,378

Fonte: Elaborado pelo autor (2026).

A Tabela 25 de *Precision*, *Recall* e F1 por categoria é complementada pela Tabela 26, matriz de confusão a seguir, que torna explícita a origem dos falsos positivos reportados. A matriz considera a execução representativa do Cenário A (as sete execuções produziram classificações idênticas, com F1 Geral = 0,8674 e DP = 0).

Tabela 26 – Matriz de confusão do Cenário A (execução representativa, N = 43 indicadores)

Esperado ↓ / Previsto →	Crítico	Médio	Baixo	NSA	Total esperado
Crítico	14	0	0	0	14
Médio	0	11	0	0	11
Baixo	0	0	10	0	10
NSA	6	4	5	0	8*
Total previsto	20	15	15	0	50*

Fonte: Elaborado pelo autor (2026), a partir dos resultados dos cenários controlados. *A soma "Total previsto" (50) excede o total de indicadores rastreados pelo *ground truth* (N = 43) porque o agente incluiu no diagnóstico indicadores cujo perfil MCA os classifica como não aplicáveis (NSA) ao curso avaliado (Turismo - Bacharelado). Esses indicadores, que deveriam ter sido omitidos pelo agente segundo a lógica da MCA, foram distribuídos entre as categorias Crítico (6), Médio (4) e Baixo (5), gerando 15 falsos positivos. A estrutura resultante não configura classificação multilabel no sentido estrito, cada indicador foi classificado pelo agente em uma única categoria, mas a inclusão de indicadores fora do escopo do *ground truth* produz a soma excedente observada. As métricas de *Precision*, *Recall* e *F1-Score* foram calculados por categoria com base na interseção entre output do agente e *ground truth*, tratamento que permanece válido para classificação multiclasse quando a população inclui tanto indicadores rastreados quanto indicadores indevidamente promovidos.

É importante esclarecer a origem dos falsos positivos. O *ground truth* do Cenário A rastreia 43 indicadores: 35 com *gap* atribuído (14 críticos, 11 médios, 10 baixos) e 8

classificados como NSA pela MCA (indicadores 1.13, 1.14, 2.7, 3.2, 3.7, 3.8, 3.9, 3.10). Os 15 falsos positivos identificados pelo agente (indicadores como 1.8, 1.9, 1.18, 1.22, 1.24 e 3.18, entre outros) são indicadores que também deveriam ser NSA segundo a MCA, por exemplo, o indicador 1.8 (Estágio curricular supervisionado) é NSA porque a variável V11 = false (curso não é licenciatura). O agente, ao não reconhecer a precedência lógica do perfil MCA, incluiu esses indicadores no diagnóstico e atribuiu-lhes criticidade com base no conceito registrado, gerando os falsos positivos. Nenhum indicador com *gap* real foi suprimido (zeros fora da diagonal nas três primeiras linhas). Esse padrão, *Recall* perfeito para categorias aplicáveis associado à promoção indevida de indicadores NSA, indica que o agente é conservador na detecção de fragilidades, mas apresenta limitação na aplicação integral do perfil MCA.

Exemplo qualitativo 1 – Falso positivo por ignorar o perfil MCA (Cenário A, execução 1).

Indicador 1.8 – Estágio curricular supervisionado – relação com a rede de escolas da educação básica.

Perfil MCA do curso (Turismo - Bacharelado - Nova Xavantina): V11 = false (curso não é licenciatura; indicador 1.8 é NSA segundo a lógica determinística da MPC/MCA).

Ground truth: NSA (não se aplica).

Saída do agente (trecho literal do diagnóstico):

- **Indicador:** 1.8 Estágio curricular supervisionado - relação com a rede de escolas da educação básica.
- **Conceito atual** → alvo: 2 → 5.
- **Prioridade:** crítico.
- **Gaps:** Ausência de práticas inovadoras para gestão da relação IES-rede de escolas, com resultados.
- **Evidências necessárias:** Documentação de práticas inovadoras na relação IES-escolas.
- **Interpretação.** O agente atribuiu criticidade a um indicador que a MCA sinaliza como inaplicável ao curso (Turismo, bacharelado, sem estágio em rede de escolas). A falha não está na análise do conteúdo, o texto gerado é coerente com a descrição do indicador, mas na precedência lógica: o agente deveria ter reconhecido a variável MCA V11 = false antes de emitir classificação. O mesmo padrão ocorreu com os indicadores 1.9, 1.18, 1.22, 1.24 e 3.18 nesta execução, totalizando seis falsos positivos na categoria Crítico.

A análise das métricas por categoria revela padrões relevantes sobre o comportamento do *framework*:

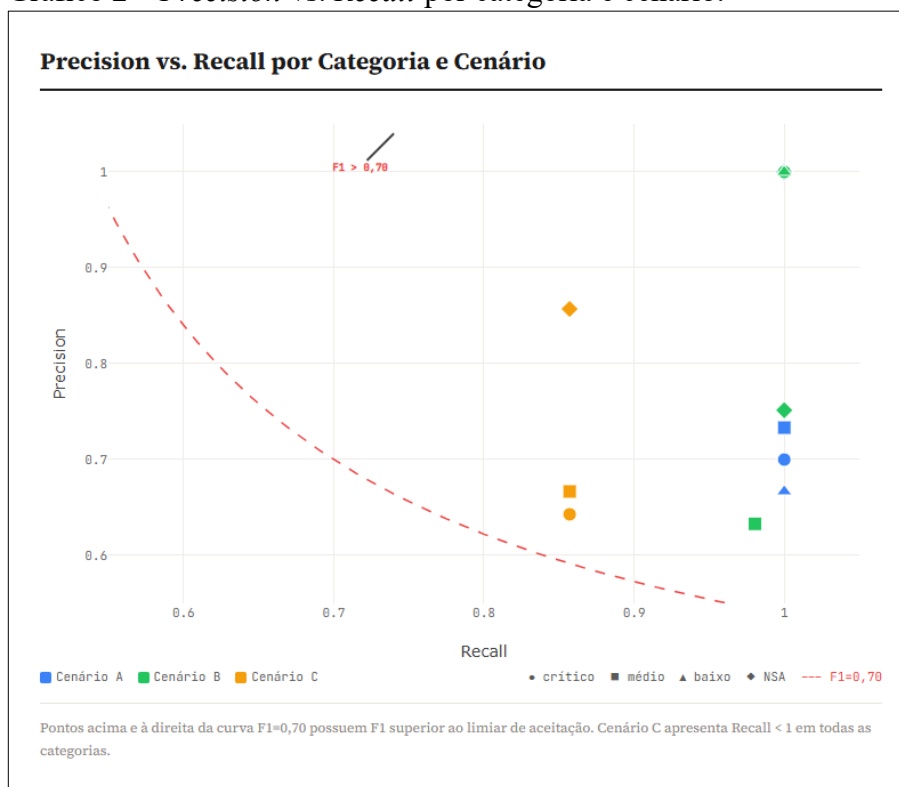
Padrão de *Recall* elevado com *Precision* moderada. No Cenário A, o *framework* obteve *Recall* de 1,0000 em todas as categorias, indicando que identificou todos os indicadores que deveriam ser classificados em cada categoria, sem falsos negativos. A *Precision* mais

moderada (entre 0,6667 e 1,0000) revela tendência de sobreclassificação: o agente classificou alguns indicadores em categorias nas quais eles não pertenciam segundo o *ground truth*. Dos 14 indicadores esperados como críticos, o agente identificou 20, incluindo 6 falsos positivos (indicadores 1.8, 1.9, 1.18, 1.22, 1.24 e 3.18). Padrão semelhante ocorreu nas categorias "médio" (11 esperados, 15 identificados) e "baixo" (10 esperados, 15 identificados). Esse comportamento sugere que o agente adota postura conservadora, preferindo sinalizar potenciais fragilidades a omiti-las. No contexto de avaliação educacional, esse comportamento é consistente com a priorização de sensibilidade sobre especificidade: a omissão de um *gap* crítico (falso negativo) tem consequências potencialmente mais severas do que a sinalização excessiva (falso positivo), cuja verificação posterior pode ser realizada pelo coordenador de curso. A adequação dessa priorização depende, contudo, do contexto específico de uso, cuja verificação posterior pode ser realizada pelo coordenador de curso.

Tratamento de indicadores NSA. O *framework* obteve $F1 = 1,0000$ para NSA no Cenário A (8 indicadores NSA, todos corretamente omitidos), demonstrando capacidade de reconhecer indicadores não aplicáveis quando o perfil MCA é inequívoco. No Cenário B, com 14 indicadores NSA em perfil MCA mínimo (todas as variáveis desativadas), o F1 de NSA foi 0,8294 ($Precision = 0,7515$, $Recall = 1,0000$), indicando que, embora todos os indicadores NSA tenham sido corretamente omitidos ($Recall = 1,0$), alguns indicadores aplicáveis foram indevidamente tratados como NSA ($Precision < 1,0$).

Capacidade discriminativa no Cenário B (*edge-case*). O $F1 = 1,0000$ na categoria "crítico" do Cenário B confirma que nenhum indicador foi erroneamente classificado como crítico em um cenário onde nenhum *gap* crítico existe. Este resultado é significativo, pois demonstra que o *framework* não gera falsos alarmes de criticidade na ausência de indicadores com conceitos 1 ou 2. A capacidade de não emitir alertas quando não há fragilidades severas é relevante para a credibilidade do artefato, uma vez que um sistema que consistentemente gera falsos alarmes tenderia a ser desconsiderado pelos usuários.

Desempenho no Cenário C com instrumento de Autorização. As métricas do Cenário C apresentam dispersão mais elevada (F1 DP entre 0,324 e 0,378), influenciada pela execução anômala. Nas execuções válidas, o *framework* demonstrou capacidade de operar com o instrumento de Autorização e seus pesos específicos (40/35/25), identificando corretamente a dimensão D3 como prioritária. A categoria "crítico" obteve $Recall$ de 0,8571, com 6 dos 7 indicadores críticos corretamente identificados na média das execuções válidas.

Gráfico 2 – *Precision vs. Recall* por categoria e cenário.

Fonte: Elaborado pelo autor (2026).

6.1.3 Resultados das métricas de reconhecimento de evidências (MCE)

A Tabela 27 apresenta os resultados consolidados das sete métricas MCE para cada cenário, obtidos a partir das execuções que incluíram a validação de artefatos de evidência.

Tabela 27 – Métricas MCE por cenário

Cenário	TRA (M1)	TAO (M2)	TDE (M3)	ICC-N2 (M4)	TFNM (M5)	TDR (M6)	Qualidade e N3 (M7)
A	1,0000 (DP=0)	0,5960 (DP=0,0463)	0,6465 (DP=0,0463)	0 (DP=0)	0 (DP=0)	0 (DP=0)	0 (DP=0)
B	0,9047 (DP=0,0825)	0,5128 (DP=0,0222)	0,7923 (DP=0,0453)	0 (DP=0)	0 (DP=0)	0 (DP=0)	0 (DP=0)
C	0,9583 (DP=0,0417)	0,6111 (DP=0,1531)	0,5794 (DP=0,0728)	0 (DP=0)	0,0417 (DP=0,0417)	0 (DP=0)	0 (DP=0)

Fonte: Elaborado pelo autor (2026).

A análise das métricas MCE revela capacidades e limitações do *framework* no reconhecimento e orientação de evidências institucionais:

Taxa de Reconhecimento de Artefatos (TRA). O *framework* atingiu o limiar de aceitação ($\geq 80\%$) em todos os cenários, com TRA variando entre 0,9047 e 1,0000. No Cenário A, a TRA foi de 1,0000 (DP = 0), indicando que o agente reconheceu todos os artefatos existentes em todas as execuções. Nos Cenários B e C, a TRA foi de 0,9047 (DP = 0,0825) e 0,9583 (DP = 0,0417), respectivamente. Esses resultados demonstram que o agente consulta efetivamente a base de evidências antes de formular recomendações, atendendo ao requisito fundamental da MCE.

Taxa de Adequação de Orientação (TAO). O *framework* não atingiu o limiar de aceitação de 70% em nenhum cenário, com TAO variando entre 0,5128 e 0,6111. A análise detalhada das divergências revela um padrão recorrente: quando o *ground truth* espera classificação "V" (vincular, indicando que o artefato existente é completo e suficiente), o agente frequentemente recomenda "VP" (vincular com complemento parcial), sugerindo a produção de documentação complementar mesmo quando o artefato existente seria suficiente. Essa tendência revela postura conservadora do agente na orientação sobre evidências. Do ponto de vista prático, a recomendação de complemento não invalida a vinculação ao artefato existente e pode, em alguns casos, contribuir para o fortalecimento das evidências. Contudo, a métrica TAO captura essa divergência como erro de classificação, reduzindo o valor observado.

Taxa de Diferenciação de Escopo (TDE). O *framework* superou o limiar de aceitação ($\geq 60\%$) nos Cenários A (64,65%) e B (79,23%), mas ficou ligeiramente abaixo no Cenário C (57,94%). O resultado do Cenário B (79,23%) é particularmente relevante, indicando que o agente diferencia adequadamente evidências de escopo institucional (que mencionam compartilhamento entre cursos), parcial (que mencionam complemento específico do curso) e de curso (sem menção a compartilhamento) na maioria dos casos.

Índice de Compartilhamento Cross-Curso (ICC-N2) e Taxa de Descoberta RAG (TDR). Ambas as métricas permaneceram em zero para todos os cenários. O resultado do ICC-N2 indica que o agente não conseguiu identificar artefatos compartilhados entre cursos via navegação em relações *cross-curso* (Nível 2), o que representa limitação relevante do *framework* na sua configuração atual. A TDR em zero indica que a descoberta semântica de artefatos via busca na base de conhecimento RAG (Nível 3) não foi alcançada em nenhuma execução. Conforme definido na metodologia, essas métricas possuem caráter exploratório, e os resultados apontam para oportunidades de refinamento na engenharia de *prompts* e na configuração das ferramentas de busca disponíveis aos agentes.

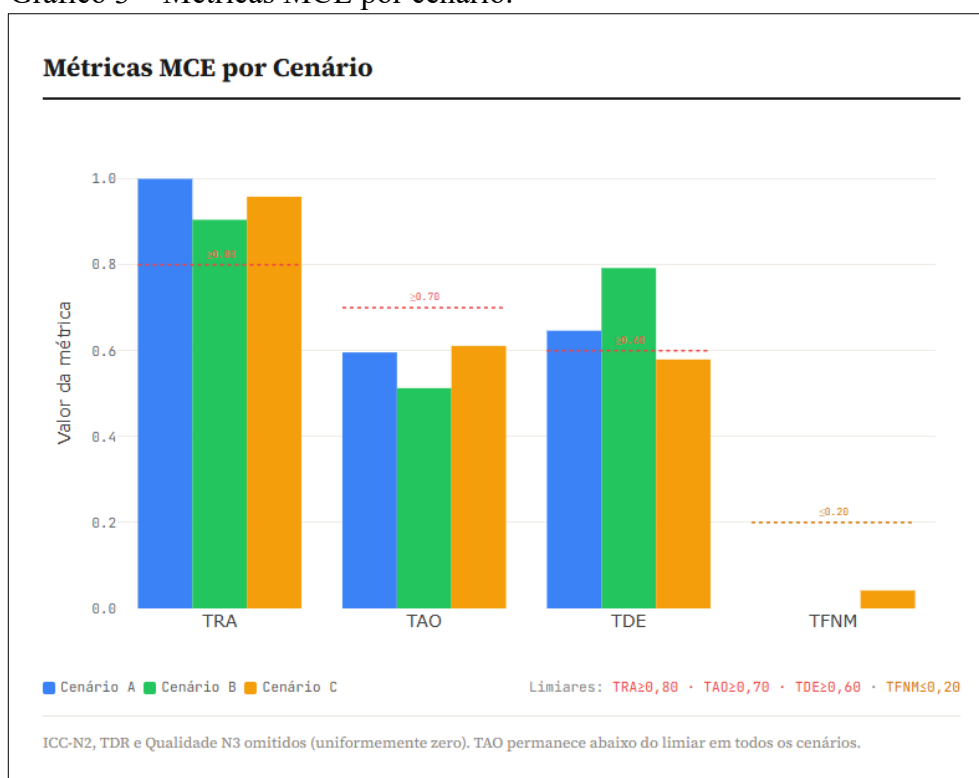
Taxa de Falsos Negativos MCE (TFNM). O *framework* atendeu ao limiar de aceitação ($\leq 20\%$) em todos os cenários, com TFNM entre 0 e 0,0417. Este resultado indica que o agente raramente recomendou produção de novo artefato quando já existia um disponível na base de evidências, corroborando a elevada TRA observada.

Tabela 27 – Síntese do atendimento aos limiares MCE

Métrica	Limiar	Cenário A	Cenário B	Cenário C
TRA ($\geq 80\%$)	$\geq 0,80$	1,0000 (atendido)	0,9047 (atendido)	0,9583 (atendido)
TAO ($\geq 70\%$)	$\geq 0,70$	0,5960 (não atendido)	0,5128 (não atendido)	0,6111 (não atendido)
TDE ($\geq 60\%$)	$\geq 0,60$	0,6465 (atendido)	0,7923 (atendido)	0,5794 (não atendido)
ICC-N2 ($\geq 70\%$)	$\geq 0,70$	0 (não atendido)	0 (não atendido)	0 (não atendido)
TFNM ($\leq 20\%$)	$\leq 0,20$	0 (atendido)	0 (atendido)	0,0417 (atendido)
TDR	Exploratório	0	0	0

Fonte: Elaborado pelo autor (2026).

Gráfico 3 – Métricas MCE por cenário.



Fonte: Elaborado pelo autor (2026).

6.1.4 Estabilidade dos resultados entre execuções

A execução repetida dos cenários controlados ao longo de quatro sessões distintas permite observar a estabilidade dos resultados produzidos pelo *framework*. A análise a seguir reporta a variabilidade observada nas métricas entre as 7 execuções de cada cenário, como elemento complementar à avaliação da correção funcional. A análise formal de consistência interna, com procedimento e métricas próprias, é apresentada em seção dedicada.

Tabela 28 – Variabilidade das métricas de classificação entre execuções

Cenário	F1 Geral (média)	F1 Geral (DP)	Coefficiente de Variação	Execuções com resultado estável
A	0,8674	0	0%	7/7 (100%)
B	0,8968	0,0247	2,75%	7/7 (variação marginal: 0,8698 a 0,9219)
C	0,7997	0,3526	44,09%	6/7 (1 outlier com F1 = 0)

Fonte: Elaborado pelo autor (2026).

No Cenário A, todas as sete execuções produziram F1 Geral idêntico de 0,8674, com os mesmos indicadores classificados nas mesmas categorias de prioridade e a mesma dimensão prioritária identificada. Esse resultado indica que, para cenários com perfil de fragilidade inequívoco (alta concentração de *gaps* críticos em dimensão específica), o *framework* produz diagnósticos estáveis entre execuções.

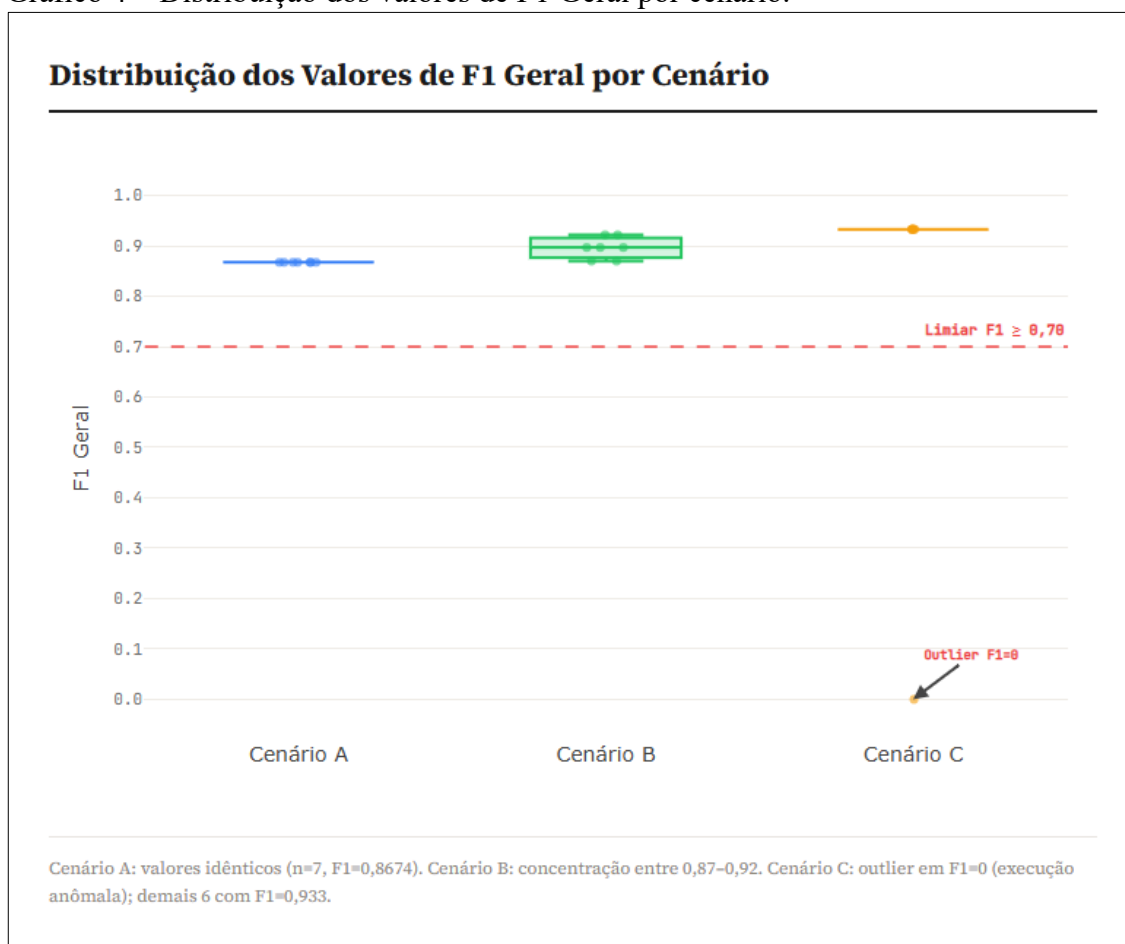
No Cenário B, a variação foi marginal (DP = 0,0247, CV = 2,75%), com F1 oscilando entre 0,8698 e 0,9219. A variação concentra-se nas categorias "médio" (F1 DP = 0,1051) e "NSA" (F1 DP = 0,2026), sugerindo que cenários homogêneos (todos os indicadores com conceito 3) oferecem menos sinais discriminativos ao agente, resultando em maior sensibilidade à variabilidade inerente ao modelo.

No Cenário C, o coeficiente de variação elevado (44,09%) é inteiramente atribuível a uma execução anômala com F1 = 0. As demais seis execuções produziram F1 = 0,933 de forma estável (DP = 0). A inspeção da saída textual da execução anômala (Exemplo qualitativo 2) revela que o agente adotou variação estilística no formato dos cabeçalhos e marcadores de

prioridade, resultando em falha do *parser regex* na extração dos elementos estruturais. O conteúdo semântico do diagnóstico, identificação de D3 como dimensão prioritária e classificação do indicador 1.4 como crítico, é coerente com o *ground truth*, porém a verificação completa do conteúdo contra todos os indicadores esperados não foi conduzida de forma exaustiva, de modo que a atribuição da falha exclusivamente ao *parser* deve ser tratada como hipótese mais provável, e não como conclusão definitiva. A taxa de ocorrência dessa falha (1 em 21 execuções, equivalente a 4,76%) indica que, para operação em produção, mecanismos de validação de formato, re-tentativa automática ou adoção de output estruturado nativo seriam recomendáveis.

Estabilidade das métricas MCE. As métricas MCE apresentaram desvios-padrão próximos de zero na maioria dos cenários: TRA (DP entre 0 e 0,0825), TAO (DP entre 0,0222 e 0,1531) e TDE (DP entre 0,0453 e 0,0728). A TRA apresentou DP = 0 no Cenário A (todas as execuções com TRA = 1,0000), indicando estabilidade no comportamento de consulta à base de evidências.

Gráfico 4 – Distribuição dos valores de F1 Geral por cenário.



Fonte: Elaborado pelo autor (2026).

Exemplo qualitativo 2 – Execução anômala do Cenário C (F1 = 0 por falha de *parsing*)

Sessão: 20260410_153457, Cenário C, execução 1.

Diagnóstico esperado: D3 como dimensão prioritária; 9 críticos, 14 médios, 7 baixos, 7 NSA.

Resultado: F1 Geral = 0; nenhum indicador foi identificado em qualquer categoria (Identificados = 0 em todas as linhas); dimensão prioritária registrada como 0 (indefinida).

Saída textual do agente (cabeçalho literal):

- **Relatório de Análise de Gaps de Progressão Conceitual**
- **Curso:** Medicina – Bacharelado Presencial
- **Instrumento:** Autorização | CC Calculado: 2,9
- **Indicador:** 1.4 – Estrutura Curricular
- **Conceito Atual:** 2 → **Conceito Alvo:** 5
- **Prioridade:** CRÍTICO
- **Diagnóstico técnico.** O conteúdo gerado pelo agente está semanticamente correto: identifica D3 como dimensão crítica (1,82 de média, peso 40%) e classifica o indicador 1.4 como crítico. A falha encontra-se no pipeline de *parsing*, que espera o marcador literal - ****Prioridade:**** ****crítico**** (minúsculas, sem *emoji*, *bullet* com dois asteriscos). Nesta execução, o agente adotou variação estilística - ****Prioridade: CRÍTICO**** com cabeçalho **### 1.4 - Estrutura Curricular** em lugar de **### 1.4 Estrutura Curricular** - que o *parser regex* não consegue processar. O resultado é zero classificações extraídas, ainda que o diagnóstico textual esteja correto.
- **Implicação para o pipeline.** A ocorrência (1 em 21 execuções, equivalente a 4,76%) não caracteriza falha do agente na compreensão do cenário, mas vulnerabilidade do *parser* a variações de formato. Estratégias de mitigação discutidas na seção 6.1.5 (*structured output*, re-tentativa automática, validação prévia do formato) são diretamente motivadas por este caso.

6.1.5 Discussão dos resultados

Os resultados da validação por cenários controlados permitem realizar considerações em quatro eixos: correção funcional, reconhecimento de evidências, robustez do pipeline e alinhamento com a DSRM.

Correção funcional. O F1 Geral médio de 0,8546 supera o limiar de aceitação (0,70) e situa-se em patamar que, na literatura de classificação de textos, é considerado "bom" (SOKOLOVA; LAPALME, 2009). O padrão de *Recall* elevado com *Precision* moderada observado no Cenário A indica que o *framework* prioriza a detecção de fragilidades (sensibilidade) em detrimento da especificidade. No contexto de avaliação educacional, esse comportamento é pertinente: a omissão de uma fragilidade real (falso negativo) pode levar o

coordenador de curso a desconsiderar uma dimensão que demanda intervenção, ao passo que a sinalização de uma fragilidade inexistente (falso positivo) pode ser verificada e descartada mediante análise posterior.

O desempenho no Cenário B ($F1 = 0,8968$) demonstra capacidade discriminativa relevante: o *framework* não gera alertas de criticidade em cenários onde não existem *gaps* críticos. Um sistema de apoio à decisão que consistentemente gera falsos alarmes tenderia a ser desconsiderado pelos usuários, e os resultados do Cenário B atenuam esse risco. A convergência entre os três cenários, que exercitam diferentes perfis avaliativos (predominância de críticos, homogeneidade, contraste extremo) e diferentes instrumentos (Reconhecimento e Autorização), indica que o *framework* é capaz de se adaptar a configurações variadas sem degradação significativa de desempenho.

Reconhecimento de evidências (MCE). Os resultados MCE revelam capacidades consolidadas e limitações específicas. A TRA elevada (0,9047 a 1,0000) indica que o agente consulta efetivamente a base de evidências antes de formular recomendações, atendendo ao requisito fundamental da MCE. A TFNM dentro do limiar (0 a 4,17%) corrobora esse achado, demonstrando que o agente raramente ignora artefatos existentes. Contudo, a TAO abaixo do limiar em todos os cenários (0,5128 a 0,6111) sinaliza que a orientação produzida nem sempre corresponde ao esperado. A análise qualitativa das divergências indica que o padrão predominante é a recomendação de complementação (VP) quando o *ground truth* espera vinculação simples (V). Essa discrepância pode decorrer de dois fatores: (i) o agente, treinado para ser cauteloso em suas recomendações, prefere sugerir complementação como medida preventiva; e (ii) a distinção entre artefato "suficiente" e artefato "que pode ser complementado" envolve julgamento semântico que o *ground truth* trata de forma binária, mas o agente aborda de forma gradual.

