

UNIVERSIDADE FEDERAL DO TRIÂNGULO MINEIRO - UFTM



AV. DR. RANDOLFO BORGES, 1400 UNIVERDECIDADE - 38064-200 - UBERABA/MG

MESTRADO PROFISSIONAL EM MATEMÁTICA EM REDE NACIONAL - PROFMAT



PROFMAT

DISSERTAÇÃO DE MESTRADO

UTILIZAÇÃO DE IA NA EXTRAÇÃO DE HABILIDADES DA
BNCC EM MATERIAIS DIDÁTICOS DE MATEMÁTICA NO
ENSINO FUNDAMENTAL II: UMA EXPERIÊNCIA BASEADA EM
ENGENHARIA DE PROMPTS

BRUNA CRISTINA MORAES ARAÚJO GONÇALVES

UBERABA-MG

2026

BRUNA CRISTINA MORAES ARAÚJO GONÇALVES

UTILIZAÇÃO DE IA NA EXTRAÇÃO DE HABILIDADES DA
BNCC EM MATERIAIS DIDÁTICOS DE MATEMÁTICA NO
ENSINO FUNDAMENTAL II: UMA EXPERIÊNCIA BASEADA EM
ENGENHARIA DE PROMPTS

Dissertação apresentada ao Curso de Pós-Graduação em Mestrado Profissional em Matemática, área de concentração “Matemática”, da Universidade Federal do Triângulo Mineiro, como requisito parcial para obtenção do título de mestre em Matemática.

Orientador: Prof. Dr. Leandro Cruvinel Lemes

Uberaba-MG

2026

**Catálogo na fonte: Biblioteca da Universidade Federal do
Triângulo Mineiro**

G624u Gonçalves, Bruna Cristina Moraes Araújo
Utilização de IA na extração de habilidades da BNCC em materiais didáticos de matemática no ensino fundamental II: uma experiência baseada em engenharia de prompts / Bruna Cristina Moraes Araújo Gonçalves. -- 2026.
78 p.: il., fig., graf., tab.

Dissertação (Mestrado Profissional em Matemática em Rede Nacional)
-- Universidade Federal do Triângulo Mineiro, Uberaba, MG, 2026
Orientador: Prof. Dr. Leandro Cruvinel Lemes

1. Matemática - Estudo e ensino. 2. Base Nacional Comum Curricular. 3. Processamento de linguagem natural (Computação). 4. Inteligência artificial.
I. Lemes, Leandro Cruvinel. II. Universidade Federal do Triângulo Mineiro.
III. Título.

CDU 51:004.8

BRUNA CRISTINA MORAES ARAÚJO GONÇALVES

Utilização de IA na extração de habilidades da BNCC em materiais didáticos de Matemática no Ensino Fundamental II: uma experiência baseada em engenharia de prompts

Dissertação apresentada ao Programa de Pós-Graduação em Mestrado Profissional em Matemática, área de concentração Matemática da Universidade Federal do Triângulo Mineiro como requisito parcial para obtenção do título de mestre.

Uberaba, 23 de janeiro de 2026.

Banca Examinadora:

Dr. Leandro Cruvinel Lemes – Orientador
Universidade Federal do Triângulo Mineiro

Dr. Daniel Fernando Bovolenta Ovigli
Universidade Federal do Triângulo Mineiro

Dra. Hellen Monção de Carvalho Santana
Universidade Federal de São Carlos



Documento assinado eletronicamente por **LEANDRO CRUVINEL LEMES, Professor do Magistério Superior**, em 23/02/2026, às 14:17, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#) e no art. 34 da [Portaria Reitoria/UFTM nº 215, de 16 de julho de 2024](#).



Documento assinado eletronicamente por **DANIEL FERNANDO BOVOLENTA OVIGLI, Professor do Magistério Superior**, em 23/02/2026, às 17:01, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#) e no art. 34 da [Portaria Reitoria/UFTM nº 215, de 16 de julho de 2024](#).



Documento assinado eletronicamente por **Hellen Monção de Carvalho Santana, Usuário Externo**, em 24/02/2026, às 10:50, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#) e no art. 34 da [Portaria Reitoria/UFTM nº 215, de 16 de julho de 2024](#).



A autenticidade deste documento pode ser conferida no site http://sei.ufmt.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **1710902** e o código CRC **200D351F**.

*À Penélope, minha filha, por me tornar mais forte e resiliente.
Ao Igor, meu esposo, por me apoiar e sempre acreditar que eu conseguiria.*

AGRADECIMENTOS

O percurso do Mestrado Profissional, embora desafiador, foi uma jornada de profundo crescimento, possível apenas com o apoio e a colaboração de muitas pessoas e instituições. A elas manifesto minha eterna gratidão.

Agradeço, primeiramente, a **Deus** pela saúde, pela força e pelas oportunidades que guiaram meus passos até a conclusão desta etapa.

À **Universidade Federal do Triângulo Mineiro (UFTM)** por oferecer ambiente de excelência, professores inspiradores e infraestrutura necessários para o meu desenvolvimento.

Ao **Mestrado Profissional em Matemática em Rede Nacional (PROFMAT)** pela proposta de qualificação e aperfeiçoamento profissional que tanto contribuiu para minha formação como educadora.

Ao Professor **Leandro Cruvinel**, meu orientador, a quem sou grata pela confiança e pela fundamental orientação científica. Sua paciência, seu rigor metodológico e seu incentivo constante foram o alicerce para o sucesso desta dissertação.

Aos **professores do Profmat UFTM** que generosamente compartilharam seus conhecimentos e experiências em cada disciplina: Adriana de Campos Inforzato, Osmar Aléssio, Leandro Cruvinel Lemes, Flávio Molina da Silva, Heron Martins Félix, Rafael Peixoto, Nelson Fernando Inforzato, Marcelo Ferreira, Fábio Antônio Araújo de Campos, Bruno Nunes de Souza, Rafael Rodrigo Ottoboni e Marcela Luciano Vilela de Souza. Cada aula foi fundamental para a minha formação pessoal e profissional. Sinto muito orgulho por ter conhecido cada um de vocês e ter compartilhado essa caminhada juntos! Muito obrigada!

Aos **membros da Banca Examinadora** que dedicaram seu tempo à leitura e avaliação da dissertação, contribuindo com críticas construtivas que, indubitavelmente, aprimoraram este texto.

À Penélope, minha filha, e ao Igor, meu esposo, que são minha base e meu maior incentivo. A compreensão nos momentos de dedicação integral ao estudo, o suporte incondicional e o amor demonstrado foram a força motriz para superar todos os obstáculos e alcançar este objetivo.

A todos que, de forma direta ou indireta, contribuíram para a realização deste sonho.

Muito obrigada!!

*Os professores ideais são os que se fazem de pontes,
que convidam os alunos a atravessarem,
e depois, tendo facilitado a travessia,
desmoram-se com prazer,
encorajando-os a criarem as suas próprias pontes.
Nikos Kazantzakis - Escritor grego*

RESUMO

Esta pesquisa investiga, de forma comparativa, a eficácia de diferentes estratégias de formulação de prompts de Inteligência Artificial Generativa na identificação e extração de habilidades da Base Nacional Comum Curricular (BNCC) em materiais didáticos de Matemática para os Anos Finais do Ensino Fundamental. Foram avaliados quatro modelos de IA: Google Gemini 3 Pro, Google Gemini 2.5 Flash, Chat GPT-5 e Claude AI Sonnet 4.5. A metodologia baseou-se em métricas de recuperação de informação (Precisão, Recall, F_1 -Macro e Coeficiente de Jaccard), utilizando quatro estratégias distintas de prompt aplicadas a cinco materiais didáticos selecionados e comparando os resultados gerados pela Inteligência Artificial com um gabarito de referência construído manualmente pela pesquisadora e mais 5 professores de Matemática dos anos finais do Ensino Fundamental. A pesquisa evidencia a viabilidade técnica da aplicação de Inteligência Artificial (IA) generativa no alinhamento curricular, embora requeira supervisão docente qualificada. Os achados contribuem para o desenvolvimento de ferramentas de apoio à curadoria de recursos educacionais e para a discussão sobre letramento em IA na formação de professores.

Palavras-chave: Inteligência Artificial Generativa; Engenharia de Prompts; Base Nacional Comum Curricular; Ensino da Matemática; Processamento de Linguagem Natural.

ABSTRACT

This research comparatively investigates the effectiveness of different Generative Artificial Intelligence prompt formulation strategies in identifying and extracting skills from the Common National Curricular Base (BNCC) in Mathematics teaching materials for the Final Years of Elementary School. Four AI models were evaluated: Google Gemini 3 Pro, Google Gemini 2.5 Flash, Chat GPT-5, and Claude AI Sonnet 4.5. The methodology was based on information retrieval metrics (Precision, Recall, F_1 -Macro, and Jaccard Coefficient), using four distinct prompt strategies applied to five selected teaching materials and comparing the results generated by AI with a reference answer key manually constructed by the researcher and five additional Mathematics teachers from the final years of Elementary School. The research demonstrates the technical feasibility of applying generative Artificial Intelligence (AI) in curricular alignment, although it requires qualified teacher supervision. The findings contribute to the development of support tools for the curation of educational resources and to the discussion about AI literacy in teacher training.

Keywords: Generative Artificial Intelligence; Prompt Engineering; National Common Curricular Base; Teaching Mathematics; Natural Language Processing.

Lista de Figuras

3.1	Exemplo da complexidade multimodal do material analisado (Frações.PNG) . . .	21
4.1	Comparação do desempenho médio entre as estratégias de prompt considerando todos os modelos de IA e materiais didáticos	37
4.2	Evolução temporal das métricas de desempenho ao longo das quatro estratégias de prompt	38
4.3	Comparação do desempenho médio entre os modelos de IA considerando todas as estratégias de prompt e materiais didáticos	39
4.4	Desempenho da melhor combinação (Gemini 3 Pro com Prompt 4) por material didático nas quatro métricas avaliadas	40

Lista de Tabelas

3.1	Caracterização dos materiais didáticos utilizados no corpus da pesquisa	20
4.1	Resultados da análise com Prompt 1	32
4.2	Resultados da análise com Prompt 2	33
4.3	Resultados da análise com Prompt 3	34
4.4	Resultados da análise com Prompt 4	35
4.5	Síntese comparativa do desempenho médio dos modelos de IA e estratégias de prompt	41

Sumário

1	INTRODUÇÃO	1
2	REFERENCIAL TEÓRICO	5
2.1	CURADORIA DE MATEMÁTICAS DIDÁTICAS	5
2.2	A BNCC E O ENSINO DA MATEMÁTICA	8
2.3	PROCESSAMENTO DE LINGUAGEM NATURAL E EDUCAÇÃO	12
2.4	MODELOS DE LINGUAGEM E ENGENHARIA DE PROMPTS	14
3	METODOLOGIA	19
3.1	CARACTERIZAÇÃO DA PESQUISA	19
3.2	CORPUS DE ANÁLISE	19
3.2.1	Materiais didáticos selecionados	19
3.2.2	Modelos de inteligência artificial Generativa	22
3.3	ESTRATÉGIAS DE FORMULAÇÃO DE PROMPTS	22
3.3.1	Prompt 1 – identificação direta	22
3.3.2	Prompt 2 – com exemplos e estrutura detalhada	23
3.3.3	Prompt 3 – raciocínio passo a passo com reflexão	23
3.3.4	Prompt 4 – revisão em múltiplas etapas	24
3.4	PROCEDIMENTOS DE COLETA DE DADOS	24
3.4.1	Construção do gabarito de referência	24
3.4.2	Aplicação dos prompts aos modelos de IA	25
3.5	MÉTRICAS DE AVALIAÇÃO	26
3.5.1	Precisão	27
3.5.2	Recall (revocação)	27
3.5.3	F₁-Macro	27
3.5.4	Coeficiente de Jaccard	27
3.6	PROCEDIMENTOS DE ANÁLISE DE DADOS	28
3.6.1	Análise por estratégia de prompt	28
3.6.2	Análise comparativa entre modelos	28
3.6.3	Análise gráfica e visualização	28
3.7	LIMITAÇÕES METODOLÓGICAS	29
3.8	ASPECTOS ÉTICOS	29

4	RESULTADOS	31
4.1	ANÁLISE POR ENGENHARIA DE PROMPTS	31
4.1.1	Síntese comparativa entre prompts	36
4.2	ANÁLISE DE DESEMPENHO MÉDIO GLOBAL	36
4.2.1	Desempenho médio por formulação de prompt	37
4.2.2	Análise gráfica de desempenho médio entre modelos e prompts . . .	38
5	DISCUSSÕES	43
5.1	ANÁLISE COMPARATIVA DOS RESULTADOS	43
5.1.1	Aprofundamento da análise comparativa: o trade-off pedagógico . .	44
5.2	IMPLICAÇÕES DOS RESULTADOS PARA A PRÁTICA DOCENTE	45
5.2.1	Precisão e recall na curadoria automatizada	45
5.2.2	Influência da estrutura do material no desempenho dos modelos . .	45
5.2.3	Avanços metodológicos em engenharia de prompts	46
5.2.4	Evolução tecnológica e obsolescência	46
5.2.5	Implicações para formação docente e letramento em IA	46
5.3	DIREÇÕES PARA PESQUISAS FUTURAS	47
5.4	SÍNTESE DOS ACHADOS PRINCIPAIS	47
6	CONCLUSÃO	49
	APÊNDICE	55

1 INTRODUÇÃO

A Base Nacional Comum Curricular (BNCC) constitui-se como documento normativo que define o conjunto essencial de aprendizagens a serem asseguradas a todos os estudantes da Educação Básica brasileira, orientando a elaboração de currículos, a formação docente e a produção de materiais didáticos (12). Esse documento estabelece um conjunto orgânico de aprendizagens essenciais, estruturadas em competências e habilidades que devem ser desenvolvidas ao longo de toda a Educação Básica. No contexto do ensino de Matemática, a BNCC propõe o desenvolvimento de competências e habilidades que ultrapassam a mera reprodução de procedimentos, enfatizando o raciocínio lógico, a resolução de problemas, a argumentação e o uso de tecnologias para explorar conceitos e representações. Entretanto, apesar de sua centralidade, o alinhamento efetivo entre os materiais didáticos disponíveis (livros, apostilas, recursos digitais) e as habilidades prescritas ainda constitui um desafio prático e temporal significativo para professores e pesquisadores, como apontam (36).

Nas últimas décadas, a expansão acelerada de repositórios e plataformas digitais de materiais pedagógicos trouxe novas possibilidades de acesso a recursos, mas também instaurou um cenário de sobrecarga informacional. Por um lado, as ferramentas de busca tradicionais, frequentemente baseadas em descritores superficiais ou metadados pouco estruturados, não oferecem mecanismos semânticos adequados para filtrar conteúdos de acordo com as competências e habilidades específicas da BNCC. Por outro, observa-se, no contexto da prática docente, que a abundância de informações impõe ao professor o desafio de selecionar, em tempo reduzido, materiais pertinentes e pedagogicamente alinhados com os objetivos de ensino. Assim, o processo de planejamento de aulas de Matemática, em vez de se tornar mais eficiente, pode ser prejudicado pela ausência de sistemas inteligentes capazes de apoiar a curadoria de recursos.

Neste contexto, o advento de modelos de linguagem de grande escala (Large Language Models – LLMs), conhecidos popularmente como Inteligência Artificial Generativa (IAG), representa uma inflexão tecnológica com potencial para transformar processos educacionais. Essas ferramentas são capazes de processar, interpretar e gerar linguagem natural de forma sofisticada e emergem como poderosos aliados na otimização de tarefas intelectuais complexas. Especificamente, o uso de LLMs pode automatizar o mapeamento curricular, realizando a complexa tarefa de correlacionar textos de materiais didáticos com as habilidades detalhadas na BNCC, aliviando a carga de trabalho do docente e aprimorando a precisão do alinhamento.

As técnicas de Processamento de Linguagem Natural (PLN) e os modelos de linguagem

baseados em Inteligência Artificial (IA) oferecem novas formas de analisar e classificar textos educacionais, possibilitando identificar relações entre conteúdos, objetivos e competências de maneira mais precisa e automatizada. Contudo, a precisão e a utilidade da saída gerada por um LLM dependem crucialmente da formulação da instrução de entrada, um processo conhecido como Prompt Design ou Engenharia de Prompts. O desempenho da IA na identificação e extração de habilidades da BNCC não é uniforme; ele é altamente sensível à clareza, especificidade e estruturação da instrução que lhe é fornecida.

Sob a ótica da relevância social, esta pesquisa responde à necessidade urgente de instrumentalizar o docente da Educação Básica frente às exigências normativas da BNCC. Como apontam Macêdo, Brandão e Nunes (36), o alinhamento curricular manual constitui um desafio prático significativo, frequentemente limitado pela dependência exclusiva do livro didático. Nesse cenário, o desenvolvimento de estratégias automatizadas de curadoria não visa substituir o professor, mas atuar como uma 'inteligência assistiva' capaz de mitigar a sobrecarga informacional inerente à cultura digital descrita por Kenski (33). Ao otimizar a triagem de materiais, busca-se devolver ao docente o tempo necessário para o planejamento pedagógico criativo e a mediação efetiva em sala de aula.

Já a relevância científica deste estudo reside na análise sistemática da Engenharia de Prompts aplicada a documentos curriculares brasileiros, preenchendo uma lacuna na literatura de Processamento de Linguagem Natural e Ensino da Matemática. Diferentemente de abordagens genéricas, esta investigação compara empiricamente a eficácia de estratégias cognitivas — como Chain-of-thought (50) e Few-shot learning (15) — mensurando objetivamente a precisão e a revocação (recall) dos modelos. Contribui-se, assim, para o entendimento técnico de como diferentes arquiteturas de LLMs interpretam a semântica pedagógica da BNCC, oferecendo métricas rigorosas para validar o uso dessas ferramentas na educação.

Epistemologicamente, a pesquisa fundamenta-se na concepção do Ensino da Matemática mediada por tecnologias digitais, compreendendo a inteligência artificial não como uma ferramenta externa ou neutra, mas como parte integrante de um sistema de reorganização do pensamento. Alinha-se, portanto, à perspectiva de 'seres-humanos-com-mídias' proposta por Borba e Villarreal (8), onde o conhecimento matemático e curricular é produzido na simbiose entre a cognição docente e o processamento algorítmico, expandindo as fronteiras da sala de aula e da autoria docente (9).

Dessa forma, a presente dissertação tem como objetivo central comparar a eficácia de diferentes estratégias de formulação de prompts de Inteligência Artificial Generativa na tarefa de extração e alinhamento de habilidades da BNCC em materiais didáticos de Matemática. Para tanto, busca-se avaliar métricas de Precisão e Recall em diferentes modelos (Gemini, GPT, Claude) e analisar a influência da estruturação do prompt (direto vs. raciocínio passo a passo) na qualidade da resposta. Para cumprir este propósito, a pesquisa está estruturada em cinco capítulos: no Capítulo 2, revisa-se o referencial teórico sobre a BNCC, o desafio do alinhamento, o processamento de linguagem natural e a engenharia de prompts, bem como o uso da IA na

Educação; no Capítulo 3, detalha-se a metodologia de comparação das diferentes estratégias de prompt e de avaliação dos resultados; o Capítulo 4 descreve os resultados obtidos através da pesquisa; o Capítulo 5 apresenta a análise e discussão dos dados; e, por fim, o Capítulo 6 conclui o trabalho com as considerações finais e sugestões para pesquisas futuras.

