

UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE CIÊNCIA DA INFORMAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO & ORGANIZAÇÃO DO
CONHECIMENTO

JORDANA RABELO SOARES

**BASE DE CONHECIMENTO PARA RECUPERAÇÃO E ACESSO À INFORMAÇÃO
ACADÊMICA: PROVIMENTO PARA UM *CHATBOT* NA ECI UFMG**

Belo Horizonte
2025

**BASE DE CONHECIMENTO PARA RECUPERAÇÃO E ACESSO À INFORMAÇÃO
ACADÊMICA: PROVIMENTO PARA UM *CHATBOT* NA ECI UFMG**

Dissertação apresentada ao Programa de Pós-Graduação em Gestão & Organização do Conhecimento, Escola de Ciência da Informação da Universidade Federal de Minas Gerais para obtenção do grau de Mestre, área de concentração Ciência da Informação.

Linha de Pesquisa: Gestão & Tecnologia da Informação e Comunicação (GETIC)

Orientadora: Patrícia Nascimento Silva

Belo Horizonte

2025

Dados Internacionais de Catalogação-na-Publicação (CIP)

S676b Soares, Jordana Rabelo
Base de conhecimento para recuperação e acesso à informação acadêmica [recurso eletrônico] : provimento para um *chatbot* na ECI UFMG / Jordana Rabelo Soares Gonçalves. - 2025.
1 recurso online (205 f. : il., color) pdf.

Orientadora: Patrícia Nascimento Silva.

Dissertação (mestrado) - Universidade Federal de Minas Gerais, Escola de Ciência da Informação.

Referências: f.178-197.

Anexo: f. 198-201.

Apêndice: f.202-205.

1. Ciência da informação - Teses. 2. Base do Conhecimento - Teses. 3. Recuperação de Informação - Teses. 4. Curadoria informacional - Teses. 5. Chatbots - Teses. 6. Informação acadêmica - Teses. I. Nascimento Silva, Patrícia. II. Universidade Federal de Minas Gerais. Escola de Ciência da Informação.III. Título.

CDD. – 23. ed. 025.043

Ficha catalográfica: Suellen Souza Gonçalves – CRB6: 003757/O

Como citar: SOARES, Jordana Rabelo. **Base de conhecimento para recuperação e acesso à informação acadêmica**: provimento para um *chatbot* na ECI UFMG. 2025. 205f. Dissertação (Mestrado em Gestão e Organização do Conhecimento) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2025.



UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE CIÊNCIA DA INFORMAÇÃO - ECI
PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO
- PPGGOC

FOLHA DE APROVAÇÃO

BASE DE CONHECIMENTO PARA RECUPERAÇÃO E ACESSO À INFORMAÇÃO ACADÊMICA: PROVIMENTO PARA UM CHATBOT NA ECI UFMG

JORDANA RABELO SOARES

Dissertação submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, como requisito para obtenção do grau de Mestre em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, área de concentração CIÊNCIA DA INFORMAÇÃO, linha de pesquisa Gestão e Tecnologia da Informação e Comunicação.

APROVADA COM LOUVOR EM 12 DE NOVEMBRO DE 2025, POR VIDEOCONFERÊNCIA, COM INDICAÇÃO PARA PARTICIPAR DE PREMIAÇÕES, PELA BANCA CONSTITUÍDA PELOS MEMBROS:

Prof(a). Patricia Nascimento Silva (Orientadora)
ECI/UFMG

Prof(a). José Eduardo Santarem Segundo
USP

Prof(a). Gercina Ângela de Lima
ECI/UFMG

Prof(a). Eduardo Valadares da Silva
ECI/UFMG

Belo Horizonte, 12 de novembro de 2025.



Documento assinado eletronicamente por **Patricia Nascimento Silva, Professora do Magistério Superior**, em 12/11/2025, às 17:42, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Gercina Angela de Lima, Professora do Magistério Superior**, em 20/03/2024, às 16:25, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Eduardo Valadares da Silva, Professor do Magistério Superior**, em 13/11/2025, às 11:21, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Jose Eduardo Santarem Segundo, Usuário Externo**, em 16/11/2025, às 22:11, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **4726877** e o código CRC **987F9193**.

Referência: Processo nº 23072.269527/2025-75

SEI nº 4726877



UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE CIÊNCIA DA INFORMAÇÃO - ECI
PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO
- PPGGOC

ATA DA DEFESA DA DISSERTAÇÃO DA ALUNA

JORDANA RABELO SOARES

Realizou-se, no dia 12 de novembro de 2025, às 14:00 horas, por videoconferência, da Universidade Federal de Minas Gerais, a defesa de dissertação, intitulada *BASE DE CONHECIMENTO PARA RECUPERAÇÃO E ACESSO À INFORMAÇÃO ACADÊMICA: PROVIMENTO PARA UM CHATBOT NA ECI UFMG*, apresentada por JORDANA RABELO SOARES, número de registro 2024659017, graduada no curso de BIBLIOTECONOMIA, como requisito parcial para a obtenção do grau de Mestre em GESTÃO E ORGANIZAÇÃO DO CONHECIMENTO, à seguinte Comissão Examinadora: Prof(a). Patricia Nascimento Silva - ECI/UFMG (Orientadora), Prof(a). José Eduardo Santarem Segundo - USP, Prof(a). Gercina Ângela de Lima - ECI/UFMG, Prof(a). Eduardo Valadares da Silva - ECI/UFMG.

A Comissão considerou a dissertação:

Aprovada

Reprovada

APROVADA COM LOUVOR, COM INDICAÇÃO DOS MEMBROS DA BANCA PARA PARTICIPAR DE PREMIAÇÕES.

Finalizados os trabalhos, lavrei a presente ata que, lida e aprovada, vai assinada por mim e pelos membros da Comissão.

Belo Horizonte, 12 de novembro de 2025.

Assinatura dos membros da banca examinadora:



Documento assinado eletronicamente por **Patricia Nascimento Silva, Professora do Magistério Superior**, em 12/11/2025, às 17:40, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Gercina Angela de Lima, Professora do Magistério Superior**, em 13/11/2025, às 10:46, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Eduardo Valadares da Silva, Professor do Magistério Superior**, em 13/11/2025, às 11:21, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Jose Eduardo Santarem Segundo, Usuário Externo**, em 16/11/2025, às 22:11, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **4726859** e o código CRC **70CB39AC**.

Referência: Processo nº 23072.269527/2025-75

SEI nº 4726859

AGRADECIMENTOS

A Deus, fonte de todo conhecimento, pela força e serenidade que me sustentaram ao longo desta caminhada.

À minha orientadora, Prof^a. Dr^a. Patrícia Nascimento Silva, pela dedicação, paciência e incentivo constante. Sua orientação criteriosa foi fundamental para a realização deste trabalho, não apenas pela contribuição acadêmica, mas também pelo exemplo de profissionalismo e humanidade.

Aos integrantes do projeto Recuperação de Informação em modelos de Inteligência Artificial: o papel da Ciência da Informação, coordenado pela minha orientadora, Profa. Patrícia, especialmente aos colegas Maria Eduarda Pereira (Dudha), Amanda Carvalho Fernandes, Christiano Benvindo dos Santos e João Paulo Kroger Macedo Neves. Registro meus agradecimentos por todas as enriquecedoras trocas, pelo apoio e pelas oportunidades de crescimento que me foram generosamente concedidas nessa equipe multidisciplinar.

Meus agradecimentos aos membros da banca de avaliação de mestrado, Prof.^a Dr^a. Gercina Ângela de Lima, Prof. Dr. José Eduardo Santarém Segundo, Prof. Dr. Eduardo Valadares e Prof. Dr. Renato Rocha Souza, por suas valiosas contribuições e pelo incentivo ao desenvolvimento de minha pesquisa.

Aos professores e colegas da Escola de Ciência da Informação da UFMG, pelos debates, aprendizados e trocas que enriqueceram minha formação e ampliaram meu olhar sobre a Ciência da Informação.

À toda a Equipe da Escola de Enfermagem da UFMG, REME e BVS Enfermagem, em especial ao Prof. Dr. Francisco Carlos Félix Lana e à Prof^a Dra^a Tânia Couto Machado Chianca, pelo incentivo e apoio.

Aos estudantes que participaram do questionário, pela disponibilidade e pelas contribuições, que foram essenciais para a construção desta pesquisa.

À minha família, em especial aos meus pais, pelo amor, apoio incondicional e incentivo em todos os momentos.

Aos amigos que me acompanharam durante o mestrado, especialmente Suellen, Elaine, Danielle e Leonora, pelo companheirismo, pelas palavras de encorajamento e pelo apoio nos momentos mais difíceis.

Por fim, à UFMG e ao Programa de Pós-Graduação em Gestão e Organização do Conhecimento - PPGGOC, pela oportunidade de desenvolvimento acadêmico e pessoal.

RESUMO

A recuperação de informações acadêmicas (normas, resoluções e procedimentos) em universidades públicas constitui um desafio para os estudantes, diante da dispersão de documentos em múltiplos canais e por sua linguagem técnica. Nesse cenário, a curadoria informacional surge como estratégia para organizar, e a base de conhecimento, uma maneira estruturada para subsidiar a recuperação em sistemas, sendo esses os questionamentos que norteiam esta pesquisa. A partir dessa lacuna, esta dissertação teve como objetivo estruturar e implementar uma base de conhecimento para subsidiar a construção de um *chatbot* voltado à recuperação de informações acadêmicas (normas, resoluções e procedimentos) sobre as graduações na Escola de Ciência da Informação da Universidade Federal de Minas Gerais. Especificamente, buscou-se: (i) mapear métodos e técnicas utilizados no desenvolvimento de *chatbots* com foco na informação; (ii) identificar o comportamento de busca dos estudantes em relação a documentos normativos; (iii) identificar e analisar *chatbots* educacionais implementados; e (iv) realizar a curadoria informacional para subsidiar a implementação de uma base de conhecimento. A pesquisa caracteriza-se como um estudo de caso, com natureza aplicada, descritiva e exploratória e abordagem quali quantitativa, delineada em quatro etapas. A primeira consistiu em uma revisão de literatura sobre o desenvolvimento de *chatbots*, identificando metodologias, arquiteturas, linguagens, plataformas e técnicas de Processamento de Linguagem Natural. Observou-se que os estudos privilegiam aspectos técnicos, negligenciando o tratamento da informação. A segunda etapa envolveu a análise de um questionário aplicado no âmbito do projeto Recuperação da Informação em modelos de Inteligência Artificial, em 2025, com respostas de estudantes da graduação da UFMG, que permitiu compreender como eles buscam normas, quais dificuldades enfrentam e quais são suas dúvidas e opiniões. Observou-se que a fragmentação das fontes e a linguagem normativa complexa configuram os principais desafios. Na terceira etapa, foram analisados estudos sobre *chatbots* educacionais, previamente identificados na revisão de literatura, selecionados por apresentarem modelos documentados e replicáveis. A análise forneceu exemplos de fluxos de interação, estruturas de *intents*, usos de entidades e bases de conhecimento, servindo como referência prática. Por fim, a quarta etapa concentrou-se na curadoria informacional da base, que foi desenvolvida em cinco fases: levantamento de necessidades informacionais; seleção de fontes e documentos; estruturação e criação da base; desenvolvimento (popular e validar); e modelagem de fluxos informacionais. Essa etapa resultou em uma base de conhecimento com 31 conceitos centrais, organizados em uma estrutura taxonômica com 225 intenções (perguntas e respostas) fundamentadas em documentos oficiais, acompanhadas da localização exata em artigos, parágrafos, além do *link* dos documentos. Apesar da validação envolver uma especialista, a ausência de validação com usuários finais foi uma limitação da pesquisa e segue como trabalho futuro. A pesquisa contribuiu de forma prática ao apresentar um estudo de caso, que gerou um produto: uma base de conhecimento estruturada e escalável que oferece uma fundamentação teórica e metodológica, no contexto informacional, para o desenvolvimento de soluções inovadoras, fomentando a transformação digital e promovendo impacto social. Este produto estratégico subsidiará a implementação de um *chatbot* institucional, fornecendo respostas contextualizadas, confiáveis e alinhadas às necessidades informacionais dos estudantes.

Palavras-chave: Base de conhecimento. Recuperação da Informação. Curadoria informacional. *Chatbots*. Informação acadêmica.

ABSTRACT

The retrieval of academic information (regulations, resolutions and procedures) in public universities poses a challenge for students, given the dispersion of documents across multiple channels and their technical language. In this context, informational curation emerges as a strategy for organization, while the knowledge base provides a structured means to support retrieval in systems, these being the core questions guiding this research. In light of this gap, the present dissertation aimed to structure and implement a knowledge base to support the development of a chatbot designed for the retrieval of academic information (regulations, resolutions and procedures) related to undergraduate programs at the School of Information Science of the Federal University of Minas Gerais. Specifically, it sought to: (i) map methods and techniques used in the development of information-oriented chatbots; (ii) identify the information-seeking behavior of the students regarding normative documents; (iii) identify and analyze implemented educational chatbots; and (iv) conduct informational curation to support the implementation of a knowledge base. This research is characterized as a case study, with an applied, descriptive and exploratory nature, adopting a mixed-methods approach, structured in four stages. The first stage consisted of a literature review on chatbot development, identifying methodologies, architectures, languages, platforms and Natural Language Processing techniques. It was observed that existing studies privilege technical aspects, often neglecting information treatment. The second stage involved the analysis of a questionnaire administered within the project Information Retrieval in Artificial Intelligence Models, in 2025, with responses from undergraduate students at UFMG, which allowed for an understanding of how students seek normative documents, the challenges that they face, as well as their doubts and opinions. The results indicated that the fragmentation of information sources and the complexity of normative language constitute the main difficulties. In the third stage, studies on educational chatbots, previously identified in the literature review, were analyzed and selected based on the presentation of documented and replicable models. The analysis provided examples of interaction flows, intent structures, uses of entities and knowledge bases, serving as practical references. Finally, the fourth stage focused on the informational curation of the knowledge base, developed in five phases: identification of informational needs; selection of sources and documents; structuring and creation of the base; development (population and validation); and modeling of informational flows. This stage resulted in a knowledge base comprising 31 core concepts, organized into a taxonomic structure with 225 intents (questions and answers) grounded in official documents, each accompanied by exact references to articles, paragraphs and document links. Although validation involved an expert, the absence of end-user validation was a limitation of the study and remains as future work. This research made a practical contribution by presenting a case study that generated a product: a structured and scalable knowledge base that offers theoretical and methodological grounding, within the informational context, for the development of innovative solutions, fostering digital transformation and promoting social impact. This strategic product will support the implementation of an institutional chatbot capable of providing contextualized, reliable and user-centered responses aligned with the informational needs of the students.

Keywords: Knowledge base. Information retrieval. Informational curation. Chatbots. Academic information.

LISTA DE FIGURAS

Figura 1 - Representação do processo de recuperação	34
Figura 2 - Evolução dos chatbots ao longo da história	49
Figura 3 - Arquitetura geral de um chatbot	52
Figura 4 - Fluxo básico de funcionamento de um chatbot	53
Figura 5 - Delineamento da pesquisa	66
Figura 6 - Fluxograma da seleção e elegibilidade dos documentos da revisão	73
Figura 7 - Mapa de calor (afiliação dos autores)	79
Figura 8 - Síntese da revisão de literatura	100
Figura 9 - Chatbot - arquitetura	123
Figura 10 - Sistema de busca	126
Figura 11 - Arquitetura de um sistema de chatbot voltado à recuperação de informações e serviços	129
Figura 12 - Arquitetura UNYSA	132
Figura 13 - Principais componentes do chatbot	135
Figura 14 - Etapas para a construção da base de conhecimento	142
Figura 15 - Base de conhecimento (visão geral)	160
Figura 16 - Base de conhecimento (detalhamento para o conceito truncamento - definição)	161
Figura 17 - Fluxograma	165

LISTA DE QUADROS

Quadro 1 - Classificação dos SOCs	32
Quadro 2 - Paralelo entre as atividades realizadas pelo bibliotecário acadêmico e o curador de conteúdo	42
Quadro 3 - Tipologias de chatbots	49
Quadro 4 - Proposta de pesquisa	59
Quadro 5 - Protocolo da revisão de literatura	70
Quadro 6 - Documentos selecionados	74
Quadro 7 - Lista de categorias identificadas nas dúvidas dos estudantes	114
Quadro 8 - Estudos selecionados	121
Quadro 9 - Relação das fontes e dos documentos utilizados	145
Quadro 10 - Estrutura da base de conhecimento	151
Quadro 11 - Curadoria base de conhecimento	158
Quadro 12 - Critérios de seleção da ferramenta chatbot	169
Quadro 13 - Ferramentas para chatbot com IA	170
Quadro 14 - Recorte do questionário da pesquisa sobre o comportamento de busca dos estudantes de graduação e pós-graduação da UFMG	198

LISTA DE GRÁFICOS

Gráfico 1 - Distribuição das publicações por ano	78
Gráfico 2 - Área de conhecimento (categorização do autor)	80
Gráfico 3 - Cursos com maior participação	107
Gráfico 4 - Graduação X Período	108
Gráfico 5 - Momentos de acesso	109
Gráfico 6 - Ambientes e fontes mais acessados	110
Gráfico 7 - Estratégias diante das dificuldades	111
Gráfico 8 - Confiabilidade das fontes não oficiais	112
Gráfico 9 - Avaliação geral do acesso	113
Gráfico 10 - Principais dúvidas apresentadas	115
Gráfico 11 - Percepção sobre acessibilidade	116
Gráfico 12 - Detalhamento da base de conhecimento	156
Gráfico 13 - Quantidade de perguntas criadas	157

LISTA DE ABREVIATURAS

AACR2 – Anglo-American Cataloguing Rules
API – Interface de Programação de Aplicações
BDTD – Biblioteca Digital Brasileira de Teses e Dissertações
BERT – Bidirectional Encoder Representations From Transformers
BPM – Gestão de Processos de Negócio
BRASED – Thesaurus Brasileiro da Educação
CEPE – Conselho de Ensino, Pesquisa e Extensão
CI – Ciência da Informação
CSV – Valores Separados Por Vírgulas
DRCA – Departamento de Registro e Controle Acadêmico
ECI – Escola de Ciência da Informação
FCA – Formação Complementar Aberta
FI – Formação Livre
FUMP – Fundação Universitária Mendes Pimentel
HTML – Hypertext Markup Language
IAG – Inteligência Artificial Generativa
LLAMA – Large Language Model Meta AI
LLM – Modelo de Linguagem de Grande Escala
MI – Aprendizado de Máquina
NAI – Núcleo de Acessibilidade e Inclusão
NGG – Norma Geral de Graduação
NIPS – Números de Identificação Pessoal
NLG – Geração de Linguagem Natural
NSG – Nota Semestral Global
NLU – Compreensão de Linguagem Natural
OI – Organização da Informação
PLN – Processamento de Linguagem Natural
RAG – Retrieval Augmented Generation
RDA – Resource Description And Access
RI – Recuperação da Informação
RU – Restaurante Universitário
SKOS – Simple Knowledge Organization System
SOCS – Sistemas de Organização do Conhecimento

SIGA – Sistema de Gestão Acadêmica

TF-IDF – Term Frequency – Inverse Document Frequency

TI – Tecnologia da Informação

UFMG – Universidade Federal de Minas Gerais

SUMÁRIO

1. INTRODUÇÃO	20
1.1 Problema e Justificativa	22
1.2 Objetivos	25
1.2.1 Objetivo geral	25
1.2.2 Objetivos específicos	25
1.3 Estrutura da dissertação	26
2. REFERENCIAL TEÓRICO	27
2.1 Organização da Informação e do Conhecimento	28
2.2 Representação da informação e do conhecimento	29
2.3 Recuperação da Informação	33
2.3.1 Modelos de Recuperação da Informação	35
2.3.1.1 Modelos clássicos de Recuperação da Informação	36
2.3.1.1.1 Modelo booleano	36
2.3.1.1.2 Modelo vetorial	36
2.3.1.1.3 Modelo probabilístico	36
2.3.1.2 Modelos dinâmicos de Recuperação da Informação	37
2.3.2.1 Lógica fuzzy	37
2.3.2.2 Redes neurais	37
2.3.2.3 Algoritmos Genéticos	38
2.3.4 Inteligência Artificial e Inteligência Artificial Generativa	38
2.4 Curadoria da informação	41
2.4.1 Definições e tipologias	41
2.4.2 Aplicações da curadoria na Ciência da Informação	44
3. CHATBOTS	46
3.1 Histórico e evolução	47
3.2 Tipos de chatbots	49
3.3 Arquitetura e funcionamento de chatbots	51
4 METODOLOGIA	56
4.1 Caracterização da pesquisa	56
4.2 Locus da pesquisa	57
4.3 Etapas da pesquisa	58
4.4 Etapa 1: Revisão da literatura	61
4.5 Etapa 2: Comportamento de busca dos estudantes de graduação	62

4.6 Etapa 3: Análise de chatbots da educação	62
4.7 Etapa 4 - Curadoria informacional	63
5. RESULTADOS	67
5.1 REVISÃO DE LITERATURA	67
5.2 Resultados	77
5.2.1 Metodologias utilizadas	80
5.2.2 Abordagens ágeis	81
5.2.3 Abordagens tradicionais	82
5.2.4 Modelo	85
5.2.5 Linguagens	90
5.2.6 Plataforma	93
5.2.6.1 Técnicas de Processamento de Linguagem Natural (PLN)	95
5.2.6.2 Term Frequency-Inverse Document Frequency (TF-IDF)	96
5.2.6.3 Seq2seq	97
5.2.6.4 Bag of words	98
5.3 Considerações da revisão	99
6. COMPORTAMENTO DE BUSCA DOS ESTUDANTES DE GRADUAÇÃO DA UFMG	104
6.1 Metodologia	104
6.1.1 População e amostra	104
6.1.2 Aspectos éticos e instrumentos de coleta	105
6.1.3 Fluxo do questionário	105
6.1.4 Divulgação do questionário	106
6.1.5 Análise dos resultados	106
6.1.6 Descrição da amostra	106
6.2 Seção 1 - Quando, Como e Onde os estudantes de graduação acessam as informações acadêmicas	108
6.3 Seção 2 - Como identificam e qual a percepção sobre as informações acadêmicas	110
6.4 Seção 3 - Dúvidas, opiniões e sugestões dos estudantes de graduação	113
7. CHATBOTS EDUCACIONAIS	120
7.1 Resultados	122
7.2 Discussão (contexto informacional)	135
7.3 Considerações sobre os chatbots educacionais	138
8. BASE DE CONHECIMENTO	140

8.1 Levantamento de necessidades informacionais	142
8.2 Seleção de fontes e documentos	143
8.3 Estruturação e criação da base de conhecimento	148
8.4 Desenvolvimento da base de conhecimento (popular e validar)	152
8.5 Modelagem de fluxos informacionais	162
8.6 Recuperação da informação e aplicações da base de conhecimento	166
8.7 Trabalhos futuros para o desenvolvimento do chatbot	167
9. CONCLUSÕES	172
9.1 Contribuições da dissertação	176
9.2 Desafios e limitações	176
9.3 Trabalhos futuros	177
9.4 Considerações finais	177
REFERÊNCIAS	178
ANEXO A	198
APÊNDICE A	202

1. INTRODUÇÃO

A Ciência da Informação (CI) configura-se como um campo interdisciplinar dedicado à investigação das propriedades, do comportamento e do fluxo da informação, abrangendo seus processos de organização, armazenamento, recuperação e uso (Borko, 1968; Saracevic, 2007). Em sua essência, a CI busca compreender como a informação é criada, organizada, disseminada e utilizada por indivíduos e organizações, permeando desde registros documentais até sistemas e tecnologias que viabilizam o acesso ao conhecimento (Buckland, 1991; Capurro; Hjørland, 2003).

Com o avanço da digitalização e das tecnologias de comunicação em rede (Castells, 2000), a CI passou a incorporar, em seus estudos, novas interfaces de acesso à informação, entre elas os *chatbots*, que se destacam como agentes de mediação automatizada. Esses sistemas conversacionais são projetados para simular interações humanas em linguagem natural, operando com base em técnicas de Processamento de Linguagem Natural (PLN) (Jurafsky; Martin, 2023), utilizando técnicas e algoritmos da Recuperação da Informação (RI) (Baeza-Yates; Ribeiro-Neto, 1999) e representando, nesse sentido, uma alternativa às interfaces tradicionais, baseadas em menus ou comandos técnicos (Nielsen, 2000; Shneiderman, 2020).

O uso de *chatbots* tem crescido no contexto educacional, administrativo e institucional, além de vários outros setores, e vem sendo valorizado pela sua capacidade de oferecer atendimento automatizado, respostas personalizadas e suporte contínuo à demanda por informação (Moraes Neto *et al.*, 2020). Seu potencial se manifesta, especialmente, em ambientes marcados pela dispersão de conteúdos e multiplicidade de fontes, como ocorre nas universidades públicas, onde normas, editais, calendários e regulamentos estão frequentemente distribuídos entre diferentes sistemas, setores e formatos. Estudos recentes apontam que a mediação automatizada pode contribuir para reduzir a sobrecarga dos serviços administrativos, facilitar o acesso a informações relevantes e mitigar as dificuldades enfrentadas por estudantes na busca e interpretação de documentos institucionais (Okonkwo; Ade-Ibijola, 2021).

Para que os sistemas conversacionais atuem de forma eficaz como mediadores da informação, é necessário considerar os princípios da CI, sobretudo no que diz respeito à curadoria da informação e à organização, representação e

atualização das bases de conhecimento que os sustentam (Reina; Cruz, 2020; Del Sole *et al.*, 2025). A integração entre a organização da informação e a Inteligência Artificial (IA), sendo esta última uma tecnologia em ascensão para o desenvolvimento de sistemas e agentes conversacionais, configura-se, nesse contexto, como um desafio tanto metodológico quanto teórico, ainda pouco explorado nas abordagens voltadas ao desenvolvimento de *chatbots* no âmbito da CI.

Esta pesquisa insere-se nesse cenário ao investigar como os fundamentos da CI podem contribuir para o desenho de sistemas conversacionais mais eficazes, especialmente no que se refere à organização, à recuperação e ao acesso a informações normativas e institucionais no contexto das universidades públicas. Parte-se da premissa de que a mediação informacional, para ser efetiva, deve estar alinhada às necessidades reais dos usuários e ancorada em práticas de curadoria que assegurem a confiabilidade e relevância da informação.

A curadoria da informação, entendida como a atividade de selecionar, organizar, contextualizar e manter atualizados conteúdos relevantes, é uma das atribuições centrais do profissional da informação. Nesse sentido, esse profissional desempenha um papel estratégico na concepção de sistemas que promovam a qualidade, acessibilidade e integridade das informações em ambientes institucionais complexos, como o universitário (Jorente; Padua; Nakano, 2019).

A curadoria de conteúdo informacional, curadoria de informações, ou de conteúdo, ou, ainda, curadoria informacional, diz respeito à seleção de fontes de informações com o objetivo de direcionar o usuário na localização de recursos que possibilitem a otimização no tempo de suas buscas. Seu escopo inclui a importância do refinamento de informações para grupos de usuários com necessidades informacionais específicas, remetendo à ideia de um consumo mais preciso e específico de informação frente ao excesso de fontes (Maia; Santos Neto, 2024; Siebra; Borba; Miranda, 2016).

Dessa forma, ao realizar a curadoria informacional, por meio da construção de uma base de conhecimento, considerando todo esse processo de estruturação, desenvolvimento e curadoria, para um *chatbot* voltado à recuperação de documentos normativos, esta pesquisa busca aplicar os princípios da organização e representação do conhecimento à construção de um sistema conversacional, garantindo que os conteúdos mediados por qualquer tipo de tecnologia sejam relevantes, compreensíveis e atualizados, contribuindo, assim, para o

aprimoramento da comunicação institucional e para o atendimento das demandas informacionais do público alvo das universidades: os estudantes.

Tendo em vista a velocidade das ferramentas tecnológicas e das plataformas para a construção de *chatbots*, esta solução se destaca por ser manter ativa em diversas ferramentas, inclusive em soluções que envolvam IA como recurso de apoio, garantindo, sobretudo, a confiabilidade das informações, independentemente da tecnologia utilizada. Este último elemento, a confiabilidade, é essencial quando se trata de um serviço a ser implementado em uma universidade pública que preza pela qualidade das informações disponibilizadas aos seus discentes.

1.1 Problema e Justificativa

A adoção de *chatbots* no contexto educacional tem se expandido de forma significativa, impulsionada pelo potencial desses agentes conversacionais em promover interatividade e automatizar serviços. Estudos apontam sua utilização em diferentes funções, como ensino e aprendizagem, avaliação, aconselhamento, pesquisa e atividades administrativas (Okonkwo; Ade-Ibijola, 2021)

A literatura também evidencia aplicações específicas, como o atendimento institucional (Ranoliya; Raghuwashi; Singh, 2017; Lee *et al.*, 2019), a oferta de avaliação e *feedback* (Chen; Chen; Lin, 2020) e o suporte à gestão acadêmica em processos de admissão, concessão de bolsas e cobrança de taxas (Huang; Zhu; Gao, 2020). Essa versatilidade pode ser observada, ainda, na organização de conteúdos, na disponibilização de exercícios práticos, na emissão de lembretes de prazos, na explicação de critérios de avaliação, na oferta de orientações acadêmicas e no fornecimento de informações logísticas e administrativas. Diferentes estudos reforçam essa diversidade de aplicações, destacando contribuições voltadas à aprendizagem (Benotti; Martinez; Schapachnik, 2017), à gestão acadêmica (Ismail; Ade-Ibijola, 2019; Mabunda; Ade-Ibijola, 2019) e à usabilidade em interfaces (Rohrig; Heb, 2019).

Em síntese, os *chatbots* têm se mostrado ferramentas promissoras para qualificar os processos comunicacionais no ambiente universitário, promovendo uma mediação mais eficiente entre estudantes e instituições (Clarizia *et al.*, 2018).

Mesmo com todas as evoluções nos sistemas e na infraestrutura, um desafio persistente no contexto das universidades públicas brasileiras é a fragmentação das informações normativas e acadêmicas. A universidade federal é uma pessoa jurídica

de direito público instituída e mantida pela União que possui autonomia didático-científica, administrativa, de gestão financeira e patrimonial, obedecendo ao princípio da indissociabilidade entre ensino, pesquisa e extensão (Andifes, 2002). A dispersão das informações nas universidades pode ser resultado da estrutura organizacional descentralizada, em que múltiplos setores, como colegiados, departamentos, pró-reitorias e secretarias acadêmicas, são responsáveis pela produção, atualização e disponibilização desses conteúdos, sem uma política de padronização ou integração informacional. Essa configuração fragmentada pode representar um desafio adicional ao acesso e à compreensão das informações por parte dos estudantes, especialmente daqueles com menor familiaridade com os trâmites acadêmicos.

Nesse contexto, é possível observar a presença de práticas informacionais que envolvem o uso de redes informais, como consultas a colegas mais experientes ou a grupos em redes sociais, complementando ou antecedendo a busca por fontes institucionais. Esse tipo de mediação informal pode estar relacionado não apenas à dispersão das informações, mas também à natureza técnica da linguagem adotada em muitos documentos institucionais, o que tende a demandar maior esforço de interpretação por parte dos usuários (Kuhlthau, 2004; Case; Given, 2016).

Apesar dos avanços tecnológicos em algoritmos, nas ferramentas de IA e no PLN, o desenvolvimento de *chatbots* ainda carece de metodologias estruturadas para o tratamento da informação. Observa-se uma escassez de *frameworks* que orientem a coleta, organização, representação e curadoria dos dados que alimentam esses sistemas (Reina; Cruz, 2020). Essa ausência de base metodológica compromete a precisão e o contexto das respostas, gerando interações superficiais, mesmo em plataformas tecnicamente sofisticadas (Ferneda, 2021). Além disso, sem modelos de gestão da informação, torna-se complexo constituir bases de conhecimento consolidadas, capazes de integrar dados confiáveis e atualizados e, assim, garantir relevância e coerência nas respostas (Lima, 2020).

Nesse cenário, observa-se a possibilidade de aplicar práticas de curadoria da informação, voltadas à organização e apresentação mais acessível de conteúdos normativos e acadêmicos. O uso de tecnologias baseadas em IA, como os *chatbots* acadêmicos, tem sido explorado como uma alternativa para auxiliar nesse processo, oferecendo respostas automatizadas em linguagem natural, com potencial para ampliar o alcance e a efetividade da comunicação institucional (Mendoza *et al.*, 2022; Villanueva; Aguilar-Alonso, 2021).

Nesse contexto, há questionamentos que envolvem diferentes pilares, como os usuários (Quais são as dúvidas? Como localizam e onde acessam as informações?) e a recuperação e organização da informação (Como estruturar essa informação? A curadoria informacional pode apoiar a recuperação de informações acadêmicas normativas?), que geraram a seguinte questão de pesquisa: Como estruturar e implementar uma base de conhecimento para subsidiar a construção de um chatbot na Escola de Ciência da Informação da Universidade Federal de Minas Gerais (ECI UFMG)?

A justificativa para essa investigação baseia-se na compreensão de que a informação desempenha um papel fundamental no funcionamento das instituições públicas. A qualidade dos serviços educacionais está relacionada à oferta de informações claras, acessíveis e confiáveis, desde o atendimento inicial até a resposta a demandas mais complexas. Elementos como celeridade, transparência, fluidez e segurança ao longo do ciclo de vida da informação contribuem para a efetividade da gestão pública (Santana, 2016).

No âmbito das universidades, observa-se a existência de extensos repositórios e bases de dados institucionais que, quando organizados de forma apropriada, podem subsidiar a tomada de decisão, facilitar o encaminhamento de demandas acadêmicas e contribuir para um ambiente mais eficiente e participativo (Guimarães, 2022).

Nesse cenário, as tecnologias da informação, em especial os *chatbots*, têm sido exploradas como alternativas para apoiar a comunicação e o atendimento no setor público educacional. A literatura da CI tem apontado a importância de integrar seus fundamentos teóricos ao desenvolvimento de tecnologias de mediação da informação em ambientes digitais, embora ainda haja poucos estudos aplicados ao contexto de agentes conversacionais, especificamente (Nascimento Silva, 2023b).

Compreender melhor essa lacuna pode auxiliar no desenvolvimento de soluções tecnológicas mais alinhadas às necessidades informacionais dos usuários e aos princípios que orientam a organização da informação em ambientes acadêmicos (Shawar; Atwell, 2007; Srivastava; Prabhakar, 2020).

Diante disso, esta pesquisa visa contribuir para a elaboração de soluções tecnológicas aplicáveis em contextos institucionais similares, com foco no usuário e no aprimoramento dos serviços acadêmicos.

Esta pesquisa justifica-se em três ordens complementares: profissional, científica e pessoal.

Quanto à ordem profissional, a pesquisa responde a uma demanda institucional por soluções que qualifiquem os serviços de informação acadêmica, considerando os desafios relacionados à fragmentação de conteúdos normativos e à dificuldade de mediação entre usuários e documentos institucionais. Além disso, a demanda parte de uma necessidade real da ECI UFMG, local que forma profissionais da informação e deve ser exemplo de inovações dessa natureza e soluções para a Universidade.

Quanto à ordem científica, pretende-se contribuir com o campo da CI ao se investigar a aplicação de princípios teóricos da área no desenvolvimento de agentes conversacionais, especialmente no que se refere à representação, organização e recuperação, para o efetivo acesso à informação.

Quanto à ordem pessoal, esta pesquisa tem como ponto de partida a experiência prática da autora como bibliotecária na UFMG. Essa vivência permite a articulação entre o conhecimento técnico da área e o desenvolvimento de ferramentas que possam ampliar o acesso à informação e apoiar a comunidade acadêmica em suas demandas cotidianas, considerando as competências do bibliotecário em equipes multidisciplinares e apoiando profissionais da tecnologia.

1.2 Objetivos

Nos próximos tópicos, são apresentados o objetivo geral e os objetivos específicos que orientaram esta dissertação.

1.2.1 Objetivo geral

Estruturar e implementar uma base de conhecimento para subsidiar a implementação de um *chatbot* voltado à recuperação de informações acadêmicas (normas, resoluções e procedimentos) sobre as graduações na Escola de Ciência da Informação da Universidade Federal de Minas Gerais.

1.2.2 Objetivos específicos

Para alcançar o objetivo geral proposto, foram traçados os seguintes objetivos específicos:

- I. Mapear métodos e técnicas utilizados no desenvolvimento e na construção de *chatbots* com foco na informação;
- II. Identificar o comportamento de busca dos estudantes quanto a documentos normativos acadêmicos na UFMG;
- III. Identificar e analisar *chatbots* já implementados no contexto educacional;
- IV. Realizar a curadoria informacional para subsidiar a implementação de uma base de conhecimento voltada à recuperação de informações acadêmicas.

Entende-se como curadoria informacional, no contexto desta pesquisa e para o desenvolvimento de um *chatbot* na ECI UFMG, a construção de uma base de conhecimento com informações acadêmicas (normas, resoluções e procedimentos) para o estabelecimento de regras e definições (perguntas e respostas) e o mapeamento de fluxos de informação para o contexto específico de cursos de graduação.

1.3 Estrutura da dissertação

Esta dissertação está organizada em seis seções. A [Seção 1](#), dedicada à introdução, contextualiza o tema no campo da CI, apresenta o problema de pesquisa, as justificativas e os objetivos geral e específicos. A [Seção 2](#) corresponde ao referencial teórico, no qual são discutidos aspectos relacionados à organização, representação, recuperação e curadoria da informação. A [Seção 3](#) trata dos *chatbots*, apresentando seu histórico, suas diferentes tipologias e arquiteturas, além de suas aplicações. A [Seção 4](#) descreve a metodologia da pesquisa, detalhando as etapas desenvolvidas, bem como as técnicas de coleta, tratamento e análise dos dados utilizados. A [Seção 5](#) apresenta os resultados da revisão de literatura, a [Seção 6](#), o levantamento das necessidades informacionais dos estudantes, a [Seção 7](#), a análise de *chatbots* educacionais com boas práticas e metodologias replicáveis e, por fim, a [Seção 8](#) trata da estruturação e criação da base de conhecimento. Já a [Seção 9](#) reúne as considerações finais, sintetizando o trabalho e destacando suas principais contribuições, limitações e possibilidades de continuidade em estudos futuros.

2. REFERENCIAL TEÓRICO

A CI fundamenta-se em princípios que orientam a mediação entre a informação e seus usuários, buscando garantir que a informação relevante seja disponibilizada ao usuário adequado, no momento certo e pelo meio mais eficaz (Lancaster, 1979; Saracevic, 2007). Entre os aspectos centrais da área relacionados a essa pesquisa, destacam-se a organização e a representação da informação, essenciais para a recuperação e para o processo de curadoria da informação.

A Organização da Informação (OI) é um processo de arranjo de acervos, tradicionais ou eletrônicos, realizado por meio da descrição física e de conteúdo (assunto) de seus objetos informacionais. O produto desse processo descritivo é a representação da informação, entendida como um conjunto de elementos descritivos que representam os atributos de um objeto informacional específico (Brascher e Café, 2008).

A representação da informação diz respeito à forma como a informação é descrita e codificada para possibilitar sua compreensão e seu uso, envolvendo a criação de sistemas de representação que traduzem o conteúdo informacional em formatos acessíveis e interoperáveis, essenciais para a comunicação eficaz entre sistemas e usuários (Campos, 2001; Fujita, 2012; Lima, 2020).

A Recuperação da Informação (RI) trata-se do conjunto de técnicas e métodos que permitem localizar e acessar informações relevantes de maneira rápida e eficaz, especialmente diante do crescimento exponencial dos dados disponíveis. A RI envolve o processamento, a indexação e a seleção da informação, sendo fundamental para atender às necessidades específicas dos usuários e ao contexto das buscas (Salton, 1968; Baeza-Yates; Ribeiro-Neto, 2013; Ferneda, 2012, 2021).

A curadoria da informação consiste no processo de selecionar, organizar, contextualizar e compartilhar informações de modo a atender a um objetivo específico. Trata-se da transformação de dados brutos em conteúdos úteis, estratégicos e de qualidade. Mais do que coletar informações, envolve a filtragem criteriosa para destacar o que é realmente relevante, assegurando que o material disponibilizado seja adequado ao público ou ao propósito. Esse processo facilita tanto a compreensão quanto a tomada de decisões (Pereira; Carvalho, 2023). No contexto digital, a curadoria também compreende a gestão e preservação de conteúdos, assegurando que estejam acessíveis e atualizados para o uso atual e futuro. Assim, a curadoria é um trabalho essencial para garantir a precisão,

relevância e utilidade das informações em diversas áreas, desde o marketing e o jornalismo até a gestão empresarial e a pesquisa científica (Siebra; Borba; Miranda, 2016).

A CI destaca a interlocução e a interdisciplinaridade com outras áreas, como a Ciência da Computação (CC) e a IA, considerando a informação como o motor do ciclo informacional e a disponibilidade da informação como seu objetivo final (Borko, 1968; Coneglian; Gonzalez; Santarém Segundo, 2017), reforçando a importância da relação de seus aspectos para a resolução de problemas da sociedade contemporânea.

2.1 Organização da Informação e do Conhecimento

Brascher e Café (2008) diferenciam a Organização do Conhecimento (OC) da OI a partir do momento em que o processo de descrição se aplica a unidades do pensamento (conceitos) e visa à construção de modelos de mundo que se constituem em abstrações da realidade.

Dahlberg (1993, p. 211, tradução nossa) define a OC como "a ciência que estrutura e organiza sistematicamente unidades do conhecimento (conceitos) segundo seus elementos de conhecimento (características) inerentes e a aplicação desses conceitos e classes de conceitos ordenados a objetos/assuntos".

Nesta pesquisa, foram utilizadas tanto a OI quanto a OC, ao se manusear informações acadêmicas e, também, conceitos desse domínio.

O objetivo principal da OI é tornar a informação acessível, compreensível e prontamente utilizável, visando auxiliar tanto indivíduos quanto instituições na tomada de decisões embasadas e na aquisição de conhecimento de maneira mais eficiente. Segundo Álvares (2013) e Saracevic (2007), a OI vai além de simplesmente criar índices ou listas de títulos; ela requer uma análise minuciosa do conteúdo, do contexto de produção da informação e das necessidades específicas dos usuários. Por conseguinte, a organização não se limita meramente ao armazenamento, mas abrange a estruturação e a representação, que viabilizam um acesso contextualizado e efetivo aos materiais informativos, conforme observado por Lima (2020).

Para que a informação seja organizada de maneira apropriada, é fundamental que ela seja descrita de forma detalhada, o que implica expor suas características e suas relações com outros elementos, possibilitando sua identificação e

contextualização. Essa descrição pode abranger aspectos físicos, relacionados ao suporte da informação, bem como aspectos de conteúdo, que envolvem temas, conceitos e atributos. Nesse contexto, Lancaster (2004) ressalta a importância da uniformidade na utilização de terminologias padronizadas e esquemas conceituais, pois isso garante a eficácia dos sistemas organizacionais. A adoção de vocabulários controlados e de padrões internacionais contribui para a interoperabilidade entre sistemas e para a precisão na RI.

Além disso, a OI deve ser orientada pelo usuário, levando em consideração seus padrões de comportamento, suas motivações e seus objetivos. De acordo com Bates (2002), compreender como os usuários exploram e interagem com os sistemas informativos é crucial para planejar estratégias organizacionais que atendam às suas reais necessidades. Saracevic (2007) complementa essa visão ao afirmar que a usabilidade e a acessibilidade dos sistemas dependem não apenas de uma estrutura de informação bem elaborada, mas também de um entendimento aprofundado do contexto social e cognitivo dos usuários.

2.2 Representação da informação e do conhecimento

A representação da informação, conforme Hjørland (2016) e Brascher e Café (2008), engloba a descrição física e intelectual de documentos, bem como a organização dessas descrições para facilitar o acesso e a recuperação pelos usuários. Fujita (2012) explica que representar a informação envolve traduzir o conteúdo conceitual de um documento para um modelo de dados padronizado, seja em formato bibliográfico ou em metadados, levando em consideração a relevância e o contexto de uso. Conforme Novellino (1996):

A principal característica do processo de representação da informação é a substituição de uma entidade linguística longa e complexa - o texto do documento - por sua descrição abreviada. O uso de tal sumarização não é apenas uma consequência de restrições práticas quanto ao volume de material a ser armazenado e recuperado. Essa sumarização é desejável pois sua função é demonstrar a essência do documento. Ela funciona então como um artifício para enfatizar o que é essencial no documento considerando sua recuperação, sendo a solução ideal para organização e uso da informação (Novellino, 1996, p.38).

A representação da informação é um produto da OI que envolve um conjunto de elementos descritivos que representam os atributos de um objeto informacional específico. Nesse mesmo sentido, a representação do conhecimento é o resultado

da organização de unidades do pensamento (conceitos), refletindo em um modelo de abstração do mundo real, construído para determinada finalidade (Brascher; Café, 2008).

Dessa forma, a representação do conhecimento busca estruturar conceitualmente a realidade, visando formalizar conhecimentos que possam ser compartilhados em diferentes contextos. Diversas ferramentas e métodos, como ontologias, mapas conceituais e Sistemas de Organização do Conhecimento (SOCs), como taxonomias e tesouros, auxiliam nesse processo, mapeando conceitos e significados em um determinado domínio.

É relevante destacar, conforme Dahlberg (1978), o papel fundamental dos (SOCs), como ontologias, taxonomias e tesouros, que possibilitam estruturar conceitos de forma coesa e abrangente, contribuindo para a representação de temas e a conexão entre assuntos relacionados. Bates (2017) ressalta que a representação do conhecimento vai além da simples identificação de temas em documentos, abrangendo, também, as relações epistemológicas entre conceitos e a maneira como as pessoas os compreendem e utilizam.

Retomando a representação da informação, a concepção da informação envolve uma série de procedimentos que visam reconhecer, resumir e documentar as propriedades de um item informativo, de forma a facilitar seu acesso e sua utilização (Lancaster, 2004). Dentro desse contexto, estudiosos da CI identificam, de modo geral, dois tipos principais de representação: a representação que descreve (descritiva) e a representação por temas (temática). Cada uma delas aborda diferentes aspectos do material ou recurso informativo, sendo complementares em ambientes como bibliotecas, bancos de dados e outros SOCs.

Conforme Lima (2020), para representar a informação, é necessário transformar a linguagem natural do item em uma síntese do conteúdo temático. A linguagem natural frequentemente apresenta ambiguidades, nas quais uma palavra pode ter múltiplos significados (polissemia) e um significado pode ser expresso por diversas palavras (sinonímia) (Lima, 2020).

A representação descritiva se concentra na identificação física e formal do recurso. Nesse tipo de representação, são registrados detalhes como título, autor, edição, data de publicação, editora, local de publicação e outros elementos que facilitam a localização e diferenciação de um documento em um acervo ou sistema bibliográfico (Cleveland; Cleveland, 2001).

Normas como Resource Description and Access (RDA) e Anglo-American Cataloguing Rules (AACR2) são exemplos de diretrizes utilizadas para padronizar esses metadados descritivos, buscando facilitar a troca de registros e a interoperabilidade entre catálogos (Cunha; Cavalcanti, 2008).

Por outro lado, a representação por temas (ou de conteúdo) foca nos tópicos e nas ideias principais presentes no recurso, com o propósito de auxiliar os usuários a encontrar documentos pertinentes, conforme suas necessidades de informação (Lancaster, 2004).