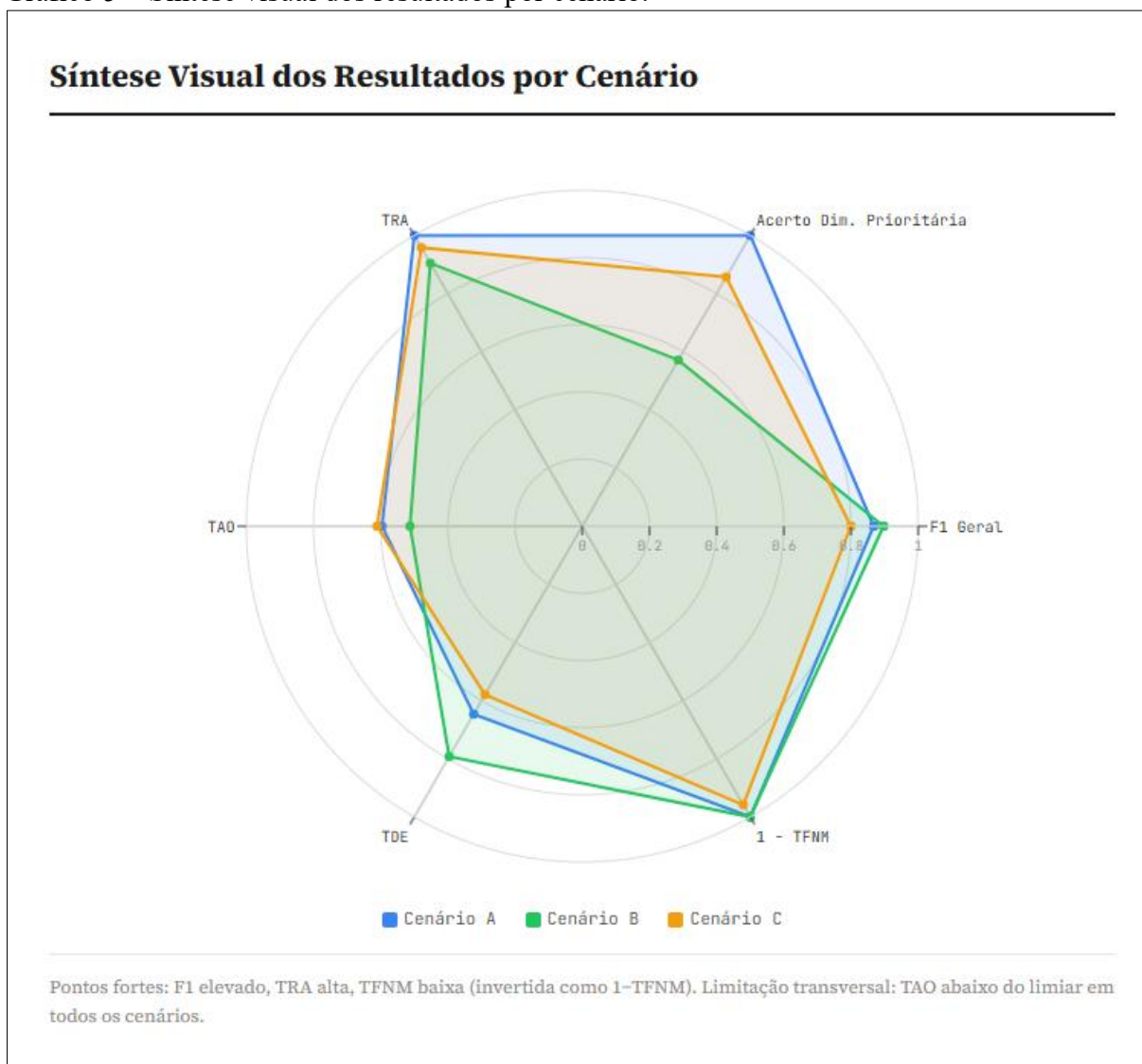
As métricas ICC-N2 e TDR em zero em todos os cenários controlados representam limitação estrutural do componente MCE. A ausência de navegação *cross-curso* (Nível 2) e de descoberta semântica via RAG (Nível 3) indica que o agente, na configuração atual dos *prompts* e ferramentas, restringe sua busca ao plano MPC do curso em avaliação, sem explorar relações com outros cursos da instituição ou a base de conhecimento indexada. Combinadas com a TAO abaixo do limiar em todos os cenários (0,5128 a 0,6111), essas métricas indicam que a MCE, como proposta, não atingiu o nível de maturidade necessário para uso operacional em sua função de orientar o compartilhamento de evidências entre cursos. A MCE permanece como componente conceitual válido da arquitetura do *framework*, cuja operacionalização demanda

refinamento na engenharia de *prompts* e na configuração das ferramentas de busca dos agentes em iterações futuras.

Robustez do *pipeline* de extração. A ocorrência de uma execução com $F1 = 0$ no Cenário C (4,76% das execuções) revela fragilidade no *pipeline* de extração estruturada. A análise da execução anômala indica que o diagnóstico textual gerado pelo agente, embora potencialmente correto em conteúdo, não seguiu o formato esperado pelo *parser* de extração, resultando em falha completa na comparação com o *ground truth*. Essa vulnerabilidade decorre da dependência do pipeline em *parsers* que interpretam a saída textual dos agentes como dados estruturados (JSON). Estratégias de mitigação incluem: (i) implementação de re-tentativa automática quando o *parsing* falha; (ii) validação do formato do output antes do cômputo das métricas; e (iii) utilização de modelos com suporte nativo a output estruturado (*structured output*), que restringem a geração textual a um *schema* JSON predefinido, eliminando a necessidade de *parsing* posterior.

Alinhamento com a DSRM. Os resultados demonstram que o artefato atende ao requisito de conformidade funcional estabelecido na etapa de avaliação da *Design Science Research Methodology* (PEFFERS *et al.*, 2007). O *framework* produz diagnósticos que identificam corretamente a maioria dos *gaps* de progressão conceitual (*Recall* $\geq 0,85$ para categorias críticas nas execuções válidas), diferencia cenários com e sem fragilidades severas, e identifica a dimensão prioritária na maioria das execuções. A estabilidade observada entre execuções, com reprodutibilidade perfeita no Cenário A e variação marginal no Cenário B, reforça a confiabilidade dos resultados obtidos. É importante registrar, contudo, que esses resultados constituem verificação de conformidade, o agente segue as regras de classificação do *framework* com alta fidelidade, e não demonstração de eficácia prática, que requereria validação por especialistas humanos. Conforme Venable, Pries-Heje e Baskerville (2016), artefatos em DSR podem ser avaliados em diferentes estágios de maturidade, e o presente estágio corresponde à verificação de viabilidade técnica e conformidade funcional. Essas evidências sustentam a transição para os procedimentos de validação subsequentes, nos quais o *framework* é exercitado com dados reais de avaliações do CEE/SECITECI-MT.

Gráfico 5 – Síntese visual dos resultados por cenário.



Fonte: Elaborado pelo autor (2026).

6.2 Validação por Análise Retrospectiva com Dados Oficiais

A validação por análise retrospectiva constitui o segundo procedimento de avaliação do *framework*, conforme definido na seção 3.2.2.2. Este procedimento utiliza dados oficiais de avaliações *in loco* para verificar a relevância prática dos diagnósticos e recomendações produzidos pelo *framework*, operando com conceitos reais do Conceito de Curso (CC) tratados como autoavaliação.

Foram executadas oito sessões de validação, totalizando 39 processamentos de cursos com o modelo Claude Sonnet 4.6 (Anthropic). Cinco cursos da Universidade do Estado de Mato Grosso (UNEMAT) foram selecionados segundo os critérios de diversidade definidos

na metodologia. O custo total de processamento foi de USD 35,47 (média de USD 0,91 por curso processado).

6.2.1 Caracterização dos cursos selecionados

A Tabela 29 apresenta os cinco cursos selecionados para a análise retrospectiva, com suas características e a distribuição de indicadores por faixa de conceito, com origem nos relatórios de visita *in loco* das comissões designadas pela SECITECI, homologados pelo CEE/MT.

Tabela 29 – Cursos selecionados para a análise retrospectiva

Curso	Código INEP	CC Oficial	Grau Acadêmico	Total de Execuções
Tecnologia em Alimentos	1662941	3,45	Tecnologia	8
Educação Física	96415	4,44	Licenciatura/ Bacharelado	7
Pedagogia	80922	3,39	Licenciatura	8
Tecnologia em Gestão Turismo	58512	3,33	Tecnologia	8
Agronomia	1621064	3,63	Bacharelado	8

Fonte: Elaborado pelo autor (2026), com base em dados oficiais da UNEMAT.

A seleção contemplou diversidade de graus acadêmicos (uma licenciatura, um bacharelado, um sendo licenciatura e bacharelado e dois cursos de tecnologia) e faixas de CC (de 3,33 a 4,44). Cabe registrar que o portfólio avaliado da UNEMAT não inclui cursos com CC nas faixas extremas da escala (1, 2 e 5), de modo que a amostra reflète a amplitude efetivamente disponível no universo institucional. O curso de Educação Física (CC = 4,44) foi incluído como controle positivo: por possuir CC elevado e nenhum indicador com conceitos 1 ou 2, espera-se que o *framework* não gere alertas de criticidade, verificando a capacidade discriminativa do sistema em situação de baixa fragilidade. O curso de Tecnologia em Gestão Turismo (CC = 3,33) representa o perfil com maior concentração de fragilidades, enquanto os cursos de Tecnologia em Alimentos (CC = 3,45) e Agronomia (CC = 3,63) situam-se em faixa intermediária.

A distribuição representativa de indicadores por curso é detalhada na Tabela 30 a seguir, obtida a partir de uma sessão de referência.

Tabela 30 – Distribuição de indicadores por categoria no *ground truth*

Curso	Críticos (C1-C2)	Médios (C3)	Baixos (C4)	NSA	Total
Tecnologia em Alimentos	6	10	10	24	50
Educação Física	0	7	12	16	35
Pedagogia	4	8	17	19	48
Tecnologia em Gestão Turismo	7	12	11	19	49
Agronomia	6	12	15	15	48

Fonte: Elaborado pelo autor (2026), com base em dados oficiais da UNEMAT e classificação determinística pela MPC/MCA.

Observa-se que o curso de Tecnologia em Alimentos possui a maior proporção de indicadores NSA (48%), em razão de seu perfil MCA (curso de tecnologia com variáveis contextuais específicas). O curso de Educação Física é o único sem indicadores críticos, configurando o controle positivo previsto na metodologia.

6.2.2 Resultados das métricas automáticas de classificação

6.2.2.1 Desempenho geral

A Tabela 31 apresenta o desempenho consolidado do *framework* considerando todas as 39 execuções.

Tabela 31 – Desempenho geral do *framework* na análise retrospectiva

Métrica	Valor
Total de execuções (cursos processados)	39
F1 Geral (média)	0,9582
F1 Geral (DP)	0,0872
F1 Geral (mínimo)	0,6111
F1 Geral (máximo)	1,0000

Métrica	Valor
Limiar de aceitação (F1 \geq 0,70)	Atendido
Acurácia do Diagnóstico (AD) (média)	1,0000
AD (DP)	0
Execuções com indicadores críticos	32
Dimensão prioritária (acertos/total)	22/39
Taxa de acerto dimensão prioritária	56,41%
Custo total (USD)	35,47
Custo por curso (USD)	0,91

Fonte: Elaborado pelo autor (2026).

O *framework* obteve F1 Geral médio de 0,9582 (DP = 0,0872), superando significativamente o limiar de aceitação de 0,70 e apresentando desempenho superior ao observado nos cenários controlados (F1 = 0,8546). A Acurácia do Diagnóstico (AD), que corresponde ao *Recall* da categoria crítico, foi de 1,0000 em todas as 32 execuções que envolveram cursos com indicadores de conceitos 1 ou 2, indicando que o *framework* identificou corretamente todos os indicadores com fragilidades severas presentes nos dados oficiais da UNEMAT. O limiar de AD \geq 0,80 foi, portanto, plenamente atendido.

A taxa de acerto da dimensão prioritária (56,41%) merece análise contextualizada, apresentada na seção 6.2.2.2.

A Tabela 32 e 33 a seguir, construídas a partir de uma execução de referência por curso, exemplificam os padrões reportados nas métricas agregadas da seção.

Tabela 32 – Matriz de confusão: Curso de Tecnologia em Alimentos (execução com F1 Geral = 1,0)

Esperado ↓ / Previsto →	Crítico	Médio	Baixo	NSA	Total esperado
Crítico	6	0	0	0	6
Médio	0	10	0	0	10
Baixo	0	0	10	0	10
NSA	0	0	0	24	24
Total previsto	6	10	10	24	50

Fonte: Elaborado pelo autor (2026).

Todos os 50 indicadores foram classificados na categoria correspondente do *ground*

truth. Apesar disso, a dimensão prioritária foi identificada como D2 quando o esperado era D1, evidenciando que F1 Geral elevado não implica acerto automático da dimensão prioritária, e justificando a reportagem dessas métricas em separado.

Tabela 33 – Matriz de confusão: Educação Física (controle positivo, F1 Geral = 0,6111)

Esperado ↓ / Previsto →	Crítico	Médio	Baixo	NSA	Total esperado
Crítico	–	–	–	–	0
Médio	5	2	0	0	7
Baixo	0	0	12	0	12
NSA	0	0	0	16	16
Total previsto	5	2	12	16	35

Fonte: Elaborado pelo autor (2026).

O curso de Educação Física (CC oficial 4,44) foi selecionado como controle positivo: o *ground truth* determinístico não aponta nenhum indicador crítico. O agente, contudo, classificou cinco indicadores de D3 como críticos (3.1, 3.6, 3.7, 3.8 e 3.10), todos esperados como médios. Nenhum falso negativo ocorreu nas demais categorias. O resultado tensiona a caracterização de controle positivo: o *framework* não passa no teste de "não emitir alerta quando não há fragilidade severa" neste curso específico, diferentemente do que ocorreu no Cenário B (onde o F1 de Crítico foi 1,0).

Exemplo qualitativo 3 – Falsos positivos críticos no controle positivo (Educação Física)

Indicadores envolvidos: 3.1, 3.6, 3.7, 3.8, 3.10 (todos da Dimensão 3 - Infraestrutura).

Ground truth: os cinco indicadores foram classificados como Médio pela lógica determinística da MPC/MCA, a partir de conceito 3 (suficiente) no Conceito de Curso oficial do INEP.

Saída do agente: classificou os cinco como Crítico, elevando indevidamente o patamar de fragilidade.

Padrão observado. A divergência é sistemática dentro da dimensão: dos 35 indicadores do *ground truth*, apenas os cinco indicadores de infraestrutura oscilaram entre categorias ao longo das seis execuções analisadas pela consistência interna. Esse agrupamento sugere que o agente aplica um viés específico sobre indicadores de infraestrutura com conceito 3, interpretando "suficiência" como "fragilidade prioritária".

Impacto prático. Em cenário de produção, esse comportamento geraria alertas desnecessários de criticidade para um coordenador de curso cujo CC é 4,44, a maior nota entre os cinco cursos selecionados. A taxa de falsos positivos críticos (5 em 35 indicadores = 14,3%) é baixa em números absolutos, mas qualitativamente relevante por contradizer a expectativa estabelecida para o controle positivo na metodologia.

6.2.2.2 Desempenho por curso

A Tabela 34 detalha o desempenho do *framework* para cada curso, considerando todas as execuções realizadas ao longo das oito sessões.

Tabela 34 – Métricas de classificação por curso (consolidado de todas as sessões)

Curso	CC	Exec.	F1 (média)	F1 (DP)	F1 Min	F1 Max	AD	AD Expand.	Precision Críticos	Dim. Prioritária (acertos)
Tecnologia em Alimentos	3,45	8	1,0000	0	1,0000	1,0000	1,0000	1,0000	1,0000	0/8 (0%)
Educação Física	4,44	7	0,8889	0,1898	0,6111	1,0000	N/A	*	*	0/7 (0%)
Pedagogia	3,39	8	0,9221	0,0044	0,9112	0,9237	1,0000	1,0000	1,0000	8/8 (100%)
Tecnologia em Gestão Turismo	3,33	8	0,9722	0	0,9722	0,9722	1,0000	1,0000	1,0000	8/8 (100%)
Agronomia	3,63	8	0,9989	0,003	0,9914	1,0000	1,0000	1,0000	1,0000	6/8 (75%)

Fonte: Elaborado pelo autor (2026). N/A = não aplicável (curso sem indicadores C1-C2). * = valores variáveis entre sessões (ver análise detalhada).

Tecnologia em Alimentos (CC = 3,45). O *framework* obteve F1 = 1,0000 em todas as oito execuções, com AD = 1,0000, AD Expandida = 1,0000 e *Precision* de Críticos = 1,0000. Todos os 6 indicadores críticos, 10 médios, 10 baixos e 24 NSA foram corretamente classificados em todas as execuções, sem falsos negativos nem falsos positivos. Este resultado perfeito indica que o *framework* opera com máxima eficácia em cursos de tecnologia com perfil MCA específico, possivelmente beneficiado pela proporção elevada de indicadores NSA (48%) que reduz a ambiguidade classificatória. A dimensão prioritária, contudo, foi incorretamente identificada em todas as execuções: o *ground truth* esperava a Dimensão 1, porém o agente

indicou a Dimensão 2. Este padrão persistente sugere que, embora o *framework* identifique corretamente os indicadores problemáticos, a lógica de agregação por dimensão do agente pode divergir do critério determinístico utilizado pelo *ground truth*.

Educação Física (CC = 4,44, controle positivo). O F1 médio de 0,8889 (DP = 0,1898) apresenta a maior dispersão entre os cursos, com F1 oscilando entre 0,6111 e 1,0000. Em 5 das 7 execuções, o *framework* obteve F1 = 1,0000, classificando corretamente todos os indicadores. Contudo, em 2 das 7 execuções (28,6%), o *framework* classificou 5 indicadores de conceito 3 como críticos (indicadores 3.1, 3.6, 3.7, 3.8 e 3.10, todos da Dimensão 3 – Infraestrutura), gerando exatamente os falsos alarmes de criticidade que o controle positivo foi concebido para detectar. A AD não é aplicável para este curso, dado que não possui indicadores com conceitos 1 ou 2. Esse resultado indica que o *framework* não atende plenamente à função de controle positivo: embora na maioria das execuções (71,4%) o sistema não gere falsos alarmes, a taxa de 28,6% de falha compromete a confiança no *framework* para cursos com desempenho avaliativo elevado. A discrepância com o Cenário B dos cenários controlados, no qual o F1 da categoria Crítico foi 1,0000 em todas as execuções, sem nenhum falso alarme, sugere que a complexidade dos dados reais introduz instabilidade classificatória não capturada pelos cenários sintéticos, particularmente em indicadores de infraestrutura com conceito 3.

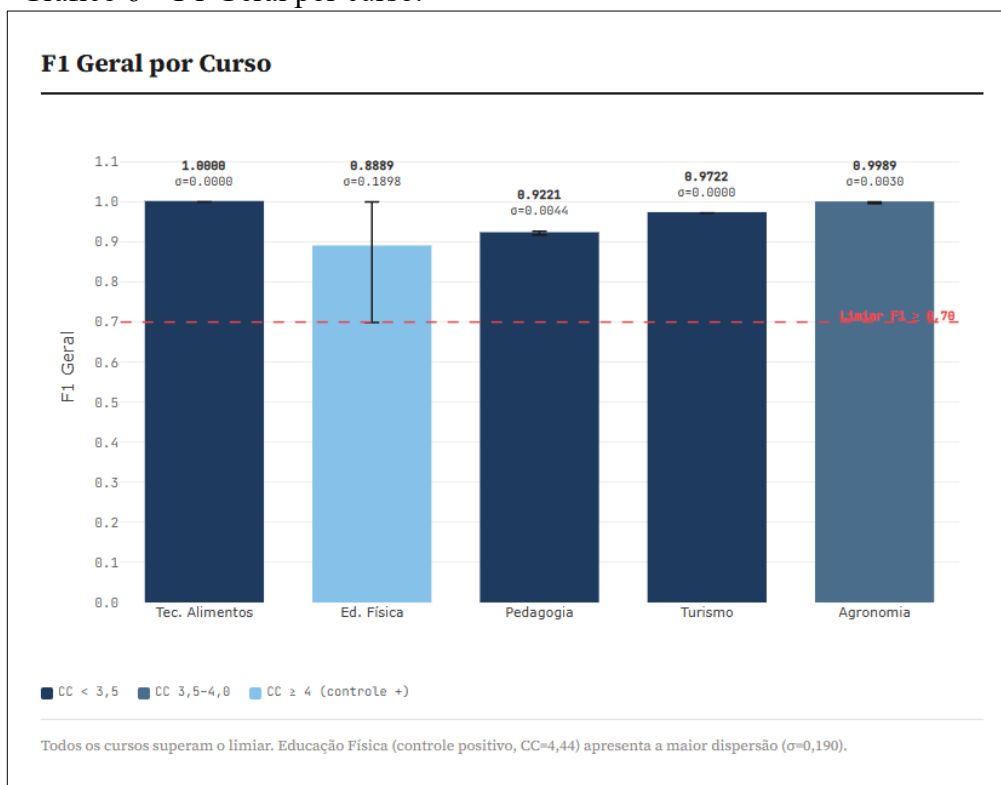
Pedagogia (CC = 3,39). O F1 médio de 0,9221 (DP = 0,0044) indica desempenho elevado e estável. A AD de 1,0000 confirma que os 4 indicadores críticos foram identificados em todas as execuções. A variação marginal do F1 (entre 0,9112 e 0,9237) concentra-se na categoria "médio", onde *Precision* moderada (0,6667 na sessão de referência) indica que alguns indicadores de prioridade baixa ou NSA foram classificados como médios. A dimensão prioritária foi corretamente identificada em 100% das execuções (Dimensão 2).

Tecnologia em Gestão Turismo (CC = 3,33). O *framework* obteve F1 = 0,9722 em todas as oito execuções (DP = 0), com AD = 1,0000, AD Expandida = 1,0000 e *Precision* de Críticos = 1,0000. Os 7 indicadores críticos foram identificados sem falha, e a dimensão prioritária (Dimensão 2) foi acertada em 100% das execuções. A única fonte de desvio do F1 perfeito encontra-se na categoria "médio" (F1 = 0,8889 na sessão de referência, com *Precision* = 0,8000): o *framework* identificou 15 indicadores como médios quando o *ground truth* esperava 12, resultando em 3 falsos positivos.

Agronomia (CC = 3,63). O F1 médio de 0,9989 (DP = 0,003) indica desempenho próximo ao perfeito. Em 6 das 8 execuções, o F1 foi 1,0000; nas 2 execuções restantes, o F1 foi 0,9914, com variação mínima. A AD de 1,0000 confirma a identificação de todos os 6

indicadores críticos. A dimensão prioritária foi acertada em 6 das 8 execuções (75%), com os erros concentrados em sessões em que o agente indicou dimensão diferente da esperada (Dimensão 1 esperada).

Gráfico 6 – F1 Geral por curso.



Fonte: Elaborado pelo autor (2026).

6.2.2.3 Análise da dimensão prioritária

A taxa global de acerto da dimensão prioritária (56,41%) requer análise detalhada, dado que o desempenho varia significativamente entre cursos, conforme pode ser observado na Tabela 35.

Tabela 35 – Acerto de dimensão prioritária por curso

Curso	Dimensão esperada	Acertos	Total	Taxa
Tecnologia em Alimentos	D1	0	8	0%
Educação Física	D1	0	7	0%
Pedagogia	D2	8	8	100%
Tecnologia em Gestão	D2	8	8	100%

Curso	Dimensão esperada	Acertos	Total	Taxa
Turismo				
Agronomia	D1	6	8	75%

Fonte: Elaborado pelo autor (2026).

Observam-se dois padrões distintos. Os cursos de Pedagogia e Tecnologia em Gestão Turismo apresentaram acerto de 100% na identificação da dimensão prioritária (D2), enquanto os cursos de Tecnologia em Alimentos e Educação Física apresentaram acerto de 0%, com o agente indicando dimensão distinta da esperada em todas as execuções. O curso de Agronomia situou-se em posição intermediária, com 75% de acerto.

A análise dos erros de dimensão prioritária nos cursos com taxa de 0% requer distinção entre dois padrões. No curso de Tecnologia em Alimentos, que obteve $F1 = 1,0000$ em todas as execuções (classificação perfeita por indicador), a dimensão prioritária esperada (D1, com maior concentração de *gaps* críticos) deveria ser derivável sem ambiguidade a partir dos próprios indicadores corretamente classificados. O fato de o agente indicar consistentemente D2 sugere que a lógica de agregação por dimensão do agente diverge sistematicamente do critério determinístico do *ground truth*, constituindo limitação funcional que não pode ser atribuída a ambiguidade dos dados.

No caso da Educação Física, que não possui indicadores críticos ($CC = 4,44$), a definição de dimensão prioritária recai sobre a dimensão com maior volume de *gaps* médios, contexto no qual a ambiguidade é genuína: na ausência de fragilidades severas, a hierarquização de dimensões admite múltiplas perspectivas.

O resultado global de 56,41% de acerto, pouco acima do acaso para três dimensões (33%), indica que, na configuração atual, o *framework* é confiável para identificar quais indicadores são problemáticos e em que nível de prioridade, mas não para indicar qual dimensão priorizar. Essa limitação é relevante para a prática de gestão, na qual a identificação da dimensão prioritária orienta a alocação de recursos e esforços de melhoria.

6.2.2.4 Métricas detalhadas por categoria (sessão de referência)

A Tabela 36 apresenta as métricas por categoria de prioridade para cada curso, extraídas de uma sessão de referência representativa do comportamento típico do *framework*.

Tabela 36 – *Precision, Recall* e F1 por categoria de prioridade e curso

Curso	Categoria	<i>Precision</i>	<i>Recall</i>	F1	Esperados	Identificados
Tecnologia em Alimentos	Crítico	1,0000	1,0000	1,0000	6	6
Tecnologia em Alimentos	Médio	1,0000	1,0000	1,0000	10	10
Tecnologia em Alimentos	Baixo	1,0000	1,0000	1,0000	10	10
Tecnologia em Alimentos	NSA	1,0000	1,0000	1,0000	24	24
Educação Física	Crítico	1,0000	1,0000	1,0000	0	0
Educação Física	Médio	1,0000	1,0000	1,0000	7	7
Educação Física	Baixo	1,0000	1,0000	1,0000	12	12
Educação Física	NSA	1,0000	1,0000	1,0000	16	16
Pedagogia	Crítico	1,0000	1,0000	1,0000	4	4
Pedagogia	Médio	0,6667	1,0000	0,8000	8	12
Pedagogia	Baixo	0,8095	1,0000	0,8947	17	21
Pedagogia	NSA	1,0000	1,0000	1,0000	19	19
Tecnologia em Gestão Turismo	Crítico	1,0000	1,0000	1,0000	7	7
Tecnologia em Gestão Turismo	Médio	0,8000	1,0000	0,8889	12	15
Tecnologia em Gestão Turismo	Baixo	1,0000	1,0000	1,0000	11	11
Tecnologia em Gestão Turismo	NSA	1,0000	1,0000	1,0000	19	19
Agronomia	Crítico	1,0000	1,0000	1,0000	6	6
Agronomia	Médio	1,0000	1,0000	1,0000	12	12
Agronomia	Baixo	1,0000	1,0000	1,0000	15	15
Agronomia	NSA	1,0000	1,0000	1,0000	15	15

Fonte: Elaborado pelo autor (2026), dados da sessão de referência (20260412_175432).

Os resultados por categoria confirmam os padrões observados nos cenários

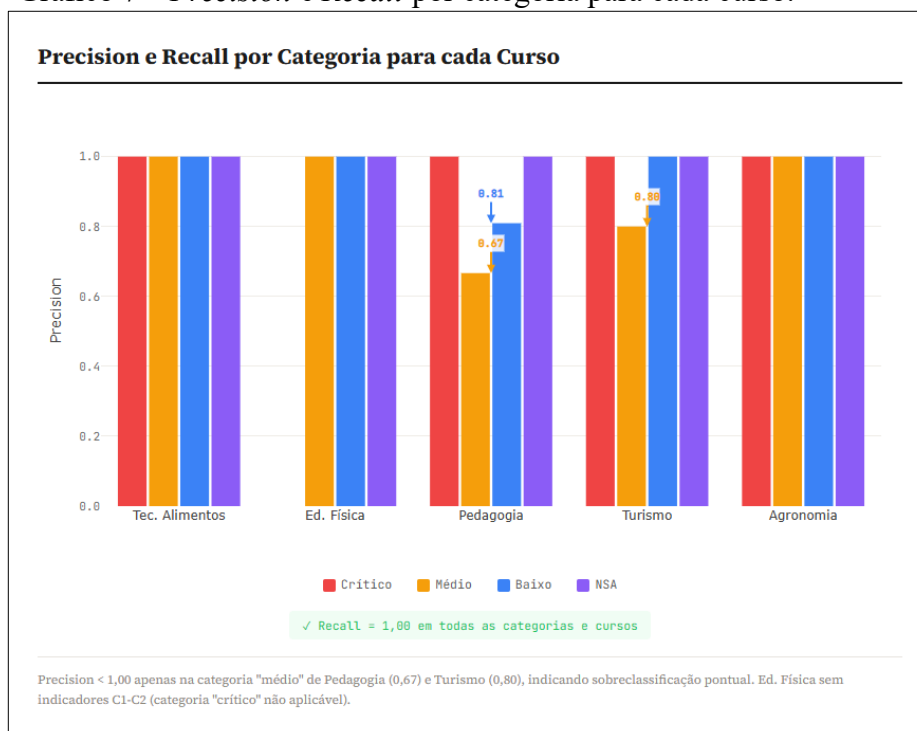
controlados e permitem aprofundar a análise:

Recall perfeito para indicadores críticos. Em todos os cursos com indicadores C1-C2 (Tecnologia em Alimentos, Pedagogia, Tecnologia em Gestão Turismo e Agronomia), o *Recall* da categoria "crítico" foi 1,0000 com *Precision* de 1,0000, indicando que o *framework* identificou todos os indicadores com fragilidades severas sem gerar falsos positivos de criticidade. Este resultado é significativo, pois demonstra que, com dados reais de avaliações oficiais, o *framework* mantém a capacidade de detecção observada nos cenários controlados.