Busca-se, com isso, contribuir para o desenvolvimento de ferramentas de apoio ao professor, capazes de aproximar a lógica curricular normativa da BNCC dos contextos reais de prática docente e de seleção de recursos. A pesquisa insere-se, portanto, no campo do Ensino da Matemática mediada por tecnologias digitais, articulando aspectos de análise curricular, linguística computacional e design pedagógico. Ao discutir os resultados comparativos entre prompts e estratégias de extração, pretende-se refletir não apenas sobre a eficiência técnica das abordagens testadas, mas também sobre suas implicações didático-pedagógicas. Espera-se que este estudo contribua para subsidiar o desenvolvimento de estratégias de prompt eficazes e aplicáveis às práticas da educação alinhadas à BNCC e mais responsivas às necessidades formativas dos estudantes da Educação Básica.

2 REFERENCIAL TEÓRICO

Este capítulo apresenta os fundamentos teóricos que sustentam a investigação, organizados em quatro eixos conceituais interconectados. Inicialmente, discute-se a curadoria de materiais didáticos (Seção 2.1), abordando seu papel na cultura digital e as políticas públicas estruturantes, como o PNLD. Em seguida, analisa-se a Base Nacional Comum Curricular (BNCC) (Seção 2.2) sob a ótica do desenvolvimento de competências matemáticas. O terceiro eixo explora o Processamento de Linguagem Natural (PLN) na educação (Seção 2.3), contextualizando o uso de inteligência artificial como ferramenta de apoio docente. Por fim, aprofunda-se a teoria sobre Engenharia de Prompts (Seção 2.4), detalhando as estratégias técnicas e cognitivas que fundamentam a metodologia deste estudo.

2.1 CURADORIA DE MATERIAIS DIDÁTICOS

A curadoria de materiais didáticos na Educação Básica é um processo crucial para a qualidade do ensino, que vai muito além de uma simples seleção, pois trata-se de uma prática intencional que envolve identificar, selecionar, organizar e avaliar recursos, garantindo a coerência pedagógica e o alinhamento com as diretrizes curriculares vigentes. Essa prática exige rigor científico, sensibilidade pedagógica e um compromisso constante com a melhoria da educação, servindo como um pilar fundamental para a construção de um currículo vibrante e significativo.

No Brasil, o Programa Nacional do Livro e do Material Didático (PNLD) é o principal mecanismo para a avaliação e seleção de materiais educativos, assegurando que os recursos distribuídos às escolas públicas estejam alinhados aos critérios pedagógicos, técnicos e éticos estabelecidos pelo MEC. Como aponta a própria descrição do programa, ele “é destinado a avaliar e a disponibilizar obras didáticas, pedagógicas e literárias de forma sistemática, regular e gratuita”(10). Embora historicamente focado no livro impresso, o programa expandiu seu escopo, como evidenciado pelos editais recentes para Obras Pedagógicas e Recursos Educacionais Digitais (RED) e Obras Literárias, de acordo com os editais do programa (11). Essa ampliação reflete a necessidade de diversificar os suportes de aprendizagem, embora a curadoria de materiais digitais abertos na web ainda careça de diretrizes tão consolidadas quanto as do programa oficial.

A busca por materiais é ampla e envolve diversas fontes: livros didáticos aprovados pelo PNLD, recursos digitais, plataformas educacionais, aplicativos, vídeos, simuladores, jogos, materiais manipuláveis, artigos científicos e até mesmo produções culturais que possam ser

contextualizadas adaptados para o uso em sala de aula. Cada material é submetido a uma rigorosa avaliação, que considera a adequação curricular, qualidade didática e metodológica, relevância, atualidade, diversidade, inclusão, confiabilidade, veracidade, engajamento, potencial de motivação, custo, acessibilidade, facilidade de acesso e uso pelos professores e alunos, sempre alinhado aos objetivos de aprendizagem e às competências da BNCC.

Em alguns casos, a curadoria vai além da simples seleção, pois pode ser necessário adaptar materiais existentes para torná-los mais adequados ao contexto local, ou até mesmo produzir novos materiais quando não se encontra algo que atenda às necessidades específicas, o que exige criatividade e profundo conhecimento pedagógico.

As obras didáticas devem ser projetadas para garantir que o processo de aprendizagem ocorra de forma contínua e organizada, levando em conta a função do professor como o responsável por organizar as ações pedagógicas e os estudantes como protagonistas de suas aprendizagens tendo o livro didático como um instrumento cultural que medeia a aprendizagem do estudante. (10)

A curadoria não termina na seleção, pois é um processo contínuo que envolve o acompanhamento do uso dos materiais em sala de aula, a coleta de feedback dos professores e alunos, bem como a realização de ajustes e aprimoramentos. Conforme apresentado por Fiscarelli, “os materiais didáticos selecionados e escolhidos pelos professores são aqueles que, primeiramente, dão segurança a eles quanto à maneira de usá-los e à receptibilidade dos alunos.”(28). É essencial que os curadores estejam abertos a revisar suas escolhas e a incorporar novas descobertas e tecnologias.

A curadoria de materiais é frequentemente um processo colaborativo, pois os professores podem compartilhar suas seleções, avaliações e adaptações com outros colegas através de plataformas online e redes de professores, enriquecendo o trabalho de toda a equipe pedagógica e fortalecendo o desenvolvimento da interdisciplinaridade e multidisciplinaridade. “A tecnologia tem o potencial para motivar os alunos a realizar projetos e aumentar o seu interesse, podendo ajudá-los ao realizarem suas investigações, comunicarem e colaborarem com colegas e reunirem informações relevantes para gerar artefatos que representem seu entendimento.”(19).

Mesmo sem uma validação extensiva, alguns modelos ou abordagens de curadoria são frequentemente discutidos ou propostos na literatura. A validação aqui muitas vezes significa que o modelo foi testado ou aplicado em um contexto, mas não necessariamente que sua eficácia foi comprovada e generalizada para diferentes cenários.

Os professores reconhecem a curadoria como uma habilidade essencial para o ensino no século XXI, e a veem como um meio para otimizar o tempo, personalizar o aprendizado e aumentar o engajamento dos alunos. A curadoria sempre foi uma ferramenta utilizada no meio educacional, direcionando o estudo dos alunos de forma variada, cujo objetivo principal é enriquecer com uma diversidade de recursos e conteúdos.

Com a era digital, temos uma variedade de recursos tecnológicos disponíveis, jogos, aplicativos, vídeos e podcast que podem ser usados como complementos, ponto de partida para

discussões ou parte integrante de metodologias ativas, por exemplo a sala de aula invertida. De acordo com (43), “a internet possibilitou a disponibilização de uma plataforma global de criação e acesso a uma imensa variedade de recursos e conteúdos de ‘muitos para muitos’.”, porém entre os principais obstáculos para a integração de vídeos nas práticas escolares estão a falta de tempo para a seleção criteriosa, a ausência de formação docente específica e as dificuldades de acesso a materiais de qualidade.

No cenário da cultura digital, cabe ao professor o papel de curador, filtrando, avaliando e integrando os recursos audiovisuais em propostas pedagógicas que façam sentido para os estudantes. Assim, o professor utiliza diferentes recursos em sua aula, tornando-a mais atrativa, motivadora, interativa, dinâmica, acessível e criativa, atraindo, assim, a concentração e interesse de todos envolvidos. A pesquisa de (20) identificou cinco competências digitais essenciais para educadores: letramento digital; curadoria de conteúdos digitais; criação de experiências de aprendizagem digital; avaliação em ambientes digitais; ética e segurança no uso de tecnologias.

Muitos estudos já realizados buscam entender como os professores selecionam, avaliam e integram vídeos em suas aulas, bem como os desafios e facilidades que encontram. (40) define a etapa de curadoria da informação digital em etapas que são achar; selecionar de acordo com qualidade, originalidade e relevância; editorializar; arranjar; criar; compartilhar; engajar e monitorar.

Outra abordagem bastante conhecida é “The 5 Models Of Content Curation”, a qual considera como etapas agrupar os conteúdos em um único local de forma automatizada; filtrar informações divulgando apenas as ideias mais relevantes; identificar tendências importantes dos conteúdos; sistematizar a curadoria baseado na mistura de conteúdo e apresentar os conteúdos em uma linha do tempo organizada (7).

A IA desempenha um papel importante no YouTube, tanto no ranqueamento de vídeos quanto nas sugestões de novos vídeos aos usuários. Para que se possam recomendar itens a um usuário, é necessário se ter conhecimento sobre quem é este usuário, logo é necessário capturar e armazenar os dados pessoais e comportamentais relativos. Portanto, faz-se crucial identificar o usuário quando este acessa o sistema onde as rotinas de recomendação foram implantadas.

Implantação da IA generativa para melhorar a moderação de conteúdo: A IA generativa já está nos ajudando a ampliar rapidamente o conjunto de informação usado para treinar nossos classificadores, o que permite a nós identificar conteúdo abusivo mais rapidamente. A maior velocidade e precisão dos nossos sistemas também nos permite diminuir a quantidade de conteúdo nocivo a que os revisores humanos estão expostos. (51).

Os sistemas de IA enfrentam desafios técnicos e éticos significativos na distinção entre nuances linguísticas, expressões irônicas, manifestações de opinião e desinformação deliberada, constituindo uma área de pesquisa complexa em Processamento de Linguagem Natural e ética computacional. Além disso, há o risco de a moderação de IA ser vista como censura, assim o equilíbrio entre a liberdade de expressão e a necessidade de proteger os usuários de conteúdo prejudicial é delicado.

O algoritmo do YouTube é um sistema inteligente que analisa o comportamento dos usuários em tempo real para recomendar vídeos que sejam mais relevantes e mantenham o público engajado por mais tempo. Ou seja, um mecanismo complexo que age como um “curador” personalizado, conectando cada espectador a conteúdos que combinem com seus interesses. (45).

Os algoritmos de IA são treinados com base em vastos conjuntos de dados de desempenho que ajudam o algoritmo a identificar se o vídeo é relevante e atraente para o público e de dados de personalização que garantem que as recomendações sejam adaptadas a cada usuário. Esses fatores incluem cliques no vídeo; número de impressões (vezes que a miniatura é exibida) em relação às visualizações (logo miniaturas e títulos atrativos são essenciais aqui); tempo assistido do vídeo; interações no vídeo incluindo curtidas, descurtidas, comentários e compartilhamentos; popularidade do vídeo; se o vídeo é novo, ele pode receber um impulso inicial para alcançar o público certo; frequência de envio do canal; avaliação do tempo que o usuário continua navegando no YouTube após assistir ao vídeo; recomendação de vídeos similares ao histórico do espectador; envolvimento com vídeos anteriores; tempo assistido em vídeos semelhantes indicando interesse no tema; frequência de exposição; vídeos marcados como “não tenho interesse”.

O algoritmo, ao tentar otimizar o engajamento, pode inadvertidamente priorizar conteúdo que gera mais cliques e visualizações, que por vezes são conteúdos mais polarizadores ou sensacionalistas. O grande desafio para as plataformas é que identificar e corrigir esses vieses é extremamente difícil, pois os algoritmos são sistemas de baixa interpretabilidade complexos, e o volume de dados é gigantesco. “Esses algoritmos utilizam Inteligência Artificial (IA) para prever e sugerir conteúdo que pode interessar ao usuário com base em seu comportamento e preferências anteriores” (1). A busca por algoritmos com maior equidade, o combate à desinformação em larga escala e a proteção da privacidade do usuário constituem desafios multidimensionais que demandam inovação contínua, regulamentação apropriada e diálogo permanente entre desenvolvimento tecnológico, considerações éticas e demandas sociais.

2.2 A BNCC E O ENSINO DA MATEMÁTICA

A Base Nacional Comum Curricular (BNCC) materializa, no contexto brasileiro, a transição para um currículo orientado por competências. Essa abordagem dialoga com a perspectiva de (39), para quem a competência se manifesta na capacidade de mobilizar recursos cognitivos diversos para solucionar situações complexas. No ensino da Matemática, isso implica uma ruptura com o modelo conteudista tradicional. Conforme discutido por (37) no âmbito de avaliações internacionais como o PISA (Programme for International Student Assessment), o letramento matemático não reside no domínio isolado de algoritmos, mas na capacidade de matematizar a realidade. Assim, ao prescrever o desenvolvimento de habilidades como argumentar e modelar, a BNCC exige do professor uma curadoria de materiais que fomentem a investigação ativa, e não apenas a reprodução mecânica descrita por (23) como insuficiente para o cenário atual.

A estruturação da BNCC em competências e habilidades também impacta diretamente os processos avaliativos, exigindo que instrumentos de avaliação considerem não apenas a reprodução de conhecimentos, mas a capacidade de aplicá-los em situações-problema autênticas. Nesse sentido, o alinhamento entre materiais didáticos e habilidades da BNCC torna-se crucial para assegurar coerência entre o que se ensina, como se ensina e como se avalia.

Segundo o documento, a área de Matemática deve promover o desenvolvimento de competências específicas que possibilitem aos estudantes “desenvolver o raciocínio lógico, o espírito de investigação e a capacidade de produzir argumentos convincentes, recorrendo aos conhecimentos matemáticos para compreender e atuar no mundo” e “utilizar processos e ferramentas matemáticas, inclusive tecnologias digitais disponíveis, para modelar e resolver problemas cotidianos, sociais e de outras áreas de conhecimento, validando estratégias e resultados” (12). Essa perspectiva amplia o papel da Matemática escolar, deslocando-o de uma abordagem centrada na repetição de algoritmos para um enfoque de mobilização de conhecimentos em situações-problema. Tal movimento é coerente com a concepção de Ensino da Matemática como campo que articula o domínio conceitual ao desenvolvimento de práticas sociais de significação (23).

A noção de competência adotada pela BNCC fundamenta-se na perspectiva de que o conhecimento deve ser mobilizável em contextos diversos, transcendendo a mera memorização de conteúdos. (39) define competência como a capacidade de mobilizar recursos cognitivos para solucionar situações complexas, o que implica articular conhecimentos teóricos, habilidades práticas e atitudes adequadas ao contexto.

A BNCC também estrutura a área em unidades temáticas — Números, Álgebra, Geometria, Grandezas e Medidas e Probabilidade e Estatística —, dentro das quais são definidos objetos de conhecimento e habilidades associadas a cada ano escolar. Essa estrutura favorece a progressão das aprendizagens e o desenvolvimento de competências transversais, na medida em que os objetos de conhecimento não se reduzem a conteúdos, mas expressam processos e práticas que articulam saberes conceituais, procedimentais e atitudinais.

diferentes campos que compõem a Matemática reúnem um conjunto de ideias fundamentais que produzem articulações entre eles: equivalência, ordem, proporcionalidade, interdependência, representação, variação e aproximação. Essas ideias fundamentais são importantes para o desenvolvimento do pensamento matemático dos alunos e devem se converter, na escola, em objetos de conhecimento (12).

Essa concepção de competência, entendida como “a mobilização de conhecimentos (conceitos e procedimentos), habilidades (práticas, cognitivas e socioemocionais), atitudes e valores para resolver demandas complexas da vida cotidiana, do pleno exercício da cidadania e do mundo do trabalho.” (12) alinha-se ao paradigma de competência discutido por (39), segundo o qual o ensino deve proporcionar aos alunos condições de transferir saberes e aplicá-los em novos contextos, tal perspectiva implica deslocar o foco da mera execução de técnicas para o desenvolvimento do raciocínio, da argumentação e da modelagem, dimensões que ampliam o

sentido social e formativo dessa área.

Diversos autores brasileiros têm apontado que a adoção de competências e habilidades como eixo estruturante da BNCC representa uma oportunidade de repensar a prática docente e o papel do currículo. (35) argumenta que ensinar Matemática não deve ser apenas “fazer o aluno resolver contas”, mas levá-lo a compreender a lógica interna e a aplicabilidade da disciplina em situações reais. De modo convergente, (27) defendem que o ensino de Matemática precisa promover uma aprendizagem significativa e problematizadora, o que requer que o professor se aproprie criticamente dos referenciais curriculares e os traduza em práticas coerentes com o contexto de sua turma.

Um dos principais desafios é garantir que os recursos selecionados por meio da curadoria estejam, de fato, integrados ao currículo escolar e não apenas utilizados de forma complementar e fragmentada. Para isso, é preciso focar em formação inicial e continuada tanto dos professores quanto da equipe pedagógica, pois a ausência de programas de formação em curadoria digital dificulta que os professores desenvolvam competências críticas para avaliar a qualidade e a pertinência pedagógica dos recursos.

A construção de um repertório variado de referências por meio de ações de curadoria é crucial para que o professor possa direcionar – com intencionalidade – a sua criatividade no momento de construir ou adaptar materiais didáticos, auxiliando também para que o processo de autoria se torne mais prazeroso e assim entusiasme o professor a seguir criando. (4).

A curadoria não termina na seleção, pois é um processo contínuo que envolve o acompanhamento do uso dos materiais em sala de aula. Em consonância com essa perspectiva, (4) argumentam que o professor, ao selecionar e adaptar recursos, assume uma identidade de ‘professor-autor’. Expandindo essa visão para o contexto das interfaces digitais, (17) destacam que a curadoria de conteúdo, quando associada à reflexão sobre a prática e ao uso de portfólios digitais, potencializa a autoria docente. Segundo os autores, ao organizar e articular conteúdos digitais, o docente deixa de ser apenas consumidor para arquitetar trilhas de aprendizagem, desenvolvendo competências críticas essenciais para a cultura digital.

As percepções e as práticas de curadoria de materiais didáticos por professores de matemática na educação básica constituem um campo de estudo específico dentro da tecnologia educacional. Embora o foco em matemática possa ser menos abordado na literatura em comparação com a curadoria em geral, é possível inferir informações valiosas a partir de estudos relacionados para identificar tendências e resultados.

O valor dos vídeos para o ensino de matemática é um dos pilares investigados, uma vez que os professores reconhecem o potencial dos vídeos para visualizar conceitos abstratos, apresentar aplicações do mundo real, engajar os alunos e atender a diferentes estilos de aprendizagem. No entanto, o desafio central é a dependência do livro didático, que “se faz presente nos processos de ensino e de aprendizagem, funcionando na maioria das vezes como o único material didático de que professores e alunos têm acesso.”(36).

Os critérios que os professores de matemática utilizam para selecionar vídeos são qualidades gerais, precisão, clareza e adequação à idade. Eles buscam a correção dos conceitos matemáticos, a clareza das demonstrações, a relevância para o currículo e a forma como o vídeo aborda a resolução de problemas. Essas práticas de busca, avaliação e organização ocorrem em diversas plataformas, e a integração dos vídeos curados nas aulas de matemática pode acontecer em variados momentos, desde a introdução a um novo tópico, a explicação de um conceito ou até mesmo a revisão de conteúdo.