Esse processo de representação engloba a análise do conteúdo, a seleção de termos, a indexação e a categorização, podendo ser realizado por meio de vocabulários controlados (como tesouros, taxonomias e ontologias como Simple Knowledge Organization System (SKOS), classificações bibliográficas (tal como a Classificação Decimal de Dewey) e linguagens documentárias específicas (Currás, 2010; Foskett, 1996).

A representação por temas requer tanto habilidades técnicas em indexação quanto a compreensão do conteúdo temático do documento, de modo a associar os conceitos presentes ao vocabulário mais apropriado (Lima, 2020). Dessa maneira, a representação que descreve e a representação por temas desempenham funções complementares na organização e na busca da informação (Currás, 2010; Cleveland; Cleveland, 2001; Foskett, 1996).

Em se tratando de instrumentos para o processo de descrição, os metadados podem ser compreendidos como dados estruturados que descrevem as características essenciais de um recurso, seja ele um documento, produto ou serviço, possibilitando sua identificação, organização e recuperação em diferentes contextos informacionais. Elementos como autor, título e assunto constituem exemplos clássicos de metadados aplicados tanto em repositórios digitais quanto em páginas *web* ou bases de dados (Gilliland, 2016).

No ambiente digital, destaca-se a adoção de padrões consolidados, como o Dublin Core (2020), que define um conjunto mínimo de elementos de descrição com características de simplicidade, universalidade semântica, abrangência internacional e extensibilidade, atributos que favorecem a padronização e a interoperabilidade entre sistemas. Nesse sentido, os metadados assumem papel central na representação e no acesso à informação, funcionando como um elo entre os recursos documentais e os mecanismos de recuperação que os disponibilizam ao usuário (Robredo; Brascher, 2010).

No contexto desta pesquisa, serão considerados, especialmente, os elementos da classificação (taxonomias) e as relações dos conceitos, bem como os SOCs. Os SOCs são objetos que cumprem um importante papel de padronização da terminologia adotada para a organização e recuperação de informações ao delimitarem o uso de termos e definirem os conceitos e as relações de uma determinada área do conhecimento de forma compartilhada e consensual. Esses sistemas são fruto do processo de organização do conhecimento (Brascher e Café, 2008). Conforme Dahlberg (2006), os SOCs são tipos de sistemas conceituais ou, também, tipos de representações do conhecimento, frutos do processo de organização do conhecimento.

De acordo com a classificação de Hodge (2000), os SOCs são divididos em três categorias: (1) listas de termos; (2) classificações e categorias; e (3) listas de relacionamentos, conforme apresentado no Quadro 1.

Quadro 1 - Classificação dos SOCs

Classificação	Tipos de SOCs			
	Lista de autoridades	Glossários	Dicionários	Gazetteers
Listas de termos	Listas de termos que controlam as variações de nomes para entidades, como nomes de países, indivíduos ou instituições.	Lista de termos com definições, geralmente de um assunto ou domínio específico.	Lista alfabética de palavras e suas definições. Seu escopo é mais geral que o dos glossários.	Dicionário de nomes de lugares e acidentes geográficos, tais como cidades, rios, vulcões. Quando georreferenciados, os <i>gazetteers</i> apresentam as coordenadas para a localização de lugares na superfície da Terra.
Classificações e categorias	Cabeçalhos de assunto	Esquemas de classificação, taxonomias e esquemas de categorização		
	Conjunto de termos controlados que representam os assuntos de uma coleção.	Esquemas utilizados para agrupar entidades em classes mais gerais. Os esquemas de categorização são geralmente usados para reunir os termos de tesouros em tópicos. As taxonomias são usadas em modelos orientados a objeto e em sistemas de gestão do conhecimento para indicar grupos de objetos baseados em características particulares.		
Listas de relacionamentos	Tesouros	Redes semânticas	Ontologia	
	Conjunto de termos que representam	Estrutura de conceitos e	Estrutura de conceitos e representação dos relacionamentos	

	conceitos e as relações de equivalência, hierárquicas e associativas que se estabelecem entre eles.	termos em forma de rede ou teia. Os conceitos são nós e os relacionamentos expandem-se a partir desses nós.	complexos entre eles, incluindo regras de inferência e axiomas.
--	---	---	---

Fonte: adaptado de Hodge (2000).

2.3 Recuperação da Informação

A RI pode ser compreendida a partir de múltiplas perspectivas teóricas. Ferneda (2003) propõe que a RI funciona como um processo comunicacional, estabelecendo uma ponte entre produtores de conteúdo e usuários da informação, caracterizando-se pela interação dinâmica entre emissor e receptor na satisfação de demandas informacionais específicas. Complementarmente, Saracevic (2007) situa a RI no contexto de um campo tecnológico emergente, resultante da convergência interdisciplinar entre a CI e a (CC), evidenciando sua natureza híbrida e seu potencial inovador.

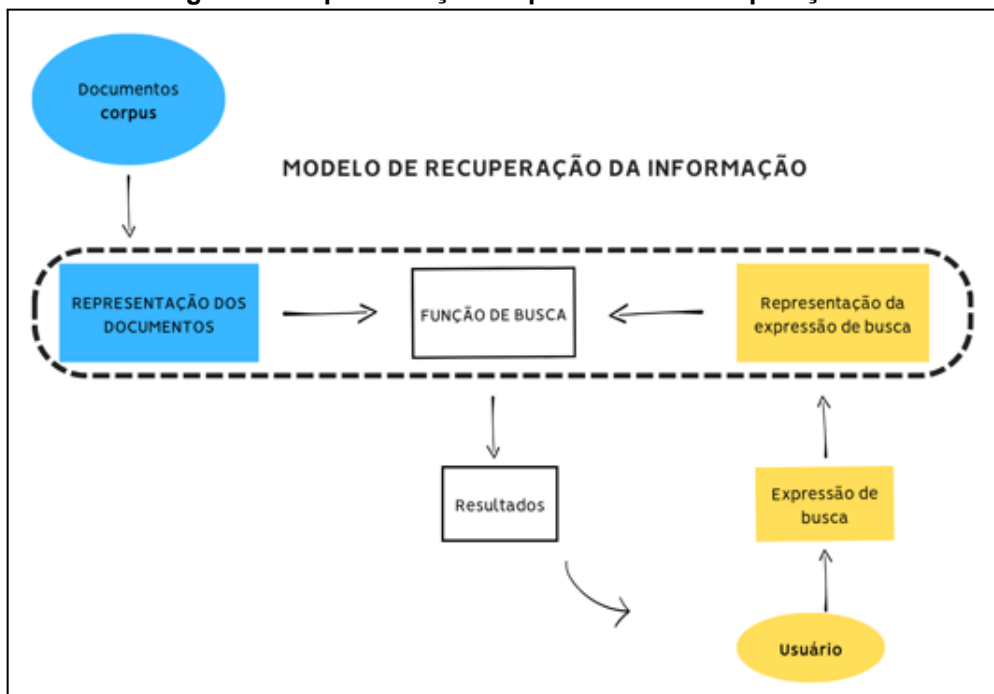
Saracevic (2007) explica, ainda, que a RI envolve os aspectos intelectuais da descrição da informação e suas especificações para recuperação, abrangendo quaisquer sistemas, técnicas ou máquinas usadas para esse fim. Lancaster (1979) esclarece que a RI é semelhante à pesquisa bibliográfica, definindo-a como um processo para localizar uma coleção específica de documentos. Robredo (2003) afirma que a RI engloba a finalidade do trabalho documental, incluindo procedimentos de seleção, aquisição, descrição bibliográfica, análise e indexação.

Com o recente aumento no volume de dados digitais, a RI tornou-se um domínio crucial para organizar, armazenar e acessar rapidamente os metadados. A RI é um campo de pesquisa e aplicação prática que visa compreender e utilizar métodos e técnicas para identificar os documentos, textos ou conteúdos mais relevantes para um determinado conjunto de termos ou consultas de pesquisa (Manning; Schütze, 1999).

Em termos mais simples, sua função é atender com eficiência e precisão às necessidades de informação dos usuários em meio a repositórios e bancos de dados cada vez mais vastos e diversificados. Ferneda (2012) explica que o processo

de recuperação de informações busca identificar, em um conjunto de documentos, aqueles que atendem a uma determinada necessidade de informações. A Figura 1 ilustra uma representação do processo de recuperação de informações.

Figura 1 - Representação do processo de recuperação



Fonte: adaptado de Fernalda (2012, p.14).

Um *corpus* é definido como um artefato que representa ou expressa um objeto, uma ideia ou uma informação por meio de sinais gráficos e icônicos, como palavras, imagens, mapas e símbolos, bem como elementos sonoros e visuais, que pode ser armazenado eletronicamente. A representação de documentos visa descrever e identificar, individualmente, cada elemento de um conjunto documental, com base em seus conteúdos específicos.

Na RI, a expressão de busca é a forma pela qual o usuário comunica sua necessidade de informação ao sistema, sendo composta por termos representativos dessa necessidade. A representação da expressão de busca deve assemelhar-se àquela utilizada para os documentos, facilitando a comparação entre a consulta do usuário e os documentos do *corpus*. Essa semelhança é essencial para que o sistema de recuperação de informação localize com precisão os documentos relevantes que correspondam à necessidade de informação do usuário.

A função de busca compara a expressão de busca com as representações dos documentos no *corpus* para determinar a relevância desses documentos em relação à consulta do usuário, permitindo a identificação e a recuperação dos

documentos mais adequados. O resultado da busca deve consistir em documentos que satisfaçam as necessidades informacionais dos usuários.

O usuário, por sua vez, é aquele que interage com o sistema em busca de documentos que atendam às suas necessidades informacionais, sendo fundamental compreender como eles se envolvem nesse processo. Para atingir tal compreensão, é essencial conhecer e entender as necessidades informacionais dos usuários, como menciona Le Coadic (2004).

2.3.1 Modelos de Recuperação da Informação

Os modelos de RI formam a base teórica e prática para encontrar, categorizar e classificar documentos conforme as necessidades de quem os busca. Tipicamente, eles são categorizados em modelos convencionais (ou baseados em números) e modelos ativos, cada um com maneiras diferentes de estabelecer a ligação entre as consultas e o conteúdo, possibilitando a descoberta de dados relevantes em coleções de informações (Baeza-Yates; Ribeiro-Neto, 1999).

Os modelos tradicionais de RI, de acordo com Ferneda (2003; 2012), utilizam métodos matemáticos e analíticos para avaliar a similaridade entre os termos da consulta e os termos encontrados nos documentos, empregando técnicas de ponderação, análise de frequência de termos e avaliação de importância.

A diversidade dessas estratégias, que progrediu desde as operações booleanas até as técnicas probabilísticas e analíticas, reflete o avanço contínuo do campo e a sinergia entre a Inteligência Artificial e as análises avançadas de linguagem natural (Manning; Raghavan; Schütze, 2009). Essa integração entre disciplinas, combinando CI, Tecnologia da Informação (TI), entre outras áreas, tem como objetivo melhorar a maneira como a informação é estruturada, acessada e utilizada (Ferneda, 2003; Saracevic, 2007).

Esta seção foi incluída na pesquisa com o objetivo de apoiar a compreensão sobre os *chatbots* (tipologias, arquitetura e funcionamento), que serão apresentados na [Seção 3](#), os quais estão diretamente relacionados à RI, uma vez que esses modelos são utilizados em sua estrutura interna. Assim, nas seções subsequentes, serão apresentados os modelos clássicos e dinâmicos que fundamentam os estudos da área.

2.3.1.1 Modelos clássicos de Recuperação da Informação

2.3.1.1.1 Modelo booleano

O modelo booleano, uma das abordagens sistemáticas iniciais da RI, baseia-se na lógica proposicional e emprega operadores booleanos (AND, OR e NOT) para combinar termos de pesquisa. Ele categoriza os documentos em relevantes ou irrelevantes, com base em operações binárias, recuperando apenas aqueles que atendem precisamente aos critérios especificados. Sua estrutura é considerada adequada para sistemas que exigem precisão em consultas estruturadas. No entanto, sua limitação está na falta de ponderação ou adaptação às nuances semânticas, o que o torna menos eficaz para consultas flexíveis ou expressões de linguagem natural. Apesar de sua simplicidade e facilidade de implementação, o modelo booleano carece de mecanismos para lidar com variações ou gradientes de relevância para refinar as respostas em sistemas interativos (Souza, 2006).

2.3.1.1.2 Modelo vetorial

Introduzido por Salton (1968), o modelo vetorial apresentou uma evolução significativa na RI ao propor a representação de documentos e consultas como vetores em um espaço multidimensional. A relevância é avaliada por meio da similaridade entre esses vetores, calculada a partir de medidas como a similaridade de cosseno, que permite a classificação dos resultados por grau de relevância (Baeza-Yates; Ribeiro-Neto, 2013). Esse modelo se mostra eficaz para consultas mais naturais, pois considera a importância do termo por meio de pesos (Ferneda, 2012) como TF-IDF (Term Frequency-Inverse Document Frequency). A classificação por similaridade aumenta a relevância dos resultados, alinhando-se melhor às necessidades do usuário de expressar consultas de forma menos estruturada (Belkin; Croft, 1992).

2.3.1.1.3 Modelo probabilístico

O modelo probabilístico atribui probabilidades de relevância a documentos, com base em consultas, oferecendo uma abordagem mais adaptável aos sistemas de RI. Baseado no Teorema de Bayes, esse modelo estima as probabilidades de relevância dos documentos com relação à presença ou ausência de termos em documentos classificados anteriormente (Baeza-Yates; Ribeiro-Neto, 2013). Seu

recurso central é o aprendizado contínuo, que ajusta as probabilidades de relevância dinamicamente à medida que novas informações são fornecidas, aumentando a precisão ao longo do tempo e oferecendo uma solução flexível para demandas complexas (Ferneda, 2012).

2.3.1.2 Modelos dinâmicos de Recuperação da Informação

A evolução dos modelos clássicos levou ao surgimento de modelos dinâmicos que incorporam técnicas de IA. A IA, segundo Russel e Norvig (2021), tem como objetivo simular as habilidades cognitivas humanas para resolver problemas e aprender. Esses modelos reconhecem a função ativa do usuário na definição da relevância do documento, entendendo-a como contextual e dependente das necessidades e preferências do usuário em um determinado momento (Baeza-Yates; Ribeiro-Neto, 2013). Consequentemente, a representação de documentos torna-se dinâmica e responsiva às necessidades emergentes de informações nas interações com os sistemas de RI (Saracevic, 2007; Ingwersen; Jarvelin, 2005).

2.3.2.1 Lógica fuzzy

A lógica difusa aplicada à RI representa graus de relevância, em vez de classificações binárias, introduzindo o conceito de “conjuntos difusos”. Essa abordagem permite aos sistemas lidar com incertezas e ambiguidades linguísticas, alinhando a recuperação mais de perto com a avaliação de relevância humana em interações com sistemas interativos (Ferneda, 2012). Os pesos dos termos de indexação são transformados em graus de relevância, refletindo escalas graduais de relevância na classificação (Souza, 2006).

2.3.2.2 Redes neurais

As Redes Neurais Artificiais (RNAs), inspiradas em neurônios biológicos, aprendem padrões complexos em grandes volumes de dados. Aplicadas à RI, as RNAs ajustam os pesos dos termos e identificam as relações semânticas de forma adaptativa, com base em exemplos de documentos relevantes ou *feedback* do usuário (Baeza-Yates; Ribeiro-Neto, 2013). Esse método de aprendizado se mostra eficiente em bancos de dados textuais heterogêneos e em constante evolução,

aumentando a robustez da classificação e alinhando-se às necessidades do usuário (Ferneda, 2012).

2.3.2.3 Algoritmos Genéticos

Os Algoritmos Genéticos (GAs), em RI, representam uma abordagem de otimização inspirada na seleção natural. As representações de documentos são tratadas como “códigos genéticos”, com os AGs sendo usados para otimizar a seleção de termos, refinar as estratégias de indexação ou ajustar os pesos nos modelos de classificação (Ferneda, 2003, 2012; Baeza-Yates; Ribeiro-Neto, 2013). Essa abordagem facilita a exploração de espaços de pesquisa complexos, identificando as configurações mais adequadas às demandas dos usuários em cenários com significativa heterogeneidade de dados.

Vale ressaltar que os modelos de RI são frequentemente usados em conjunto com outras abordagens e tecnologias em sistemas interativos. A integração entre IA, Web Semântica e Ciência de Dados surge como uma tendência para redesenhar os processos de recuperação e organização de informações (Coneglian; Segundo, 2022). Nesse sentido, o desenvolvimento de modelos informacionais que consideram a linguagem natural e a representação de dados semânticos é fundamental para aprimorar a pesquisa em sistemas interativos (Lemos; Martins; Souza, 2023).

2.3.4 Inteligência Artificial e Inteligência Artificial Generativa

O campo da IA surgiu oficialmente na década de 1950, em um contexto de avanços tecnológicos impulsionados pelo pós-guerra. O termo foi utilizado pela primeira vez em 1956, durante a Conferência de Dartmouth, quando pesquisadores como John McCarthy, Marvin Minsky, Allen Newell e Herbert Simon propuseram investigar a possibilidade de construir máquinas capazes de simular aspectos do raciocínio humano (Russell; Norvig, 2021). Naquele período, a IA foi fortemente marcada por abordagens simbólicas, também chamadas de IA clássica, que utilizavam regras lógicas, sistemas especialistas e representações formais do conhecimento como base para a resolução de problemas.

Durante as décadas seguintes, o campo passou por diferentes fases de entusiasmo e descrédito, conhecidas como os “invernos da IA”, motivadas por limitações técnicas e altas expectativas não correspondidas. A partir da década de 1980, com os avanços em *hardware* e o desenvolvimento de sistemas especialistas,

a IA retomou espaço em aplicações específicas, como diagnósticos médicos e sistemas de suporte à decisão. Contudo, foi apenas no início do século XXI que a área alcançou um novo patamar, sendo impulsionada pelo aumento do poder computacional, pela disponibilidade de grandes volumes de dados (*big data*) e pelo avanço das técnicas de aprendizado de máquina (*machine learning*) e aprendizado profundo (*deep learning*) (Goodfellow; Bengio; Courville, 2016).

Nesse novo cenário, a IA passou a ser entendida de forma mais ampla como o estudo e desenvolvimento de agentes inteligentes, capazes de perceber o ambiente, processar informações e agir de modo a maximizar resultados (Russell; Norvig, 2021). Essa definição inclui desde sistemas baseados em regras até modelos estatísticos e conexionistas, capazes de aprender padrões a partir de dados. O aprendizado profundo, em especial, representou um marco ao permitir o treinamento de RNAs com múltiplas camadas, possibilitando avanços expressivos em áreas como reconhecimento de imagens, PLN e sistemas de recomendação.

Para Ferneda (2012), a aplicação da IA à RI introduz uma dimensão dinâmica antes ausente nos modelos tradicionais. Ao incorporar técnicas como as RNAs, os sistemas passam a simular processos cognitivos humanos, permitindo que as representações documentais sejam continuamente ajustadas conforme o *feedback* dos usuários. Essa capacidade de aprendizado adaptativo amplia a precisão e a relevância da busca, uma vez que possibilita lidar com a ambiguidade da linguagem natural, reconhecer padrões complexos e responder de forma mais personalizada às necessidades informacionais, superando os limites dos modelos estáticos de recuperação.

A partir de 2017, com a introdução dos *transformers* (Vaswani *et al.*, 2017), consolidou-se uma nova fase da IA, marcada por modelos de linguagem de larga escala, Large Language Models (LLMs). Esses modelos são capazes de aprender representações semânticas complexas, permitindo que sistemas computacionais processem, interpretem e gerem linguagem natural com alto grau de sofisticação. Esse movimento abriu caminho para o desenvolvimento da chamada Inteligência Artificial Generativa (IAG), caracterizada pela capacidade de criar novos conteúdos, textos, imagens, sons ou vídeos a partir de dados previamente aprendidos.

Conceitualmente, a IAG pode ser compreendida como uma vertente da IA que utiliza técnicas avançadas de aprendizado profundo para gerar novos conteúdos que preservam coerência, relevância e adequação ao contexto (Trindade; Oliveira, 2024). A IAG diferencia-se das formas anteriores de IA por não apenas classificar,

recomendar ou prever informações, mas também por produzir saídas inéditas. Ferramentas como o GPT (Generative Pre-trained Transformer), da OpenAI, exemplificam esse avanço. No caso dos LLMs, a arquitetura baseada em *transformers* permite treinar sistemas em grandes *corpus* textuais, possibilitando a generalização de tarefas complexas e a adaptação a diferentes contextos de uso (Bommasani *et al.*, 2022). Assim, a IAG amplia as fronteiras da automação, permitindo aplicações em diversas áreas, como educação, saúde, ciência e comunicação, ao mesmo tempo em que suscita debates éticos sobre autoria, confiabilidade e uso responsável.

Em síntese, tanto a IA quanto a IAG refletem a evolução de um campo que, desde suas origens simbólicas até os atuais modelos generativos, busca reproduzir e ampliar as capacidades humanas de processar informação. No âmbito da CI, essas tecnologias podem ser compreendidas como instrumentos de mediação, organização e recuperação, oferecendo novas possibilidades, mas também desafios quanto à transparência, à confiabilidade e ao controle dos sistemas.

Assim, a IA, em especial a generativa, representa um avanço tecnológico, um novo paradigma na relação entre humanos e máquinas. Sua capacidade de criar, simular e interagir a partir de dados e informações utilizados no seu treinamento, considerando vários contextos e de maneira cada vez mais próxima à cognição humana, amplia o potencial de inovação, ao mesmo tempo em que impõe o desafio de desenvolver práticas de governança, transparência e responsabilidade em sua aplicação. No campo da informação acadêmica, esses avanços oferecem novas possibilidades de mediação e acesso a conteúdos normativos, aspecto diretamente vinculado ao escopo desta pesquisa.

Por fim, retomando o assunto principal desta seção, a RI constitui um elemento essencial para o acesso efetivo ao conhecimento na contemporaneidade. Conforme destacado por Ferneda (2021), a efetividade dos sistemas de informação está intrinsecamente relacionada à sua capacidade de organizar, indexar e recuperar conteúdos relevantes de acordo com as demandas dos usuários. Assim, a RI transcende a mera localização de documentos, constituindo-se como uma mediação fundamental entre o usuário e os conteúdos, viabilizando respostas contextualizadas e ajustadas às necessidades específicas de cada situação. Este processo, além de técnico, revela-se também social, pois depende do entendimento das práticas informacionais e das características dos públicos atendidos.

Nessa perspectiva, Baeza-Yates e Ribeiro-Neto (2013) enfatizam que a RI representa o elemento-chave para transformar grandes volumes de dados em conhecimento acessível e útil. Sem mecanismos eficientes de busca e filtragem, a mera disponibilidade de informação não se converte em acesso efetivo, uma vez que os usuários podem se perder diante do excesso de dados dispersos e desorganizados.

Portanto, a RI revela-se indispensável para garantir que a grande quantidade de informações disponíveis na atualidade seja efetivamente acessível e útil aos seus usuários.

2.4 Curadoria da informação

O termo "curadoria" tem suas raízes na tradição museológica, referindo-se à prática de cuidado, organização e mediação de acervos culturais. O curador, nesse contexto, atua como um elo entre os objetos e o público, atribuindo-lhes valor simbólico e histórico por meio de uma gestão criteriosa de coleções (Barbosa, 2013). Com o advento da internet e a explosão de informações em ambientes digitais, o conceito expandiu seu escopo, tornando-se uma prática estratégica e interdisciplinar, especialmente no campo da CI.

Nesse novo cenário, a curadoria é vista como um processo crítico de seleção, organização, contextualização e disseminação de informações. Seu objetivo central é transformar o volume massivo de dados em conhecimento significativo, confiável e acessível para diferentes públicos (Reina; Cruz, 2020; Aleixo *et al.*, 2020). A prática, no ambiente digital, envolve ações como a filtragem de conteúdos, o uso de metadados, o controle de vocabulários e a preservação digital, garantindo a qualidade, autenticidade e recuperabilidade da informação (Bates, 2005; Siebra, Borba; Miranda, 2016).

2.4.1 Definições e tipologias

A expansão do conceito de curadoria resultou no surgimento de diferentes modalidades, que refletem objetivos e contextos distintos. Para o propósito deste trabalho, é fundamental distinguir e delimitar a curadoria da informação de conceitos correlatos, como a curadoria digital e a curadoria de conteúdo, para justificar a abordagem teórica adotada.

A curadoria digital está vinculada à dimensão técnica de gestão e preservação de objetos digitais. Seu foco principal é a infraestrutura e os procedimentos sistemáticos, que asseguram que os dados e conteúdos digitais mantenham sua integridade, autenticidade e acessibilidade a longo prazo. Conforme Siebra, Borba e Miranda (2016), o desafio central dessa modalidade é combater a obsolescência tecnológica, garantindo a longevidade dos acervos digitais por meio de estratégias de metadados e preservação.

A curadoria de conteúdo é uma prática mais ligada às mídias digitais e ao marketing cujo objetivo é selecionar, agregar e redistribuir informações para públicos específicos, visando ao engajamento e à promoção de uma marca ou um serviço (Pereira; Carvalho, 2023).

No entanto, é a curadoria da informação que melhor representa o escopo e a profundidade necessários para esta pesquisa. Ela ultrapassa a mera gestão técnica da curadoria digital e a agregação de conteúdo da curadoria de conteúdo. A curadoria informacional é definida como a gestão qualificada do ciclo da informação, em que a prioridade não é apenas a preservação, mas a relevância, a confiabilidade e a adequação do conteúdo às necessidades dos usuários (Aleixo *et al.*, 2020). Essa prática enfatiza o papel do profissional da informação como um mediador qualificado, que filtra a sobrecarga informacional e a desinformação, oferecendo um conteúdo confiável e contextualizado.

O Quadro 2, proposto por Guerra González (2017, p. 102), estabelece um paralelo entre as atividades realizadas pelo bibliotecário acadêmico e aquelas atribuídas ao curador de conteúdo digital. A comparação evidencia que, embora atuem em ambientes distintos, ambos compartilham práticas relacionadas à seleção, avaliação e disseminação da informação.

Quadro 2 - Paralelo entre as atividades realizadas pelo bibliotecário acadêmico e o curador de conteúdo

Atividades do curador de conteúdo digitais	Descrição da atividade	Atividades tradicionais dos bibliotecários acadêmicos
Busca por informação relevante	A curadoria de conteúdo busca a relevância, o que se evidencia nos julgamentos de valor manifestos na seleção de um conteúdo em detrimento de outro. O bibliotecário acadêmico torna-se um agente ativo ao declarar: "na minha opinião, isso é o que você deve saber".	Busca por informações relevantes que atendam a solicitações de informações específicas por parte do usuário.

Seleção de conteúdo a partir do contexto do usuário	Uma das funções da curadoria de conteúdo é dar significado ao conteúdo que outros criam; isso faz com que seja necessário aproximar os julgamentos sobre a relevância de um conteúdo de acordo com o contexto em que os usuários se encontram. A escolha do conteúdo, seu tratamento intelectual e sua apresentação serão condicionados pelo contexto em que a comunicação ocorrer.	Seleção de conteúdo considerando-se os metadados.
Avaliação sistemática de conteúdo	O curador de conteúdo avalia a qualidade de forma sistemática para escolher o melhor em um tema.	Avaliação dos recursos de informação.
Criação de coleções de conteúdo digital ou mídia que os reúna	Mesclar conteúdo digital em diferentes formatos para criar uma coleção ou um meio que forneça informações coerentes em uma espécie de narrativa sobre um tema, uma área de conhecimento etc.	Recuperação da informação e criação de catálogos.
Difusão dos conteúdos curados	Escolher plataformas ou canais online específicos para a divulgação do conteúdo digital já curado, levando em consideração as características dos diferentes tipos de utilizadores e o seu grau de literacia informacional.	Desenvolvimento de programas de atividades e exposições em que o acervo da biblioteca se destaca.

Fonte: Guerra González (2017, p. 102)

No caso do curador de conteúdo, as ações estão direcionadas à busca, filtragem e organização de materiais digitais, valorizando a relevância e o contexto em que a informação será consumida. O bibliotecário acadêmico, em suas funções tradicionais, exerce um papel semelhante ao realizar a mediação informacional, respondendo a demandas específicas dos usuários, selecionando conteúdos com base em critérios técnicos e criando instrumentos de organização e recuperação, como catálogos e coleções.

Observa-se, portanto, que a atividade de curadoria digital, conforme descrita por Guerra González (2017), aproxima-se das práticas consolidadas da biblioteconomia, sobretudo no que se refere à avaliação crítica da informação, ao tratamento intelectual e à adequação às necessidades do usuário. Essa correspondência reforça a ideia de que a curadoria informacional, aplicada em ambientes digitais e, mais especificamente, no desenvolvimento de *chatbots*, pode ser compreendida como uma atualização e expansão das funções históricas do bibliotecário acadêmico, adaptadas às novas demandas tecnológicas e comunicacionais.

Neste trabalho, adota-se o termo curadoria informacional para destacar sua abrangência no âmbito da CI, englobando as dimensões técnica, cognitiva, social e ética da gestão de informações. Essa escolha reforça o papel estratégico do profissional da informação como um agente ativo na construção do conhecimento, capaz de conectar as demandas dos usuários com a oferta de conteúdos qualificados.

2.4.2 Aplicações da curadoria na Ciência da Informação

No campo da CI, a curadoria tornou-se um pilar central para o tratamento, a organização e a recuperação do conhecimento, especialmente diante do crescimento exponencial de dados digitais. A curadoria informacional não se limita à organização passiva de dados; ela se manifesta como um processo sistemático de seleção, avaliação, organização e disseminação, com o objetivo de garantir a autenticidade e a relevância dos conteúdos para diferentes contextos (Reyes *et al.*, 2017).

O profissional da informação atua como curador, definindo critérios de qualidade, monitorando fontes, filtrando conteúdos e disseminando-os em formatos que atendam às necessidades específicas dos usuários (Aleixo *et al.*, 2020). Essa prática ganha ainda mais relevância no ambiente digital, onde se integram tecnologias como as *big data* e sistemas de recomendação para aprimorar a recuperação da informação e a experiência do usuário.

Além disso, a curadoria informacional assume um papel ético e social, especialmente no enfrentamento da desinformação. Ao promover o acesso a conteúdos de qualidade e valorizar a transparência, o curador contribui para o desenvolvimento do pensamento crítico e para a construção de um ambiente informacional mais robusto (Reina; Cruz, 2020).

Assim, a curadoria informacional se apresenta como uma prática essencial para a sustentabilidade de ambientes acadêmicos, institucionais e corporativos. Ela permite que a informação qualificada seja acessada de forma eficiente, apoiando a tomada de decisão e o desenvolvimento do conhecimento.

No contexto contemporâneo, sistemas conversacionais como os *chatbots* vêm assumindo um papel crescente na mediação entre usuários e conteúdos informacionais. Sua atuação não se limita à simulação de diálogo, mas envolve a identificação de intenções comunicativas, a interpretação de linguagem natural e, em

alguns casos, a adaptação das respostas ao contexto do usuário (Adamopoulou; Moussiades, 2020). Avanços recentes em IA permitiram o desenvolvimento de modelos mais sofisticados, baseados em técnicas de aprendizado de máquina, que identificam padrões de comportamento e ajustam respostas com base no histórico de interações (Griol *et al.*, 2008).

Contudo, antes da implementação de um *chatbot*, destaca-se a importância da curadoria humana como etapa crítica para garantir a qualidade, relevância e contextualização dos dados que alimentarão o sistema. Em contextos marcados por fluxos massivos de informação e pela diversidade de fontes, como apontado por Coneglian, Gonzalez e Santarém Segundo (2017), em estudos sobre Big Data Analytics, o papel do profissional da informação torna-se central. Sua atuação não se limita à organização de dados, mas envolve uma mediação qualificada entre as necessidades dos usuários e os recursos tecnológicos disponíveis, atuando em sinergia com profissionais da computação e gestores (Coneglian; Gonzalez; Santarém Segundo, 2017). Nesse sentido, esta pesquisa aborda aspectos da curadoria informacional para a construção de *chatbots*.

3. CHATBOTS

Este capítulo apresenta alguns dos fundamentos teóricos sobre os sistemas conversacionais, abordando a evolução histórica dos *chatbots*, suas funções enquanto mediadores informacionais, suas classificações técnicas e aspectos relacionados à sua arquitetura e ao seu funcionamento.

O termo *chatbot* consolidou-se na literatura especializada com o avanço de sistemas computacionais capazes de interagir com usuários por meio da linguagem natural. Segundo Labadze, Grigolia e Machaidze (2023), *chatbots* são aplicativos de *software* ou interfaces *web* projetados para imitar conversas humanas por meio de interações de texto ou voz utilizando (PLN) e algoritmos de IA para simular diálogos naturais. Comarella e Café (2008, p.12) definem esses sistemas como programas que podem funcionar tanto como interlocutores informais quanto como ferramentas comerciais especializadas, enfatizando sua interatividade e a diversidade de propósitos práticos.

A diversidade terminológica encontrada na literatura científica reflete diferentes perspectivas teóricas e aplicações práticas desses sistemas. Shawar e Atwell (2007) utilizam o termo agentes conversacionais para destacar a ênfase na linguagem natural e na utilidade das respostas em diálogos estruturados. Essa abordagem fundamenta-se na teoria dos atos de fala, de Austin (1965), e na prática conversacional, de Grice (1975), estabelecendo bases linguísticas sólidas para compreender a comunicação humano-computador.

Brandtzæg e Følstad (2017) adotam a denominação "assistentes virtuais" ou "inteligentes", acentuando o papel desses sistemas em tarefas específicas como atendimento ao cliente e suporte técnico. Esta perspectiva alinha-se aos sistemas de apoio à decisão, enfatizando a funcionalidade prática sobre a simulação conversacional. Adamopoulou e Moussiades (2020) observam que essas variações terminológicas revelam diferentes enfoques conceituais e aplicações técnicas, moldadas pela evolução das tecnologias subjacentes e pelas teorias de interação humano-computador que as fundamentam.

Na CI, os *chatbots* são compreendidos como mediadores tecnológicos que transformam as práticas de organização, acesso e uso da informação. Esta perspectiva alinha-se com a teoria da mediação da informação desenvolvida por Almeida Júnior (2015), que considera a mediação como um processo ativo de

transformação tanto do usuário quanto da própria informação durante o processo comunicacional.

Os *chatbots*, nesse contexto teórico, atuam como interfaces inteligentes que transcendem a simples recuperação de informação. Eles contextualizam dados por meio da conversação, personalizando respostas conforme o perfil e as necessidades específicas do usuário (Pimenta; Almeida Júnior, 2009). Isso cria novas possibilidades de acesso informacional em ambientes institucionais, educacionais e públicos, democratizando o conhecimento a partir de interfaces mais naturais e intuitivas (Soares; Nascimento Silva, 2024).

3.1 Histórico e evolução

A trajetória dos *chatbots* está diretamente associada à evolução da IA, do PLN e das práticas de mediação digital (Adamopoulou; Moussiades, 2020). O teste de Turing, em 1950, foi a primeira experiência para se avaliar a capacidade dos computadores em entender os usuários. Desde os primeiros sistemas baseados em regras até os modelos atuais de linguagem generativa, diferentes gerações de *chatbots* foram desenvolvidas, com características técnicas, propósitos e contextos distintos. Nesta dissertação, adota-se a periodização proposta por Adamopoulou e Moussiades (2020) e Attigeri, Agrawal e Kolekar (2024), que organiza a evolução dos *chatbots* em quatro fases principais:

1. *Chatbots* baseados em regras (1966-1990);
2. Sistemas estruturados via *web* (1990-2010);
3. Assistentes inteligentes (2010-2020);
4. Modelos generativos (2020-presente).

Cada uma dessas fases evidencia transformações tecnológicas e mudanças no papel dos *chatbots* como intermediadores da interação entre usuários e sistemas informacionais.

No primeiro período: *Chatbots* baseados em regras (1966-1990), o marco inicial é o ELIZA, desenvolvido por Joseph Weizenbaum (1966), no Massachusetts Institute of Technology (MIT). O sistema simulava um terapeuta rogeriano e operava por meio da reformulação de perguntas com base em padrões simples de texto. Em

1972, foi criado o PARRY, projetado para simular uma pessoa com esquizofrenia paranoide, incorporando estados internos rudimentares. Ainda naquela época, ELIZA e PARRY interagiram via ARPANET, demonstrando, simbolicamente, o potencial comunicacional dessas tecnologias (Adamopoulou; Moussiades, 2020).

Em seguida, no segundo período: Sistemas estruturados via *web* (1990-2010), com a popularização da internet, os *chatbots* começaram a ser aplicados em ambientes virtuais. Destaca-se o ALICE (1995), que utilizava a linguagem AIML para estruturar conversas com base em padrões hierárquicos. Outros exemplos, como o SmarterChild (2001) e o Mitsuku (2005), contribuíram para a expansão da personalização e da interação contextualizada, abrindo caminho para o uso em plataformas de mensagens e ambientes comerciais.

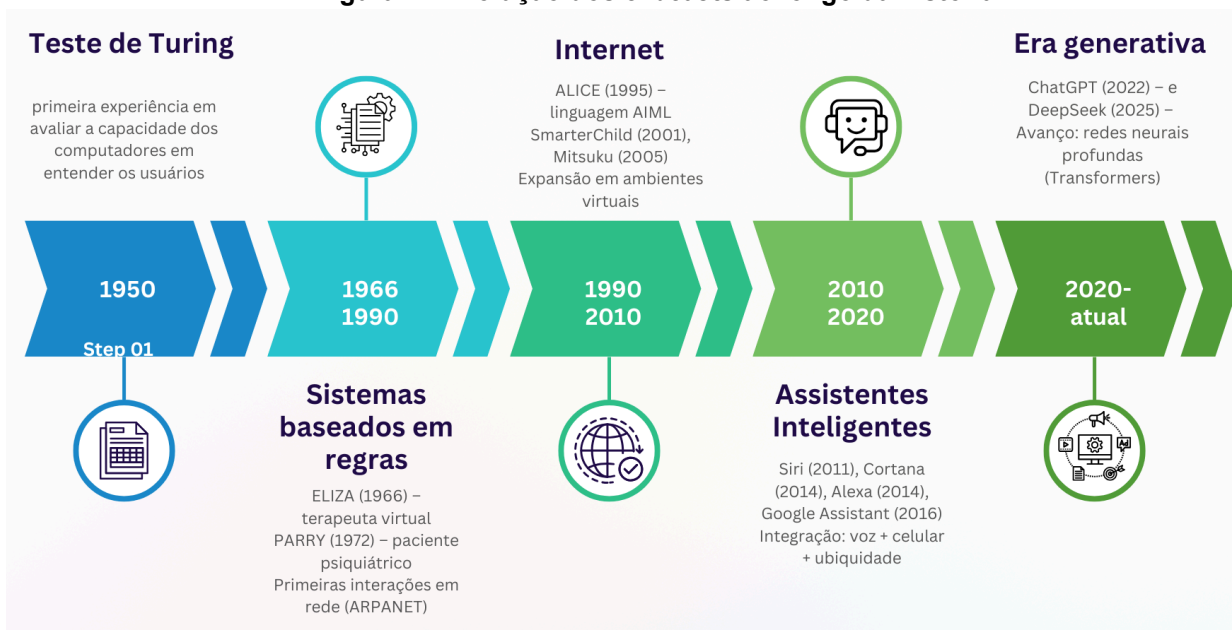
O terceiro período: Assistentes inteligentes (2010-2020), é marcado pela integração dos *chatbots* a dispositivos móveis e sistemas operacionais. Assistentes como Siri (2011), Cortana (2014), Alexa (2014) e Google Assistant (2016) passaram a combinar aprendizado de máquina, reconhecimento de voz e conectividade em tempo real. Essa nova geração reflete o paradigma da computação ubíqua (Weiser, 1991), operando de forma integrada e contínua no cotidiano dos usuários. A partir de 2020, observa-se um avanço tecnológico relevante em relação aos períodos anteriores, representado pelos modelos generativos.

Por fim, no quarto período, a Modelos generativos (2020-presente), surgiu uma nova classe de *chatbots* movidos por arquiteturas de redes neurais profundas, em especial o modelo Transformer (Vaswani *et al.*, 2017). Nesse estágio, sistemas como o ChatGPT (lançado em 2022 pela OpenAI) consolidaram o conceito de *chatbot* generativo, combinando grandes volumes de dados com aprendizado profundo para produzir textos fluentes e contextualizados (OpenAI, 2025). Outro marco recente foi o DeepSeek (janeiro/2025), criado pela DeepSeek AI (Hangzhou DeepSeek Artificial Intelligence Basic Technology Research Co., Ltd.) sobre o modelo *open-source* DeepSeek-R1, que, até 27 de janeiro, já liderava os aplicativos gratuitos mais baixados na App Store dos EUA, graças ao seu baixo custo de treinamento e código aberto.

A Figura 2 apresenta os principais marcos evolutivos organizados cronologicamente, destacando as transformações técnicas e conceituais relacionadas à mediação comunicacional entre humanos e sistemas computacionais. Além disso, a figura exemplifica alguns dos *chatbots* mais representativos desenvolvidos em cada período, ilustrando como essas tecnologias

acompanharam e refletiram tais avanços, perpassando períodos clássicos de diálogo estruturado, sistemas baseados em regras, *chatbots* estatísticos e, finalmente, agentes generativos, como ChatGPT e DeepSeek. A evolução dos *chatbots* retrata o aperfeiçoamento de seus recursos técnicos e de sua capacidade de interagir com humanos, além das mudanças conceituais sobre como a comunicação mediada por sistemas passou a ser compreendida e aplicada ao longo dos anos.

Figura 2 - Evolução dos *chatbots* ao longo da história



Fonte: adaptado de Adamopoulou e Moussiades (2020) e Attigeri, Agrawal, Kolekar (2024).

3.2 Tipos de *chatbots*

Compreender os diferentes tipos de *chatbots* é essencial para definir estratégias de curadoria informacional adequadas a cada contexto de aplicação. Contudo, na literatura, existem diversas classificações e dimensões. Assim, tendo em vista as diferentes abordagens, foi elaborado o Quadro 3, que apresenta as tipologias identificadas, bem como sua definição, conforme os autores.

Quadro 3 - Tipologias de *chatbots*

Classificação	Tipologia	Descrição	Autores
	FAQ Bots	“Bots de Perguntas Frequentes (FAQ Bots) são sistemas de conversação que funcionam por mapeamento léxico de termos-chave a um repositório de perguntas pré-definidas, de onde extraem a resposta exata ao usuário. Não geram texto livremente, mas devolvem a resposta estática associada à melhor	Mataloni Neto (2023, p.

Funcionalidade		correspondência encontrada na base”.	71-73)
	<i>Chatbots</i> conversacionais	“Chatbots conversacionais combinam busca por similaridade semântica com módulos de PLN, identificando intenções (‘intents’), extraíndo entidades, mantendo contexto e gerando respostas adaptadas ao fluxo dialogal, ainda que fragmentado, permitindo diálogos abertos e não-lineares”.	
Estrutura de interação	Interação fechada	Sistemas estruturados em fluxos baseados em botões, menus hierárquicos ou perguntas pré-definidas. O usuário não digita livremente, mas seleciona opções que guiam todo o diálogo, garantindo caminhos controlados e minimizando entradas inválidas.	Reina e Cruz (2020)
	Interação aberta	Modalidade que permite ao usuário digitar livremente no <i>chat</i> , sendo a interpretação realizada por módulos de PLN que lidam com ambiguidade, contexto e variações linguísticas.	
	Modelo híbrido	Combina, em diferentes partes do diálogo, fluxos estruturados por regras com interação livre via <i>chat</i> , acessando o componente de PLN sempre que não houver correspondência exata com as regras definidas.	
Tecnologia	Baseados em menus ou botões	<i>Chatbots</i> elementares que operam como árvores de decisão: apresentam menus de botões a partir dos quais cada escolha apresenta o próximo conjunto de opções até a resposta desejada. Não há campo de texto livre – toda a navegação é guiada por cliques.	IBM (2025)
	Baseados em regras	Sistemas que empregam a lógica condicional “se-então” (<i>decision trees</i> avançadas); fazem detecção básica de palavras-chave para mapear <i>inputs</i> a respostas predefinidas, atuando como FAQs interativas, mas sem a capacidade de lidar com perguntas não previstas.	
	IAG	<i>Chatbots</i> que usam LLMs para gerar, livremente, textos, imagens e sons de alta qualidade, adaptando-se ao estilo do usuário e criando conteúdo novo em resposta às consultas, com maior fluidez e personalização.	
	<i>Voice bots</i>	Sistemas multimodais que integram Speech-to-Text e Text-to-Speech, permitindo interação por voz; incluem desde IVR simples até <i>bots</i> avançados, com IAs capazes de entender e sintetizar fala em conversas naturais.	
	<i>Chatbots</i> híbridos	Integram regras condicionais com aprendizado de máquina, oferecendo estrutura determinística para cenários	

		previsíveis e flexibilidade adaptativa para variações não previstas, combinando confiabilidade e capacidade de generalização.	
--	--	---	--

Fonte: adaptado de Reina e Cruz (2020), Mataloni Neto (2023) e IBM (2025).

A síntese apresentada no Quadro 1 evidencia a diversidade de abordagens que caracterizam os *chatbots*, tanto em termos de funcionalidade quanto de estrutura de interação e tecnologias subjacentes. Essa tipologia, fundamentada em diferentes autores, como Reina e Cruz (2020), Mataloni Neto (2023) e IBM (2025), permite compreender que a escolha do modelo mais adequado deve considerar o contexto de uso, os objetivos institucionais e as necessidades específicas dos usuários. Reconhecer as distinções entre FAQ Bots, sistemas baseados em regras, *chatbots* conversacionais, híbridos e soluções integradas por IAG contribui para o desenho de estratégias informacionais mais eficientes e alinhadas às demandas contemporâneas de mediação e acesso à informação.

3.3 Arquitetura e funcionamento de *chatbots*

Em linhas gerais, a estrutura dos *chatbots* pode ser vista como um conjunto organizado de elementos que engloba interfaces de interação, módulos de PLN e bases de conhecimento subjacentes, que são responsáveis por organizar e recuperar informações.