Padrão de sobreclassificação na categoria médio. Nos cursos de Pedagogia e Tecnologia em Gestão Turismo, a categoria "médio" apresentou *Precision* inferior a 1,0 (0,6667 e 0,8000, respectivamente) com *Recall* de 1,0000, repetindo o padrão observado nos cenários controlados: o *framework* identifica todos os indicadores médios, porém inclui alguns indicadores de outras categorias (tipicamente "baixo") como médios, inflando a contagem de indicadores que demandam atenção. Em Pedagogia, 8 indicadores eram esperados como médios, mas 12 foram identificados; em Tecnologia em Gestão Turismo, 12 esperados e 15 identificados.

Tratamento de indicadores NSA. O *Recall* de 1,0000 e *Precision* de 1,0000 para NSA em todos os cursos na sessão de referência indica tratamento correto dos indicadores não aplicáveis, com nenhum indicador NSA incluído indevidamente no diagnóstico e nenhum indicador aplicável classificado erroneamente como NSA.

Gráfico 7 – *Precision* e *Recall* por categoria para cada curso.



Fonte: Elaborado pelo autor (2026).

6.2.3 Resultados das métricas de reconhecimento de evidências (MCE)

A Tabela 37 apresenta os resultados consolidados das métricas MCE obtidos a partir de 37 execuções que incluíram a validação de artefatos de evidência.

Tabela 37 – Métricas MCE consolidadas (análise retrospectiva)

Métrica	Média	DP	Mín.	Máx.	Limiar	Atendimento
TRA (M1)	0,4910	0,3578	0	1,0000	$\geq 0,80$	Não atendido
TAO (M2)	0,2432	0,2511	0	1,0000	$\geq 0,70$	Não atendido
TDE (M3)	0,6622	0,3486	0	1,0000	$\geq 0,60$	Atendido (na média)
ICC-N2 (M4)	0,1757	0,2514	0	1,0000	$\geq 0,70$	Não atendido
TFNM (M5)	0,0270	0,1212	0	0,6667	$\leq 0,20$	Atendido
TDR (M6)	0,1847	0,3574	0	1,0000	Exploratório	-
Qualidade N3 (M7)	0,2432	0,4350	0	1,0000	Exploratório	-

Fonte: Elaborado pelo autor (2026), n = 37 execuções com MCE.

Os resultados MCE no contexto retrospectivo apresentam desempenho inferior ao observado nos cenários controlados, o que é coerente com a maior complexidade dos dados reais e a distribuição graduada dos artefatos entre os cursos. A TRA de 0,4910 indica que o agente reconheceu, em média, metade dos artefatos existentes, com alta variabilidade entre cursos (DP = 0,3578). A TFNM de 0,0270, dentro do limiar, indica que, embora o agente nem sempre reconheça os artefatos, raramente recomenda produção de novo artefato quando um já existe.

A análise por curso revela o efeito do gradiente progressivo de complexidade MCE definido na metodologia, conforme observado na Tabela 38:

Tabela 38 – Métricas MCE por curso (sessão de referência)

Curso	TRA	TAO	TDE	ICC-N2	TFN M	TDR	Qual. N3	Nível MCE predominante
Tecnologia em Alimentos	0,6667	0,3333	1,0000	0	0	0	0	N1

Curso	TRA	TAO	TDE	ICC-N2	TFN M	TDR	Qual. N3	Nível MCE predominante
Educação Física	0,3333	1,0000	0,5000	0,3333	0	0	0	N1 + N2
Pedagogia	0	0,2500	0,2500	0	0	1,0000	1,0000	N3
Tecnologia em Gestão Turismo	0	0	0,5000	0	0	0	0	N1
Agronomia	0,5000	0,3333	1,0000	0,5000	0	1,0000	1,0000	N1 + N2 + N3

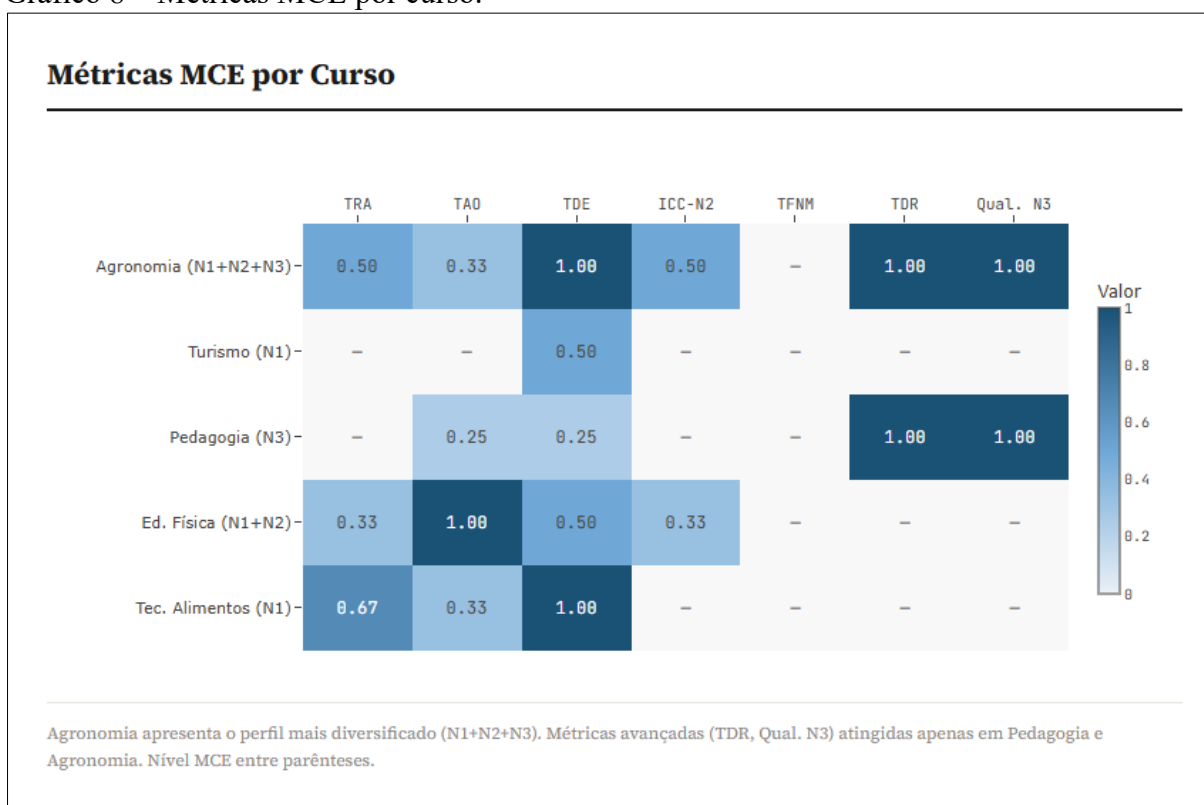
Fonte: Elaborado pelo autor (2026), dados da sessão de referência (20260412_175432).

O curso de Agronomia, configurado para exercitar os três níveis de acesso (N1, N2 e N3), apresentou o desempenho MCE mais diversificado: TDR = 1,0000 (descoberta total via RAG), ICC-N2 = 0,5000 (identificação parcial de artefatos *cross-curso*) e Qualidade N3 = 1,0000 (orientação adequada para artefatos descobertos via RAG). Este resultado contrasta com os cenários controlados, nos quais TDR e ICC-N2 foram uniformemente zero, sugerindo que a configuração de artefatos no contexto retrospectivo proporcionou condições mais favoráveis para o exercício dessas capacidades.

O curso de Pedagogia, cujos artefatos foram configurados para descoberta exclusiva via RAG (Nível 3), obteve TDR = 1,0000 e Qualidade N3 = 1,0000 em algumas sessões, indicando que o agente é capaz de descobrir e orientar adequadamente artefatos indexados na base de conhecimento quando a busca semântica é o único caminho disponível. Contudo, a TRA = 0 nesse curso indica que o agente não reconheceu artefatos via consulta direta, o que é esperado dado que os artefatos estavam disponíveis apenas via RAG.

A variabilidade elevada das métricas MCE (DP entre 0,1212 e 0,4350) indica que o comportamento de reconhecimento de evidências é menos determinístico do que a classificação de prioridade, o que pode decorrer da dependência de buscas semânticas cujos resultados variam conforme a formulação das consultas pelo agente.

Gráfico 8 – Métricas MCE por curso.



Fonte: Elaborado pelo autor (2026).

6.2.4 Resultados da classificação de relevância

Conforme descrito na seção 3.2.2.2, a classificação de relevância das recomendações foi realizada exclusivamente por agente de IA auxiliar. Quatro sessões de classificação foram conduzidas, cada uma com três rodadas independentes, utilizando o modelo claude-sonnet-4-6 como agente classificador. A Tabela 39 sintetiza os resultados.

Tabela 39 – Resultados da classificação automática de relevância

Sessão	Rodadas	Recomendações	Estáveis (%)	Instáveis (%)	DP Médio	<i>Kappa de Fleiss</i> (autoconsistência)	Interpretação
1	3	148	80,4	19,6	0,1131	0,739	Autoconsistência substancial
2	3	147	83,0	17,0	0,0982	0,7724	Autoconsistência substancial
3	3	129	86,8	13,2	0,0761	0,8318	Autoconsistência elevada
4	3	148	80,4	19,6	0,1199	0,7165	Autoconsistência substancial

Sessão	Rodadas	Recomendações	Estáveis (%)	Instáveis (%)	DP Médio	Kappa de Fleiss (autoconsistência)	Interpretação
Média	-	-	82,7	17,4	0,1018	0,7649	Autoconsistência substancial

Fonte: Elaborado pelo autor (2026).

O Kappa de Fleiss médio de 0,7649 (DP = 0,0502) indica autoconsistência elevada do agente classificador entre as três rodadas de cada sessão. É necessário registrar que, neste contexto, o κ opera como índice de autoconsistência de um processo estocástico, e não como métrica de concordância inter-avaliador no sentido clássico da estatística. As três rodadas de classificação são instâncias do mesmo agente com os mesmos parâmetros (modelo, temperatura, *prompt*), e não fontes de julgamento independentes. A escala de Landis e Koch (1977), originalmente concebida para avaliadores independentes, deve ser interpretada com cautela neste contexto: o valor obtido indica que o agente produz classificações estáveis entre execuções, mas essa autoconsistência é condição necessária, não suficiente, para validar a classificação de relevância. Um agente que classificasse consistentemente todas as recomendações como "plenamente relevantes" obteria κ elevado sem ser informativo. A proporção de recomendações estáveis (82,7%, isto é, com a mesma classificação em todas as três rodadas) reforça a estabilidade interna do instrumento classificador.

Tabela 40 – Distribuição da classificação por nível de relevância (moda das rodadas)

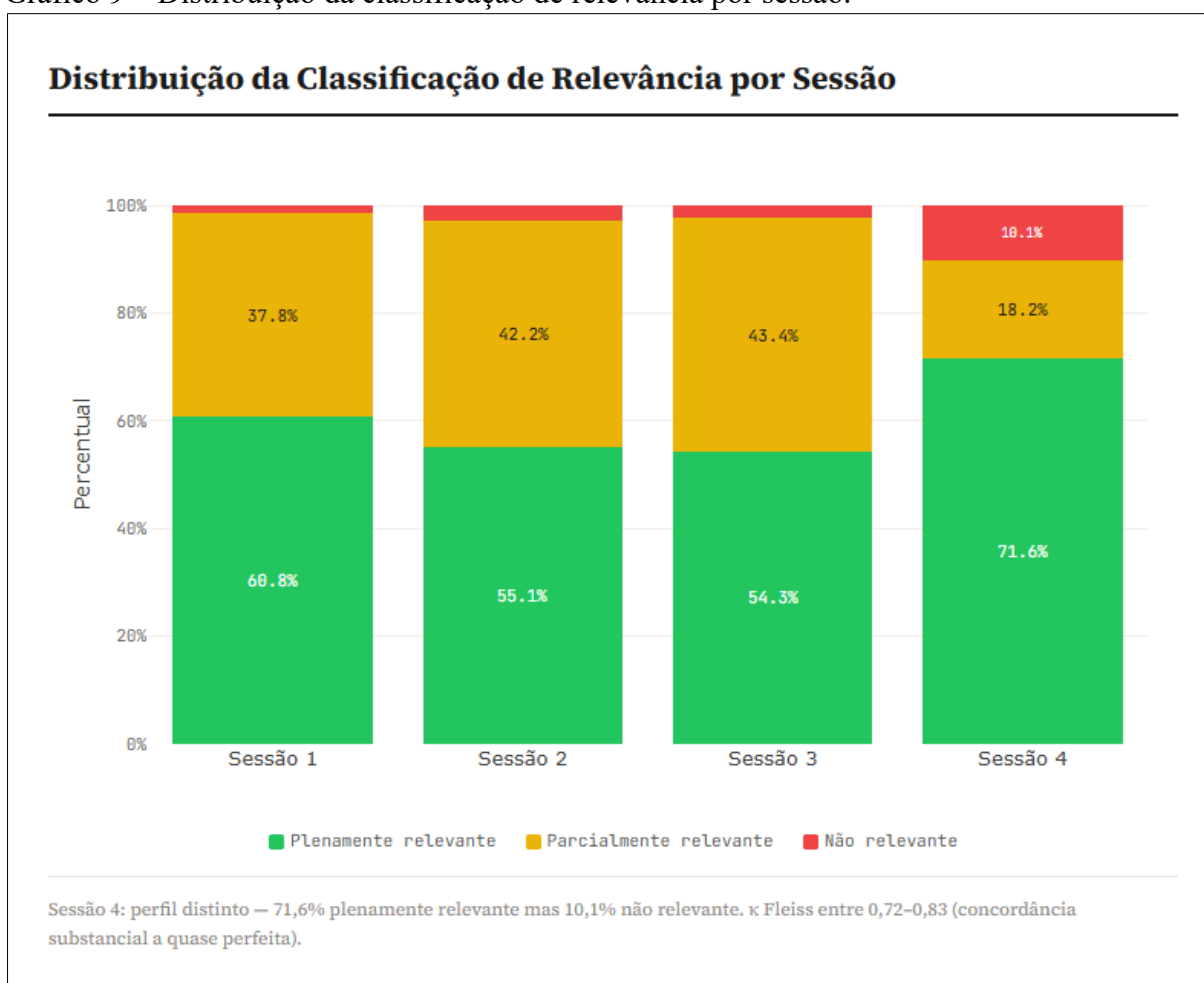
Sessão	Plenamente relevante (3)	Parcialmente relevante (2)	Não relevante (1)
1	60,8%	37,8%	1,4%
2	55,1%	42,2%	2,7%
3	54,3%	43,4%	2,3%
4	71,6%	18,2%	10,1%

Fonte: Elaborado pelo autor (2026). Percentuais calculados sobre a moda de cada recomendação.

Nas três primeiras sessões, a distribuição apresentou padrão consistente: 54 a 61% das recomendações classificadas como plenamente relevantes, 38 a 43% como parcialmente relevantes e 1 a 3% como não relevantes. A sessão 4 apresentou perfil distinto, com 71,6% de plenamente relevantes e 10,1% de não relevantes, o que pode decorrer de variação nos *outputs* do *framework* nessa sessão específica ou de diferenças na formulação textual das recomendações que influenciaram o julgamento do agente classificador.

A consistência de cruzamentos entre sessões, que verifica quantos indicadores mantiveram a classificação (moda) entre todas as quatro sessões, foi de 49,3% (73 de 148 indicadores únicos). Essa taxa inferior à consistência intra-sessão (82,7%) indica que, embora o agente classificador seja estável dentro de cada sessão (entre rodadas), as recomendações geradas em sessões distintas apresentam variações textuais suficientes para alterar a classificação de relevância atribuída.

Gráfico 9 – Distribuição da classificação de relevância por sessão.



Fonte: Elaborado pelo autor (2026).

6.2.5 Estabilidade dos resultados entre execuções

A execução repetida dos cinco cursos ao longo de oito sessões permite observar a estabilidade dos resultados do *framework* com dados reais. A análise a seguir reporta a variabilidade observada como elemento complementar à avaliação de relevância prática. A análise formal de consistência interna é apresentada em seção dedicada.

Tabela 41 – Variabilidade do F1 Geral por curso entre sessões

Curso	F1 (média)	F1 (DP)	Coefficiente de Variação
Tecnologia em Alimentos	1,0000	0	0%
Educação Física	0,8889	0,1898	21,35%
Pedagogia	0,9221	0,0044	0,48%
Tecnologia em Gestão Turismo	0,9722	0	0%
Agronomia	0,9989	0,003	0,30%

Fonte: Elaborado pelo autor (2026).

Os cursos de Tecnologia em Alimentos e Tecnologia em Gestão Turismo apresentaram estabilidade perfeita ($DP = 0$), com resultados idênticos em todas as sessões. Agronomia e Pedagogia exibiram variação marginal ($CV < 0,5\%$), indicando que pequenas oscilações na classificação de indicadores de prioridade média não comprometem a estabilidade geral do diagnóstico. A Educação Física, como controle positivo com CC elevado, apresentou a maior instabilidade ($CV = 21,35\%$), atribuível às duas execuções em que indicadores de conceito 3 foram sobreclassificados como críticos.

A AD manteve-se em 1,0000 ($DP = 0$) em todas as 32 execuções de cursos com indicadores críticos, demonstrando que a detecção de fragilidades severas é o componente mais estável do *framework*.

6.2.6 Discussão dos resultados

Os resultados da análise retrospectiva permitem avaliar a relevância prática do *framework* em três dimensões: eficácia diagnóstica com dados reais, reconhecimento de evidências em contexto institucional e implicações para a gestão de cursos.

Eficácia diagnóstica com dados reais. O F1 Geral médio de 0,9582 na análise retrospectiva supera o observado nos cenários controlados (0,8546), indicando que o *framework* opera com maior eficácia em dados reais do que em cenários sintéticos. Este resultado, embora contraintuitivo à primeira vista, pode ser explicado por dois fatores: (i) os cenários controlados foram deliberadamente concebidos para exercitar casos extremo e situações-limite (como o Cenário B com todos os conceitos iguais a 3 ou o Cenário C com contraste extremo), enquanto

os dados reais tendem a apresentar distribuições de conceitos mais moderadas; e (ii) a proporção de indicadores NSA nos dados reais (média de 37% a 48% dos indicadores dependendo do curso) reduz a ambiguidade classificatória, dado que o tratamento de indicadores NSA apresenta F1 consistentemente elevado.

A Acurácia do Diagnóstico (AD) de 1,0000 em todas as execuções com indicadores críticos constitui o resultado central deste procedimento. Em termos práticos, significa que, quando os conceitos atribuídos pela comissão avaliadora são inseridos no *framework*, todos os indicadores com conceitos 1 ou 2 são corretamente identificados como fragilidades severas. Nenhum falso negativo crítico foi observado em 32 execuções, atendendo plenamente ao limiar de aceitação ($AD \geq 0,80$) e superando a expectativa metodológica.

A taxa de acerto da dimensão prioritária (56,41%) constitui limitação relevante do *framework* na sua configuração atual. Conforme analisado na seção 6.2.2.3, os erros concentram-se em dois padrões distintos: cursos com classificação perfeita por indicador (como Tecnologia em Alimentos, $F1 = 1,0000$) nos quais a divergência na dimensão prioritária indica limitação na lógica de agregação por dimensão do agente; e cursos sem indicadores críticos (como Educação Física, $CC = 4,44$) nos quais a ambiguidade na hierarquização de dimensões é genuína. Na prática, essa taxa indica que o *framework* não deve ser utilizado como referência isolada para a priorização de dimensões, devendo o coordenador de curso considerar o diagnóstico detalhado por indicador, componente no qual o *framework* demonstra desempenho elevado ($F1 = 0,9582$, $AD = 1,0000$), para orientar suas próprias decisões de priorização estratégica.

Reconhecimento de evidências em contexto institucional. Os resultados MCE no contexto retrospectivo apresentam desempenho insatisfatório em relação aos limiares definidos na metodologia. A TRA de 0,4910 (limiar $\geq 0,80$) e a TAO de 0,2432 (limiar $\geq 0,70$) ficaram significativamente abaixo dos limiares. A TDR diferente de zero (0,1847) e o ICC-N2 de 0,1757, embora representem variação em relação ao zero absoluto dos cenários controlados, permanecem distantes dos limiares de aceitação. Esses resultados indicam que a MCE, nas condições experimentais desta pesquisa, não atingiu o nível de desempenho necessário para orientar de forma confiável o compartilhamento de evidências entre cursos. A variação positiva observada na análise retrospectiva em relação aos cenários controlados, particularmente nos cursos de Agronomia ($TDR = 1,0000$ na sessão de referência) e Pedagogia (Qualidade N3 = 1,0000), sugere que determinadas configurações de artefatos propiciam condições mais favoráveis para o exercício dessas capacidades, o que poderá orientar o refinamento dos

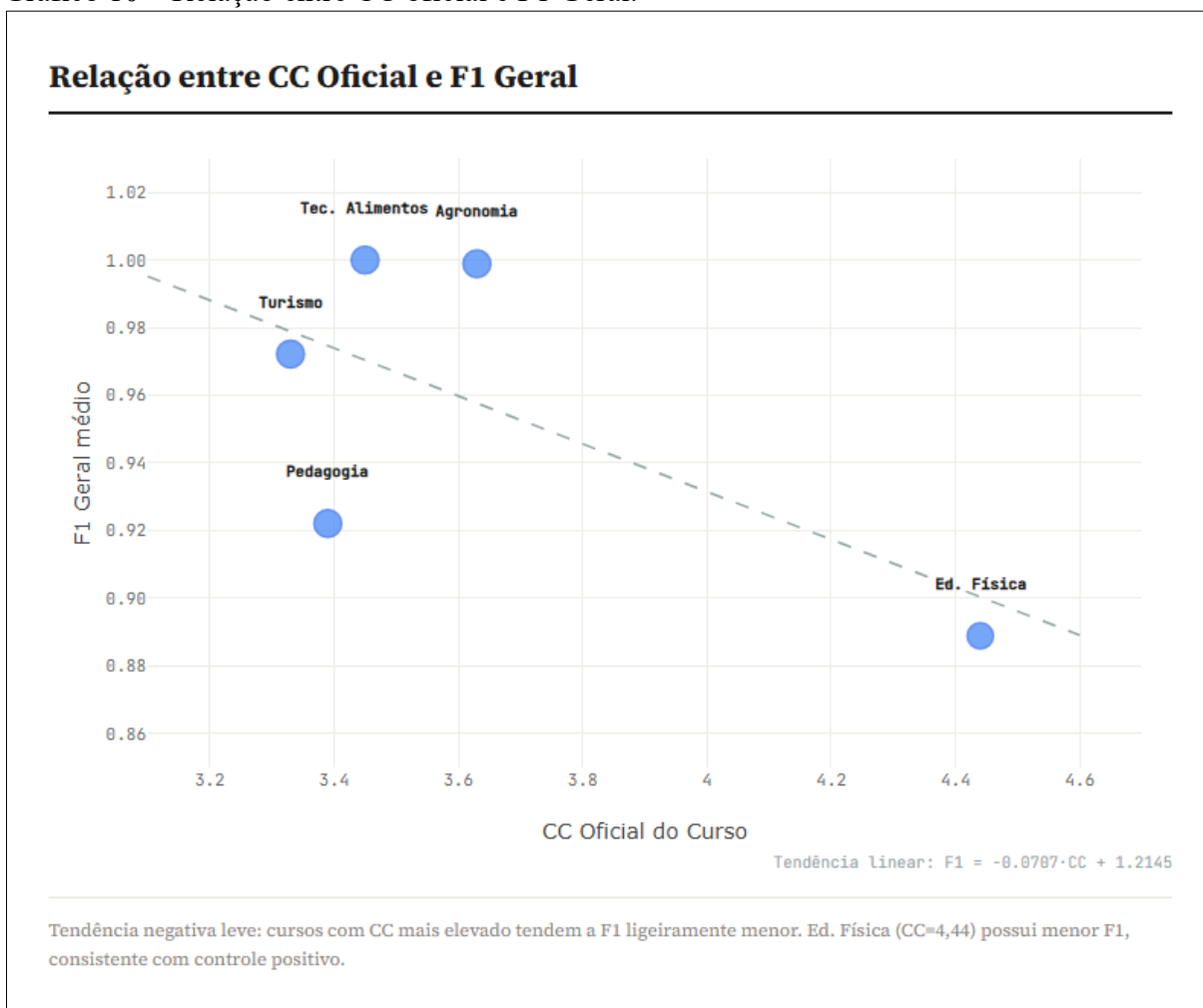
prompts e ferramentas em iterações futuras.

A TRA inferior no contexto retrospectivo (0,4910 contra 0,9543 nos cenários controlados) reflete a maior complexidade dos artefatos reais e a distribuição graduada por níveis de acesso. Cursos cujos artefatos são acessíveis apenas via RAG (como Pedagogia, com TRA = 0) demonstram que a consulta direta à base de evidências, que funciona de forma confiável no Nível 1, não se aplica a artefatos de Nível 3, para os quais a descoberta semântica é o caminho exclusivo.

Classificação de relevância e pertinência prática. A classificação por agente auxiliar indica que, na maioria das sessões, 54 a 61% das recomendações foram avaliadas como plenamente relevantes e 38 a 43% como parcialmente relevantes, com apenas 1 a 3% classificadas como não relevantes. A estabilidade interna do agente classificador, mensurada pelo índice de autoconsistência detalhado na seção 6.2.4, indica que as classificações são reprodutíveis entre rodadas. Esses resultados devem ser interpretados, contudo, como indicadores de coerência interna do sistema, um agente de IA avaliando os *outputs* de agentes da mesma família, e não como evidência de relevância prática das recomendações para coordenadores de curso e gestores de IES. A ausência de validação por especialistas humanos constitui limitação metodológica central desta pesquisa: sem participação de avaliadores humanos, não é possível aferir se as recomendações são efetivamente úteis para a prática de gestão educacional. Essa validação, que permitiria calibrar a concordância entre o julgamento do agente classificador e o de profissionais da área, é tratada como trabalho futuro prioritário na seção 7.2.

Alinhamento com a DSRM e implicações. A análise retrospectiva complementa os cenários controlados ao demonstrar que o *framework* opera de forma eficaz com dados oficiais de avaliações dos cursos da UNEMAT. Os resultados sustentam a viabilidade do artefato como ferramenta de apoio à gestão de cursos de graduação, conforme objetivo definido na *Design Science Research Methodology* (PEFFERS *et al.*, 2007). A combinação de AD = 1,0000 (detecção perfeita de fragilidades severas), F1 Geral elevado (0,9582) e recomendações classificadas como relevantes pelo agente auxiliar na maioria dos casos indica que o *framework* produz *outputs* que podem orientar ações de melhoria em contextos reais de avaliação educacional.

Gráfico 10 – Relação entre CC oficial e F1 Geral.



Fonte: Elaborado pelo autor (2026).

6.3 Validação por Consistência Interna

A validação por consistência interna constitui o terceiro procedimento de avaliação do *framework*, conforme definido na seção 3.2.2.3. Este procedimento verifica a confiabilidade operacional do *framework*, mensurando a estabilidade dos resultados produzidos quando os mesmos dados de entrada são submetidos ao sistema em diferentes momentos. A relevância deste procedimento decorre da natureza estocástica dos modelos de Inteligência Artificial generativa, que introduz variabilidade textual nas respostas mesmo quando os parâmetros de geração são controlados (AMATRIAIN, 2024).

A análise de consistência foi operacionalizada como pós-processamento das sete execuções da análise retrospectiva (seção 6.2), utilizando os mesmos cinco cursos da UNEMAT. Para cada curso, os elementos estruturais dos diagnósticos e recomendações gerados em cada

execução foram comparados automaticamente, sem necessidade de execuções adicionais dos agentes de IA. A configuração dos agentes foi mantida constante em todas as execuções: o agente de Diagnóstico operou com temperatura 0,1, e o agente de Recomendação com temperatura 0,3, conforme descrito na seção 3.2.2.3.

6.3.1 Métricas de consistência

Conforme definido na metodologia, a análise distingue dois níveis de consistência: estrutural e semântica. A consistência estrutural foi avaliada por duas métricas complementares: o Coeficiente de Consistência Estrutural (CCE) e o Índice de Consistência Semântica (ICS).

O CCE mensura a proporção dos elementos estruturais do diagnóstico que permanecem idênticos entre as execuções, considerando cinco componentes: (a) conjunto de indicadores críticos e prioritários; (b) classificação de prioridade atribuída a cada indicador; (c) identificação da dimensão prioritária; (d) categorias de prazo das recomendações; e (e) classificações MCE de reconhecimento de evidências. Cada componente produz uma taxa de estabilidade entre 0 e 1, e o CCE corresponde à média ponderada desses componentes.

O CCE Expandido inclui, além dos componentes (a) a (d) do CCE, a estabilidade das classificações MCE (componente e), que avalia se as orientações sobre evidências (vincular, produzir, vincular com complemento, revisar) permanecem estáveis entre execuções.

O ICS mensura a equivalência de sentido das recomendações textuais geradas em diferentes execuções, por meio de *embeddings* vetoriais e similaridade por cosseno. Este índice captura a consistência do conteúdo prescritivo mesmo quando a redação varia, o que é esperado e aceitável em modelos generativos.

O limiar de aceitação estabelecido na metodologia é $CCE \geq 0,80$.

6.3.2 Resultados do Coeficiente de Consistência Estrutural (CCE)

6.3.2.1 Resultado consolidado

A Tabela 42 apresenta os resultados consolidados das métricas de consistência, obtidos a partir da comparação das sete execuções para os cinco cursos.

Tabela 42 – Métricas de consistência consolidadas

Métrica	Valor	Limiar	Atendimento
CCE Médio	0,9759	$\geq 0,80$	Atendido
CCE Desvio-padrão	0,0476	-	-
CCE Expandido Médio	0,7776	-	-
ICS Médio	0,9509	-	-
ICS Desvio-padrão	0,006	-	-

Fonte: Elaborado pelo autor (2026).