Além dos obstáculos comuns, como a falta de tempo e a dificuldade em encontrar conteúdo de qualidade, surgem desafios específicos da área, como a complexidade em encontrar vídeos que expliquem conceitos complexos de forma clara e precisa. (36) ressaltam a falta de criatividade, a elevada carga horária, as péssimas condições de trabalho e a desvalorização profissional como fatores que limitam a inovação no uso de recursos.

Adicionalmente, o uso de recursos digitais sem critérios pedagógicos sólidos pode levar a um ensino superficial, focado mais no entretenimento do que na construção de conhecimento. Como apontam (16), professores relatam dificuldades em conciliar o planejamento da disciplina com as práticas em laboratórios de informática, e notam que os alunos tendem a associar essas aulas ao entretenimento.

Por fim, há um consenso de que os professores se sentem despreparados para a curadoria de materiais didáticos de matemática e necessitam de formação contínua que lhes possibilite desenvolver competências de análise crítica e integração pedagógica das mídias. Há um reconhecimento geral de que vídeos curtos e bem planejados aumentam o engajamento dos alunos, especialmente quando contextualizados com situações do cotidiano.

Contudo, a implementação da BNCC em sala de aula apresenta desafios como a estrutura do documento, embora coerente com uma visão por competências, não se traduz automaticamente em materiais didáticos ou práticas pedagógicas alinhadas. Persistem dificuldades de interpretação dos descritores de habilidades e de articulação entre o texto normativo e os objetos de aprendizagem efetivamente trabalhados em livros e plataformas digitais. Esse hiato reforça a necessidade de pesquisas que proponham formas de analisar, classificar e alinhar materiais didáticos às habilidades da BNCC, o que justifica o presente estudo.

Em síntese, a BNCC introduz uma concepção de ensino de Matemática que valoriza o raciocínio e a resolução de problemas. Ao mesmo tempo, impõe desafios de natureza interpretativa e operacional, especialmente no que se refere à curadoria e à adequação de recursos pedagógicos aos objetivos curriculares. Tais desafios constituem o plano de fundo para o desenvolvimento de metodologias automáticas de extração e classificação de habilidades, como a proposta nesta pesquisa, que busca integrar a leitura curricular à mediação tecnológica, favorecendo uma prática docente mais eficiente e coerente com as diretrizes nacionais.

2.3 PROCESSAMENTO DE LINGUAGEM NATURAL E EDUCAÇÃO

O avanço das tecnologias de IA e, em especial, das técnicas de Processamento de Linguagem Natural (PLN), tem produzido transformações significativas no campo educacional. O PLN, entendido como uma área da ciência da computação que utiliza aprendizado de máquina para possibilitar que computadores reconheçam, entendam, processam e gerem linguagem humana de forma natural e contextual (31), oferece um conjunto de métodos e ferramentas capazes de extrair, classificar e representar informações semânticas de corpora textuais extensos como documentos curriculares, planos de aula ou materiais didáticos.

Na perspectiva educacional, essas tecnologias inauguram possibilidades de análise automatizada de discursos, currículos e produções textuais que, até recentemente, dependiam exclusivamente de procedimentos manuais. O PLN tem potencial para auxiliar pesquisadores e professores a compreender padrões de linguagem presentes em textos educacionais, favorecendo a construção de instrumentos de análise mais objetivos e escaláveis. Essa perspectiva é particularmente relevante no contexto da Base Nacional Comum Curricular (BNCC), em que a estrutura textual altamente normativa e descritiva requer métodos de extração capazes de identificar e agrupar competências e habilidades de modo coerente com sua semântica.

No Brasil, a discussão sobre o uso de IA na educação tem ganhado destaque crescente. Segundo pesquisa realizada pelo Instituto Semesp em 2024, três em cada quatro professores concordam com o uso da tecnologia como ferramenta de ensino, embora reconheçam desafios relacionados à dispersão dos estudantes e à necessidade de formação adequada (2). O Ministério da Educação tem desenvolvido iniciativas para orientar o uso responsável dessas tecnologias, incluindo a elaboração do Referencial para Uso e Desenvolvimento Responsáveis de Inteligência Artificial na Educação (13).

Nos últimos anos, a emergência dos modelos de linguagem de grande escala (LLMs), como o BERT (Bidirectional Encoder Representations from Transformers), desenvolvido por pesquisadores do Google em 2018, redefiniu o panorama do PLN. O BERT é um modelo pré-treinado de código aberto que utiliza a arquitetura transformer para aprender representações semânticas profundas e contextualizadas de forma bidirecional (26). Diferentemente de modelos anteriores que processavam texto sequencialmente, o BERT analisa simultaneamente o contexto completo de uma palavra, permitindo uma compreensão mais sofisticada das relações linguísticas (25). Essas representações viabilizam o reconhecimento de relações sutis entre conceitos e expressões, ampliando consideravelmente a capacidade dos sistemas de compreender documentos educacionais complexos.

No contexto da Ensino da Matemática, o uso do PLN apresenta especificidades relevantes, de acordo com (9), textos didáticos dessa área frequentemente combinam linguagem natural, notação simbólica e representações visuais, o que exige abordagens híbridas que integrem

análise textual e reconhecimento estrutural. Além disso, a terminologia matemática apresenta polissemias contextuais — termos como razão, função ou relação — que demandam estratégias semânticas sensíveis ao domínio conceitual.

Marcelo de Carvalho Borba, professor do Programa de Pós-Graduação em Ensino da Matemática da UNESP e coordenador do Grupo de Pesquisa em Informática, outras Mídias e Ensino da Matemática (GPIMEM), tem desenvolvido pesquisas há mais de duas décadas sobre o uso de tecnologias digitais na Ensino da Matemática (9). Seu constructo teórico “seres-humanos-com-mídias” propõe compreender a produção de conhecimento matemático como resultado da interação entre humanos e tecnologias, concebendo as mídias digitais como coautoras do pensamento (8).

Pesquisas em Learning Analytics e Educational Data Mining também têm explorado o potencial do PLN na compreensão de processos de ensino-aprendizagem. (42), em estudo publicado no periódico Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, destacam que essas áreas evoluíram enormemente na última década, aplicando técnicas de mineração de dados e análise automática de textos para oferecer indicadores sobre práticas pedagógicas, competências discursivas e engajamento estudantil. (6) argumentam que, embora Educational Data Mining e Learning Analytics compartilhem objetivos comuns, diferem em seus métodos: enquanto EDM foca no desenvolvimento de novos modelos e algoritmos, Learning Analytics aplica modelos preditivos conhecidos em sistemas instrucionais.

A análise automatizada da BNCC por meio de PLN insere-se, assim, em um campo emergente de interseção entre Linguística Computacional e Ensino da Matemática, no qual o objetivo não é substituir o olhar docente, mas ampliar suas possibilidades de leitura e organização de informações. Esse processo pode ser entendido como uma forma de inteligência assistiva, que potencializa a capacidade dos educadores de navegar por documentos extensos e extrair relações entre habilidades, conteúdos e práticas pedagógicas.

No que se refere à curadoria de recursos educacionais digitais, o Centro de Inovação para a Educação Brasileira (CIEB) publicou em 2019 um estudo sobre modelos de curadoria, destacando a importância de critérios pedagógicos e tecnológicos para seleção de materiais (18). O estudo propõe que professores atuem como curadores, responsáveis por selecionar e classificar recursos digitais adequados aos objetivos de aprendizagem e ao contexto curricular. Nesse sentido, sistemas baseados em PLN podem apoiar o processo de curadoria docente, facilitando o alinhamento entre recursos educacionais e as habilidades previstas na BNCC.

Do ponto de vista epistemológico, essa integração tecnológica dialoga com a concepção de docência aumentada por tecnologias, segundo a qual a IA deve ser vista como mediadora da reflexão docente, e não como substituta (33). Essa mediação implica compreender a IA como participante do coletivo cognitivo, no qual humanos e tecnologias constroem significados e decisões pedagógicas colaborativamente.

No âmbito internacional, iniciativas como as orientações da UNESCO sobre IA na educação têm consolidado diretrizes éticas, enfatizando princípios de transparência, equidade e

interpretabilidade (46) (47). No Brasil, pesquisadores têm destacado a necessidade de garantir que o desenvolvimento tecnológico tenha o ser humano no gerenciamento dos processos e como agente principal nas tomadas de decisões, desenvolvendo habilidades para a resolução de problemas que integrem a colaboração humano-IA e a avaliação crítica do conteúdo gerado (3).

No cenário brasileiro, a relação entre Inteligência Artificial e o trabalho docente exige um olhar atento às implicações éticas e à soberania humana na tomada de decisão. (5) alertam, com base na Teoria Crítica da Tecnologia, que a implementação da IA não deve ser vista apenas como inovação metodológica, mas analisada sob a ótica de suas consequências para a autonomia docente. Para os pesquisadores, é fundamental garantir que a tecnologia atue como suporte à atividade intelectual do professor, evitando que a mediação pedagógica seja substituída por automatismos ou mecanismos de precarização do trabalho docente.

Em síntese, o Processamento de Linguagem Natural aplicado à Educação oferece caminhos promissores para compreender e organizar o conhecimento curricular. No caso da Matemática, essa abordagem permite aproximar o texto normativo da BNCC dos materiais de ensino efetivamente disponíveis, contribuindo para o planejamento pedagógico e a formação docente. Contudo, o uso educativo da IA requer uma postura crítica que reconheça tanto suas potencialidades quanto seus limites, especialmente no que diz respeito à transparência algorítmica e à mediação humana.

2.4 MODELOS DE LINGUAGEM E ENGENHARIA DE PROMPTS

Os modelos de linguagem baseados em IA têm se tornado elementos centrais na pesquisa contemporânea em Processamento de Linguagem Natural (PLN) e, mais recentemente, em aplicações educacionais. Esses modelos, conhecidos pela sigla inglesa *Large Language Models* (LLMs), são arquiteturas de redes neurais profundas treinadas sobre vastos corpora textuais, capazes de reconhecer padrões complexos de sintaxe, semântica e pragmática da linguagem humana.

Os LLMs funcionam essencialmente como modelos probabilísticos que preveem a sequência de palavras mais provável em um dado contexto. Como explica (30), o modelo GPT processa palavras em unidades discretas chamadas *tokens*, sendo que algumas palavras são divididas em múltiplos tokens, e utiliza mecanismos de autoatenção para estabelecer dependências e relacionamentos entre os elementos de uma sequência. Diferentemente da compreensão humana, esses modelos internalizam estatisticamente relações entre palavras, frases e contextos, modelando padrões linguísticos sem necessariamente compreender o significado no sentido humano do termo. Ainda assim, sua capacidade de generalização contextual os torna ferramentas poderosas para tarefas de análise semântica, classificação textual e extração de informações, especialmente em domínios extensos e estruturados, como o texto da BNCC.

Modelos como GPT (*Generative Pre-trained Transformer*), desenvolvido pela OpenAI, baseiam-se na arquitetura *Transformer*, introduzida por (48) no artigo seminal “*Attention Is All You Need*”, onde apresenta como essa arquitetura revolucionou o campo do PLN ao introduzir o mecanismo de autoatenção (*self-attention*), que permite ao modelo identificar relações entre palavras distantes em uma sequência textual e processar dados de forma paralela, superando as limitações das redes neurais recorrentes (RNNs) anteriormente utilizadas. O mecanismo de autoatenção capacita os *transformers* a processar o contexto e responder com detalhes em linguagem natural, em vez de apenas prever a próxima palavra em uma frase de forma sequencial.

O GPT-1, lançado pela OpenAI em 2018, foi o primeiro modelo da família GPT e possuía 0,12 bilhão de parâmetros. O GPT-2, lançado em 2019, expandiu para 1,5 bilhão de parâmetros, e o GPT-3, revelado em 2020, atingiu 175 bilhões de parâmetros, tornando-se um marco na história dos modelos de linguagem (15). Segundo o (24), o GPT-3 foi treinado com mais de 45 terabytes de informações, incluindo textos de mais de 7 mil livros, todo o conteúdo da Wikipédia e bancos de dados como WebText e Common Crawl. Essa vastidão de dados de treinamento é o que permite ao GPT-3 realizar diversas tarefas de PLN sem necessidade de ajuste fino específico para cada tarefa, essa característica é conhecida como aprendizado *few-shot* ou *zero-shot*.

Em ambientes educacionais, tais modelos têm sido empregados para geração automática de feedback, análise de currículos, avaliação de coerência semântica de textos de estudantes e assistência ao planejamento docente. No entanto, a efetividade desses modelos depende fortemente da formulação linguística de *prompts*, que consiste nas instruções que orientam sua execução.

No âmbito da administração pública e institucional brasileira, a importância da formulação correta de instruções tem sido normatizada. Segundo o *Guia prático de prompt e pesquisa com IA* publicado pela Secretaria de Governo Digital, o prompt é definido como “uma instrução clara e contextual que orienta modelos de linguagem generativa a produzirem respostas mais precisas e relevantes” (14). O documento destaca que a engenharia de prompt atua como a disciplina dedicada a otimizar essas instruções, funcionando como o ponto de partida que baliza as expectativas sobre a resposta da IA. Para maximizar a utilidade das respostas, o Guia recomenda diretrizes como: evitar ambiguidades, atribuir um papel específico ao modelo (persona) e utilizar técnicas de refinamento iterativo (14).

A engenharia de prompts (*prompt engineering*) emergiu como um novo campo de estudo dedicado à concepção de estratégias linguísticas e estruturais que maximizem o desempenho de modelos de linguagem. Conforme (29), a engenharia de prompts é uma disciplina relativamente nova que visa desenvolver e otimizar prompts para utilizar eficientemente modelos de linguagem em uma ampla variedade de aplicações e tópicos de pesquisa. As habilidades de engenharia de prompt ajudam a entender melhor os recursos e as limitações dos LLMs, não se limitando apenas a projetar prompts, mas abrangendo uma ampla gama de técnicas úteis para interagir e desenvolver com LLMs. A qualidade da saída gerada depende não apenas do modelo, mas também da clareza, estrutura e propósito do prompt.

O Guia de Engenharia de Prompts sistematiza as principais categorias de prompts, entre as quais destacam-se:

- **Zero-shot prompts:** o modelo é instruído a realizar a tarefa sem exemplos prévios. Segundo a (32), o *zero-shot prompting* é um método que se baseia no pré-treinamento do LLM para inferir uma resposta apropriada, aproveitando o conhecimento incorporado durante o treinamento para abordar novos problemas sem exemplos específicos.
- **Few-shot prompts:** incluem exemplos de entrada e saída para orientar o comportamento do modelo. (15) demonstraram que fornecer alguns exemplos (tipicamente de 3 a 5) melhora significativamente o desempenho em tarefas complexas, pois os exemplos servem como condicionamento para as respostas subsequentes.
- **Chain-of-thought prompts:** estimulam a explicitação do raciocínio passo a passo. Segundo (50), o *chain-of-thought* (CoT) envolve a geração de sentenças curtas que explicam os passos de raciocínio um a um, levando à resposta final. A técnica CoT funciona melhor para tarefas de raciocínio complexas quando se utilizam modelos grandes com muitos parâmetros.
- **Zero-shot chain-of-thought:** uma variante que não requer exemplos prévios, apenas a adição de frases como “Vamos pensar passo a passo” ao final do prompt. (34) demonstraram que essa abordagem simples pode melhorar significativamente o desempenho em tarefas de raciocínio aritmético, senso comum e raciocínio simbólico.
- **Instructional prompts:** combinam instruções diretas, contexto e formato de resposta esperada, fornecendo orientações claras sobre como o modelo deve abordar a tarefa.

No contexto educacional, a escolha e a formulação de prompts também adquirem uma dimensão pedagógica. A interação com sistemas de IA requer competências para compreender seus mecanismos e avaliar os resultados gerados, denominada literacia de IA. Essa literacia implica reconhecer que os prompts são mais do que comandos técnicos: são atos de mediação linguística e pedagógica, que refletem intenções curriculares, concepções de aprendizagem e critérios de avaliação.

No campo da Ensino da Matemática, essa mediação assume relevância particular. A linguagem matemática, por sua natureza formal e simbólica, apresenta desafios à interpretação automática. (9) destacam que a IA, quando integrada ao ensino de Matemática, deve ser compreendida como participante de um coletivo de pensamento, no qual humanos e tecnologias constroem conjuntamente significados. Nessa perspectiva, o prompt é o elo discursivo que traduz a intencionalidade pedagógica humana em comandos compreensíveis pelo modelo de linguagem.

Além da dimensão técnica, a engenharia de prompts envolve questões éticas e epistemológicas. Os sistemas de IA refletem padrões presentes nos dados de treinamento e podem reproduzir vieses linguísticos e culturais. No caso da BNCC, cuja formulação se baseia em

princípios de equidade e inclusão, é fundamental assegurar que os processos de extração automática preservem esses valores, evitando reducionismos semânticos ou interpretações descontextualizadas das habilidades curriculares. Assim, o desenho de prompts para análise de textos educacionais deve combinar rigor técnico, transparência e sensibilidade pedagógica.

Do ponto de vista metodológico, a comparação entre diferentes tipos de prompt permite compreender de que modo a variação na formulação linguística impacta a capacidade de inferência semântica dos modelos de linguagem. Essa compreensão é essencial para o desenvolvimento de sistemas de curadoria automatizada que sejam pedagogicamente consistentes, pois, como observa (33), a tecnologia não é neutra: ela media e transforma os modos de pensar e ensinar.

A interação eficaz com esses modelos demanda mais do que habilidades técnicas; exige o desenvolvimento de um “Letramento Digital para Inteligências Artificiais”. Conforme teoriza (49), a formação docente não deve se limitar ao manuseio instrumental (o “como usar”), mas deve ser multidimensional, abrangendo aspectos éticos, de autoria e de segurança de dados. O autor argumenta que, diante da velocidade exponencial dos avanços tecnológicos, o letramento em IA deve capacitar o professor a compreender as implicações discursivas e culturais dessas ferramentas, superando a visão ingênua de que a tecnologia é apenas um suporte neutro (49).

Em síntese, os modelos de linguagem contemporâneos inauguram uma nova etapa na relação entre tecnologia e Ensino da Matemática, em que o prompt se configura como elemento estratégico para integrar o raciocínio humano à interpretação algorítmica. Investigar sua eficácia na extração de habilidades da BNCC não é apenas uma questão técnica, mas também epistemológica e pedagógica, pois envolve compreender como a linguagem, humana e artificial, pode ser mobilizada para promover uma leitura curricular mais inteligente, ética e formativa.

3 METODOLOGIA

Este capítulo descreve os procedimentos metodológicos adotados para investigar a eficácia de diferentes estratégias de formulação de prompts de Inteligência Artificial Generativa na tarefa de extração e alinhamento de habilidades da BNCC em materiais didáticos de Matemática. A pesquisa caracteriza-se como aplicada, de abordagem quantitativa, com objetivo descritivo e comparativo, utilizando procedimento experimental.