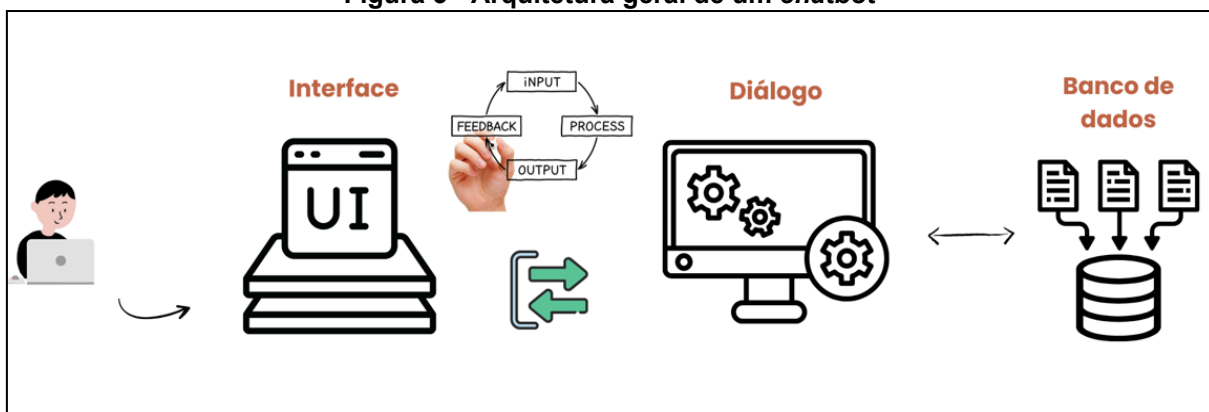
A arquitetura de um *chatbot* combina técnicas de PLN com modelos de inferência, muitas vezes baseados em redes neurais profundas, para analisar mensagens e produzir respostas. Villanueva e Aguilar-Alonso (2021) descrevem esse procedimento como um fluxo de trabalho que se inicia com a entrada do usuário, passando por etapas de tokenização e vetorização. Jurafsky e Martin (2009) explicam que a tokenização converte texto em unidades analíticas, enquanto a vetorização cria representações numéricas para análise semântica. Essas fases, em conformidade com técnicas de pré-processamento de RI, preparam o texto para a próxima etapa.

A estrutura desses sistemas envolve processos de pré-processamento, identificação de intenções e geração de respostas, integrando recursos de PLN e aprendizado computacional, para garantir interações naturais e contextualmente pertinentes. Lancaster (2004) destaca a importância da organização do

conhecimento por meio de índices e vocabulários controlados para uma recuperação eficaz de informações, essencial em cenários como o suporte. O aprimoramento constante dos modelos, alimentado por interações reais ou simuladas, possibilita que o *chatbot* adapte suas respostas às nuances da linguagem humana, atendendo às expectativas dos usuários.

A seguir, é apresentada uma breve explicação das tecnologias e dos elementos que compõem a arquitetura e o funcionamento de um *chatbot*, conforme ilustrado nas Figuras 3 e 4. Segundo Yang *et al.* (2023), o sistema se conecta a três elementos essenciais: a interface do usuário, o gerenciador de diálogo e o banco de dados, conforme demonstrado na Figura 3 e detalhado a seguir.

Figura 3 - Arquitetura geral de um *chatbot*



Fonte: adaptado de Yang *et al.* (2023).

A) Interface do usuário: facilita a troca de mensagens entre os usuários e o *chatbot*, no qual os usuários inserem perguntas e recebem as respostas. É o ambiente visual e funcional por onde o usuário interage com o sistema, inserindo perguntas e recebendo respostas. Ela traduz as ações humanas em comandos compreensíveis pelo *chatbot* e apresenta as respostas de forma clara e acessível.

B) Gerenciador de diálogo: trata-se do sistema responsável por realizar operações para compreender as perguntas dos usuários e fornecer uma resposta apropriada. Nessa arquitetura, a plataforma é utilizada para processar a linguagem natural, o que viabiliza o desenvolvimento de um diálogo fluente.

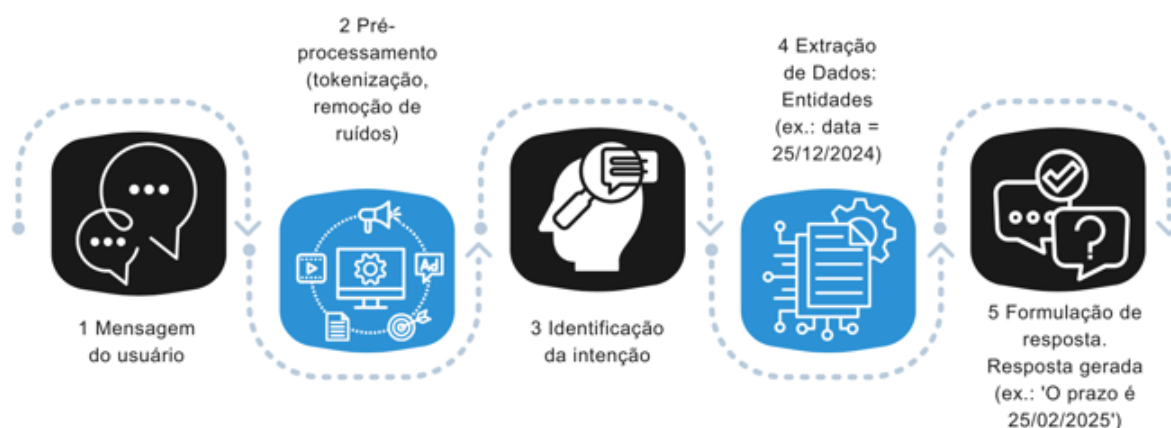
C) Banco de dados: armazena um extenso conjunto de dados organizados e relacionados que podem ser acessados de forma eficiente. A base de informações contém o conhecimento do *chatbot*, armazenado na nuvem e integrado ao sistema.

O funcionamento de um *chatbot* inicia-se com a análise semântica e sintática da mensagem do usuário, transformando sequências de caracteres em representações internas que codificam o significado do texto por meio de técnicas como tokenização e vetorização (Jurafsky; Martin, 2009). Em seguida, modelos de aprendizado de máquina, frequentemente redes neurais profundas, inferem a intenção do interlocutor e selecionam ou geram a resposta mais adequada, utilizando bases de conhecimento ou estratégias de similaridade semântica (Safi *et al.*, 2020). Ainda segundo os autores, a precisão desse processo depende de algoritmos treinados em grandes volumes de dados, permitindo ao *chatbot* refinar sua compreensão ao longo do tempo.

O funcionamento desses sistemas depende da integração entre entrada do usuário, processamento algorítmico e saída informacional, sendo influenciado pela qualidade da modelagem de dados e pela capacidade de aprendizado contínuo.

O funcionamento de um *chatbot* envolve uma sequência de etapas que transformam a entrada do usuário em uma resposta coerente, apoiando-se em princípios de PLN e RI. A Figura 4 apresenta o funcionamento do *chatbot* e, em seguida, cada etapa é descrita conforme o autor.

Figura 4 - Fluxo básico de funcionamento de um *chatbot*



Fonte: adaptado de Villanueva e Aguilar-Alonso (2021).

(1) Mensagem do usuário: a interação com o *chatbot* inicia-se quando o usuário envia um enunciado ou comando por meio de canais como aplicativos, *websites* e plataformas de mensagens instantâneas (ex.: WhatsApp). Esse conteúdo chega como texto bruto, pronto para ser processado. Jurafsky e Martin (2009)

destacam que a análise linguística começa com a captura precisa da entrada, que serve como base para toda a *pipeline* de PLN. Shawar e Atwell (2007) complementam, dizendo que a escolha do canal influencia a estrutura da mensagem, exigindo flexibilidade do sistema para lidar com variações de formato e linguagem natural.

(2) Pré-processamento/Análise de conteúdo: na etapa de pré-processamento, o *chatbot* remove ruídos (ex.: pontuação irrelevante, abreviações e caracteres especiais) e realiza a tokenização, segmentando o texto em unidades analíticas, como palavras ou frases. Manning, Raghavan e Schütze (2009) indicam que esse processo é essencial para converter texto bruto em uma representação estruturada, de forma adequada aos métodos de PLN. A tokenização, segundo Bengio *et al.* (2003), facilita a identificação de padrões linguísticos, enquanto técnicas como *stemming* ou lematização podem ser aplicadas para normalizar termos, alinhando-se às ideias de Lancaster (2004) sobre padronização em RI.

(3) Identificação da intenção (*intents*): o sistema de PLN analisa a mensagem para determinar a intenção do usuário, utilizando algoritmos de classificação treinados em grandes conjuntos de dados. Hirschberg e Manning (2015, p. 261) afirmam que "a identificação de intenções depende de modelos estatísticos que categorizam enunciados com base em exemplos prévios", permitindo ao *chatbot* compreender a finalidade comunicativa (ex.: consulta ou solicitação). Jurafsky e Martin (2009) reforçam a etapa em sistemas conversacionais, pois conectam a entrada do usuário a ações específicas no *backend* do *chatbot*.

(4) Extração de dados: após identificar a intenção, o *chatbot* extrai entidades relevantes da mensagem, como datas, nomes ou locais, em um processo conhecido como "extração de entidades nomeadas" (NER). Jurafsky e Martin (2009) observam que essa etapa enriquece a interpretação semântica ao fornecer contexto adicional ao diálogo. Por exemplo, em "Qual o prazo da matrícula em 25/02/2025?", a data é uma entidade crítica. Belkin e Croft (1992) destacam que a extração de dados é uma ponte entre a RI tradicional e os sistemas interativos, ampliando a capacidade de resposta personalizada.

(5) Formulação de resposta: ao final, o *chatbot* gera ou recupera uma resposta com base na intenção e nas entidades extraídas, utilizando uma base de conhecimento, Application Programming Interfaces (APIs) ou regras predefinidas. Jurafsky e Martin (2009) explicam que o mecanismo de geração de respostas pode variar desde *scripts* simples até modelos de linguagem avançados, como os

descritos por Villanueva e Aguilar-Alonso (2021). A resposta é, então, retornada ao usuário pelo mesmo canal, concluindo o ciclo interativo. Adamopoulou e Moussiades (2020) acrescentam que a integração com bases de conhecimento estruturadas, como ontologias, potencializa a pertinência das respostas em contextos complexos.

Com isso, é possível identificar diversas tipologias, arquiteturas e formatações, que deverão ser analisadas para aplicação conforme o domínio e as ferramentas utilizadas para a construção do *chatbot*.

4 METODOLOGIA

Esta seção descreve os procedimentos metodológicos adotados para o desenvolvimento da pesquisa, apresentando a caracterização, o delineamento, os métodos de coleta e análise de dados e as etapas que compõem o processo investigativo. Buscou-se proporcionar transparência e possibilitar a replicabilidade do estudo, explicitando as escolhas técnicas e teóricas que fundamentam os resultados obtidos.

4.1 Caracterização da pesquisa

Esta pesquisa adota o método de estudo de caso, voltado à investigação do ambiente informacional da Escola de Ciência da Informação da Universidade Federal de Minas Gerais (ECI UFMG). Segundo Gil (2002) e Lakatos e Marconi (2003), o estudo de caso é apropriado quando se busca examinar um fenômeno contemporâneo em seu contexto real, permitindo compreender fatores estruturais, tecnológicos e comunicacionais que impactam o acesso à informação.

A escolha deste contexto específico fundamenta-se na necessidade de compreender, em profundidade, o problema informacional situado no ambiente universitário: a recuperação e o acesso a documentos normativos por parte dos estudantes. O ambiente da ECI UFMG apresenta características típicas das universidades públicas brasileiras, onde normas acadêmicas, regulamentos, editais e procedimentos encontram-se dispersos em múltiplos ambientes digitais institucionais, refletindo a estrutura organizacional descentralizada da instituição.

Esta pesquisa caracteriza-se como aplicada, com abordagem qualitativa e quantitativa, classificando-se como exploratória e descritiva. O desenho metodológico articula quatro etapas complementares, correspondentes aos objetivos específicos da pesquisa. Cada etapa contribui para a construção progressiva da base de conhecimento: a primeira estabelece o referencial teórico-metodológico; a segunda identifica as necessidades informacionais dos usuários; a terceira analisa soluções existentes; e a quarta integra esses elementos para a estruturação e implementação de uma base de conhecimento aplicável ao contexto estudado.

Quanto ao método, o estudo de caso permite a investigação detalhada do problema informacional no contexto específico da ECI/UFMG, considerando suas particularidades organizacionais, sua estrutura documental e o perfil dos usuários.

Esta abordagem possibilita compreender como diferentes setores (colegiados, departamentos e secretarias acadêmicas) produzem e disponibilizam conteúdos normativos, bem como identificar as dificuldades enfrentadas pelos estudantes no acesso a essas informações.

Quanto à natureza, a pesquisa é aplicada, pois busca a resolução de um problema prático identificado no contexto universitário: a dificuldade de acesso à informação normativa. Conforme Gil (2002), a pesquisa aplicada visa gerar conhecimentos com finalidades práticas e direcionados à solução de problemas específicos da realidade institucional, propondo uma solução concreta e replicável para o problema identificado.

Quanto à abordagem, adota-se uma perspectiva qualitativa, pois esta permite analisar o fenômeno em sua complexidade, considerando aspectos subjetivos, contextuais e institucionais envolvidos na mediação informacional. São incorporados, também, elementos quantitativos para dimensionar a frequência das necessidades informacionais identificadas, caracterizando uma abordagem mista pontual (Gil, 2002).

Quanto aos objetivos, a pesquisa é exploratória por investigar um tema ainda pouco desenvolvido de forma integrada na literatura da CI: a curadoria da informação para o desenvolvimento de uma base de conhecimento para subsidiar sistemas conversacionais aplicados ao contexto acadêmico. Simultaneamente, ela é descritiva ao mapear as características dos *chatbots* acadêmicos existentes e as necessidades informacionais dos estudantes relacionadas ao acesso a documentos regulatórios no contexto específico da universidade estudada.

4.2 Lócus da pesquisa

O lócus desta pesquisa é a Escola de Ciência da Informação da Universidade Federal de Minas Gerais (ECI UFMG), unidade acadêmica que oferece cursos de graduação em Biblioteconomia, Arquivologia e Museologia, além de dois programas de pós-graduação em CI (mestrado e doutorado). A Escola completou 75 anos em 2025 e segue sendo referência nacional e internacional para a área de CI.

A escolha da ECI/UFMG como contexto de investigação fundamenta-se em critérios estratégicos que fortalecem a relevância e aplicabilidade da pesquisa. Primeiramente, a ECI constitui um ambiente informacional complexo e representativo das universidades públicas brasileiras, onde coexistem múltiplos

níveis de ensino (graduação e pós-graduação). A natureza da própria Escola de Ciência da Informação oferece um contexto privilegiado para investigar questões relacionadas à organização, representação e recuperação da informação.

O ambiente informacional da UFMG e da ECI caracteriza-se pela diversidade de fontes normativas, como: normas, resoluções, editais de seleção e bolsas, regulamentos específicos dos cursos e programas, manuais de procedimentos acadêmicos, calendários institucionais e formulários para solicitações diversas. Essas informações encontram-se em diferentes *sites* e portais institucionais da UFMG e da ECI, refletindo uma fragmentação informacional.

A população-alvo da pesquisa compreende estudantes de graduação da ECI/UFMG, que incluem discentes dos cursos de Biblioteconomia, Arquivologia e Museologia. Essa população foi selecionada por representar usuários de diferentes perfis formativos, apesar da proximidade temática, e demandas informacionais distintas, o que possibilita uma análise dos comportamentos de busca por informações normativas no contexto universitário. O período de investigação corresponde ao ano de 2025, contemplando tanto a coleta de dados quanto a análise documental das normas vigentes.

4.3 Etapas da pesquisa

A pesquisa foi estruturada em quatro etapas principais, articuladas de forma complementar, conforme detalhado no Quadro 4:

Quadro 4 - Proposta de pesquisa

Etapa	Objetivo específico	Natureza	Abordagem	Finalidade	Técnica de pesquisa	Procedimentos	Artefato
1	Mapear métodos e técnicas utilizados no desenvolvimento e construção de <i>chatbots</i> com foco na informação.	APLICADA	Qualitativa e quantitativa	Exploratória Descritiva	Pesquisa bibliográfica Revisão de literatura	<ul style="list-style-type: none"> → Adaptação e execução de protocolo para revisão de literatura; → Busca e seleção de trabalhos em bases de dados reconhecidas (Scopus, Web of Science, Scielo e BDTD); → Análise quantitativa dos documentos selecionados; → Análise qualitativa com base na técnica de análise de conteúdo de Bardin (2016); → Sistematização e apresentação dos resultados. 	Seção 5
2	Identificar e analisar o comportamento de busca dos estudantes universitários em relação ao acesso a documentos normativos acadêmicos.		Qualitativa e quantitativa	Exploratória Descritiva	Pesquisa documental	<ul style="list-style-type: none"> → Análise quantitativa e qualitativa dos dados coletados em um questionário com perguntas fechadas e abertas; → Sistematização e apresentação dos resultados. 	Seção 6

3	Identificar e analisar <i>chatbots</i> desenvolvidos para o contexto acadêmico.		Qualitativa	Exploratória Descritiva	Pesquisa bibliográfica Análise de conteúdo	<ul style="list-style-type: none"> → Leitura exploratória dos 26 estudos (recorte da revisão) sobre <i>chatbots</i> do contexto educacional; → Definição de categorias de análise (metodologia explícita, fluxos, ferramentas e documentação); → Análise do conteúdo para identificar padrões, temas e discussões presentes nos artigos; → Sistematização e apresentação dos resultados. 	Seção 7
4	Realizar a curadoria informacional para subsidiar a implementação de uma base de conhecimento voltada à recuperação de informações acadêmicas		Qualitativa e quantitativa	Exploratória Descritiva	Aplicada	<p>Realização da curadoria propondo um modelo com as etapas necessárias para a construção da base de conhecimento:</p> <ul style="list-style-type: none"> → Levantamento e análise documental dos regulamentos institucionais da UFMG e da ECI UFMG; → Seleção e organização das informações relevantes para o estudo de caso, considerando a perspectiva dos estudantes; → Estruturação da base de conhecimento, discutida com a literatura, identificando conceitos, termos controlados, expressões e equivalências, bem como a aplicação de técnicas de organização da informação; → Implementação (popular e validar) da base de conhecimento com informações acadêmicas; → Indicação de possíveis fluxos de navegação → Sistematização e apresentação dos resultados da curadoria informacional. 	Seção 8

Fonte: elaborado pela autora (2025).

Os procedimentos metodológicos foram realizados em quatro etapas. A primeira etapa consolida o referencial teórico-metodológico a partir da revisão da literatura sobre metodologias para a construção de *chatbots*, mapeando como eles vêm sendo construídos, quais técnicas e metodologias são utilizadas, e identificando, sobretudo, a lacuna existente quanto ao tratamento da informação, bem como as metodologias adotadas para realizar essa curadoria.

A segunda etapa consistiu em analisar os resultados de um questionário aplicado aos estudantes da UFMG, com o objetivo de compreender o comportamento de busca dos estudantes de graduação, a fim de identificar como buscam normas acadêmicas, quais dificuldades enfrentam nesse processo e quais documentos geram o maior número de dúvidas. Esses resultados originaram as categorias que orientaram a construção da base de conhecimento.

A terceira etapa envolveu a análise de *chatbots* educacionais já implementados, identificados na revisão de literatura realizada na Etapa 1, a fim de identificar padrões, limitações e boas práticas replicáveis ao contexto universitário, especialmente no que se refere ao contexto informacional, e, também, a estrutura, como a modelagem de intenções, entidades e fluxos documentados, que serviram de referência prática para a modelagem da base de conhecimento.

Por fim, a quarta etapa sintetiza os achados anteriores, concentrando-se na modelagem da curadoria e na implementação da base de conhecimento, em que os dados coletados e as boas práticas identificadas foram sistematizadas em conceitos, perguntas, respostas e possíveis fluxos conversacionais para subsidiar a construção de um *chatbot* voltado à recuperação de informações acadêmicas (normas, resoluções e procedimentos) sobre as graduações da ECI UFMG.

A seguir, são detalhadas as quatro etapas que constituem o desenvolvimento desta dissertação, organizadas de forma a apresentar uma visão estruturada do processo adotado.

4.4 Etapa 1: Revisão da literatura

Para a adequada condução da pesquisa, foi necessário, inicialmente, investigar a construção de *chatbots* no contexto da CI, bem como as ferramentas, plataformas, linguagens e metodologias utilizadas, por meio do desenvolvimento de

uma revisão de literatura. Esse estudo corresponde ao primeiro objetivo específico: mapear métodos e técnicas utilizados no desenvolvimento e construção de *chatbots* com foco na informação, cujos procedimentos e resultados foram apresentados na [Seção 5.1](#), compondo, também, o referencial teórico desta dissertação, uma vez que a revisão foi essencial para a identificação das lacunas a serem investigadas.

4.5 Etapa 2: Comportamento de busca dos estudantes de graduação

A segunda etapa da pesquisa corresponde ao segundo objetivo específico: identificar o comportamento de busca dos estudantes quanto a documentos normativos acadêmicos na UFMG. Para isso, foi realizada uma pesquisa desenvolvida no âmbito do projeto Recuperação de Informação em modelos de Inteligência Artificial: o papel da Ciência da Informação, ao qual a autora desta dissertação está vinculada como membro.

O questionário da pesquisa de opinião pública (ANEXO A) foi aplicado entre 16 de dezembro de 2024 e 15 de março de 2025, por meio da plataforma Google Forms, escolhida por sua acessibilidade, popularidade entre os estudantes e facilidade de integração com diferentes meios de divulgação. Três estratégias principais foram utilizadas para ampliar o alcance da pesquisa: (1) compartilhamento em grupos de WhatsApp vinculados à UFMG; (2) envio via listas de e-mail institucionais da graduação e pós-graduação; e (3) afixação de materiais em locais de grande circulação no *campus*, como bibliotecas, restaurantes universitários e praças de serviço. Essa combinação buscou assegurar a diversidade e representatividade da amostra, além de considerar perfis diferenciados de estudantes.

A análise dos resultados dessa coleta de dados, referente ao comportamento de busca dos alunos de graduação, apresentada na [Seção 6](#), foi utilizada, nesta etapa da pesquisa, como subsídio para a curadoria informacional da base de conhecimento, tendo em vista seu público alvo: alunos da graduação da UFMG.

4.6 Etapa 3: Análise de *chatbots* da educação

A realização desta etapa correspondeu ao cumprimento do terceiro objetivo específico da pesquisa: identificar e analisar *chatbots* já implementados no contexto

educacional. Para isso, foram analisados 26 estudos sobre *chatbots* acadêmicos, previamente identificados na revisão de literatura ([Seção 5.1](#)). O objetivo principal foi selecionar os estudos que apresentassem modelos, metodologias, documentação detalhada dos passos e fluxos de informação considerados mais replicáveis para a autora, com foco na curadoria da informação e no potencial de adaptação ao contexto desta pesquisa.

Para avaliar a replicabilidade dos estudos analisados, foram consideradas as seguintes categorias de análise:

- Metodologia explícita: a clareza e o detalhamento na descrição das etapas de desenvolvimento do *chatbot* ou a adoção de *frameworks* e *softwares* reconhecidos na área, incluindo descrições passo a passo dos métodos utilizados;
- Fluxo de informação estruturado: detalhamento dos componentes essenciais do *chatbot*, como módulos de PLN, diálogo e bases de conhecimento, além de representação visual do fluxo de informação por meio de diagramas e fluxogramas;
- Ferramentas e documentação disponível: utilização de plataformas de desenvolvimento abertas e disponibilização de documentação clara das etapas de desenvolvimento.

4.7 Etapa 4 - Curadoria informacional

Esta etapa corresponde ao quarto e último objetivo específico: realizar a curadoria informacional para subsidiar a implementação de uma base de conhecimento voltada à recuperação de informações acadêmicas, ou seja, normas, resoluções e procedimentos sobre as graduações na ECI UFMG. Foram utilizados como subsídios os resultados da Etapa 1 ([Seção 5.1](#)), revisão de literatura sobre desenvolvimento e construção de *chatbots*, Etapa 2 ([Seção 6](#)), resultados sobre a pesquisa survey realizada para investigar comportamento de busca dos estudantes da UFMG, e os achados apresentados na Etapa 3 ([Seção 7](#)), que analisa experiências com *chatbots* educacionais, destacando metodologias, fluxos de informação e aspectos replicáveis para o desenvolvimento da solução proposta. A

curadoria informacional para estruturação e implementação da base de conhecimento foi constituída de cinco etapas, conforme apresentado a seguir:

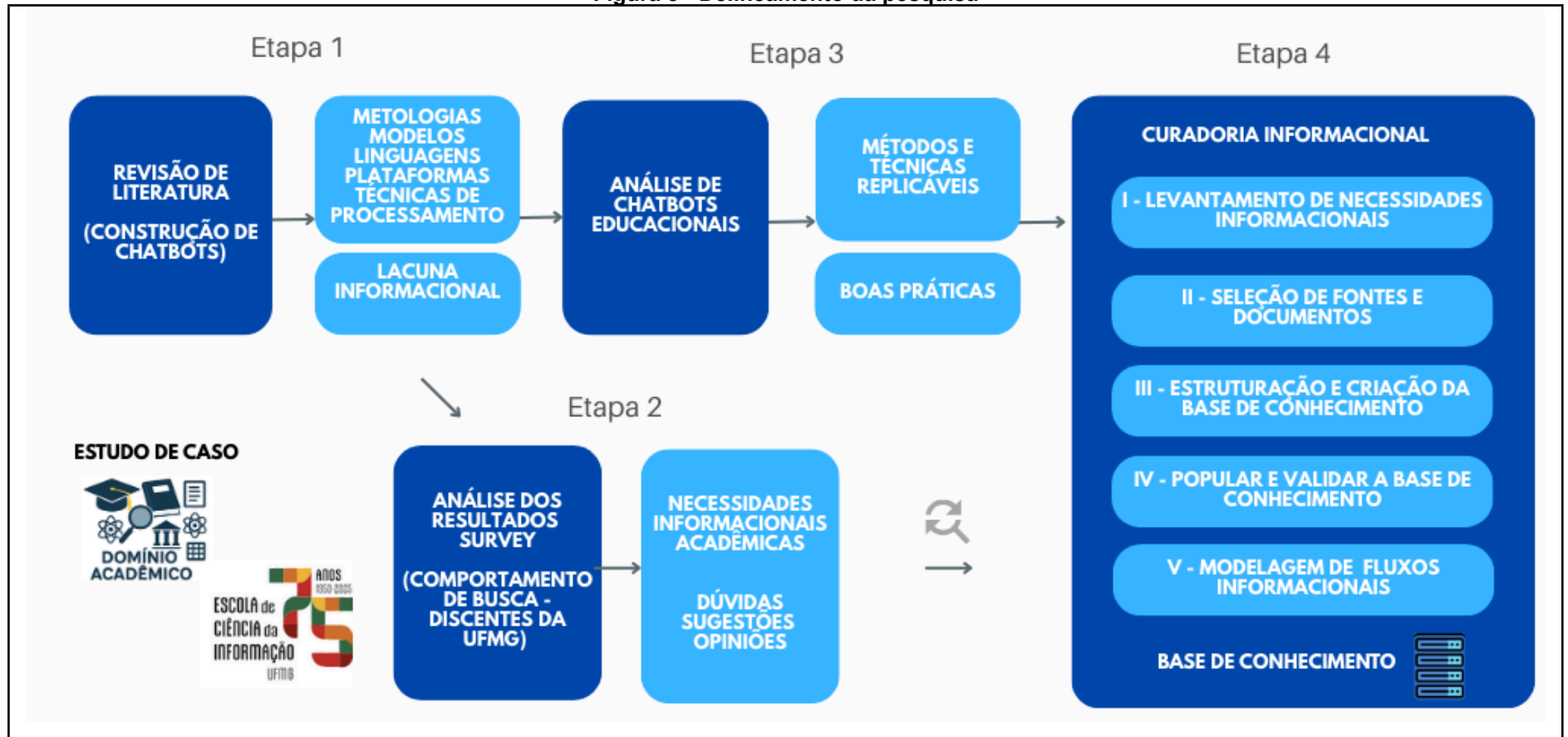
- I. **Levantamento de necessidades informacionais:** realizado com base nas respostas obtidas na pesquisa com estudantes de graduação da UFMG, permitindo identificar as principais dúvidas e dificuldades relacionadas ao acesso às informações acadêmicas (normas, resoluções e procedimentos);
- II. **Seleção de fontes e documentos:** identificação, coleta e análise das normas mais relevantes para o contexto estudado, com ênfase na Resolução Complementar nº 01/2018 (Normas Gerais da Graduação - NGG), nos documentos e conteúdos disponíveis no *site* da Pró-Reitoria de Graduação da UFMG e no *site* da Central de Graduação da ECI UFMG;
- III. **Estruturação e criação da base de conhecimento:** organização dos documentos selecionados, identificando conceitos e termos normativos, em tesouros e documentos normativos existentes, incluindo definições, a localização em artigos de resoluções e referências cruzadas, bem como categorias e sinônimos, com base nas NGG, nas resoluções e nas regras procedimentais da ECI UFMG, com o intuito de construir a estrutura da base de conhecimento a ser populada com as informações identificadas na curadoria. Para a organização da base de conhecimento, foi criada uma taxonomia que organizou todos os conceitos de maneira hierárquica e organizou, também, a base de conhecimento em formato de perguntas e respostas. Com isso, foi definida uma estrutura padrão que se repetiu para diversos conceitos e suas especificações.
- IV. **Popular e validar a base de conhecimento:** após a definição da estrutura, as informações devem ser incluídas seguindo a estrutura taxonômica definida, bem como as categorizações propostas. A validação foi realizada a partir da conferência das informações e, também, por meio de conferência por um especialista do domínio. Nesta pesquisa, a orientadora do trabalho, que possui experiência no domínio acadêmico, tendo atuado por mais de oito

anos nos sistemas acadêmicos da UFMG, exerceu esse papel. Contudo, essa validação interna foi visualizada como uma limitação da pesquisa.

- V. **Modelagem de fluxos informacionais:** a partir das informações coletadas e organizadas, foi sugerido um fluxo padrão, com a definição dos caminhos de navegação e orientação do usuário a partir de dúvidas recorrentes, considerando a organização dos termos normativos em estrutura hierárquica, do mais geral ao mais específico, bem como os vínculos entre termos, temas e procedimentos acadêmicos.

A Figura 5 apresenta todo o delineamento da pesquisa, evidenciando que as etapas não se configuram de maneira linear, mas se complementam, convergindo para alcançar o objetivo central desta pesquisa.

Figura 5 - Delineamento da pesquisa



Fonte: elaborado pela autora (2025).

5. RESULTADOS

A partir desta seção, são apresentados os resultados das quatro etapas delineadas para a pesquisa, iniciando pela revisão de literatura (Etapa 1), seguida de análise da pesquisa sobre o comportamento de busca dos estudantes de graduação (Etapa 2), análise de *chatbots* educacionais (Etapa 3) e curadoria informacional (Etapa 4).

5.1 REVISÃO DE LITERATURA

A pesquisa bibliográfica é a etapa inicial de todo o trabalho científico, tendo como objetivo reunir as informações e os dados que servirão de base para a construção da investigação proposta a partir de um determinado tema (Macedo, 1994).

Quanto aos diferentes tipos de pesquisa bibliográfica, a Revisão Sistemática de Literatura (RSL) configura-se como uma modalidade de revisão que busca identificar estudos relacionados a um determinado tema, valendo-se de métodos explícitos e organizados de busca, com o objetivo de analisar a qualidade, a validade e a aplicabilidade das pesquisas encontradas (De-La-Torre-Ugarte-Guanilo; Takahashi; Bertolozzi, 2011).

Na área de CI, existem diferentes abordagens para revisões sistemáticas, não havendo um protocolo único e padronizado, como em outras áreas. Assim, destaca-se que esta pesquisa foi classificada como uma revisão de literatura, pois, embora tenha seguido um protocolo e critérios de qualidade, foi conduzida pela orientanda, sob a supervisão da orientadora, sem o apoio de uma equipe específica de revisores, tampouco utilizando seleção às cegas e reuniões de consenso, etapas que são obrigatórias em uma RSL.

De acordo com Gil (2002), a revisão da literatura é um componente crítico dos trabalhos acadêmicos, fornecendo uma visão geral, um resumo e uma avaliação de pesquisas relevantes sobre um tópico, problema ou período de tempo específico. Corroborando a ideia, Kitchenham (2004) aponta que a revisão de literatura se conceitua como um método que busca identificar, avaliar e interpretar pesquisas disponíveis e relevantes para uma determinada questão de pesquisa, temática ou fenômeno de interesse.

Para realizar a revisão da literatura sobre *chatbots*, um protocolo foi adaptado e considerou os seguintes critérios: objetivo geral; questões a serem resolvidas; critérios de elegibilidade; critérios de inclusão e exclusão; termos de busca; expressões e *strings*; fontes de informações pesquisadas; procedimentos de seleção dos documentos recuperados; procedimentos de análise; e critério de exclusão após análise dos documentos.

A pesquisa foi delineada, primeiramente, pela definição das fontes de pesquisa, optando-se por bases de dados com ampla cobertura em diversas áreas do conhecimento, como Scopus, Web of Science, Scielo e Biblioteca Digital de Teses e Dissertações (BDTD). A Base de Dados em Ciência da Informação (BRAPCI), específica da CI no contexto brasileiro, foi excluída da pesquisa devido ao baixo retorno de estudos sobre *chatbots*. As bases de dados Web of Science, Scopus e Scielo foram selecionadas por se tratarem de bases que indexam periódicos de destaque na área da CI e em disciplinas relacionadas. O repositório da BDTD também foi utilizado, por compilar e divulgar textos completos de teses e dissertações defendidas em instituições de ensino e pesquisa brasileiras, abrangendo a produção acadêmica e científica nacional relacionada ao tema dos *chatbots*.

Posteriormente, foram estabelecidos os critérios de elegibilidade, inclusão e exclusão dos documentos. A definição da expressão e das *strings* de busca foi realizada considerando a sintaxe de cada base de dados, utilizando-se o termo "*chatbot*" e termos correlatos (assistente virtual e agente conversacional), além dos termos "metodologia" e "construção" e os seus respectivos sinônimos, além de "informação" para relacionar aspectos sobre dados e informação. A partir dos resultados, os documentos foram analisados e sistematizados por afiliação, ano, objetivo, métodos utilizados, modelos, linguagem e área de aplicação.

O critério de seleção das fontes de pesquisa considerou bases de dados multidisciplinares, inter-relacionadas com CI e TI. Para o critério de elegibilidade, foram selecionados os idiomas inglês, por abrangerem grande parte da produção científica na área, espanhol e português. Não foi estabelecida uma delimitação de datas nas buscas, a fim de não restringir a pesquisa e permitir a recuperação de questões históricas sobre o tema.

Quanto à tipologia documental da amostra, foram selecionados artigos de periódicos, publicações em eventos, teses e dissertações e capítulos de livros. Os critérios de inclusão foram definidos para que fosse realizada a seleção dos documentos de forma adequada ao estudo e consideraram: documentos que possuíam resumo; documentos nos idiomas inglês — pelo fato de a maioria dos artigos ser no idioma inglês e essa língua abranger grande parte da produção científica na área —, espanhol e português; e documentos relacionados ao desenvolvimento e à construção de *chatbots*. Foram definidos três critérios de exclusão: documentos que não foram escritos nos idiomas selecionados; documentos cujos títulos, resumos e palavras-chave não abordam o tema *chatbot*; e documentos não disponíveis na íntegra.

Com relação aos campos de busca, foram utilizadas *strings* para os campos título, resumo e palavras-chave, nas bases *Scopus*, *Web of Science*, BDTD, e título e palavras-chave na base *Scielo*. As expressões de busca definidas para a pesquisa envolviam os termos: *chatbot*, assistente virtual, agente conversacional, metodologia, informação e construção, por melhor representarem os objetivos da pesquisa. É importante esclarecer que as expressões foram utilizadas nos três idiomas selecionados e combinadas a partir de operadores booleanos. Inicialmente, as *strings* foram construídas de forma genérica, mas, depois, foram adaptadas conforme a sintaxe utilizada em cada base de dados, já que cada base possui parâmetros de recuperação próprios.

O critério definido em procedimentos de seleção para os documentos recuperados consistiu na realização da leitura do título e dos resumos dos documentos recuperados, com o intuito de verificar a pertinência do conteúdo ao objetivo geral do protocolo de revisão. Nos procedimentos de análise, procurou-se identificar conceitos, definições e aplicações dos *chatbots*, além de elementos como construção e metodologia, a fim de recuperar documentos estritamente pertinentes ao estudo. Os critérios de exclusão, após a análise dos documentos, tiveram como intuito descartar trabalhos que não possuíssem abordagem conceitual, teórica ou metodológica em seu conteúdo.

O procedimento de análise dos resultados envolveu o *software* Rayyan¹ para a triagem de resumos, localização e exclusão de estudos duplicados.

¹Disponível em: <https://www.rayyan.ai/>. Acesso em 07 out. 2025.

Posteriormente, seguiu-se com a leitura analítica dos documentos pela pesquisadora, considerando o título e o resumo, a fim de verificar a relevância do documento para a pesquisa.

Considerando os aspectos e as categorias com o intuito de entender os fenômenos investigados, a técnica de análise de conteúdo foi utilizada no tratamento de resultados, inferências e interpretações. Conforme definido por Bardin (2016), a análise de conteúdo consiste em um conjunto de técnicas de pesquisa que visam interpretar e analisar o conteúdo e significado de tipos de documentos. A análise de conteúdo pode utilizar tanto indicadores quantitativos (frequência de palavras) quanto indicadores qualitativos (temas recorrentes e relações entre conceitos). Esses indicadores são utilizados para descrever o conteúdo das mensagens e para fazer inferências sobre o contexto em que foram produzidas (Bardin, 2016).

Assim, durante a leitura completa dos documentos, foram identificados e registrados os temas recorrentes para cada documento em uma planilha Excel. Ao final, as temáticas identificadas foram agrupadas conforme a relação entre os conceitos e foi realizada uma classificação facetada dos documentos. A partir da ocorrência desses indicadores nos resultados, consideraram-se as seguintes categorias: metodologias, modelos, linguagens, plataformas e técnicas de PLN.

O protocolo de revisão de literatura foi executado entre os meses de julho e agosto de 2024, e a análise dos dados utilizou uma metodologia mista, a partir de métodos qualitativos e quantitativos. O Quadro 5 apresenta o detalhamento do protocolo utilizado.

Quadro 5 - Protocolo da revisão de literatura

Protocolo	
Critérios	Descrição
Objetivo geral	Identificar artigos científicos, teses, dissertações e trabalhos publicados em eventos relacionados ao desenvolvimento/construção de <i>chatbots</i> .
Questões a serem resolvidas	Quais são as metodologias utilizadas e empregadas no desenvolvimento de <i>chatbots</i> ?
	A quais áreas do conhecimento estão relacionados?
	Quais são os principais domínios de aplicação dos <i>chatbots</i> na literatura atual?
	Quais são as principais tecnologias e plataformas utilizadas no desenvolvimento de <i>chatbots</i> ?
	Quais aspectos sobre organização, representação e recuperação da informação foram identificados?

Fontes de informação pesquisadas	Bases de dados: Scopus, Web Of Science, Scielo, BDTD
Critérios de elegibilidade	Idioma: inglês, espanhol, português;
	Tipologia documental: artigos de periódicos, publicações em eventos, teses e dissertações e capítulos de livros;
Critérios de inclusão e de exclusão	Inclusão: documentos que possuem resumo;
	Documentos nos idiomas inglês, português e espanhol;
	Trabalhos relacionados ao desenvolvimento e à construção de <i>chatbots</i> ;
	Exclusão de documentos duplicados;
	Exclusão de documentos que não foram escritos em português, inglês ou espanhol;
	Documentos cujos títulos, resumos e palavras-chave não abordam o tema sobre <i>chatbots</i> ;
Documentos que não estejam disponíveis na íntegra;	
Campos de busca	título, resumo, palavras-chave (Scopus, WOS, BDTD); título, palavras-chave (Scielo);
Expressões de busca	chatbot e sinônimos E metodologia, construção e sinônimos E informação
	português, inglês e espanhol;
<i>String</i> geral	("chatbot" OR "chatbots" OR "assistente virtual" OR "agente conversacional" OR "assistente de IA" OR "assistente de inteligência artificial" OR "virtual assistant" OR "conversational agents" OR "AI assistant" OR "artificial intelligence assistant" OR "asistente virtual" OR "agente conversacional" OR "asistente de IA" OR "asistente de inteligencia artificial") AND ("construção" OR "metodologia" OR "desenvolvimento" OR "construction" OR "methodology" OR "development" OR "construcción" OR "metodologia" OR "desenvolvimento") AND ("informação" OR "information" OR "información")
Scopus	TITLE-ABS-KEY (("chatbot" OR "chatbots" OR "assistente virtual" OR "agente conversacional" OR "assistente de IA" OR "assistente de inteligência artificial" OR "virtual assistant" OR "conversational agents" OR "AI assistant" OR "artificial intelligence assistant" OR "asistente virtual" OR "agente conversacional" OR "asistente de IA" OR "asistente de inteligencia artificial") AND ("construção" OR "metodologia" OR "desenvolvimento" OR "construction" OR "methodology" OR "development" OR "construcción" OR "metodologia" OR "desenvolvimento") AND ("informação" OR "information" OR "información"))
WOS	TS= (("chatbot" OR "chatbots" OR "assistente virtual" OR "agente conversacional" OR "assistente de IA" OR "assistente de inteligência artificial" OR "virtual assistant" OR "conversational agents" OR "AI assistant" OR "artificial intelligence assistant" OR "asistente virtual" OR "agente conversacional" OR "asistente de IA" OR "asistente de inteligencia artificial") AND ("construção" OR "metodologia" OR "desenvolvimento" OR "construction" OR "methodology" OR "development" OR "construcción" OR "metodologia" OR "desenvolvimento") AND ("informação" OR "information" OR "información"))
Scielo	(ab:(("chatbot" OR "chatbots" OR "assistente virtual" OR "agente conversacional" OR "assistente de IA" OR "assistente de inteligência artificial" OR "virtual assistant" OR "conversational agents" OR "AI assistant" OR "artificial intelligence assistant" OR "asistente virtual" OR "agente conversacional" OR "asistente de IA" OR "asistente de inteligencia artificial") AND ("construção" OR "metodologia" OR "desenvolvimento" OR "construction" OR "methodology" OR "development" OR "construcción" OR

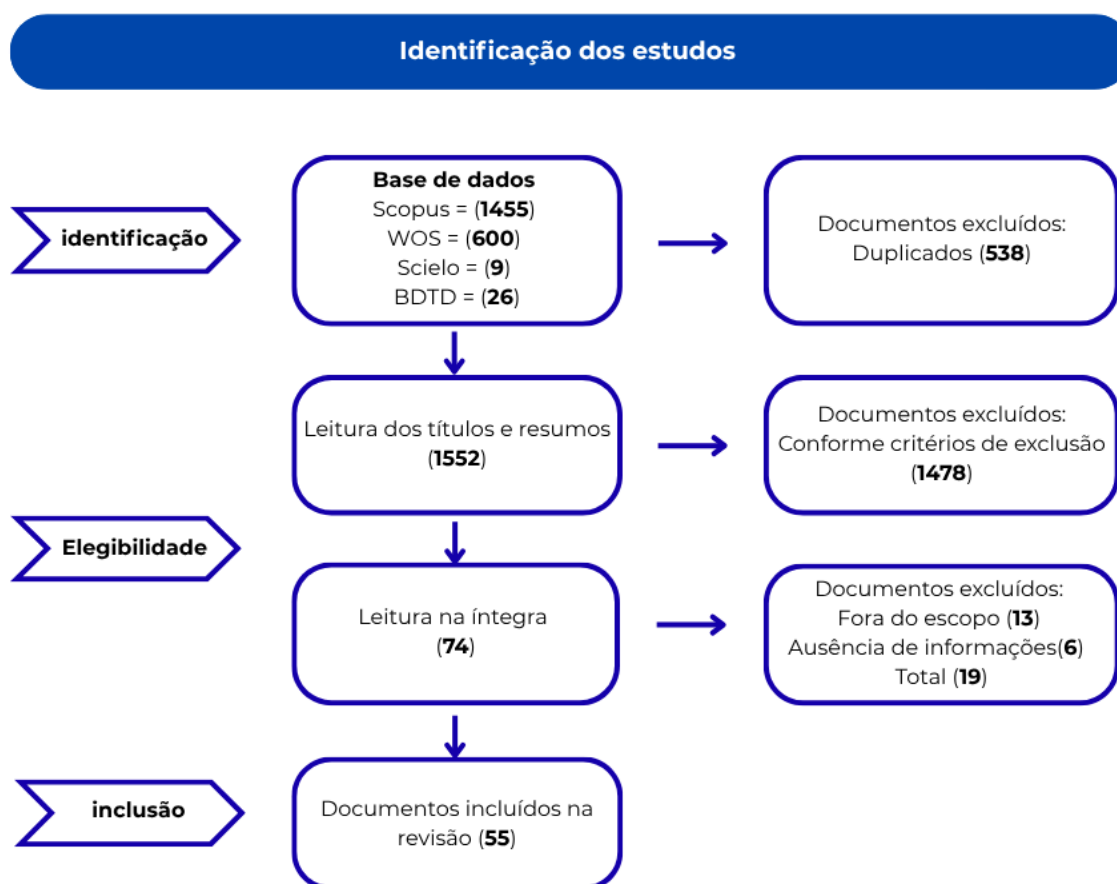
	"metodologia" OR "desenvolvimento") AND ("informação" OR "information" OR "información"))) OR (ti:(("chatbot" OR "chatbots" OR "assistente virtual" OR "agente conversacional" OR "assistente de IA" OR "assistente de inteligência artificial" OR "virtual assistant" OR "conversational agents" OR "AI assistant" OR "artificial intelligence assistant" OR "asistente virtual" OR "agente conversacional" OR "asistente de IA" OR "asistente de inteligencia artificial") AND ("construção" OR "metodologia" OR "desenvolvimento" OR "construction" OR "methodology" OR "development" OR "construcción" OR "metodologia" OR "desenvolvimento") AND ("informação" OR "information" OR "información")))
BDTD	(Título:chatbot OR chatbots assistente virtual OR agente conversacional OR assistente de IA OR assistente de inteligência artificial OR virtual assistant OR conversational agents OR AI assistant OR artificial intelligence assistant OR asistente virtual OR agente conversacional OR asistente de IA OR asistente de inteligencia artificial AND construção OR metodologia OR desenvolvimento OR construction OR methodology OR development OR construcción OR metodologia OR desenvolvimento AND informação OR information OR información E Assunto:chatbot OR chatbots assistente virtual OR agente conversacional OR assistente de IA OR assistente de inteligência artificial OR virtual assistant OR conversational agents OR AI assistant OR artificial intelligence assistant OR asistente virtual OR agente conversacional OR asistente de IA OR asistente de inteligencia artificial AND construção OR metodologia OR desenvolvimento OR construction OR methodology OR development OR construcción OR metodologia OR desenvolvimento AND informação OR information OR información E Resumo:chatbot OR chatbots assistente virtual OR agente conversacional OR assistente de IA OR assistente de inteligência artificial OR virtual assistant OR conversational agents OR AI assistant OR artificial intelligence assistant OR asistente virtual OR agente conversacional OR asistente de IA OR asistente de inteligencia artificial AND construção OR metodologia OR desenvolvimento OR construction OR methodology OR development OR construcción OR metodologia OR desenvolvimento AND informação OR information OR información)
Procedimentos de seleção dos documentos recuperados	Inicialmente, a leitura dos títulos dos documentos recuperados, com o intuito de verificar a pertinência do conteúdo ao objetivo geral da pesquisa e, em seguida, o mesmo procedimento com os resumos.
Procedimentos de análise	Leitura completa dos documentos, a fim de identificar elementos de representação e recuperação de informação em <i>chatbots</i> .
Critério de exclusão após análise dos documentos	Trabalhos que não possuem abordagem conceitual, teórica ou metodológica sobre <i>chatbots</i> .
Tratamento	<i>Softwares</i> Rayyan e Excel.

Fonte: adaptado de Nascimento Silva (2023b).