O CCE médio de 0,9759 (DP = 0,0476) supera o limiar de aceitação de 0,80, indicando consistência estrutural elevada. Em termos práticos, 97,59% dos elementos estruturais do diagnóstico permaneceram idênticos ao longo das sete execuções, demonstrando que o *framework* produz resultados estruturalmente reproduzíveis apesar da variabilidade inerente ao modelo generativo.

O ICS médio de 0,9509 (DP = 0,006) indica consistência semântica elevada nas recomendações textuais: embora a redação varie entre execuções, o conteúdo prescritivo (tipos de ação, prazos, evidências requeridas) mantém equivalência semântica de 95,09% na média, com variação marginal entre cursos.

O CCE Expandido de 0,7776, inferior ao CCE (0,9759), reflete a menor estabilidade das classificações MCE (componente e), cuja média foi de 0,2749. Esse resultado indica que, enquanto os elementos diagnósticos (indicadores, prioridades, dimensões) são altamente estáveis, as classificações de evidências apresentam maior variabilidade entre execuções.

6.3.2.2 Resultados por curso

A Tabela 43 detalha as métricas de consistência para cada curso.

Tabela 43 – Métricas de consistência por curso

Curso	CCE	CCE Expandido	ICS	F1 Geral (média)	F1 (DP)
Tecnologia em Alimentos	1,0000	0,8211	0,9512	1,0000	0
Educação Física	0,8913	0,7231	0,9414	0,8704	0,2008

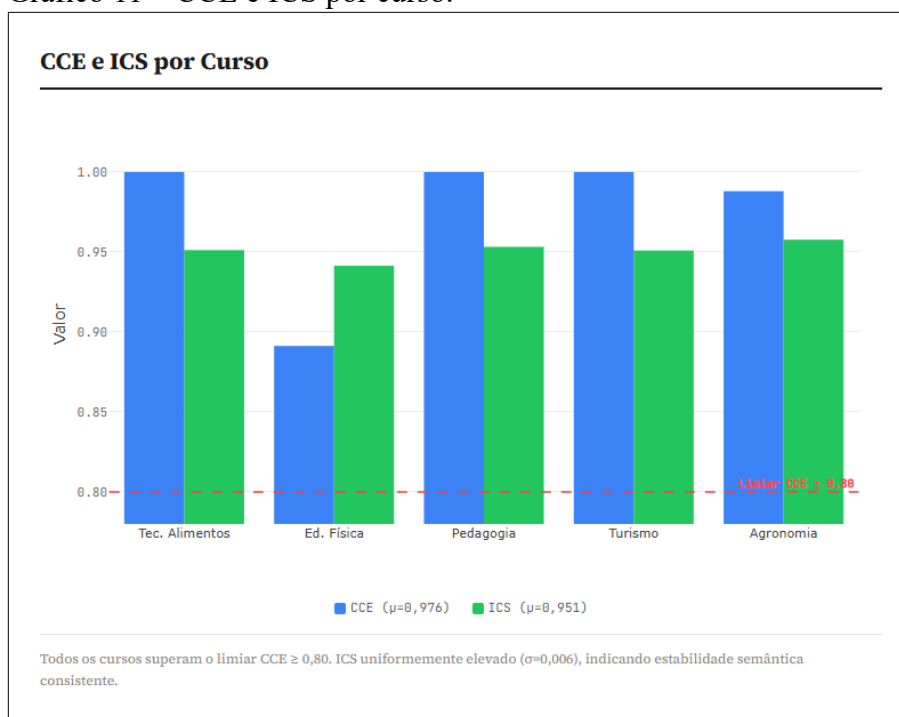
Curso	CCE	CCE Expandido	ICS	F1 Geral (média)	F1 (DP)
Pedagogia	1,0000	–	0,9532	0,9219	0,0047
Tecnologia em Gestão Turismo	1,0000	0,7750	0,9509	0,9722	0
Agronomia	0,9880	0,7913	0,9577	0,9988	0,0033

Fonte: Elaborado pelo autor (2026). "-" indica dados não disponíveis (MCE não configurada para o curso).

Os cursos de Tecnologia em Alimentos, Pedagogia e Tecnologia em Gestão Turismo apresentaram CCE = 1,0000, indicando consistência estrutural perfeita: todos os elementos do diagnóstico permaneceram idênticos em todas as sete execuções. O curso de Agronomia obteve CCE de 0,9880, com variação mínima atribuível à dimensão prioritária (estabilidade de 85,71%, conforme detalhado adiante). O curso de Educação Física apresentou o CCE mais baixo (0,8913), porém ainda acima do limiar de aceitação, com a variação concentrada na classificação de prioridade (estabilidade de 73,68%).

O ICS variou entre 0,9414 (Educação Física) e 0,9577 (Agronomia), indicando que todos os cursos mantiveram consistência semântica elevada nas recomendações, com equivalência prescritiva superior a 94% entre execuções.

Gráfico 11 – CCE e ICS por curso.



Fonte: Elaborado pelo autor (2026).

A Tabela 43 consolidada apresenta o CCE médio, mas oculta em qual nível (indicador, categoria, dimensão) a instabilidade se concentra. A Tabela 44 abaixo, construída para o curso de Educação Física (CCE = 0,8913 – o mais baixo da amostra), torna visíveis os indicadores responsáveis pela instabilidade observada.

Tabela 44 – Estabilidade da classificação de prioridade por indicador (Educação Física, 6 execuções)

Indicador	Execução 1	Execução 2	Execução 3	Execução 4	Execução 5	Execução 6	Estável
1.4	baixo	baixo	baixo	baixo	baixo	baixo	Sim
1.7	baixo	baixo	baixo	baixo	baixo	baixo	Sim
1.8	baixo	baixo	baixo	baixo	baixo	baixo	Sim
1.9	baixo	baixo	baixo	baixo	baixo	baixo	Sim
1.13	baixo	baixo	baixo	baixo	baixo	baixo	Sim
1.18	médio	médio	médio	médio	médio	médio	Sim
1.20	baixo	baixo	baixo	baixo	baixo	baixo	Sim
3.1	crítico	médio	médio	crítico	médio	médio	Não
3.3	baixo	baixo	baixo	baixo	baixo	baixo	Sim
3.5	baixo	baixo	baixo	baixo	baixo	baixo	Sim
3.6	crítico	médio	médio	crítico	médio	médio	Não
3.7	crítico	médio	médio	crítico	médio	médio	Não
3.8	crítico	médio	médio	crítico	médio	médio	Não
3.9	baixo	baixo	baixo	baixo	baixo	baixo	Sim
3.10	crítico	médio	médio	crítico	médio	médio	Não

Fonte: Elaborado pelo autor (2026). Quinze indicadores de prioridade apresentados; cinco instáveis (33,3%), todos da Dimensão 3.

A instabilidade não se distribui aleatoriamente: concentra-se nos mesmos cinco indicadores de infraestrutura (3.1, 3.6, 3.7, 3.8, 3.10) identificados como falsos positivos críticos na análise retrospectiva. Em duas das seis execuções (1 e 4), o agente classificou-os como Crítico; nas outras quatro, como Médio. A taxa de estabilidade de 73,68% (14 em 19 classificações) ainda supera o limiar implícito, mas a identidade dos indicadores instáveis é relevante: são precisamente os indicadores que a análise retrospectiva identifica como problemáticos.

Exemplo qualitativo 4 – Instabilidade concentrada na divergência Crítico ↔ Médio (Educação Física, 6 execuções)

Ground truth para os indicadores 3.1, 3.6, 3.7, 3.8, 3.10: Médio (conceito 3).

Comportamento observado. Em duas execuções (1 e 4), o agente classificou os cinco indicadores como Crítico; nas demais quatro execuções, como Médio. Não houve oscilação intermediária (por exemplo, Baixo ou NSA): a instabilidade é binária entre "subclassificação como esperado" e "sobreclassificação para Crítico".

Triangulação com 6.2. Nesta amostra de seis execuções, duas (33,3%) reproduziriam o resultado reportado na seção 6.2 (F1 Geral = 0,6111 por sobreclassificação de D3); quatro (66,7%) produziram F1 próximo de 1,0. A execução de referência usada na seção 6.2 corresponde a uma das execuções "Crítico", o que explica o desempenho atípico do controle positivo quando comparado aos outros cursos.

Consequência metodológica. A divergência entre a média reportada (CCE = 0,8913) e a identidade dos indicadores instáveis sugere que análises de consistência devem reportar não apenas o coeficiente agregado, mas também o conjunto de indicadores responsáveis pela variabilidade. Do ponto de vista de produção, a inspeção desses indicadores específicos permite ao operador do *framework* identificar classes de casos que exigem revisão humana (no caso, indicadores de infraestrutura com conceito 3 em cursos de saúde e afins).

6.3.3 Análise dos componentes de estabilidade

A Tabela 45 apresenta as taxas de estabilidade para cada componente estrutural, discriminadas por curso.

Tabela 45 – Estabilidade por componente estrutural e curso

Curso	(a) Indicadores Críticos	(b) Classificação Prioridade	(c) Dimensão Prioritária	(d) Categorias Prazo	(e) Classificação MCE
Tecnologia em Alimentos	1,0000	1,0000	1,0000	1,0000	0,3462
Educação Física	1,0000	0,7368	1,0000	1,0000	0,3158
Pedagogia	1,0000	1,0000	1,0000	1,0000	–

Curso	(a) Indicadores Críticos	(b) Classificação Prioridade	(c) Dimensão Prioritária	(d) Categorias Prazo	(e) Classificação MCE
Tecnologia em Gestão Turismo	1,0000	1,0000	1,0000	1,0000	0,1563
Agronomia	1,0000	1,0000	0,8571	1,0000	0,2813
Média	1,0000	0,9474	0,9714	1,0000	0,2749

Fonte: Elaborado pelo autor (2026). "-" indica dados não disponíveis.

Tabela 46 – Médias consolidadas dos componentes de estabilidade

Componente	Estabilidade Média	Interpretação
(a) Indicadores críticos/prioritários	1,0000	Estabilidade perfeita
(b) Classificação de prioridade	0,9474	Estabilidade elevada
(c) Dimensão prioritária	0,9714	Estabilidade elevada
(d) Categorias de prazo	1,0000	Estabilidade perfeita
(e) Classificação MCE	0,2749	Instabilidade significativa

Fonte: Elaborado pelo autor (2026).

A análise dos componentes, conforme observado na Tabela 46, revela uma hierarquia clara de estabilidade:

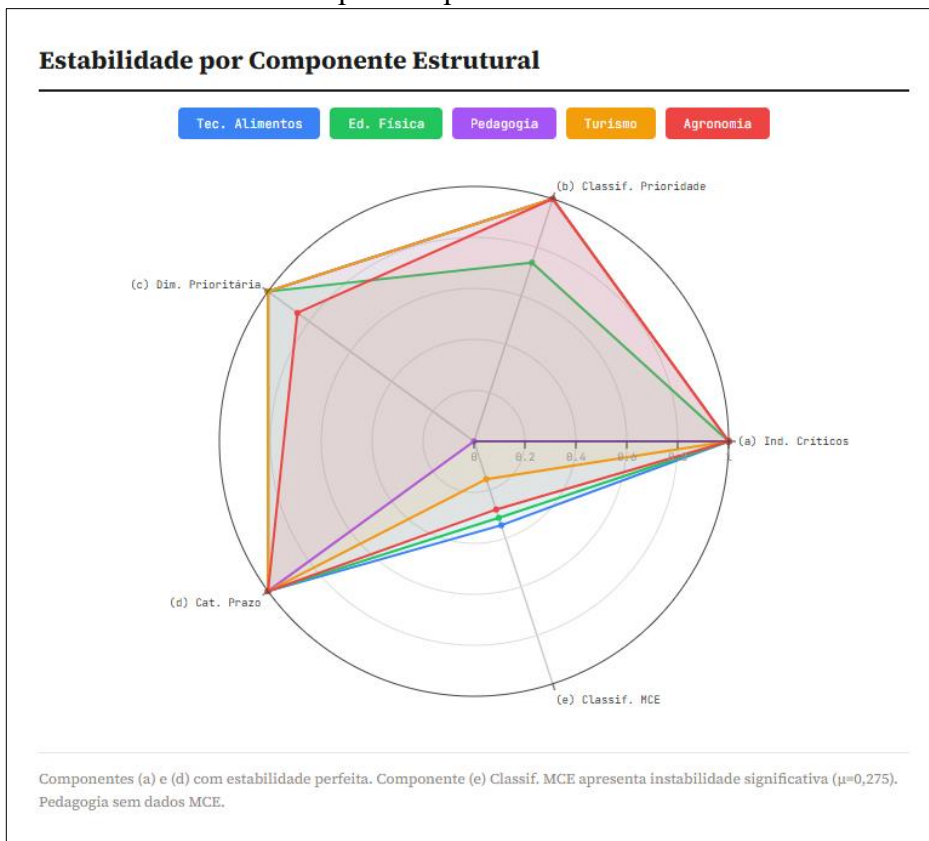
Componentes (a) e (d): estabilidade perfeita (1,0000). O conjunto de indicadores identificados como críticos ou prioritários permaneceu idêntico em todas as execuções de todos os cursos. De forma análoga, as categorias de prazo das recomendações (curto, médio e longo prazo) foram consistentes em 100% dos casos. Esses resultados indicam que os elementos mais relevantes para a gestão, isto é, quais indicadores demandam ação e qual a urgência temporal, são determinísticos na prática, apesar da natureza estocástica do modelo.

Componente (b): estabilidade elevada (0,9474). A classificação de prioridade (crítico, médio, baixo) por indicador foi estável em 94,74% dos casos. A variação concentra-se no curso de Educação Física (estabilidade = 0,7368), no qual 5 indicadores de conceito 3 foram reclassificados de "médio" para "crítico" em 2 das 7 execuções, conforme já identificado na análise retrospectiva (seção 6.2.2.2). Este padrão é coerente com a sensibilidade do agente em contextos de ausência de fragilidades severas: em um curso com $CC = 4,44$, sem indicadores C1 ou C2, o modelo apresenta maior variabilidade na atribuição de prioridades, possivelmente por ausência de sinais discriminativos claros.

Componente (c): estabilidade elevada (0,9714). A dimensão prioritária foi identificada de forma estável em 97,14% dos casos. O único curso com variação foi Agronomia (estabilidade = 0,8571), no qual a dimensão indicada mudou em 1 das 7 execuções. Nos demais cursos, a dimensão prioritária foi a mesma em todas as execuções, incluindo os cursos de Tecnologia em Alimentos e Educação Física, nos quais a dimensão indicada pelo agente divergiu do *ground truth* de forma consistente (cf. seção 6.2.2.3).

Componente (e): instabilidade significativa (0,2749). As classificações MCE de reconhecimento de evidências (vincular, produzir, vincular com complemento, revisar) apresentaram estabilidade média de apenas 27,49%, variando entre 15,63% (Tecnologia em Gestão Turismo) e 34,62% (Tecnologia em Alimentos). Este resultado indica que, embora os elementos diagnósticos sejam altamente reprodutíveis, a orientação sobre evidências é o componente mais sensível à variabilidade do modelo. A análise qualitativa sugere que a busca e a interpretação de artefatos de evidência envolvem julgamento semântico que produz classificações MCE diferentes entre execuções, especialmente para indicadores cujos artefatos requerem avaliação de completude ou adequação.

Gráfico 12 – Estabilidade por componente estrutural.



Fonte: Elaborado pelo autor (2026).

6.3.4 Variabilidade das métricas quantitativas

A Tabela 47 apresenta a variabilidade das métricas quantitativas entre as sete execuções, mensurada pelo desvio-padrão, conforme previsto na seção 3.2.2.3.

Tabela 47 – Variabilidade das métricas quantitativas por curso

Curso	F1 Geral (média)	F1 (DP)	AD (média)	AD (DP)	F1 Crítico (DP)	F1 Médio (DP)	F1 NSA (DP)
Tecnologia em Alimentos	1,0000	0	1,0000	0	0	0	0
Educação Física	0,8704	0,2008	0	0	–	–	–
Pedagogia	0,9219	0,0047	1,0000	0	0	–	–
Tecnologia em Gestão Turismo	0,9722	0	1,0000	0	0	0	0
Agronomia	0,9988	0,0033	1,0000	0	0	0	0

Fonte: Elaborado pelo autor (2026). "–" indica dados não disponíveis ou não computáveis para o perfil do curso.

Os resultados revelam que a Acurácia do Diagnóstico (AD) apresentou variabilidade zero em todos os cursos com indicadores críticos (DP = 0), confirmando que a detecção de fragilidades severas é o componente mais determinístico do *framework*. O F1 Geral apresentou variabilidade zero nos cursos de Tecnologia em Alimentos e Tecnologia em Gestão Turismo, variação marginal em Pedagogia (DP = 0,0047) e Agronomia (DP = 0,0033), e variação mais acentuada em Educação Física (DP = 0,2008). A dispersão elevada no curso de Educação Física reflete as duas execuções anômalas (F1 = 0,6111) já discutidas na seção 6.2, nas quais indicadores de conceito 3 foram sobreclassificados como críticos.

6.3.5 Consistência semântica das recomendações

O Índice de Consistência Semântica (ICS), calculado por meio de *embeddings* vetoriais e similaridade por cosseno entre os textos das recomendações, apresentou valores elevados e uniformes para todos os cursos, conforme pode ser observado na Tabela 48.

Tabela 48 – Índice de Consistência Semântica (ICS) por curso

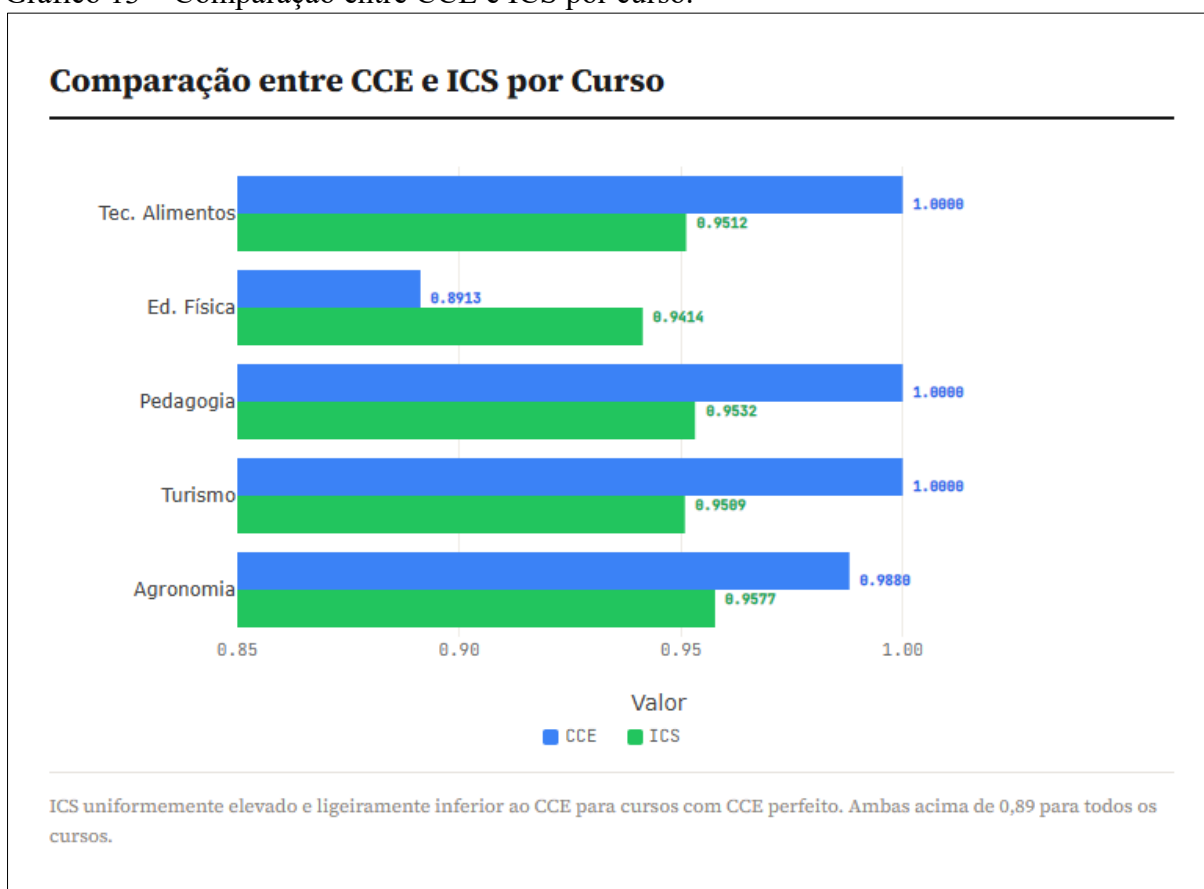
Curso	ICS	Interpretação
Tecnologia em Alimentos	0,9512	Consistência semântica elevada
Educação Física	0,9414	Consistência semântica elevada
Pedagogia	0,9532	Consistência semântica elevada
Tecnologia em Gestão Turismo	0,9509	Consistência semântica elevada
Agronomia	0,9577	Consistência semântica elevada
Média	0,9509	Consistência semântica elevada

Fonte: Elaborado pelo autor (2026).

O ICS médio de 0,9509 indica que, embora a redação das recomendações varie entre execuções (conforme esperado em modelos generativos), o conteúdo prescritivo permanece semanticamente equivalente em mais de 95% dos casos. Este resultado é coerente com a expectativa metodológica de que "variações na redação são esperadas e aceitáveis (...) desde que o conteúdo prescritivo seja equivalente" (seção 3.2.2.3).

A baixa dispersão do ICS entre cursos ($DP = 0,006$) indica que a consistência semântica é independente do perfil do curso, sugerindo que a estabilidade textual decorre das configurações dos agentes (temperatura, *prompts*) e não das características dos dados de entrada.

Gráfico 13 – Comparação entre CCE e ICS por curso.



Fonte: Elaborado pelo autor (2026).

6.3.6 Discussão dos resultados

Os resultados da validação por consistência interna permitem avaliar a confiabilidade operacional do *framework* em três dimensões: reprodutibilidade dos elementos estruturais, robustez semântica e limitações no reconhecimento de evidências.

Reprodutibilidade dos elementos estruturais. O CCE médio de 0,9759 supera o limiar de aceitação (0,80) e situa-se em patamar elevado de consistência. A estabilidade perfeita (1,0000) dos componentes (a) e (d) indica que os elementos mais relevantes para a tomada de decisão em gestão educacional, que indicadores demandam ação e em que prazo, são determinísticos na prática. Este achado é particularmente significativo no contexto de sistemas baseados em IA generativa, nos quais a reprodutibilidade é frequentemente apontada como limitação (AMATRIAIN, 2024). A combinação de temperatura baixa no agente de diagnóstico (0,1) e *prompts* estruturados com regras explícitas de classificação demonstra ser eficaz para

produzir estabilidade nos *outputs* diagnósticos.

A variabilidade observada no componente (b) para o curso de Educação Física (estabilidade = 0,7368) é consistente com os achados da seção 6.2 e reforça a interpretação de que o agente apresenta maior sensibilidade em contextos de baixa fragilidade. A ausência de indicadores C1-C2 neste curso (CC = 4,44) reduz os sinais discriminativos disponíveis ao modelo, resultando em classificações de prioridade mais instáveis para indicadores de conceito 3, que podem oscilar entre "médio" e "crítico" dependendo da variabilidade estocástica da geração textual. Em termos práticos, esta instabilidade afeta a classificação de gravidade, mas não a identificação dos indicadores que demandam atenção: o componente (a), que verifica se os mesmos indicadores são sinalizados como prioritários, permaneceu em 1,0000 para este curso.

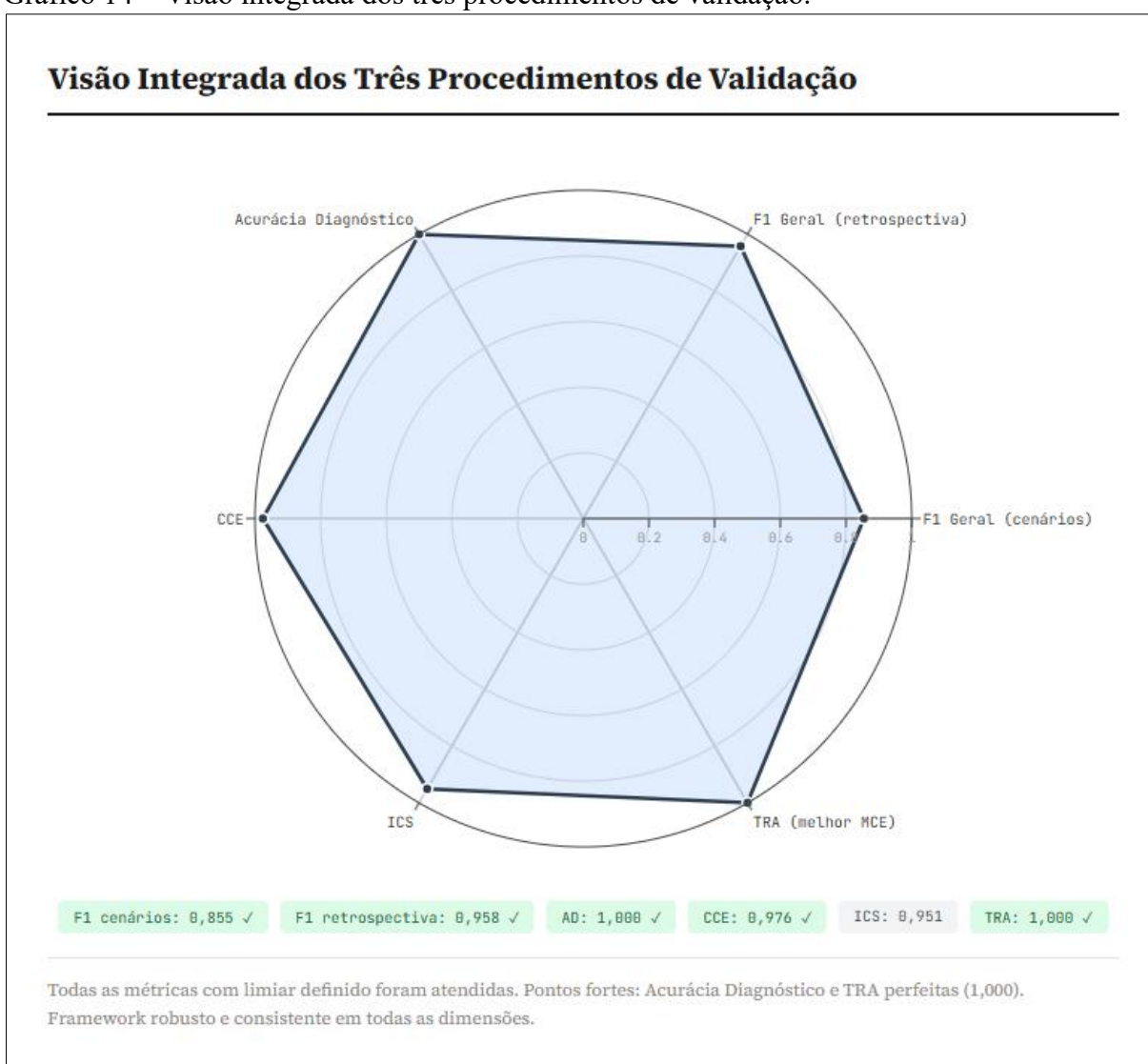
Robustez semântica das recomendações. O ICS médio de 0,9509 (DP = 0,006) demonstra que as recomendações textuais, embora variem em redação entre execuções, mantêm equivalência prescritiva elevada. Este resultado valida a estratégia de diferenciação de temperaturas adotada no *design* dos agentes: a temperatura ligeiramente mais elevada do agente de Recomendação (0,3) permite diversidade na formulação textual sem comprometer a consistência do conteúdo orientador. A uniformidade do ICS entre cursos (variação de apenas 0,6 pontos percentuais entre o menor e o maior valor) indica que a consistência semântica é uma propriedade do sistema, e não uma característica dependente do perfil avaliativo do curso.

Limitações no reconhecimento de evidências. A estabilidade média do componente MCE (0,2749) constitui a limitação mais significativa identificada neste procedimento. A instabilidade das classificações de evidências entre execuções indica que o comportamento de busca e interpretação de artefatos pelos agentes é sensível à variabilidade estocástica do modelo, resultando em orientações MCE inconsistentes para o mesmo indicador em execuções distintas. Este achado é coerente com os resultados MCE da seção 6.2.3, nos quais a alta variabilidade das métricas MCE (DP entre 0,1212 e 0,4350) já sinalizava comportamento menos determinístico nessa dimensão. A implicação para o uso operacional do *framework* é que as orientações sobre evidências devem ser interpretadas como sugestões indicativas, sujeitas a verificação pelo coordenador, enquanto os diagnósticos de prioridade podem ser tratados com elevada confiança na sua reprodutibilidade.

Alinhamento com a DSRM. A validação por consistência interna complementa os procedimentos anteriores ao demonstrar que o artefato não apenas produz resultados corretos (seção 6.1) e relevantes para dados reais (seção 6.2), mas também os produz de forma

reprodutível. Conforme enfatizado por Hevner *et al.* (2004), a avaliação rigorosa de um artefato de *design science* requer demonstração de sua utilidade, qualidade e eficácia. A confiabilidade operacional, mensurada pelo CCE e pelo ICS, contribui para a dimensão de qualidade ao assegurar que usuários do *framework* obterão diagnósticos substancialmente equivalentes em execuções independentes. O CCE de 0,9759 e o ICS de 0,9509, combinados com o F1 Geral de 0,9582 (seção 6.2) e a AD de 1,0000, configuram um perfil de avaliação que sustenta a adequação do artefato para uso prático em contextos de gestão de cursos de graduação.