3.1 CARACTERIZAÇÃO DA PESQUISA

A presente investigação configura-se como uma pesquisa aplicada, uma vez que busca gerar conhecimento para aplicação prática na solução de problemas específicos do contexto educacional brasileiro: o desafio de alinhar materiais didáticos às habilidades prescritas pela BNCC. Quanto à abordagem, adota-se o método quantitativo, pois os resultados são expressos por meio de métricas numéricas derivadas de matrizes de confusão, permitindo análise estatística e comparação objetiva entre diferentes condições experimentais.

Do ponto de vista dos objetivos, a pesquisa é descritiva e comparativa: descritiva porque caracteriza sistematicamente o desempenho de modelos de IA em uma tarefa específica; comparativa porque contrasta três modelos distintos sob quatro estratégias de prompt diferentes. O procedimento adotado é experimental, uma vez que foram manipuladas variáveis independentes (tipo de modelo de IA e estratégia de prompt) para observar seus efeitos sobre variáveis dependentes (métricas de desempenho na identificação de habilidades).

3.2 CORPUS DE ANÁLISE

A seleção do corpus de análise é uma etapa fundamental para garantir a validade e a representatividade dos dados extraídos. A seguir, detalham-se os critérios de escolha dos materiais e os modelos de IA utilizados.

3.2.1 Materiais didáticos selecionados

O corpus desta pesquisa foi constituído por cinco materiais didáticos de Matemática, selecionados de forma intencional para representar diferentes formatos, níveis de complexidade

e unidades temáticas da BNCC. A diversidade do corpus visa avaliar a robustez dos modelos de IA em contextos variados, simulando a heterogeneidade de recursos encontrados na prática docente. A Tabela 3.1 apresenta a caracterização detalhada de cada material.

Tabela 3.1: Caracterização dos materiais didáticos utilizados no corpus da pesquisa

Material	Formato	Ano Escolar	Unidade Temática	Extensão
Frações.PNG (22)	Imagem	7º ano	Números	1 página
Matemática 9 ano apostila.pdf (41)	PDF	9º ano	Múltiplas	35 páginas
Média Aritmética.docx (21)	Doc. Word	6º ano	Probabilidade e Estatística	2 páginas
Operações com números na forma decimal.pdf (38)	PDF	6º ano	Números	10 páginas
Parte livro 7 ano.pdf (22)	PDF	7º ano	Números	2 páginas

Fonte: elaborado pela autora, 2025.

A constituição do corpus seguiu o critério de diversidade tipológica e não de exaustividade curricular. Por essa razão, optou-se por selecionar materiais do 6º, 7º e 9º anos que apresentassem desafios estruturais distintos para a IA (como a leitura de imagens no 7º ano ou tabelas complexas no 9º ano), em detrimento de uma progressão linear que incluísse o 8º ano. A ausência de material deste ano específico não compromete a análise, dado que o objetivo é testar a capacidade de interpretação semântica dos modelos diante de diferentes suportes.

A visualização completa desses materiais é fundamental para compreender as dificuldades impostas aos modelos, como a mistura de linguagem coloquial e formal na apostila do 9º ano (41)) ou a necessidade de OCR (reconhecimento óptico de caracteres) na imagem do 7º ano (22).

A constituição do corpus seguiu o critério de *amostragem intencional*, buscando representar a heterogeneidade de formatos com os quais o professor depara-se em sua curadoria digital cotidiana. Embora o referencial teórico destaque a centralidade dos vídeos (YouTube) na cultura digital, esta etapa da investigação optou por delimitar o escopo a materiais estáticos (PDF, DOCX e Imagem). Essa decisão metodológica justifica-se pela necessidade de isolar a capacidade interpretativa dos modelos sobre texto e imagem estática — substratos fundamentais que compõem também os vídeos — antes de introduzir a variável temporal de mídias audiovisuais.

A inclusão de um arquivo de imagem (.PNG) visa simular o cenário frequente onde o docente utiliza capturas de tela, quadros ou slides isolados de videoaulas, bem como testar a capacidade dos modelos de processar representações visuais, competência crescentemente relevante dado o uso de infográficos e esquemas na educação digital. Os materiais em PDF representam o formato mais comum de distribuição de conteúdos educacionais, enquanto o documento Word (.docx) exemplifica materiais produzidos por professores em plataformas de

edição de texto.

Para ilustrar a complexidade da tarefa imposta à IA, destaca-se o material “Frações.PNG” (ver Figura 3.1). Trata-se de uma imagem digitalizada de uma página de livro didático contendo representações híbridas: texto explicativo, notação matemática simbólica ($\frac{1}{2}$) e representações pictóricas (diagramas de barras e círculos). O processamento deste arquivo exige que o modelo realize reconhecimento óptico de caracteres (OCR) simultaneamente à interpretação semântica das imagens, uma competência de multimodalidade essencial para a curadoria moderna. Já o material “Matemática 9 ano apostila.pdf” apresenta desafios distintos: diagramação em colunas, caixas de texto laterais e uma mistura de linguagem coloquial com formalismo matemático, testando a capacidade da IA de filtrar ruídos visuais e focar no conteúdo curricular.

Figura 3.1: Exemplo da complexidade multimodal do material analisado (Frações.PNG)

Operações com frações
No Volume anterior, você já estudou algumas operações com frações. Vamos agora retomar e aprofundar esse estudo.

Adição e subtração de frações
Com denominadores iguais
Pela manhã, Paulo tomou a água correspondente a $\frac{1}{3}$ da medida de capacidade de uma jarra. No período da tarde, tomou o correspondente a $\frac{2}{3}$ dessa medida de capacidade. Qual fração da medida de capacidade da jarra ele tomou ao todo?
Para determinar essa fração, precisamos adicionar o $\frac{1}{3}$ da medida de capacidade que foi tomado de manhã com os $\frac{2}{3}$ da medida de capacidade consumidos no período da tarde.
$$\frac{1}{3} + \frac{2}{3} = \frac{3}{3}$$

Logo, Paulo bebeu $\frac{3}{3}$ da medida de capacidade da jarra.

Para adicionar ou subtrair frações com denominadores iguais, adicionamos ou subtraímos os numeradores e conservamos o denominador.

Com denominadores diferentes
A mãe de Maria tinha aproximadamente $\frac{2}{3}$ de um queijo e usou aproximadamente $\frac{1}{6}$ do queijo para fazer uma torta. Qual fração do queijo restou?
Para encontrar a resposta para esse problema, precisamos subtrair a quantidade da queijo usada da quantidade que a mãe de Maria tinha.
$$\frac{2}{3} - \frac{1}{6}$$

Como os denominadores são diferentes, precisamos determinar frações equivalentes a essas, que tenham denominadores iguais, para prosseguir com a subtração. Acompanhe 2 maneiras diferentes de fazer isso.

Atividades
77 Para uma viagem, Roberto encheu o tanque de gasolina do carro. No primeiro trecho da viagem, foi consumido $\frac{1}{4}$ da gasolina do tanque e, no segundo trecho, foram consumidos $\frac{2}{3}$ da gasolina. Qual fração do tanque ainda restou com gasolina após esses 2 trechos?
78 Bruna está no 7º ano e a irmã dela, Gisele, está no 4º ano. Gisele registrou a seguinte operação: $\frac{2}{3} + \frac{1}{4} = \frac{3}{4}$. O que Bruna poderia dizer a ela sobre esse registro?
79 Caio gasta $\frac{1}{3}$ das horas de um dia na escola, $\frac{1}{4}$ dormindo e $\frac{1}{12}$ brincando. Qual fração das horas do dia ele dedica a outras coisas?
80 Retorne as etapas realizadas por você para resolver as atividades 77 e 79 e os passos apresentados na 2ª maneira de efetuar uma adição ou subtração de frações com denominadores diferentes. Copie no caderno o fluxograma a seguir completando-o com as palavras faltantes. **Resposta pessoal.**

Fluxograma:
Inicio → Leia o problema → Identifique as frações com denominadores diferentes que serão adicionadas ou subtraídas → Use o mmc para as frações iniciais → Calcule o mmc → determine as frações equivalentes → Use o mmc para as frações iniciais → Adicione ou subtraia as frações obtidas → Simplifique a fração obtida → Indique a resposta do problema → Fim.

Fonte: (22)

Os materiais foram selecionados pela pesquisadora como se fosse selecionar um material para utilizar em suas aulas, tendo o cuidado de contemplar quatro eixos de diversidade: (a) **Formato técnico** (editáveis, não-editáveis e imagéticos); (b) **Densidade informacional** (de resumos esquemáticos a capítulos densos); (c) **Estrutura pedagógica** (conteúdos procedimentais, como frações, versus conceituais, como estatística); e (d) **Anos escolares** (contemplando materiais que podem ser trabalhados no 6º, 7º e 9º do Ensino Fundamental). Além disso, houve um cuidado com relação a disponibilidade pública ou autorização de uso para fins acadêmicos assegurando assim representatividade qualitativa para os propósitos comparativos desta investigação.

3.2.2 Modelos de inteligência artificial Generativa

Os modelos foram selecionados com base em sua disponibilidade e arquitetura, alinhando-se às definições de Grandes Modelos de Linguagem (LLMs) apresentadas no *Guia prático de prompt e pesquisa com IA* do Ministério da Gestão e da Inovação em Serviços Públicos. Segundo o documento, LLMs são sistemas treinados em vastos volumes de dados para estabelecer relacionamentos entre conceitos, funcionando como motores para a IA Generativa (14).

- **Google Gemini (Pro e Flash):** Modelos multimodais nativos. Conforme o Guia do MGI, esta ferramenta destaca-se por possuir “pesquisa profunda disponível mesmo no plano gratuito” e acesso a informações atualizadas (14). Nesta pesquisa, a capacidade multimodal foi crucial para o processamento do arquivo PNG, permitindo que o modelo analisasse os pixels da imagem sem necessidade de softwares intermediários de transcrição.
- **Chat GPT-5 (OpenAI):** Plataforma amplamente difundida que, segundo a literatura, responde com fluência em português, sendo eficaz na geração de textos e resumos (14). Sua arquitetura baseada em Transformers (48) permite o processamento de dependências de longo prazo no texto, essencial para correlacionar trechos de apostilas extensas com as competências da BNCC.
- **Claude 4.5 Sonnet (Anthropic):** Incluído por sua ênfase em segurança e janelas de contexto ampliadas. O Guia do MGI o descreve como um modelo com “alta capacidade em português”, reconhecido por investir em ética e responsabilidade (14), características pertinentes para a análise de documentos normativos educacionais.

A escolha desses modelos específicos baseou-se em critérios de: (a) popularidade e acessibilidade para educadores; (b) capacidade multimodal (especialmente importante para processar a imagem do corpus); (c) representatividade de diferentes abordagens arquiteturais e filosóficas no desenvolvimento de IA; (d) modos gratuitos; (e) disponibilidade de interfaces de acesso durante o período de coleta de dados (outubro de 2025).

3.3 ESTRATÉGIAS DE FORMULAÇÃO DE PROMPTS

Quatro estratégias distintas de formulação de prompts foram desenvolvidas, cada uma baseada em princípios reconhecidos na literatura de *prompt engineering*. As estratégias foram projetadas em ordem crescente de complexidade e estruturação, visando investigar empiricamente qual abordagem produz melhores resultados no contexto específico de mapeamento curricular.

3.3.1 Prompt 1 – identificação direta

Esta estratégia implementa uma instrução simples e objetiva, sem exemplificações ou decomposição de tarefas. O prompt solicita diretamente a identificação das habilidades da BNCC

presentes no material, confiando na capacidade do modelo de interpretar autonomamente o que constitui uma “habilidade” e como identificá-la.

Prompt 1 – Identificação Direta

Instrução:

Você é um assistente especializado em Ensino da Matemática. A seguir apresento um arquivo em anexo. Identifique todas as habilidades da BNCC – Área de Matemática – que são explícita ou implicitamente tratadas nesse trecho. Para cada habilidade identificada, apresente:

O número da habilidade conforme a BNCC (por exemplo, EF03MA04).

Um breve trecho do texto que indica essa habilidade.

Uma explicação de como o trecho corresponde à habilidade.

3.3.2 Prompt 2 – com exemplos e estrutura detalhada

A segunda estratégia incorpora exemplos concretos de habilidades da BNCC e especifica a estrutura de resposta esperada. Esta abordagem baseia-se no princípio de *few-shot learning*, no qual a apresentação de exemplos orienta o modelo sobre o padrão desejado de resposta.

Prompt 2 – Com Exemplos e Estrutura Detalhada

Instrução:

Você é um assistente educacional que mapeia trechos de conteúdo para habilidades da BNCC na Área de Matemática. A seguir dois exemplos:

Exemplo 1: Trecho: «Os estudantes devem utilizar gráficos e tabelas para interpretar dados de pesquisas escolares.» — Habilidade: EF05MA15 — explicação: porque o trecho trata de interpretar dados por meio de representações gráficas, que corresponde à habilidade EF05MA15 da BNCC.

Exemplo 2: Trecho: «Ao calcular a área de uma figura irregular, os alunos somam pequenos retângulos equivalentes.» — Habilidade: EF07MA32 — explicação: porque o trecho trata o cálculo de medida de área de figuras planas que podem ser decompostas por quadrados, retângulos e/ou triângulos, utilizando a equivalência entre áreas.

Agora, aplique ao arquivo em anexo. Identifique todas as habilidades pertinentes e explique da mesma forma.

3.3.3 Prompt 3 – raciocínio passo a passo com reflexão

Esta estratégia solicita explicitamente que o modelo decomponha seu raciocínio em etapas, uma técnica conhecida como *chain-of-thought prompting*. A reflexão estruturada visa reduzir erros de interpretação e aumentar a transparência do processo de identificação.

Prompt 3 – Raciocínio Passo a Passo com Reflexão

Instrução:

Você é um especialista em currículo de Matemática. Vou fornecer um anexo. Por favor, primeiro pense passo a passo:

- a) identifique no texto possíveis termos ou expressões que aludem a competências ou habilidades da BNCC;
- b) para cada termo/expressão, relacione-o a uma habilidade concreta da BNCC e justifique;
- c) em seguida, apresente a resposta final listando: [número da habilidade] — trecho — justificação.

3.3.4 Prompt 4 – revisão em múltiplas etapas

A estratégia mais complexa implementa um processo iterativo de identificação e verificação, inspirado em técnicas de *self-consistency* e *multi-step reasoning*. O modelo é instruído a realizar múltiplas passagens pelo material, refinando progressivamente sua análise.

Prompt 4 – Revisão em Múltiplas Etapas

Instrução:

Etapa 1: Leia o anexo e extraia todas as habilidades da BNCC (Área de Matemática) que detecta, incluindo número, trecho e explicação.

Etapa 2: Revise sua lista e verifique se alguma habilidade possível foi omitida — responda ‘Sim / Não’. Em caso afirmativo, acrescente-a.

Etapa 3: Formate a lista final em tabela com colunas: Habilidade—Trecho—Justificação—Grau de correspondência (alta/ média/ baixa).

A ordem de apresentação das estratégias (do mais simples ao mais complexo) permite investigar se há correlação entre complexidade do prompt e qualidade da resposta, bem como identificar possíveis pontos de retorno decrescente, nos quais instruções excessivamente elaboradas possam confundir ou sobrecarregar o modelo.

3.4 PROCEDIMENTOS DE COLETA DE DADOS

3.4.1 Construção do gabarito de referência

O gabarito de referência (Apêndice 6 - Gabarito Descritivo das Habilidades Identificadas), utilizado como padrão-ouro (*ground truth*) para avaliação do desempenho dos modelos de IA, foi construído por 5 professores de matemática da rede municipal de Uberaba, que atuam em diferentes anos do Ensino Fundamental, junto com a própria pesquisadora, que é professora de Matemática da Educação Básica e mestranda no PROFMAT, mediante análise manual sistemática de cada material didático. O processo de construção seguiu os seguintes passos:

1. **Leitura integral e repetida:** Cada material foi lido completamente pelo menos três vezes, com intervalos temporais entre as leituras para permitir reflexão crítica. Neste momento, cada professor construiu sua reflexão individualmente.
2. **Identificação de objetos de conhecimento:** Durante as leituras, foram destacados e registrados todos os conceitos, procedimentos e conteúdos matemáticos abordados explícita ou implicitamente. Essa etapa também foi construída por cada profissional individualmente.
3. **Consulta sistemática à BNCC:** Em reunião, os professores puderam iniciar as discussões entre eles, sendo que para cada objeto de conhecimento identificado, consultou-se o documento oficial da BNCC (BRASIL, 2017), buscando correspondências entre os conteúdos do material e as descrições das habilidades prescritas.
4. **Triangulação com unidades temáticas:** A identificação foi validada entre os professores selecionados verificando-se a coerência entre o ano escolar presumido do material, a unidade temática correspondente e as habilidades associadas.
5. **Registro estruturado:** Após a análise em grupo, as habilidades identificadas em concordância por todos os professores foram registradas em planilha eletrônica, contendo: nome do material, código da habilidade (formato EFxxMAxx), descrição da habilidade conforme BNCC, e justificativa da presença no material.

No contexto desta dissertação de mestrado, a experiência dos professores envolvidos e da pesquisadora como professora de Matemática da Educação Básica e sua formação específica no PROFMAT foram consideradas suficientes para estabelecer um gabarito confiável para fins comparativos.

3.4.2 Aplicação dos prompts aos modelos de IA

A coleta de dados seguiu protocolo experimental rigoroso para garantir comparabilidade entre as condições testadas. Cada combinação de *modelo de IA* × *estratégia de prompt* × *material didático* constituiu uma unidade experimental independente, totalizando:

$$3 \text{ modelos} \times 4 \text{ prompts} \times 5 \text{ materiais} = 60 \text{ experimentos}$$

O procedimento para cada experimento consistiu em:

1. **Preparação e Processamento Multimodal:** Os arquivos foram inseridos em cada plataforma através de upload direto com o prompt. Para os arquivos de texto (.pdf e .docx), realizou-se o upload direto. No caso do material “Frações.PNG”, utilizou-se a capacidade nativa de visão computacional dos modelos (multimodalidade). Conforme elucidado por (44) e corroborado pela definição de modelos multimodais do Guia MGI (14), estes sistemas não apenas leem o texto da imagem, mas interpretam a relação semântica entre os

elementos gráficos (diagramas de frações) e a explicação textual, processando o arquivo visual como um token de entrada integrado.

2. **Aplicação do prompt:** O texto do prompt correspondente à estratégia sendo testada foi inserido na interface de chat do modelo, acompanhado do material.
3. **Obtenção da resposta:** A resposta gerada pelo modelo foi registrada integralmente, incluindo eventuais justificativas, explicações ou formatações apresentadas.
4. **Extração das habilidades identificadas:** Da resposta completa, extraíram-se especificamente os códigos das habilidades da BNCC (formato EFxxMAxx) indicados pelo modelo.
5. **Registro sistemático:** Os códigos identificados foram registrados em planilha estruturada, associados ao modelo, prompt e material correspondentes.