A busca nas quatro bases de dados selecionadas resultou na recuperação de 2.090 documentos. No *software* Rayyan, foram removidos 538 documentos duplicados, restando 1.552 trabalhos para a triagem inicial. Nesta fase, foram analisados títulos e resumos, e uma quantidade expressiva de 1.478 documentos foi descartada por não atender aos critérios de inclusão/exclusão definidos no protocolo. As razões para a exclusão incluíram: foco em análises e avaliação de

modelos de linguagem generativa (como ChatGPT, Bard e Copilot); idioma diferente de inglês, português ou espanhol; e documentos sem acesso ao texto completo. Portanto, 74 documentos foram selecionados para a leitura na íntegra, visando verificar sua pertinência ao objetivo geral da revisão. Os procedimentos de análise buscaram identificar conceitos, definições, aspectos metodológicos e aplicações de *chatbots*. Após a leitura completa, 19 documentos foram excluídos por apresentarem lacunas e ausência de dados considerados relevantes para a revisão, por não descreverem os detalhes e as ferramentas de implementação e por trazerem apenas informações sobre a finalidade e a aplicação da ferramenta *chatbot*, totalizando 55 documentos para análise e discussão, conforme apresentado no fluxograma de seleção de documentos (Figura 6) e detalhado no Quadro 6.

Figura 6 - Fluxograma da seleção e elegibilidade dos documentos da revisão



Fonte: elaborado pelos autores (2025).

Quadro 6 - Documentos selecionados

ID	Título	Autor / Ano	Elementos	Categoria
D1	Developing a Chatbot to Support Individuals With Neurodevelopmental Disorders: Tutorial	Singla <i>et al.</i> (2024)	Python Django Software Foundation	Linguagem Plataforma
D2	Development Of A Natural Language Chatbot Interface For Website Users	Lynnyk, Krestyanpol e Rozvod (2024)	Javascript HEROKU	Linguagem Plataforma
D3	Development and Research of a Chatbot Using the Linguistic Core of Amazon Lex V2	Hnatushenko, Ostrovska e Nosov (2024)	Amazon Lex	Plataforma
D4	Artificial Intelligence Chatbot Platform: AI Chatbot Platform for Educational Recommendations in Higher Education	Kingchang, Chatwattana e Wannapiroon (2024)	SDLC LINE	Métodos Plataforma
D5	A comprehensive solution to retrieval-based chatbot construction	Moore <i>et al.</i> (2023)	BERT aprendizagem contrastiva Python	Modelo Linguagem
D6	Movie recommender chatbot based on Dialogflow	Baizal <i>et al.</i> (2023)	N-grama Dialog Flow TF-IDF	Modelo Plataforma PLN
D7	Advancements in Conversational AI: Building Mental Health Chatbot with BERT Model	Mishra <i>et al.</i> (2023)	BERT Python	Modelo Linguagem
D8	Enhancing Pipeline-Based Conversational Agents with Large Language Models	Foosherian <i>et al.</i> (2023)	LLM	Modelo
D9	The Development Of Chatbot Provided Registration Information Services For Students In Distance Learning	Chumkaew (2023)	Dialog Flow	Plataforma
D10	Intelligent agent for incident management	Iparraguirre-Villanueva <i>et al.</i> (2023)	Scrum PHP Dialog Flow	Métodos Linguagem Plataforma
D11	AI Based Chatbot for Healthcare using Machine Learning	Tanwar <i>et al.</i> (2023)	Dialog Flow TF-IDF	Plataforma PLN
D12	Chatbot as a Virtual Assistant to Retrieve Information from Datasheets Using Memory Controllers Domain Knowledge	Elsayed <i>et al.</i> (2023)	BERT RASA	Modelo Plataforma
D13	Using Artificial Intelligence Methods to Create a Chatbot for University Questions and Answers	Ramalakshmi <i>et al.</i> (2023)	RASA Python RASA	Modelo Linguagem Plataforma
D14	Introducing a Chatbot to the Web Portal of a Higher Education Institution to Enhance Student Interaction	Oliveira e Matos (2023)	LLM Dialog Flow	Modelo Plataforma

D15	Design and Development of AI-Powered Healthcare WhatsApp Chatbot	Prakasam <i>et al.</i> (2023)	TF-IDF	PLN
D16	Development of a Class Materials Search System using LINE Chatbot	Thaiprasert <i>et al.</i> (2023)	Dialog Flow LINE TF-IDF	Plataforma PLN
D17	Designing an Educational Chatbot with Joint Intent Classification and Slot Filling	Wang, Liu e Song (2022)	BERT	Modelo
D18	The Automation of Client Servicing in University and College Admission Office	Stepanov <i>et al.</i> (2023)	Javascript	Linguagem
D19	Automated information retrieval and services of graduate school using chatbot system	Rukhiran e Netinant (2022)	RAD Dialog Flow	Métodos Plataforma
D20	Adapting intelligent information services in libraries: a case of smart AI chatbots	Panda e Chakravarty (2022)	ENGATI	Plataforma
D21	Language-Agnostic Knowledge Graphs for Smarter Multilingual Chatbots	Vasilevich <i>et al.</i> (2022)	RASA LLM RASA	Modelo Plataforma
D22	Applying cognitive computing to legal services	Whittle e Hall (2022)	IBM Watson	Plataforma
D23	WASABI Contextual BOT	Agrawal <i>et al.</i> (2022)	RASA RASA	Modelo Plataforma
D24	Knowledge-Based Chatbot For Humanoid Robot In Restaurant For Question And Answering System	Yossy e Budiharto (2022)	Dialog Flow	Plataforma
D25	A support system for formal college students: A case study of a community-based app augmented with a chatbot	Alqaidi, Alharbi e Almatrafi (2021)	Scrum Dialog Flow	Métodos Plataforma
D26	Development of a chatbot using machine learning algorithms to automate educational processes	Alekseev, Shagalova e Sokolova (2021)	Python	Linguagem
D27	Designing an Educational Chatbot: A Case Study of CikguAIBot	Nasharuddin <i>et al.</i> (2021)	RAD	Métodos
D28	Development of Information Technology Telecom Chatbot: An Artificial Intelligence and Machine Learning Approach	Gowda <i>et al.</i> (2021)	Javascript Discord	Linguagem Plataforma
D29	Proposal for a Chatbot Prototype in the Plant Health Department of Brazilian Ministry of Agriculture	Feitosa <i>et al.</i> (2020)	Facebook	Plataforma
D30	Development of Intelligent Telegram Chatbot Using Natural Language	Gunawan <i>et al.</i> (2021)	Python Dialog Flow	Linguagem Plataforma

	Processing			
D31	Arisa nest - A cloud-based platform for development of virtual assistant	Zambiasi e Rabelo (2020)	Javascript PHP Arisa Nest	Linguagem Plataforma
D32	A Chatbot to Support Basic Students Questions	Santana <i>et al.</i> (2021)	TF-IDF	PLN
D33	The Combination of Natural Language Processing and Entity Extraction for Academic Chatbot	Tommy, Kirana e Riska (2020)	OOP	Métodos
D34	Developing UNYSA Chatbot as Information Services about Yogyakarta State University	Muhyidin e Setiawan (2021)	Scrum SDLC Dialog Flowt	Métodos Linguagem Plataforma
D35	From Process Models to Chatbots	López <i>et al.</i> (2019)	BPMN	Métodos
D36	AI and Web-Based Human-Like Interactive University Chatbot (UNIBOT)	Patel <i>et al.</i> (2019)	PHP CSS	Linguagem
D37	Dinus Intelligent Assistance (DINA) chatbot for University Admission Services	Agus Santoso <i>et al.</i> (2018)	Dialog Flow	Plataforma
D38	Cecílio: um chatbot para automação do atendimento aos usuários nas Instituições Federais de Ensino Superior	Feitoza (2021)	BPMN NEORON	Métodos Plataforma
D39	Technical Aspects of Developing Chatbots for Medical Applications: Scoping Review	Safi <i>et al.</i> (2020)	Dialog Flow	Plataforma
D40	Uncoding library chatbots: deploying a new virtual reference tool at the San Jose State University library	Rodriguez e Mune (2022)	Kommunicate	Plataforma
D41	Development and evaluation of a chatbot for the Regional Museum of Sao Joao del-Rei	Alves <i>et al.</i> (2018)	IBM Watson	Plataforma
D42	Deep Learning Techniques for Implementation of Chatbots	Karri e Kumar (2020)	Seq2seq Bag of Words	PLN
D43	Initial study of a chatbot for virtual modality students from the Inter-American School of librarianship	Múnera Torres, Álvarez e Osorio (2022)	Design Thinking	Métodos
D44	Developing a Chatbot for College Student Programme Advisement	Chan <i>et al.</i> (2018)	build-and-test PHP IBM Watson	Métodos Linguagem Plataforma
D45	Chatbots in the library: is it time?	Allison (2012)	PHP	Linguagem
D46	Multi-platform Chatbot Modeling and Deployment with the Jarvis Framework	Daniel <i>et al.</i> (2019)	MDE Dialog Flow	Modelo Plataforma

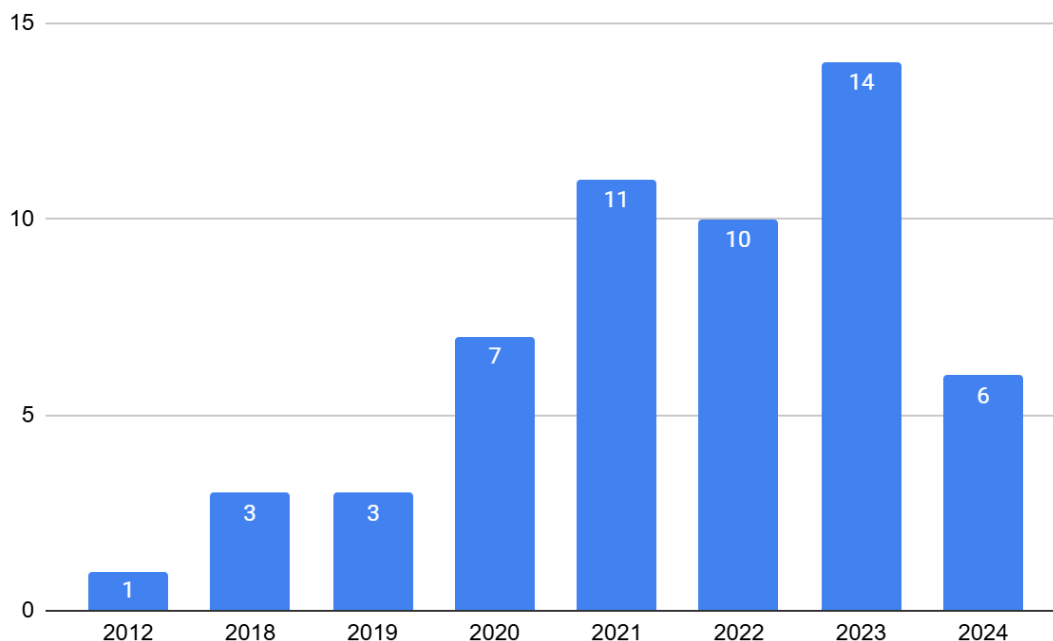
D47	Advanced NLP Models for Technical University Information Chatbots: Development and Comparative Analysis	Attigeri, Agrawal e Kolekar (2024)	TF-IDF	PLN
D48	Academic Libraries Can Develop AI Chatbots for Virtual Reference Services with Minimal Technical Knowledge and Limited Resources	Chase (2024)	Dialog Flow Kommunicate	Plataforma
D49	A Model to Develop Chatbots for Assisting the Teaching and Learning Process	Mendoza <i>et al.</i> (2022)	Design Thinking Dialog Flow	Métodos Plataforma
D50	Virtual Assistant Prototype for Property Sales	Sanjaya e Santoso (2021)	Dialog Flow	Plataforma
D51	Development and evaluation of the GISSA Mother-Baby ChatBot application in promoting child health	Barreto <i>et al.</i> (2021)	Python	Linguagem
D52	A Chatbot as a Support System for Educational Institutions	Villanueva e Aguilar-Alonso (2021)	Dialog Flow	Plataforma
D53	Design of an Educational Chatbot Using Artificial Intelligence in Radiotherapy	Chow, Sanders e Li (2023)	IBM Watson	Plataforma
D54	An Intelligent Chatbot System Based on Entity Extraction Using RASA NLU and Neural Network	Jiao (2020)	RASA Python Cython RASA	Modelo Linguagem Plataforma
D55	Intelligent Chatbot-LDA Recommender System	Touimi <i>et al.</i> (2020)	LDA MOOC EAD	Modelo Plataforma

Fonte: elaborado pela autora (2025).

5.2 Resultados

Os 55 estudos da amostra que apresentaram elementos e ferramentas utilizadas para a construção dos *chatbots* consistiram artigos de periódicos, trabalhos apresentados em congressos e uma dissertação de mestrado. A data de publicação dos estudos recuperados sugere que a temática dos *chatbots* é relativamente recente, com uma tendência crescente de publicações nos últimos anos, especialmente a partir de 2020. A maior parte dos trabalhos da amostra foi publicada entre os anos de 2021 e 2023, conforme apresenta o Gráfico 1, sobre a distribuição da frequência das publicações ao longo dos anos.

Gráfico 1 - Distribuição das publicações por ano



Fonte: elaborado pelos autores (2025).

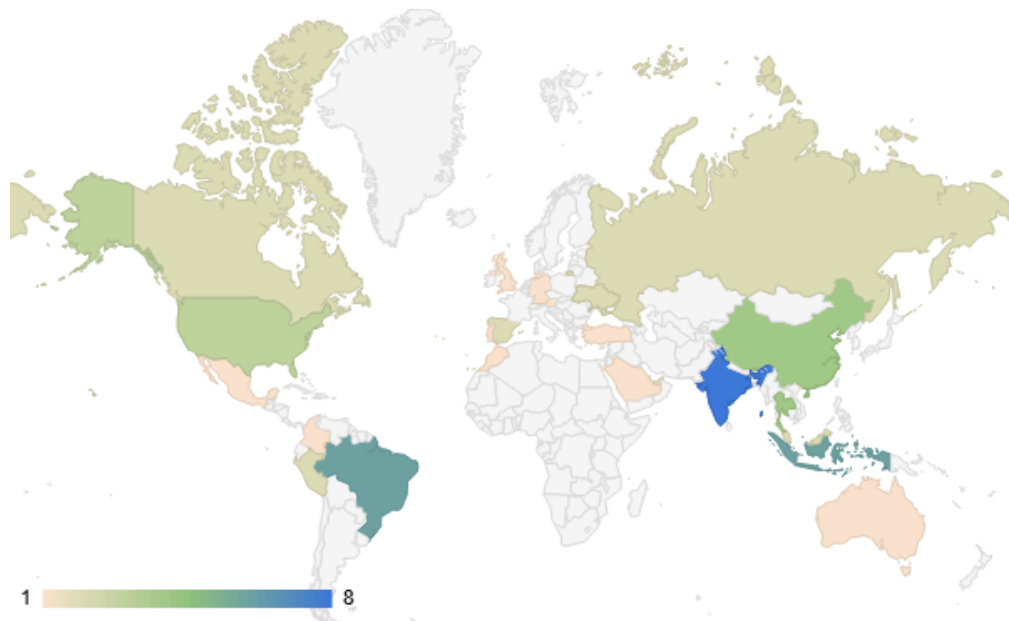
Em relação à área de aplicação dos *chatbots*, a revisão buscou responder à seguinte pergunta: Quais são os principais domínios de aplicação dos *chatbots* na literatura? Foi observado que diversas áreas têm se beneficiado da capacidade dos *chatbots* de proporcionar atendimento rápido, reduzir custos operacionais e melhorar a experiência do usuário por meio da automatização e personalização das interações. No estudo, observou-se que a maior parte dos *chatbots* descritos, 26 (47,27%), está concentrada na área da educação, sendo empregada para auxiliar alunos com dúvidas, fornecer *feedback* instantâneo, disponibilizar materiais de estudo e apoiar o processo de ensino a distância.

Em segundo lugar, foram identificados, em maior número, domínios aplicados à saúde, desenvolvidos para atividades como triagem de sintomas, agendamento de consultas, lembretes de medicação, cessação de tabagismo e fornecimento de informações sobre bem-estar e saúde mental, com sete (12,72%) dos *chatbots* apresentados nos estudos. A CI aparece na terceira posição, com quatro (7,27%) estudos, seguida da (CC), que apresentou três (5,45%) estudos. As demais áreas de aplicação contaram com um estudo cada (1,81%). No caso da CI, os assistentes virtuais aplicados em bibliotecas e serviços de informação abrangem desde a curadoria de conteúdo, o suporte à navegação em catálogos e a orientação em

pesquisas até a alfabetização informacional, orientando os usuários sobre como realizar pesquisas, avaliar a credibilidade das fontes e compreender práticas de uso ético da informação.

O local de filiação dos autores é uma informação relevante quanto à análise de contribuições acadêmicas, pois pode identificar a distribuição geográfica dos estudos, além de indicar as instituições que lideram as pesquisas em uma determinada área. Na Figura 7, o mapa de calor ilustra a distribuição geográfica das publicações, evidenciando a Índia, com oito estudos, como o país com a maior concentração de pesquisas, seguida por Brasil e Indonésia (seis). A intensidade da cor azul no mapa reflete a concentração de estudos, indicando uma atividade acadêmica mais intensa na Índia e um foco significativo em pesquisa na Indonésia e no Brasil. Países como Tailândia, China e EUA, representados em verde, demonstram uma produção moderada de publicações, enquanto países em bege apresentam menor número de estudos, sugerindo uma contribuição limitada ao tema, possivelmente devido à escassez de recursos ou menor interesse na área. Por fim, a cor branca indica países que não contribuíram com publicações na amostra analisada.

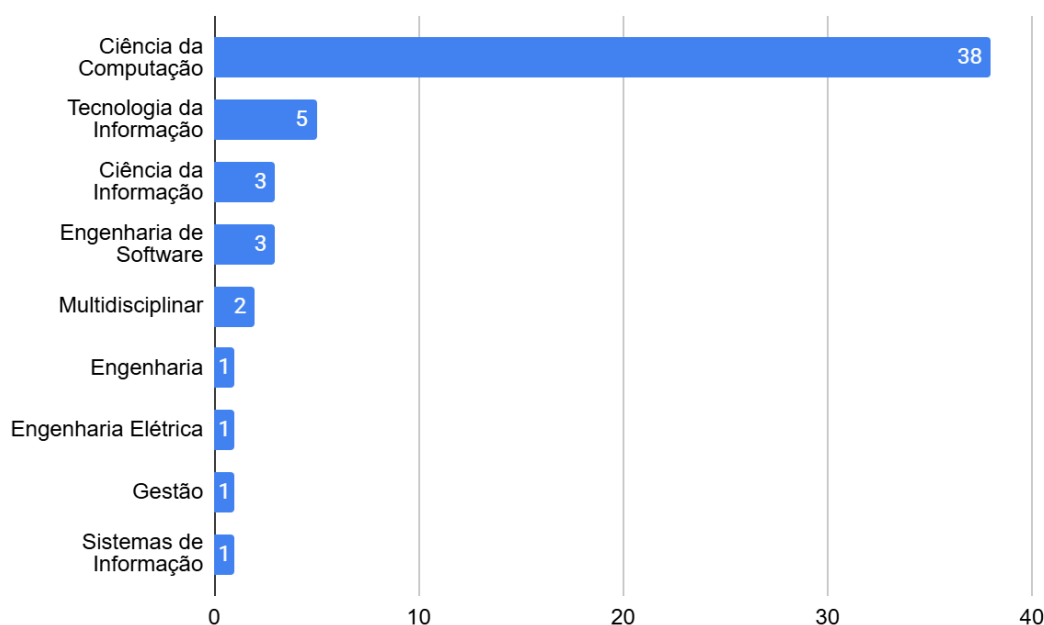
Figura 7 - Mapa de calor (afiliação dos autores)



Fonte: elaborado pela autora (2025).

A área de conhecimento dos estudos (Gráfico 2) foi utilizada a fim de se buscar a compreensão das diferentes abordagens adotadas na pesquisa sobre *chatbots* e permitir reconhecer, de forma mais clara, em que disciplina essa tecnologia tem sido mais explorada. A Ciência da Computação foi identificada como a área que mais desenvolveu e publicou estudos sobre *chatbots*, demonstrando seu papel central no avanço e na inovação das ferramentas tecnológicas. A (CC) tem se destacado pela criação de novas arquiteturas de PLN, técnicas de aprendizado de máquina e abordagens que permitem a melhoria contínua dos *chatbots*, tanto em termos de eficácia quanto de aplicabilidade em diferentes contextos (Mitchell, 2019).

Gráfico 2 - Área de conhecimento (categorização do autor)



Fonte: elaborado pela autora (2025).

A seguir, nas seções 5.2.1 a 5.2.5, serão discutidas as seguintes categorias identificadas: Metodologias, Modelos, Linguagens, Plataformas e Técnicas de PLN.

5.2.1 Metodologias utilizadas

Com o intuito de investigar as metodologias empregadas na implementação de *chatbots*, identificou-se, a partir da análise dos documentos recuperados, o uso de algumas metodologias e processos, destacando-se não só métodos ágeis, a exemplo do Scrum e do Design Thinking, como também os métodos tradicionais de

desenvolvimento de *software*, como o Software Development Life Cycle (SDLC), e ferramentas de modelagem, como a Business Process Model and Notation (BPMN).

De acordo com Moore *et al.* (2023), o desenvolvimento de *chatbots* envolve uma gama de metodologias e abordagens que buscam facilitar sua criação e implementação. Oliveira e Matos (2023) afirmam que os métodos empregados na construção de *chatbots* podem variar de acordo com os objetivos e o público-alvo da ferramenta e incluir desde abordagens ágeis (que favorecem iterações rápidas e *feedback* contínuo) até métodos mais tradicionais (que seguem etapas bem definidas de planejamento, desenvolvimento e avaliação).

5.2.2 Abordagens ágeis

As abordagens ágeis surgiram como uma resposta à necessidade de uma maior flexibilidade e adaptabilidade no desenvolvimento de *software*, em contraposição aos métodos tradicionais, que se mostravam lentos e pouco eficientes diante das constantes mudanças no cenário (Schwaber; Sutherland, 2020; Jurafsky; Martin, 2023).

Em 2001, um grupo de 17 desenvolvedores de *software* se reuniu, em Utah, para explorar novas abordagens que pudessem substituir os métodos tradicionais de desenvolvimento. O resultado desse encontro foi o Manifesto para Desenvolvimento Ágil de Software, um documento que delinea os valores e princípios fundamentais das metodologias ágeis (Beck *et al.*, 2001).

Essas metodologias têm sido aplicadas em diversos campos, oferecendo uma abordagem flexível e iterativa para lidar com projetos complexos que envolvem a colaboração de desenvolvedores, pesquisadores, analistas, cientistas e outros profissionais (Schwaber; Sutherland, 2020). A capacidade de adaptação e o foco na entrega contínua de valor tornam os métodos ágeis uma escolha frequente em contextos que exigem respostas rápidas e soluções alinhadas às necessidades dos usuários. O estudo de Nascimento Silva (2023a) reitera essa contribuição na aplicação dos métodos ágeis na CI. Neste cenário, o Scrum e o Design Thinking se destacaram como metodologias ágeis, sendo abordadas em cinco estudos.

O método Scrum se mostrou eficaz para o desenvolvimento de *chatbots*, conforme destacado por Iparraguirre-Villanueva *et al.* (2023), que demonstraram a aplicação bem-sucedida no desenvolvimento de um *chatbot* para otimizar o suporte

de TI (Tecnologia da Informação) e a gestão de incidentes. De maneira similar, Alqaidi, Alharbi e Almatrafi (2021) descreveram o Assistente do Estudante, um *chatbot* com funcionalidades que incluem respostas automatizadas, fóruns para achados e perdidos, anúncios e recrutamento de voluntários, desenvolvido sob a metodologia Scrum, e, também, o *chatbot* UNYSA, apresentado por Muhyidin e Setiawan (2021), que visa melhorar o desempenho acadêmico e diminuir a carga de trabalho dos orientadores, igualmente desenvolvido com base no Scrum.

O Scrum adota uma abordagem iterativa e incremental que ajuda no gerenciamento de riscos, na melhora da previsibilidade do projeto, além de destacar a importância da transparência, da inspeção e da adaptação como princípios fundamentais (Schwaber; Sutherland, 2020).

A metodologia Design Thinking, explorada nos estudos de Chan *et al.* (2018) e Daniel *et al.* (2019), destaca-se como uma abordagem complementar que coloca o usuário no centro do processo de *design*. Essa metodologia enfatiza a empatia e a prototipagem rápida, facilitando a criação de *chatbots* mais eficazes e centrados na experiência do usuário. A relevância do Design Thinking no desenvolvimento de *chatbots* se alinha aos princípios da CI, conforme apontado por Guimarães e Rocha (2021), em sua pesquisa sobre práticas informacionais e Design Thinking, que aborda a importância de compreender as necessidades e os contextos dos usuários para criar sistemas de informação eficazes.

O Design Thinking é uma abordagem que envolve a definição do problema, ideação, prototipagem e testes, com foco na empatia e na compreensão profunda do usuário, a fim de criar experiências mais intuitivas (Combelles; Ebert; Lucena, 2020).

5.2.3 Abordagens tradicionais

As abordagens tradicionais, também conhecidas como modelos "clássicos", ou "prescritivos", são abordagens mais estruturadas e sequenciais para o desenvolvimento de *software*. Modelos Cascata, ou Waterfall, por exemplo, são considerados alguns dos mais utilizados e são metodologias que seguem uma sequência linear de fases: coleta de requisitos, *design*, implementação, testes e manutenção (Sommerville, 2016). Os métodos tradicionais de desenvolvimento de *software* são caracterizados por um planejamento detalhado e uma documentação extensiva antes do início do desenvolvimento.

Os autores Kingchang, Chatwattana e Wannapiroon (2024) e Feitosa *et al.* (2020) optaram, em seus estudos, pelo uso de métodos tradicionais no desenvolvimento de *software* para implementar os *chatbots*, demonstrando que essas abordagens mais estruturadas também são eficazes e aprimoram o seu desenvolvimento, especialmente em projetos que exigem maior controle e previsibilidade. O Software Development Life Cycle (SDLC) possui fases bem definidas de planejamento, análise, *design*, implementação, testes e manutenção e oferece um *framework* estruturado, que garante maior controle e previsibilidade ao projeto. A integração com sistemas existentes também é um fator importante a ser considerado, visto que o SDLC ajuda a garantir que o *chatbot* se integre perfeitamente aos sistemas da organização, evitando conflitos e falhas.

Setores como saúde e finanças, nos quais a confiabilidade é essencial e erros podem ter consequências graves, se beneficiam da abordagem estruturada do SDLC para minimizar riscos e garantir a qualidade do *chatbot* (Feitosa *et al.*, 2020). Os estudos de López *et al.* (2019) e Feitosa (2021) descreveram a metodologia BPMN no desenvolvimento de *chatbots*. López *et al.* (2019) utilizaram a BPMN para orientar processos, enquanto Feitosa (2021) aplicou a mesma metodologia no desenvolvimento do *chatbot* Cecílio para o atendimento a usuários em instituições federais de ensino superior.

A BPMN é uma notação gráfica que padroniza a modelagem de processos de negócios, oferecendo fluxogramas claros que facilitam a compreensão dos processos por qualquer pessoa, independentemente de seus conhecimentos técnicos, oferecendo uma linguagem visual padronizada que possibilita a modelagem, análise e otimização de processos informacionais, de forma semelhante à CI, abrangendo o ciclo completo da informação: geração, coleta, organização, armazenamento, recuperação, disseminação e uso (López *et al.*, 2019).

Com menor frequência nos estudos analisados e sendo identificadas em somente um trabalho cada, estão: Rapid Application Development (RAD), ou Desenvolvimento Rápido de Aplicações, descrito nos estudos de Rukhiran e Netinant (2022), Recuperação automatizada de informações e serviços da pós-graduação usando o sistema *chatbot*, e de Nasharuddin *et al.* (2021), Designing an Educational Chatbot: A Case Study of CikguAIBot. O RAD se caracteriza como uma metodologia de desenvolvimento de *software* que prioriza a prototipagem

rápida e o *feedback* regular do usuário para entregar aplicações de forma mais rápida e eficiente.

A Programação Orientada a Objetos (POO) foi descrita em *The Combination of Natural Language Processing and Entity Extraction for Academic Chatbot*, trabalho publicado pelos autores Tommy, Kirana e Riska (2020). Segundo Silveira (2015), a POO é semelhante a uma metodologia, porém é considerada um paradigma de programação que organiza o *software* em torno de objetos, que representam abstrações do mundo real. Esses objetos combinam dados (atributos) e comportamentos (métodos) em uma única estrutura. O seu principal objetivo é criar sistemas modulares, reutilizáveis e fáceis de manter. A POO pode ser utilizada em conjunto com diversas metodologias de desenvolvimento, como, por exemplo, o Scrum, para gerenciar o projeto, e a POO, para estruturar o código. A escolha da metodologia depende de fatores como o tamanho do projeto, a complexidade do problema e as preferências da equipe

Por fim, a metodologia build-and-test, que pode ser traduzida como "construir e testar", é uma abordagem iterativa para o desenvolvimento de *software* que se concentra em construir e testar pequenas partes do sistema de informação continuamente, priorizando e colocando o usuário no centro do processo, permitindo a criação de produtos de qualidade para atender às suas necessidades e expectativas. O estudo *Developing a Chatbot for College Student Programme*, de Chan *et al.* (2018), utilizou o método para desenvolver o *chatbot* EaSelective para auxiliar os alunos na seleção de cursos eletivos.

Na categoria metodologias, observou-se que metodologias ágeis e iterativas, em especial o Scrum, destacaram-se como as mais utilizadas na construção de *chatbots*, reforçando a importância de processos flexíveis que permitam rápidas adaptações. Em paralelo, abordagens centradas no usuário, como o Design Thinking, e metodologias tradicionais, a exemplo do SDLC e BPMN, também aparecem de forma relevante, sugerindo a combinação de práticas ágeis e modelos mais estruturados. Essas escolhas indicam a busca por uma constante interação com o usuário e por ciclos de desenvolvimento contínuos, fatores consideráveis para o sucesso de projetos de *chatbot* em um cenário tecnológico em rápida evolução.

A partir da relação das metodologias encontradas e sua ligação com a área da CI, foi observada a forma como essas metodologias lidam com o ciclo de vida

das informações, desde sua criação e organização até a recuperação e o uso efetivo pelos usuários. Ao incorporar abordagens ágeis e centradas no usuário, como o Scrum e o Design Thinking, potencializa-se a interação humano-computador, destacando os usuários, uma temática relevante nessa área. Da mesma forma, práticas estruturadas, como o SDLC e o BPMN, promovem uma visão sistemática dos processos de desenvolvimento e gestão de dados, refletindo a natureza interdisciplinar da CI (Borko, 1968), que integra técnicas de classificação, análise e utilização de dados para aprimorar a tomada de decisão e a acessibilidade da informação.

A CI tem potencial de destaque na criação de sistemas de *chatbot*, uma vez que as metodologias de desenvolvimento aplicadas são diretamente influenciadas pela maneira como os dados são organizados, representados e recuperados. Com isso, a organização lógica da informação, por meio de uma estrutura planejada, é vital para que o *chatbot* compreenda e utilize as informações de forma precisa, garantindo interações mais eficazes com os usuários (Carvalho, 2022).

A CI oferece princípios teóricos e práticos que garantem a estruturação adequada, como os metadados (Alves *et al.*, 2018), possibilitando que os *chatbots* compreendam e respondam com precisão às interações dos usuários.

Para Guimarães (2022), a relação da CI com o desenvolvimento de artefatos pode ser vista como uma aproximação transdisciplinar, pois permite a expansão das fronteiras de contato dessa área com campos que vão além dos científicos, como o desenvolvimento de artefatos tecnológicos nas áreas do *Design* e da Engenharia.

5.2.4 Modelo

Modelos baseados em PLN são utilizados para possibilitar que os *chatbots* compreendam e respondam às perguntas dos usuários de forma contextual e relevante. Os modelos aplicados na construção e no desenvolvimento de *bots*, assim como as metodologias descritas no item 5.2.1, devem ser definidos com base nas necessidades específicas do projeto, como a complexidade, o custo e a natureza da interação desejada (Hyrmet; Arbana, 2021).

Nos resultados da revisão, foi identificado o modelo RASA em quatro estudos: os autores Ramalakshmi *et al.* (2023), no estudo Using Artificial Intelligence Methods to Create a Chatbot for University Questions and Answers, destacam a eficácia do

uso do RASA para o desenvolvimento de *chatbots*, a importância da recuperação dinâmica de informações e o papel dos *chatbots* no aprimoramento do envolvimento do usuário em contextos educacionais. O estudo de Vasilevich *et al.* (2022), Language-Agnostic Knowledge Graphs for Smarter Multilingual Chatbots, utilizou o modelo RASA para desenvolver o SmartBot, *chatbot* que tem como objetivo fornecer recomendações de subsídios relevantes com base nas contribuições do usuário.

O estudo WASABI Contextual BOT, de Agrawal *et al.* (2022), e o estudo de Jiao (2020), An Intelligent Chatbot System Based on Entity Extraction Using RASA NLU and Neural Network, apresentaram o RASA como um *framework open-source* para a construção de *chatbots* e assistentes virtuais com foco em PLN e Machine Learning (ML). De acordo com os autores, o RASA permite a criação de interfaces conversacionais robustas e complexas, com capacidade de entender a linguagem natural e manter diálogos mais naturais e engajadores.

Nesta revisão, o RASA foi categorizado tanto como Modelo (5.2.4) quanto como Plataforma (5.2.6), pois, segundo Meshram *et al.* (2021), o RASA pode ser compreendido de ambas as formas para a construção de *chatbots*. Como modelo, o RASA utiliza algoritmos avançados de *machine learning*, como aprendizado por reforço e redes neurais, para processar a linguagem natural, entender a intenção do usuário e gerar respostas relevantes. Como plataforma, o RASA oferece um conjunto de ferramentas e recursos que facilitam o desenvolvimento, a implantação e o gerenciamento de *chatbots*, além de oferecerem integração com diversas plataformas de mensagens, APIs e bancos de dados, tornando-o uma solução para a construção de *chatbots* robustos e escaláveis (Meshram *et al.*, 2021).

O modelo BERT (Bidirectional Encoder Representations from Transformers) foi abordado em quatro estudos recentes: por Moore *et al.* (2023), no trabalho A Comprehensive Solution to Retrieval-Based Chatbot Construction; por Mishra *et al.* (2023), em Advancements in Conversational AI: Building Mental Health Chatbot with BERT Model; por Elsayed *et al.* (2023), em Chatbot as a Virtual Assistant to Retrieve Information from Datasheets Using Memory Controllers Domain Knowledge; e por Wang, Liu e Song (2022), em Designing an Educational Chatbot with Joint Intent Classification and Slot Filling. O BERT é um modelo de linguagem baseado na arquitetura Transformer e destaca-se por permitir o processamento paralelo de sequências de texto e por capturar relações bidirecionais entre palavras,

considerando o contexto tanto à esquerda quanto à direita. Essa abordagem inovadora marca um avanço significativo no campo do PLN, oferecendo flexibilidade e precisão para compreender o contexto linguístico. Sua capacidade de ser ajustado para uma ampla variedade de tarefas de PLN sem a necessidade de alterações substanciais em sua arquitetura reforça sua relevância tanto em pesquisas quanto em aplicações práticas.

Outro modelo destacado foi o LLM, ou Modelo de Linguagem de Grande Escala. O modelo LLM é um tipo de modelo avançado de IA, projetado para compreender e gerar textos de maneira semelhante à linguagem humana. Os modelos LLM são baseados em arquiteturas de aprendizado profundo e são treinados em grandes volumes de dados, incluindo textos de livros, artigos, *sites* e outras fontes (Sejnowski, 2023).

Foosherian *et al.* (2023), em *Enhancing Pipeline-Based Conversational Agents with Large Language Models*, destacam o aprimoramento de *chatbots* baseados em *pipelines* com LLMs. Oliveira e Matos (2023), em *Introducing a Chatbot to the Web Portal of a Higher Education Institution to Enhance Student Interaction*, apresentam a implementação de um *chatbot* com o objetivo de melhorar a interação dos estudantes em uma instituição de ensino, e Vasilevich *et al.* (2022), em *Language-Agnostic Knowledge Graphs for Smarter Multilingual Chatbots*, destacam o *chatbot* SmartBot, que utiliza um modelo de compreensão de linguagem natural para extrair intenções e entidades das entradas dos usuários por meio de grafos de conhecimento. Enquanto os modelos tradicionais se baseiam em estruturas fixas e regras explícitas, os LLMs dependem de aprendizado profundo, sendo treinados em grandes volumes de dados, o que os capacita a interpretar consultas mais complexas e fornecer respostas mais contextualizadas (Sejnowski, 2023). Todavia, a flexibilidade semântica dos modelos LLMs apresenta desafios, como possíveis alucinações e maior demanda computacional.

O modelo de aprendizagem contrastiva se mostrou eficiente para o desenvolvimento de *chatbots* baseados em recuperação, conforme destacaram Moore *et al.* (2023) em *A comprehensive solution to retrieval-based chatbot construction*. O estudo destaca o potencial da aprendizagem contrastiva em superar desafios como a desambiguação, o reconhecimento de entidades e a análise de

sentimentos, permitindo, assim, que os *chatbots* respondam de forma mais precisa e natural às interações dos usuários.

Os modelos de aprendizagem contrastiva se assemelham a fundamentos da RI, especialmente na tarefa de identificar padrões semânticos em grandes volumes de dados. Segundo Baeza-Yates e Ribeiro-Neto (2013), a eficácia dos sistemas de RI depende diretamente da capacidade de representar informações de maneira adequada e ranquear documentos com base na relevância para as consultas realizadas pelos usuários. A qualidade da recuperação se associa à modelagem eficiente dos aspectos semânticos. Em paralelo à aprendizagem contrastiva, ao focar na distinção entre exemplos similares e dissimilares, ela permite aprimorar a representação semântica das informações, facilitando a compreensão da linguagem natural pelo sistema.

O Latent Dirichlet Allocation (LDA) é um modelo probabilístico que emprega técnicas de PLN para estimar a probabilidade de ocorrência de sequências de palavras em um texto. Conforme descrito por Touimi *et al.* (2020), no estudo Intelligent Chatbot-LDA Recommender System, essa abordagem pode ser aplicada à construção de *chatbots* de recomendação, permitindo identificar tópicos e padrões linguísticos relevantes para aprimorar as respostas.

O modelo probabilístico N-grama foi destacado por Baizal *et al.* (2023), no estudo Movie recommender chatbot based on Dialogflow, ao implementar um *chatbot* de entretenimento criado para a indicação de filmes. Esse modelo, amplamente utilizado no PLN, estima a probabilidade de uma palavra ocorrer com base nas N-1 palavras anteriores em uma sequência textual. O N-grama analisa grandes volumes de texto (*corpus*) e identifica padrões de frequência, permitindo prever os termos ou frases mais prováveis em determinados contextos.

Do ponto de vista da CI, a adoção do modelo N-grama contribui diretamente para a organização e recuperação da informação em sistemas automatizados, como *chatbots* e repositórios digitais. Isso porque a modelagem por N-gramas possibilita o reconhecimento de padrões linguísticos que favorecem a identificação de intenções do usuário e a recomendação precisa de conteúdos, otimizando a busca e o acesso à informação relevante. Em ambientes como bibliotecas digitais e sistemas de recomendação, essa técnica pode ser utilizada para aprimorar a representação de

documentos e facilitar a mediação entre o usuário e os recursos informacionais disponíveis, promovendo uma recuperação mais contextualizada e eficiente.

Tanto o modelo N-grama quanto o modelo LDA utilizam a abordagem probabilística: o N-grama estima a probabilidade de sequências de palavras, enquanto o LDA calcula probabilidades de distribuição de tópicos em documentos (Touimi *et al.*, 2020).

Por fim, no estudo de Daniel *et al.* (2019), Multi-platform Chatbot Modeling and Deployment with the Jarvis Framework, foi destacado o modelo Model Driven Engineering (MDE). Este modelo de desenvolvimento de *software* oferece vantagens em relação às abordagens tradicionais. Ao colocar os modelos no centro do processo, o MDE permite que as equipes de desenvolvimento criem sistemas complexos de forma mais eficiente, com maior qualidade e menor custo.

Para a categoria modelos, verificou-se, nos trabalhos analisados, que os modelos RASA e BERT, cada um com quatro ocorrências, foram os mais adotados, demonstrando a relevância de *frameworks* abertos e de modelos pré-treinados para o desenvolvimento de *chatbots*. O crescimento de aplicações baseadas em LLM, presentes em três documentos, reforça a tendência de uso de modelos de linguagem em larga escala, alinhada ao avanço das técnicas de IA. Já métodos mais específicos, como aprendizagem contrastiva, LDA, N-gramas probabilísticos e MDE, aparecem em menor frequência, indicando abordagens pontuais para necessidades particulares de modelagem. Em conjunto, esses dados denotam a diversidade de práticas e a busca por soluções robustas na construção de ferramentas de conversação.

Pode-se observar que os modelos encontrados possuem relação com a CI no que diz respeito à organização e recuperação da informação. Por meio da capacidade de compreender contextos linguísticos complexos, o modelo aprimora a recuperação de informações. De acordo com Saracevic (1996), a relação entre a CI e as tecnologias da informação se encontra na aplicação de recursos computacionais voltados para a recuperação da informação, bem como na criação de produtos e serviços que promovam a satisfação do usuário. Em concordância, Guimarães (2022) aponta que a construção de artefatos tecnológicos, especialmente os informacionais, está profundamente vinculada à representação, à organização, ao

uso e à recuperação da informação, aspectos que são tradicionalmente tratados pela CI.

Ferneda (2012) destaca, ainda, que os avanços das tecnologias computacionais têm impulsionado significativamente a recuperação da informação, possibilitando uma gestão mais eficiente e o acesso a grandes volumes de dados. Nos *chatbots*, a recuperação da informação se torna essencial para fornecer respostas precisas e relevantes aos usuários, reforçando a convergência entre os autores sobre a relevância da CI no desenvolvimento de artefatos tecnológicos.

5.2.5 Linguagens

As linguagens de programação são conceituadas como sistemas formais de comunicação que permitem aos programadores instruir computadores a realizar tarefas específicas, desempenhando um papel na criação de *chatbots*. Essas linguagens fornecem as ferramentas para expressar algoritmos, manipular dados e controlar o comportamento do *hardware* e *software* (Mitchell, 2019). Cada linguagem possui sua própria sintaxe e semântica, que definem como os comandos são escritos e interpretados. No contexto do desenvolvimento de *chatbots*, as linguagens de programação são componentes essenciais no processo de modelagem e implementação desses *chatbots* e de sistemas computacionais.

Elas desempenham um papel nesse processo, determinando como a lógica, a compreensão da linguagem natural e as interfaces de interação são implementadas. A escolha da linguagem depende de uma série de fatores, incluindo o tipo de *chatbot*, o ambiente de uso e o grau de inteligência necessário (Srivastava; Prabhakar, 2020).

Tendo em vista investigar as tecnologias utilizadas, essa categoria e a próxima foram criadas para responder à seguinte questão: Quais são as principais tecnologias e plataformas utilizadas no desenvolvimento de *chatbots*?

A análise dos resultados revelou que o Python se destaca como uma linguagem popular para o desenvolvimento de *chatbots*, devido à sua robustez e à ampla gama de bibliotecas disponíveis para PLN e ML. Na área da saúde, existem diversos estudos, como os de Singla *et al.* (2024), Mishra *et al.* (2023) e Bendotti *et al.* (2023), que demonstraram a eficácia do Python no desenvolvimento de *chatbots* para assistência médica e monitoramento de pacientes. Já no contexto educacional,

pesquisadores como Ramalakshmi *et al.* (2023), Nasharuddin *et al.* (2021), Muhyidin e Setiawan (2021), e Touimi *et al.* (2020) exploraram o uso do Python em *chatbots* voltados para ensino e aprendizagem.

Destaca-se, em particular, o estudo de Nasharuddin *et al.* (2021), que investigou a integração do Python com a plataforma Telegram, permitindo o desenvolvimento de assistentes virtuais com a capacidade de aprendizado contínuo, o que contribui para a adaptação dinâmica às necessidades dos usuários. Essas características tornam os *chatbots* mais flexíveis e eficazes, aprimorando a interação e a personalização das respostas ao longo do tempo, seja na saúde ou na educação.

O JavaScript foi a segunda linguagem que se destacou nos estudos e se revelou utilizada no desenvolvimento de *chatbots* por sua versatilidade e capacidade de criar aplicações web interativas. Estudos como os de Lynnyk, Krestyanpol e Rozvod (2024), Wang, Liu e Song (2022), Santana *et al.* (2021), López *et al.* (2019) e Bendotti *et al.* (2023) evidenciam a aplicação do JavaScript em diversos contextos, incluindo *chatbots* para atendimento ao cliente, suporte técnico e interação social. Esses estudos demonstram a ampla diversidade de usos da linguagem, desde a criação de interfaces interativas até a implementação de funcionalidades complexas no backend, o que ressalta sua capacidade de atender às crescentes demandas tecnológicas do desenvolvimento de *software* moderno.

A linguagem PHP (Hypertext Preprocessor), encontrada em cinco estudos da revisão, se caracteriza como uma linguagem de programação popular no desenvolvimento *web*, especialmente para a criação de conteúdos dinâmicos e interativos. No contexto de *chatbots*, a PHP tem sido utilizada para integrar essas ferramentas em páginas *web* e plataformas *online*. Estudos de Iparraguirre-Villanueva *et al.* (2023), López *et al.* (2019), Agus Santoso *et al.* (2018), Mendoza *et al.* (2022) e Sanjaya e Santoso (2021) destacam a eficácia da PHP em diversas aplicações, como e-commerce, serviços *online* e comunicação interativa. Sua integração com servidores *web* permite a geração de páginas dinâmicas e interativas, fundamentais para plataformas modernas e responsivas. A literatura mostra que, mesmo com a evolução das tecnologias, a PHP continua sendo uma escolha eficaz, devido à sua capacidade de criar experiências interativas e facilitar a

interação contínua entre usuários e sistemas, especialmente em projetos que exigem flexibilidade e dinamismo.

Outras linguagens, como Cascading Style Sheets (CSS) e Cython, também se destacaram pontualmente nos estudos. O chat UNIBOT, desenvolvido para auxiliar consultas e fornecer respostas rápidas aos usuários de uma instituição de ensino, utilizou o CSS para o desenvolvimento do *front-end* do assistente. Por sua vez, Cython foi a linguagem empregada no desenvolvimento das funcionalidades de um *bot* criado para cotação, consultas relacionadas à busca de preços, limites e volume de ações.

Portanto, na categoria linguagem, pôde-se observar que o Python se destaca como a linguagem mais utilizada (oito ocorrências), possivelmente devido à sua ampla gama de bibliotecas e ferramentas voltadas à IA e ao aprendizado de máquina, aspectos essenciais para o desenvolvimento de *chatbots*. Em seguida, JavaScript e PHP aparecem com cinco referências cada, refletindo sua relevância tanto no *front-end* quanto no *back-end* de aplicações *web*, ou seja, *full-stack*. Já CSS e Cython, cada uma com apenas uma ocorrência, indicam usos pontuais para a estilização de interfaces e otimização de desempenho, respectivamente. Esses dados reforçam a predominância de linguagens que oferecem robustez e flexibilidade para lidar com a complexidade inerente a sistemas de conversação.