Gráfico 14 – Visão integrada dos três procedimentos de validação.



Fonte: Elaborado pelo autor (2026).

6.4 Prova de Conceito (Demonstração Integral)

A prova de conceito materializa a atividade de Demonstração prevista no modelo DSRM (PEFFERS *et al.*, 2007), comprovando que o artefato funciona conforme projetado em ambiente operacional real. Diferentemente dos procedimentos de validação documentados nas seções 6.1 a 6.3, que mensuram propriedades específicas do *framework* por meio de métricas quantitativas, a prova de conceito documenta a execução integral do sistema desde a entrada de dados até a geração do relatório final com plano de ação, evidenciando a viabilidade prática da solução proposta.

Conforme estabelecido na seção 3.2.2.4 da Metodologia, o propósito desta atividade não é empregar métricas de avaliação, mas demonstrar que o *framework* é tecnicamente viável e operacionalmente funcional. Essa distinção entre demonstração e avaliação, enfatizada por Peffers *et al.* (2007) e reiterada por Venable, Pries-Heje e Baskerville (2016), é fundamental para a correta interpretação dos resultados aqui apresentados: enquanto as seções 6.1 a 6.3 fundamentaram a qualidade analítica do artefato, a presente seção complementa a avaliação global ao evidenciar que todas as etapas do fluxo operacional podem ser executadas de forma coerente e integrada em cenário real de uso.

A presente seção sintetiza os aspectos mais relevantes da execução, destacando os elementos que comprovam a operacionalidade *end-to-end* do sistema.

6.4.1 Curso Selecionado e Dados de Entrada

Para a realização da prova de conceito, selecionou-se o Curso de Bacharelado em Medicina do Campus Universitário de Cáceres, pertencente ao conjunto de cursos da UNEMAT utilizados ao longo da validação. A escolha do curso fundamentou-se em três critérios: (a) disponibilidade de documentos institucionais na base de conhecimento RAG, incluindo Projeto Pedagógico de Curso (PPC) e Diretrizes Curriculares Nacionais (DCN); (b) perfil avaliativo com variações conceituais entre dimensões, permitindo a demonstração de diagnósticos diferenciados por indicador; e (c) presença de indicadores condicionais filtrados pela MCA, o que evidencia a operação do motor de regras de aplicabilidade durante a execução.

6.4.2 Processo de Análise

A execução da prova de conceito seguiu o fluxo operacional completo do *framework*, estruturado em nove etapas sequenciais que envolvem três classes de atores: a Diretoria de Regulação da Educação Superior (DGRES/UNEMAT), responsável pelo cadastro institucional dos cursos e pela validação dos artefatos produzidos; o gestor do curso, responsável pela configuração do perfil de aplicabilidade, pelo lançamento dos conceitos e pela elaboração do plano de ação; e a cadeia multiagente de IA, constituída pelos Agentes de Diagnóstico, Recomendação e Síntese. Cada etapa foi registrada por meio de capturas de tela da interface da aplicação. A presente subseção descreve o percurso operacional e os elementos observados em cada fase.

6.4.2.1 Cadastro institucional do curso e inserção das DCN

A primeira etapa é de responsabilidade da Diretoria de Regulação e consiste no cadastro institucional do curso na plataforma Notacinco, de modo a refletir os dados registrados na instituição e no sistema e-MEC/INEP. Nessa etapa, registram-se atributos estruturantes do curso como denominação, grau acadêmico, modalidade, carga horária, unidade de oferta e situação regulatória vigente, que constituem a identidade formal do objeto avaliativo. Em seguida, procede-se ao *upload* dos arquivos das Diretrizes Curriculares Nacionais (DCN) pertinentes ao curso, documentos normativos que subsidiarão tanto a extração de metadados quanto a fundamentação das recomendações produzidas pela cadeia multiagente. A centralização dessa operação na DGRES assegura padronização cadastral, evita divergências entre o registro da plataforma e o registro oficial da IES e delimita a fronteira entre as atribuições institucionais e as atribuições de gestão acadêmica do curso.

Figura 14 – Captura de tela do cadastro/edição de curso e *upload* da DCN.

The figure consists of three side-by-side screenshots of the UNEMAT web application interface.

- Left Screenshot: 'Editar: Medicina' (Edit Course)**
 - Identificação (Identification):**
 - Campus: CAC — Campus Universitário de Cáceres
 - Localidade: Avenida São João, S/N — Cáceres/MT
 - Código INEP: 1188849
 - Nome do Curso: Medicina
 - Turno: Integral, Matutino, Vespertino, Noturno
 - Classificação Acadêmica (Academic Classification):**
 - Nível: Graduação
 - Grau: Bacharelado
 - Modalidade: (empty)
- Middle Screenshot: 'Documentos Institucionais' (Institutional Documents)**
 - Buttons: + Novo Documento, Filtrar, Exportar
 - Search bar: Pesquisar...
 - Table of DCN documents:

Curso (DCN)	Re
(DCN) Agromonia / Engenharia Agronômica	RE
(DCN) Educação Física	RE
(DCN) Tecnológicos - Cursos Superiores	RE
(DCN) Pedagogia	RE
(DCN) Turismo	RE
 - Footer: notacinco.application.1.inst.88a v17.755.88038 2026 © Notacinco
- Right Screenshot: 'Novo Documento Institucional' (New Institutional Document)**
 - Tipo de Documento: Diretrizes Curriculares Nacionais (DCN)
 - Curso da DCN: Medicina
 - Título: DCN Medicina
 - Referência Legal: RESOLUÇÃO CNE/CES Nº 3, DE 30 DE SETEMBRO DI
 - Descrição: Institut Diretrizes Curriculares Nacionais do Curso de Graduação em Medicina.
 - Data de Publicação: 30/09/2025
 - Vigência Início: dd/mm/aaaa
 - Vigência Fim: dd/mm/aaaa
 - Arquivo (PDF ou DOCX): Escolher arquivo

Fonte: Elaborado pelo autor (2026).

6.4.2.2 Verificação dos dados cadastrais do curso

A segunda etapa é de responsabilidade do gestor do curso e consiste na verificação dos dados previamente cadastrados pela Diretoria de Regulação. Ao acessar a plataforma, o gestor confere os atributos registrados, válida a correspondência com os dados oficiais do curso e sinaliza eventuais inconsistências à DGRES. Essa conferência antecede qualquer ação analítica, pois os dados cadastrais alimentam o motor de regras da MCA e o contexto dos agentes de IA, de modo que inconsistências cadastrais propagam-se, se não corrigidas, para todas as etapas subsequentes. Constitui, portanto, ponto de controle que preserva a integridade do fluxo analítico desde sua origem.

Figura 15 – Captura de tela da visualização de curso.

Fonte: Elaborado pelo autor (2026).

6.4.2.3 Upload do PPC e constituição da base de conhecimento RAG

A terceira etapa compreende o *upload* do Projeto Pedagógico de Curso (PPC) pelo gestor e a subsequente constituição da base de conhecimento RAG do curso. Uma vez submetido o documento, é acionado o Agente de Diagnóstico para executar a extração de metadados do PPC e das DCN previamente carregadas pela DGRES, produzindo representação estruturada dos elementos normativos e pedagógicos que caracterizam o curso. Paralelamente, os documentos são submetidos ao *pipeline* de processamento descrito na seção 5.2.8 – extração textual, segmentação em *chunks*, geração de *embeddings* vetoriais e armazenamento no PostgreSQL com extensão *pgvector*, compondo a base semântica que será consultada pelos agentes nas etapas seguintes. A interface da plataforma exhibe ao gestor os documentos indexados, o número de fragmentos gerados e o status do processamento, assegurando que a base de conhecimento esteja completa antes do avanço para as etapas analíticas.

Figura 16 – Captura de tela do *upload* do PPC do curso.

The figure consists of four screenshots from a web application interface:

- Leftmost screenshot:** A dialog box titled "Enviar PPC" (Send PPC) for the course "Medicina". It shows a file "5893_res_c...16_2024.pdf" selected for upload. A message at the bottom states: "A DCN é gerenciada no módulo de Documentos Institucionais, pois uma mesma DCN vale para todos os cursos de uma mesma área de conhecimento." (The DCN is managed in the Institutional Documents module, as the same DCN applies to all courses in the same knowledge area.)
- Second screenshot:** The "Metadados do PPC" (PPC Metadata) page for "Medicina". It indicates "Extração em andamento" (Extraction in progress) and notes: "A IA está analisando o PPC. A página será atualizada automaticamente quando concluir." (The AI is analyzing the PPC. The page will be updated automatically when complete.)
- Third screenshot:** A table titled "Disciplinas Extraídas" (Extracted Disciplines) showing 73 disciplines. The table has columns for "Disciplina" (Discipline), "CH" (Credits), and "Ementa" (Content). Examples include:

Disciplina	CH	Ementa
UCI-Introdução ao Estudo da Medicina	60h	Capacita para o resumo ativo e introduz política epidemiológica.
UCII- Concepção e Formação do Ser Humano	60h	Estuda morfológica e fisiologia reprodutiva, sexual, familiar e pré-natal.
UCIII- Metabolismo	60h	Aborda processos metabólicos, endócrinos, metabólicos e metabólicos.
CM I (Ciências Morfofuncionais I)	120h	Anatomia e histomorfologia locomotor, reprodutor e fisiologia da membrana.
IESC I	60h	SUS, SUS, território, PNEIS, DATASUS e visitas.
Habilidades Médicas I	90h	Treina anamnese, sinais, introdução ao exame físico.
Habilidades de Comunicação	60h	Fundamentos de comunicação clínica e comunicação de.
Core Curriculum I (Educação das Relações Étnico-Raciais)	30h	Reflexões étnico-raciais, estereótipos sociais e.
UCIV-Funções Biológicas I	60h	Homeostase, fisiologia e sistema endócrino e mic.
UCV-Funções Biológicas II	60h	Fisiologia cardiovascular miócito.
UCVI-Mecanismo de Agressão e Defesa	60h	Resposta do organismo a mecanismos de lesão.
CM II	120h	Anatomia/histologia de sistema respiratório, renal/urinário.
- Rightmost screenshot:** The "Conformidade PPC-DCN - Medicina" (PPC-DCN Compliance - Medicine) dashboard. It shows an overall compliance rate of 56.8%. A breakdown shows:
 - Conformes (Conformant): 16 (21.4%)
 - Parcialmente (Partially Conformant): 18 (24.3%)
 - Não Conformes (Non-Conformant): 10 (13.5%)
 The dashboard also includes a section for "Análise por Dispositivo" (Analysis by Device) and a "Formação" (Formation) status of "Parcial" (Partial).

Fonte: Elaborado pelo autor (2026).

6.4.2.4 Configuração do perfil de aplicabilidade com a Matriz MCA

A quarta etapa consiste na configuração do perfil da Matriz de Condições de Aplicabilidade (MCA), por meio da qual o gestor define quais variáveis caracterizam o curso e, por consequência, quais indicadores do instrumento de avaliação são aplicáveis. Ao registrar as características do curso, grau acadêmico, modalidade, existência de estágio curricular obrigatório, oferta de Trabalho de Conclusão de Curso (TCC), presença de laboratórios especializados, entre outras variáveis, o motor de regras da MCA executa automaticamente a filtragem contextual, removendo do escopo de análise os indicadores classificados como não aplicáveis (NSA). Essa filtragem constitui funcionalidade crítica do *framework*, pois assegura que os agentes de IA operem apenas sobre o subconjunto de indicadores efetivamente pertinentes ao perfil do curso, evitando diagnósticos genéricos ou recomendações para indicadores irrelevantes.

Como funcionalidade assistiva, o gestor pode solicitar ao Agente de Diagnóstico a sugestão automatizada do perfil MCA, com base nos metadados extraídos do PPC e das DCN na etapa anterior. O agente analisa os documentos via base RAG e recomenda, para cada variável MCA, o valor mais aderente ao curso, acompanhado de grau de confiança, justificativa textual e referência ao trecho documental que fundamenta a sugestão, conforme descrito na seção 5.2.7. A decisão final de habilitação ou desabilitação das variáveis permanece sob

controle do gestor, em conformidade com o princípio de supervisão humana que orienta a arquitetura do *framework*. Confirmada a configuração, o perfil é salvo e passa a reger as etapas subsequentes.

Figura 17 – Captura de tela da configuração da Matriz de Condições de Aplicabilidade (MCA).

Fonte: Elaborado pelo autor (2026).

6.4.2.5 Lançamento dos conceitos no Instrumento de Avaliação

A quinta etapa consiste no registro dos conceitos atribuídos ao curso pela comissão avaliadora *in loco*. O gestor inicia o cadastro da avaliação informando ao sistema a nota global recebida pelo curso, a fim de estabelecer o valor de referência que será confrontado com o cálculo interno do *framework*. Em seguida, procede ao lançamento conceito a conceito, registrando, para cada indicador aplicável do Instrumento de Avaliação de Cursos de Graduação, a nota atribuída pela comissão.

Ao final do lançamento, a plataforma exibe a nota calculada a partir dos conceitos informados e a confronta com a nota global registrada no cadastro da avaliação. A conferência tem por finalidade mitigar erros humanos de digitação no lançamento indicador a indicador, garantindo que os valores inseridos sejam consistentes com a nota efetivamente atribuída pela comissão, sem constituir juízo sobre o mérito da avaliação original. Divergências identificadas nessa etapa demandam revisão dos lançamentos antes do avanço para o cadastro da Matriz de Progressão Conceitual (MPC).

Figura 18 – Captura de tela do cadastro de avaliação do curso.

A imagem apresenta uma interface de usuário dividida em quatro painéis para o cadastro de avaliação de curso. O primeiro painel, 'Nova Avaliação', permite selecionar o tipo de instrumento (Autorização ou Reconhecimento), definir a data da visita (12/07/2018) e o número do processo MEC (202401234567). Também é possível definir pesos para as dimensões (40, 35, 25) e o conceito de curso (3). O segundo painel, 'Instrumento de Avaliação de Cursos de Graduação (Presencial e a Distância)', mostra a configuração de cinco dimensões de avaliação, cada uma com uma escala de 1 a 5 e um botão 'NSA'. O terceiro painel, 'Dimensão 2 — Corpo Docente e Tutorial', detalha seis indicadores de avaliação, como 'Núcleo Docente Estruturante - NDE' e 'Equipe multidisciplinar', com uma escala de 1 a 5 e um botão 'NSA'. O quarto painel, 'CONCEITO DE CURSO', apresenta o resultado final: um conceito ajustado de 3,00, com pesos de 40% para D1, 35% para D2 e 25% para D3. O cálculo das notas resulta em 3,37, com um diagnóstico MPC de 3,76 (1,50) para a Organização Didático-Pedagógica e 3,22 (1,13) para o Corpo Docente e Tutorial.

Fonte: Elaborado pelo autor (2026).

6.4.2.6 Elaboração dos planejamentos com a Matriz MPC

A sexta etapa consiste na elaboração do planejamento de melhoria por meio da MPC. O gestor seleciona, inicialmente, o tipo de instrumento que orientará o cadastro, autorização ou reconhecimento de curso, e, em seguida, prioriza os indicadores sobre os quais deseja atuar. Não há exigência de seleção integral: o gestor pode elaborar múltiplos planejamentos simultâneos, cada qual contemplando um subconjunto distinto de indicadores, o que viabiliza intervenções focadas por dimensão, por grau de urgência ou por natureza da fragilidade identificada. Para cada indicador selecionado, o gestor define o conceito desejado, e a plataforma calcula automaticamente o diagnóstico MPC correspondente, identificando os *gaps* entre a situação verificada e a situação almejada. A persistência da MPC materializa-se como plano de ação, entrada primária da cadeia de agentes de IA na etapa seguinte.

Figura 19 – Captura de tela de configuração da Matriz de Progressão Conceitual (MPC).

Fonte: Elaborado pelo autor (2026).

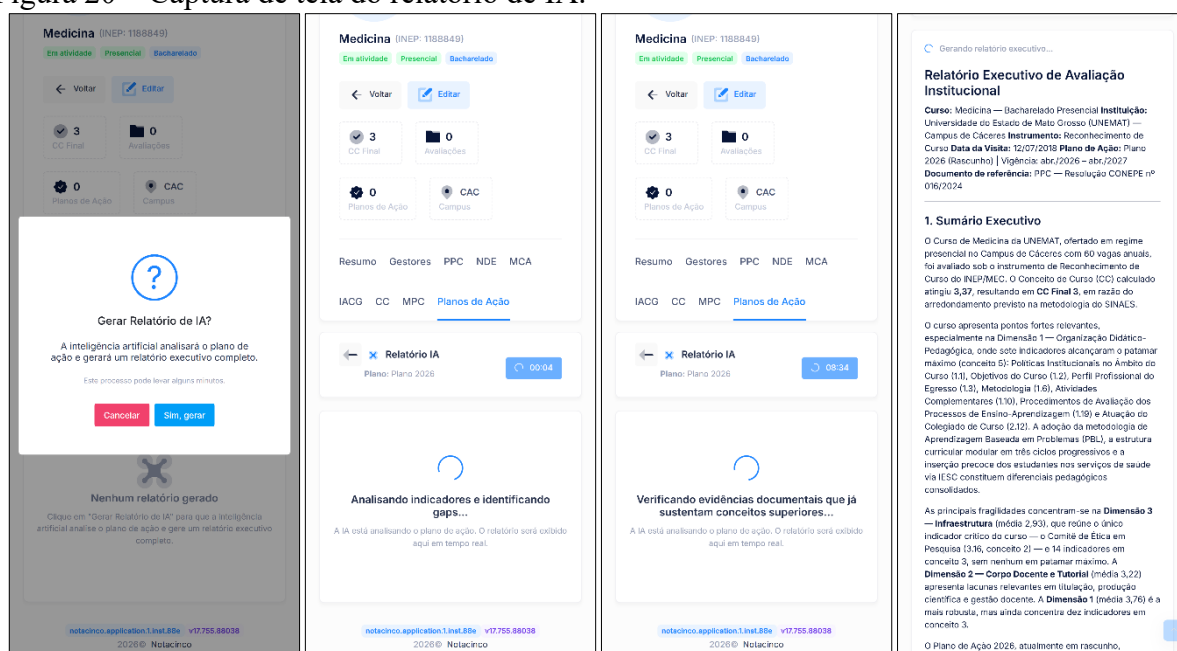
6.4.2.7 Execução sequencial dos agentes de IA e geração do relatório

A sétima etapa constitui o núcleo analítico da prova de conceito: a execução sequencial dos três agentes especializados do *framework* sobre o plano de ação previamente elaborado. O processamento é iniciado a partir da interface do plano de ação, com o acionamento da geração do relatório de avaliação integrado. A cadeia de agentes opera conforme a orquestração sequencial determinística descrita na seção 5.2.3: o Agente de Diagnóstico processa os indicadores selecionados, comparando conceitos atuais com os atributos de progressão da MPC e classificando prioridades; seu *output* alimenta o Agente de Recomendação, que consulta ativamente a base RAG para fundamentar orientações específicas; por fim, o Agente de Síntese consolida diagnóstico e recomendações em relatório executivo estruturado.

A execução é processada de forma assíncrona, com *streaming* da resposta em tempo real na interface por meio de *Server-Sent Events* (SSE), permitindo ao gestor acompanhar a geração progressiva do relatório. O tempo total de execução da cadeia completa dos três agentes é registrado pelo painel de transparência de IA e apresentado na subseção 6.4.4. O relatório produzido pelo Agente de Síntese é disponibilizado em formato *Markdown*, renderizado na interface da aplicação e exportável em PDF, organizado nas seis seções padronizadas definidas no *prompt* do agente: (a) sumário executivo; (b) diagnóstico por dimensão, com médias e

indicadores prioritários; (c) plano de ação consolidado com cronograma e responsáveis; (d) análise de artefatos e evidências; (e) recomendações estratégicas com identificação de impacto imediato; e (f) considerações finais com projeção de melhoria no CC. Observou-se que o relatório atendeu aos requisitos de linguagem formal adequada a documento institucional, fundamentando afirmações em dados concretos do curso e em fragmentos recuperados da base RAG, com recomendações contextualizadas ao perfil específico do curso avaliado.

Figura 20 – Captura de tela do relatório de IA.



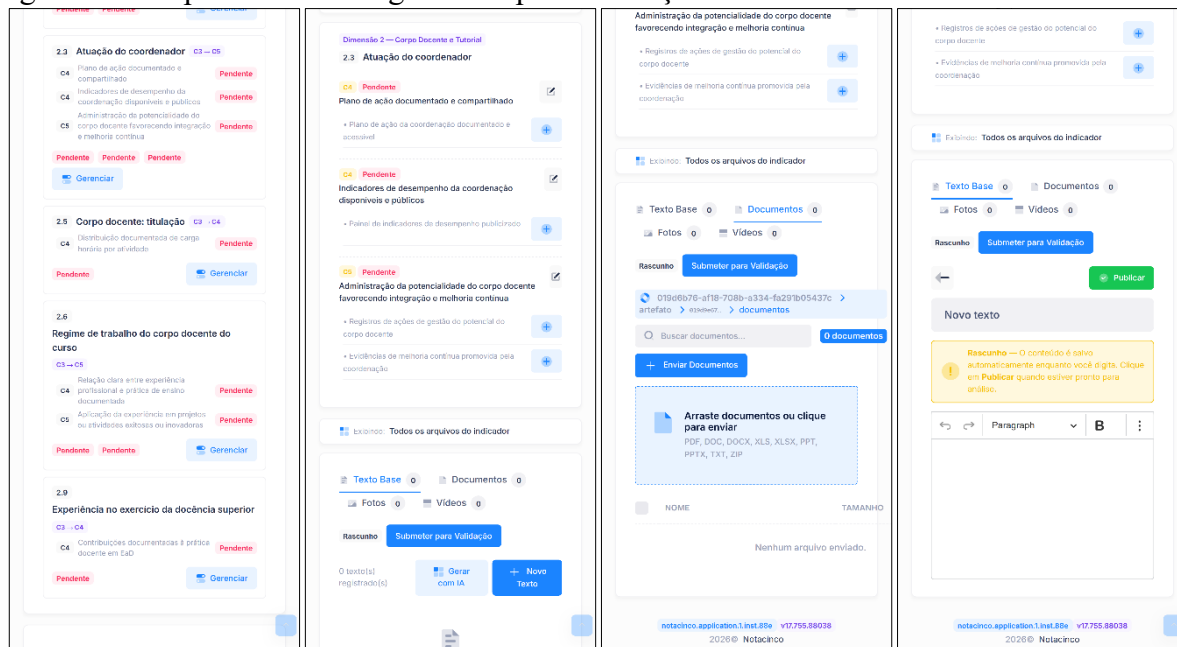
Fonte: Elaborado pelo autor (2026).

6.4.2.8 Preenchimento do plano de ação com evidências por indicador

A oitava etapa consiste na operacionalização do plano de ação pelo gestor do curso, indicador a indicador. Para cada indicador contemplado no planejamento, o gestor deve fornecer evidências que demonstrem o cumprimento integral do que o indicador requer para o alcance do conceito-alvo. O relatório gerado pelos agentes na etapa anterior opera como instrumento orientador, indicando ao gestor o que produzir, quais responsáveis envolver e quais evidências documentais priorizar. A plataforma aceita evidências em múltiplos formatos, textos estruturados, documentos em PDF e DOC, imagens e registros em vídeo, com vistas a contemplar a diversidade de artefatos institucionais exigidos pelas dimensões do instrumento de avaliação. Concluído o cadastro das evidências de todos os indicadores contemplados no planejamento, o gestor submete o conjunto para validação pela Diretoria de Regulação,

encerrando o ciclo operacional sob sua responsabilidade.

Figura 21 – Captura de tela da gestão de planos de ação e cadastro de evidências.



Fonte: Elaborado pelo autor (2026).

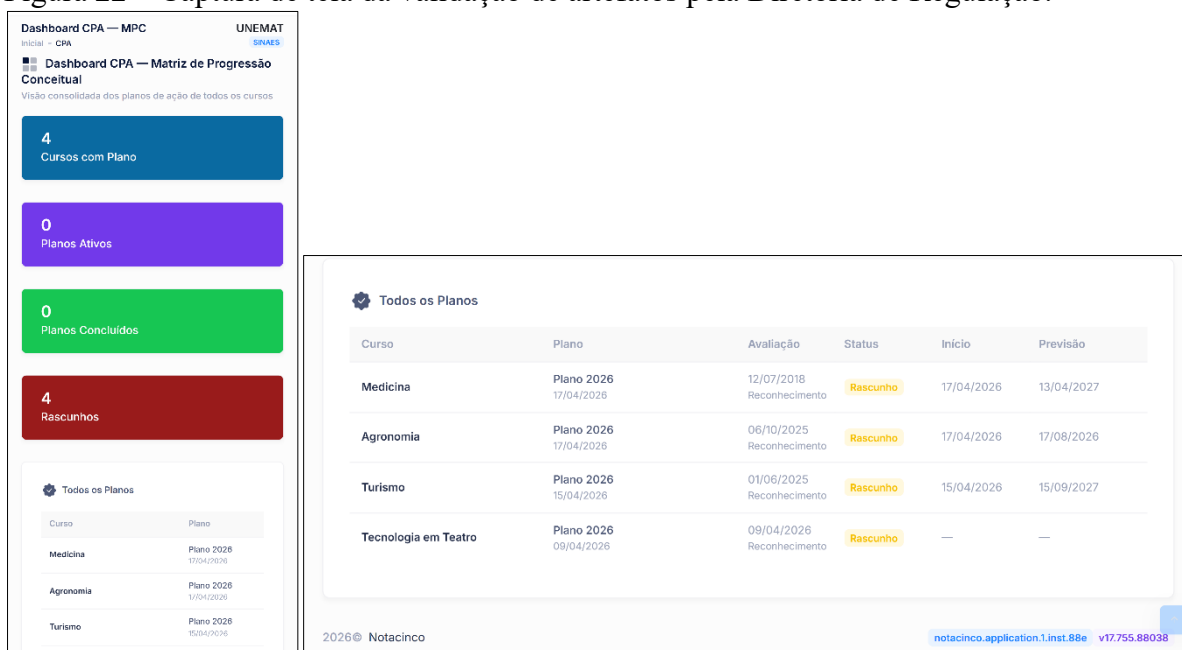
6.4.2.9 Validação dos artefatos pela DGRES e operação da Matriz MCE

A nona e última etapa é de responsabilidade da Diretoria de Regulação e consiste na validação dos artefatos submetidos pelo gestor do curso. A DGRES examina cada evidência quanto à aderência ao indicador correspondente, à suficiência documental e à conformidade com os requisitos normativos aplicáveis. As evidências validadas são incorporadas à base de conhecimento institucional, compondo o acervo de artefatos homologados que passam a ser referenciáveis por outros cursos da IES.

É nesse momento que se manifesta a operação da Matriz de Compartilhamento de Evidências (MCE). Ao classificar cada artefato conforme seu escopo de compartilhamento, institucional, de curso ou parcial, conforme descrito na seção 4.2.8, a MCE atua diretamente sobre o status do artefato e subsidia a plataforma a sinalizar, aos gestores de outros cursos, a existência de evidências previamente validadas que podem ser reaproveitadas, total ou parcialmente, em seus próprios planos de ação. Essa lógica materializa o ganho de eficiência sistêmico pretendido pelo *framework*: em vez de replicar a produção de artefatos já homologados institucionalmente, como programas de acolhimento estudantil, regulamentos de estágio ou políticas de apoio ao discente, o gestor pode vinculá-los diretamente ao curso,

concentrando seu esforço na produção das evidências que efetivamente dependem de sua atuação. Enquanto a MCA delimita o escopo de habilitação de variáveis e indicadores para cada curso, a MCE atua sobre o escopo de compartilhamento dos artefatos, configurando o par de matrizes que articula, respectivamente, a aplicabilidade analítica e a economia de esforço documental.

Figura 22 – Captura de tela da validação de artefatos pela Diretoria de Regulação.



Fonte: Elaborado pelo autor (2026).

6.4.3 Outputs Gerados

A execução integral do *framework* produziu três *outputs* intermediários e um *output* final consolidado, correspondentes às saídas de cada agente da cadeia sequencial. A Tabela 49 sintetiza os *outputs* gerados, com indicação de conteúdo principal.

Tabela 49 – Síntese dos *outputs* gerados na prova de conceito

Agente	Conteúdo Principal
Diagnóstico	Análise por dimensão; identificação de <i>gaps</i> entre conceitos atuais e desejados; classificação de prioridades.
Recomendação	Orientações com ações concretas, responsáveis, prazos, evidências e nível de esforço por indicador.
Síntese	Relatório executivo estruturado em seis seções, com sumário, diagnóstico, plano de ação, artefatos, recomendações

estratégicas e projeção de melhoria do CC.

Fonte: Elaborado pelo autor (2026).

O diagnóstico produzido pelo primeiro agente identificou os indicadores com maior *gap* conceitual, organizando-os por dimensão e classificando a prioridade de intervenção conforme a distância entre o conceito atual e o conceito-alvo definido no plano de ação. As recomendações do segundo agente demonstraram capacidade de fundamentação contextual, referenciando trechos específicos do PPC e das DCN recuperados via busca semântica na base RAG. O relatório final, produzido pelo Agente de Síntese, consolidou as análises em documento coeso, com linguagem formal adequada ao contexto institucional e estrutura padronizada que facilita a leitura e a tomada de decisão pelo gestor.

Figura 23 – Captura de tela do relatório executivo.