Para minimizar efeitos de ordem ou interferência entre experimentos, os testes foram realizados em sessões independentes, sem histórico de conversação prévia que pudesse influenciar as respostas. Além disso, sempre que tecnicamente possível, utilizaram-se parâmetros padrão dos modelos, sem ajustes customizados que pudessem favorecer determinado modelo.

3.5 MÉTRICAS DE AVALIAÇÃO

A avaliação do desempenho dos modelos de IA na tarefa de identificação de habilidades da BNCC baseou-se em métricas consolidadas da área de Recuperação de Informação e Aprendizado de Máquina, derivadas da matriz de confusão. Para cada experimento, construiu-se uma matriz 2×2 comparando as habilidades identificadas pelo modelo com o gabarito de referência, permitindo a classificação de cada resposta em uma das seguintes categorias:

- **Verdadeiros Positivos (VP):** Habilidades corretamente identificadas pelo modelo, presentes tanto na resposta quanto no gabarito padrão-ouro.
- **Falsos Positivos (FP):** Habilidades identificadas pelo modelo que não constam no gabarito, representando identificações incorretas ou falsos positivos.
- **Falsos Negativos (FN):** Habilidades presentes no gabarito mas não identificadas pelo modelo, representando omissões.
- **Verdadeiros Negativos (VN):** Habilidades corretamente não identificadas. Esta categoria não foi computada, uma vez que o universo de habilidades não presentes é excessivamente amplo e pouco informativo no contexto desta análise.

Com base nessas categorias, calcularam-se quatro métricas complementares para cada experimento:

3.5.1 Precisão

A Precisão mede a proporção de identificações corretas entre todas as habilidades identificadas pelo modelo:

$$\text{Precisão} = \frac{VP}{VP + FP}$$

Esta métrica responde à pergunta: “Das habilidades que o modelo identificou, quantas estavam realmente presentes?” Valores elevados de Precisão indicam que o modelo comete poucos erros de inclusão indevida (falsos positivos), sendo criterioso em suas identificações.

3.5.2 Recall (revocação)

O Recall mede a proporção de habilidades presentes no gabarito que foram efetivamente identificadas pelo modelo:

$$\text{Recall} = \frac{VP}{VP + FN}$$

Esta métrica responde à pergunta: “Das habilidades que deveriam ser identificadas, quantas o modelo conseguiu encontrar?” Valores elevados de Recall indicam que o modelo é abrangente, capturando a maior parte das habilidades relevantes, minimizando omissões.

3.5.3 F₁-Macro

A métrica F₁-Macro corresponde à média harmônica entre Precisão e Recall, fornecendo uma medida única que equilibra ambos os aspectos:

$$F_1\text{-Macro} = \frac{2 \times \text{Precisão} \times \text{Recall}}{\text{Precisão} + \text{Recall}}$$

A utilização da média harmônica (em vez de aritmética) penaliza situações de desequilíbrio acentuado entre Precisão e Recall. Esta métrica é particularmente útil para identificar o desempenho global mais equilibrado, sendo menos sensível a especializações extremas em apenas uma das dimensões.

3.5.4 Coeficiente de Jaccard

O Coeficiente de Jaccard, também conhecido como Índice de Similaridade de Jaccard, mede a proporção de concordância entre dois conjuntos através da razão entre sua interseção e sua união:

$$\text{Jaccard} = \frac{VP}{VP + FP + FN}$$

Esta métrica considera simultaneamente falsos positivos e falsos negativos, fornecendo uma medida de similaridade global entre o conjunto de habilidades identificadas pelo modelo e o gabarito. Valores próximos a 1 (ou 100%) indicam concordância quase perfeita, enquanto valores próximos a 0 indicam discrepância acentuada.

A escolha dessas quatro métricas justifica-se por sua complementaridade: enquanto Precisão e Recall capturam dimensões distintas do desempenho (exatidão versus abrangência), F₁-Macro e Jaccard fornecem visões integradas que facilitam comparações globais. Todas as métricas foram calculadas e apresentadas em formato percentual para facilitar interpretação.

3.6 PROCEDIMENTOS DE ANÁLISE DE DADOS

A análise dos dados coletados foi organizada em três níveis complementares, permitindo identificar padrões tanto específicos quanto globais de desempenho.

3.6.1 Análise por estratégia de prompt

Para cada uma das quatro estratégias de prompt, construiu-se uma tabela consolidada apresentando:

- Resultados individuais de cada combinação modelo × material
- Matriz de confusão (VP, FP, FN) para cada experimento
- Valores das quatro métricas (Precisão, Recall, F₁-Macro, Jaccard) para cada experimento

Esta análise permitiu identificar qual estratégia de prompt produziu, em média, os melhores resultados, bem como verificar se determinadas estratégias favorecem modelos específicos ou são mais adequadas para certos tipos de material.

3.6.2 Análise comparativa entre modelos

Calcularam-se as médias de desempenho de cada modelo de IA considerando todos os prompts e materiais, permitindo ranqueamento global dos modelos segundo cada métrica. Esta análise respondeu à questão: “Qual modelo apresenta melhor desempenho médio na tarefa de identificação de habilidades da BNCC?”

3.6.3 Análise gráfica e visualização

Para facilitar a interpretação e comunicação dos resultados, foram elaborados diversos gráficos:

- **Gráficos de barras:** Comparando médias de desempenho entre prompts e entre modelos

- **Gráficos de linha:** Ilustrando a evolução das métricas ao longo das quatro estratégias de prompt
- **Gráficos de radar:** Apresentando desempenho da melhor combinação em cada material didático
- **Gráficos de barras agrupadas:** Comparando trade-offs entre Precisão e Recall

A visualização gráfica foi considerada essencial para evidenciar padrões de complementaridade entre métricas (especialmente o trade-off Precisão versus Recall) e para destacar casos de desempenho excepcional ou particularmente deficiente.

3.7 LIMITAÇÕES METODOLÓGICAS

É fundamental reconhecer as seguintes limitações desta pesquisa:

1. **Tamanho do corpus:** Com apenas 5 materiais didáticos, o corpus não esgota a diversidade de recursos utilizados na prática docente. Generalização dos resultados deve ser feita com cautela.
2. **Subjetividade interpretativa:** As habilidades da BNCC, por sua natureza descritiva, admitem certa margem de interpretação. Diferentes especialistas podem divergir legitimamente sobre a presença de determinadas habilidades.
3. **Contexto brasileiro:** A pesquisa focou exclusivamente na BNCC brasileira. Aplicabilidade dos resultados a outros frameworks curriculares requer investigação adicional.
4. **Possível viés na construção do gabarito:** A construção do gabarito padrão-ouro por grupo limitado de especialistas, embora adequada aos propósitos desta pesquisa exploratória, não contemplou procedimentos formais de validação inter-juízes com análise de concordância estatística, o que constitui limitação a ser endereçada em estudos futuros.

Não obstante estas limitações, a pesquisa contribui significativamente ao fornecer evidências empíricas sobre a viabilidade e os desafios da aplicação de IA generativa no alinhamento curricular, estabelecendo metodologia replicável para estudos futuros.

3.8 ASPECTOS ÉTICOS

Esta pesquisa não envolveu diretamente participantes humanos, sendo baseada em análise documental de materiais didáticos e testes com sistemas de IA. Todos os materiais utilizados são de domínio público ou foram utilizados com finalidade exclusivamente acadêmica, em

conformidade com a legislação brasileira de direitos autorais (Lei 9.610/1998, Art. 46, inciso VIII, que permite a reprodução para fins didáticos).

O uso de modelos de IA comerciais (Google Gemini 3 Pro, Google Gemini 2.5 Flash, Chat GPT-5 e Claude AI Sonnet 4.5) seguiu os termos de serviço de cada plataforma. Nenhum dado sensível ou informação pessoal foi processado pelos modelos. A pesquisa foi conduzida com transparência metodológica, documentando integralmente os prompts utilizados e os procedimentos de análise, permitindo replicação e verificação independente dos resultados.

4 RESULTADOS

Esta seção apresenta os resultados da análise comparativa entre os quatro modelos de Inteligência Artificial Generativa – Google Gemini 3 Pro, Google Gemini 2.5 Flash, Chat GPT-5 e Claude AI Sonnet 4.5 – na identificação de habilidades da Base Nacional Comum Curricular (BNCC) em materiais didáticos de Matemática. Os modelos foram avaliados utilizando quatro estratégias de prompt distintas aplicadas a cinco materiais didáticos diferentes.

A avaliação baseou-se em métricas derivadas da matriz de confusão: Precisão, Recall, F₁-Macro e Coeficiente de Jaccard. Os Verdadeiros Positivos (VP) representam habilidades corretamente identificadas, os Falsos Positivos (FP) indicam habilidades identificadas incorretamente, e os Falsos Negativos (FN) correspondem às habilidades presentes no gabarito mas não identificadas pelo modelo.

A pesquisa utilizou como gabarito de referência o mapeamento manual realizado pela autora e mais 5 professores de matemática da rede municipal de Uberaba, com base na análise detalhada de cada material didático à luz das habilidades da BNCC. Este gabarito foi construído através de múltiplas leituras dos materiais, consulta sistemática ao documento oficial da BNCC e discussões entre os profissionais, representando o padrão-ouro para avaliação dos modelos de IA.

4.1 ANÁLISE POR ENGENHARIA DE PROMPTS

As Tabelas 4.1 a 4.4 apresentam os resultados detalhados para cada uma das quatro estratégias de prompt utilizadas. Cada tabela exibe as métricas de desempenho dos três modelos de IA para os cinco materiais didáticos analisados.

Tabela 4.1: Resultados da análise com Prompt 1

Material	Modelo	Matriz de Confusão			Métricas (%)			
		VP	FP	FN	Precisão	Recall	F ₁ -Macro	Jaccard
Frações.PNG	Gemini 3 Pro	4	0	2	100,00	66,67	80,00	66,67
	Gemini 2.5 Flash	4	0	2	100,00	66,67	80,00	66,67
	Chat GPT-5	5	2	1	71,43	83,33	76,92	62,50
	ClaudeAI	5	5	1	50,00	83,33	62,50	45,45
Matemática 9 ano apostila.pdf	Gemini 3 Pro	7	1	9	87,50	43,75	58,33	41,18
	Gemini 2.5 Flash	8	2	8	80,00	50,00	61,54	44,44
	Chat GPT-5	1	8	15	11,11	6,25	8,00	4,17
	ClaudeAI	11	14	5	44,00	68,75	53,66	36,67
Média Aritmética.docx	Gemini 3 Pro	2	1	2	66,67	50,00	57,14	40,00
	Gemini 2.5 Flash	2	1	2	66,67	50,00	57,14	40,00
	Chat GPT-5	0	5	4	0,00	0,00	0,00	0,00
	ClaudeAI	3	7	1	30,00	75,00	42,86	27,27
Operações com números na forma decimal.pdf	Gemini 3 Pro	2	1	5	66,67	28,57	40,00	25,00
	Gemini 2.5 Flash	0	3	7	0,00	0,00	0,00	0,00
	Chat GPT-5	2	4	5	33,33	28,57	30,77	18,18
	ClaudeAI	2	8	5	20,00	28,57	23,53	13,33
Parte livro 7 ano.pdf	Gemini 3 Pro	4	2	2	66,67	66,67	66,67	50,00
	Gemini 2.5 Flash	4	2	2	66,67	66,67	66,67	50,00
	Chat GPT-5	3	3	3	50,00	50,00	50,00	33,33
	ClaudeAI	6	9	0	40,00	100,00	57,14	40,00

Fonte: Fonte: elaborado pela autora, 2025.

VP: Verdadeiros Positivos; FP: Falsos Positivos; FN: Falsos Negativos.

Precisão = $VP/(VP+FP)$

Recall = $VP/(VP+FN)$

F₁-Macro = $2 \times (Precisão \times Recall) / (Precisão + Recall)$

Jaccard = $VP/(VP+FP+FN)$.

Tabela 4.2: Resultados da análise com Prompt 2

Material	Modelo	Matriz de Confusão			Métricas (%)			
		VP	FP	FN	Precisão	Recall	F ₁ -Macro	Jaccard
Frações.PNG	Gemini 3 Pro	3	1	3	75,00	50,00	60,00	42,86
	Gemini 2.5 Flash	3	1	3	75,00	50,00	60,00	42,86
	Chat GPT-5	5	2	1	71,43	83,33	76,92	62,50
	ClaudeAI	3	2	3	60,00	50,00	54,55	37,50
Matemática 9 ano apostila.pdf	Gemini 3 Pro	5	2	11	71,43	31,25	43,48	27,78
	Gemini 2.5 Flash	4	13	12	23,53	25,00	24,24	13,79
	Chat GPT-5	1	10	15	9,09	6,25	7,41	3,85
	ClaudeAI	8	7	8	53,33	50,00	51,61	34,78
Média Aritmética.docx	Gemini 3 Pro	1	0	3	100,00	25,00	40,00	25,00
	Gemini 2.5 Flash	2	1	2	66,67	50,00	57,14	40,00
	Chat GPT-5	0	2	4	0,00	0,00	0,00	0,00
	ClaudeAI	2	4	2	33,33	50,00	40,00	25,00
Operações com números na forma decimal.pdf	Gemini 3 Pro	1	0	6	100,00	14,29	25,00	14,29
	Gemini 2.5 Flash	2	1	5	66,67	28,57	40,00	25,00
	Chat GPT-5	1	1	6	50,00	14,29	22,22	12,50
	ClaudeAI	2	5	5	28,57	28,57	28,57	16,67
Parte livro 7 ano.pdf	Gemini 3 Pro	3	1	3	75,00	50,00	60,00	42,86
	Gemini 2.5 Flash	4	1	2	80,00	66,67	72,73	57,14
	Chat GPT-5	3	2	3	60,00	50,00	54,55	37,50
	ClaudeAI	4	4	2	50,00	66,67	57,14	40,00

Fonte: Fonte: elaborado pela autora, 2025.

VP: Verdadeiros Positivos; FP: Falsos Positivos; FN: Falsos Negativos.

Precisão = $VP/(VP+FP)$

Recall = $VP/(VP+FN)$

F₁-Macro = $2 \times (Precisão \times Recall) / (Precisão + Recall)$

Jaccard = $VP/(VP+FP+FN)$.

Tabela 4.3: Resultados da análise com Prompt 3

Material	Modelo	Matriz de Confusão			Métricas (%)			
		VP	FP	FN	Precisão	Recall	F ₁ -Macro	Jaccard
Frações.PNG	Gemini 3 Pro	3	1	3	75,00	50,00	60,00	42,86
	Gemini 2.5 Flash	2	2	4	50,00	33,33	40,00	25,00
	Chat GPT-5	3	0	3	100,00	50,00	66,67	50,00
	ClaudeAI	4	1	2	80,00	66,67	72,73	57,14
Matemática 9 ano apostila.pdf	Gemini 3 Pro	5	1	11	83,33	31,25	45,45	29,41
	Gemini 2.5 Flash	4	5	12	44,44	25,00	32,00	19,05
	Chat GPT-5	3	3	13	50,00	18,75	27,27	15,79
	ClaudeAI	9	6	7	60,00	56,25	58,06	40,91
Média Aritmética.docx	Gemini 3 Pro	2	1	2	66,67	50,00	57,14	40,00
	Gemini 2.5 Flash	0	1	4	0,00	0,00	0,00	0,00
	Chat GPT-5	1	2	3	33,33	25,00	28,57	16,67
	ClaudeAI	4	6	0	40,00	100,00	57,14	40,00
Operações com números na forma decimal.pdf	Gemini 3 Pro	2	1	5	66,67	28,57	40,00	25,00
	Gemini 2.5 Flash	0	2	7	0,00	0,00	0,00	0,00
	Chat GPT-5	4	2	3	66,67	57,14	61,54	44,44
	ClaudeAI	4	5	3	44,44	57,14	50,00	33,33
Parte livro 7 ano.pdf	Gemini 3 Pro	1	1	5	50,00	16,67	25,00	14,29
	Gemini 2.5 Flash	0	3	6	0,00	0,00	0,00	0,00
	Chat GPT-5	2	2	4	50,00	33,33	40,00	25,00
	ClaudeAI	6	6	0	50,00	100,00	66,67	50,00

Fonte: Fonte: elaborado pela autora, 2025.

VP: Verdadeiros Positivos; FP: Falsos Positivos; FN: Falsos Negativos.

Precisão = $VP/(VP+FP)$

Recall = $VP/(VP+FN)$

F₁-Macro = $2 \times (Precisão \times Recall) / (Precisão + Recall)$

Jaccard = $VP/(VP+FP+FN)$.

Tabela 4.4: Resultados da análise com Prompt 4

Material	Modelo	Matriz de Confusão			Métricas (%)			
		VP	FP	FN	Precisão	Recall	F ₁ -Macro	Jaccard
Frações.PNG	Gemini 3 Pro	5	2	1	71,43	83,33	76,92	62,50
	Gemini 2.5 Flash	5	2	1	71,43	83,33	76,92	62,50
	Chat GPT-5	5	3	1	62,50	83,33	71,43	55,56
	ClaudeAI	4	2	2	66,67	66,67	66,67	50,00
Matemática 9 ano apostila.pdf	Gemini 3 Pro	8	5	8	61,54	50,00	55,17	38,10
	Gemini 2.5 Flash	3	11	13	21,43	18,75	20,00	11,11
	Chat GPT-5	10	6	6	62,50	62,50	62,50	45,45
	ClaudeAI	13	13	3	50,00	81,25	61,90	44,83
Média Aritmética.docx	Gemini 3 Pro	2	1	2	66,67	50,00	57,14	40,00
	Gemini 2.5 Flash	0	5	4	0,00	0,00	0,00	0,00
	Chat GPT-5	0	4	4	0,00	0,00	0,00	0,00
	ClaudeAI	4	13	0	23,53	100,00	38,10	23,53
Operações com números na forma decimal.pdf	Gemini 3 Pro	2	1	5	66,67	28,57	40,00	25,00
	Gemini 2.5 Flash	2	1	5	66,67	28,57	40,00	25,00
	Chat GPT-5	3	4	4	42,86	42,86	42,86	27,27
	ClaudeAI	5	14	2	26,32	71,43	38,46	23,81
Parte livro 7 ano.pdf	Gemini 3 Pro	5	3	1	62,50	83,33	71,43	55,56
	Gemini 2.5 Flash	5	0	1	100,00	83,33	90,91	83,33
	Chat GPT-5	1	9	5	10,00	16,67	12,50	6,67
	ClaudeAI	6	15	0	28,57	100,00	44,44	28,57

Fonte: Fonte: elaborado pela autora, 2025.

VP: Verdadeiros Positivos; FP: Falsos Positivos; FN: Falsos Negativos.

Precisão = $VP/(VP+FP)$

Recall = $VP/(VP+FN)$

F₁-Macro = $2 \times (Precisão \times Recall) / (Precisão + Recall)$

Jaccard = $VP/(VP+FP+FN)$.