A escolha da linguagem de programação adequada para o desenvolvimento de *chatbots* depende de uma série de variáveis, incluindo a demanda por interatividade, os tipos de informações que o sistema precisa processar, a necessidade de integração com outras plataformas e a complexidade das funcionalidades desejadas, o que resulta em uma variedade de opções que oferecem soluções distintas para os desafios enfrentados pelos desenvolvedores.

Na área CI, conforme apontado por Boroko (1968) e Saracevic (1996), as linguagens de programação constituem ferramentas e tecnologias indispensáveis para o desenvolvimento de sistemas de organização, recuperação e análise de dados. A linguagem de programação Python, por exemplo, destaca-se pela ampla adoção em projetos de mineração de dados e aprendizado de máquina, áreas estratégicas para lidar com grandes volumes de informação, conforme destacado por Lancaster (2004) desde os anos 1970.

Para além disso, o uso de diferentes linguagens de programação possibilita a criação de interfaces informacionais inovadoras, a exemplo de catálogos interativos, bibliotecas digitais, sistemas de busca especializada e serviços de referência virtual. Essas soluções tecnológicas permitem aprimorar a mediação e o acesso à informação, contribuindo para a experiência do usuário e para a eficiência dos processos de organização e recuperação de dados no contexto informacional.

5.2.6 Plataforma

As plataformas para a criação de *chatbots* oferecem um conjunto de ferramentas que permitem aos desenvolvedores criar, treinar, implementar e gerenciar *chatbots* com mais facilidade. Embora algumas plataformas possibilitem a personalização avançada para desenvolvedores, elas geralmente dispensam um conhecimento profundo de programação (Touimi *et al.*, 2020). Recursos como PLN, escalabilidade, análise, interface amigável, suporte para vários idiomas e acesso a informações e serviços facilitam a gestão e o desenvolvimento dos *chatbots*.

As pesquisas recuperadas neste estudo indicaram que os principais benefícios do uso de plataformas para o desenvolvimento de *chatbots* incluem uma experiência do cliente aprimorada, redução de custos e a possibilidade de modernizar processos comerciais tradicionais por meio de interações eficientes baseadas em IA (Baizal *et al.*, 2023; Chumkaew, 2023; Iparraguirre-Villanueva *et al.* 2023).

Os resultados demonstram a predominância do Dialogflow e do Google Cloud (17 ocorrências), seguidos pelo RASA (cinco ocorrências) e pelo IBM Watson (quatro ocorrências). As plataformas Kommunicate e LINE obtiveram duas ocorrências cada, enquanto ENGATI, Amazon Lex, Django Software Foundation, HEROKU, Arisa Nest, Discord, Facebook, NEORON e MOOC EAD foram mencionadas em apenas um estudo cada.

Tanwar *et al.* (2023) e Thaiprasert *et al.* (2023) concordam com isso em seus estudos sobre a aplicabilidade do Dialogflow em ambientes que exigem um alto nível de interatividade e resposta imediata. A adoção do Google Dialogflow, também evidenciada nos estudos de Alqaidi, Alharbi e Almatrafi (2021), Nasharuddin *et al.* (2021) e Erekata *et al.* (2021), se confirmam ao se relacionar a sua capacidade de integração com outras ferramentas e plataformas tecnológicas, como a Google

Cloud e a API, externas. Por exemplo, Nasharuddin *et al.* (2021) e Muhyidin e Setiawan (2021) relatam que o uso do Dialogflow, em conjunto com tecnologias de análise de dados e sistemas de *backend*, melhorou a eficácia dos *chatbots*, tornando-os mais responsivos e aptos a oferecer experiências personalizadas aos usuários.

Destaca-se que o RASA foi alocado tanto na categoria 5.2.4 Modelo quanto na categoria 5.2.6 Plataforma, descritas nos estudos de Elsayed *et al.* (2023), Chatbot as a Virtual Assistant to Retrieve Information from Datasheets Using Memory Controllers Domain Knowledge, Ramalakshmi *et al.* (2023), Using Artificial Intelligence Methods to Create a Chatbot for University Questions and Answers, e Oliveira e Matos (2023), Introducing a Chatbot to the Web Portal of a Higher Education Institution to Enhance Student Interaction.

Portanto, na categoria plataforma, percebeu-se a predominância do Google DialogFlow, evidenciando sua capacidade e ampla adoção em projetos de *chatbots*. Em seguida, RASA e IBM Watson surgem como alternativas relevantes, embora outras plataformas — como Kommunicate, LINE, Engati e Amazon Lex — também tenham sido mencionadas, contribuindo para a diversidade de opções disponíveis.

As plataformas de criação de *chatbots* se mostram alinhadas às práticas da CI ao viabilizarem a organização, a recuperação e a disseminação do conhecimento de maneira dinâmica e personalizada. Conforme pontuado por Borko (1968), a CI lida com os processos de geração, coleta, organização e uso de informações, e as tecnologias de interação conversacional, como os *chatbots*, potencializam esses processos ao criarem ambientes que respondem de forma imediata às necessidades informacionais dos usuários. Dessa forma, soluções como DialogFlow, RASA e IBM Watson exemplificam iniciativas voltadas ao aperfeiçoamento da comunicação entre pessoas e sistemas, facilitando o acesso e a interpretação de dados em múltiplos contextos. Ademais, a adoção de plataformas remete ao princípio de Lancaster (2004) sobre a relevância de sistemas automatizados para o avanço da recuperação de informações.

Ao empregarem técnicas de PLN e algoritmos de IA, os *chatbots* contribuem para tornar os fluxos de informação mais eficientes e escaláveis, atendendo às demandas crescentes na era digital (Carvalho, 2022). Tal perspectiva reforça o valor de práticas relacionadas ao planejamento, à gestão e à avaliação de sistemas

informativos, essenciais para a consolidação de metodologias e ferramentas que possibilitem uma interação cada vez mais satisfatória entre pessoas e metadados, em consonância com os objetivos da CI.

5.2.6.1 Técnicas de Processamento de Linguagem Natural (PLN)

O PLN é um campo da IA dedicado ao estudo das interações entre computadores e a linguagem humana que envolve tanto a compreensão de textos e falas quanto a geração automática de linguagem natural (Nadkarni; Ohno-Machado; Chapman, 2011). Essa área abrange técnicas e modelos computacionais que permitem a análise de estruturas linguísticas, como sintaxe e semântica, além de aplicações práticas em tarefas como tradução automática, classificação de textos, *chatbots* e resumo automático (Caseli, Nunes, 2024; Comarella; Café, 2008). Ao combinar conhecimentos de linguística, estatística e aprendizado de máquina, o PLN busca capacitar os sistemas computacionais a lidarem de forma eficiente com as ambiguidades, variações e contextos próprios da língua humana.

Nos estudos desta revisão, destacaram-se as técnicas TF-IDF (Term Frequency – Inverse Document Frequency), que medem a relevância de cada termo em um conjunto de documentos; a Seq2seq (Sequence to Sequence), uma arquitetura baseada em redes neurais voltadas para tarefas como tradução e geração de texto; e a Bag of Words, abordagem que converte textos em conjuntos de palavras, facilitando a identificação de padrões lexicais e estatísticos.

Nesta categoria, optou-se por destacar as técnicas de PLN, tendo em vista sua relevância e predominância em alguns estudos. Segundo Caseli e Nunes (2024), o PLN é um campo de pesquisa que tem como objetivo investigar e propor métodos e sistemas de processamento computacional da linguagem humana. O termo “Natural”, na sigla, se refere às línguas faladas pelos humanos, distinguindo-as das demais linguagens (matemáticas, visuais, gestuais, de programação etc).

O PLN pode ser dividido em duas grandes subáreas: Interpretação de Linguagem Natural (NLU) e Geração de Linguagem Natural (NLG). O PLN foca em entender a entrada do usuário e gerar a saída correspondente, incluindo componentes como o reconhecimento automático de texto e o gerenciamento de diálogos (Caseli; Nunes, 2024). Essas abordagens estão alinhadas com as ideias de

Rukhiran e Netinant (2022), que, no estudo Recuperação automatizada de informações e serviços da pós-graduação usando o sistema chatbot, destacam que o PLN e a CI compartilham objetivos relacionados à organização, representação e recuperação de informações textuais e linguísticas. Segundo os autores Taskin e Al (2019), o PLN desenvolve métodos e ferramentas para que máquinas compreendam, processem e gerem linguagem humana, ampliando as interações entre usuários e sistemas informacionais. Por sua vez, a CI oferece a base teórica necessária para estruturar e recuperar a informação, levando em consideração aspectos como relevância, contexto e as necessidades dos usuários.

Essa interação entre PLN e CI é especialmente evidente em aplicações como motores de busca, sistemas de recomendação e *chatbots*. Nesses sistemas, técnicas de PLN, como análise semântica, extração de entidades e sumarização automática, são utilizadas para aprimorar a precisão e a acessibilidade da informação (Caseli; Nunes, 2024). Assim, PLN e CI se complementam, unindo os avanços tecnológicos da linguística computacional com os fundamentos da organização e recuperação de informação, o que resulta em sistemas mais eficientes e alinhados às necessidades informacionais dos usuários. Cada uma dessas abordagens atende a necessidades específicas do PLN e pode ser combinada a outras soluções, o que resulta em um desempenho mais aperfeiçoado e adequado às demandas dos usuários (Yang *et al.*, 2023).

5.2.6.2 Term Frequency-Inverse Document Frequency (TF-IDF)

Matthews (2020) apresenta o TF-IDF como um método clássico e ainda útil para recuperação de informações em sistemas de *chatbots* simples ou baseados em palavras-chave. O método serve como um ponto de contraste para mostrar a evolução das técnicas no desenvolvimento de *chatbots*, revelando-se eficiente e muito implementado, sendo relevante em contextos em que os custos computacionais de LLMs não são viáveis. Os modelos baseados em indexação, como TF-IDF, têm sido pilares na RI. O método mede a relevância de um termo em um documento com base em sua frequência (TF) e raridade em todo o *corpus* (IDF) (Manning; Raghavan; Schütze, 2009). O TF-IDF é uma técnica bastante utilizada para representar texto em forma vetorial, considerando a frequência de termos em um documento e sua raridade no *corpus*. Nos *chatbots* baseados em recuperação,

técnicas como TF-IDF são utilizadas para medir a similaridade entre a consulta do usuário e os documentos pré-existent no banco de dados (Salton, 1968). Apesar de sua simplicidade, o TF-IDF é eficaz para consultas que envolvem palavras-chave exatas e bases de conhecimento específicas.

Nos estudos da revisão, os autores Baizal *et al.* (2023), com o título Movie recommender chatbot based on Dialogflow, Tanwar *et al.* (2023), com AI Based Chatbot for Healthcare using Machine Learning, Thaiprasert *et al.* (2023), com Development of a Class Materials Search System using LINE Chatbot, Santana *et al.* (2021), com A Chatbot to Support Basic Students Questions, Prakasam *et al.* (2023), com Design and Development of AI-Powered Healthcare WhatsApp Chatbot, e Attigeri, Agrawal e Kolekar (2024), com Advanced NLP Models for Technical University Information Chatbots: Development and Comparative Analysis, destacaram a utilização do TF-IDF para implementar os *chatbots* em educação, saúde e entretenimento, respectivamente. Os estudos mencionam que os *chatbots* podem se beneficiar do TF-IDF, principalmente nas áreas de compreensão da linguagem natural e RI.

5.2.6.3 Seq2seq

A técnica Seq2seq (Sequence-to-Sequence) foi introduzida por Sutskever, Vinyals e Le (2014) e descreve uma arquitetura de redes neurais capaz de transformar uma sequência de entrada em outra sequência de saída. A Seq2seq faz uso de redes neurais para transformar uma sequência de entrada em outra sequência de saída, sendo aplicada em atividades como tradução automática, sumarização e geração de legendas para imagens (Sutskever; Vinyals; Le, 2014). De modo geral, o codificador (*encoder*) lê os elementos fornecidos (palavras, caracteres ou tokens) e condensa as informações em um vetor de contexto. Em seguida, o decodificador (*decoder*) utiliza esse vetor para gerar a sequência de saída, elemento por elemento. Essa estratégia tem sido aplicada em tarefas de PLN, como tradução automática, sumarização e geração de texto (Karri; Kumar, 2020). As versões aprimoradas do “Seq2seq” fazem uso de mecanismos de atenção (*attention*), o que permite ao decodificador focar em partes específicas do vetor de contexto e produzir resultados mais precisos na geração de sequências.

A técnica Seq2seq foi mencionada no estudo de Karri e Kumar (2020), *Deep Learning Techniques for Implementation of Chatbots*, que destacou, ainda, parâmetros que auxiliam na criação de um *chatbot* útil, como: escalabilidade, interoperabilidade, velocidade e Teste de Turing, sendo este último a capacidade do *chatbot* em simular o comportamento humano, ou seja, testar a capacidade de uma máquina que exibe comportamento inteligente equivalente a um humano.

5.2.6.4 Bag of words

A técnica “Bag of Words” converte o texto em um conjunto de palavras, sem considerar a ordem ou as relações de dependência entre elas (Rudkowsky *et al.*, 2018). Embora simples, essa técnica é valiosa em classificações e na detecção de tópicos, uma vez que fornece uma representação estatística dos dados linguísticos (Jurafsky; Martin, 2023). Nesse contexto, cada palavra torna-se uma unidade, e o texto é representado por meio de frequências de ocorrência desses termos ou pela presença/ausência de cada um (Manning; Raghavan; Schütze, 2009). Essa abordagem, embora simples, se mostrou eficaz em aplicações de PLN, como classificação de documentos, análise de sentimentos e detecção de tópicos, pois fornece uma representação estatística a partir do vocabulário utilizado no texto, e foi utilizada em conjunto ao Seq2seq por Karri e Kumar (2020), em *Deep Learning Techniques for Implementation of Chatbots*.

No contexto da CI, a abordagem Bag of Words pode ser útil para a organização e recuperação da informação em ambientes que lidam com grandes volumes de dados textuais, como repositórios institucionais, bibliotecas digitais e sistemas de busca documental (Manning; Raghavan; Schütze, 2009). Ao possibilitar a indexação automática de conteúdos e a classificação por similaridade textual, a técnica contribui para o mapeamento de temas, o agrupamento de documentos e a recomendação de materiais relevantes aos usuários (Jurafsky; Martin, 2023).

Por outro lado, a principal limitação semântica da técnica está no fato de ela desconsiderar o contexto em que as palavras aparecem, o que pode resultar em interpretações superficiais ou imprecisas, especialmente em casos em que o significado depende da relação entre termos (Manning; Raghavan; Schütze, 2009). Dessa forma, embora eficiente para a análise exploratória e a organização inicial de

grandes coleções textuais, a Bag of Words pode não capturar nuances semânticas importantes para a recuperação precisa da informação.

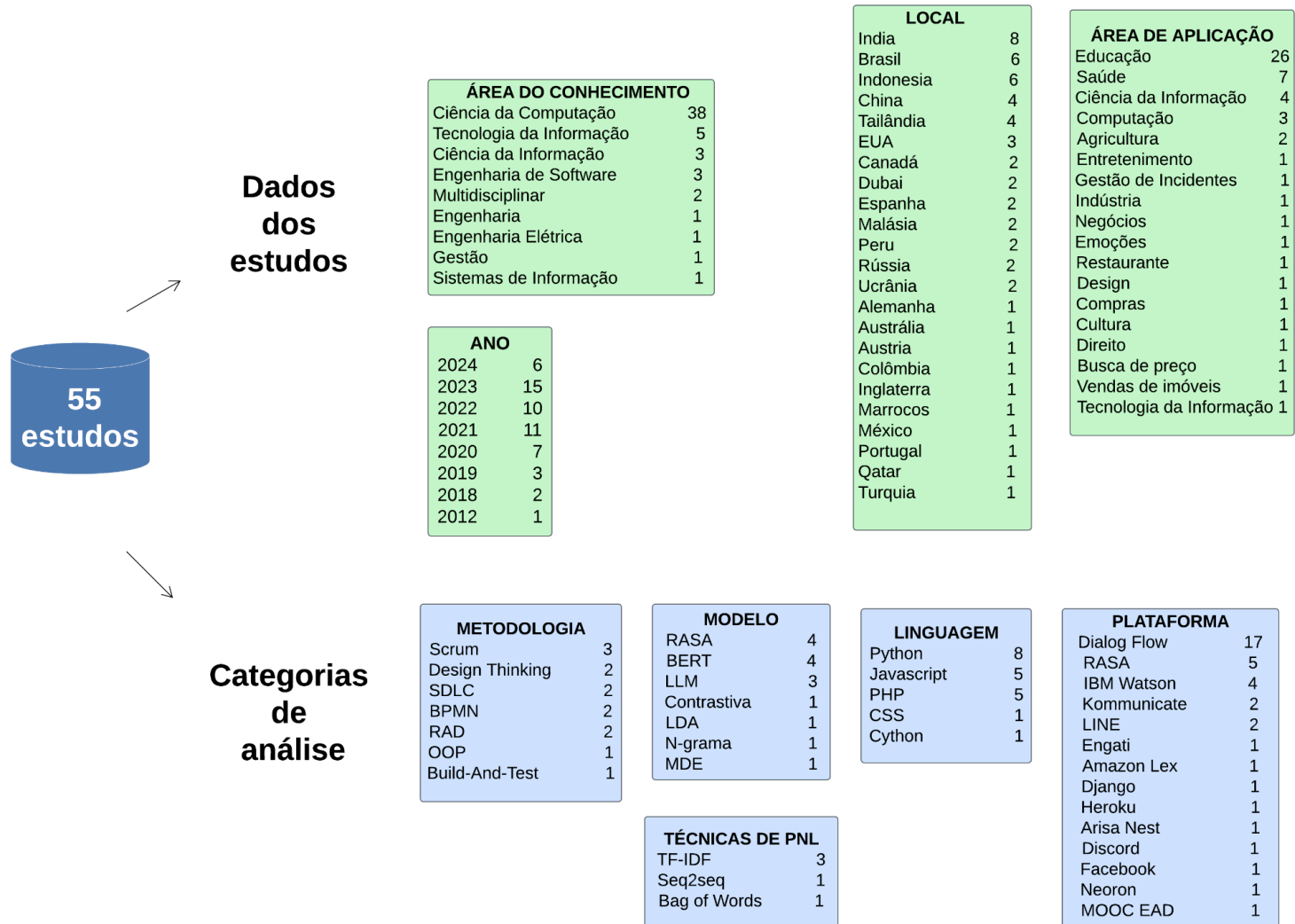
5.3 Considerações da revisão

O intuito da revisão de literatura foi identificar a produção científica sobre o desenvolvimento/construção de *chatbots*, permitindo a elaboração de um mapeamento, abordando aspectos como metodologias, modelos, linguagens de programação, plataformas e técnicas empregadas na construção de *bots* de conversação. Para orientar a revisão, foram formuladas cinco questões principais: (1) Quais são as metodologias utilizadas e empregadas no desenvolvimento de *chatbots*?; (2) A quais áreas do conhecimento estão relacionados?; (3) Quais são os principais domínios de aplicação dos *chatbots* na literatura atual?; (4) Quais são as principais tecnologias e plataformas utilizadas no desenvolvimento de *chatbots*? e; (5) Quais aspectos sobre organização, representação e recuperação da informação foram identificados?, que guiaram a análise dos estudos selecionados, bem como as categorias de análise, a saber: Metodologias, Modelos, Linguagens, Plataformas e Técnicas de PLN.

As categorias foram desenhadas com o intuito de explorar a diversidade de abordagens e técnicas presentes na literatura, além de compreender como as diferentes áreas estão utilizando os *chatbots* para otimizar processos e serviços.

Analisando todas as cinco categorias, foi possível observar que, apesar de estar relacionada com várias técnicas e processos, a CI ganha pouco destaque, ainda que a informação seja o elemento fundamental dos *chatbots*. Para representar a visão geral da revisão, a Figura 8 foi elaborada, com base em diferentes aspectos relacionados a (1) dados sobre as publicações e (2) análise de conteúdo. Na análise dos dados sobre as publicações, as informações estão divididas em: local, ano de publicação e área do conhecimento. Para a análise de conteúdo, foram identificadas cinco categorias, considerando uma classificação facetada, tendo em vista o fato de que algumas soluções estão em mais de uma categoria.

Figura 8 - Síntese da revisão de literatura



Fonte: elaborado pela autora (2025).

A revisão de literatura buscou compreender as técnicas e os métodos empregados no desenvolvimento e na implementação de *chatbots*, com ênfase nas contribuições da CI. Os resultados revelam que esses sistemas conversacionais têm se consolidado como ferramentas relevantes de mediação e recuperação da informação (Patel *et al.*, 2019; Panda; Chakravarty, 2022; Chumkaew, 2023; Moore *et al.*, 2023), sobretudo diante da crescente complexidade informacional (Karri; Kumar, 2020) e da demanda por agilidade, acessibilidade e personalização no acesso ao conhecimento (Villanueva; Aguilar-Alonso, 2021; Stepanov *et al.*, 2023; Attigeri, Agrawal e Kolekar, 2024).

Nesse contexto, observou-se um alinhamento entre os fundamentos clássicos da CI, como a organização, representação e recuperação da informação, e as técnicas empregadas na construção de *chatbots* (Rodriguez; Mune, 2022; Elsayed *et al.*, 2023). As abordagens identificadas variam desde modelos baseados em regras até arquiteturas sofisticadas de aprendizado profundo, como *transformers* e sistemas híbridos (Baizal *et al.*, 2023; Foosherian *et al.*, 2023; Wang; Liu; Song, 2022; Whittle; Hall, 2022).

Apesar dos avanços tecnológicos, permanecem lacunas significativas no que diz respeito à estruturação e ao tratamento das bases informacionais que alimentam esses sistemas (Guimarães, 2022; Buckland, 1991; Lima, 2020), objetos de estudo desta pesquisa. Muitos estudos trabalham com documentos dispersos, como arquivos PDF, planilhas e conteúdos extraídos de *sites* institucionais, os quais nem sempre seguem padrões estruturais compatíveis com a leitura automatizada (Lancaster, 2004; Kingchang; Chatwattana; Wannapiroon, 2024; Santana *et al.*, 2021; Rukhiran; Netinant, 2022). Essa desorganização afeta diretamente a precisão da recuperação e a confiabilidade das respostas oferecidas aos usuários.

Também se constatou a escassez de estratégias de curadoria da informação, ancoradas em modelos consolidados de organização e representação. A ausência dessas práticas dificulta a construção de sistemas mais robustos e sustentáveis, sobretudo em contextos institucionais, como o universitário, em que há alta rotatividade normativa e necessidade constante de atualização (Prakasam *et al.*, 2023; Mishra *et al.*, 2023; Chase, 2024).

Outro ponto recorrente refere-se à baixa adoção de estruturas semânticas, como vocabulários controlados, taxonomias, tesouros e ontologias. Tais

instrumentos são reconhecidos por promover a desambiguação e a padronização terminológica, além de aumentar a eficácia dos processos de recuperação (Allison, 2012; Agus Santoso *et al.*, 2018; Touimi *et al.*, 2020). Entretanto, sua incorporação prática nos projetos analisados ainda é limitada, o que compromete a coerência dos diálogos automatizados e a consistência das respostas (Hodge, 2000; Soergel, 1999).

No que se refere à manutenção dos sistemas, observou-se a ausência de procedimentos bem definidos para atualização automatizada, controle de versão ou validação periódica das bases de conhecimento (Oliveira; Matos, 2023; Mendoza *et al.*, 2022; Singla *et al.*, 2024). Essa lacuna é especialmente preocupante em ambientes com dinâmica informacional acelerada, como universidades públicas.

Além disso, a avaliação dos *chatbots* permanece centrada em métricas computacionais tradicionais (ex: acurácia, F-score), enquanto critérios informacionais mais qualitativos, como clareza, relevância contextual, satisfação do usuário e efetividade da mediação, são menos abordados (Feitoza, 2021; Wang; Liu; Song, 2022).

Por fim, constatou-se que a maioria dos estudos analisados concentra-se predominantemente em aspectos técnico-computacionais, com pouca incorporação das dimensões conceituais e metodológicas próprias da CI (Borko, 1968). Esses trabalhos apresentam *frameworks* bem estruturados, metodologias replicáveis e diretrizes claras que favorecem a reutilização institucional, a consolidação de boas práticas e a evolução integrada dos sistemas (Rosenfeld; Morville; Arango, 2015).

Um aspecto que ajuda a explicar essa lacuna é a própria heterogeneidade interdisciplinar dos projetos de *chatbots*, frequentemente desenvolvidos por equipes voltadas a questões técnicas e da área de Computação, que priorizam o desempenho algorítmico e a implementação técnica em detrimento de fundamentos informacionais. A ausência de profissionais ou referências diretas da CI nas etapas de concepção, curadoria e avaliação pode contribuir para que elementos estruturantes, como políticas de atualização, padronização semântica e critérios qualitativos de mediação sejam pouco contemplados. Assim, a integração mais sólida entre práticas informacionais e soluções tecnológicas torna-se um requisito para garantir que esses sistemas avancem não só em eficiência computacional, mas

também em confiabilidade, coerência conceitual e aderência às necessidades reais dos usuários.

6. COMPORTAMENTO DE BUSCA DOS ESTUDANTES DE GRADUAÇÃO DA UFMG

A pesquisa realizada no âmbito do projeto Recuperação de Informação em modelos de Inteligência Artificial: o papel da Ciência da Informação buscou identificar o comportamento de busca dos estudantes de graduação da UFMG em relação ao acesso a informações acadêmicas, com ênfase em normas, regimentos e resoluções institucionais.

6.1 Metodologia

A pesquisa caracterizou-se como aplicada, exploratória e descritiva, com abordagem mista (quantitativa e qualitativa). A natureza aplicada busca solucionar problemas práticos relacionados ao acesso e à recuperação de informações acadêmicas na UFMG. O caráter exploratório busca maior familiaridade com o comportamento informacional dos discentes, enquanto o aspecto descritivo apresenta características dessa população específica (Gil, 2017).

A abordagem mista segue os preceitos de Creswell (2013), combinando a mensuração de padrões comportamentais (quantitativa) com o aprofundamento nas percepções e experiências dos participantes (qualitativa), proporcionando uma compreensão mais abrangente do fenômeno investigado (Prodanov; Freitas, 2013).

O delineamento configura-se como um *survey* transversal, método adequado para investigar comportamentos e opiniões de uma população em um momento específico (Severino, 2013). Segundo Fowler Jr. (2011), essa estratégia permite identificar características populacionais a partir de amostras representativas, possibilitando generalizações estatisticamente válidas.

6.1.1 População e amostra

A pesquisa amostral foi realizada por meio de coleta direta, com questionários contendo perguntas abertas e fechadas, sendo direcionada à população de estudantes de graduação da UFMG. Segundo dados oficiais de maio de 2021 (UFMG, 2021), a Universidade possui 34.482 alunos de graduação e 10.556 alunos de pós-graduação, totalizando uma população de 45.038 alunos.

O cálculo do tamanho amostral para o percentual de frequência em uma população considerou o tamanho da população = 45.038, a frequência (p) antecipada de 50%, o limite de confiança igual a 5 (precisão absoluta) e o efeito de desenho igual a 1.0 (para amostras aleatórias). O cálculo foi realizado utilizando a ferramenta OpenEpi², que calculou o tamanho da amostra de 381 para um intervalo de confiança de 95%. Considerando uma perda de 30%, o valor foi ajustado para uma amostra de 496 participantes. Além disso, será considerada a amostra estratificada, considerando a representatividade conforme a vinculação dos discentes (graduação e pós-graduação), as áreas de conhecimento, bem como outra categorização possível de ser identificada na amostra. Apesar dos cálculos e indicadores de representatividade, destaca-se que a pesquisa utilizou uma amostragem por conveniência, tendo em vista que há um grande quantitativo de cursos de graduação na UFMG, e essa representatividade não considerou os aspectos de cada curso.

6.1.2 Aspectos éticos e instrumentos de coleta

A pesquisa obedeceu às diretrizes éticas previstas na Resolução CNS nº 510/2016, garantindo o anonimato e o caráter voluntário da participação, caracterizando-se como uma pesquisa de opinião pública. Para isso, foi elaborado o Termo de Consentimento Livre e Esclarecido (TCLE), cujo aceite foi condição para iniciar o preenchimento do questionário. O instrumento foi aplicado via Google Forms, sendo configurado para não coletar informações pessoais, como e-mail, e escolhido pela ampla acessibilidade e facilidade de integração com canais de divulgação internos à UFMG. Os dados foram armazenados em computadores pessoais e serão descartados após cinco anos, conforme recomendado na legislação vigente.

6.1.3 Fluxo do questionário

O questionário foi construído seguindo uma categorização própria, com a caracterização do perfil e três seções. Assim, o questionário se inicia com a identificação do nível acadêmico (graduação ou pós-graduação), restringindo-se,

² Disponível em: <https://www.openepi.com>. Acesso em 11 out. 2025.

nesta análise, aos estudantes de graduação. Estes são categorizados em três grupos, conforme o período cursado: 1º ao 4º; 5º ao 8º; e 9º ou mais. Após a identificação do curso e do período cursado, o questionário investiga, na Seção 1: Quando, como e onde acessam as informações acadêmicas; na Seção 2: Como identificam e qual a percepção sobre as informações acadêmicas; e na Seção 3: Quais são as dúvidas, opiniões e sugestões.

O instrumento combina perguntas fechadas e abertas, permitindo análises quantitativas e qualitativas abrangentes sobre o acesso às informações acadêmicas (normas, resoluções e procedimentos), conforme apresentado no Anexo A.

6.1.4 Divulgação do questionário

A divulgação do questionário foi realizada entre 16 de dezembro de 2024 e 15 de março de 2025, utilizando três estratégias principais: grupos de WhatsApp direcionados aos estudantes; envio via listas institucionais de e-mail, por meio do serviço de distribuição em massa da Diretoria de Tecnologia da Informação (DTI UFMG), serviço que garantiu a distribuição para todos alunos ativos de graduação; e a fixação de materiais impressos em locais estratégicos da Universidade. Essa abordagem buscou maximizar o alcance e a diversidade da amostra.

6.1.5 Análise dos resultados

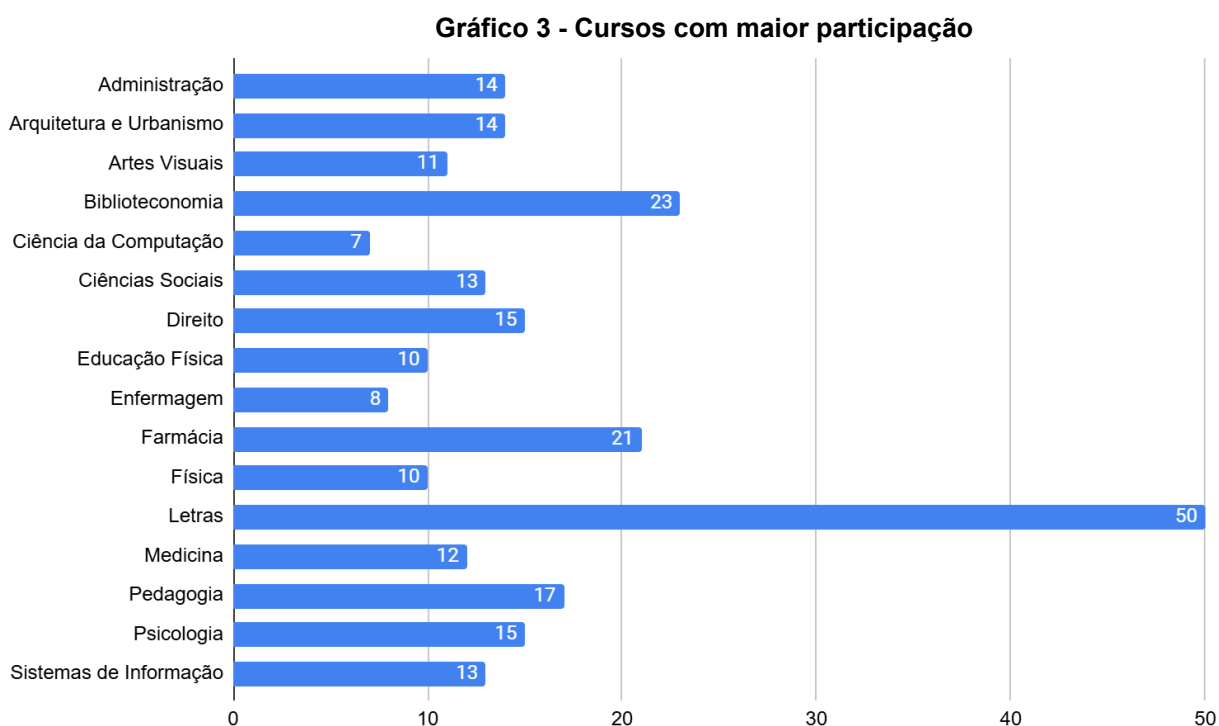
Os resultados foram apresentados por estatística descritiva, enquanto as respostas abertas foram analisadas por meio da técnica de análise de conteúdo, de Bardin (2016). Essa análise compreendeu pré-análise, codificação temática e interpretação dos dados, adaptando o vocabulário dos discentes ao contexto normativo acadêmico. A interpretação desses dados subsidia as etapas posteriores da pesquisa, destacando os desafios e as necessidades informacionais da graduação.

6.1.6 Descrição da amostra

Foram obtidas 795 respostas ao questionário, das quais 527 (66,3%) referem-se a estudantes de graduação, superando o tamanho amostral estimado

para essa população, o que reforça a representatividade e diversidade dos dados coletados.

Entre os graduandos, foram identificados respondentes de 73 cursos distintos, com apenas uma resposta sem identificação de curso. Os cursos com maior número de participantes pertencem às áreas de Humanidades e Ciências Sociais Aplicadas, destacando-se: Letras (50), Biblioteconomia (23), Pedagogia (17), Psicologia (15), Administração (14), Arquitetura e Urbanismo (14), Ciências Sociais (13) e Sistemas de Informação (13), conforme apresentado no Gráfico 3, ordenados pelos nomes do cursos.

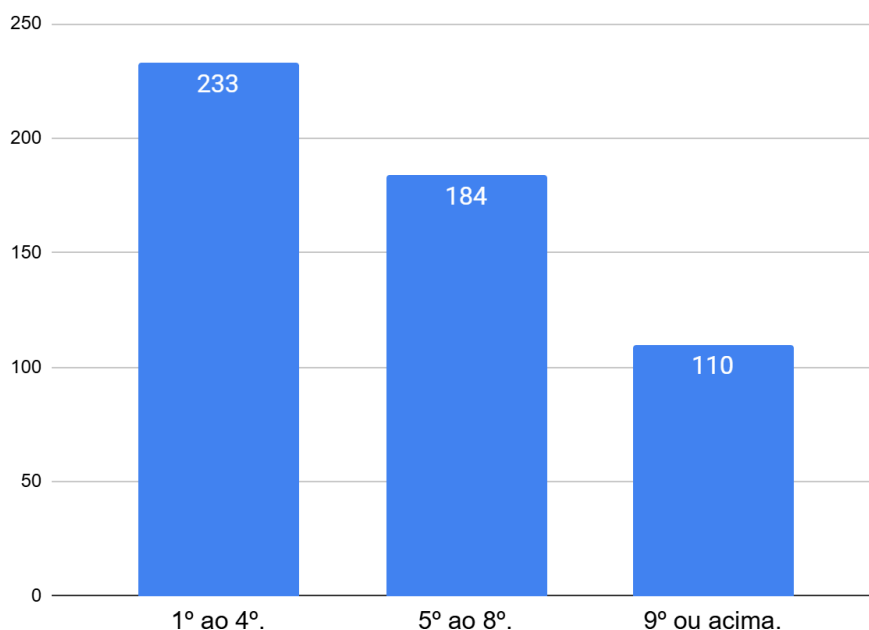


Fonte: dados da pesquisa (2025).

Quanto à distribuição dos estudantes pelos períodos do curso, a amostra contempla todos os períodos da graduação, classificados em três faixas: 1º ao 4º período, 5º ao 8º período e 9º período ou superior, conforme apresentado no Gráfico 4, assegurando a representatividade em diferentes fases do percurso acadêmico. A distribuição por períodos acadêmicos revelou representatividade equilibrada: 233 estudantes entre o 1º e o 4º período (44,2%), 184 entre o 5º e o 8º período (34,9%) e 110 no 9º período em diante (19,0%). Essa diversidade permite uma análise abrangente dos comportamentos informacionais ao longo da trajetória acadêmica,

considerando que a complexidade das demandas por informação normativa reduz com o avanço no curso.

Gráfico 4 - Graduação X Período



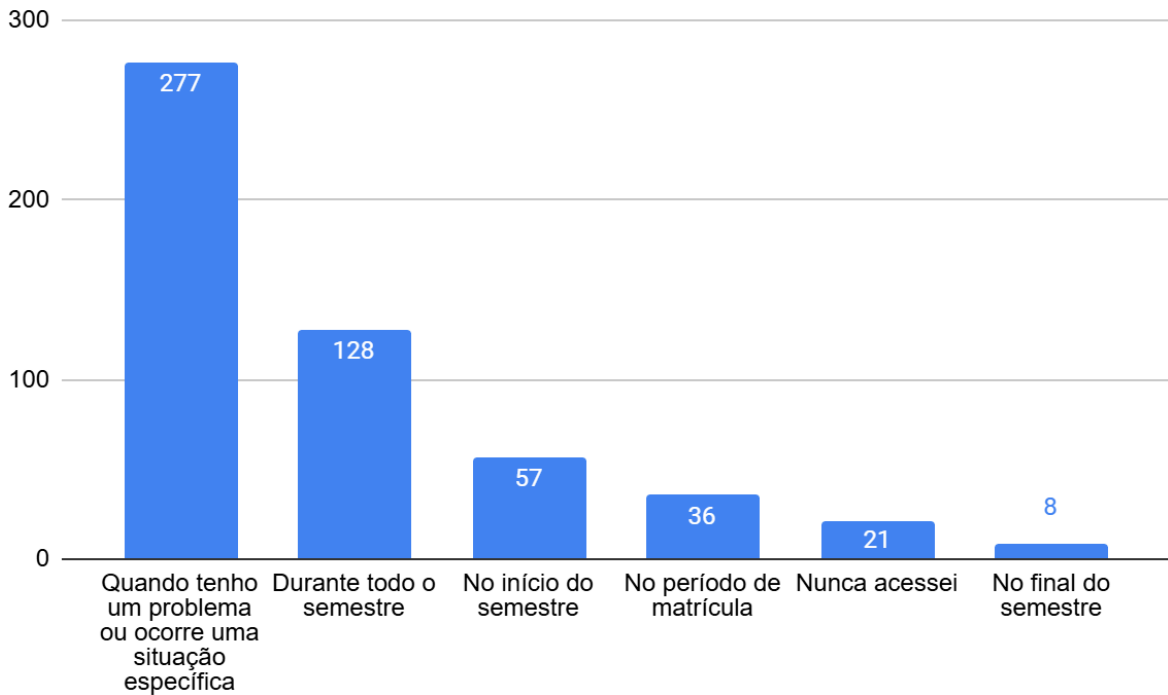
Fonte: dados da pesquisa (2025).

Nas próximas seções, as respostas foram analisadas conforme as seções do formulário.

6.2 Seção 1 - Quando, Como e Onde os estudantes de graduação acessam as informações acadêmicas

Em relação ao momento de acesso, a análise dos dados quantitativos sobre o momento da busca informacional revelou um comportamento predominantemente reativo entre os graduandos. Conforme apresentado no Gráfico 5, mais da metade dos participantes, 277 (52,56%), acessa informações institucionais quando tem um problema ou em alguma situação específica. Esse padrão contrasta com a busca proativa observada em parcelas menores da amostra, que realizam consultas ao longo do semestre, 128 (24,28%), ou no início do período letivo, 57 (10,81%).

Gráfico 5 - Momentos de acesso



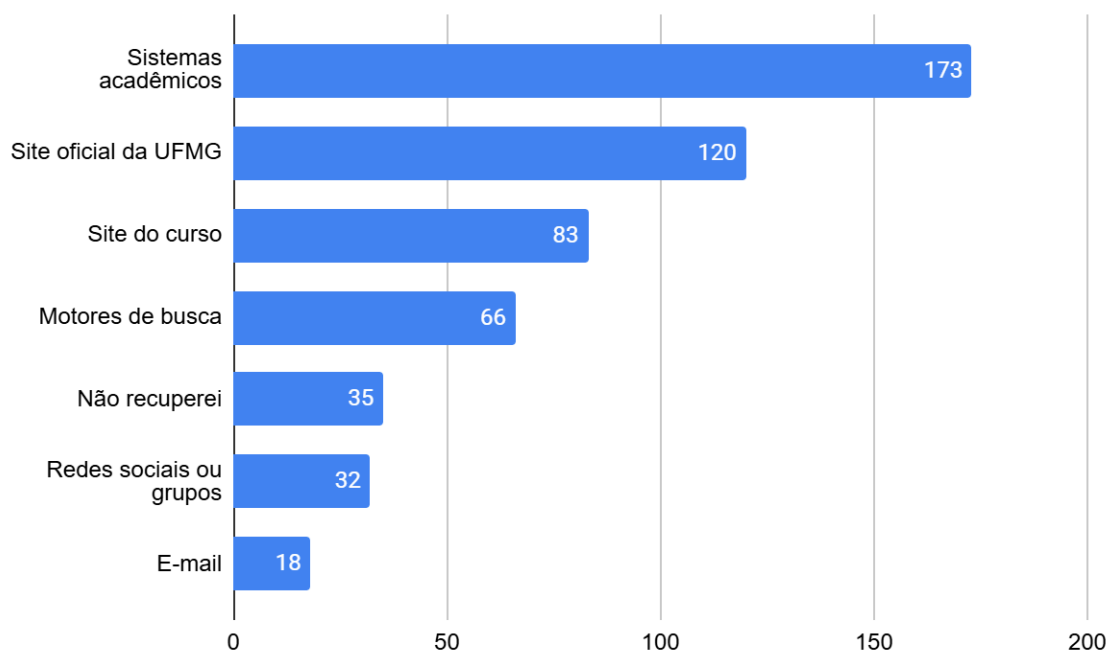
Fonte: dados da pesquisa (2025).

Em relação às formas de acesso, predomina o acesso em ambiente digital, com 88,24% (465) utilizando computadores ou dispositivos móveis. Quanto ao local de acesso, 32,83% (173) indicaram sistemas acadêmicos institucionais (Sistema de Gestão Acadêmica (SIGA), Diário e Moodle), 22,58% (120) acessam o *site* oficial da UFMG e 15,56% (83), o *site* específico do curso.

Em relação ao local de acesso às informações acadêmicas (ambientes e fontes utilizadas), os resultados apontam para a ausência de uma plataforma centralizada de informação, exigindo que os discentes transitem por múltiplos ambientes digitais, conforme apresentado no Gráfico 6.

A heterogeneidade das fontes de informação, como discutido por Savolainen (1995), potencializa a sobrecarga informacional, elevando o risco de os estudantes abandonarem a tarefa ou optarem por caminhos menos confiáveis. Assim, em vez de promover a autonomia, a multiplicidade de canais de comunicação pode criar um cenário de incerteza, que impacta particularmente os ingressantes e aqueles em etapas críticas do percurso acadêmico.

Gráfico 6 - Ambientes e fontes mais acessados

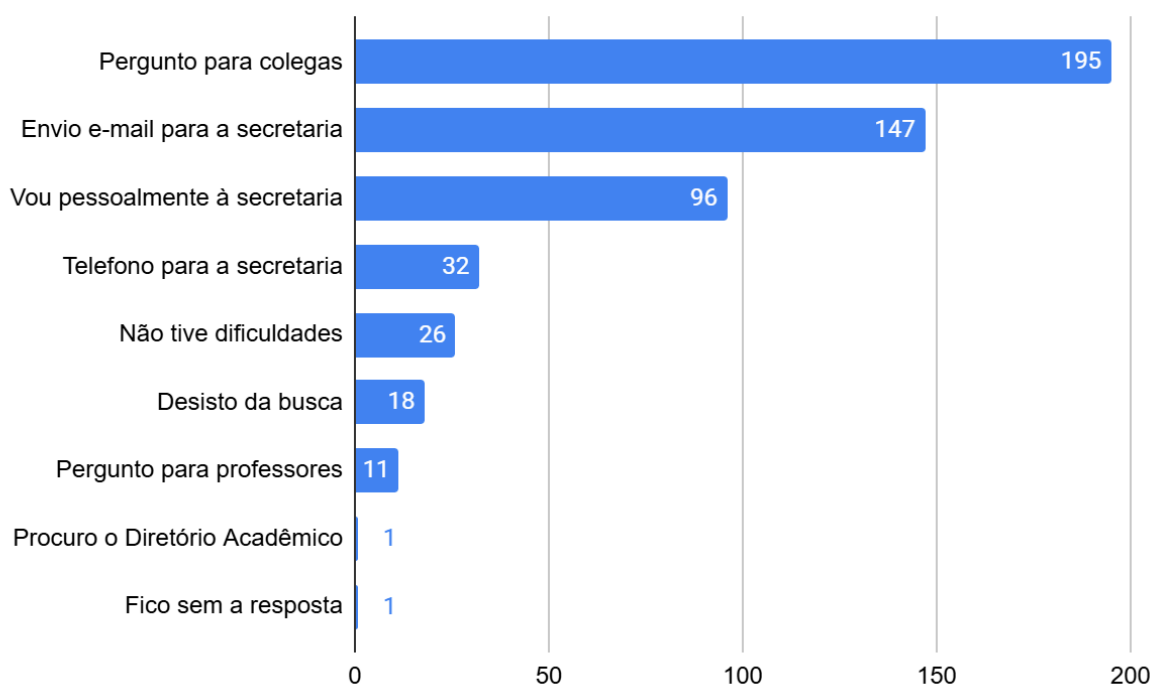


Fonte: dados da pesquisa (2025).

6.3 Seção 2 - Como identificam e qual a percepção sobre as informações acadêmicas

Em situações de dificuldade para encontrar informações acadêmicas, os estudantes priorizam a mediação informal, sendo que 36,04% (195) recorrem a colegas, 27,89% (147) enviam e-mail para secretarias ou centrais institucionais e 18,21% (96) procuram atendimento presencial (Gráfico 7). Perguntar a professores, telefonar ou desistir da busca aparecem com frequência reduzida. Tais dados reforçam o papel das redes sociais informais e a importância da mediação humana diante das falhas ou barreiras institucionais (Savolainen, 1995; Wilson, 1999), sinalizando que a informação oficial, ainda que disponível, nem sempre está facilmente acessível ou inteligível.