<p>Com todos os dados consolidados, o relatório executivo está pronto.</p> <p>Relatório Executivo de Avaliação Institucional</p> <p>Curso: Medicina — Bacharelado Presencial Instituição: Universidade Federal de Minas Gerais (UFMG) — Campus de Ciências Instrumento: RAG Curso Data da Visita: 10/07/2026 Plano de Ação: Plano 2026 (Res. Abr. 2026 — Abr. 2027) Documento de referência: PPC — Resolução 016/2024</p> <p>1. Sumário Executivo</p> <p>O curso de Medicina da UFMG, ofertado em regime presencial em Cáceres com 60 vagas anuais, foi avaliado sob o Instrumento de Curso do INEP/MELC. O Conceito do Curso (CC) calculado atingiu o CC Final 3, em razão do arredondamento previsto na metodologia.</p> <p>O curso apresenta pontos fortes relevantes, especialmente na Organização Didático-Pedagógica, onde sete indicadores alcançaram o conceito 5 (conceito 5): Políticas Institucionais no Âmbito do Curso (1.1), Objetivos do Curso (1.2), Perfil Profissional do Egresso (1.3), Metodologia (1.8), Atividades Complementares (1.10), Procedimentos de Avaliação dos Processos de Ensino-Aprendizagem (1.19). Procedimentos de Avaliação dos Processos de Ensino-Aprendizagem (1.19), Procedimentos de Avaliação dos Processos de Ensino-Aprendizagem (1.19) e Atuação do Colegiado de Curso (1.12), a metodologia de Aprendizagem Baseada em Problemas (PBL), a modular em três ciclos progressivos e a inserção precoce dos estudantes nos serviços de saúde via IESC constituem diferenciais pedagógicos.</p> <p>As principais fragilidades concentram-se na Dimensão 3 — Infraestrutura (3.3), conceito 3 — e 14 indicadores em conceito 3, sem nenhum conceito 4. A Dimensão 2 — Corpo Docente e Tutorial (Índice 3.22) relevantes em titulação, produção científica e gestão docente. A (3.27) é a mais robusta, mas ainda concentra dez indicadores em conceito 3.</p> <p>O Plano de Ação 2026, atualizado em rascunho, contempla 37 Itens entre as três dimensões. Nenhum anexo ou evidência foi produzida para o plano de ação. A análise de ganhos imediatos identificou 4 indicadores com conceito 3 superior ao atual, sem necessidade de produção de novos dados representando oportunidade de progresso imediato mediante evidências já existentes no PPC.</p> <p>A projeção de melhoria do CC, caso as ações prioritárias sejam implementadas para um CC calculado entre 3,8 e 4,2, com potencial de CC Final próximo de implementação e da regularização do CEP.</p> <p>2. Diagnóstico por Dimensão</p> <p>2.1 Dimensão 1 — Organização Didático-Pedagógica</p>	<p>2. Diagnóstico por Dimensão</p> <p>2.1 Dimensão 1 — Organização Didático-Pedagógica</p> <p>Média da dimensão: 3,78 Peso: 40%</p> <p>A Dimensão 1 é a mais desenvolvida do curso, sustentada por seu conceito 5 e pela robustez do projeto pedagógico fundamentado. Interesse estruturado em todos por áreas de complexidade crescente.</p> <p>Indicadores em patamar máximo (conceito 5):</p> <table border="1"> <thead> <tr> <th>Indicador</th> <th>Descrição</th> </tr> </thead> <tbody> <tr> <td>1.1</td> <td>Políticas Institucionais no Âmbito do Curso</td> </tr> <tr> <td>1.2</td> <td>Objetivos do Curso</td> </tr> <tr> <td>1.3</td> <td>Perfil Profissional do Egresso</td> </tr> <tr> <td>1.8</td> <td>Metodologia</td> </tr> <tr> <td>1.10</td> <td>Atividades Complementares</td> </tr> <tr> <td>1.19</td> <td>Procedimentos de Avaliação dos Processos de Ensino-Aprendizagem</td> </tr> </tbody> </table> <p>Indicadores com conceito ≤ 3 (prioridade de intervenção):</p> <table border="1"> <thead> <tr> <th>Indicador</th> <th>Descrição</th> </tr> </thead> <tbody> <tr> <td>1.5</td> <td>Conteúdos Curriculares</td> </tr> <tr> <td>1.7</td> <td>Estágio Curricular Supervisionado</td> </tr> <tr> <td>1.11</td> <td>Trabalho de Conclusão de Curso (TCC)</td> </tr> <tr> <td>1.12</td> <td>Apoio ao Discente</td> </tr> <tr> <td>1.13</td> <td>Gestão do Curso e Processos de Avaliação Interna Externa</td> </tr> <tr> <td>1.16</td> <td>Tecnologias de Informação e Comunicação no Processo Ensino-Aprendizagem</td> </tr> <tr> <td>1.18</td> <td>Material Didático</td> </tr> <tr> <td>1.20</td> <td>Número de Vagas</td> </tr> <tr> <td>1.22</td> <td>Integração do Curso com o Sistema Local e Região Saúde (SUS)</td> </tr> <tr> <td>1.23</td> <td>Atividades Práticas de Ensino para Áreas da Saúde</td> </tr> </tbody> </table> <p>Indicador em patamar intermediário:</p>	Indicador	Descrição	1.1	Políticas Institucionais no Âmbito do Curso	1.2	Objetivos do Curso	1.3	Perfil Profissional do Egresso	1.8	Metodologia	1.10	Atividades Complementares	1.19	Procedimentos de Avaliação dos Processos de Ensino-Aprendizagem	Indicador	Descrição	1.5	Conteúdos Curriculares	1.7	Estágio Curricular Supervisionado	1.11	Trabalho de Conclusão de Curso (TCC)	1.12	Apoio ao Discente	1.13	Gestão do Curso e Processos de Avaliação Interna Externa	1.16	Tecnologias de Informação e Comunicação no Processo Ensino-Aprendizagem	1.18	Material Didático	1.20	Número de Vagas	1.22	Integração do Curso com o Sistema Local e Região Saúde (SUS)	1.23	Atividades Práticas de Ensino para Áreas da Saúde	<p>4. Plano de Ação Consolidado</p> <p>O Plano de Ação 2026 contempla 37 Itens, todos com status pendentes. A seguir, as ações são organizadas por horizonte temporal.</p> <p>4.1 Ações de Curto Prazo (1 a 3 meses)</p> <p>Ações com maior retorno imediato, predominantemente de natureza organizacional, com esforço baixo a médio.</p> <table border="1"> <thead> <tr> <th>Ação</th> <th>Indicador</th> <th>Conceito Atual — Meta</th> <th>Responsável</th> </tr> </thead> <tbody> <tr> <td>Formalizar convênio com instituição parceira de tutoria do CEP/CONEP</td> <td>3.16</td> <td>2 — 3</td> <td>Reitoria / Pró-Reitoria de Pesquisa</td> </tr> <tr> <td>Revisar composição do NDC em regime integral e representatividade por área</td> <td>2.1</td> <td>3 — 4</td> <td>Coordenador do Curso</td> </tr> <tr> <td>Elaborar relatório semestral de gestão do corpo docente</td> <td>2.3</td> <td>3 + 4</td> <td>Coordenador do Curso</td> </tr> <tr> <td>Formalizar portaria de designação do coordenador com carga horária de gestão</td> <td>2.4</td> <td>3 — 4</td> <td>Coordenador + Direção da Faculdade</td> </tr> <tr> <td>Elaborar mapa/orçário dos cenários de prática do IESC e do internato</td> <td>1.22</td> <td>3 — 4</td> <td>Coordenador da IESC + Coordenador do Internato</td> </tr> <tr> <td>Documentar estratégias de integração ensino-hospitalar</td> <td>1.7</td> <td>3 — 4</td> <td>Coordenador de Integração</td> </tr> </tbody> </table>	Ação	Indicador	Conceito Atual — Meta	Responsável	Formalizar convênio com instituição parceira de tutoria do CEP/CONEP	3.16	2 — 3	Reitoria / Pró-Reitoria de Pesquisa	Revisar composição do NDC em regime integral e representatividade por área	2.1	3 — 4	Coordenador do Curso	Elaborar relatório semestral de gestão do corpo docente	2.3	3 + 4	Coordenador do Curso	Formalizar portaria de designação do coordenador com carga horária de gestão	2.4	3 — 4	Coordenador + Direção da Faculdade	Elaborar mapa/orçário dos cenários de prática do IESC e do internato	1.22	3 — 4	Coordenador da IESC + Coordenador do Internato	Documentar estratégias de integração ensino-hospitalar	1.7	3 — 4	Coordenador de Integração	<p>7. Considerações Finais</p> <p>7.1 Projeção de Melhoria no Conceito de Curso</p> <p>O CC calculado atual é de 3,37, resultando em CC Final 3. A projeção considera dois cenários:</p> <p>Cenário conservador — execução das ações de curto prazo e ganhos imediatos identificados (indicadores 1.7, 1.22, 1.23 e 3.12) e conceito 4, regularização do CEP para conceito 3 e elevação de 3 a Dimensão 2 para conceito 4; CC calculado projetado entre 3,8 e 4.</p> <p>Cenário otimista — execução integral do plano, incluindo ações de longo prazo, com progressão de indicadores-chave para conceito 5 (nas Dimensões 3 e elevação da produção científica docente; CC calculado entre 3,9 e 4,2, com CC Final 4 consolidado e base para futura projeção).</p> <p>A progressão do CC de 3 para 4 é factível dentro da vigência do Plano de Ação (2026-2027), desde que as ações de curto prazo sejam iniciadas em prazos sejam formalizados antes do início de vigência do plano.</p> <p>7.2 Próximos Passos Sugeridos</p> <ol style="list-style-type: none"> Imediato (até 30 dias): Definir prazos e responsáveis para o Plano de Ação 2026, priorizando as ações de curto prazo para iniciar o processo de convênio com instituição parceira de tutoria CEP/CONEP. Curto prazo (até 60 dias): Produzir o conjunto de documentos docente (Instituição de corpo docente, portarias do NDC, atas e mapa de cenários de prática do IESC e do internato, e as portarias designação de equipes técnicas para os laboratórios. Formalizar os quatro indicadores com ganho imediato identificados. Médio prazo (até 180 dias): Contratar acervo virtual para a base de dados do TCC, criar o repositório institucional de trabalhos, projeto de autoavaliação do curso e iniciar o processo de certificação junto ao CONEP. Longo prazo (até 360 dias): Obter homologação do CEP para o Programa de Incentivo à Produção Científica Docente com o intuito de iniciar o processo de credenciamento do CEUA próprio junto ao MEC. Contínuo: Registrar sistematicamente todas as evidências produzidas em nome do Plano de Ação 2026, construindo a memória e subsidiar os próximos ciclos avaliativos. <p><i>Relatório gerado com base nos dados do Plano de Ação 2026 (Res. Indicadores da avaliação de reconhecimento (visita de 12/07/2024) de Medicina da UFMG — Resolução CONEP nº 016/2024 — e Curriculares Nacionais para o Curso de Graduação em Medicina (3/2/025). Nenhum anexo foi localizado na base institucional. A D</i></p>
Indicador	Descrição																																																																		
1.1	Políticas Institucionais no Âmbito do Curso																																																																		
1.2	Objetivos do Curso																																																																		
1.3	Perfil Profissional do Egresso																																																																		
1.8	Metodologia																																																																		
1.10	Atividades Complementares																																																																		
1.19	Procedimentos de Avaliação dos Processos de Ensino-Aprendizagem																																																																		
Indicador	Descrição																																																																		
1.5	Conteúdos Curriculares																																																																		
1.7	Estágio Curricular Supervisionado																																																																		
1.11	Trabalho de Conclusão de Curso (TCC)																																																																		
1.12	Apoio ao Discente																																																																		
1.13	Gestão do Curso e Processos de Avaliação Interna Externa																																																																		
1.16	Tecnologias de Informação e Comunicação no Processo Ensino-Aprendizagem																																																																		
1.18	Material Didático																																																																		
1.20	Número de Vagas																																																																		
1.22	Integração do Curso com o Sistema Local e Região Saúde (SUS)																																																																		
1.23	Atividades Práticas de Ensino para Áreas da Saúde																																																																		
Ação	Indicador	Conceito Atual — Meta	Responsável																																																																
Formalizar convênio com instituição parceira de tutoria do CEP/CONEP	3.16	2 — 3	Reitoria / Pró-Reitoria de Pesquisa																																																																
Revisar composição do NDC em regime integral e representatividade por área	2.1	3 — 4	Coordenador do Curso																																																																
Elaborar relatório semestral de gestão do corpo docente	2.3	3 + 4	Coordenador do Curso																																																																
Formalizar portaria de designação do coordenador com carga horária de gestão	2.4	3 — 4	Coordenador + Direção da Faculdade																																																																
Elaborar mapa/orçário dos cenários de prática do IESC e do internato	1.22	3 — 4	Coordenador da IESC + Coordenador do Internato																																																																
Documentar estratégias de integração ensino-hospitalar	1.7	3 — 4	Coordenador de Integração																																																																

Fonte: Elaborado pelo autor (2026).

6.4.4 Registros de Transparência de IA

Em conformidade com os mecanismos de transparência e auditabilidade descritos na seção 5.2.11, todos os registros das interações dos agentes com o provedor de IA foram capturados pelo painel de transparência da plataforma. A Tabela 50 apresenta os registros consolidados da execução da prova de conceito, extraídos da tabela de *logs* de agentes do sistema.

Tabela 50 – Registros de transparência de IA da prova de conceito

Agente	Modelo	Custo	Tokens (output)	Tempo (seg)	Ferramentas
Diagnóstico	<i>claude-sonnet-4-6</i>	\$0.3072	53.969	158.5s	DiagnosticoAgent
Recomendação	<i>claude-sonnet-4-6</i>	\$0.7784	184.268	319.9s	RecomendacaoAgent
Síntese	<i>claude-sonnet-4-6</i>	\$0.9728	303.313	106.4s	GanhoImediatoAgent
Total	-	\$ 2,0584	541.550	584,8	-

Fonte: Elaborado pelo autor (2026), a partir dos registros do painel de transparência de IA da plataforma Notacinc.

Os registros de transparência permitem observar o consumo de recursos computacionais de cada agente ao longo da cadeia de processamento. O Agente de Diagnóstico, que opera com temperatura de 0,1 e acesso a quatro ferramentas (indicadores, diagnóstico MPC, base RAG e avaliação), realiza o processamento inicial dos indicadores selecionados. O Agente de Recomendação, configurado com temperatura de 0,3 e acesso a três ferramentas (base RAG, artefatos e plano de ação), demanda volume de *tokens* de *input* elevado em razão de receber o *output* integral do agente anterior em seu *prompt*. O Agente de Síntese, com temperatura de 0,2 e limite expandido de 16.384 *tokens* de saída, concentra o maior volume de *input* por consolidar os *outputs* dos dois agentes precedentes.

O tempo total de execução da cadeia completa constitui informação relevante para a viabilidade operacional do *framework* em contexto de uso real. A execução assíncrona com *streaming* via SSE mitiga a percepção de latência pelo usuário, pois o relatório começa a ser exibido na interface antes de sua conclusão integral. Essa estratégia de apresentação progressiva é fundamental para a experiência de uso em cenários nos quais o tempo de processamento pode atingir dezenas de segundos, mantendo o engajamento do gestor durante a geração do relatório.

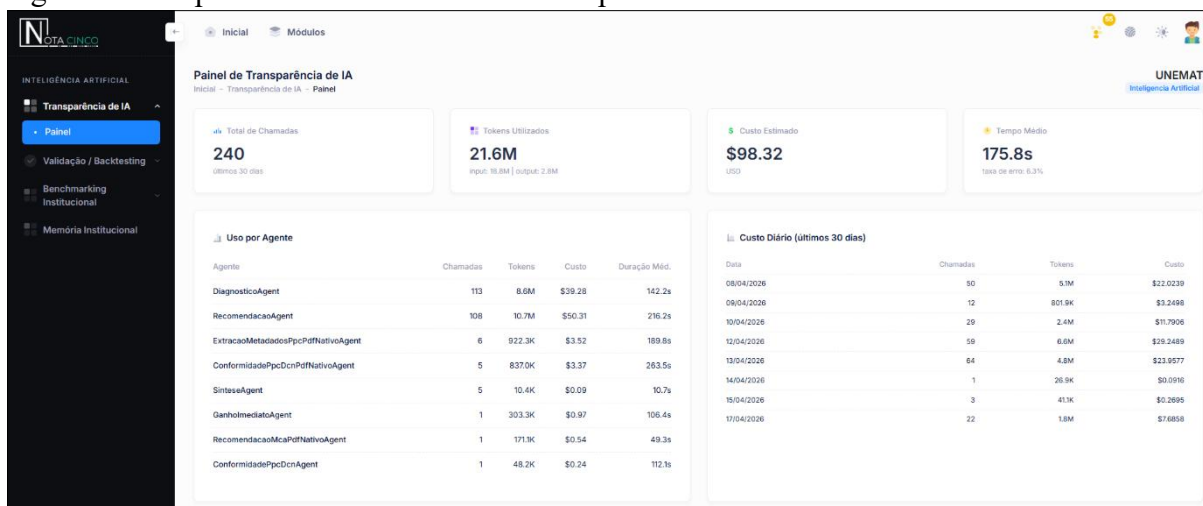
A disponibilização dos registros de transparência ao usuário final atende ao princípio de supervisão humana que permeia a arquitetura do *framework* (seção 5.2.11). Ao visualizar o modelo utilizado, os *tokens* consumidos e as ferramentas invocadas, o gestor dispõe de elementos para avaliar a confiabilidade do processo analítico e, se necessário, solicitar nova execução ou ajustar os parâmetros do plano de ação. Essa transparência operacional alinha-se aos princípios de uso responsável de IA preconizados pela Estratégia Brasileira de Inteligência Artificial (BRASIL, 2021).

Em síntese, a prova de conceito demonstrou que o *framework* é capaz de executar integralmente o ciclo analítico proposto, desde o cadastro institucional do curso pela Diretoria de Regulação até a validação dos artefatos produzidos e a sinalização de evidências

reaproveitáveis por meio da MCE. Todas as etapas operaram conforme projetado: o motor de regras da MCA filtrou corretamente os indicadores aplicáveis; a conferência entre a nota calculada e a nota informada no cadastro da avaliação atuou como salvaguarda contra erros de digitação; os conceitos foram processados pelo serviço de diagnóstico MPC; a base RAG forneceu fundamentação documental contextualizada; os três agentes de IA produziram, em cadeia sequencial, diagnóstico, recomendações e relatório coerentes entre si e aderentes ao contexto normativo do SINAES; e a MCE operou ao final do ciclo como mecanismo de governança do reaproveitamento institucional de artefatos já homologados.

Conforme enfatizado por Peffers *et al.* (2007), a demonstração não substitui a avaliação formal do artefato, mas a complementa ao evidenciar sua viabilidade técnica e operacional. Combinada com os resultados dos procedimentos de validação apresentados nas seções 6.1 a 6.3, a prova de conceito fornece evidência complementar que sustenta a adequação do *framework* para uso em contextos reais de gestão de cursos de graduação, atendendo à atividade de Demonstração do DSRM e operacionalizando a dimensão de verificação de viabilidade do Objetivo Específico V da pesquisa.

Figura 24 – Captura de tela do Painel de Transparência de IA.



Fonte: Elaborado pelo autor (2026).

6.5 Discussão Geral dos Resultados

As seções precedentes deste capítulo documentaram os resultados de três procedimentos de validação e uma prova de conceito, conforme preconizado pela estratégia multimétodo definida na Metodologia. A presente seção consolida esses resultados em

perspectiva integrada, articulando as evidências obtidas com a literatura pertinente e identificando as contribuições e limitações do *framework* proposto.

6.5.1 Síntese das Métricas

A Tabela 51 consolida as principais métricas de desempenho do *framework* obtidas nos três procedimentos de validação, permitindo a apreciação integrada dos resultados.

Tabela 51 – Síntese consolidada das métricas de validação

Métrica	Cenários Controlados (6.1)	Análise Retrospectiva (6.2)	Consistência Interna (6.3)
F1 Geral (média)	0,8546 (DP 0,1981)	0,9582 (DP 0,0872)	-
Acurácia do Diagnóstico (AD)	-	1,0000 (DP 0)	1,0000 (DP 0)
CCE	-	-	0,9759 (DP 0,0476)
ICS	-	-	0,9509 (DP 0,006)
TRA	0,9543	0,4910	-
TAO	0,5733	0,2432	-
Estabilidade MCE (componente e)	-	-	0,2749
Total de execuções	21	39	39 (pós-processamento)
Custo total (USD)	21,93	35,47	-

Fonte: Elaborado pelo autor (2026). Os valores reportados correspondem as médias consolidadas de cada procedimento.

A leitura integrada do Tabela 51 revela um padrão de desempenho consistente: o *framework* demonstra alta eficácia na classificação de indicadores e na detecção de fragilidades críticas, com confiabilidade operacional elevada, ao mesmo tempo em que apresenta limitações concentradas no reconhecimento e orientação de evidências documentais (MCE). Esse padrão permanece estável ao longo dos três procedimentos, o que confere robustez a interpretação.

A melhoria do F1 Geral entre cenários controlados (0,8546) e análise retrospectiva (0,9582) indica que o *framework* opera com maior eficácia em dados reais do que em cenários sintéticos deliberadamente concebidos para exercitar situações-limite. A Acurácia do

Diagnóstico (AD) de 1,0000 em todas as 32 execuções com indicadores críticos constitui o achado mais expressivo da validação: nenhum indicador com conceitos 1 ou 2 foi omitido pelo sistema em qualquer execução, eliminando o risco de falsos negativos críticos que poderiam comprometer a confiança do gestor no artefato.

O Coeficiente de Consistência Estrutural (CCE) de 0,9759 e o Índice de Consistência Semântica (ICS) de 0,9509 demonstram que a natureza estocástica do modelo de linguagem não compromete a reprodutibilidade dos resultados diagnósticos. Em termos práticos, um coordenador de curso que execute o *framework* em momentos distintos obterá diagnósticos estruturalmente equivalentes em mais de 97% dos elementos analisados e recomendações semanticamente equivalentes em mais de 95% dos casos.

O custo total de processamento de USD 57,40 para 60 execuções (média de USD 0,96 por execução) situa o *framework* em patamar de viabilidade econômica para operação em escala institucional. Considerando que cada execução produz diagnóstico completo, recomendações contextualizadas e relatório executivo para um curso de graduação, o custo por análise é marginal frente ao investimento de horas-técnico que a mesma tarefa demandaria em formato convencional.

6.5.2 Contribuições do *Framework*

Os resultados da validação permitem identificar cinco contribuições principais do *framework* proposto, organizadas em contribuições de natureza metodológica e contribuições de natureza prática.

Detecção exaustiva de fragilidades críticas. A AD de 1,0000 evidencia que o *framework* identifica a totalidade dos indicadores com conceitos 1 ou 2, sem exceção. Essa capacidade é particularmente relevante no contexto do SINAES, no qual a não identificação de uma fragilidade severa pode resultar em conceito insatisfatório na avaliação *in loco*, com consequências regulatórias para o curso. O desempenho observado supera a expectativa metodológica (limiar de 0,80) e sugere que a combinação entre motor de regras determinístico (MPC/MCA) e agente de IA com temperatura baixa (0,1) constitui abordagem eficaz para tarefas analíticas que exigem sensibilidade máxima.

Diagnóstico contextualizado e reprodutível. O F1 Geral elevado (0,9582 na retrospectiva) indica que o *framework* não apenas detecta fragilidades, mas classifica corretamente a prioridade de intervenção para a grande maioria dos indicadores. A

reprodutibilidade, atestada pelo CCE de 0,9759, diferencia o artefato de abordagens puramente generativas, nas quais a variabilidade dos *outputs* e frequentemente apontada como limitação para uso em contextos decisórios (AMATRIAIN, 2024). A ancoragem dos agentes em *ground truth* determinístico, gerado pelo motor de regras MPC/MCA, confere estabilidade estrutural sem eliminar a capacidade generativa necessária para produzir recomendações textuais contextualizadas.

Recomendações fundamentadas em evidências documentais. A integração do *pipeline* RAG permite que os agentes de Recomendação e Síntese fundamentem suas orientações em trechos específicos do PPC, das DCN e de planejamentos anteriores do curso, superando a limitação de recomendações genéricas frequentemente observada em sistemas de apoio a decisão educacional. A consistência semântica elevada (ICS = 0,9509) assegura que a fundamentação documental se mantem estável entre execuções, mesmo com variação na redação.

Filtragem contextual por perfil do curso. A MCA, operacionalizada como motor de regras integrado ao *framework*, demonstrou capacidade de filtrar corretamente os indicadores não aplicáveis em todos os cenários e cursos avaliados. O tratamento de indicadores NSA obteve F1 consistentemente elevado, incluindo $F1 = 1,0000$ no Cenário A. Essa contribuição é significativa considerando que aproximadamente 58% dos indicadores no instrumento de Autorização e 59% no instrumento de Reconhecimento possuem condições de aplicabilidade, conforme identificado na construção da MCA (Capítulo 4).

Operacionalização da Análise de Conteúdo em artefato computacional. Do ponto de vista metodológico, o *framework* demonstra a viabilidade de transformar os resultados de Análise de Conteúdo Categórica Temática (BARDIN, 2016), materializados nas matrizes MPC, MCA e MCE, em base de conhecimento normativo operacionalizável por agentes de IA. Essa contribuição extrapola o domínio da avaliação educacional, sugerindo que a abordagem pode ser adaptada a outros contextos regulatórios nos quais instrumentos normativos estruturados orientam processos de avaliação e certificação.

6.5.3 Limitações Identificadas

A análise dos resultados permite identificar limitações em quatro dimensões, cuja explicitação é necessária para a correta interpretação do alcance do *framework* e para a orientação de trabalhos futuros.

Reconhecimento e orientação de evidências (MCE). A limitação mais significativa reside no desempenho das métricas MCE. A TAO abaixo do limiar em todos os cenários (0,2432 a 0,6111), o ICC-N2 próximo de zero (0,1757 na retrospectiva) e a instabilidade do componente MCE na consistência interna (0,2749) indicam que o *framework*, em sua configuração atual, não atinge o nível de sofisticação necessário para orientar de forma confiável o compartilhamento de evidências entre cursos. Essa limitação não compromete a utilidade primária do artefato (diagnóstico e recomendações por indicador), porém restringe o potencial da MCE como instrumento de otimização de recursos em contextos *multicampi*.

Identificação da dimensão prioritária. A taxa de acerto da dimensão prioritária foi de 56,41% na análise retrospectiva, com desempenho concentrado em cursos cujo perfil de fragilidades favorece a identificação inequívoca (Pedagogia e Tecnologia em Gestão Turismo com 100%). Nos cursos em que a distribuição de *gaps* entre dimensões é mais equilibrada, o agente tende a divergir do critério determinístico do *ground truth*. Essa divergência pode refletir tanto limitação do agente quanto rigidez excessiva da métrica, dado que a priorização de dimensões admite múltiplas perspectivas validas em contextos de gestão.

Sensibilidade em perfis de baixa fragilidade. O comportamento observado no curso de Educação Física (CC = 4,44) revela que o *framework* apresenta instabilidade classificatória em cursos com poucos sinais discriminativos. Em duas das sete execuções, indicadores de conceito 3 foram sobreclassificados como críticos, elevando o desvio-padrão do F1 para 0,2008. Embora esse comportamento não afete a identificação dos indicadores prioritários (componente (a) do CCE = 1,0000), pode gerar ruído na comunicação de resultados ao coordenador.

Fragilidade no *pipeline* de extração estruturada. A ocorrência de uma execução com F1 = 0 no Cenário C (4,76% das execuções nos cenários controlados) decorre de falha no *parser* de extração, não de erro analítico do agente. Essa vulnerabilidade, embora pontual, sinaliza a necessidade de mecanismos de validação de formato e tentativa automática no *pipeline* de processamento, ou ainda a adoção de *output* estruturado nativo (*structured output*) disponível nos modelos contemporâneos.

Ausência de validação por usuários finais. Os procedimentos de validação adotados nesta pesquisa mensuram propriedades objetivas do artefato (correção, relevância, consistência) por meio de métricas automatizadas e classificação por agente auxiliar. A validação da utilidade percebida por coordenadores de curso e gestores de IES, que constitui dimensão complementar relevante no *Design Science Research* (VENABLE; PRIES-HEJE;

BASKERVILLE, 2016), não foi conduzida no escopo desta dissertação, constituindo oportunidade para trabalhos futuros.

Tamanho amostral e representatividade. A validação foi conduzida com 3 cenários sintéticos, 5 cursos reais da mesma instituição (UNEMAT), 21 execuções controladas e 39 execuções retrospectivas. Nenhum teste de significância estatística foi aplicado, e nenhum intervalo de confiança foi calculado para as métricas reportadas. As métricas apresentadas (médias, desvios-padrão, taxas de acerto) são, portanto, descritivas e não inferenciais: afirmações como "supera o limiar" não são acompanhadas de significância estatística. Adicionalmente, a amostra de cursos é circunscrita a uma única instituição pública estadual, com cursos presenciais cujos CC variam entre (3,33 e 4,44), sem representação de cursos com CC nas faixas extremas (1, 2 ou 5), de instituições privadas, de universidades federais ou de cursos na modalidade a distância. Essa circunscrição limita a generalização dos resultados para o universo das IES brasileiras.