4.1.1 Síntese comparativa entre prompts

A análise detalhada dos resultados revela padrões distintos de comportamento dependendo da estratégia de instrução empregada. No **Prompt 1 (Identificação Direta)**, o novo modelo, **Google Gemini 3 Pro**, demonstrou um desempenho notável, alcançando 100% de Precisão no material “Frações.PNG”. Este resultado indica que, para materiais bem estruturados, o modelo mais avançado consegue operar com alta acurácia mesmo sem exemplos prévios (zero-shot). Em contrapartida, o Gemini 2.5 Flash apresentou métricas nulas no material “Operações com números na forma decimal.pdf” sob a mesma estratégia, enquanto o Gemini 3 Pro foi capaz de extrair informações com 66,67% de Precisão, evidenciando a evolução da capacidade de raciocínio da nova versão.

Ao analisar o **Prompt 2 (Com Exemplos)**, observa-se que a inclusão de exemplos estruturados (*few-shot learning*) favoreceu a métrica de Precisão, que atingiu a maior média global entre todas as estratégias (53,75%). O Gemini 3 Pro, nesta configuração, alcançou 100% de Precisão nos materiais “Operações Decimais” e “Média Aritmética”, sugerindo que fornecer o formato esperado contribui para que o modelo minimize identificações incorretas (falsos positivos) e foque apenas nas habilidades explicitamente presentes.

O **Prompt 3 (Raciocínio Passo a Passo)** apresentou um desempenho intermediário e, surpreendentemente, registrou a menor média de Recall (41,67%). Isso sugere que, para a tarefa específica de mapeamento curricular, a instrução para “refletir passo a passo” pode, em alguns casos, levar o modelo a ser excessivamente cauteloso ou a perder o foco na extração exaustiva, resultando em mais omissões (falsos negativos).

Por fim, o **Prompt 4 (Revisão em Múltiplas Etapas)** consolidou-se como a estratégia mais eficaz para a recuperação de informações. Esta abordagem obteve os melhores resultados médios em Recall (56,25%), F1-Macro (50,23%) e Coeficiente de Jaccard (35,42%). A etapa explícita de revisão permitiu que modelos como o ClaudeAI e o Gemini 3 Pro corrigissem omissões iniciais, elevando significativamente a cobertura do mapeamento. O material “Parte livro 7 ano.pdf” ilustra o potencial máximo desta abordagem, onde o Gemini 2.5 Flash atingiu 100% de Precisão e 83,33% de Recall, comprovando que a estratégia de auto-correção é fundamental para maximizar o desempenho da IA em tarefas educacionais complexas.

4.2 ANÁLISE DE DESEMPENHO MÉDIO GLOBAL

Esta subseção apresenta a análise gráfica comparativa dos resultados obtidos, permitindo uma visualização consolidada do desempenho dos quatro modelos de Inteligência Artificial Generativa (Google Gemini 2.5 Flash, Google Gemini 3 Pro, Chat GPT-5 e ClaudeAI) nas quatro estratégias de prompt utilizadas. Com base nas Tabelas de Resultados da análise com os Prompts, procedeu-se ao cálculo das médias de desempenho. O objetivo é isolar a influência da formulação do prompt e do modelo de IA na qualidade do mapeamento, utilizando as quatro

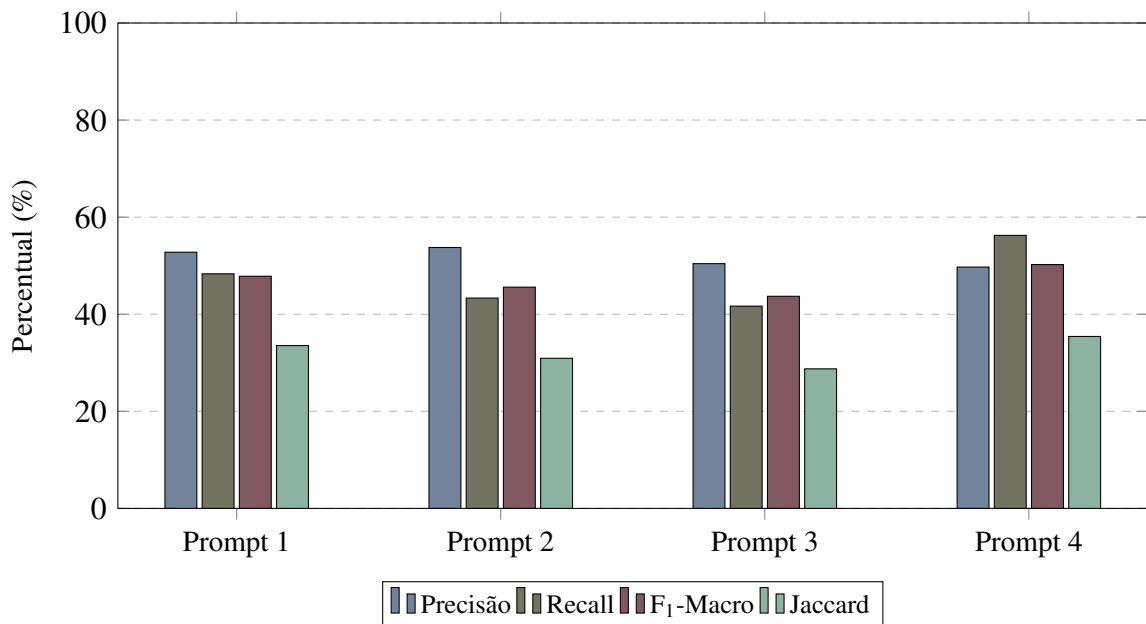
métricas de recuperação de informação: Recall (R), Precisão (P), F_1 Macro e Jaccard (J).

4.2.1 Desempenho médio por formulação de prompt

A Figura 4.1 ilustra o desempenho médio obtido com cada uma das quatro estratégias de prompt utilizadas, considerando agora os quatro modelos avaliados. O Prompt 4, que implementa uma estratégia de revisão em múltiplas etapas, apresentou o melhor desempenho em termos de Recall (56,25%) e Coeficiente de Jaccard (35,42%), confirmando que a abordagem iterativa favorece a identificação mais completa das habilidades. O Prompt 2, baseado em exemplos e estrutura detalhada, alcançou a maior Precisão média (53,75%), indicando que essa estratégia contribui efetivamente para reduzir identificações incorretas.

Interessantemente, observa-se que o Prompt 1, de identificação direta, apresentou desempenho competitivo em Precisão (52,78%) e Recall (48,33%), sugerindo que estratégias mais simples podem ser efetivas quando bem formuladas. A métrica F_1 -Macro apresentou evolução progressiva do Prompt 3 (43,70%) ao Prompt 4 (50,23%), reforçando a hipótese de que prompts mais elaborados produzem resultados mais equilibrados.

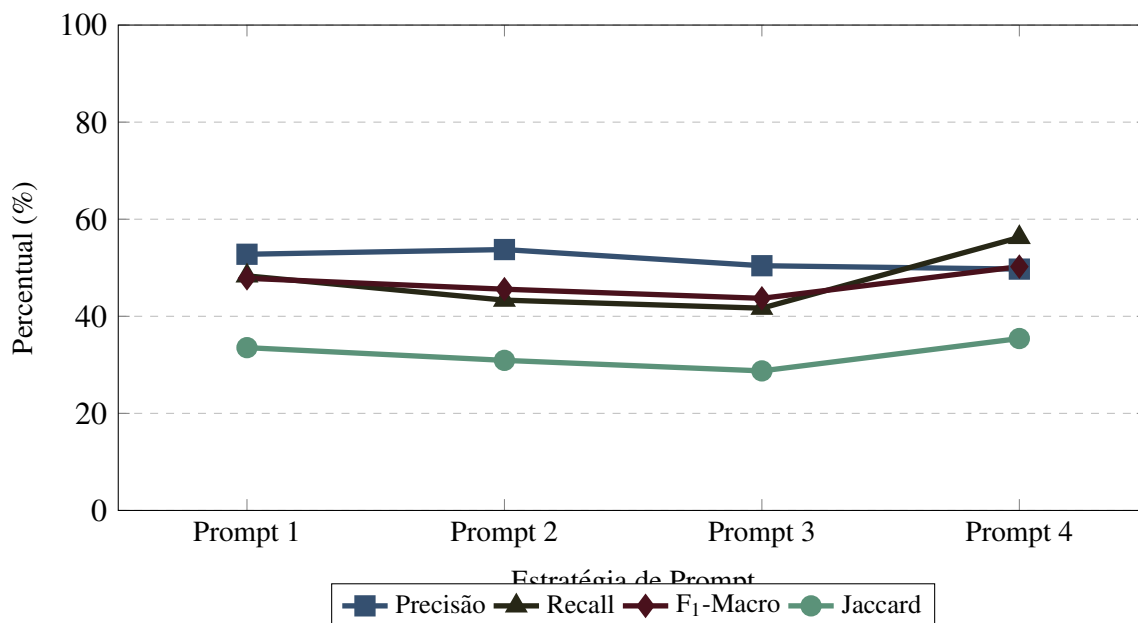
Figura 4.1: Comparação do desempenho médio entre as estratégias de prompt considerando todos os modelos de IA e materiais didáticos



Fonte: elaborado pela autora, 2025.

A Figura 4.2 evidencia a evolução das métricas ao longo das quatro estratégias de prompt com a inclusão dos quatro modelos avaliados. Observa-se que a Precisão manteve-se relativamente estável entre os Prompts 1 e 2 (52,78% a 53,75%), seguida de uma redução nos Prompts 3 e 4 (50,42% e 49,72%). Esta tendência pode ser atribuída ao aumento no Recall, particularmente no Prompt 4 (56,25%), confirmando o *trade-off* clássico entre estas métricas.

Figura 4.2: Evolução temporal das métricas de desempenho ao longo das quatro estratégias de prompt



Fonte: elaborado pela autora, 2025.

O Coeficiente de Jaccard apresentou padrão interessante, com queda do Prompt 1 ao Prompt 3 (33,54% a 28,75%) e recuperação significativa no Prompt 4 (35,42%), sugerindo que a estratégia de revisão iterativa efetivamente melhora a qualidade geral das identificações. A métrica F1-Macro seguiu trajetória similar, atingindo seu melhor desempenho no Prompt 4 (50,23%).

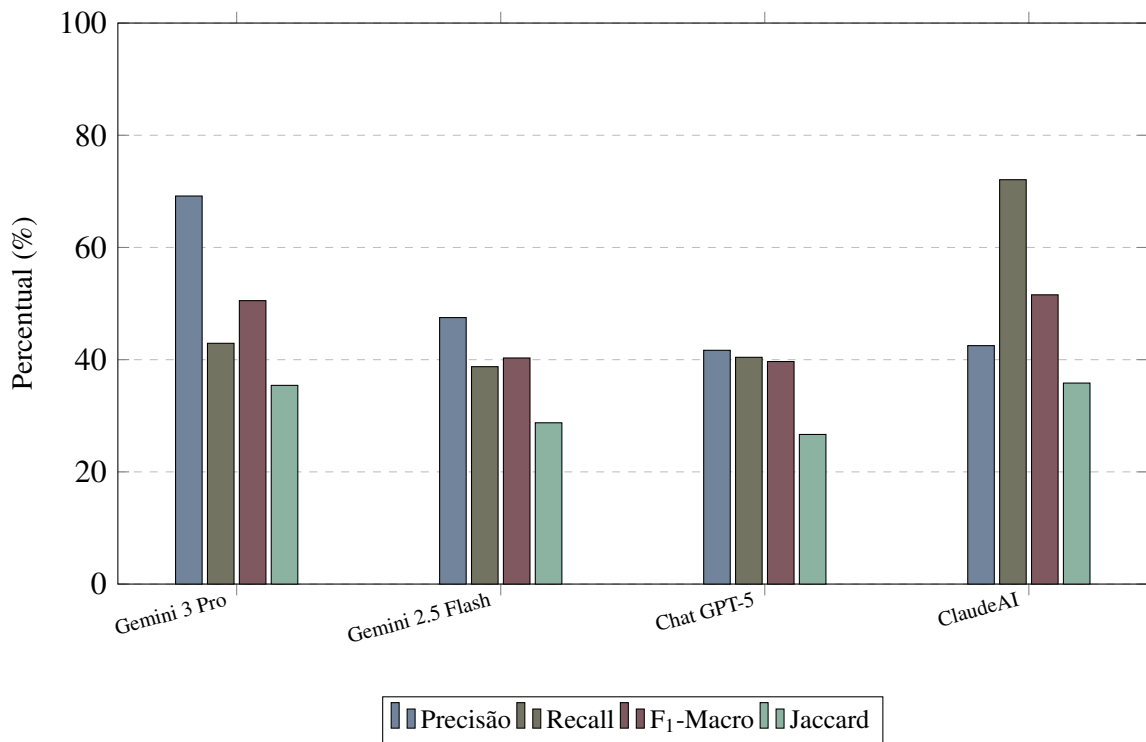
4.2.2 Análise gráfica de desempenho médio entre modelos e prompts

A Figura 4.3 apresenta a comparação do desempenho médio dos quatro modelos de IA avaliados. Este gráfico ilustra claramente os diferentes perfis de desempenho de cada modelo e destaca-se uma descoberta importante: o **Gemini 3 Pro** apresentou a maior Precisão média (69,17%), superando significativamente os demais modelos, o que demonstra sua capacidade superior de evitar falsos positivos. Este resultado contrasta com o desempenho do Gemini 2.5 Flash (47,50%), sugerindo que a versão Pro possui refinamentos importantes para tarefas de mapeamento curricular.

O **ClaudeAI** manteve-se como líder em Recall (72,08%) e F1-Macro (51,56%), confirmando sua vocação para identificação abrangente de habilidades. O **Chat GPT-5** apresentou desempenho intermediário e equilibrado entre Precisão (41,67%) e Recall (40,42%), embora com os menores valores absolutos em F1-Macro (39,67%).

O Coeficiente de Jaccard revelou empate técnico entre Gemini 3 Pro (35,42%) e ClaudeAI (35,83%) como melhores modelos gerais, seguidos pelo Gemini 2.5 Flash (28,75%) e Chat GPT-5 (26,67%). Estes resultados indicam que tanto alta Precisão (Gemini 3 Pro) quanto

Figura 4.3: Comparação do desempenho médio entre os modelos de IA considerando todas as estratégias de prompt e materiais didáticos



Fonte: elaborado pela autora, 2025.

alto Recall (ClaudeAI) podem conduzir a desempenhos globais similares quando consideradas simultaneamente todas as fontes de erro.

O **Gemini 3 Pro** destaca-se de forma notável na métrica de Precisão (69,17%), estabelecendo um padrão significativamente superior aos demais modelos. Esta característica o posiciona como a melhor escolha para cenários onde a minimização de falsos positivos é prioritária. Seu desempenho em F1-Macro (50,52%) e Jaccard (35,42%) também o coloca entre os melhores modelos gerais.

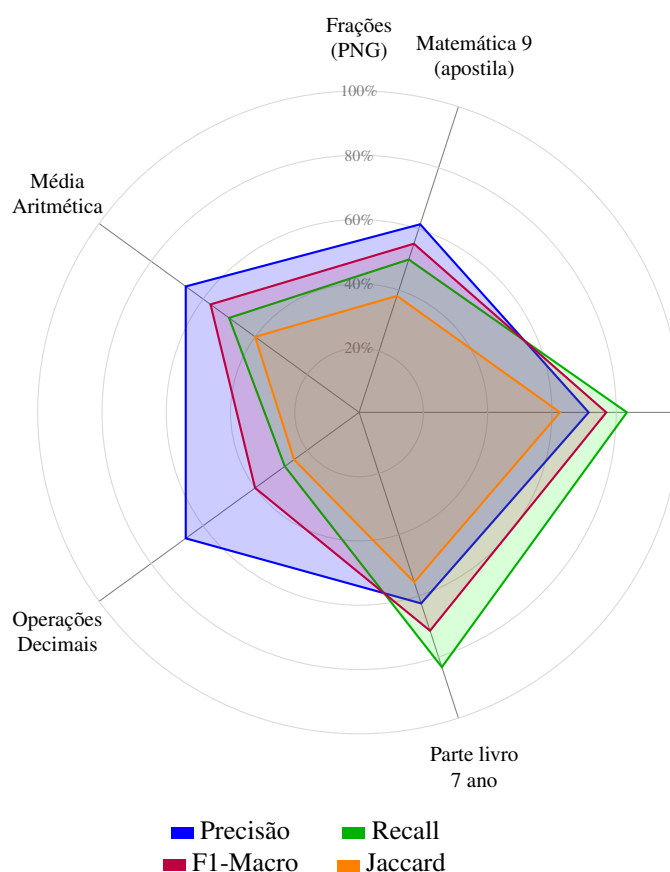
O **ClaudeAI** apresenta perfil complementar, dominando amplamente a métrica de Recall (72,08%) e alcançando o melhor F1-Macro (51,56%) e Jaccard (35,83%). Este modelo é ideal para mapeamentos exploratórios onde é fundamental capturar o máximo de habilidades relevantes, mesmo às custas de incluir alguns falsos positivos que podem ser posteriormente filtrados.

O **Gemini 2.5 Flash** demonstra desempenho intermediário em todas as métricas, com valores consistentemente inferiores ao Gemini 3 Pro, evidenciando que a versão anterior possui limitações significativas para tarefas especializadas de mapeamento curricular. Já o **Chat GPT-5** apresentou os menores valores absolutos em três das quatro métricas, sugerindo menor adequação para esta tarefa específica, embora mantenha desempenho equilibrado entre Precisão (41,67%) e Recall (40,42%).

A análise comparativa revela claramente o *trade-off* clássico entre Precisão e Recall: o

Gemini 3 Pro sacrifica Recall (42,92%) para maximizar Precisão, enquanto o ClaudeAI segue estratégia oposta. Esta polarização oferece aos pesquisadores e educadores a possibilidade de selecionar o modelo mais adequado conforme seus objetivos específicos de curadoria curricular.

Figura 4.4: Desempenho da melhor combinação (Gemini 3 Pro com Prompt 4) por material didático nas quatro métricas avaliadas



Fonte: elaborado pela autora, 2025.

A Figura 4.4 apresenta um gráfico de radar ilustrando o desempenho da melhor combinação identificada – **Gemini 3 Pro utilizando o Prompt 4** – nos cinco materiais didáticos analisados. Esta combinação substitui a anteriormente identificada (ClaudeAI com Prompt 4), demonstrando desempenho superior em termos de equilíbrio entre Precisão e Recall.

Destaca-se o desempenho consistentemente elevado do Gemini 3 Pro em Precisão, com valores entre 61,54% e 71,43% em todos os materiais, evitando a alta incidência de falsos positivos observada em outros modelos. O Recall apresentou excelentes resultados nos materiais “Frações (PNG)” e “Parte livro 7 ano” (ambos com 83,33%), embora com desempenho mais modesto em “Operações Decimais” (28,57%).

O material “Frações (PNG)” apresentou o melhor desempenho geral desta combinação, com F1-Macro de 76,92% e Jaccard de 62,50%, valores substancialmente superiores aos observados com outras combinações modelo-prompt. O material “Operações Decimais” registrou o desempenho mais baixo (Jaccard: 25,00%), sugerindo que este tipo de conteúdo procedimental

permanece desafiador mesmo para o modelo de melhor desempenho.