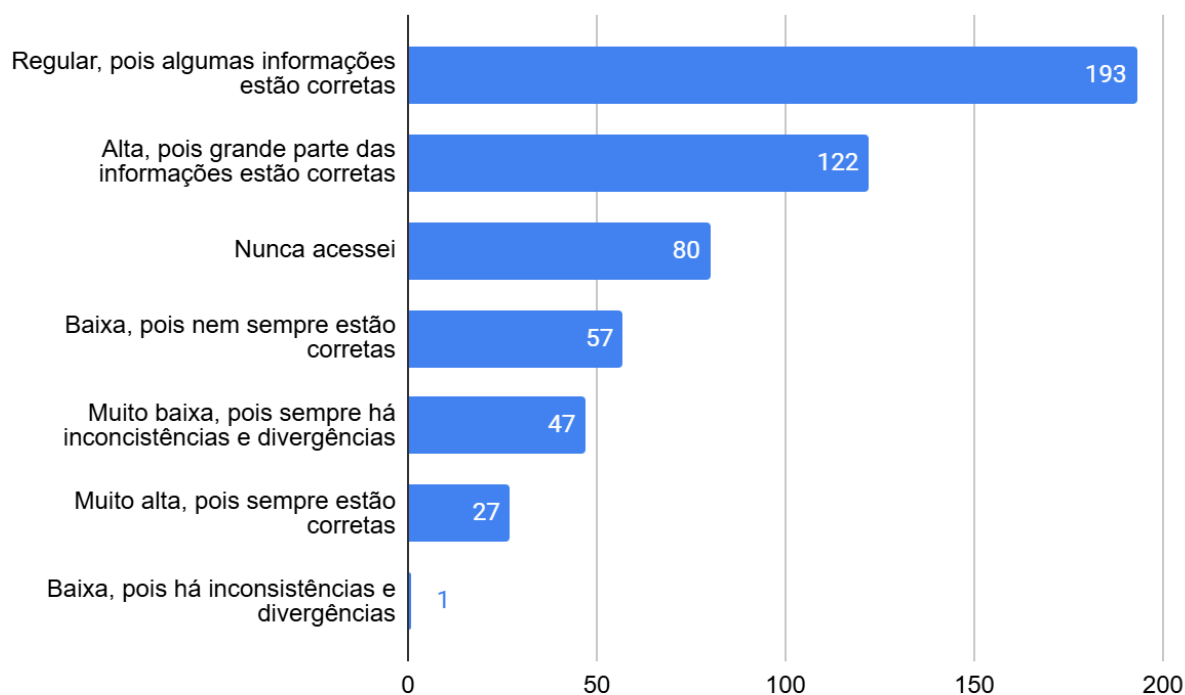
Gráfico 7 - Estratégias diante das dificuldades



Fonte: dados da pesquisa (2025).

Ao avaliarem a confiabilidade das informações obtidas em canais não oficiais, 36,62% (193) dos estudantes classificam como “regular”, 23,15% (122) como “alta” e 15,18% (80) relatam nunca ter utilizado tais fontes (Gráfico 8). Essa distribuição revela cautela na adoção de canais alternativos e também pode evidenciar uma lacuna de letramento informacional para parte do público. A literatura (Wilson, 1999; Savolainen, 1995; Kuhlthau, 2004) destaca que, em ambientes institucionais fragmentados, critérios como autoridade, transparência e consistência são fundamentais para a construção da confiança e que a falta de integração estimula julgamentos compensatórios e busca social.

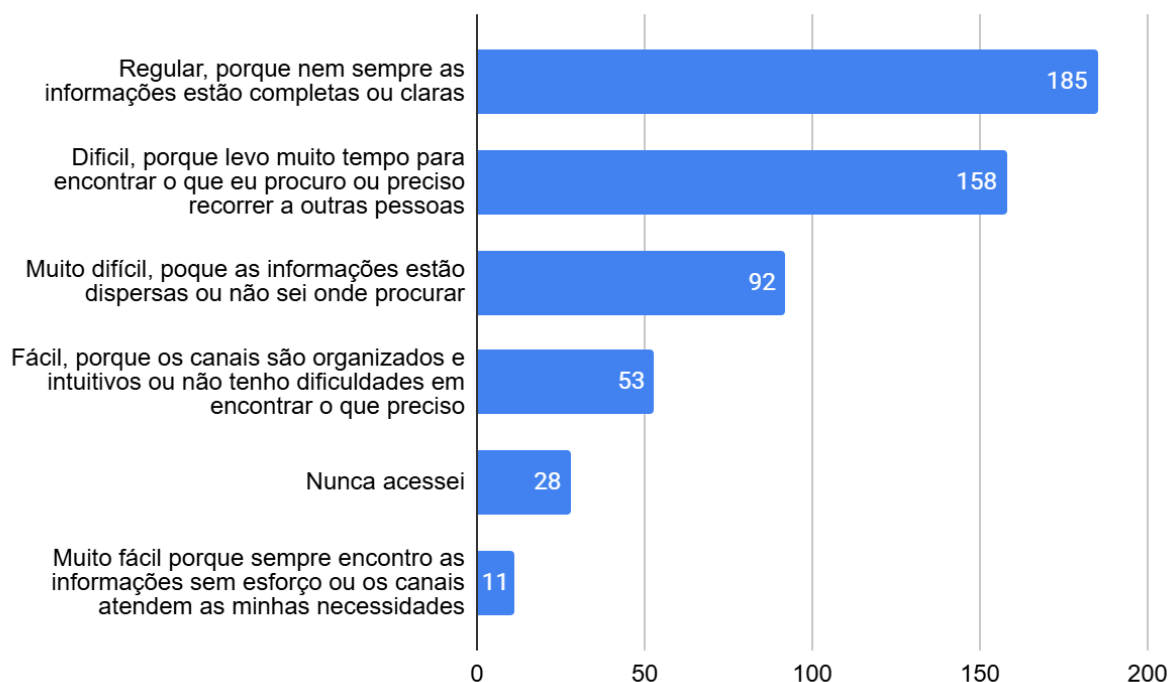
Gráfico 8 - Confiabilidade das fontes não oficiais



Fonte: dados da pesquisa (2025).

Em relação à avaliação do acesso e da recuperação das informações acadêmicas, 35,11% (185) os consideram regulares, destacando que as informações nem sempre são completas ou claras; 29,98% (158) os consideram difíceis, devido ao tempo demandado para encontrar o que procuram ou à necessidade de recorrer a terceiros; e 17,46% (92) os avaliam como muito difíceis, citando dispersão das informações e falta de clareza sobre onde buscá-las. A avaliação geral do acesso e da recuperação de informações acadêmicas reforça o quadro de dificuldades enfrentado pelos estudantes, ou seja, 47,23% avaliam o acesso como difícil ou muito difícil (Gráfico 9). Conforme Wilson (1999), Kuhlthau (2004) e Savolainen (1995), barreiras estruturais e contextuais produzem estratégias de busca compensatórias, insegurança, ansiedade e, muitas vezes, abandono da demanda.

Gráfico 9 - Avaliação geral do acesso



Fonte: dados da pesquisa (2025).

6.4 Seção 3 - Dúvidas, opiniões e sugestões dos estudantes de graduação

As respostas abertas referentes às dúvidas, opiniões e sugestões dos estudantes de graduação foram analisadas qualitativamente por meio da técnica de análise de conteúdo (Bardin, 2016), realizada manualmente pela autora, utilizando planilhas do Excel. Essa abordagem se mostrou necessária devido à variação linguística presente nas respostas, que, muitas vezes, empregavam termos informais ou regionais para conceitos acadêmicos, como, por exemplo, o termo “jubramento”, usado popularmente para se referir ao desligamento do curso. Na graduação, foram extraídas 497 respostas relativas às dúvidas mais recorrentes ou à última dúvida que os estudantes recordam. As categorias mais frequentes identificadas incluem: normas acadêmicas (60 menções), integralização curricular (46), localização de informações institucionais (45), procedimentos de matrícula (42) e prazos acadêmicos (28). Esses tópicos refletem não apenas o desconhecimento do funcionamento normativo básico, mas também dificuldades em acompanhar a própria trajetória acadêmica, localizar fontes oficiais de informação, interpretar

etapas formais do processo de matrícula e compreender as datas-chave do calendário universitário.

Além dessas categorias centrais, surgem dúvidas relacionadas a procedimentos complementares, tais como atividades complementares, estágio, aproveitamento de estudos, trancamento e acesso a plataformas digitais. Esse conjunto revela que, para além das questões estruturais, existe uma demanda por orientação prática contínua, especialmente em situações específicas ao longo do percurso acadêmico.

Do ponto de vista teórico, esse padrão de respostas reforça o entendimento de que o contexto informacional percebido e as barreiras presentes no sistema institucional moldam estratégias diversificadas de busca (Savolainen, 1995). Muitas vezes, diante de fragmentação ou baixa transparência das fontes oficiais, estudantes recorrem a mediações alternativas ou buscam apoio em redes informais. A categorização das principais dúvidas está apresentada no Quadro 7, com sua distribuição visual detalhada no Gráfico 10.

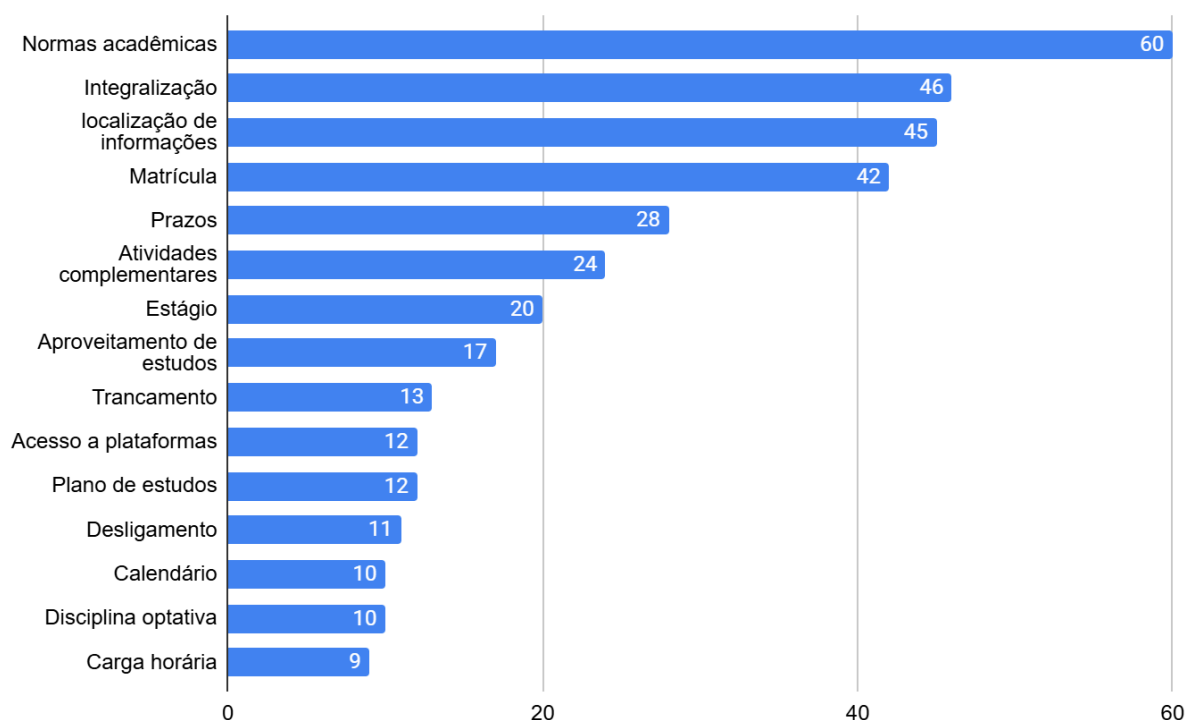
Quadro 7 - Lista de categorias identificadas nas dúvidas dos estudantes

Categorias	Quantidade
Normas Acadêmicas	60
Integralização	46
Localização de Informações	45
Matrícula	42
Prazos	28
Atividades Complementares	24
Estágio	20
Aproveitamento de Estudos	17
Trancamento	13
Acesso a Plataformas	12
Plano de Estudos	12
Desligamento	11
Calendário	10
Disciplina Optativa	10
Carga Horária	9
Documentos e Formulários	8
Reopção	8
Grade Curricular	7
Intercâmbio	6

Regime Especial	6
Trancamento Parcial de Matrícula	6
Colaão de Grau	5

Fonte: dados da pesquisa (2025).

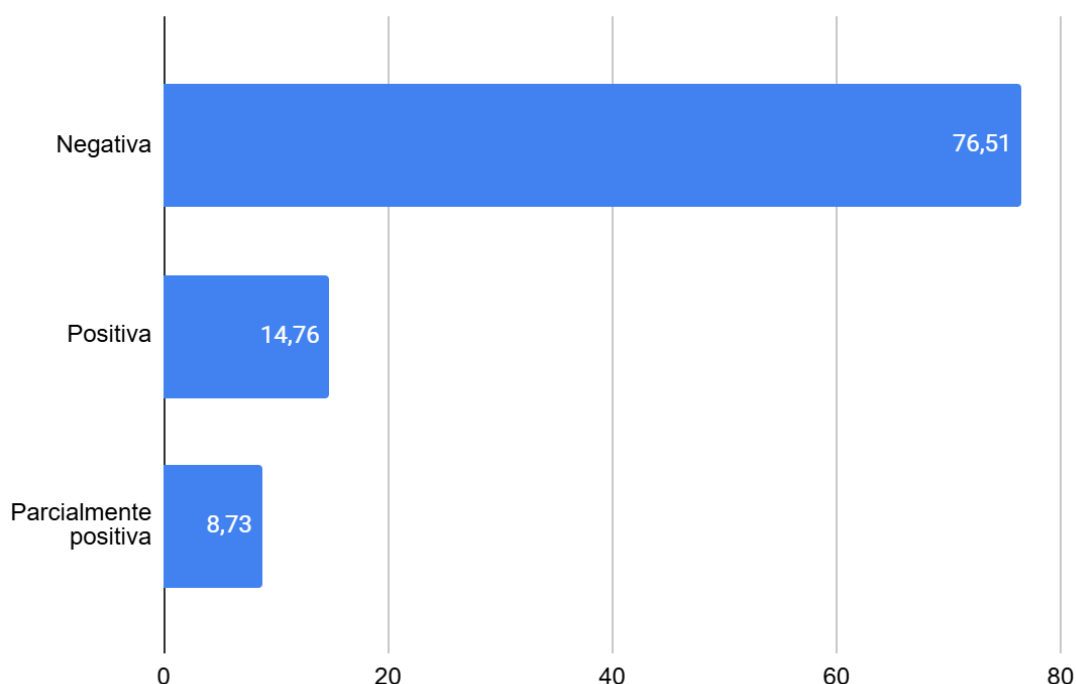
Gráfico 10 - Principais dúvidas apresentadas



Fonte: dados da pesquisa (2025).

Das 481 respostas válidas referentes à percepção sobre a acessibilidade e organização das normas acadêmicas, 76,51% (368) apresentaram avaliação negativa, 14,76% (71), positiva e 8,73% (42), parcial (parcialmente positiva), conforme apresentado no Gráfico 11. Este predomínio das avaliações negativas indica uma insatisfação significativa com a forma como as normas são disponibilizadas e organizadas para os graduandos. As críticas recorrentes destacam a dispersão das fontes, a ausência de padronização, o uso de linguagem excessivamente técnica e dificuldades de usabilidade nos sistemas institucionais.

Gráfico 11 - Percepção sobre acessibilidade



Fonte: dados da pesquisa (2025).

As respostas abertas aprofundaram esse diagnóstico, trazendo exemplos de experiências e percepções recorrentes entre os discentes:

Não. São dispersas, em linguagem complexa e em geral é muito mais fácil conversar com veteranos do que procurar por conta própria nos sites da ufmg (Respondente 1).

[...] elas precisam ser acessíveis e bem organizadas, pelo fato variadas pessoas acessarem, de diferentes pensamentos, uma forma mais fácil certamente agrega todos que acessam e que vão acessar (Respondente 2).

Não, pois não sei onde se encontram nem possuo conhecimento para interpretar, assim como quais normas o meu curso segue, que diversas situações por vezes são resolvidas em conversas de corredor (Respondente 3).

De modo geral, os dados evidenciam que a fragmentação das fontes institucionais e a complexidade normativa do ambiente universitário impactam negativamente a experiência do estudante, indicando a percepção de dificuldade no acesso e na recuperação das informações acadêmicas.

Em relação a sugestões para o aprimoramento do acesso e da recuperação das informações acadêmicas, na questão: “Você tem alguma sugestão para

aprimorar o acesso e a recuperação de informações acadêmicas (normas, regimentos, resoluções) da UFMG?”, a pergunta teve o objetivo de coletar propostas dos estudantes para a melhoria dos sistemas e canais institucionais. Assim, das 356 respostas válidas recebidas, três grandes categorias de sugestões se destacaram, evidenciando diferentes dimensões.

A principal demanda, expressa por 56,74% dos participantes, refere-se à necessidade de melhorias nas plataformas institucionais. As sugestões concentram-se na integração, centralização e padronização dos sistemas, podendo indicar que a dispersão informacional prejudica o acesso eficiente aos conteúdos acadêmicos e reforçando a importância de ambientes digitais mais unificados e intuitivos.

A segunda categoria, apontada por 17,98% das respostas, diz respeito à organização e localização das informações. Entre as propostas, estão a criação de portais unificados, menus temáticos, mecanismos de busca mais eficientes e o agrupamento dos documentos normativos por assunto ou departamento. Tais recomendações indicam a preocupação dos estudantes com a navegabilidade e a facilidade de acesso direcionado às informações relevantes.

Essas duas frentes de sugestões evidenciam que os próprios estudantes reconhecem os principais obstáculos e apontam caminhos concretos para a evolução dos sistemas acadêmicos da UFMG. O alinhamento dessas propostas com tendências tecnológicas e com o foco no usuário reforçam a necessidade de estratégias institucionais que priorizem a integração, a acessibilidade e a personalização dos serviços informacionais. A seguir, algumas sugestões destacadas pelos participantes.

Ter um sistema unificado entre os colegiados e enviar todas as informações, normas, regimentos etc no e-mail do aluno assim que ele entra e depois periodicamente (Respondente 1).

Focar tudo em um mesmo site, com normas/regimentos e resoluções de todos os prédios/departamentos da UFMG (Respondente 2).

Chat para tirar dúvidas, manuais com mais instruções, um *site* só para essas resoluções organizado por assunto/departamento mas tudo em um único site para facilitar (Respondente 3).

Proponho a criação de um assistente virtual, tipo *chatbot*, que pudesse responder dúvidas acadêmicas relacionadas a normas e regulamentos (Respondente 4).

A dispersão das fontes e a falta de padronização nos canais institucionais contribuem para uma experiência de busca fragmentada e desmotivadora. Bates (2017) e Case e Given (2016) destacam que a fragmentação informacional é um dos principais desafios da gestão acadêmica, sobretudo em universidades que operam com múltiplas plataformas não conectadas, tornando o acesso às informações mais dispendioso e pouco eficiente. Outro ponto relevante é a dependência dos estudantes de fontes informais, como colegas e secretarias acadêmicas, para suprir dificuldades na localização autônoma das informações. Wilson (1999) explica que, diante de barreiras de acesso, os indivíduos tendem a recorrer a redes próximas. Savolainen (2007) complementa que redes sociais informais podem atuar como sistemas de apoio informacional em ambientes institucionais complexos, o que se confirma no contexto observado.

Com relação às avaliações negativas, indicada por 76% dos respondentes, sobre a organização das normas acadêmicas, elas reforçam a necessidade de repensar os canais institucionais, sobretudo pela dispersão de conteúdos e pela linguagem excessivamente técnica. A ausência de uma mediação sistemática faz com que colegas, grupos de mensagens e secretarias assumam o papel de mediadores informais, suprimindo lacunas deixadas pelos sistemas oficiais. Nesse contexto, a curadoria informacional assume papel estratégico como mediação implícita, ao organizar e contextualizar previamente conteúdos normativos de forma acessível e confiável (Maia; Santos Neto, 2024; Jorente; Padua; Nakano, 2019).

Tendo em vista a proposta desta dissertação em relação à construção de uma base de conhecimento, os resultados desta pesquisa de opinião subsidiam a construção de um *chatbot* ao emergirem como alternativa concreta de mediação, identificando as necessidades do público a ser atendido: os estudantes. O estudante poderá interagir diretamente com o sistema, mas só receberá respostas úteis porque, anteriormente, ocorreu o trabalho de curadoria, que selecionou, organizou e traduziu a linguagem normativa para o vocabulário cotidiano. A literatura reforça que a efetividade da recuperação depende da integração entre organização, representação e recuperação da informação (Ferneda, 2012, 2021; Lima, 2020), assegurando clareza conceitual e precisão semântica nas respostas automatizadas.

Por fim, as sugestões dos próprios estudantes, que demandam centralização de plataformas, padronização dos documentos e criação de assistentes virtuais,

indicam a urgência de soluções inovadoras. A base de conhecimento proposta nesta pesquisa, estruturada a partir da curadoria informacional, poderá responder a essas demandas, ao mesmo tempo em que contribuirá para fortalecer a mediação institucional e reduzir os riscos de desinformação, ampliando o acesso autônomo às informações normativas da UFMG.

7. CHATBOTS EDUCACIONAIS

Esta seção corresponde à Etapa 3 da pesquisa, dedicada à análise de boas práticas no desenvolvimento de *chatbots* educacionais. Seu objetivo é examinar experiências no contexto acadêmico e identificar elementos que orientem a construção da base informacional do *chatbot* no estudo de caso desta dissertação.

Com base na revisão de literatura ([Seção 5.1](#)) dos 26 estudos identificados, foram selecionados cinco projetos com alto potencial de replicação, segundo três critérios: (i) metodologia explícita (etapas de desenvolvimento bem descritas ou uso de *frameworks* reconhecidos); (ii) fluxo de informação estruturado; e (iii) disponibilidade/acessibilidade de ferramentas, dados e documentação. Os estudos selecionados, alinhados a esses critérios, estão sintetizados no Quadro 8.

Quadro 8 - Estudos selecionados

Estudo	Título	Autores	País	Perfil / usuário	Metodologia explícita	Estrutura de fluxo informacional	Ferramentas e documentação
D14	Introducing a Chatbot to the Web Portal of a Higher Education Institution to Enhance Student Interaction	Oliveira e Matos (2023)	Portugal	Estudantes de graduação	Passo-a-passo de indexação	Documento → <i>embedding</i> → busca	Código exemplificado Python, Streamlit e LangChain
D16	Development of a Class Materials Search System using LINE Chatbot	Thaiprasert <i>et al.</i> (2023)	Tailândia	Estudantes Eng. Elétrica	Pipeline PLN + métricas RI	<i>Intent-match</i> → BD materiais	LINEAPI, Dialogflow ES, Firebase, fórmulas TF IDF e cosseno, código não anexado e passos detalhados
D19	Automated information retrieval and services of graduate school using chatbot system	Rukhiran e Netinant (2022)	Tailândia	Candidatos, alunos e egressos	RAD + TAM detalhado	<i>Intent</i> → GoogleSheets/SIS → resposta	LINE Official Account, Dialogflow, Google Forms; exemplos de JSON e menus.
D34	Developing UNYSA Chatbot as Information Services about Yogyakarta State University	Muhyidin e Setiawan (2021)	Indonésia	Público externo e alunos	Scrum + ISO/IEC25010	Agente NLU → <i>scraping/spreadsheet</i> → resposta	Dialogflow, Flask, Google Cloud, Dataset de cenários (395) e repositório público
D49	A Model to Develop Chatbots for Assisting the Teaching and Learning Process	Mendoza <i>et al.</i> (2022)	México	Professores, alunos e equipe	Google Design <i>print</i> , personas, fluxograma de intents e avaliação UEQ.	<i>Role-based intents</i> + BD relacional/NoSQL	Código GitHub, <i>stack</i> Angular–Dialogflow-Fi rebase e <i>scripts</i> SQL

Fonte: elaborado pela autora (2025).

7.1 Resultados

Em uma comparação realizada, foi identificado que todos os estudos selecionados documentam etapas, mas os estudos D19, D34, D49 incluem ciclos de avaliação de usuários, que faltam nos estudos D14 e D16. Os documentos D34 e D49 apresentam fluxogramas e diagramas de arquitetura completos. A seguir, os estudos são descritos com destaque para as fontes de informação utilizadas, bem como a estrutura do fluxo informacional elaborado.

O estudo D14, de Oliveira e Matos (2023), intitulado Introduzindo um Chatbot na Plataforma Web de uma Instituição de Ensino Superior para Melhorar o Envolvimento do Aluno, delineou a integração de um *chatbot* na plataforma *web* de uma instituição de ensino superior para melhorar a interação e os serviços de suporte ao aluno. O tratamento de dados e informações ocorreu por meio de análise semântica e estágios de recuperação. Foram apresentados códigos em Python para cada etapa e a seguinte estrutura de fluxo:

2. Configuração: preparo do ambiente Streamlit com as bibliotecas necessárias.
3. Carregamento: importação de documentos de texto como base de conhecimento.
4. Preparação: segmentação dos documentos em unidades menores.
5. Representação: conversão dos segmentos em embeddings semânticos.
6. Indexação: armazenamento e otimização dos *embeddings* para uma pesquisa eficiente no Pinecone.
7. Interface: desenvolvimento de uma interface de bate-papo interativa usando o Streamlit.
8. Processamento: incorporação de um modelo de linguagem para lidar com consultas e gerar respostas com base na pesquisa semântica dentro da base de conhecimento.

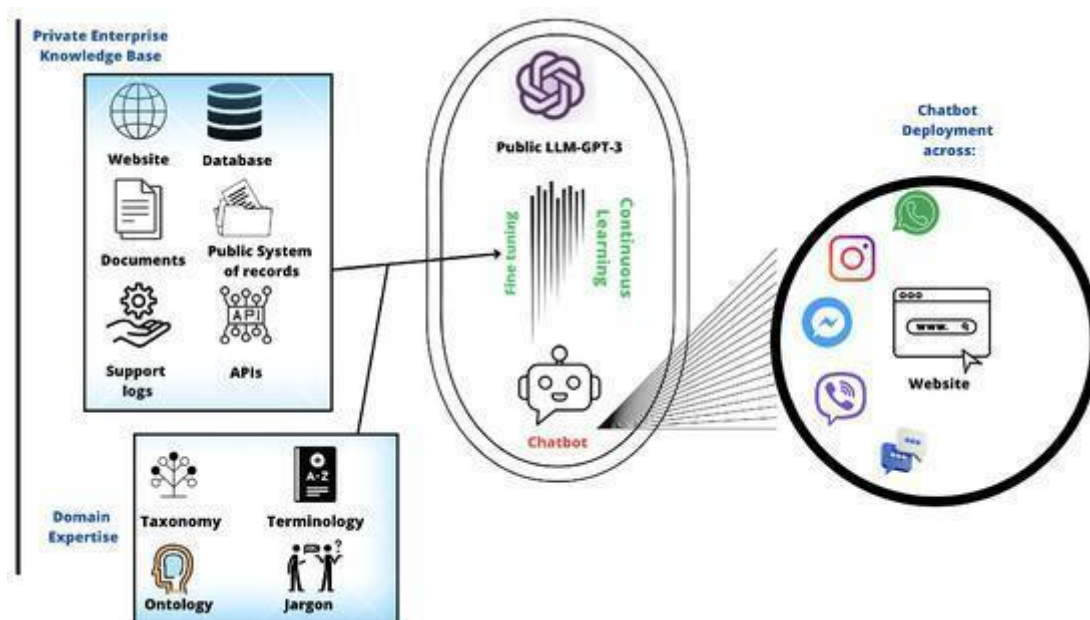
O *chatbot* combina uma base de conhecimento textual com uma interface de usuário interativa, empregando técnicas de PLN e ferramentas. O processo abrange desde a preparação dos dados até a interação com o usuário, englobando os seguintes estágios principais: inicialmente, a configuração do ambiente de desenvolvimento do Streamlit, seguida da utilização das bibliotecas necessárias

para a criação da interface da *web* e, por fim, a integração com outras ferramentas de PLN.

Com relação aos dados utilizados, eles são resultantes de documentos de texto carregados de um diretório específico, conforme descrito no fluxo de trabalho. O estudo não especifica exatamente a origem desses documentos, mas é possível supor, pelo contexto da informação no texto do estudo, que eles são fornecidos pela própria instituição de ensino como regulamentos acadêmicos, materiais de aula, FAQs etc.

Posteriormente, uma base de conhecimento é estabelecida com o carregamento de documentos de texto de um diretório específico. Para aprimorar o processamento e a representação semântica, esses documentos são divididos em unidades menores. Os modelos de linguagem pré-treinados são usados para criar *embeddings* de vetores para cada segmento de texto, aprendendo o significado semântico das informações. Essas incorporações são, então, indexadas em um serviço de pesquisa vetorial eficiente (Pinecone), permitindo a recuperação rápida de informações semanticamente semelhantes para consultas futuras.

Figura 9 - Chatbot - arquitetura



Fonte: Oliveira e Matos (2023).

A Figura 9 apresenta a arquitetura do sistema de *chatbot* descrito por Oliveira e Matos (2023), estruturada em três módulos interconectados. O primeiro módulo, de carregamento e indexação de documentos, utiliza o Directory Loader para agregar arquivos de texto a partir de um diretório. Esses documentos são, então, fragmentados em "*chunks*", por meio do Recursive Character Text Splitter, e convertidos em *embeddings* semânticos, com o modelo all-MiniLM-L6-v2. Os vetores gerados são armazenados em um banco vetorial Pinecone, viabilizando consultas por similaridade.

O segundo módulo, de recuperação e geração de respostas, recebe a consulta do usuário, busca os trechos mais relevantes no índice vetorial e os envia a um modelo de linguagem, que sintetiza a resposta em linguagem natural.

Por fim, o terceiro módulo, de interface, é implementado em Streamlit, com o componente *streamlit-chat*, e é responsável por coordenar todo o fluxo de interação, do recebimento da pergunta até a exibição da resposta ao usuário.

O estudo D16, de Thaiprasert *et al.* (2023), intitulado: Development of a Class Materials Search System using LINE Chatbot, apresentou um *chatbot* desenvolvido para localizar recursos educacionais por meio da plataforma LINE³, com o objetivo de aprimorar o autoaprendizado entre os estudantes de Engenharia, inicialmente, com foco em duas disciplinas: Sinais e Sistemas, e Sistemas de Controle. Ele gira em torno de uma abordagem de tratamento de informações centrada na interação e colaboração do usuário para a criação e recuperação de conteúdo.

A base de dados foi construída especificamente para o sistema de busca de materiais de aula. Os dados (materiais de aula) não vêm de fontes públicas preexistentes; o estudo menciona somente que não há um *dataset* público para esse tipo de sistema. Em vez disso, o *chatbot* coleta e armazena materiais sugeridos pelos próprios usuários (no caso, estudantes e professores). Esses materiais são sugestões dos usuários sobre o conteúdo relevante, como textos, vídeos ou *blogs*, que são então armazenados na base de dados do sistema.

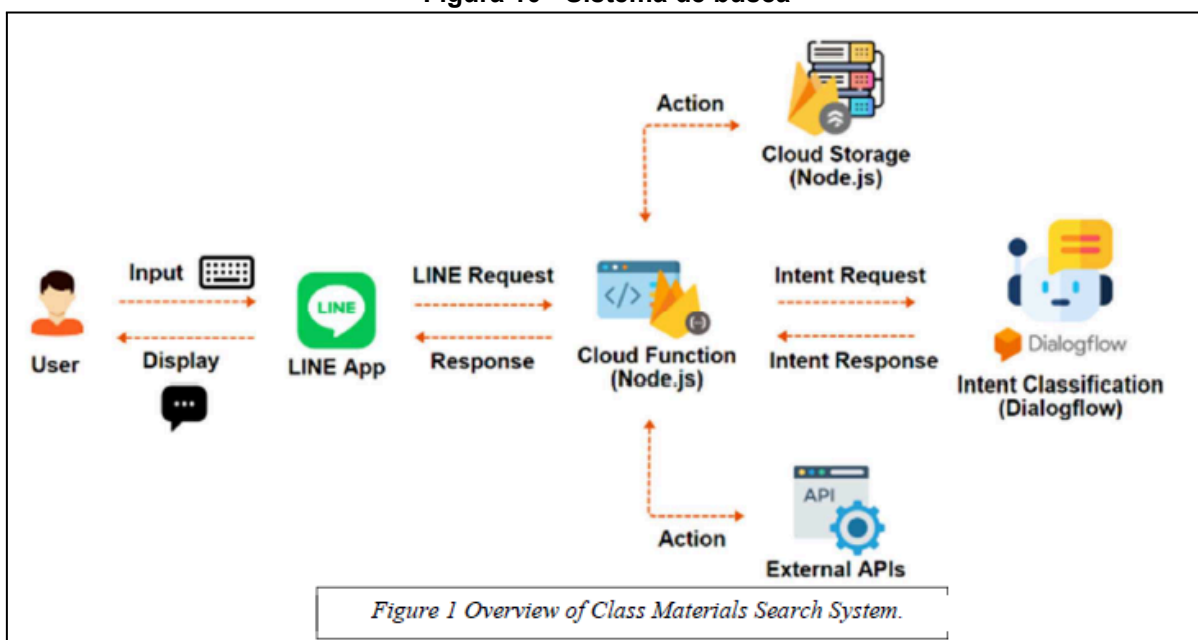
A arquitetura do sistema que incorpora o *chatbot* no aplicativo LINE, o Dialogflow ES e o Firebase (Firestore), foi criada com um fluxo bidirecional. O *design* desse sistema é segmentado em três componentes principais: classificação de intenção, banco de dados baseado em nuvem e servidor de *backend*. O

³ LINE: plataforma de mensagens instantâneas e chamadas de voz/vídeo — mais popular na Ásia.

processamento de informações se baseia em uma arquitetura que se apoia na classificação de intenções do Dialogflow ES. A seguir, a estrutura do fluxo (intenções - processamento e indexação) é apresentada:

1. Intenções (Dialogflow ES): o Dialogflow decodifica a fala do usuário, categorizando a intenção como sugestão de material (“Sugerir ao TAE”) ou solicitação de material (“Sugerir ao usuário”), juntamente com as intenções padrão. As entidades (curso, conteúdo, tipo de fonte e URL) são extraídas.
2. Processamento (*back-end*):
 - Sugestão: permite que usuários contribuam com novos materiais relevantes, enriquecendo a base de dados para futuras buscas. O *backend* recebe dados de sugestões e os armazena no Firestore.
 - Solicitação: permite que usuários busquem materiais específicos. O *backend* consulta o Firestore usando as entidades fornecidas, recuperando materiais relevantes, classificados por uma “pontuação média de utilidade”. O material é enviado ao usuário, que é incentivado a avaliá-lo. A avaliação atualiza a pontuação no Firestore.
3. Indexação (Firestore): o Firestore armazena os metadados do material (incluindo entidades e pontuação). A recuperação e a pesquisa são realizadas por meio de consultas diretas a esses metadados, com a pontuação atuando como um critério de classificação para priorizar materiais altamente avaliados.

Figura 10 - Sistema de busca



Fonte: Thaiprasert *et al.* (2023).

A Figura 10, de Thaiprasert *et al.* (2023), ilustra o fluxo de funcionamento de um sistema de busca de materiais de aula integrado ao aplicativo LINE, estruturado em três camadas principais. Quando o estudante envia uma mensagem por meio do LINE, a requisição é encaminhada a uma função em nuvem, que atua como *backend*. Essa função envia o conteúdo textual ao Dialogflow ES, responsável por identificar a intenção do usuário (*intent*) e extrair os parâmetros necessários. A partir do *intent* detectado, o sistema executa uma das duas ações possíveis: (i) registra uma nova sugestão de material no banco de dados Firestore, incluindo informações como curso, tipo de fonte e URL; ou (ii) recupera, do mesmo banco de dados, o recurso mais bem avaliado relacionado à solicitação, envia-o ao usuário e solicita uma avaliação de 1 a 5, a fim de atualizar o *ranking* de relevância do material. Por fim, o *backend* processa e retorna a resposta final ao aplicativo LINE, concluindo o ciclo de interação.

O estudo D19, de Rukhiran e Netinant (2022), intitulado Automated Information Retrieval and Services of Graduate School using Chatbot System, descreve o desenvolvimento de um *chatbot* integrado à plataforma de mensagens LINE, com o objetivo de automatizar a recuperação de informações e serviços para estudantes de pós-graduação. O sistema utiliza o Dialogflow para interpretar a linguagem natural do usuário e identificar suas intenções, permitindo o acesso a

informações predefinidas armazenadas em diversas fontes. Essa abordagem integra diferentes sistemas, como Google Forms e servidores *web* da Universidade, para fornecer respostas e serviços personalizados. A estrutura do fluxo é organizada da seguinte forma:

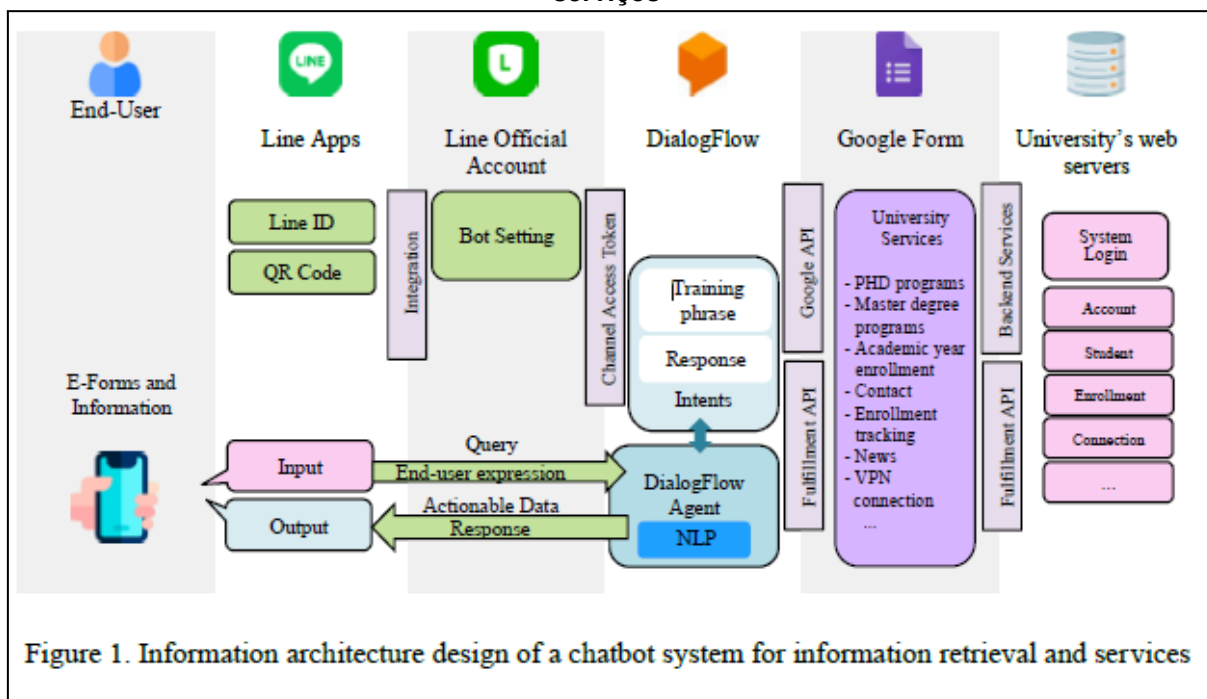
1. Entrada do usuário (LINE): o processo começa quando o usuário envia uma mensagem de texto ou escolhe uma opção no "*rich menu*", dentro da aplicação LINE. O *rich menu* oferece acesso direto a serviços específicos, como inscrição em cursos e acesso ao *site* da Universidade.
2. Classificação de intenções (Dialogflow): a mensagem do usuário é enviada para o Dialogflow, que utiliza técnicas de PLN e *training phrases* para identificar a intenção da mensagem. O sistema compara a entrada do usuário com as intenções previamente configuradas, como busca de informações ou ações específicas.
3. Extração de entidades (Dialogflow): além da intenção, o Dialogflow extrai entidades relevantes da mensagem do usuário, como palavras-chave relacionadas a cursos, programas acadêmicos, documentos e outros parâmetros necessários para fornecer a resposta adequada.
4. Recuperação da informação (fontes): dependendo da intenção identificada, o *chatbot* acessa informações de diversas fontes. O DialogFlow foi configurado para interagir com essas fontes externas por meio de APIs de integração. O artigo menciona a Fulfillment API para se conectar aos servidores da Universidade e Google API para acessar os dados no Google Forms.
 - Google Forms: utilizado como um banco de dados em nuvem para armazenar informações sobre cursos, anos acadêmicos, requisitos de documentos e dados de contato. O acesso aos dados é feito por meio da Google API e um *token* de autenticação.
 - Servidores *web* da Universidade: o Dialogflow interage com os servidores *web* da Universidade por meio de uma Fulfillment API, permitindo realizar ações como *login* de estudantes e recuperar informações específicas relacionadas a cada aluno.

- Respostas predefinidas no Dialogflow: o sistema possui diálogos de resposta predefinidos para perguntas frequentes (FAQ) e informações de contato, associados a intenções específicas.
5. Formatação da resposta (Fulfillment API e JSON): após recuperar a informação, especialmente via Fulfillment API, os dados são formatados usando a linguagem JSON, que é o formato preferido para a interface do *chatbot*. O JSON permite estruturar as respostas com texto, números e formatação simples (usando HyperText Markup Language (HTML) para cores, embora de forma limitada).
 6. Apresentação da resposta (LINE): a resposta formatada é enviada de volta ao usuário como uma mensagem na aplicação LINE, simulando uma comunicação bidirecional entre o *chatbot* e o usuário.

A Figura 11, de Rukhiran e Netinant (2022), apresenta a arquitetura de um sistema de *chatbot* voltado à recuperação de informações e serviços, organizado em três camadas principais: interface com o usuário, processamento da linguagem natural e integração com sistemas institucionais. O fluxo tem início com a interação do usuário por meio do aplicativo LINE, que encaminha a mensagem ao agente DialogFlow por meio de um *token* de acesso. Nessa etapa, o NLU identifica a intenção do usuário (*intent*) e extrai as entidades relevantes. Com base na *intent* detectada, a Fulfillment API direciona a requisição para o serviço apropriado, que pode incluir formulários Google (com informações sobre cursos e contatos) ou os servidores *web* da Universidade (para acesso a dados acadêmicos e autenticação).

Por fim, os dados obtidos são organizados no DialogFlow e enviados ao LINE, completando o ciclo de resposta em linguagem natural.

Figura 11 - Arquitetura de um sistema de *chatbot* voltado à recuperação de informações e serviços



Fonte: Rukhiran e Netinant (2022).

O estudo D34, de Muhyidin e Setiawan (2021), intitulado Developing UNYSA Chatbot as Information Services about Yogyakarta State University, detalha o desenvolvimento e a implementação do *chatbot* UNYSA (UNY Smart Assistant) para fornecer serviços de informação na Yogyakarta State University. O tratamento da informação é realizado por meio do Dialogflow, que interpreta a linguagem natural dos usuários, permitindo ao sistema acessar e apresentar informações das fontes. O estudo apresenta duas fontes de dados principais para recuperar as informações: Web Scraping, do *site* oficial da UNY, e Google Sheets, com uma arquitetura que facilita a integração com várias plataformas de *frontend*. O fluxo de interação do sistema é estruturado da seguinte forma:

1. Entrada do usuário (múltiplas plataformas): o usuário interage com o *chatbot* por meio de diversas plataformas de *frontend*, como Google Assistant, WhatsApp, Telegram, LINE e Facebook Messenger, enviando entradas em formato de texto.
2. PLN (Dialogflow): a mensagem do usuário é encaminhada para o Dialogflow, que atua como o motor principal de NLU. Dentro do Dialogflow, são definidos:

- Agentes: módulos que realizam todo o processamento da linguagem natural.
 - Entidades: usadas para identificar o significado de termos específicos nas perguntas dos usuários (como "acadêmicos", "registro", "bolsas" e "programas da UNY"). As entidades são, principalmente, "developer entities", ou seja, definidas manualmente.
 - *Intents*: mapeiam as perguntas dos usuários para respostas específicas, baseadas em tópicos comuns (ex: "acadêmicos" e "programas de estudo").
 - *Training phrases*: exemplos de perguntas associadas a *intents*, que ajudam a identificar as entidades presentes nas consultas.
 - Parâmetros: ajustados com base nas entidades extraídas das *training phrases*, exigindo que o usuário forneça palavras-chave para receber respostas precisas.
3. RI (Web Scraping e Google Sheets): para gerar as respostas, o *chatbot* utiliza duas principais fontes de dados:
- Web Scraping, do *website* oficial da UNY: informações relevantes são extraídas diretamente do *site* da Universidade, permitindo acesso a conteúdos atualizados.
 - Google Sheets: uma planilha é utilizada como fonte de dados para algumas respostas. As informações são alimentadas manualmente.

A integração é feita utilizando o Flask, um micro *framework* Python que gerencia a coleta dos dados de ambas as fontes e os formata para serem enviados ao Dialogflow para a elaboração da resposta ao usuário. As planilhas são alimentadas manualmente, com informações sobre cursos, inscrições ou outros dados administrativos. O estudo não especifica detalhadamente o processo de alimentação dessas planilhas.

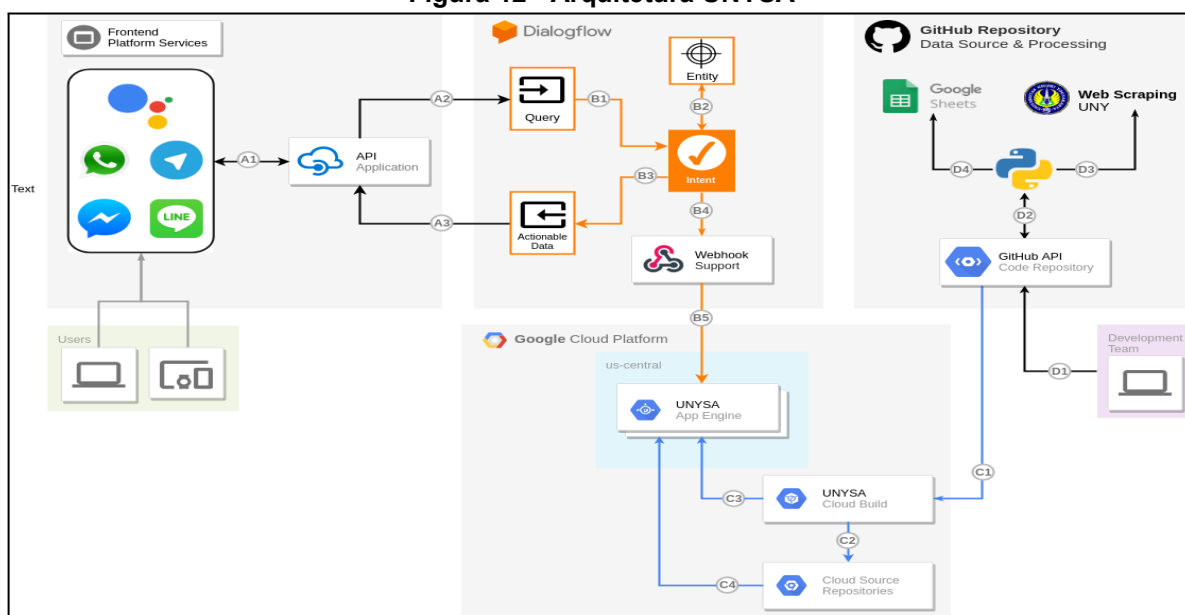
4. Gerenciamento da resposta (Flask): o Flask, um micro *framework* baseado em Python, é utilizado para gerenciar as respostas do Dialogflow. O Flask é responsável por:
- Coletar dados das fontes de informações, como *web scraping* e Google Sheets.