6.5.4 Comparação com a Literatura

A comparação dos resultados do *framework* proposto com a literatura apresenta desafio metodológico: não foram identificados, na revisão de literatura conduzida, sistemas que combinem agentes de IA generativa, matrizes normativas derivadas de Análise de Conteúdo e arquitetura RAG aplicados especificamente a avaliação de cursos de graduação no contexto do SINAES. Essa lacuna, que justificou a própria pesquisa, implica que a comparação direta de métricas com sistemas equivalentes não é possível. Não obstante, é viável posicionar os resultados em relação a referenciais da literatura em três dimensões.

No que concerne a classificação automática de textos normativos, o F1 Geral de 0,9582 obtido na análise retrospectiva situa-se em patamar compatível com os resultados reportados na literatura de classificação de textos, onde valores acima de 0,85 são tipicamente considerados indicadores de desempenho elevado (SOKOLOVA; LAPALME, 2009). A AD de 1,0000, comparável ao *Recall* de classe crítica, supera os valores frequentemente reportados em sistemas de classificação multiclasse, nos quais o *Recall* de classes minoritárias tende a ser inferior ao de classes majoritárias. No contexto do *framework* proposto, esse desempenho decorre do *ground truth* determinístico fornecido pelo motor de regras MPC/MCA, que ancora o raciocínio do agente em critérios normativos explícitos.

Em relação a consistência de sistemas baseados em LLMs, os resultados obtidos

(CCE = 0,9759, ICS = 0,9509) alinham-se a evidências recentes de que a combinação de temperaturas baixas com *prompts* estruturados e conhecimento de domínio explícito mitiga a variabilidade inerente aos modelos generativos. Amatriain (2024) observa que a calibração de temperatura constitui parâmetro crítico para aplicações que demandam consistência, e os resultados desta pesquisa corroboram essa orientação, especialmente no contraste entre a estabilidade do agente de Diagnóstico (temperatura 0,1) e a variabilidade ligeiramente maior do agente de Recomendação (temperatura 0,3).

No domínio da avaliação institucional brasileira, o *framework* proposto representa abordagem distinta da predominante na literatura, que historicamente concentrou-se em análises estatísticas de indicadores do ENADE e do CPC (POLIDORI; MARINHO-ARAUJO; BARREYRO, 2006; BITTENCOURT *et al.*, 2008). Enquanto essas abordagens operam sobre dados agregados de desempenho discente e indicadores quantitativos, o *framework* atua sobre os descritores qualitativos dos instrumentos de avaliação *in loco*, transformando-os em base de conhecimento operacionalizável. Essa complementaridade sugere que o artefato proposto não substitui, mas amplia o repertório analítico disponível para a gestão de cursos de graduação.

A adoção de *Design Science Research* como abordagem metodológica posiciona o trabalho em alinhamento com a orientação de Hevner *et al.* (2004) para pesquisas que produzem artefatos tecnológicos. A estratégia de validação multimétodo empregada, combinando cenários controlados, análise retrospectiva e consistência interna, atende a diretriz de avaliação rigorosa preconizada por esses autores, ao mensurar múltiplas dimensões de qualidade do artefato (correção, relevância, confiabilidade) por meio de procedimentos complementares e reproduzíveis.

6.6 Implicações para a Gestão de IES

Os resultados apresentados e discutidos nas seções precedentes transcendem o âmbito estritamente técnico da validação do artefato e projetam implicações concretas para a prática de gestão de Instituições de Educação Superior. A presente seção articula essas implicações, considerando o contexto regulatório do SINAES e os desafios operacionais enfrentados por coordenadores de curso e Comissões Próprias de Avaliação (CPA).

Redução da assimetria informacional no processo de autoavaliação. A literatura sobre avaliação institucional identifica recorrentemente a dificuldade dos coordenadores em compreender os critérios qualitativos que diferenciam os níveis conceituais nos instrumentos

do INEP (DIAS SOBRINHO, 2010). Os descritores, formulados em linguagem normativa e com nuances semânticas sutis entre conceitos adjacentes, exigem domínio especializado para sua correta interpretação. O *framework* proposto, ao decompor sistematicamente esses descritores em elementos constitutivos, atributos de progressão e evidências requeridas (MPC), e ao operacionalizar essa decomposição por meio de agentes de IA que produzem diagnósticos específicos por indicador, reduz a assimetria informacional entre o que o instrumento exige e o que o coordenador compreende. A AD de 1,0000 evidencia que essa tradução normativa ocorre sem perda de informação crítica.

Apoio a priorização estratégica de ações de melhoria. A gestão de cursos de graduação opera sob restrições de recursos humanos, financeiros e temporais que impõe a necessidade de priorização. O *framework* contribui para essa priorização ao classificar indicadores por nível de urgência (crítico, médio, baixo), associar responsáveis e prazos a cada recomendação e projetar o impacto das ações no Conceito de Curso. A estabilidade da classificação de prioridades (componentes (a) e (b) do CCE com média de 0,9737) assegura que as decisões de priorização não serão influenciadas por variabilidade estocástica do modelo, conferindo previsibilidade ao planejamento de ações.

Continuidade da memória institucional. Andriola (2009) identifica a rotatividade de gestores e a descontinuidade dos processos de autoavaliação como obstáculos recorrentes a efetividade da gestão da qualidade em IES brasileiras. O *framework* proposto endereça essa fragilidade por dois mecanismos: a base de conhecimento RAG, que preserva documentos institucionais indexados e consultáveis independentemente da permanência de gestores específicos; e a base de conhecimento global, que acumula experiências de cursos da plataforma em formato anonimizado e consultável. Em termos práticos, um novo coordenador que assuma a gestão de um curso pode acessar, por meio do sistema, diagnósticos anteriores, recomendações implementadas e experiências de outros cursos que enfrentaram desafios semelhantes, sem depender de transferência informal de conhecimento.

Otimização de recursos em contextos *multicampi*. A UNEMAT, como instituição de estrutura *multicampi* distribuída pelo território mato-grossense, enfrenta desafios de coordenação e padronização que se beneficiam diretamente da abordagem proposta. A MCA assegura que cada curso seja avaliado conforme seu perfil específico, enquanto a MCE, apesar das limitações identificadas na validação, delinea um caminho para o compartilhamento de evidências documentais entre cursos do mesmo *campus* ou da mesma área de conhecimento. A viabilidade econômica do sistema (custo médio de USD 0,96 por análise completa) torna

factível sua aplicação ao conjunto de cursos da instituição, sem concentração de custos que inviabilize a adoção em escala.

Preparação proativa para avaliações externas. O ciclo avaliativo do SINAES opera em periodicidade definida, e a preparação para avaliações *in loco* constitui atividade estratégica para as IES. O *framework* possibilita que coordenadores conduzam autoavaliações periódicas com suporte analítico automatizado, identificando fragilidades e implementando ações corretivas antes da visita da comissão avaliadora. A prova de conceito (seção 6.4) demonstrou que o fluxo completo, da configuração do curso a geração do relatório com plano de ação, pode ser executado de forma integral e documentada, fornecendo ao gestor instrumento concreto de preparação.

Transparência e auditabilidade do processo decisório. A integração de IA em processos de gestão institucional suscita questões legítimas sobre a confiabilidade e a explicabilidade das recomendações geradas. O *framework* responde a essas questões por meio de três mecanismos: o painel de transparência de IA, que registra cada interação com o provedor de IA (modelo, *tokens*, tempo, ferramentas); a rastreabilidade garantida pela orquestração sequencial dos agentes, na qual cada recomendação pode ser rastreada ao diagnóstico que a originou; e o princípio de supervisão humana (*human-in-the-loop*), que assegura que nenhuma ação é implementada automaticamente. Esses mecanismos alinham-se aos princípios da Estratégia Brasileira de Inteligência Artificial (BRASIL, 2021) e contribuem para a construção de confiança institucional na adoção de tecnologias de IA.

E necessário ponderar, contudo, que as implicações aqui articuladas decorrem dos resultados de validação obtidos no contexto institucional da UNEMAT, com cinco cursos e 60 execuções. A generalização para outros contextos institucionais requer cautela e, idealmente, validação complementar em IES com perfis distintos (universidades federais, centros universitários, faculdades privadas). A arquitetura *multi-tenant* da plataforma Notacinco foi concebida para viabilizar essa expansão, porém a adequação do *framework* a contextos significativamente diferentes do ambiente de estudo constituirá objeto de investigação futura.

7 CONCLUSÕES

A presente dissertação teve como objetivo geral desenvolver e testar um *framework* baseado em agentes de Inteligência Artificial para apoiar os processos de avaliação de cursos de graduação e gerar planos de ações acadêmicas e administrativas para aprimorar a qualidade educacional, conforme as diretrizes do SINAES. A pesquisa foi conduzida sob a abordagem do *Design Science Research* (PEFFERS *et al.*, 2007), percorrendo integralmente as seis atividades do modelo DSRM: identificação do problema, definição de objetivos, *design* e desenvolvimento, demonstração, avaliação e comunicação.

O Objetivo Específico I, referente ao levantamento bibliográfico, foi operacionalizado no Capítulo 2, que examinou sistematicamente a literatura sobre políticas públicas de avaliação, *Design Science Research*, Análise de Conteúdo e Inteligência Artificial generativa aplicada à gestão educacional. Essa revisão permitiu identificar a lacuna central que motivou a pesquisa: a inexistência de *frameworks* que articulem a decomposição analítica dos instrumentos normativos do INEP com as capacidades de processamento de modelos generativos.

O Objetivo Específico II foi atendido pela construção das três matrizes analíticas que estruturam a base de conhecimento normativo do *framework* (Capítulo 4). A Matriz de Progressão Conceitual (MPC), desenvolvida mediante Análise de Conteúdo Categórica Temática (BARDIN, 2016), decompôs os descritores dos cinco níveis conceituais de cada indicador nos instrumentos de Autorização (55 indicadores, 275 registros) e de Reconhecimento (58 indicadores, 290 registros). A MCA codificou as condições de aplicabilidade dos indicadores, abrangendo 25 regras lógicas que determinam quais indicadores se aplicam a cada perfil de curso. A MCE classificou o escopo de compartilhamento de evidências entre cursos, distinguindo artefatos institucionais, parciais e específicos.

O Objetivo Específico III foi concretizado pela integração de três agentes de IA generativa ao *framework* (Capítulo 5): os agentes de Diagnóstico, Recomendação e Síntese, orquestrados de forma sequencial determinística via API da Anthropic (modelo Claude Sonnet 4.6), com *pipeline* RAG para processamento de documentos institucionais e engenharia de *prompt* estruturada para aderência normativa.

O Objetivo Específico IV materializou-se na plataforma Notacinco (Capítulo 5), aplicação *web* com arquitetura *multi-tenant* em Laravel/PHP com PostgreSQL, que incorpora módulos de cadastro de cursos, motor de regras da MCA, calculadora do Conceito de Curso,

visualização de dados, execução dos agentes de IA e interface conversacional.

O Objetivo Específico V foi integralmente operacionalizado no Capítulo 6 por meio de estratégia multimétodo. A validação por cenários controlados (21 execuções, 3 cenários) obteve F1 Geral de 0,8546. A validação por análise retrospectiva com dados oficiais do INEP (39 execuções, 5 cursos da UNEMAT) elevou o F1 Geral para 0,9582 e registrou Acurácia do Diagnóstico (AD) de 1,0000 em todas as execuções com indicadores críticos. A validação por consistência interna demonstrou Coeficiente de Consistência Estrutural (CCE) de 0,9759 e Índice de Consistência Semântica (ICS) de 0,9509. A prova de conceito documentou a execução integral do *framework* em curso real da UNEMAT, comprovando sua viabilidade operacional.

Os resultados obtidos permitem concluir que o *framework* proposto atende ao objetivo geral da pesquisa: constitui artefato tecnológico capaz de transformar dados avaliativos do SINAES em diagnósticos estruturados, recomendações contextualizadas e planos de ação acionáveis para a gestão de cursos de graduação. A combinação de AD = 1,0000, F1 Geral de 0,9582 e CCE de 0,9759 configura perfil de avaliação que, nos termos de Hevner *et al.* (2004), sustenta a utilidade, a qualidade e a eficácia do artefato produzido.

7.1 O Produto Tecnológico e a Contribuição para a Inovação

No âmbito de programas de mestrado profissional, a produção de artefato tecnológico com potencial de aplicação prática constitui requisito e diferencial em relação ao mestrado acadêmico. A presente pesquisa produziu como produto tecnológico a plataforma Notacinco, sistema *web* que operacionaliza integralmente o *framework* proposto e que foi submetido a processo de registro de programa de computador junto à Agência de Inovação (AGINOV) da Universidade do Estado de Mato Grosso, para posterior encaminhamento ao Instituto Nacional da Propriedade Industrial (INPI).

O pedido de registro foi formalizado em 6 de abril de 2026, com a denominação *Notacinco*, tendo a Universidade do Estado de Mato Grosso (CNPJ 01.367.770/0001-30) como titular dos direitos de propriedade intelectual. O programa foi classificado nas seguintes categorias do INPI: Aplicativo (AP01), Planejamento (AP02), Controle (AP03), Inteligência Artificial (IA01), Sistemas Especialistas (IA02), Gerenciador de Informações (GI01), Gerador de Relatórios (GI04) e Avaliação de Desempenho (AV01). Os campos de aplicação abrangem Administração e Processo de Ensino (ED03), Ensino Regular (ED01), Currículo (ED05), Função Administrativa (AD02), Documentação (IF02) e Ciência da Informação (IF07).

A integridade do código-fonte foi assegurada por resumo criptográfico SHA-512, e o relatório de registro documenta os componentes principais incluídos no *hash*: os artefatos analíticos (MPC, MCA e MCE), as classes de serviço *core* (motor de regras, calculadora de CC e serviço de diagnóstico MPC), as classes dos quatro agentes de IA (Diagnóstico, Recomendação, Síntese e Chat) e as ferramentas (*tools*) que fornecem acesso estruturado aos dados para os agentes.

O estágio de desenvolvimento foi classificado como *parcialmente desenvolvido*, indicando que a plataforma poderá ser levada ao mercado com investimento razoável. A data de criação registrada é 13 de outubro de 2025, com protótipo em operação a partir de 22 de janeiro de 2026. A concepção do projeto remonta a 25 de julho de 2024, coincidindo com o período das atividades de pesquisa no POLEDUC/UFC.

O registro de *software* constitui contribuição tangível do mestrado profissional à inovação institucional. Ao formalizar a proteção da propriedade intelectual, a UNEMAT assegura os direitos sobre o artefato e viabiliza, no futuro, a exploração comercial da plataforma no modelo SaaS (*Software as a Service*) para atendimento a múltiplas IES, bem como a celebração de parcerias de cooperação técnica e contratos de transferência de tecnologia. A arquitetura *multi-tenant* do sistema, concebida desde o início do desenvolvimento, foi projetada para suportar essa escalabilidade.

7.2 Limitações do Estudo

A explicitação das limitações da pesquisa é requisito de rigor científico no *Design Science Research* (HEVNER *et al.*, 2004) e contribui para a correta delimitação do alcance dos resultados obtidos.

A limitação mais fundamental da pesquisa reside na circularidade da estratégia de validação. O *ground truth* utilizado em todos os procedimentos de validação, cenários controlados e análise retrospectiva, é gerado deterministicamente pelo motor de regras MPC/MCA, componente integrante do próprio *framework* avaliado. As regras de classificação codificadas no *ground truth* (conceitos 1-2 = crítico, conceito 3 = médio, conceito 4 = baixo) são idênticas às regras instruídas nos *prompts* dos agentes de IA. Essa identidade implica que os procedimentos de validação mensuram a conformidade funcional do agente, isto é, se ele segue as regras internas do sistema com fidelidade, e a estabilidade operacional, se os resultados são reproduzíveis entre execuções. O que não foi demonstrado é a eficácia prática: se as regras

do *framework*, e conseqüentemente os diagnósticos por ele produzidos, são percebidos como úteis e corretos por coordenadores de curso, membros de CPA e gestores institucionais. Essa distinção é coerente com o modelo DSRM (PEFFERS *et al.*, 2007), que separa demonstração de avaliação, e com Venable, Pries-Heje e Baskerville (2016), que reconhecem que artefatos podem ser avaliados em diferentes estágios de maturidade. O presente trabalho situa-se no estágio de verificação de viabilidade técnica; a validação de eficácia prática por especialistas humanos constitui trabalho futuro prioritário.

A segunda limitação refere-se ao escopo institucional e ao tamanho amostral da validação. Os procedimentos de avaliação foram conduzidos exclusivamente com dados de cursos da Universidade do Estado de Mato Grosso (UNEMAT), instituição pública estadual multicampi, totalizando 5 cursos reais, 3 cenários sintéticos, 21 execuções controladas e 39 execuções retrospectivas. Embora a seleção tenha contemplado diversidade de graus acadêmicos (licenciaturas, bacharelados e tecnólogo), modalidades (presencial) e faixas de Conceito de Curso (3,33 a 4,44), a ausência de cursos com CC nas faixas extremas da escala (1, 2 e 5) e de cursos de instituições privadas limita a generalização dos resultados. As métricas reportadas na dissertação são descritivas (médias, desvios-padrão, taxas); nenhum teste de significância estatística foi aplicado nem intervalos de confiança calculados, dado que o tamanho amostral não é suficiente para inferência estatística robusta. A generalização dos resultados para o universo das IES brasileiras requer validação complementar com amostras maiores e mais diversificadas.

A terceira limitação concerne ao desempenho do *framework* no reconhecimento e na orientação de evidências documentais (MCE). Conforme documentado na seção 6.5, a Taxa de Adequação de Orientação (TAO) permaneceu abaixo do limiar de aceitação em todos os cenários, e a estabilidade do componente MCE na consistência interna foi de apenas 0,2749. Essa limitação restringe a confiabilidade das orientações sobre compartilhamento de evidências entre cursos, componente que constitui funcionalidade diferenciadora da MCE.

A quarta limitação, de importância central, refere-se à ausência de validação por especialistas humanos. Os procedimentos de avaliação empregados nesta pesquisa mensuraram propriedades de conformidade funcional e estabilidade operacional por meio de métricas automatizadas; a classificação de relevância das recomendações foi conduzida exclusivamente por agente de IA auxiliar da mesma família de modelos utilizada pelo *framework* (Claude Sonnet 4.6). Essa abordagem constitui proxy experimental que mensura a coerência interna do sistema, mas não demonstra a utilidade percebida das recomendações por coordenadores de

curso, membros da CPA e gestores institucionais. A validação por especialistas humanos, que permitiria aferir se os diagnósticos e recomendações produzidos são efetivamente úteis para a prática de gestão educacional, constitui trabalho futuro prioritário, conforme preconizado por Venable, Pries-Heje e Baskerville (2016) para artefatos em estágio avançado de desenvolvimento no *Design Science Research*.

A quinta limitação relaciona-se à dependência de provedor externo de IA. O *framework* utiliza a API da Anthropic (Claude Sonnet 4.6) como provedor de IA generativa, o que implica dependência de serviço de terceiro para a funcionalidade analítica central do sistema. Alterações na disponibilidade, nos custos ou no comportamento dos modelos podem afetar o desempenho operacional do *framework* sem que haja controle por parte dos desenvolvedores.

A sexta limitação refere-se à versão dos instrumentos de avaliação. As matrizes MPC, MCA e MCE foram construídas a partir dos Instrumentos de Avaliação de Cursos de Graduação do INEP nas versões vigentes à época da pesquisa (2017). Eventuais atualizações nos instrumentos normativos exigirão revisão e atualização correspondente das matrizes, processo que, embora metodologicamente replicável, demanda esforço analítico significativo.

Por fim, a sétima limitação refere-se à reprodutibilidade da pesquisa. Os *prompts* dos agentes de Diagnóstico, Recomendação e Síntese, que determinam o comportamento dos modelos de IA e, conseqüentemente, a qualidade dos diagnósticos e recomendações produzidos, não foram publicados integralmente na dissertação. Essa decisão decorreu de considerações de propriedade intelectual associadas ao registro do *software* Notacinco junto ao INPI. A ausência dos *prompts* limita a capacidade de replicação dos resultados por outros pesquisadores e impede a avaliação independente de potenciais vieses introduzidos pelas instruções de classificação. Recomenda-se, como medida de mitigação, a publicação da estrutura e das instruções-chave de classificação como material suplementar, preservando os elementos proprietários, ou a disponibilização integral dos *prompts* em repositório institucional após a formalização do registro de propriedade intelectual.

7.3 Recomendações para Trabalhos Futuros e Alinhamento Estratégico

As limitações identificadas e os achados da pesquisa delineiam uma agenda de trabalhos futuros organizada em três eixos: aprimoramento do artefato, expansão do escopo de validação e evolução estratégica da plataforma.

No eixo de aprimoramento do artefato, a limitação mais premente é o desempenho MCE. Recomenda-se investigar estratégias de refinamento na engenharia de *prompts* dos agentes para melhorar a navegação *cross-curso* (ICC-N2) e a descoberta semântica via RAG (TDR), incluindo a experimentação com técnicas de *few-shot prompting* e instruções explícitas sobre os níveis de acesso da MCE. Adicionalmente, a adoção de *output* estruturado nativo (*structured output*) pode mitigar a fragilidade do *pipeline* de extração identificada nos cenários controlados e contribuir para a estabilidade das classificações MCE.

No eixo de expansão da validação, três frentes são prioritárias. Em primeiro lugar, a condução de estudo de percepção de utilidade com coordenadores de curso e membros da CPA, empregando instrumentos validados de avaliação de sistemas de informação (por exemplo, o *Technology Acceptance Model*) para mensurar a aceitação e a intenção de uso do *framework* por seus destinatários. Em segundo lugar, a validação em IES de perfis distintos, incluindo instituições privadas, universidades federais e centros universitários, para verificar a generalização dos resultados obtidos na UNEMAT. Em terceiro lugar, a avaliação longitudinal do impacto do *framework* sobre os resultados efetivos das avaliações *in loco*, comparando o Conceito de Curso de programas que utilizaram o sistema com aqueles que não o utilizaram, em ciclos avaliativos subsequentes.

No eixo de evolução estratégica da plataforma, o formulário de registro do *software* já antecipa direções relevantes: a expansão comercial no modelo SaaS B2B para atendimento a múltiplas IES; a exploração da base de conhecimento global como recurso de *benchmarking* anônimo entre instituições; a adaptação do motor de regras e das matrizes para avaliação de programas de pós-graduação *stricto sensu* (métricas da CAPES); e a integração com ecossistemas externos via servidor MCP (*Model Context Protocol*), permitindo que sistemas acadêmicos tradicionais consumam o motor de inteligência do Notacinco como serviço integrado.

A presente pesquisa demonstrou que a integração entre técnicas de Análise de Conteúdo, agentes de Inteligência Artificial generativa e arquitetura *web* orientada a serviço produz artefato capaz de apoiar efetivamente a gestão de cursos de graduação no contexto do SINAES. A Matriz de Progressão Conceitual, como contribuição metodológica original, evidencia que instrumentos normativos estruturados podem ser decompostos em bases de conhecimento operacionalizáveis por agentes inteligentes, abordagem cuja aplicabilidade extrapola o domínio da avaliação educacional e se estende a outros contextos regulatórios graduados. O caminho percorrido, da formulação do problema à validação do artefato, reafirma

a pertinência do *Design Science Research* como paradigma metodológico para pesquisas que produzem soluções tecnológicas com rigor acadêmico e relevância prática, em consonância com a vocação do mestrado profissional em Políticas Públicas e Gestão da Educação Superior.

REFERÊNCIAS

- ADRIAANS, P.; ZANTINGE, D. **Data Mining**. England: Addison Wesley Longman, 1996.
- AEROSPIKE. Gen AI 2.0 is here: why agentic AI runs on real-time data infrastructure. **Aerospike Blog**, 2025. Disponível em: <https://aerospike.com/blog/agentic-ai-explained/>. Acesso em: 12 jan. 2026.
- AFONSO, A. J. Para uma concetualização alternativa de accountability em educação. **Educação & Sociedade**, [s. l.], v. 33, n. 119, p. 471-484, 2012. Disponível em: <https://doi.org/10.1590/s0101-73302012000200008>. Acesso em: nov. 2025.
- AGÊNCIA BRASIL. **Sete a cada dez estudantes usam IA na rotina de estudos**. Brasília: Agência Brasil, 6 ago. 2024. Disponível em: <https://agenciabrasil.ebc.com.br/educacao/noticia/2024-08/sete-cada-dez-estudantes-usam-ia-na-rotina-de-estudos>. Acesso em: 12 jan. 2026.
- ALBASHRAWI, M. Generative AI for decision-making: a multidisciplinary perspective. **Journal of Innovation & Knowledge**, [s. l.], v. 10, n. 4, p. 100751, 2025. Disponível em: <https://doi.org/10.1016/j.jik.2025.100751>. Acesso em: mar. 2026.
- ALEXANDRU, A. *et al.* Enhanced education by using intelligent agents in multi-agent adaptive e-learning systems. **Studies in Informatics and Control**, [s. l.], v. 24, n. 1, p. 13-22, 2015.
- ALNAHHAS, N.; YOUSEF, D. GAI as a catalyst in national technology sovereignty: evaluating the influence of GAI on government policy. *In: ANNUAL INTERNATIONAL CONFERENCE ON DIGITAL GOVERNMENT RESEARCH*, 25., 2024, Taipei. **Proceedings** [...]. New York: ACM, 2024. p. 618-626. DOI: <https://doi.org/10.1145/3657054.3657126>. Disponível em: <https://dl.acm.org/doi/10.1145/3657054.3657126>. Acesso em: 11 mar. 2026.
- AMATRIAIN, X. **Prompt Design and Engineering**: introduction and advanced methods. [S. l.]: arXiv, 2024. Disponível em: <http://arxiv.org/abs/2401.14423>. Acesso em: nov. 2025.
- AMAZON WEB SERVICES (AWS). **What are AI Agents?** Agents in Artificial Intelligence explained. [S. l.]: AWS, 2025. Disponível em: <https://aws.amazon.com/what-is/ai-agents/>. Acesso em: jan. 2026.
- ANALYTICS VIDHYA. **Generative AI in Decision-Making**: potential, pitfalls and practical solutions. [S. l.]: Analytics Vidhya, 7 nov. 2024. Disponível em: <https://www.analyticsvidhya.com/blog/2024/11/generative-ai-in-decision-making/>. Acesso em: 12 jan. 2026.
- ANDRIOLA, W. B. Avaliação institucional na Universidade Federal do Ceará (UFC): organização de sistema de dados e indicadores de qualidade institucional. **Avaliação**, Sorocaba, v. 9, n. 4, p. 33-54, 2004a. Disponível em: <http://periodicos.uniso.br/ojs/index.php/avaliacao/article/view/1286/1277>. Acesso em: out. 2025.

ANDRIOLA, W. B. Desafios, necessidades e consequências da atuação das Comissões Próprias de Avaliação (CPA's), com vistas a implementar o Sistema Nacional de Avaliação da Educação Superior (SINAES): o caso da Universidade Federal do Ceará (UFC). *In: FÓRUM DE PRÓ-REITORES DE GRADUAÇÃO*, 21., 2004, Fortaleza. **Trabalho apresentado [...]**. Fortaleza: [s. n.], 2004b.

ANDRIOLA, W. B. Propostas estatais voltadas à avaliação do ensino superior brasileiro: breve retrospectiva histórica do período 1983-2008. **REICE. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación**, [s. l.], v. 6, n. 4, p. 127-148, 2008. Disponível em: <https://revistas.uam.es/reice/article/view/5434>. Acesso em: out. 2025.

ANDRIOLA, W. B. Planejamento estratégico e gestão universitária como atividades oriundas da autoavaliação de instituições de ensino superior (IES): o exemplo da Universidade Federal do Ceará (UFC). **Revista Iberoamericana de Evaluación Educativa**, [s. l.], v. 2, n. 2, p. 82, 2009. Disponível em: <https://doi.org/10.15366/rie2009.2.2.005>. Acesso em: dez. 2025.

ARTERO, A. O. **Inteligência artificial: teórica e prática**. São Paulo: Livraria da Física, 2009.

ASC OHIO STATE UNIVERSITY. **Agentic AI in Higher Education**. [S. l.]: ASC Office of Distance Education, 2025. Disponível em: <https://ascode.osu.edu/news/agentic-ai-higher-education>. Acesso em: 12 jan. 2026.

BAKER, R. S. J. D.; YACEF, K. The state of educational data mining in 2009: a review and future visions. **Journal of Educational Data Mining**, [s. l.], v. 1, n. 1, p. 3-17, 2009.

BANDI, A. *et al.* The rise of agentic AI: a review of definitions, frameworks, architectures, applications, evaluation metrics, and challenges. **Future Internet**, [s. l.], v. 17, n. 9, p. 404, 2025. Disponível em: <https://doi.org/10.3390/fi17090404>. Acesso em: fev. 2026.

BARDIN, L. **Análise de conteúdo**. Tradução de Luís Antero Reto e Augusto Pinheiro. São Paulo: Edições 70, 2016.

BARREYRO, G. B. A avaliação da educação superior em escala global: da acreditação aos rankings e os resultados de aprendizagem. **Avaliação: Revista da Avaliação da Educação Superior**, Campinas; Sorocaba, v. 23, n. 1, p. 5-22, mar. 2018. DOI: <https://doi.org/10.1590/s1414-40772018000100002>. Disponível em: <https://www.researchgate.net/publication/324301259>. Acesso em: 11 mar. 2026.

BASKERVILLE, R. *et al.* Design science research contributions: finding a balance between artifact and theory. **Journal of the Association for Information Systems**, [s. l.], v. 19, n. 5, p. 358, 2018. Disponível em: <https://doi.org/10.17705/1jais.00495>. Acesso em: out. 2025.

BAZERMAN, M. H. **Processo decisório: para cursos de Administração, Economia e MBAs**. Rio de Janeiro: Elsevier, 2014.