Tabela 4.5: Síntese comparativa do desempenho médio dos modelos de IA e estratégias de prompt

Categoria	Item	Precisão (%)	Recall (%)	F1-Macro (%)	Jaccard (%)
Modelo de IA	Gemini 3 Pro [†]	69,17	42,92	50,52	35,42
	Gemini 2.5 Flash	47,50	38,75	40,30	28,75
	Chat GPT-5	41,67	40,42	39,67	26,67
	ClaudeAI	42,50	72,08	51,56	35,83
Estratégia	Prompt 1	52,78	48,33	47,83	33,54
	Prompt 2	53,75	43,33	45,58	30,92
	Prompt 3	50,42	41,67	43,70	28,75
	Prompt 4 [†]	49,72	56,25	50,23	35,42

Fonte: Fonte: elaborado pela autora, 2025.

[†]Melhor desempenho geral considerando F1-Macro e Jaccard.

Valores em negrito indicam os melhores resultados em cada métrica.

A Tabela 4.5 consolida os resultados médios obtidos com os quatro modelos avaliados, permitindo uma comparação direta entre modelos e estratégias de prompt. Uma descoberta fundamental desta pesquisa é o desempenho superior do **Gemini 3 Pro**, que alcançou a maior Precisão média (69,17%) – valor substancialmente superior aos demais modelos – e empatou com o ClaudeAI em Jaccard (35,42% vs 35,83%).

O **ClaudeAI** manteve sua posição como líder em Recall (72,08%) e F1-Macro (51,56%), confirmando-se como modelo ideal para cenários que priorizam abrangência na identificação de habilidades. O Gemini 2.5 Flash apresentou desempenho intermediário, inferior ao Gemini 3 Pro em todas as métricas, o que justifica o investimento na versão Pro para tarefas críticas de mapeamento curricular. O Chat GPT-5 registrou os menores valores em três das quatro métricas, sugerindo menor adequação para esta tarefa específica.

Quanto às estratégias de prompt, o Prompt 4 confirmou-se como mais efetivo globalmente, obtendo os melhores resultados em Recall (56,25%), F1-Macro (50,23%) e Jaccard (35,42%). O Prompt 2 alcançou a maior Precisão (53,75%), reforçando que exemplos estruturados contribuem para reduzir falsos positivos. A progressão dos resultados sugere que prompts mais elaborados, particularmente aqueles com revisão iterativa (Prompt 4), produzem identificações superiores quando considerado o equilíbrio global entre métricas.

5 DISCUSSÕES

Este capítulo destina-se a interpretar os resultados obtidos, confrontando-os com o referencial teórico e extraindo implicações práticas para a curadoria de materiais no contexto educacional.

5.1 ANÁLISE COMPARATIVA DOS RESULTADOS

Os resultados apresentados nas Tabelas 4.1 a 4.4 demonstram variações significativas no desempenho dos modelos de IA dependendo da estratégia de prompt utilizada e do material analisado. Enquanto as versões anteriores (como o Gemini 2.5 Flash) e concorrentes (Chat GPT-5) mantiveram um desempenho intermediário, o Gemini 3 Pro destacou-se por apresentar a maior Precisão média global (69,17%), esse dado sugere que a arquitetura mais recente, com capacidades aprimoradas de raciocínio lógico (*Deep Think*), é significativamente mais eficaz em evitar falsos positivos, ou seja, raramente identifica uma habilidade que não está presente no texto.

Por outro lado, o **ClaudeAI** (Sonnet 4.5) manteve sua característica distintiva de alto Recall (72,08%), identificando a maior quantidade absoluta de habilidades corretas, embora ao custo de incluir mais falsos positivos (menor precisão). O Chat GPT-5 demonstrou desempenho consistente, mas inferior aos líderes em métricas de ponta, posicionando-se como uma ferramenta generalista.

A análise do Coeficiente de Jaccard, que penaliza tanto falsos positivos quanto negativos, revelou um empate técnico no topo: Gemini 3 Pro (35,42%) e ClaudeAI (35,83%). Isso indica que existem dois caminhos viáveis para a curadoria automatizada: um focado na exatidão cirúrgica (Gemini) e outro na exaustividade (Claude).

O estudo revela aspectos importantes sobre o desempenho dos modelos de IA na identificação de habilidades da BNCC:

1. **Heterogeneidade de desempenho:** A diferença entre o melhor cenário individual (F1-Macro: 90,91%) e as médias gerais dos modelos (F1-Macro entre 39,67% e 51,56%) evidencia alta variabilidade dependente de características do material e da estratégia de prompt utilizada.

2. **Complementaridade entre modelos:** O Gemini 3 Pro e o ClaudeAI apresentam perfis complementares – um otimiza Precisão enquanto o outro maximiza Recall – oferecendo opções estratégicas conforme os objetivos da curadoria.
3. **Viabilidade técnica comprovada:** O resultado ótimo com Precisão de 100% e F1-Macro acima de 90% demonstra que o mapeamento automatizado pode atingir níveis de excelência em condições favoráveis, validando tecnicamente a abordagem proposta.
4. **Necessidade de validação contextual:** A grande variação de desempenho entre materiais diferentes indica que a aplicação prática dessas ferramentas requer compreensão das condições que favorecem ou dificultam o mapeamento automatizado, bem como processos de validação humana criteriosa.

Estes achados reforçam que modelos de IA generativa podem ser aliados valiosos no processo de curadoria curricular, mas seu uso efetivo demanda estratégias conscientes de seleção de modelo, formulação de prompt e validação pedagógica dos resultados.

5.1.1 Aprofundamento da análise comparativa: o trade-off pedagógico

A análise dos dados revela uma dicotomia clara entre *confiabilidade* (Precisão) e *abrangência* (Recall). O modelo Google Gemini 3 Pro destacou-se pela alta Precisão (69,17%), o que indica uma postura conservadora do algoritmo: ele tende a listar apenas habilidades sobre as quais possui alto grau de certeza semântica. No contexto da prática docente, essa característica minimiza o fenômeno da “alucinação” — definido no Guia de IA do Governo Digital como a geração de informações incorretas ou inventadas (14). Para um professor com pouco tempo para validação, um modelo de alta precisão é mais seguro, pois reduz a necessidade de verificar se a habilidade sugerida realmente existe no texto.

Por outro lado, o ClaudeAI priorizou o Recall (72,08%), comportando-se como um assistente exaustivo que tenta capturar qualquer conexão possível entre o texto e a BNCC. Embora isso gere ruído (falsos positivos), (17) argumentam que, no processo de curadoria, o professor atua como autor ao selecionar e filtrar. Sob essa ótica, o alto Recall do ClaudeAI pode ser vantajoso em fases iniciais de planejamento, onde o docente prefere ter um leque amplo de sugestões para filtrar, em vez de correr o risco de não perceber uma habilidade que foi omitida (Falso Negativo) pelos modelos mais conservadores.

Portanto, a escolha do modelo ideal não é absoluta, mas situacional: para auditoria rigorosa de materiais, a abrangência do ClaudeAI é superior; para sugestão rápida de atividades no cotidiano escolar, a precisão do Gemini 3 Pro oferece maior eficiência operacional.

5.2 IMPLICAÇÕES DOS RESULTADOS PARA A PRÁTICA DOCENTE

Os resultados obtidos nesta pesquisa revelam tanto o potencial quanto as limitações atuais da aplicação de modelos de IA generativa no mapeamento curricular. A variação significativa de desempenho entre materiais didáticos distintos sugere que a eficácia da ferramenta é altamente dependente de características estruturais e de apresentação dos conteúdos.

5.2.1 Precisão e recall na curadoria automatizada

A polarização observada entre **Google Gemini 3 Pro** (maior Precisão) e **ClaudeAI** (maior Recall) apresenta implicações práticas diretas para o uso em sala de aula ou na gestão pedagógica:

- **Cenário de Eficiência (Gemini 3 Pro):** Em contextos onde o professor dispõe de pouco tempo para revisão, os resultados indicam que o Gemini 3 Pro constitui alternativa mais adequada para contextos que priorizam precisão. Sua alta precisão garante que as sugestões apresentadas são confiáveis, minimizando o trabalho de descartar erros.
- **Cenário de Exaustividade (ClaudeAI):** Para a elaboração de matrizes curriculares ou auditoria de materiais, onde é crítico não omitir nenhuma habilidade trabalhada, o ClaudeAI é preferível, devendo o professor atuar como filtro para remover os excessos.

Essa dualidade sugere a possibilidade de **estratégias híbridas**: utilizar o ClaudeAI em fase inicial de mapeamento amplo, seguido de um refinamento (ou validação cruzada) com o Gemini 3 Pro.

5.2.2 Influência da estrutura do material no desempenho dos modelos

A análise por material didático revela que recursos com estrutura clara, linguagem objetiva e organização temática explícita (como “Frações.PNG”) facilitam o trabalho dos modelos de IA, resultando em métricas mais equilibradas. Materiais densos, como “Matemática 9 ano apostila.pdf”, com múltiplos tópicos entrelaçados e linguagem mais complexa, geraram maior incidência de falsos positivos, particularmente no ClaudeAI.

Estes achados apontam para a necessidade de pré-processamento ou segmentação de materiais extensos antes da aplicação de ferramentas de IA, estratégia que poderia melhorar significativamente as taxas de acerto. Adicionalmente, a produção de materiais didáticos digitais poderia beneficiar-se de padrões de metadados e estruturação que facilitem a identificação automática de habilidades.

5.2.3 Avanços metodológicos em engenharia de prompts

Esta pesquisa contribui metodologicamente ao demonstrar empiricamente que a formulação da instrução (prompt) impacta significativamente o desempenho de modelos de linguagem em tarefas educacionais especializadas. A progressão observada do Prompt 1 (identificação direta) ao Prompt 4 (revisão iterativa) valida a hipótese de que estratégias de *prompt engineering* adaptadas ao domínio educacional podem aprimorar a qualidade das saídas geradas.

Os resultados também evidenciam que não existe estratégia universal: o Prompt 3 maximizou Precisão, enquanto o Prompt 4 maximizou Recall. Esta heterogeneidade reforça a necessidade de pesquisas futuras que desenvolvam *frameworks* adaptativos, capazes de selecionar dinamicamente a estratégia de prompt mais adequada com base em características do material e nos objetivos do usuário (mapeamento exploratório versus validação final).

5.2.4 Evolução tecnológica e obsolescência

A comparação entre o Gemini 2.5 Flash e o Gemini 3 Pro revela a rapidez da evolução das LLMs. O salto de qualidade em raciocínio matemático e interpretação de texto em um curto intervalo de tempo indica que as ferramentas de apoio ao professor tornar-se-ão viáveis mais rapidamente do que o previsto na literatura anterior. Isso reforça a necessidade urgente de letramento em IA na formação de professores, para que estes possam acompanhar e selecionar as melhores ferramentas disponíveis.

5.2.5 Implicações para formação docente e letramento em IA

Os achados desta dissertação indicam que a introdução de ferramentas de IA generativa na curadoria curricular demanda, não apenas acesso tecnológico, mas desenvolvimento de competências específicas. Professores precisarão compreender os fundamentos de Precisão, Recall e *trade-offs* associados para avaliar criticamente as sugestões automatizadas e decidir quais validações manuais são necessárias.

Tal necessidade aponta para a inclusão, em programas de formação continuada, de módulos sobre letramento em IA. Conforme discute (49), a formação de professores para a inteligência artificial não deve se restringir ao manuseio instrumental das ferramentas (o como usar), mas deve ser multidimensional. (49) defende que é indispensável promover uma compreensão interdisciplinar que abarque dimensões éticas, de privacidade, autoria e segurança, capacitando o docente a entender as IAs não apenas como softwares, mas como elementos que impactam a estrutura social e cognitiva da escola.

5.3 DIREÇÕES PARA PESQUISAS FUTURAS

Estudos futuros deveriam investigar sistematicamente quais tipos de habilidades (conceituais, procedimentais ou atitudinais) são mais ou menos suscetíveis a identificação precisa por IA, e se há padrões sistemáticos de erros que possam indicar vieses algorítmicos específicos. Os resultados obtidos abrem múltiplas avenidas de investigação:

1. **Abordagens híbridas multi-modelo:** Investigar sistemas que combinem as previsões de múltiplos modelos (*ensemble learning*) para maximizar simultaneamente Precisão e Recall.
2. **Fine-tuning específico para BNCC:** Explorar o ajuste fino de modelos de linguagem com corpus especializado de documentos curriculares brasileiros, potencialmente melhorando a compreensão das nuances terminológicas da BNCC.
3. **Validação inter-juízes:** Ampliar o estudo com múltiplos especialistas construindo gabaritos independentes, permitindo avaliar concordância humana e comparar o desempenho da IA com a variabilidade natural da interpretação curricular.
4. **Estudo longitudinal de uso real:** Acompanhar professores utilizando essas ferramentas em contexto autêntico de planejamento de aulas, investigando como as sugestões da IA influenciam suas decisões pedagógicas e se há efetivo ganho de tempo sem comprometimento da qualidade.
5. **Extensão para outras áreas curriculares:** Replicar a metodologia em disciplinas como Língua Portuguesa, Ciências e História, verificando se os padrões observados em Matemática se generalizam ou se cada área demanda estratégias específicas de prompt.
6. **Ampliar o número de amostras para estatísticas mais confiáveis:** Incluir análises de materiais textuais processados de pdf's e imagens para tentar aumentar a assertividade dos prompts.

5.4 SÍNTESE DOS ACHADOS PRINCIPAIS

Esta pesquisa investigou sistematicamente a eficácia de diferentes estratégias de prompt aplicadas a quatro modelos de IA generativa na identificação de habilidades da BNCC em materiais didáticos de Matemática. Os principais achados podem ser sintetizados em quatro conclusões centrais:

1. **Viabilidade técnica consolidada:** Os modelos avaliados demonstraram capacidade de identificar habilidades da BNCC de forma automatizada, com desempenho superior ao acaso. Com a chegada de modelos como o Gemini 3 Pro, a identificação automática

pode atingir níveis de precisão da ordem de 70% em materiais estruturados, o que sugere viabilidade técnica para uso como ferramenta de apoio, mediante supervisão especializada.

2. **Superioridade da Revisão Iterativa:** O Prompt 4 (Revisão em Múltiplas Etapas) provou ser a estratégia de engenharia de prompt mais robusta, superando instruções diretas e *few-shot* na métrica de Recall e Jaccard.
3. **Dependência de características do material:** O desempenho variou substancialmente entre materiais didáticos, materiais estruturados e objetivos favoreceram todos os modelos, enquanto conteúdos densos e multitemáticos geraram maior incidência de erros, especialmente falsos positivos.
4. **Necessidade de Supervisão Humana:** Apesar dos avanços, o Jaccard médio em torno de 35% indica que os resultados indicam que, nas condições avaliadas, os sistemas de IA ainda requerem validação especializada, não constituindo substitutos autônomos para análise curricular.

Estes achados confirmam a hipótese inicial de que a formulação do prompt influencia significativamente a qualidade do mapeamento automático, mas também revelam que esta variável interage complexamente com características do modelo de IA e do material didático analisado.

6 CONCLUSÃO

A presente dissertação demonstrou que modelos de Inteligência Artificial Generativa, especialmente os de última geração como o **Google Gemini 3 Pro**, podem auxiliar significativamente o processo de curadoria e alinhamento de materiais didáticos de Matemática à BNCC como um filtro, quando é necessário escolher o que analisar e não há tempo hábil dentre inúmeras opções de materiais. Os resultados evidenciam que estas ferramentas evoluíram de meros geradores de texto para assistentes de análise lógica capazes de interpretar critérios curriculares complexo. Portanto, não substituem o professor, mas podem contribuir para otimizar processos de planejamento curricular, conforme evidenciado pelas métricas de desempenho obtidas.

A contribuição principal deste estudo reside na sistematização metodológica de estratégias de *prompt engineering* aplicadas ao domínio educacional brasileiro, demonstrando empiricamente que diferentes formulações de instrução produzem resultados qualitativamente distintos. O Gemini 3 Pro estabeleceu-se como referência em precisão, ideal para otimização de tempo, enquanto o ClaudeAI mantém sua relevância para varreduras exaustivas de conteúdo. Tal achado possui implicações diretas para o desenvolvimento de interfaces educacionais baseadas em IA, sugerindo que sistemas futuros deveriam permitir seleção ou adaptação dinâmica da estratégia de prompt.

Do ponto de vista prático, recomenda-se que educadores e gestores educacionais considerem a adoção dessas tecnologias em caráter complementar, utilizando-as em fases iniciais de triagem de materiais, seguidas de validação pedagógica criteriosa. A economia de tempo potencial (especialmente em corpus extensos) justifica o investimento, desde que acompanhada de formação docente adequada sobre as capacidades e limitações dos sistemas. Além disso, é fundamental promover a formação continuada em letramento digital, capacitando os docentes a avaliar criticamente as sugestões algorítmicas.

Com base nos resultados obtidos, apresentam-se as seguintes recomendações práticas para educadores e gestores educacionais interessados na aplicação de IA generativa para curadoria de materiais didáticos:

1. **Seleção contextualizada do modelo:** Utilizar ClaudeAI em fases exploratórias de mapeamento amplo (priorizando Recall), e Google Gemini em fases de validação final (priorizando Precisão).
2. **Segmentação de materiais extensos:** Dividir apostilas e livros densos em seções temáticas

antes de aplicar ferramentas de IA, o que pode reduzir significativamente os falsos positivos.

3. **Revisão humana obrigatória:** Estabelecer como protocolo institucional que todas as identificações automatizadas sejam submetidas a validação por professor especialista, especialmente em casos de materiais multitemáticos.
4. **Formação docente em prompt engineering:** Incorporar módulos básicos sobre formulação eficaz de prompts em programas de formação continuada, capacitando professores a interagir produtivamente com sistemas de IA.
5. **Documentação transparente:** Manter registro sistemático das decisões tomadas (modelo utilizado, prompt aplicado, ajustes manuais realizados), permitindo auditoria e aprimoramento contínuo do processo.

Como agenda de pesquisa futura, sugere-se a investigação de sistemas híbridos que combinem a precisão do Gemini com a abrangência do Claude, bem como estudos longitudinais sobre o impacto dessas ferramentas no tempo de planejamento docente em condições reais de trabalho. A integração crescente de IA na educação é inevitável, cabe à comunidade acadêmica garantir que esta integração seja crítica, fundamentada em evidências e orientada pelos princípios de equidade e qualidade educacional preconizados pela própria BNCC.