- Formatar as respostas que serão enviadas de volta ao Dialogflow.
 - Integrar as diferentes fontes de dados para construir respostas coesas e completas.
5. Fulfillment responses (Dialogflow para *frontend*): o Dialogflow formata a resposta final para ser exibida na plataforma de *frontend* utilizada pelo usuário. O estudo menciona diferentes tipos de respostas, como simples, mídia, listas e respostas básicas, com o Google Assistant suportando todos os tipos, enquanto outras plataformas podem ter limitações. As *fulfillment responses* são as respostas acionadas quando uma *intent* é identificada.
 6. Implantação (Google Cloud App Engine): o código de *fulfillment*, que define a lógica de como o *chatbot* busca e formata as respostas, é implantado no Google Cloud App Engine, garantindo a escalabilidade e acessibilidade do sistema.

Na Figura 12, de Muhyidin e Setiawan (2021), é apresentada a arquitetura do UNYSA Chatbot, estruturada em três camadas integradas: apresentação, processamento e entrega. Na camada de apresentação, os estudantes interagem por meio de diversos canais, como WhatsApp, Telegram, LINE, Facebook Messenger e Google Assistant, enviando mensagens que são encaminhadas ao agente do *chatbot*. Na camada de processamento, o Dialogflow realiza a interpretação da linguagem natural, identificando as *intents* e extraindo as entidades relevantes. Em seguida, a requisição é direcionada ao componente de *fulfillment*, implementado com Flask. Conforme a *intent* identificada, o sistema realiza uma de duas ações: consulta dados estáticos em uma planilha do Google Sheets ou executa *scraping* nos servidores da Universidade para recuperar informações dinâmicas. Os dados coletados são estruturados em formato JSON e retornados ao Dialogflow.

Por fim, na camada de entrega, o Dialogflow formata a resposta em texto ou em *cards* e a envia de volta ao canal de origem, concluindo o ciclo de interação.

Figura 12 - Arquitetura UNYSA



Fonte: Muhyidin e Setiawan (2021).

O estudo D49, de Mendoza *et al.* (2022), intitulado *A Model to Develop Chatbots for Assisting the Teaching and Learning Process*, descreve um modelo para o desenvolvimento de *chatbots* educacionais, com foco no suporte ao processo de ensino e aprendizagem. O tratamento da informação no sistema é centrado na interação em linguagem natural, permitindo que o *chatbot* compreenda e responda às solicitações dos usuários. Além disso, o modelo organiza a gestão de informações específicas para diferentes personas (alunos, professores ou equipe administrativa), atendendo às suas necessidades e aos seus papéis distintos dentro do ambiente educacional. O fluxo é estruturado da seguinte forma:

1. Entrada do usuário: o modelo oferece suporte para interação por meio de texto e voz, garantindo flexibilidade na comunicação, independentemente da familiaridade do usuário com interfaces digitais.
2. Definição de *roles* (papéis): o *chat* é projetado para identificar três papéis principais:
 - Professor: o professor pode criar materiais extras, gerar anúncios de eventos, definir lembretes, receber trabalhos de casa e avaliar relatórios de desempenho dos alunos.

- Aluno: o aluno tem acesso a materiais extras, recebe anúncios e lembretes de eventos, pode enviar trabalhos de casa e interagir com outros membros da comunidade escolar.
 - Equipe administrativa: a equipe administrativa, incluindo psicólogos, pedagogos e administradores de sistema, tem a capacidade de gerenciar perfis de usuário, grupos e cursos, gerar anúncios de eventos, receber alertas sobre alunos com problemas e acessar relatórios de desempenho dos alunos.
3. PLN (Dialogflow): o Dialogflow é utilizado para processar e entender a linguagem natural dos usuários. O sistema identifica:
- Intenção do usuário: determina o que o usuário deseja realizar, como consultar prazos, enviar um trabalho ou acessar material de aula.
 - Entidades relevantes: o sistema extrai informações específicas da consulta, como nome do curso, tópico, data e nome do aluno, para garantir respostas precisas e contextualizadas.
4. Gestão da informação: essa configuração é personalizada para cada perfil dentro da Universidade. A informação é organizada de acordo com os perfis. Por exemplo: professores podem criar e associar materiais extras, gerar anúncios de eventos e enviar lembretes. Alunos podem acessar materiais, enviar trabalhos e receber anúncios. Equipe administrativa gerencia dados de usuários, cursos e pode gerar relatórios de desempenho, assim o *chat* é organizado de acordo com as necessidades de cada persona, nas seguintes funcionalidades:
- Materiais extras: professores podem associar *links* para *websites*, arquivos na nuvem e vídeos do YouTube a cursos e tópicos, enriquecendo o conteúdo disponibilizado aos alunos.
 - Eventos e lembretes: o sistema gerencia datas, horários e locais de eventos acadêmicos e esportivos, enviando lembretes automáticos aos usuários relevantes, como alunos e professores.
 - Trabalhos de casa: o sistema permite que os alunos enviem trabalhos e que os professores recebam e avaliem esses trabalhos.

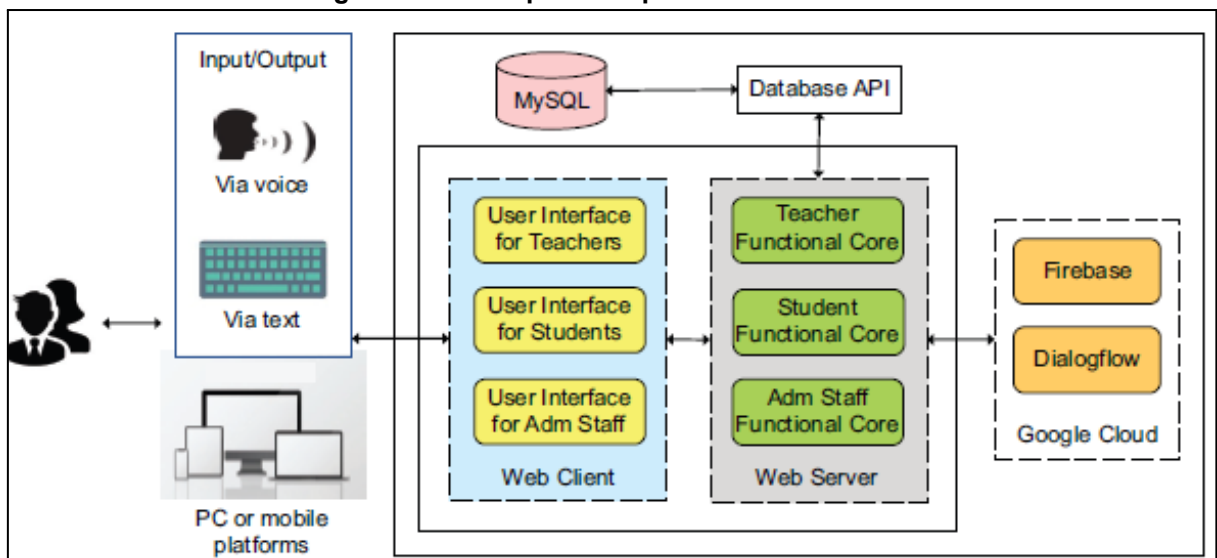
- Comunicação: a comunicação entre os usuários é facilitada pelo envio e recebimento de mensagens de texto e voz, proporcionando uma interação flexível.
 - Perfis de usuário: a equipe administrativa tem o controle sobre a gestão de perfis de usuários, permitindo o acesso a dados e registros acadêmicos de alunos e professores.
 - Relatórios de desempenho: gera relatórios estatísticos sobre o desempenho dos alunos, oferecendo informações detalhadas para professores e equipe administrativa, ajudando no acompanhamento acadêmico.
5. Armazenamento: utiliza o Firebase para armazenar informações relacionadas ao *chatbot*, como perguntas e respostas, e o MySQL para dados mais estruturados, como dados de usuários, lembretes, materiais, mensagens e cursos. A Interface de Programação de Aplicações (API) de banco de dados Object-Relational Mapping⁴ (ORM) facilita o acesso aos dados no MySQL, garantindo uma gestão eficiente.
 6. Ações proativas do *chatbot*: o *chatbot* é capaz de iniciar conversas proativas, como, por exemplo, com alunos com baixo desempenho ou com professores sobre problemas de compreensão dos alunos. Além disso, o sistema alerta a equipe administrativa sobre alunos com problemas emocionais ou dificuldades acadêmicas, promovendo uma abordagem proativa e preventiva.
 7. Interface: o modelo oferece flexibilidade na linguagem natural e integra interfaces gráficas (*widgets*), o que facilita a interação, especialmente para usuários com pouca familiaridade com sistemas digitais.

Na Figura 13, de Mendoza *et al.* (2022), são descritos os principais componentes tecnológicos utilizados na implementação do *chatbot*, organizados segundo um modelo cliente-servidor *web*. No lado cliente, foram desenvolvidas três interfaces distintas, voltadas para professores, estudantes e equipe administrativa, utilizando AngularJS. Essas interfaces permitem a interação por texto ou voz, tanto em plataformas *desktop* quanto *mobile*. No lado servidor, um Web Server hospeda

⁴Object-Relational Mapping (ORM) é uma técnica de programação que mapeia objetos de um sistema orientado a objetos para tabelas de um banco de dados relacional. Isso facilita a manipulação dos dados do banco de dados a partir de métodos e atributos dos objetos, sem a necessidade de escrever, diretamente, código SQL na maioria das operações.

uma Database API responsável por mapear os objetos da aplicação para um banco de dados relacional MySQL. Esse servidor também se conecta ao Firebase (utilizado para armazenamento em tempo real e de arquivos) e ao Dialogflow, que atua como o módulo de PLN e gerenciamento de *intents*. Quando o usuário envia uma mensagem, o Dialogflow realiza o *intent matching*, a extração de entidades e o controle do diálogo. Com base nas demandas identificadas, o sistema executa consultas ao banco MySQL ou realiza operações de leitura e escrita no Firebase. O conteúdo recuperado é, então, formatado e enviado de volta ao canal de origem, completando o ciclo de interação.

Figura 13 - Principais componentes do *chatbot*



Fonte: Mendoza *et al.* (2022).

7.2 Discussão (contexto informacional)

Os cinco estudos analisados apresentam uma diversidade de abordagens para o tratamento da informação no desenvolvimento de *chatbots* educacionais. Embora todos busquem facilitar o acesso à informação e à interação no contexto acadêmico, cada estudo adota metodologias distintas no PLN, nas fontes de dados utilizadas, nas estratégias de armazenamento e recuperação da informação e no nível de interação e personalização oferecido.

O estudo de Oliveira e Matos (2023) adota uma abordagem de busca semântica, utilizando *embeddings* vetoriais e indexação no Pinecone. A informação textual é carregada, segmentada e transformada em representações numéricas, que

capturam o significado semântico. Essa estratégia permite ao *chatbot* compreender a similaridade entre as consultas dos usuários e o conteúdo da base de conhecimento, mesmo que as palavras exatas não correspondam. Essa abordagem é flexível, permitindo que o sistema realize uma busca mais inteligente.

Destaca-se que as informações acadêmicas utilizadas (materiais de estudo, regulamentos, detalhes sobre cursos, entre outros dados da Universidade) foram organizadas em documentos de texto, que foram carregados de uma pasta específica.

Por outro lado, o estudo de *Thaiprasert et al. (2023)* adota uma estratégia baseada em classificação de *intents* com o Dialogflow ES e recuperação de respostas pré-definidas de um banco de dados Firestore. O estudo menciona apenas que o banco de dados foi construído utilizando o Firestore, onde os materiais são armazenados junto com suas entidades e sua pontuação média de utilidade, atribuída pelos usuários.

Ainda neste estudo, a informação é tratada por meio da identificação da intenção do usuário e da extração de entidades, usadas para buscar a resposta mais apropriada de um conjunto de dados existentes. Embora a abordagem de PLN seja eficaz para mapear intenções, ela é limitada à recuperação de respostas pré-configuradas. O estudo também inclui um sistema de avaliação de usuários, no qual a “pontuação média” dos materiais é atualizada com base nas interações, permitindo um mecanismo de aprendizado colaborativo.

O estudo de *Rukhiran e Netinant (2022)* segue uma linha similar ao estudo de *Thaiprasert et al. (2023)*, utilizando o Dialogflow para a compreensão da linguagem natural. No entanto, este estudo integra múltiplas fontes de dados, como Google Forms e servidores *web* da Universidade, ampliando o alcance e a diversidade das informações acessíveis. A principal diferença é que a recuperação de informações depende diretamente da organização das fontes de dados, sem a presença de uma indexação semântica explícita, o que limita um pouco a flexibilidade da busca. O estudo de *Muhyidin e Setiawan (2021)* também adota o Dialogflow para PLN, mas expande as fontes de dados, com *web scraping* e Google Sheets, e gerencia as respostas por meio de um *backend* em Flask. Essa abordagem proporciona maior flexibilidade, permitindo a integração de dados de diferentes origens de forma mais dinâmica, ao passo que a integração com múltiplos *frontends*, como o WhatsApp e

Telegram, amplia o alcance e facilita o acesso para usuários com diferentes níveis de familiaridade com tecnologia.

Por fim, o estudo de Mendoza *et al.* (2022) adota uma abordagem mais abrangente, focada no perfil de usuário e na gestão de informações específicas para diferentes funções dentro da instituição, como alunos, professores e equipe administrativa. Embora ela utilize o Dialogflow para PLN, o principal foco está na organização e no acesso a informações armazenadas em bancos Firebase e MySQL. O sistema é projetado para fornecer respostas contextualmente relevantes para diferentes tipos de usuários e pode até iniciar conversas proativamente, o que o torna um modelo interessante para *chatbots* educacionais.

Uma característica comum entre os estudos é a diversidade de fontes de dados e a necessidade de personalização da interação. Oliveira e Matos (2023) e Thaiprasert *et al.* (2023) apresentam abordagens mais diretas para a recuperação de informações específicas, como materiais de aula e dados regulatórios, enquanto os estudos de Muhyidin e Setiawan (2021) e Mendoza *et al.* (2022) expandem o escopo para incluir múltiplas fontes e usuários, com foco na escalabilidade e na flexibilidade. A indexação semântica e a recuperação baseada em pontuação média, nos estudos como Oliveira e Matos (2023) e Mendoza *et al.* (2022), são vantajosas para melhorar a precisão e relevância das respostas, enquanto os modelos baseados em *intents* são mais eficientes em contextos com tarefas bem definidas. Além disso, todos os estudos reconhecem a importância de uma avaliação contínua do sistema, seja por meio de *feedback* dos usuários (Thaiprasert *et al.*, 2023; Rukhiran; Netinant, 2022), testes automatizados (Muhyidin; Setiawan, 2021) ou atualização dinâmica da pontuação média (Mendoza *et al.*, 2022).

A privacidade e a segurança dos dados também são temas recorrentes, embora a implementação de políticas claras de dados ainda seja uma lacuna, especialmente em ambientes acadêmicos que lidam com dados pessoais sensíveis. Em termos de acessibilidade, a abordagem *role-based* (por funções), observada no estudo de Mendoza *et al.* (2022), e a adaptação a múltiplas plataformas de *frontend*, como em Muhyidin e Setiawan (2021), são particularmente relevantes no contexto de universidades públicas brasileiras, onde a diversidade no acesso a tecnologias é um fator crítico para garantir a inclusão digital.

Por fim, as estratégias de curadoria e atualização de informações foram observadas de diferentes formas nos estudos (*feedback* do usuário, testes automatizados e previsão de treinamento contínuo). Elas são essenciais para manter os sistemas de *chatbot* atualizados e relevantes em um ambiente dinâmico como o educacional.

7.3 Considerações sobre os chatbots educacionais

A análise dos estudos de Oliveira e Matos (2023), Thaiprasert *et al.* (2023), Rukhiran e Netinant (2022), Muhyidin e Setiawan (2021) e Mendoza *et al.* (2022) destacam *frameworks* bem definidos e tecnologias acessíveis, que contribuem para o fortalecimento de sistemas conversacionais em contextos educacionais. Esses projetos demonstram uma capacidade de automatizar respostas, facilitar o acesso à informação e de se adaptar a diferentes perfis de usuários.

No entanto, sob a ótica da CI, observam-se lacunas relevantes. Em geral, os estudos descrevem fluxos técnicos sofisticados, mas ainda com pouco detalhamento do fluxo informacional. Conceitos fundamentais, como curadoria, seleção, estruturação semântica, rastreabilidade e atualização das fontes, são pouco explorados.

Essa ausência compromete tanto a replicabilidade dos modelos quanto sua efetividade como mediadores informacionais. Mesmo quando há organização técnica de *intents*, bancos de dados ou arquitetura funcional, não fica claro como os conteúdos foram selecionados, validados ou mantidos. Essa limitação impede a avaliação crítica dos sistemas e sua adaptação a outros contextos institucionais.

Por outro lado, o estudo D49 (Mendoza *et al.*, 2022) se destaca por adotar a definição de personas com papéis distintos e funcionalidades específicas. A modelagem por perfis permite uma personalização mais eficaz da linguagem, das respostas e da mediação da informação, alinhando-se a autores como Almeida Júnior (2015), que defendem a adaptação da linguagem e da entrega informacional ao contexto de uso.

Em síntese, os estudos analisados contribuem significativamente do ponto de vista técnico, mas reforçam a necessidade de se integrar lógicas informacionais e critérios da curadoria nos projetos de *chatbot*. Essa articulação entre tecnologia e

organização da informação é o que orienta a proposta desta pesquisa, cuja intenção é trabalhar em aspectos informacionais na construção de *chatbots*, para subsidiar soluções confiáveis e fundamentadas nos princípios da CI.

8. BASE DE CONHECIMENTO

Inicialmente, é importante destacar que toda a construção da base de conhecimento está apoiada na curadoria informacional. A curadoria informacional, conceito que se origina da tradição museológica e foi ampliado pela CI, é compreendida como o processo de seleção, organização, contextualização e disseminação de informações, com foco em sua relevância, confiabilidade e adequação às necessidades dos usuários (Reina; Cruz, 2020; Aleixo *et al.*, 2020). Essa curadoria se diferencia da curadoria digital, centrada em aspectos técnicos de preservação, e da curadoria de conteúdo, voltada ao engajamento em mídias digitais, ao assumir um caráter mais abrangente e crítico. Nesse sentido, constitui-se um processo sistemático que organiza dados e garante a qualidade e a recuperabilidade da informação em ambientes complexos e dinâmicos, como o acadêmico (Bates, 2005; Siebra; Borba; Miranda, 2016).

No contexto dos sistemas conversacionais, a curadoria representa uma etapa importante para assegurar a qualidade das bases de conhecimento que alimentam os *chatbots*. A literatura destaca que a atuação humana na filtragem, validação e estruturação de conteúdos é decisiva para garantir que a mediação informacional realizada por esses sistemas seja efetiva (Del Sole *et al.*, 2025).

Esta etapa metodológica teve como objetivo realizar a curadoria informacional para subsidiar a implementação de uma base de conhecimento voltada à recuperação de informações acadêmicas, destinada à construção de um *chatbot* para atendimento dos estudantes da ECI, contemplando especificamente os cursos de Arquivologia, Biblioteconomia e Museologia.

A construção de uma base de conhecimento partiu da constatação, percebida nas respostas ao questionário ([Seção 6](#)), de que os estudantes apresentam dificuldades na localização, compreensão e aplicação das informações normativas, especialmente daquelas contidas na Resolução Complementar nº 01/2018 (UFMG, 2018) e das experiências identificadas na construção de *chatbots* educacionais ([Seção 7](#)), reunindo informações para a estruturação da base de conhecimento.

Assim, para orientar a organização dos conteúdos nessa base, foram integrados três eixos, que se complementam: (i) o mapeamento do comportamento informacional dos estudantes ([Seção 6](#)), que evidenciou dúvidas recorrentes e

padrões de busca, em consonância com os modelos de comportamento informacional de Wilson (1999) e Kuhlthau (2004), que ressaltam a influência da incerteza e das necessidades informacionais na mediação; (ii) a análise de modelos de *chatbots* educacionais ([Seção 7](#)), que forneceu boas práticas metodológicas e interativas, como as propostas por Mendoza *et al.* (2022) e Thaiprasert *et al.* (2023); (iii) os fundamentos da Ciência da Informação, especialmente no que se refere à organização, representação e recuperação da informação.

A base de conhecimento articula diretamente os fundamentos da CI com a prática da curadoria informacional. A organização da informação foi realizada a partir do mapeamento e da sistematização das normas acadêmicas, estruturando conteúdos dispersos em uma base coerente. Lemos, Martins e Souza (2023) destacam a importância das arquiteturas informacionais como mediadoras entre sistemas e usuários, sendo, nesta pesquisa, a organização dos conteúdos um desafio constante frente à fragmentação existente. A representação da informação ocorreu na tradução da linguagem normativa para termos acessíveis, por meio de descritores, sinônimos e categorias, aproximando o vocabulário técnico da linguagem cotidiana dos estudantes. Ferneda (2012, 2021) enfatiza a representação como núcleo da eficácia nos sistemas de recuperação. A recuperação da informação foi operacionalizada pela estrutura da base de conhecimento e na estrutura taxonômica, visando futuros fluxos conversacionais, que possibilitam o acesso rápido e preciso às respostas.

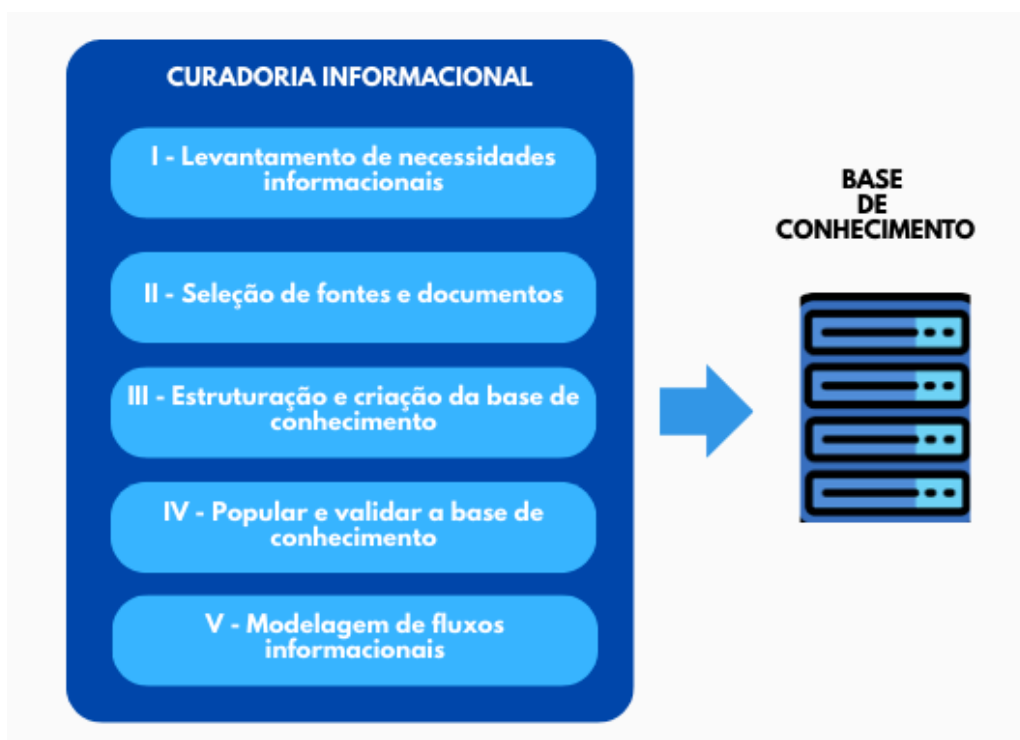
O uso da informação se concretizará na interação dos estudantes com o sistema, subsidiado pela base de conhecimento, transformando conteúdos complexos em orientações práticas e aplicáveis. Baeza-Yates e Ribeiro-Neto (2013) ressaltam que a representação adequada amplia a precisão e a abrangência da busca, enquanto Siebra, Borba e Miranda (2016) situam a curadoria informacional como prática essencial de mediação para o acesso qualificado à informação.

Com base na tipologia apresentada no Quadro 3, na [Seção 3.2](#), a base de conhecimento apoia a construção de um *chatbot* de modelo híbrido, em termos de funcionalidade, estrutura de interação e tecnologia. Em termos funcionais, ela combina elementos de FAQ Bots, que, segundo Mataloni Neto (2023), operam por mapeamento de perguntas e respostas pré-definidas, com recursos de *chatbots* conversacionais, capazes de lidar com ambiguidade e múltiplas variações

linguísticas por meio de PLN. Do ponto de vista da interação, essa base alinha-se ao modelo descrito por Reina e Cruz (2020), no qual fluxos estruturados por regras são combinados a interações abertas, garantindo precisão sem abrir mão da flexibilidade. Na dimensão tecnológica, ela corresponde ao *chatbot* híbrido descrito pela IBM (2025), que integra regras determinísticas e técnicas de IA, como o aprendizado adaptativo, unindo confiabilidade e capacidade de generalização.

A construção da base de conhecimento seguiu os procedimentos metodológicos definidos na [Seção 4.7](#), detalhados nas etapas da Figura 14, sendo apresentados na mesma ordem (I – Levantamento de necessidades informacionais; II – Seleção de fontes e documentos; III – Estruturação e criação da base de conhecimento; IV – Popular e validar a base de conhecimento; e V – Modelagem de fluxos informacionais), nas subseções 8.1 a 8.5, a seguir.

Figura 14 - Etapas para a construção da base de conhecimento



Fonte: elaborado pela autora (2025).

8.1 Levantamento de necessidades informacionais

Considerando o estudo de caso desta pesquisa, o levantamento de necessidades informacionais teve como objetivo identificar os aspectos das

informações acadêmicas (normas, resoluções e procedimentos) que geram maior dúvida entre os estudantes de graduação da UFMG. Os resultados da [Seção 6](#) foram essenciais para esse levantamento, que foi complementado com a literatura sobre *chatbots* educacionais na [Seção 7](#).

O levantamento foi realizado a partir da análise das respostas do questionário aplicado aos estudantes de graduação da UFMG. Esse instrumento buscou compreender como os discentes procuram informações acadêmicas, quais estratégias utilizam, que dificuldades enfrentam nesse processo e quais documentos institucionais geram maior número de dúvidas. A partir das respostas, foi possível identificar padrões de comportamento de busca, como a dependência de redes informais (colegas e grupos em redes sociais) diante da dificuldade de localizar as informações (normas, resoluções e procedimentos) nos canais oficiais, além da recorrência de problemas relacionados à dispersão das fontes, à linguagem técnica dos documentos e à ausência de um canal centralizado de acesso.

Assim, as análises indicaram que a compreensão das dúvidas e do comportamento de busca dos estudantes é pré-requisito essencial para a construção de uma base de conhecimento confiável e acessível. O cruzamento entre categorias empíricas, construídas a partir da análise das respostas dos estudantes e que refletem suas dúvidas mais recorrentes, como, por exemplo, o trancamento, o desligamento, a matrícula e a integralização, foi utilizado como um conceito a ser incluído na base. A literatura especializada orientou a etapa seguinte do projeto, dedicada à seleção das fontes normativas que alimentam a base de conhecimento. Esse procedimento reforça princípios centrais da CI ao articular organização, representação e recuperação da informação no processo de mediação tecnológica.

8.2 Seleção de fontes e documentos

Nesta pesquisa, consideram-se como fontes os canais institucionais responsáveis pela produção, disponibilização e atualização das informações que compõem a base de conhecimento. Incluem-se documentos normativos centrais, como a Norma Geral da Graduação (UFMG, 2018), e regulamentos complementares, além de manuais, guias e páginas institucionais, que orientam procedimentos acadêmicos. Também foram utilizados materiais de apoio, como fluxos de processos e portais informativos, que, embora não tenham caráter

normativo, cumprem papel de mediação ao traduzirem e operacionalizarem as regras para o cotidiano discente.

A construção da base de conhecimento exigiu selecionar fontes que garantissem confiabilidade, atualidade e aplicabilidade. Para isso, reuniram-se documentos normativos, guias institucionais e registros procedimentais, utilizados como referência tanto na gestão acadêmica quanto na mediação da informação.

A busca exploratória ocorreu nos portais institucionais da UFMG, abrangendo *sites* da universidade, unidades acadêmicas e sistemas internos, com o objetivo de mapear documentos oficiais, instruções operacionais e fluxos processuais relevantes para o cotidiano estudantil. Observou-se uma grande dispersão informacional, o que reforçou a importância de consolidar essas fontes em uma base integrada.

O núcleo da base documental foi constituído pela Norma Geral da Graduação (UFMG, 2018), que reúne as disposições gerais sobre a vida acadêmica dos discentes de graduação, bem como procedimentos comuns no percurso do aluno, como: matrícula, trancamento, desligamento, integralização curricular, dentre outros. Em complemento, o manual do sistema acadêmico SIGA foi incorporado como guia operacional, oferecendo instruções práticas e orientações para a utilização do sistema acadêmico da UFMG.

Além das normas centrais, foram incluídas fontes institucionais complementares, que ampliam a cobertura e favorecem a aplicação prática das regras, com destaque para:

- o *site* da ECI UFMG, com instruções específicas para Arquivologia, Biblioteconomia e Museologia;
- a página de Regras Acadêmicas da UFMG, que apresenta direitos e deveres em linguagem acessível;
- a página do Departamento de Registro e Controle Acadêmico da UFMG;
- os fluxos processuais em notação Business Process Management (BPM), que descrevem rotinas administrativas de forma padronizada e visual, disponíveis no *site* da ECI UFMG;
- o portal Viver UFMG, que reúne informações úteis para calouros e veteranos sobre a vida acadêmica,

Esses materiais, embora não façam parte, diretamente, das normas, funcionam como apoio, auxiliando os estudantes na compreensão e aplicação das regras. Assim, a base de conhecimento foi organizada de modo a equilibrar a exatidão técnica de normas, regulamentos e procedimentos com a usabilidade prática, constituindo um acervo integrado que sustenta o processo de curadoria informacional.

No que se refere às fontes e aos documentos utilizados, o Quadro 9 apresenta uma relação detalhada, incluindo formato, descrição e abrangência. Destaca-se que esse quadro está alinhado com a estrutura de metadados da base de conhecimento.

Quadro 9 - Relação das fontes e dos documentos utilizados

Fonte	Link	Documento	Formato	Descrição	Abrangência
Site da pró-reitoria de graduação	https://www.ufmg.br/prograd/wp-content/uploads/2024/07/ResCepeComp0118.pdf	NGG	pdf	Documento normativo central que regulamenta direitos, deveres e procedimentos acadêmicos da graduação na UFMG. Base principal da curadoria, por reunir disposições gerais que orientam todos os cursos.	Todos os cursos de graduação da UFMG
Site da pró-reitoria de graduação	https://www.ufmg.br/prograd/normas-gerais-de-graduacao-e-resolucoes/	Resolução CEPE nº 01/2018	pdf	Regulamenta o processo de matrícula em atividades acadêmicas curriculares, conforme previsto no art. 94 das NGG.	Todos os cursos de graduação da UFMG
Site da pró-reitoria de graduação	https://www.ufmg.br/prograd/normas-gerais-de-graduacao-e-resolucoes/	Resolução CEPE nº 02/2018	pdf	Regulamenta a oferta e definição do número de vagas ofertadas nas atividades acadêmicas curriculares, conforme previsto no art. 18 das NGG.	Todos os cursos de graduação da UFMG
Site da pró-reitoria de graduação	https://www.ufmg.br/prograd/normas-gerais-de-graduacao-e-resolucoes/	Resolução CEPE nº 01/2019	pdf	Regulamenta a criação e o registro de atividades acadêmicas curriculares, conforme previsto no art. 8 das	Todos os cursos de graduação da UFMG

	es/			NGG, bem como disposições específicas sobre exame especial, trabalho de conclusão de curso e estágio.	
Site da pró-reitoria de graduação	https://www.ufmg.br/prograd/normas-gerais-de-graduacao-e-resolucoes/	Resolução CEPE nº 04/2019	pdf	Regulamenta a realização de exame de comprovação de conhecimentos por estudantes de graduação da UFMG, conforme previsto no art 10 das NGG.	Todos os cursos de graduação da UFMG
Site da pró-reitoria de graduação	https://www.ufmg.br/prograd/normas-gerais-de-graduacao-e-resolucoes/	Resolução CEPE nº 06/2019	pdf	Regulamenta o aproveitamento de estudos realizados por estudantes de graduação em outras instituições de ensino superior, inclusive aqueles realizados em mobilidade acadêmica nacional e internacional, conforme previsto no art. 10 das NGG.	Todos os cursos de graduação da UFMG
Site da pró-reitoria de graduação	https://www.ufmg.br/prograd/normas-gerais-de-graduacao-e-resolucoes/	Resolução CEPE nº 10/2019	pdf	Estabelece diretrizes curriculares para a integralização de atividades acadêmicas curriculares de Formação em Extensão Universitária nos cursos de graduação da UFMG.	Todos os cursos de graduação da UFMG
Site da pró-reitoria de graduação	https://www.ufmg.br/prograd/normas-gerais-de-graduacao-e-resolucoes/	Resolução CEPE nº 14/2019	pdf	Regulamenta o regime especial para estudantes de graduação da UFMG, conforme previsto no art. 16 das NGG.	Todos os cursos de graduação da UFMG
Site da pró-reitoria de graduação	https://www.ufmg.br/prograd/normas-gerais-de-graduacao-e-resolucoes/	Resolução CEPE nº 01/2023	pdf	Regulamenta o Regime Acadêmico Especial para Permanência (RAEP) de estudantes da graduação da UFMG.	Todos os cursos de graduação da UFMG
Site da pró-reitoria de graduação	https://www.ufmg.br/prograd/normas-gerais-de-graduacao-e-resolucoes/	Resolução CEPE nº 05/2025	pdf	Regulamenta os processos de mobilidade acadêmica nacional e internacional no âmbito da graduação na UFMG.	Todos os cursos de graduação da UFMG

Site ECI Graduações	https://graduacoes.eci.ufmg.br/procedimentos-academicos/	Procedimentos acadêmicos	html	Reúne instruções específicas para trancamento, matrícula, colação de grau e outros procedimentos, com foco nos cursos da ECI. Fonte essencial para mapear variações de aplicação das normas gerais.	Arquivologia, Biblioteconomia e Museologia
Site UFMG	https://ufmg.br/vida-academica/regras-academicas	Regras acadêmicas	html	Página institucional que sintetiza direitos, oportunidades e obrigações do estudante. Apresenta linguagem acessível, funcionando como ponte entre documentos normativos e o cotidiano discente.	Todos os cursos de graduação da UFMG
Manual SIGA	https://www.ufmg.br/dti/wp-content/uploads/2016/08/Manual_Aplicativo_Siga_UFMG.pdf	Manual de Funcionalidades	pdf	Guia de uso do Sistema de Gestão Acadêmica (SIGA), com instruções para matrícula, emissão de documentos e acompanhamento da vida acadêmica. Complementa as normas ao explicar o funcionamento prático do sistema.	Todos os cursos de graduação da UFMG
Site UFMG Departamento de Registro e Controle Acadêmico	https://www2.ufmg.br/drca/drca/Home/Graduacao	Procedimentos administrativos	html	Canal oficial para acesso a informações sobre registro, matrícula, histórico escolar e outros trâmites formais.	Todos os cursos de graduação da UFMG
Site Viver UFMG	https://www.ufmg.br/viverufmg/primeiros-passos/perguntas-frequentes/	Guia Viver UFMG	html	Plataforma da UFMG destinada a calouros e estudantes veteranos, reunindo informações sobre os primeiros passos para a vida acadêmica.	Todos os cursos de graduação da UFMG
Fluxos BPM	https://graduacoes.eci.ufmg.br/procedimentos-academicos/	Procedimentos acadêmicos	html	Representação de processos acadêmicos em fluxos informacionais.	Arquivologia, Biblioteconomia e Museologia

Fonte: elaborado pela autora (2025).

8.3 Estruturação e criação da base de conhecimento

Esta seção apresenta uma breve discussão da estrutura da base de conhecimento com o referencial teórico da pesquisa, explicando como cada elemento foi fundamentado na literatura da CI e como essa fundamentação orienta a construção da base. O objetivo aqui é mostrar a coerência metodológica na construção da base e de que maneira suas escolhas impactarão o desempenho, a precisão e a transparência dos sistemas que a utilizarão.

A base de conhecimento pode ser definida como um repositório estruturado de informações, organizado para apoiar processos de recuperação e tomada de decisão. Para Ferneda (2012), trata-se de um conjunto sistematizado de dados e informações que, quando organizados de forma inteligível, permitem gerar conhecimento aplicável em diferentes contextos.

No campo da CI, destaca-se, ainda, a relevância da representação adequada para facilitar a recuperação (Baeza-Yates; Ribeiro-Neto, 2013). Complementarmente, a organização semântica e o uso de estruturas controladas são apontados como elementos centrais para bases capazes de oferecer respostas contextualizadas e precisas (Ferneda, 2021; Lemos; Martins; Souza, 2023).

Com base nessa compreensão, esta pesquisa desenvolveu uma base de conhecimento voltada ao contexto acadêmico da UFMG, estruturada a partir do conteúdo normativo institucional. O objetivo foi apresentar perguntas e respostas sobre informações acadêmicas (normas, resoluções e procedimentos), tendo como elemento central o conteúdo das NGG e de documentos complementares da ECI UFMG em informação estruturada e acessível, em um ambiente organizado e controlado, mantendo a precisão técnica e a rastreabilidade das fontes.

(a) Conceitos e estrutura taxonômica

A taxonomia organizou os conteúdos normativos do mais amplo ao mais específico, favorecendo fluxos conversacionais coerentes. Foram definidos os conceitos normativos centrais, organizados em uma hierarquia de três níveis (categoria, subcategoria e sub-subcategoria). Essa estrutura apoiou-se na teoria do conceito (Dahlberg, 1978) e na necessidade de padronização terminológica para a recuperação precisa (Lima; Fujita; Redigolo, 2021), em consonância com a perspectiva de organização e representação como condições para a recuperação (Lancaster, 2004). Com isso, a taxonomia orienta o fluxo conversacional e a

navegação por botões, conduzindo o estudante a termos amplos e a níveis mais específicos, o que favorece a desambiguação desde o início do diálogo.

(b) Vocabulário controlado e tratamento terminológico

O vocabulário utilizado na construção da base articula descritores controlados, ou seja, termos genéricos (TG) e termos específicos (TE), do Thesaurus Brasileiro da Educação (BRASED) e termos coloquiais mapeados nas respostas do questionário, reconhecendo a diferença entre linguagem normativa e linguagem natural, falada pelos estudantes. A escolha se apoia na literatura sobre linguagens documentárias (Ferneda, 2012, 2021; Lima; Fujita; Redigolo, 2021) e nas discussões sobre relevância e contexto (Capurro; Hjørland, 2003). Nesse sentido, referenciais clássicos, como Dewey (1876), e Lancaster (2004), evidenciam a indexação como processo técnico e cognitivo central à mediação entre diferentes linguagens.

(c) Perguntas / expressões e intenções de busca

Com base nos conceitos, foram modeladas as intenções de busca, representando padrões de perguntas dos estudantes. A formulação das questões dialoga com Kuhlthau (2004) e Savolainen (1995) ao situar a busca no contexto informacional do usuário. Dessa forma, as intenções de busca funcionam como perguntas-modelo que treinam o classificador a reconhecer variações de linguagem natural (ex: “posso trancar tudo?”, “como tranco este semestre?” ou “qual o prazo do trancamento total?”). Esse recurso pode aproximar o *chatbot* de um modelo de mediação (Almeida Júnior, 2008), transformando dúvidas reais em pontos de acesso estruturados (Thaiprasert *et al.*, 2023).

(d) Respostas e entidades

Na base de conhecimento, as entidades normativas foram definidas como elementos que qualificam o contexto das dúvidas dos estudantes. Essa modelagem segue os princípios da RI (Baeza-Yates; Ribeiro-Neto, 2013; Muhyidin; Setiawan, 2021), permitindo a distinção entre situações semelhantes na linguagem dos usuários, mas que possuem regras normativas distintas.

Na lógica de interação em sistemas conversacionais, as entidades funcionam como condições necessárias para a formulação da resposta. Cada resposta do *chatbot* depende da combinação correta de entidades identificadas na pergunta do estudante. Quando alguma entidade está ausente, o sistema não consegue gerar,

de imediato, a resposta normativa específica. Nesses casos, o sistema conversacional poderá acionar um processo de refinamento, solicitando as informações faltantes até reunir todos os parâmetros obrigatórios. Assim, a relação entre entidades e respostas é de dependência direta: as entidades delimitam a situação e, a partir delas, é possível recuperar e entregar a resposta normativa exata, mantendo a aderência às regras institucionais e a consistência do diálogo.

(e) Referências normativas e rastreabilidade

Todas as respostas foram redigidas com base em documentos oficiais (Resolução Complementar nº 01/2018 e resoluções do Conselho de Ensino, Pesquisa e Extensão (CEPE) correlatas), acompanhadas de citação explícita do artigo, parágrafo, alínea e ou *link* institucional. Esse procedimento responde aos princípios de transparência e confiabilidade (Lemos; Martins; Souza, 2023). Assim, cada resposta possui um metadado que contém a “Localização na NGG (Artigo de Referência)” e/ou um *link* institucional, o que torna a base de conhecimento auditável e facilita atualizações quando ocorrem mudanças normativas.

(f) Metadados, controle e manutenção

Para operacionalizar a base, inicialmente, foi elaborada uma planilha em Excel, que facilitou sua exportação em outros formatos não proprietários, como o Comma-Separated Values (CSV), composta por metadados que organizam, contextualizam e conectam os conteúdos normativos e todos os elementos da estrutura da base. Os metadados contemplam: identificação única (ID), conceito, termos do Thesaurus BRASED, selecionado por ser um vocabulário controlado da área, indicando termos gerais e termos específicos. Também foram descritos termos coloquiais e sinônimos, perguntas/expressões, respostas, referência normativa, além de *links* para documentos institucionais e complementares.

Essa arquitetura traduz princípios dos padrões de metadados (Gilliland, 2016) e curadoria informacional (Lemos; Martins; Souza, 2023) em um artefato de gestão. Com isso, os metadados viabilizam rastreabilidade, versionamento e expansão, permitindo a inclusão de novos conceitos e perguntas sem comprometer a estrutura, apoiando a validação contínua, com correções e inclusão de perguntas complementares.

O detalhamento dos metadados, bem como a finalidade e técnicas empregadas, são apresentadas no Quadro 10.

Quadro 10 - Estrutura da base de conhecimento

Metadado	Finalidade	Técnica de organização da informação aplicada
ID	Identificar, de forma única, cada registro, viabilizando o controle, a rastreabilidade e a referência cruzada.	Atribuição de identificador único.
Tipologia geral	Classificar os registros pelo tipo de informação (Definição, Prazo, Pré-requisitos, Limite de solicitações ou Processo), padronizando o conteúdo e apoiando a identificação de entidades no <i>chatbot</i> .	Classificação tipológica baseada na análise conceitual e funcional da informação (Dahlberg, 1978), estruturada em facetas, conforme o papel informacional no fluxo acadêmico.
Conceito	Representar o núcleo semântico de cada entrada, descrevendo de forma sintética o tema ou processo acadêmico (ex.: trancamento, desligamento ou integralização)	As entradas foram organizadas por indexação, assunto e descritores, estando fundamentadas na teoria do conceito, de Dahlberg (1978), que define conceito como unidade de conhecimento formada por características essenciais.
Termos Thesaurus TG e TE	Assegurar a padronização terminológica por meio de TG e TE extraídos do BRASED.	Vocabulário controlado; hierarquia terminológica; relações semânticas de equivalência; e subordinação.
Termos coloquiais/dicionário (sinônimos)	Mapear sinônimos, variações linguísticas e expressões estudantis para aproximar a linguagem normativa da linguagem natural do usuário.	Controle de sinônimos; pontos de acesso em linguagem natural; e mediação terminológica.
Expressão/pergunta	Reunir múltiplas formas de formulação de consulta, representando as possíveis entradas do usuário.	Expansão de pontos de acesso e ampliação do escopo.
Resposta	Oferecer respostas claras, acessíveis e normativamente fundamentadas, traduzindo o texto técnico da norma para uma linguagem compreensível.	Curadoria informacional e tradução de linguagem técnica em linguagem de uso (Reina; Cruz, 2020).
Definição das NGG ou do documento norteador	Oferecer as respostas conforme o documento oficial (norma, resolução ou procedimento)	Referência normativa.
Localização nas NGG (artigo de referência) ou no documento norteador	Garantir a rastreabilidade das respostas por meio da citação de artigo, parágrafo e alínea.	Referência normativa.
Documentação ECI - Localização <i>site</i> ECI	Direcionar o usuário à versão integral do documento no portal institucional, assegurando acesso à fonte primária.	Ligação hipertextual e vinculação à fonte original.
<i>Link</i> para documentos	Direcionar o usuário à versão integral do	Ligação hipertextual e

complementares (Manual SIGA, <i>site</i> UFMG (regras acadêmicas), <i>site</i> Viver UFMG, <i>site</i> Departamento de Registro e Controle Acadêmico (DRCA), <i>site</i> Prograd e PRAE	documento em <i>sites</i> complementares, assegurando acesso à fonte primária.	vinculação à fonte original.
Fluxo BPM (quando aplicável)	Representar de processos acadêmicos que envolvem etapas sequenciais (ex.: trancamento, matrícula e colação de grau).	Modelagem de Processos.

Fonte: elaborado pela autora (2025).

A integração de múltiplas fontes, NGG, resoluções do CEPE e regulamentos da ECI seguiu boas práticas de curadoria, harmonizando dados estáticos e dinâmicos e adotando arquiteturas flexíveis, que combinam documentos, planilhas, bancos de dados e *web scraping* (Rukhiran; Netinant, 2022; Muhyidin; Setiawan, 2021). Com a curadoria e a organização dos conteúdos normativos estabelecidos, tornou-se possível estruturar a base de conhecimento, garantindo que os fluxos conversacionais reflitam tanto a precisão normativa quanto a compreensão acessível aos estudantes.

8.4 Desenvolvimento da base de conhecimento (popular e validar)

A base de conhecimento foi estruturada a partir do mapeamento das principais dúvidas dos estudantes de graduação da UFMG, tomando como referência os conceitos normativos identificados na etapa de curadoria. O objetivo foi assegurar precisão normativa, aplicabilidade prática e mediação entre a linguagem técnica de normas, regulamentos e procedimentos e o vocabulário cotidiano dos discentes.