BELCIC, I.; STRYKER, C. **Agentes de IA em 2025: expectativas versus realidade**. [S. l.]: IBM Think, 2025. Disponível em: <https://www.ibm.com/br-pt/think/insights/ai-agents-2025-expectations-vs-reality>. Acesso em: 11 mar. 2026.

BITTENCOURT, H. R. *et al.* Uma análise da relação entre os conceitos ENADE e IDD. **Estudos em Avaliação Educacional**, São Paulo, v. 19, n. 40, p. 247-262, maio/ago. 2008.

Disponível em: <https://publicacoes.fcc.org.br/eae/article/view/2078>. Acesso em: jan. 2026.

BRAGA, A.; LUDERMIR, T.; CARVALHO, A. **Redes neurais artificiais: teorias e aplicações**. Rio de Janeiro: LTC, 2000.

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). **Instrumento de Avaliação de Cursos de Graduação: presencial e a distância: reconhecimento e renovação de reconhecimento**. Brasília, DF: Inep, 2017.

BRASIL. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). **Sinaes: Sistema Nacional de Avaliação da Educação Superior: da concepção à regulamentação**. 5. ed. rev. e ampl. Brasília, DF: Inep, 2009.

BRASIL. **Lei nº 9.131, de 24 de novembro de 1995**. Altera dispositivos da Lei nº 4.024, de 20 de dezembro de 1961, e dá outras providências [Cria o Conselho Nacional de Educação e o Exame Nacional de Cursos]. Brasília, DF: Presidência da República, [1995]. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/19131.htm. Acesso em: 11 mar. 2026.

BRASIL. **Lei nº 9.394, de 20 de dezembro de 1996**. Estabelece as diretrizes e bases da educação nacional. Brasília, DF: Presidência da República, [1996]. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/19394.htm. Acesso em: 11 mar. 2026.

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: Presidência da República, [2004]. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2004-2006/2004/lei/110.861.htm. Acesso em: 11 mar. 2026.

BRASIL. **Lei nº 13.709, de 14 de agosto de 2018**. Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília, DF: Diário Oficial da União, 15 ago. 2018.

BRASIL. Ministério da Ciência, Tecnologia e Inovações. **Portaria MCTI nº 4.617, de 6 de abril de 2021**. Altera a Portaria nº 3.522, de 4 de setembro de 2020, que dispõe sobre a Estrutura Regimental e o Quadro Demonstrativo dos Cargos em Comissão e das Funções de Confiança do Ministério da Ciência, Tecnologia e Inovações. Brasília, DF: MCTI, 2021. Disponível em: <https://www.in.gov.br/>. Acesso em: 11 mar. 2026.

BRASIL. Ministério da Educação. **Portaria nº 2.051, de 9 de julho de 2004**. Regulamenta os procedimentos de avaliação do Sistema Nacional de Avaliação da Educação Superior (SINAES), instituído pela Lei nº 10.861, de 14 de abril de 2004. Brasília, DF: MEC, 2004. Disponível em: <http://portal.mec.gov.br/>. Acesso em: 11 mar. 2026.

BRASIL. Ministério da Educação. Secretaria de Educação Superior. **Portaria SESu nº 1.050, de 28 de abril de 2003**. [Dispõe sobre a organização e o funcionamento das Comissões Próprias de Avaliação – CPA]. Brasília, DF: SESu, 2003. Disponível em: <http://portal.mec.gov.br/>. Acesso em: 11 mar. 2026.

CARVALHO, A.; MALAVASI, M. M. S. Avaliação e acreditação da educação superior na América Latina e Caribe. **Revista Internacional de Educação Superior**, Campinas, v. 6, p. e020043, 2020. DOI: <https://doi.org/10.20396/riesup.v6i0.8657782>. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/riesup/article/view/8657782>. Acesso em: 5 jan. 2026.

CARVALHO, H. A. de; OLIVEIRA, O. S. de; LIMA, I. A. de. Avaliação institucional em uma universidade pública brasileira multicâmpus: processos e desafios na qualificação da gestão. **Avaliação: Revista da Avaliação da Educação Superior**, Campinas, v. 23, n. 1, p. 217, 2018. Disponível em: <https://doi.org/10.1590/s1414-40772018000100012>. Acesso em: nov. 2025.

CONAES. **Diretrizes para a autoavaliação das Instituições de Educação Superior**. Brasília, DF: MEC/CONAES, 2004.

CONAES. **Resolução nº 01, de 17 de junho de 2010**. Normatiza o Núcleo Docente Estruturante (NDE) e dá outras providências. Brasília, DF: MEC/CONAES, 2010.

CRESWELL, J. W. **Projeto de pesquisa: métodos qualitativo, quantitativo e misto**. 3. ed. Porto Alegre: Artmed, 2010.

DAVENPORT, T. H.; HARRIS, J. G. **Competing on Analytics: the new science of winning**. [S. l.: s. n.], 2007. Disponível em: <http://www.loc.gov/catdir/toc/ecip073/2006035422.html>. Acesso em: nov. 2025.

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

DIAS SOBRINHO, J. Avaliação educativa: produção de sentido com valor ético. **Avaliação**, Campinas, v. 10, n. 4, 2008a.

DIAS SOBRINHO, J. Qualidade, avaliação: do SINAES a índices. **Avaliação: Revista da Avaliação da Educação Superior**, Campinas; Sorocaba, v. 13, n. 3, p. 817-825, nov. 2008b.

DOS-SANTOS-OLIVEIRA, I.; ROTHEN, J.-C. Autoavaliação institucional das Instituições de Ensino Superior no Brasil. **Revista Iberoamericana de Educación Superior**, [s. l.], p. 20, 2024. Disponível em: <https://doi.org/10.22201/iissue.20072872e.2024.44.1887>. Acesso em: out. 2025.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. **AI Magazine**, [s. l.], v. 17, n. 3, p. 37-54, 1996.

FELDMANN, T.; SOUZA, O. de; HEINZLE, M. R. S. As posições-sujeito (Estado e gestores) diante das avaliações em larga escala no Ensino Superior brasileiro. **Praxis Educativa**, [s. l.], v. 11, n. 3, p. 605, 2016. Disponível em: <https://doi.org/10.5212/praxeduc.v.11i3.0005>. Acesso em: set. 2025.

FESTA, M. **Autoavaliação institucional participativa da educação infantil da cidade de São Paulo**. 2019. Tese (Doutorado em Educação) – Faculdade de Educação, Universidade de São Paulo, São Paulo, 2019. Disponível em: <https://doi.org/10.11606/t.48.2019.tde-01082019-154513>. Acesso em: set. 2025.

FOWLER, M. **Patterns of Enterprise Application Architecture**. Boston: Addison-Wesley, 2002. Disponível em: https://sar.ac.id/stmik_ebook/prog_file_file/EFCofwzsj0.pdf. Acesso em: 12 mar. 2026.

FRANKLIN, S.; GRAESSER, A. Is it an agent, or just a program? A taxonomy for autonomous agents. *In*: MÜLLER, J. P.; WOOLDRIDGE, M. J.; JENNINGS, N. R. (ed.).

Intelligent Agents III. ATAL 1996. Berlin: Springer, 1997. p. 21-35. (Lecture Notes in Computer Science, v. 1193).

GAO, Y. *et al.* **Retrieval-Augmented Generation for Large Language Models: a survey**. [S. l.]: arXiv, 2023. Disponível em: <https://arxiv.org/abs/2312.10997>. Acesso em: mar. 2026.

GENRE. **Decision-Making in the Age of Generative Artificial Intelligence**. [S. l.]: Genre, 2025. Disponível em: <https://www.genre.com/us/knowledge/publications/2025/april/decision-making-in-the-age-of-generative-artificial-intelligence-en>. Acesso em: mar. 2026.

GERHARDT, T. E.; SILVEIRA, D. T. **Métodos de pesquisa**. Porto Alegre: Editora da UFRGS, 2009.

GIL, A. C. **Como elaborar projetos de pesquisa**. 7. ed. São Paulo: Atlas, 2022.

GÓMEZ, L. B. Estado regulador = Regulatory state. **EUNOMÍA. Revista en Cultura de la Legalidad**, [s. l.], p. 249, 2019. Disponível em: <https://doi.org/10.20318/eunomia.2019.5030>. Acesso em: nov. 2025.

GOOGLE CLOUD. **What are AI agents?** Definition, examples, and types. [S. l.]: Google Cloud, 2025. Disponível em: <https://cloud.google.com/discover/what-are-ai-agents>. Acesso em: 12 jan. 2026.

GREGOR, S.; HEVNER, A. R. Positioning and presenting design science research for maximum impact. **MIS Quarterly**, [s. l.], v. 37, n. 2, p. 337-355, jun. 2013. Disponível em: <https://doi.org/10.25300/MISQ/2013/37.2.01>. Acesso em: 11 mar. 2026.

GRIBOSKI, C. M. O SINAES e a avaliação institucional: dez anos de concepção e implementação. **Estudos em Avaliação Educacional**, São Paulo, v. 23, n. 53, p. 166-184, set./dez. 2012. Disponível em: <http://educa.fcc.org.br/pdf/eae/v23n53/v23n53a09.pdf>. Acesso em: 11 mar. 2026.

GRIBOSKI, C. M.; FUNGHETTO, S. S. O Sinaes e a qualidade da educação. **Retratos da Escola**, [s. l.], v. 7, n. 12, p. 49, 2013. Disponível em: <http://esforce.org.br/index.php/semestral/article/view/259/0>. Acesso em: out. 2025.

GRIBOSKI, C. M.; PEIXOTO, M. do C. de L.; HORA, P. M. da. Avaliação externa, autoavaliação e o PDI. **Avaliação: Revista da Avaliação da Educação Superior**, Campinas; Sorocaba, v. 23, n. 1, p. 178-193, mar. 2018. DOI: <https://doi.org/10.1590/s1414-40772018000100010>. Disponível em: <https://www.researchgate.net/publication/324309385>. Acesso em: 11 mar. 2026.

GRIBOSKI, C. M.; MENDES, D.; DAMASCENO, L. L. Avaliação institucional: articulações possíveis entre Comissão Própria de Avaliação, Núcleo Docente Estruturante e Projeto Pedagógico de Curso. **Pesquisa em Foco**, São Luís, v. 29, n. 1, jun. 2024. DOI: <https://doi.org/10.18817/pef.v29i1.3743>. Disponível em: https://ppg.revistas.uema.br/index.php/PESQUISA_EM_FOCO/article/view/3743. Acesso em: 11 mar. 2026.

GU, X.; BLACKMORE, K. L. A systematic review of agent-based modelling and simulation applications in the higher education domain. **Higher Education Research & Development**,

[s. l.], v. 34, n. 5, p. 883-898, 2015.

HAAS, C. M.; APARÍCIO, A. S. M. Avaliação, regulação e qualidade na educação superior: os desafios da gestão acadêmica. **EccoS – Revista Científica**, [s. l.], n. 51, 2019. Disponível em: <https://doi.org/10.5585/eccos.n51.15825>. Acesso em: out. 2025.

HAN, J.; KAMBER, M.; PEI, J. **Data Mining: concepts and techniques**. 3. ed. [S. l.]: Elsevier, 2011.

HEVNER, A. R. *et al.* Design science in information systems research. **MIS Quarterly**, [s. l.], v. 28, n. 1, p. 75-105, 2004. Disponível em: <http://community.mis.temple.edu/seminars/files/2009/10/hevner-et-al-misq-2004.pdf>. Acesso em: fev. 2026.

IKUTA, C. Y. S. Sobre o Conceito Preliminar de Curso: concepção, aplicação e mudanças metodológicas. **Estudos em Avaliação Educacional**, [s. l.], v. 27, n. 66, p. 938, 2016. Disponível em: <https://doi.org/10.18222/eae.v27i66.4039>. Acesso em: out. 2025.

JENNINGS, N. R. *et al.* A roadmap of agent research and development. **Autonomous Agents and Multi-Agent Systems**, [s. l.], v. 1, n. 1, p. 7-38, 1998.

JIANG, Y.-H. *et al.* AI agent for education: von Neumann multi-agent system framework. In: GLOBAL CHINESE CONFERENCE ON COMPUTERS IN EDUCATION (GCCCE), 28., 2024, Chongqing. **Proceedings [...]**. Chongqing: GCCCE, 2024. p. 77-84.

KAHNEMAN, D.; TVERSKY, A. Prospect theory: an analysis of decision under risk. **Econometrica**, [s. l.], v. 47, n. 2, p. 263, 1979. Disponível em: <https://doi.org/10.2307/1914185>. Acesso em: mar. 2026.

KNIGHT, F. H. **Risk, Uncertainty and Profit**. [S. l.: s. n.], 1921. Disponível em: <https://ideas.repec.org/b/vpr/ecbook/14.html>. Acesso em: mar. 2026.

KRIPPENDORFF, K. **Content Analysis: an introduction to its methodology**. Thousand Oaks, CA: SAGE Publications, 2018. Disponível em: <https://doi.org/10.4135/9781071878781>. Acesso em: 19 mar. 2026.

LAKATOS, E. M.; MARCONI, M. de A. **Fundamentos de metodologia científica**. São Paulo: Atlas, 2017.

LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. **Biometrics**, [s. l.], v. 33, n. 1, p. 159-174, 1977.

LEITE, D. Avaliação institucional e o trabalho docente: perspectiva emancipatória. **Revista de Gestão e Avaliação Educacional**, [s. l.], v. 1, n. 1, p. 105, 2017. Disponível em: <https://doi.org/10.5902/2318133829222>. Acesso em: fev. 2025.

LEITNER-HANETSEDER, S.; LEHNER, O. M. AI-powered information and Big Data: current regulations and ways forward in IFRS reporting. **Journal of Applied Accounting Research**, [s. l.], v. 24, n. 2, p. 282-298, 2022. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/JAAR-01-2022-0022/full/html>. Acesso em: 11 mar. 2025.

LEWIS, P. *et al.* Retrieval-Augmented Generation for knowledge-intensive NLP tasks. *In: CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS (NeurIPS)*, 34., 2020, Vancouver. **Proceedings** [...]. Vancouver: NeurIPS, 2020. Disponível em: <https://arxiv.org/abs/2005.11401>. Acesso em: mar. 2026.

LÓPEZ, Ó. *et al.* Effect of generative artificial intelligence on strategic decision-making in entrepreneurial business initiatives: a systematic literature review. **Administrative Sciences**, [s. l.], v. 15, n. 2, p. 66, 2025. Disponível em: <https://doi.org/10.3390/admsci15020066>. Acesso em: mar. 2026.

LORDÊLO, J. A. C.; DAZZANI, M. V. M. **Avaliação educacional: desatando e reatando nós**. Salvador: EDUFBA, 2009. Disponível em: <https://doi.org/10.7476/9788523209315>. Acesso em: out. 2025.

MAÇÃES, M. A. R. **Planejamento, estratégia e tomada de decisão**: vol. IV. São Paulo: Almedina Brasil, 2018. E-book. Disponível em: <https://integrada.minhabiblioteca.com.br/reader/books/9789896942274/>. Acesso em: 20 jan. 2026.

MARCHESI, T. de S. F. **A qualidade da educação superior nos instrumentos normativos sobre o SINAES**. 2016. 122 f. Dissertação (Mestrado em Educação) – Programa de Pós-Graduação em Educação, Universidade Federal do Triângulo Mineiro, Uberaba, 2016.

MARQUEZAN, F. F.; SCREMIN, G. Planejamento e autoavaliação institucional: articulações possíveis e necessárias. **Revista Espaço Pedagógico**, Passo Fundo, v. 26, n. 1, p. 139-157, 2018. Disponível em: <https://doi.org/10.5335/rep.v26i1.8356>. Acesso em: abr. 2026.

MARTIN, R. C. **Clean Architecture: a craftsman's guide to software structure and design**. New York: Prentice Hall, 2017.

MATO GROSSO. Conselho Estadual de Educação. **Resolução nº 311, de 15 de julho de 2008**. Fixa roteiro a ser observado na verificação “*in loco*” com vistas ao credenciamento, recredenciamento de instituições de Ensino Superior, à autorização e ao reconhecimento de cursos/habilitações de Ensino Superior vinculadas ao Sistema Estadual de Ensino. Cuiabá: CEE/MT, 2008. Disponível em: <https://normativasconselhos.mec.gov.br/?fonte=CEE-MT>. Acesso em: 16 abr. 2026.

MATO GROSSO. Conselho Estadual de Educação. **Resolução Normativa nº 001, de 6 de abril de 2011**. Dispõe sobre diretrizes para elaboração de Instrumentos de Avaliação a serem utilizados em processos de regulação das Instituições de Ensino Superior – IES e de seus cursos de graduação presenciais e a distância. Cuiabá: CEE/MT, 2011. Disponível em: <https://normativasconselhos.mec.gov.br/?fonte=CEE-MT>. Acesso em: 16 abr. 2026.

MATO GROSSO. Conselho Estadual de Educação. **Resolução nº 01/2017-CEE/MT**. Normatiza o exercício das funções de regulação, supervisão e avaliação das Instituições de Educação Superior e dos cursos superiores de graduação vinculados ao Sistema Estadual de Ensino do Estado de Mato Grosso. Cuiabá: CEE/MT, 2017.

MCKINSEY & COMPANY. **The State of AI in Early 2024: gen AI adoption spikes and starts to generate value**. [S. l.]: McKinsey Global Survey, 2024. Disponível em: <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-2024>.

Acesso em: ago. 2025.

MINTZBERG, H.; RAISINGHANI, D.; THEORET, A. The structure of “unstructured” decision processes. **Administrative Science Quarterly**, [s. l.], v. 21, n. 2, p. 246, 1976. Disponível em: <https://doi.org/10.2307/2392045>. Acesso em: mar. 2026.

MINTZBERG, H. Patterns in strategy formation. **Management Science**, [s. l.], v. 24, n. 9, p. 934-948, maio 1978. DOI: <https://doi.org/10.1287/mnsc.24.9.934>. Disponível em: <https://www.researchgate.net/publication/227444830>. Acesso em: 11 mar. 2026.

MITCHELL, T. M. Does machine learning really work? **AI Magazine**, [s. l.], v. 18, n. 3, p. 11, 15 set. 1997. Disponível em: <https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/1303>. Acesso em: 11 mar. 2026.

MORETTI, D. M. **A compatibilidade entre a lógica econômica e o ensino superior, após a Constituição Federal de 1988: o caso da Anhanguera Educacional Participações S.A.** 2013. Dissertação (Mestrado em Direito) – Faculdade de Direito, Universidade de São Paulo, São Paulo, 2013.

NASCIMENTO, A. D.; HETKOWSKI, T. M. (org.). **Educação e contemporaneidade: pesquisas científicas e tecnológicas.** Salvador: EDUFBA, 2009. 400 p. ISBN 978-85-232-0872-1. Disponível em: <http://books.scielo.org>. Acesso em: abr. 2026.

O’DAY, J. Complexity, accountability, and school improvement. **Harvard Educational Review**, Cambridge, v. 72, n. 3, p. 293-329, 2002.

OLIVEIRA, I. dos S.; ROTHEN, J. C. Vinte anos do Sistema Nacional de Avaliação da Educação Superior – SINAES no Brasil: trajetória, princípios, dilemas e tendências. **Avaliação: Revista da Avaliação da Educação Superior**, Campinas, v. 29, 2024. Disponível em: <https://doi.org/10.1590/1982-57652024v29id278360>. Acesso em: dez. 2025.

OPENAI. **GPT-4 Technical Report**. [S. l.]: arXiv, 2023. Disponível em: <http://arxiv.org/abs/2303.08774>. Acesso em: mar. 2026.

PAULA, A. S. do N. de; COSTA, F. J. F.; LIMA, K. R. R. Accountability e o mosaico da política de avaliação da educação superior brasileira. **Revista de Gestão e Avaliação Educacional**, [s. l.], v. 1, n. 1, p. 1, 2019. Disponível em: <https://doi.org/10.5902/2318133834902>. Acesso em: mar. 2025.

PEFFERS, K. *et al.* A design science research methodology for information systems research. **Journal of Management Information Systems**, [s. l.], v. 24, n. 3, p. 45-77, 2007.

PEREIRA, P. R.; GASQUE, K. C. G. D. Tomada de decisão do gestor escolar das escolas públicas de ensino médio no Distrito Federal e a interface com o letramento informacional. **Em Questão**, [s. l.], p. 40, 2019. Disponível em: <https://doi.org/10.19132/1808-5245253.40-60>. Acesso em: out. 2025.

PLANELLAS, M.; MUNI, A. **Strategic Decisions: the circle of strategic decisions.** Cambridge: Cambridge University Press, 2019. E-book. Disponível em: <https://www.semanticscholar.org/paper/Strategic-Decisions-Planellas->

Muni/3f8528c2afaffcccc69dc0daa12250282595105c. Acesso em: 11 mar. 2025.

POLIDORI, M. M.; FONSECA, D. G. da; LARROSA, S. F. T. Avaliação institucional participativa. **Avaliação: Revista da Avaliação da Educação Superior**, Campinas, v. 12, n. 2, p. 333, 2007. Disponível em: <https://doi.org/10.1590/s1414-40772007000200008>. Acesso em: nov. 2025.

POLIDORI, M. M. Políticas de avaliação da educação superior brasileira: Provão, SINAES, IDD, CPC, IGC e... outros índices. **Avaliação: Revista da Avaliação da Educação Superior**, Campinas; Sorocaba, v. 14, n. 2, p. 439-452, jul. 2009. Disponível em: <https://doi.org/10.1590/S1414-40772009000200009>. Acesso em: 11 mar. 2026.

RADONJIĆ, A.; DUARTE, H.; PEREIRA, N. Artificial intelligence and HRM: HR managers' perspective on decisiveness and challenges. **European Management Journal**, [s. l.], v. 42, n. 1, p. 57-66, fev. 2024. DOI: <https://doi.org/10.1016/j.emj.2022.07.001>. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0263237322000883>. Acesso em: 11 mar. 2026.

RIBEIRO, W. L.; GUERRA, M. das G. G. V. **Avaliação de cursos a partir do SINAES: uma análise para melhoria da qualidade da Universidade Federal da Paraíba**. João Pessoa: UFPB, 2019.

ROMERO, C.; VENTURA, S. Educational data mining and learning analytics: an updated survey. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, [s. l.], v. 10, n. 3, e1355, 2020. Disponível em: <https://doi.org/10.1002/widm.1355>. Acesso em: abr. 2026.

ROTTA, H. A.; SILVA, A. R.; COSTA, C. C. A estratégia brasileira de inteligência artificial na gestão universitária. **Caderno Pedagógico**, [S. l.], v. 22, n. 8, p. e17393, 2025. DOI: <https://doi.org/10.54033/cadpedv22n8-201>. Disponível em: <https://ojs.studiespublicacoes.com.br/ojs/index.php/cadped/article/view/17393>. Acesso em: 12 mar. 2026.

RUSSELL, S. J.; NORVIG, P. **Inteligência artificial: uma abordagem moderna**. 3. ed. Rio de Janeiro: Elsevier, 2013. 988 p.

RUSSELL, S. J.; NORVIG, P. **Artificial Intelligence: a modern approach**. 4. ed. Hoboken: Pearson, 2021.

SANCHES, R. C. F.; RAPHAEL, H. S. Projeto pedagógico e avaliação institucional: articulação e importância. **Avaliação: Revista da Avaliação da Educação Superior**, Sorocaba, v. 11, n. 1, 2006. Disponível em: <https://periodicos.uniso.br/avaliacao/article/view/923>. Acesso em: 11 mar. 2026.

SANCHES, R. C. F. **Avaliação institucional e projeto pedagógico: articulação imprescindível**. Americana: AECID Library, 2007. Disponível em: <http://hdl.handle.net/11449/104790>. Acesso em: out. 2025.

SANTANA, L. D.; PEREIRA, F. C. M.; MATTOS, M. C. de. Design science research na prática. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, [s. l.], v. 21, 2023. Disponível em: <https://doi.org/10.20396/rdbci.v21i00.8672896>. Acesso em: out. 2025.

SILVA, I. F. B.; GHIRALDELLI, R. A avaliação do ensino superior no Brasil a partir do Exame Nacional de Desempenho de Estudantes (ENADE) e o Serviço Social. **Serviço Social em Revista**, [s. l.], v. 25, n. 1, p. 148, 2022. Disponível em: <https://doi.org/10.5433/1679-4842.2022v25n1p148>. Acesso em: ago. 2025.

SILVA, L. R. A. B. D. Inteligência artificial em processos de extração de conhecimento KDD e KDT. **Revista de Estudos Universitários – REU**, [s. l.], v. 46, n. 1, p. 161, 2020. Disponível em: <https://doi.org/10.22484/2177-5788.2020v46n1p161-180>. Acesso em: nov. 2025.

SIMON, H. A. A behavioral model of rational choice. **The Quarterly Journal of Economics**, [s. l.], v. 69, n. 1, p. 99-118, fev. 1955. Disponível em: https://cooperative-individualism.org/simon-herbert_a-behavioral-model-of-rational-choice-1955-feb.pdf. Acesso em: 11 mar. 2026.

SIMON, H. A. Rational choice and the structure of the environment. **Psychological Review**, [s. l.], v. 63, n. 2, p. 129, 1956. Disponível em: <https://doi.org/10.1037/h0042769>. Acesso em: mar. 2026.

SIMON, H. A. **Administrative Behavior**: a study of decision-making processes in administrative organizations. 3. ed. New York: Free Press, 1976.

SIMON, H. A. **A racionalidade do processo decisório em empresas**. Rio de Janeiro: Edições Multiplic, v. 1, n. 1, p. 25-57, 1980.

SIMON, H. A. **The Sciences of the Artificial**. 3. ed. Cambridge, MA: The MIT Press, 1996.

SOKOLOVA, M.; LAPALME, G. A systematic analysis of performance measures for classification tasks. **Information Processing & Management**, [s. l.], v. 45, n. 4, p. 427-437, 2009.

SONNENBERG, C.; BROCKE, J. vom. Evaluation patterns for design science research artefacts. *In*: COMMUNICATIONS IN COMPUTER AND INFORMATION SCIENCE. [S. l.]: Springer, 2012. p. 71. Disponível em: https://doi.org/10.1007/978-3-642-33681-2_7. Acesso em: nov. 2025.

SORDI, M. R. L. de; LUDKE, M. Da avaliação da aprendizagem à avaliação institucional: aprendizagens necessárias. **Avaliação: Revista da Avaliação da Educação Superior**, Campinas; Sorocaba, v. 14, n. 2, p. 313-336, jul. 2009.

SOUZA, J. R. M. A. de. **Avaliação educacional na dimensão do SINAES**: qualidade do ensino ou requisito legal na Universidade Federal do Ceará. 2020. 278 f. Tese (Doutorado em Educação) – Faculdade de Educação, Universidade Federal do Ceará, Fortaleza, 2020. Disponível em: <https://doi.org/10.5902/2318133842336>. Acesso em: out. 2025.

SOUZA, S. R. A. de; GUERRA, M. das G. G. V. **Análise do papel da autoavaliação institucional no processo de melhoria da qualidade na Universidade Federal da Paraíba**. João Pessoa: UFPB, 2019.

SOUZA, S. R. A. de; GUERRA, M. das G. G. V. Avaliação da educação superior no Brasil. **Revista de Gestão e Avaliação Educacional**, [s. l.], p. 1, 2020. Disponível em:

<https://doi.org/10.5902/2318133842336>. Acesso em: out. 2025.

TOZZI, M. V.; TOZZI, A. R. Conceito preliminar de curso (CPC) x conceito do curso (CC): há coerência desses valores nos cursos de engenharia? **Revista Principia: Divulgação Científica e Tecnológica do IFPB**, João Pessoa, n. 34, p. 18-23, maio 2017. Disponível em: <https://doi.org/10.18265/1517-03062015v1n34p18-23>. Acesso em: nov. 2025.

TVERSKY, A.; KAHNEMAN, D. Judgment under uncertainty: heuristics and biases. **Science**, [s. l.], v. 185, n. 4157, p. 1124, 1974. Disponível em: <https://doi.org/10.1126/science.185.4157.1124>. Acesso em: mar. 2026.

UNEMAT. **Plano de Desenvolvimento Institucional 2022-2028**. Cáceres: UNEMAT, 2022.

VENABLE, J.; PRIES-HEJE, J.; BASKERVILLE, R. FEDS: a framework for evaluation in design science research. **European Journal of Information Systems**, [s. l.], v. 25, n. 1, p. 77-89, 2016.

WANG, S. *et al.* Artificial intelligence in education: a systematic literature review. **Expert Systems with Applications**, [s. l.], v. 252, article 124167, 2024. Disponível em: <https://doi.org/10.1016/j.eswa.2024.124167>. Acesso em: abr. 2026.

WEI, J. *et al.* **Chain-of-Thought Prompting Elicits Reasoning in Large Language Models**. [S. l.]: arXiv, 2022. Disponível em: <https://arxiv.org/pdf/2201.11903>. Acesso em: 11 mar. 2026.

WOOLDRIDGE, M. **An Introduction to MultiAgent Systems**. 2. ed. Chichester: Wiley, 2009.

ZHAO, W. X. *et al.* **A Survey of Large Language Models**. [S. l.]: arXiv, 2023. Disponível em: <https://arxiv.org/pdf/2303.18223>. Acesso em: 11 mar. 2026.

ZIMMERMANN, M. M. de S.; ALVES, L. Necessidade de meta-avaliação para a autoavaliação institucional: CPA do Senac/SC. **Revista de Gestão e Avaliação Educacional**, [s. l.], v. 10, n. 19, 2021. Disponível em: <https://doi.org/10.5902/2318133865887>. Acesso em: maio 2025.