REFERÊNCIAS

- [1] ADEMI. **Como os algoritmos de recomendação da IA funcionam em plataformas como Netflix e YouTube**. Disponível em: <https://ademi.com.br/como-os-algoritmos-de-recomendacao-da-ia-funcionam-em-plataformas-como-netflix-e-youtube>. Acesso em: 15 ago. 2025.
- [2] AGÊNCIA BRASIL. **Inteligência artificial pode ser ferramenta de ensino, mostra estudo**. Brasília, DF: EBC, 2024. Disponível em: <https://agenciabrasil.ebc.com.br/educacao/noticia/2024-05/inteligencia-artificial-pode-ser-ferramenta-de-ensino-mostra-estudo>. Acesso em: 16 nov. 2025.
- [3] ALMEIDA, A. P. et al. Inteligência artificial na educação: vislumbrar possibilidades e minimizar desafios. **Jornal da USP**, São Paulo, 24 jul. 2024. Disponível em: <https://jornal.usp.br/artigos/inteligencia-artificial-na-educacao-vislumbrar-possibilidades-e-minimizar-desafios/>. Acesso em: 16 nov. 2025.
- [4] ALVES, C. F.; LEFFA, V. J. Professor-autor de recursos educacionais abertos: uma identidade em construção. **Revista Interfaces**, v. 11, n. 04, p. 190–208, 2020.
- [5] ARAÚJO, C. H. dos S.; FERNANDES, J. da S.; VILAS BOAS, C. A. V. Inteligência Artificial e sua relação com o trabalho docente no Brasil. In: CONGRESSO INTERNACIONAL DE EDUCAÇÃO E TECNOLOGIAS (CIET:Horizonte), 2024, São Carlos. **Anais...** São Carlos: UFSCar, 2024. Disponível em: <https://ciet.ufscar.br/>. Acesso em: 16 nov. 2025.
- [6] BAKER, R. S. J. D.; INVENTADO, P. S. Educational Data Mining and Learning Analytics. In: LARUSSON, J. A.; WHITE, B. (Eds.). **Learning Analytics: from Research to Practice**. New York: Springer, 2014. p. 61–75.
- [7] BHARGAVA, O. **Os 5 modelos de curadoria de conteúdo**. Disponível em: <https://rohitbhargava.com/the-5-models-of-content-curation/>. Acesso em: 20 set. 2025.
- [8] BORBA, M. C.; VILLARREAL, M. E. **Humans-with-Media and the Reorganization of Mathematical Thinking**. New York: Springer, 2005.
- [9] BORBA, M. C.; SCUCUGLIA, R. R. S.; GADANIDIS, G. **Fases das Tecnologias Digitais em Ensino da Matemática: sala de aula e internet em movimento**. Belo Horizonte: Autêntica, 2018.
- [10] BRASIL. **Decreto n. 9.099, de 18 de julho de 2017**. Dispõe sobre o Programa Nacional do Livro e do Material Didático. Diário Oficial da União, Brasília, DF, 19 jul. 2017. Seção 1, p. 7.
- [11] BRASIL. Fundo Nacional de Desenvolvimento da Educação. **Editais PNLD**. Brasília, DF: FNDE. Disponível em: <https://www.gov.br/fnde/pt-br/aceso-a-informacao/acoes-e-programas/programas/programas-do-livro/pnld/editais>. Acesso em: 16 nov. 2025.

- [12] BRASIL. Ministério da Educação. **Base Nacional Comum Curricular (BNCC)**. Brasília, DF: MEC, 2017. Disponível em: <http://basenacionalcomum.mec.gov.br>. Acesso em: 16 nov. 2025.
- [13] BRASIL. Ministério da Educação. **Debate sobre IA nas políticas educacionais é foco do BRICS**. Brasília, DF: MEC, 2025. Disponível em: <https://www.gov.br/mec/pt-br/assuntos/noticias/2025/junho/debate-sobre-ia-nas-politicas-educacionais-e-foco-do-brics>. Acesso em: 16 nov. 2025.
- [14] BRASIL. Ministério da Gestão e da Inovação em Serviços Públicos. Secretaria de Governo Digital. **Guia prático de prompt e pesquisa com IA para servidores públicos**. Brasília, DF: MGI, 2024. Disponível em: <https://www.gov.br/governodigital/pt-br/inteligencia-artificial/guia-pratico-de-prompt-e-pesquisa-com-ia-para-servidores-publicos>. Acesso em: 16 nov. 2025.
- [15] BROWN, T. et al. Language Models Are Few-Shot Learners. **Advances in Neural Information Processing Systems**, v. 33, p. 1877–1901, 2020.
- [16] CARDOSO, M. C. S. do A.; FIGUEIRA-SAMPAIO, A. da S. Dificuldades para o uso da informática no ensino: percepção dos professores de matemática após 40 anos da inserção digital no contexto educacional brasileiro. **Ensino da Matemática Pesquisa**, v. 21, n. 2, 2019.
- [17] CARDOSO, I. N. A.; MERCADO, L. P. L. Curadoria de conteúdo e autoria docente: contribuição dos portfólios digitais na Universidade Federal de Alagoas. In: CONGRESSO INTERNACIONAL DE EDUCAÇÃO E TECNOLOGIAS (CIET:Horizonte), 2024, São Carlos. **Anais...** São Carlos: UFSCar, 2024. Disponível em: <https://ciet.ufscar.br/>. Acesso em: 16 nov. 2025.
- [18] CECHINEL, C. **Modelos de curadoria de recursos educacionais digitais**. São Paulo: CIEB, 2019. (CIEB Estudos, 5). Disponível em: <https://cieb.net.br/cieb-estudos-5-modelos-de-curadoria-de-recursos-educacionais-digitais/>. Acesso em: 16 nov. 2025.
- [19] CHEN, C.-H.; YANG, Y.-C. Revisiting the effects of project-based learning on students' academic achievement: A meta-analysis investigating moderators. **Educational Research Review**, v. 26, p. 71–81, 2019.
- [20] COSTA, F. A. et al. Competências digitais para professores: um quadro de referência. **Revista e-Curriculum**, v. 19, n. 1, p. 204–226, 2021.
- [21] DANTE, L. R.; VIANA, F. **Teláris Essencial: Matemática: 6º ano**. 1. ed. São Paulo: Ática, 2022. Livro eletrônico.
- [22] DANTE, L. R.; VIANA, F. **Teláris Essencial: Matemática: 7º ano**. 1. ed. São Paulo: Ática, 2022. Livro eletrônico.
- [23] D'AMBROSIO, U. **Ensino da Matemática: da teoria à prática**. 14. ed. Campinas: Papyrus, 2008.
- [24] DEEP LEARNING BOOK. **Conhecendo o Modelo GPT-3 (Generative Pre-trained Transformer)**. 2025. Disponível em: <https://www.deeplearningbook.com.br/conhecendo-o-modelo-gpt-3-generative-pre-trained-transformer/>. Acesso em: 16 nov. 2025.
- [25] DEEP LEARNING BOOK. **Modelo BERT Para Processamento de Linguagem Natural**. 2025. Disponível em: <https://www.deeplearningbook.com.br/modelo-bert-para-processamento-de-linguagem-natural/>. Acesso em: 16 nov. 2025.
- [26] DEVLIN, J. et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: PROCEEDINGS OF NAACL, 2019, Minneapolis. **Anais [...]**. Minneapolis: ACL, 2019. p. 4171–4186.

- [27] FIORENTINI, D.; LORENZATO, S. **Investigação em Ensino da Matemática**: percursos teóricos e metodológicos. Campinas: Autores Associados, 2012.
- [28] FISCARELLI, R. B. de O. Material didático e prática docente. **Revista Ibero-Americana de Estudos em Educação**, v. 2, n. 1, p. 31–39, 2007.
- [29] GUIA DE ENGENHARIA DE PROMPTS. **Prompt Engineering Guide**. 2025. Disponível em: <https://www.promptingguide.ai/pt>. Acesso em: 16 nov. 2025.
- [30] IBM. **O que é GPT (generative pretrained transformer)?** 2025. Disponível em: <https://www.ibm.com/br-pt/think/topics/gpt>. Acesso em: 16 nov. 2025.
- [31] IBM. **O que é PLN (processamento de linguagem natural)?** 2025. Disponível em: <https://www.ibm.com/br-pt/think/topics/natural-language-processing>. Acesso em: 16 nov. 2025.
- [32] IBM. **What is zero-shot prompting?** 2025. Disponível em: <https://www.ibm.com/think/topics/zero-shot-prompting>. Acesso em: 16 nov. 2025.
- [33] KENSKI, V. M. **Educação e Tecnologias**: o novo ritmo da informação. Campinas: Papirus, 2012.
- [34] KOJIMA, T. et al. Large Language Models are Zero-Shot Reasoners. **ArXiv preprint**, 2022. Disponível em: <https://arxiv.org/abs/2205.11916>. Acesso em: 16 nov. 2025.
- [35] LORENZATO, S. **Para Aprender Matemática**. Campinas: Autores Associados, 2006.
- [36] MACÊDO, J. A. de; BRANDÃO, D. P.; NUNES, D. M. Limites e possibilidades do uso do livro didático de Matemática nos processos de ensino e de aprendizagem. **Ensino da Matemática Debate**, Montes Claros, v. 3, n. 7, p. 68–86, jan./abr. 2019.
- [37] NISS, M.; BLUM, W.; GALBRAITH, P. The Concept of Competence in the Context of PISA and Other International Studies. In: BLUM, W. et al. (Eds.). **Modelling and Applications in Mathematics Education**. New York: Springer, 2007. p. 3–32.
- [38] PARENTE, U. L. **Operações com números na forma decimal - Parte 2**. Material Teórico - Módulo Operações Básicas. Sexto Ano do Ensino Fundamental. Rio de Janeiro: Portal da Matemática OBMEP, 2024. Disponível em: <http://matematica.obmep.org.br/>. Acesso em: 16 nov. 2025.
- [39] PERRENOUD, P. **Construir as Competências desde a Escola**. Porto Alegre: Artmed, 1999.
- [40] RAMOS, O. D. Anotações para a compreensão da atividade do ‘Curador de Informação Digital’. In: CORRÊA, E. N. (Org.). **Curadoria digital e o campo da comunicação**. São Paulo: ECA/USP, 2012.
- [41] RIO DE JANEIRO (Município). Secretaria Municipal de Educação. **Material Rioeduca: Matemática - 9º Ano - 1º Semestre**. Rio de Janeiro: SME, 2022.
- [42] ROMERO, C.; VENTURA, S. Educational Data Mining and Learning Analytics: An Updated Survey. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, v. 10, n. 3, e1355, 2020. DOI: 10.1002/widm.1355.
- [43] ROSSINI, C. R.; GONZALEZ, C. REA: o debate em política pública e as oportunidades para o mercado. In: PRETTO, N. de L.; ROSSINI, C.; SANTANA, B. (Orgs.). **Recursos Educacionais abertos**: práticas colaborativas e políticas públicas. 1. ed. Salvador: Edufba, 2012. p. 91–108.
- [44] SANTAELLA, L. **A inteligência artificial é inteligente?** São Paulo: Edições 70, 2023.

- [45] SANTOS, A. C. **Algoritmo do YouTube**: descubra como ele escolhe os seus favoritos. 2024. Disponível em: <https://www.mlabs.com.br/blog/algoritmo-do-youtube>. Acesso em: 15 ago. 2025.
- [46] UNESCO. **Artificial Intelligence and Education**: Guidance for Policy-Makers. Paris: UNESCO, 2022.
- [47] UNESCO. **AI and Education**: Guidance for Policy-makers. Paris: UNESCO, 2023.
- [48] VASWANI, A. et al. Attention Is All You Need. In: PROCEEDINGS OF THE 31ST INTERNATIONAL CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS (NIPS), 2017, Long Beach. **Anais [...]**. Long Beach: ACM, 2017. p. 6000–6010.
- [49] VILAÇA, M. L. C. Inteligência Artificial na educação: dimensões e estratégias para a formação de professores. In: CONGRESSO INTERNACIONAL DE EDUCAÇÃO E TECNOLOGIAS (CIET:Horizonte), 2024, São Carlos. **Anais...** São Carlos: UFSCar, 2024. Disponível em: <https://ciet.ufscar.br/>. Acesso em: 16 nov. 2025.
- [50] WEI, J. et al. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. **ArXiv preprint**, 2022. Disponível em: <https://arxiv.org/abs/2201.11903>. Acesso em: 16 nov. 2025.
- [51] YOUTUBE OFFICIAL BLOG. **Nossa abordagem quanto à inovação da IA responsável**. Disponível em: <https://blog.youtube/intl/pt-br/inside-youtube/nossa-abordagem-quanto-a-inovacao-da-ia-responsavel>. Acesso em: 05 ago. 2025.

APÊNDICE A - GABARITO DESCRITIVO DAS HABILIDADES IDENTIFICADAS

Abaixo apresentam-se as habilidades da BNCC identificadas no gabarito de referência para cada material analisado, com seus respectivos códigos e descrições integrais extraídas do documento normativo.

Material	Código	Descrição da Habilidade (BNCC)
Frações.PNG (7º Ano)	EF06MA06	Resolver e elaborar problemas que envolvam as ideias de múltiplo e de divisor.
	EF06MA07	Compreender, comparar e ordenar frações associadas às ideias de partes de inteiros e resultado de divisão, identificando frações equivalentes.
	EF06MA08	Reconhecer que os números racionais positivos podem ser expressos nas formas fracionária e decimal, estabelecer relações entre essas representações, passando de uma representação para outra, e relacioná-los a pontos na reta numérica.
	EF06MA09	Resolver e elaborar problemas que envolvam o cálculo da fração de uma quantidade e cujo resultado seja um número natural, com e sem uso de calculadora.
	EF06MA10	Resolver e elaborar problemas que envolvam adição ou subtração com números racionais positivos na representação fracionária.
	EF07MA10	Comparar e ordenar números racionais em diferentes contextos e associá-los a pontos da reta numérica.
Matemática 9 ano apostila.pdf (9º Ano)	EF07MA30	Resolver e elaborar problemas de cálculo de medida do volume de blocos retangulares, envolvendo as unidades usuais (metro cúbico, decímetro cúbico e centímetro cúbico).

Material	Código	Descrição da Habilidade (BNCC)
	EF08MA04	Resolver e elaborar problemas, envolvendo cálculo de porcentagens, incluindo o uso de tecnologias digitais.
	EF08MA05	Reconhecer e utilizar procedimentos para a obtenção de uma fração geratriz para uma dízima periódica.
	EF08MA15	Construir, utilizando instrumentos de desenho ou softwares de geometria dinâmica, mediatriz, bissetriz, ângulos de 90°, 60°, 45° e 30° e polígonos regulares.
	EF08MA20	Reconhecer a relação entre um litro e um decímetro cúbico e a relação entre litro e metro cúbico, para resolver problemas de cálculo de capacidade de recipientes.
	EF08MA21	Resolver e elaborar problemas que envolvam o cálculo do volume de recipiente cujo formato é o de um bloco retangular.
	EF09MA01	Reconhecer que, uma vez fixada uma unidade de comprimento, existem segmentos de reta cujo comprimento não é expresso por número racional (como as medidas de diagonais de um polígono e alturas de um triângulo, quando se toma a medida de cada lado como unidade).
	EF09MA02	Reconhecer um número irracional como um número real cuja representação decimal é infinita e não periódica, e estimar a localização de alguns deles na reta numérica.
	EF09MA03	Efetuar cálculos com números reais, inclusive potências com expoentes fracionários.
	EF09MA06	Compreender as funções como relações de dependência unívoca entre duas variáveis e suas representações numérica, algébrica e gráfica e utilizar esse conceito para analisar situações que envolvam relações funcionais entre duas variáveis.
	EF09MA07	Resolver problemas que envolvam a razão entre duas grandezas de espécies diferentes, como velocidade e densidade demográfica.
	EF09MA08	Resolver e elaborar problemas que envolvam relações de proporcionalidade direta e inversa entre duas ou mais grandezas, inclusive escalas, divisão em partes proporcionais e taxa de variação, em contextos socio-culturais, ambientais e de outras áreas.

Material	Código	Descrição da Habilidade (BNCC)
	EF09MA09	Compreender os processos de fatoração de expressões algébricas, com base em suas relações com os produtos notáveis, para resolver e elaborar problemas que possam ser representados por equações polinomiais do 2º grau.
	EF09MA10	Demonstrar relações simples entre os ângulos formados por retas paralelas cortadas por uma transversal.
	EF09MA14	Resolver e elaborar problemas de aplicação do teorema de Pitágoras ou das relações de proporcionalidade envolvendo retas paralelas cortadas por secantes.
	EF09MA15	Descrever, por escrito e por meio de um fluxograma, um algoritmo para a construção de um polígono regular cuja medida do lado é conhecida, utilizando régua e compasso, como também softwares.
Média Aritmética.docx (Probabilidade e Estatística)	EF07MA34	Planejar e realizar experimentos aleatórios ou simulações que envolvem cálculo de probabilidades ou estimativas por meio de frequência de ocorrências.
	EF07MA35	Compreender, em contextos significativos, o significado de média estatística como indicador da tendência de uma pesquisa, calcular seu valor e relacioná-lo, intuitivamente, com a amplitude do conjunto de dados.
	EF07MA36	Planejar e realizar pesquisa envolvendo tema da realidade social, identificando a necessidade de ser censitária ou de usar amostra, e interpretar os dados para comunicá-los por meio de relatório escrito, tabelas e gráficos, com o apoio de planilhas eletrônicas.
	EF08MA25	Obter os valores de medidas de tendência central de uma pesquisa estatística (média, moda e mediana) com a compreensão de seus significados e relacioná-los com a dispersão de dados, indicada pela amplitude.
Operações com números na forma decimal.pdf (6º Ano)	EF04MA02	Mostrar, por decomposição e composição, que todo número natural pode ser escrito por meio de adições e multiplicações por potências de dez, para compreender o sistema de numeração decimal e desenvolver estratégias de cálculo.

Material	Código	Descrição da Habilidade (BNCC)
	EF05MA02	Ler, escrever e ordenar números racionais na forma decimal com compreensão das principais características do sistema de numeração decimal, utilizando, como recursos, a composição e decomposição e a reta numérica.
	EF06MA02	Reconhecer o sistema de numeração decimal, como o que prevaleceu no mundo ocidental, e destacar semelhanças e diferenças com outros sistemas, de modo a sistematizar suas principais características (base, valor posicional e função do zero), utilizando, inclusive, a composição e decomposição de números naturais e números racionais em sua representação decimal.
	EF06MA04	Construir algoritmo em linguagem natural e representá-lo por fluxograma que indique a resolução de um problema simples (por exemplo, se um número natural qualquer é par).
	EF06MA12	Fazer estimativas de quantidades e aproximar números para múltiplos da potência de 10 mais próxima.
	EF06MA13	Resolver e elaborar problemas que envolvam porcentagens, com base na ideia de proporcionalidade, sem fazer uso da “regra de três”, utilizando estratégias pessoais, cálculo mental e calculadora, em contextos de educação financeira, entre outros.
	EF06MA24	Resolver e elaborar problemas que envolvam as grandezas comprimento, massa, tempo, temperatura, área (triângulos e retângulos), capacidade e volume (sólidos formados por blocos retangulares), sem uso de fórmulas, inseridos, sempre que possível, em contextos oriundos de situações reais e/ou relacionadas às outras áreas do conhecimento.