Foram organizados 31 conceitos centrais, que representam as categorias temáticas mais recorrentes nas demandas informacionais: Trancamento, Desligamento, Matrícula, Integralização, Aproveitamento de Estudos, Atividades Complementares, Colação de Grau, Estágio e Mobilidade Acadêmica. Esses conceitos poderão funcionar como entidades dentro do sistema conversacional, permitindo que o *chatbot* reconheça elementos específicos mencionados nas perguntas dos usuários. A taxonomia criada é apresentada a seguir:

1. Trancamento
 - 1.1 Trancamento total
 - 1.1.1 Trancamento total com justificativa
 - 1.1.2 Trancamento total sem justificativa
 - 1.2 Trancamento parcial
 - 1.2.1 Trancamento parcial com justificativa
 - 1.2.2 Trancamento parcial sem justificativa
2. Desligamento
 - 2.1 Desligamento por tempo de integralização 30%
 - 2.2 Desligamento por tempo de integralização 50%
 - 2.3 Desligamento por tempo máximo de integralização
 - 2.4 Desligamento por Nota Semestral Global (NSG)
 - 2.5 Desligamento por não matrícula
 - 2.6 Desligamento por infrequência
3. Matrícula
 - 3.1 Fases da matrícula
 - 3.1.1 1ª Fase da matrícula (obrigatória)
 - 3.1.2 2ª Fase da matrícula (opcional)
 - 3.1.3 3ª Fase da matrícula (opcional)
 - 3.1.4 4ª Fase da matrícula (opcional)
 - 3.2 Registro *a posteriori*
4. Matrícula prévia
5. Integralização
6. Disciplina isolada
7. Atividades acadêmicas curriculares
8. Aproveitamento de estudos
9. Atividades complementares geradoras de crédito
 - 9.1 Aproveitamento de créditos em eventos
 - 9.2 Aproveitamento de créditos em oficinas
 - 9.3 Aproveitamento de créditos em seminários
 - 9.4 Aproveitamento de estágio não obrigatório
 - 9.5 Aproveitamento de publicações

- 9.6 Aproveitamento de iniciação à docência ou monitoria
- 9.7 Aproveitamento de iniciação, extensão e pesquisa
- 9.8 Aproveitamento de vivência profissional complementar
- 9.9 Aproveitamento de monografia
- 9.10 Aproveitamento de iniciação à curadoria (Museologia)
- 9.11 Aproveitamento de cursos de curta duração (Museologia)
- 10. Colação de Grau
 - 10.1 Colação antecipada
- 11. Continuidade de estudos
- 12. Reopção
- 13. Estágio
 - 13.1 Estágio curricular
 - 13.2 Estágio não curricular
- 14. Diploma
- 15. Regime especial
 - 15.1 Regime acadêmico especial para permanência
- 16. Exame especial
- 17. Formação complementar
- 18. Formação Complementar Aberta (FCA)
- 19. Formação Livre (FL)
- 20. Auxílio financeiro estudantil
- 21. Mobilidade acadêmica
- 22. Mobilidade acadêmica internacional
- 23. Carteira da UFMG
- 24. Folha de Números de Identificação Pessoal (NIPs)
- 25. Fundação Universitária Mendes Pimentel (FUMP)
- 26. Restaurante Universitário (RU)
- 27. Moradia universitária
- 28. Bolsas e auxílios
- 29. Ônibus interno UFMG
- 30. Portal Minha UFMG
- 31. Núcleo de Acessibilidade e Inclusão (NAI)

É importante destacar que a base de conhecimento implementada, nesta pesquisa, priorizou as principais dúvidas dos estudantes de graduação, que foram apresentadas da maior frequência para a menor, com base nas dúvidas e opiniões encontradas na pesquisa realizada na [Seção 6](#). Assim, conceitos como Auxílio financeiro estudantil, Portal Minha UFMG, Carteira da UFMG e Folha de Nips, apesar de não estarem diretamente relacionados às normas, aos regulamentos e aos procedimentos, foram adicionados por se tratarem de conceitos importantes ligados ao percurso acadêmico dos estudantes no contexto da UFMG.

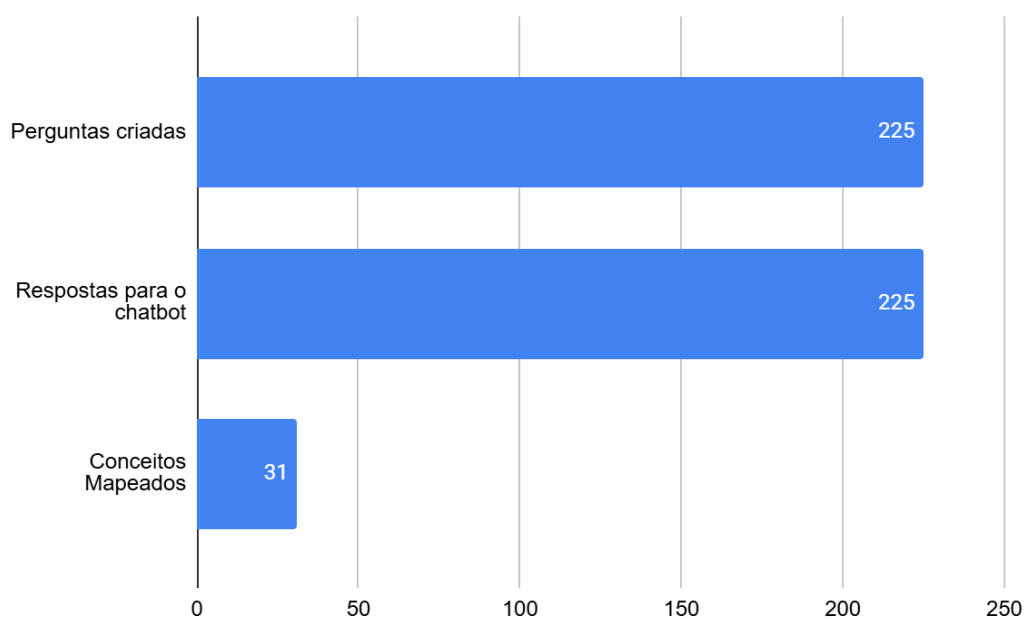
A partir desses conceitos, foram formuladas 225 perguntas, intenções do usuário, ou seja, *intents*, cada uma representando uma possível dúvida ou necessidade informacional. Com isso, foram elaboradas, também, 225 respostas. As respostas foram redigidas com base nos documentos oficiais da UFMG, sobretudo a Resolução Complementar nº 01/2018 (NGG/UFMG), complementada por regulamentos de curso e orientações institucionais. A redação buscou conciliar precisão normativa e clareza, aproximando termos técnicos da linguagem cotidiana dos estudantes, além de priorizar aplicabilidade prática e concisão, favorecendo interações rápidas no ambiente conversacional.

A elaboração da base de conhecimento representa um diferencial metodológico em relação às abordagens identificadas na literatura sobre *chatbots* educacionais, nas quais o conteúdo informacional costuma ser tratado apenas como um dado técnico a ser processado por algoritmos. Ao organizar a informação a partir de conceitos, termos controlados, expressões reais dos usuários, respostas em duas camadas (linguagem acessível e norma oficial) e vínculos diretos com artigos das NGG e documentos institucionais, esta pesquisa insere a informação como elemento estruturante. Esse modelo supera a lacuna apontada nos estudos revisados, que raramente apresentam metodologias explícitas de organização e curadoria da informação, e demonstra como os fundamentos da CI, especialmente a representação e a OI, que podem ser aplicados de forma prática para a RI, podem ser utilizados no desenvolvimento de sistemas conversacionais mais confiáveis, contextualizados e replicáveis em ambientes acadêmicos.

Os Gráficos 12 e 13 apresentam a distribuição desses quantitativos, evidenciando como cada conceito foi desdobrado em perguntas e respostas que refletem as principais demandas informacionais dos estudantes. Observa-se que

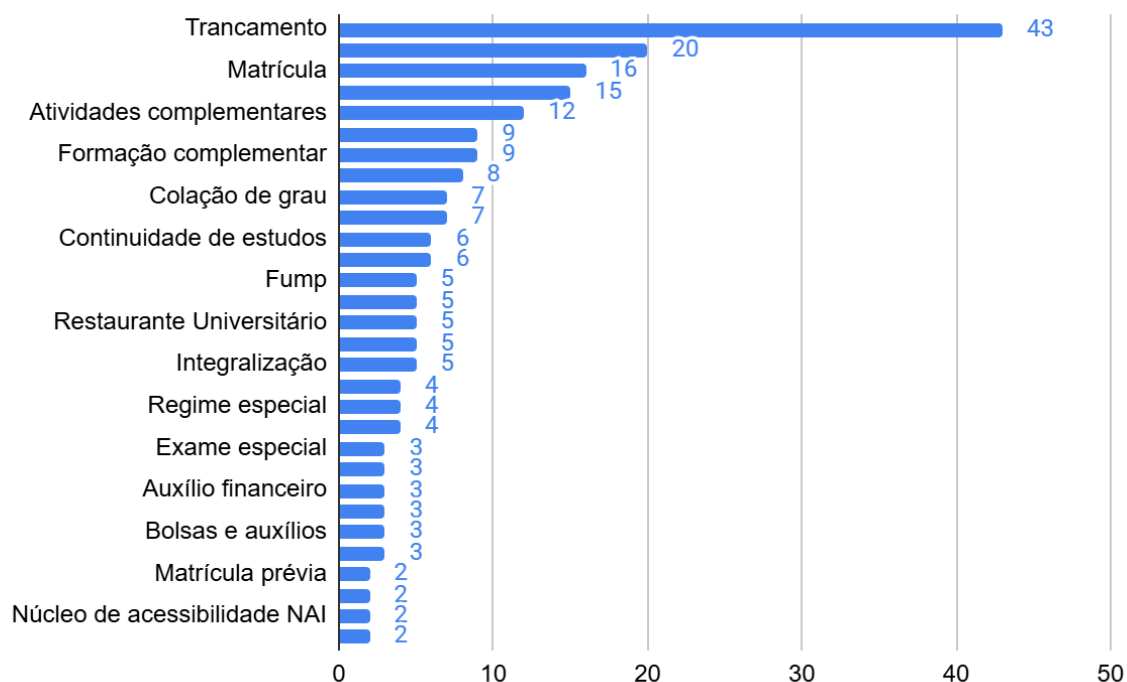
certos temas concentram maior número de dúvidas, como Trancamento, Desligamento e Matrícula, enquanto outros aparecem de forma menos recorrente, como Folha de NIPs e Moradia universitária. Essa distribuição demonstra a diversidade das necessidades informacionais e orienta a priorização dos fluxos conversacionais do *chatbot*.

Gráfico 12 - Detalhamento da base de conhecimento



Fonte: elaborado pela autora (2025).

Gráfico 13 - Quantidade de perguntas criadas



Fonte: elaborado pela autora (2025).

Em relação ao tratamento terminológico, o Quadro 14 apresenta o resultado da curadoria realizada a partir dos conceitos normativos mais relevantes para os estudantes de graduação da UFMG. Cada conceito foi associado a expressões específicas, termos presentes no thesauros e termos coloquiais, com o objetivo de ampliar a acessibilidade e a compreensão da linguagem normativa no contexto acadêmico.

Observa-se que conceitos como Trancamento apresentam maior número de expressões (43) e termos coloquiais (12), refletindo sua complexidade e frequência de questionamentos por parte dos discentes. Outros conceitos, como Matrícula Prévia ou Disciplina Isolada, apresentam menor diversidade de expressões, indicando uma menor complexidade semântica ou demanda informacional.

A presença de termos extraídos do thesauros evidencia a integração de vocabulários controlados na base, permitindo a padronização e precisão na RI. Ao mesmo tempo, a inclusão de termos coloquiais garante a mediação entre a linguagem técnica das normas e o vocabulário cotidiano dos estudantes, reforçando a abordagem centrada no usuário.

No conjunto, o Quadro 11 evidencia a estratégia adotada na curadoria: mapear conceitos normativos, desdobrá-los em expressões diversas e conectá-los a linguagens formais e informais, criando uma base de conhecimento estruturada, abrangente e acessível para a construção do *chatbot* e de outras aplicações.

Quadro 11 - Curadoria base de conhecimento

Conceitos	Perguntas/ Expressões	Termos do <i>Thesaurus</i> BRASED	Termos coloquiais e sinônimos
Trancamento	43	3	12
Desligamento	20	3	4
Matrícula	16	3	4
Estágio	15	6	1
Atividades Complementares	12	3	5
Aproveitamento de Estudos	9	4	5
Formação Complementar	9	2	1
Mobilidade Acadêmica	8	2	1
Colação de Grau	7	2	2
Formação Complementar Aberta	7	0	1
Continuidade de Estudos	6	2	3
Reopção	6	2	2
Fump	5	5	6
Atividades Acadêmicas Curriculares	5	2	4
Ru	5	2	3
Portal Minha UFMG	5	0	3
Integralização	5	3	2
Disciplina Isolada	4	2	1
Regime Especial	4	2	1
Mobilidade Acadêmica Internacional	4	3	3
Carteira da UFMG	4	0	1
Exame Especial	3	2	2
Diploma	3	2	1
Auxílio Financeiro	3	3	1
Folha De Nips	3	0	1
Bolsas E Auxílios	3	6	3
Ônibus Interno Ufmg	3	3	3
Matrícula Prévia	2	3	1

Moradia Universitária	2	3	3
NAI	2	0	3
Formação Livre	2	0	1

Fonte: elaborado pela autora (2025).

A validação da base de conhecimento foi realizada ao longo de sua construção, entre julho e setembro de 2025. Como critérios, foram considerados, principalmente: (1) a pertinência das perguntas e respostas e (2) a fonte onde a resposta foi coletada. Aspectos relativos aos termos e sinônimos, bem como o nível na estrutura taxonômica, também foram validados. Dessa forma, cada um dos 31 conceitos foram verificados, e as inconsistências, marcadas para revisão e correção. Além disso, ao longo da validação, foram indicadas perguntas para complementar conceitos e relações existentes.

Como artefato dessa dissertação, foi construída a base de conhecimento, que pode ser acessada em modo de visualização. As Figuras 15 e 16 apresentam a visualização de uma parte da base, sendo a primeira imagem uma visão geral e a segunda, um detalhamento para o conceito de trancamento (definição). A base completa, em sua versão do dia 03/10/25, está disponível no repositório da pesquisa⁵.

⁵ Disponível em: <http://doi.org/10.17632/j5x5jz7cfr.1>. Acesso em: 16 dez. 2025.

Figura 15 - Base de conhecimento (visão geral)⁶

A	B	C	D	E	F	G	H	I	
1	ID	Conceito	Tipologia geral	Termos thesaurus TG	Termos coloquiais / dicionário (sinônimos)	Expressão (pergunta)	Resposta	Definição da NGG e do documento norteador	Localização na NGG Artigo_Ref ou no documento norteador
2	1	Trancamento	Definição	Trancamento de Matrícula Matrícula Trancada No Curso	suspensão total pausa acadêmica trancar curso	O que é ? (Quais a definição?)	Trancamento de matrícula é o ato administrativo que permite ao estudante, mediante solicitação formal, interromper	O estudante poderá solicitar o trancamento total de matrícula e o trancamento parcial de matrícula.	Art. 95, caput
3	1	Trancamento	Prazo			Quais prazo trancamento?	De acordo com a NGG da UFMG, os prazos para trancamento são: Trancamento total sem justificativa: até	O estudante poderá solicitar o trancamento total de matrícula, com	Art. 96 I Art. 96 II Art. 96 III
4	1	Trancamento	Pré-Requisitos			É possível trancar a matrícula no primeiro período?	De acordo com as Normas Gerais de Graduação, não é permitido conceder trancamento total de matrícula sem	Independentemente da forma de acesso, o trancamento total de matrícula sem justificativa não	Art. 96, §3º
5	1	Trancamento	Limite Solicitações			Posso trancar mais de uma vez?	Não. O estudante terá direito ao trancamento total de matrícula uma única vez sem apresentação de	Durante sua permanência na graduação, no decorrer da vigência do seu vínculo ao curso,	Art. 96, §§1º e 2º, Art. 96, §3º
6	1	Trancamento	Processo			Como faço para pausar o curso?	É Preciso solicitar o trancamento total de matrícula. Sem justificativa: pode ser solicitado	O estudante poderá solicitar o trancamento total de matrícula, com	Art. 96 I Art. 96 II Art. 96 III
7	1.1	Trancamento Total	Definição			O que é?	O trancamento total de matrícula é conceituado como o deferimento de solicitação do estudante para não	O trancamento total de matrícula é conceituado como o deferimento de solicitação do estudante para	Art. 95 §1º
8	1.1	Trancamento Total	Prazo			Quais são os prazos para solicitar trancamento total?	Sem justificativa: até 30 dias após o início do período letivo Com justificativa: até 30 dias após à	I até 30 (trinta) dias após o início do período letivo, no caso de trancamento total sem justificativa	Art. 96 I Art. 96 II Art. 96 III
9	1.1	Trancamento Total	Limite Solicitações			À quantos trancamentos totais de matrícula cada estudante de graduação têm direito?	Cada estudante tem direito a um único trancamento total de matrícula sem justificativa. Embora não haja limite para	Cada estudante tem direito a um único trancamento total de matrícula sem justificativa.	Site UFMG: Regras acadêmicas
10			Limite			O estudante pode solicitar o	O aluno poderá solicitar o trancamento	O aluno poderá solicitar o	INSTRUÇÕES DO MÓDULO

Fonte: dados da pesquisa (2025).

⁶ Disponível em: <http://doi.org/10.17632/j5x5jz7cfr.1>. Acesso em: 16 dez. 2025.

Figura 16 - Base de conhecimento (detalhamento para o conceito trancamento - definição) ⁷

ID	Conceito	Tipologia geral	Termos theusauo TG	Termos coloquiais / dicionário (sinônimos)	Expressão (pergunta)	Resposta	Definição da NGG e do documento norteador	Localização na NGG Artigo_Ref ou no documento norteador	Formulários
1	Trancamento	Definição	Trancamento de Matrícula Matrícula Trancada No Curso Matrícula Trancada em Disciplina	suspensão total pausa acadêmica trancar curso interrupção completa pausar faculdade parar curso trancar matéria trancar Disciplina desistir da Disciplina não fazer uma Disciplina cortar uma matéria do semestre sair de uma matéria cancelar todas às matérias parar de estudar neste semestre trancar o semestre todo	O que é ? (Quais a definição?)	Trancamento de matrícula é o ato administrativo que permite ao estudante, mediante solicitação formal, interromper temporariamente sua participação em atividades acadêmicas de um curso de graduação, seja em sua totalidade (trancamento total) ou apenas em disciplinas específicas (trancamento parcial), mantendo o vínculo institucional durante o período de suspensão. "O estudante poderá solicitar o trancamento total de matrícula ou o trancamento parcial de matrícula."	O estudante poderá solicitar o trancamento total de matrícula e o trancamento parcial de matrícula. § 1o O trancamento total de matrícula é conceituado como o deferimento de solicitação do estudante para não desenvolver nenhuma atividade acadêmica curricular em um período letivo. § 2o O trancamento parcial de matrícula é conceituado como o deferimento de solicitação do estudante para não desenvolver às atividades referentes à uma determinada atividade acadêmica curricular na qual tenha se matriculado.	Art. 95, caput	

ID	Conceito	Localização na NGG Artigo_Ref ou no documento norteador	Formulários	Documentação ECI - Localização site ECI	Fluxo BPM (site ECI)	Manual SIGA	Site UFMG (regras acadêmicas)	Site Viver UFMG	Site DRCA	Site Prograd	Sites de curso UFMG
1	Trancamento	Art. 95, caput		https://ufmg.br/storage/f/0/9/0/f09021993bf696a1a49413178c1d92a4_15181056888278_1729582510.pdf			https://ufmg.br/vida-academica/regras-academicas#:~:text=â%20quantos%20trancamentos%20totais%20de%20matr%C3%ADcula%20se%20justificativa				

Fonte: dados da pesquisa (2025).

⁷ Disponível em: <http://doi.org/10.17632/j5x5jz7cfr.1>. Acesso em: 16 dez. 2025.

8.5 Modelagem de fluxos informacionais

Esta fase teve como foco integrar os principais conceitos normativos da graduação, especialmente os previstos na Resolução Complementar nº 01/2018 (NGG/UFMG), em um fluxo padrão de atendimento conversacional que oriente o estudante de forma clara, precisa e contextualizada. O desenho parte do contexto da ECI UFMG e adota, como primeiro passo, a identificação do curso (Arquivologia, Biblioteconomia ou Museologia), o que permite acomodar variações específicas (estágios, atividades complementares ou TCC), sem duplicar regras. Esse fluxo padrão não depende de uma ferramenta específica, pois foi descrito em etapas, com decisões e validações (acolhimento, seleção de temática, coleta de entidades, checagens condicionais, resposta com referência normativa e encaminhamentos), podendo ser implementado em diferentes plataformas que suportem estados de diálogo, captura de parâmetros e regras condicionais.

O fluxo de diálogo para o atendimento começa com uma mensagem de boas-vindas e com a identificação do curso do estudante (Arquivologia, Biblioteconomia ou Museologia). Esse passo inicial é necessário porque os procedimentos variam entre os cursos, como regras de estágio e atividades complementares. Ao registrar o curso, o sistema contextualiza a conversa e evita orientações genéricas, em consonância com a noção de relevância situada e mediação orientada ao contexto (Capurro; Hjørland, 2003; Wilson, 1999; Kuhlthau, 2004).

Em seguida, o *chatbot* solicita que o estudante indique a temática da dúvida. Para facilitar e orientar, são apresentados botões com os principais temas, como, por exemplo, Trancamento, Matrícula e Desligamento, mas o usuário também pode digitar livremente. As opções funcionam como uma lista de termos padronizados (vocabulário controlado), o que evita mal-entendidos, reduz o esforço e alinha a conversa aos conceitos institucionais (Dahlberg, 1978; Lima; Fujita; Redigolo, 2021). No desenho do diálogo, essa oferta de opções implementa um roteamento por regras que organiza a entrada da interação, prática recorrente em *chatbots* educacionais para melhorar a precisão e a experiência do usuário (Adamopoulou; Moussiades, 2020, Mendoza *et al.*, 2022 e Thaiprasert *et al.*, 2023).

Depois de escolhido o tema, o sistema apresenta uma definição breve para alinhar o entendimento e solicita que o estudante descreva sua situação. A partir dessa mensagem, o *chatbot* identifica o tipo de solicitação (por exemplo, “trancar o semestre”) e recupera os dados necessários para oferecer uma resposta precisa, considerando as especificidades do caso. No exemplo do Trancamento, a solicitação pode se referir a diferentes possibilidades: total ou parcial, com ou sem justificativa, e, ainda, ao período do semestre. Para lidar com essas variações, foi estruturada uma taxonomia que organiza cada especificação em níveis hierárquicos, permitindo distinguir situações semelhantes, mas normativamente distintas. Assim, por exemplo, o conceito principal de Trancamento se desdobra em categorias (total e parcial) e, dentro delas, em subcategorias, ligadas à justificativa exigida e ao período de aplicação (com justificativa, sem justificativa). Essa estrutura taxonômica assegura clareza, consistência e padronização no tratamento das informações, orientando o *chatbot* na coleta de parâmetros essenciais e garantindo respostas adequadas ao contexto informado pelo estudante.

Quando ainda estiverem faltando informações, o *chatbot* deve fazer perguntas curtas e objetivas até completar o necessário, dinâmica que espelha a negociação da pergunta e a redução de incerteza, descritas na literatura (Kuhlthau, 2004), e a importância do contexto na busca (Savolainen, 1995).

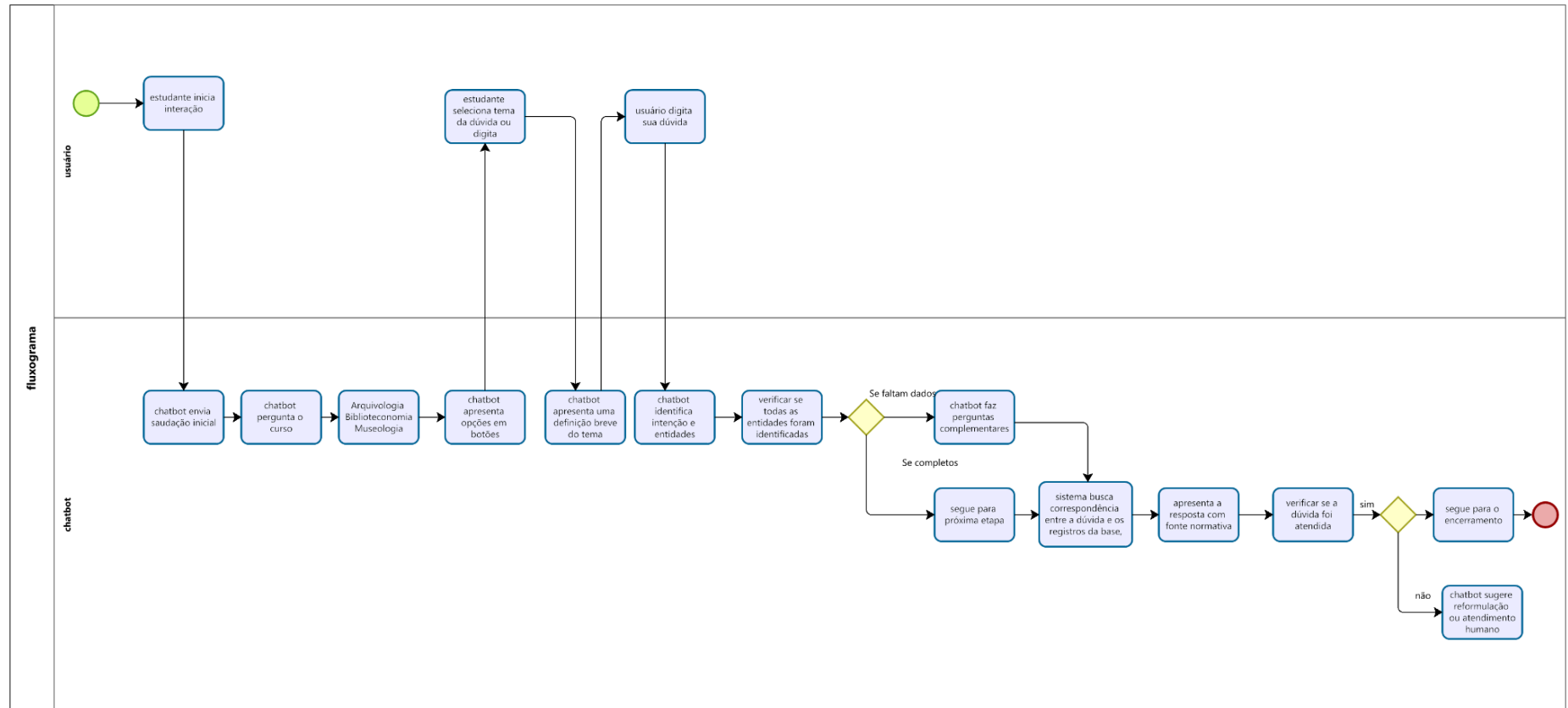
Com essas informações, o *chatbot* consulta a base de conhecimento e procura a melhor resposta, combinando a correspondência de palavras-chave com a proximidade semântica entre a pergunta do estudante e os itens já mapeados. Em termos de recuperação da informação, isso significa articular a precisão terminológica garantida pelo vocabulário controlado com a flexibilidade dos sinônimos e das expressões correntes, assegurando respostas mais relevantes e acessíveis (Ferneda, 2012, 2021; Baeza-Yates; Ribeiro-Neto, 2013).

Em relação à conformidade e à transparência, o sistema prioriza conteúdos com referência normativa explícita (artigo, parágrafo) e *links* oficiais, garantindo transparência e rastreabilidade (Lemos; Martins; Souza, 2023; Buckland, 1991; Almeida Júnior, 2008).

Após a comparação com o modelo de recuperação, a resposta é apresentada: primeiro, uma explicação clara na linguagem do estudante; em seguida, a citação da norma correspondente, com o *link* para a fonte institucional. Se

ainda houver dúvida do estudante, o *chatbot* oferecerá novas sugestões, permitindo a reformulação da pergunta ou o encaminhamento para atendimento humano. Ao final, será verificado se a questão foi resolvida, registrando-se o resultado para melhorias futuras na base e nos fluxos, prática coerente com a curadoria informacional e com o uso de metadados para manutenção e governança de conteúdo (Gilliland, 2016; Lemos; Martins; Souza, 2023). O fluxo modelado é apresentado na Figura 16.

Figura 17 - Fluxograma



Fonte: dados da pesquisa (2025).

8.6 Recuperação da informação e aplicações da base de conhecimento

A construção da base de conhecimento, fundamentada em referenciais da CI, permite discutir sua integração com diferentes arquiteturas de recuperação da informação e plataformas de implementação.

Na perspectiva da CI, os modelos clássicos de recuperação, como o modelo booleano, o modelo vetorial e o modelo probabilístico, oferecem princípios que continuam relevantes. O modelo booleano, ao estruturar a recuperação em termos, considerando a presença ou ausência de termos, encontra paralelo na lógica de intenções e entidades definidas na base. O modelo vetorial, ao considerar pesos e graus de similaridade, aproxima-se das técnicas atuais de PLN empregadas para identificar a intenção do usuário. O modelo probabilístico, por sua vez, introduz a noção de relevância estimada, aplicável no refinamento de respostas segundo o contexto e histórico de interação (Baeza-Yates; Ribeiro-Neto, 2013; Saracevic, 2007).

Esses modelos conceituais dialogam com a CI ao evidenciarem que a RI não é apenas um processo técnico de busca, mas também um fenômeno mediado por linguagem, contexto e relevância (Ferneda, 2012; Capurro; Hjørland, 2003).

No que se refere às plataformas de implementação, destacam-se soluções tradicionais para a construção de *chatbots*, como Botpress, Rasa e Dialogflow, sendo esta última a mais frequente na revisão de literatura sobre plataformas para a construção de *chatbots*, realizada na Etapa 1 desta pesquisa, que operam com arquiteturas baseadas em *intents*, entidades e fluxos de diálogo. Essas plataformas permitem mapear diretamente os elementos curados na base de conhecimento em uma estrutura operacional, sustentada por motores de regras e classificadores de linguagem natural. A escolha por uma dessas ferramentas representa um caminho alinhado à proposta metodológica da pesquisa, já que mantém o controle sobre o repositório de conhecimento e garante transparência e auditabilidade.

Entretanto, a rápida evolução das tecnologias de IA amplia o leque de possibilidades. Arquiteturas baseadas em LLMs, como GPT (OpenAI, 2025) ou LLaMA (Large Language Model Meta AI), permitem integrar bases documentais por meio de técnicas de Retrieval Augmented Generation (RAG), nas quais o sistema consulta a base normativa e insere trechos relevantes no *prompt* do modelo antes

de gerar a resposta (Lewis *et al.*, 2020). Essa abordagem combina a generalização dos LLMs com a precisão e confiabilidade da base curada, reduzindo riscos de respostas imprecisas ou alucinações. Sua arquitetura envolve três camadas: (i) a base de conhecimento estruturada, organizada em metadados e fluxos; (ii) o mecanismo de busca ou indexação, que recupera trechos pertinentes; e (iii) o modelo gerativo, responsável por produzir a resposta em linguagem natural (Lewis *et al.*, 2020).

Dessa forma, duas possibilidades se apresentam: (a) a adoção de plataformas já consolidadas, centradas em *intents* e fluxos, como Botpress, que asseguram previsibilidade e replicabilidade; e (b) a exploração de arquiteturas híbridas, que integram recuperação estruturada com IA generativa, permitindo maior flexibilidade no diálogo. Ambas as alternativas podem ser vistas como desdobramentos da mesma lógica: a base de conhecimento, fruto da curadoria informacional, continua sendo o núcleo da recuperação e o ponto de controle do sistema.

Como perspectiva de continuidade, a base de conhecimento construída para o *chatbot* da ECI apresenta potencial de expansão em diferentes dimensões. Entre elas, destaca-se a possibilidade de incorporar regulamentos específicos dos programas de pós-graduação, de modo a atender, também, às demandas desse nível de ensino. Além disso, a integração com sistemas institucionais, como o SIGA, pode ampliar a eficiência na recuperação de informações acadêmicas. Outra direção relevante é a ampliação da base para outras unidades da UFMG, permitindo a replicação do modelo em diferentes contextos universitários. A atualização contínua e a interoperabilidade com outras plataformas acadêmicas configuram-se como condições fundamentais para garantir a relevância e a usabilidade do sistema ao longo do tempo. Nesse cenário, a iniciativa contribui para fortalecer o acesso à informação por meio de novas funcionalidades, impactando na desburocratização de processos no âmbito da universidade.

8.7 Trabalhos futuros para o desenvolvimento do *chatbot*

Com a base de conhecimento estruturada e o fluxo informacional definido, a escolha da ferramenta de desenvolvimento torna-se uma decisão estratégica

orientada por critérios técnicos, metodológicos e operacionais. Embora a definição de uma plataforma específica não seja o objetivo central desta pesquisa, foram estabelecidos parâmetros que servirão de referência para escolhas futuras, considerando o contexto institucional e informacional do projeto. Ressalta-se que este levantamento foi realizado em julho de 2025, em um cenário de rápida evolução tecnológica, o que exige revisões periódicas.

A pesquisa exploratória foi conduzida no Google, abrangendo tanto fontes primárias (*websites* e documentações oficiais das ferramentas) quanto secundárias (*reviews*, comparativos técnicos e experiências de usuários). O uso do Google como ferramenta principal justifica-se pela abrangência de resultados e pela capacidade de indexação de conteúdos técnicos atualizados, fatores essenciais em áreas de constante mudança, como plataformas de *chatbot*.

O levantamento das ferramentas seguiu os seguintes procedimentos: (1) busca exploratória no Google com termos estruturados, como “chatbot development platform”, “no-code chatbot”, “conversational AI platform” e “ferramentas chatbot gratuitas”; (2) busca complementar em repositórios técnicos, como o GitHub; (3) aplicação de critérios de inclusão, priorizando soluções ativas, com documentação disponível e possibilidade de uso gratuito ou em versão trial; e (4) exclusão de ferramentas descontinuadas ou exclusivamente pagas sem período de teste.

Para a coleta de dados, foram realizadas: (1) pesquisas específicas no Google sobre cada ferramenta; (2) análise documental de *websites* oficiais e documentação técnica; (3) consulta a fontes secundárias em *blogs* especializados, comparativos independentes e plataformas como Medium; e (4) registro sistematizado das informações em planilha eletrônica.

Em relação aos critérios de seleção estabelecidos, essa definição parte da necessidade de garantir a compatibilidade entre a solução tecnológica e o perfil da equipe responsável pela implementação, que irá envolver o profissional da informação, além de considerar recursos financeiros, que, no contexto de uma universidade pública, são escassos, e a contratação não justificada é inviável. O Quadro 12 apresenta os critérios, considerando o atual contexto da ECI UFMG.

Quadro 12 - Critérios de seleção da ferramenta *chatbot*

Critério	Descrição
Gratuito	Disponibilidade de plano gratuito com funcionalidades básicas.
No-code / Low-code	Permite uso sem programação ou com mínimo de codificação.
Integrado com IA	Possui suporte nativo ou fácil integração com LLMs.
Canais	Plataformas às quais o <i>chatbot</i> pode ser integrado (ex.: Web, WhatsApp e Instagram).
Nível técnico	Grau de conhecimento necessário para configurar e manter a ferramenta.

Fonte: elaborado pela autora (2025).

As ferramentas identificadas, conforme os critérios de seleção estabelecidos, foram: Landbot, ManyChat, Tidio, Flowise, Botpress Cloud e LangChain, apresentadas no Quadro 13 e detalhadas no APÊNDICE A.

Quadro 13 - Ferramentas para chatbot com IA

Ferramenta	Endereço	Tipo	IA integrada	Nível técnico	Custo	Canais
Landbot	https://landbot.io/	<i>No-Code</i>	Parcial (via API)	Nenhum programador	Gratuito (limitado), Pago	<i>Web (widget)</i>
ManyChat	https://manychat.com/	<i>No-Code</i>	Não nativa	Nenhum programador	Gratuito (limitado), Pago	WhatsApp, Instagram, Messenger
Tidio	https://www.tidio.com/	<i>No-Code</i>	Sim (básica)	Nenhum programador	Gratuito (limitado), Pago	<i>Web (widget), Live chat</i>
Flowise	https://flowiseai.com/	<i>Low-Code, Open Source</i>	Sim (via API)	(pouco código)	custo de hospedagem /manutenção	Flexível (via APIs)
Botpress Cloud	https://botpress.com/pt	<i>Low-Code</i>	Sim (via módulo)	(pouco código)	Gratuito (limitado), Pago	Web, WhatsApp, Messenger, APIs
LangChain	https://python.langchain.com/docs/tutorials/chatbot/	<i>Framework (Open Source)</i>	Sim (avançada)	Programador (Python)	custo de desenvolvimento /manutenção	Flexível (depende da implementação)

Fonte: elaborado pela autora (2025).

Esse levantamento inicial permitiu identificar plataformas amplamente utilizadas e acessíveis, com destaque para soluções *no-code*, *low-code* e *open source*. A seleção buscou ferramentas que atendessem aos critérios definidos, considerando custo, nível técnico exigido, integração com IA e canais de atendimento suportados.

Finalmente, é importante destacar que o principal artefato produzido nesta dissertação já está sendo utilizado para o desenvolvimento do *chatbot* da ECI UFMG, no âmbito do projeto Recuperação de Informação em modelos de Inteligência Artificial: o papel da Ciência da Informação. O lançamento da versão beta do *chatbot*, que será integrado à página da escola, está previsto para dezembro de 2025.

9. CONCLUSÕES

Este trabalho partiu de um problema concreto do cotidiano dos discentes na UFMG: a dificuldade de recuperar, de maneira rápida e facilitada, informações acadêmicas (normas, regulamentos e procedimentos), especialmente aquelas consolidadas na Resolução Complementar nº 01/2018 (NGG UFMG), a Norma Geral da Graduação da UFMG, e as variações de procedimento entre cursos. No âmbito da ECI UFMG, que reúne os cursos de Arquivologia, Biblioteconomia e Museologia, esse desafio se intensifica, tendo em vista que os procedimentos gerais são adicionados aos específicos, por curso, sendo, muitas vezes, disponibilizados em linguagem técnica e em diferentes canais.

A pesquisa situou-se nesse contexto institucional e delimitou como objeto as informações acadêmicas, com foco nos estudantes de graduação da ECI UFMG, articulando fontes normativas (NGG e resoluções correlatas), documentos operacionais, páginas institucionais (UFMG, ECI e PROGRAD), mapeamento de processos realizados pela ECI UFMG e materiais de apoio ao estudante.

Neste sentido, a presente dissertação teve como objetivo estruturar e aplicar uma base de conhecimento para subsidiar a implementação de um *chatbot* voltado à recuperação de informações acadêmicas (normas, resoluções e procedimentos) sobre as graduações na ECI UFMG. O estudo concentrou-se na sistematização e modelagem dessa base, entendida como etapa essencial para futuras implementações de sistemas conversacionais no domínio acadêmico e delineada em quatro etapas.

Na Etapa 1, o objetivo foi mapear métodos e técnicas utilizados no desenvolvimento e na construção de *chatbots* com foco na informação. Para isso, realizou-se uma revisão de literatura voltada ao mapeamento de métodos, técnicas e abordagens empregados nesses sistemas. Essa etapa possibilitou identificar artigos científicos, teses, dissertações e trabalhos apresentados em eventos relacionados ao desenvolvimento e à construção de *chatbots*. A análise permitiu organizar os estudos em cinco categorias principais. A primeira refere-se às metodologias de desenvolvimento, com destaque para os *frameworks* iterativos e colaborativos, como Scrum e Design Thinking, que estruturam o processo de construção em etapas bem definidas. A segunda corresponde aos modelos de arquitetura, entre os quais se

evidenciam os sistemas baseados em *frameworks*, como o Rasa e o uso de modelos de linguagem, como o BERT, aplicados ao PLN.

A terceira categoria abrange as linguagens de programação, com predominância do Python, utilizado pela variedade de bibliotecas de PLN disponíveis e, também, do PHP, empregado em algumas soluções voltadas à integração com aplicações *web*. A quarta diz respeito às plataformas de desenvolvimento, em que se destacam soluções como Dialogflow e Rasa, utilizadas pela disponibilidade de recursos técnicos e documentações acessíveis. Por fim, a quinta categoria refere-se às técnicas de PLN, incluindo procedimentos como TF-IDF para ranqueamento de termos, bem como estratégias de classificação de *intents* e extração de entidades.

Entre as principais conclusões, observou-se que a literatura apresenta um amplo repertório de soluções técnicas, mas ainda são limitados os trabalhos que abordam de forma consistente o tratamento informacional. Poucos estudos detalham critérios de curadoria, padronização terminológica ou atualização contínua das bases de conhecimento, aspectos fundamentais para a confiabilidade e a replicabilidade dos modelos e metodologias. Essa lacuna evidenciou a importância de integrar fundamentos da CI, especialmente a organização, representação e recuperação da informação, ao desenvolvimento de *chatbots*, servindo de referência para as etapas seguintes desta pesquisa.

Na Etapa 2, o objetivo foi identificar o comportamento de busca dos estudantes quanto a documentos normativos acadêmicos na UFMG. Para isso, foram analisados os resultados de um questionário aplicado no âmbito do projeto Recuperação de Informação em Modelos de Inteligência Artificial: o papel da Ciência da Informação, com foco no acesso a normas, regimentos e resoluções (Nascimento Silva; Soares, 2025). A coleta, realizada entre dezembro de 2024 e março de 2025, buscou compreender como os discentes procuram esse tipo de informação, quais dificuldades encontram no processo e quais documentos geram maior número de dúvidas. Os resultados mostraram que os estudantes de graduação enfrentam obstáculos recorrentes, como a dispersão das fontes de informação, a complexidade da linguagem normativa e a ausência de um canal centralizado de acesso. O levantamento indicou, ainda, que, em sua maioria, os discentes recorrem, inicialmente, a buscas em *sites* institucionais e, diante de dificuldades, complementam a procura com redes informais, como colegas e grupos em redes

sociais. As principais dúvidas concentram-se em processos como trancamento, matrícula e desligamento, seguidas por questões sobre prazos e integralização curricular. Como sugestão, os estudantes destacaram a importância de criar um meio unificado, claro e de fácil acesso, capaz de reunir as normas e os procedimentos acadêmicos em uma linguagem mais objetiva e compreensível. Esses resultados permitiram identificar práticas recorrentes e obstáculos enfrentados pelos estudantes, originando as categorias e os conceitos que orientaram a construção da base de conhecimento.

Na Etapa 3, o objetivo foi identificar e analisar *chatbots* já implementados no contexto educacional. Para isso, foram examinados 26 estudos sobre *chatbots* educacionais, previamente identificados na revisão de literatura (Etapa 1), dos quais cinco apresentaram modelos documentados e replicáveis. A análise dos *chatbots* educacionais evidenciou fluxos informacionais bem estruturados e metodologias claramente documentadas, além de destacar boas práticas na construção das bases de conhecimento, que se mostraram fundamentais para a organização dos conteúdos de maneira acessível e consistente. Esses elementos indicaram a necessidade de construir uma base de conhecimento própria para a realidade da ECI UFMG, capaz de reunir conceitos normativos, intenções de busca e respostas articuladas em um repositório estruturado. A partir desses resultados, o objetivo geral da dissertação, delineado pela construção de uma base de conhecimento, foi validado por outras experiências, como boas práticas para a construção de *chatbots*.

Na Etapa 4, o objetivo foi realizar a curadoria informacional para subsidiar a implementação de uma base de conhecimento voltada à recuperação de informações acadêmicas, ou seja, normas, resoluções e procedimentos sobre as graduações na ECI UFMG. Essa etapa consistiu em um processo de curadoria informacional, no qual foi criada uma base estruturada em perguntas e respostas, contemplando tópicos prioritários baseado nas principais dúvidas dos estudantes, identificadas na Etapa 2, com o tratamento terminológico e o uso de termos de vocabulários controlados, metadados para a descrição dos elementos e estrutura taxonômica hierárquicos para apoiar a navegação. A integração dos resultados das etapas anteriores da pesquisa possibilitou a construção de uma solução informacional coerente com as necessidades dos estudantes, alinhada às normas

institucionais e com potencial de aplicação futura em diversos sistemas, especialmente para o desenvolvimento de um *chatbot*.

Para sua construção da base, inicialmente, foi proposto um modelo de curadoria informacional, dividido em quatro fases, aplicável a normas acadêmicas, envolvendo critérios de seleção de fontes, padronização terminológica, tradução da linguagem técnica para a linguagem de uso e localização das repostas nas fontes utilizadas, indicando a vinculação a artigos, parágrafos, em elementos normativos e *links* de documentos, oferecendo um procedimento replicável em outros contextos institucionais.

Em relação à padronização terminológica, os fluxos hierárquicos, modelados por meio de uma taxonomia, estão alinhados ao BRASED, contemplando remissivas e termos coloquiais mapeados no questionário, de modo a consolidar a mediação terminológica entre a linguagem institucional e a fala dos estudantes. Além disso, elaborou-se um fluxo conversacional padrão, baseado em regras e checagens condicionais e no encaminhamento humano, quando necessário, de forma a orientar a implementação do sistema em diferentes plataformas.

Entre julho e setembro de 2025, efetuou-se a inserção dos dados na base e a sua validação, com o registro de inconsistências, ajustes necessários e a incorporação de novas intenções. Um fluxo padrão também foi modelado para indicar caminhos de implementação. Por fim, para apoiar trabalhos futuros, foram indicadas plataformas de implementação, bem como os requisitos para a integração com os canais institucionais, criando as condições necessárias para a aplicação prática do *chatbot*.

A base de conhecimento foi estruturada a partir de 31 conceitos, desdobrados em 225 perguntas, com as respectivas respostas, disponibilizados em uma artefato *online*, que será disponibilizado com acesso aberto. Em conjunto com essas entregas, a base é posicionada como elemento estratégico para a recuperação da informação acadêmica e fornece parâmetros técnicos e metodológicos para futuras iniciativas de sistemas conversacionais, como *chatbots* educacionais, na ECI UFMG e em contextos afins.

Dessa forma, a pergunta de pesquisa foi respondida ser evidenciado que a curadoria informacional apoia a recuperação de informações acadêmicas normativas ao transformar documentos dispersos em uma base estruturada, transparente e

acessível, capaz de subsidiar a construção de *chatbots* em universidades públicas e de aproximar a linguagem normativa da linguagem utilizada pelos estudantes.

9.1 Contribuições da dissertação

No âmbito da CI, a dissertação contribui ao destacar a aplicação de conceitos clássicos da área, como organização e representação da informação, vocabulários controlados e RI, à construção de um artefato tecnológico concreto, por meio da curadoria informacional. Além disso, o estudo evidencia práticas estratégicas para a organização e o acesso à informação em contextos digitais, especialmente em relação às questões terminológicas, para reduzir barreiras entre a linguagem normativa e a linguagem utilizada pelos estudantes. Por fim, demonstra-se como a CI pode dialogar de forma produtiva com campos tecnológicos, como assistentes conversacionais e *chatbots*, e, também, com o uso da IA, oferecendo fundamentação teórica e metodológica para o desenvolvimento de soluções inovadoras em ambientes acadêmicos. Dessa maneira, a pesquisa contribui tanto de forma prática, ao apresentar um estudo de caso que pode ser replicável em universidades públicas, quanto de forma teórica, ao ampliar a compreensão sobre o papel da CI no acesso e na recuperação de informações e no apoio à transformação digital em instituições de ensino superior, promovendo impacto social.

9.2 Desafios e limitações

A pesquisa identificou desafios e limitações como a atualização constante das normas e dos documentos institucionais, que se apresenta como um fator crítico para a manutenção da relevância e da precisão da base de conhecimento.

Outros pontos de destaque foram a fragmentação e a complexidade da documentação normativa, muitas vezes dispersa em diferentes fontes e formatos, o que exigiu esforço adicional para a seleção, padronização e organização.

Também se observou a limitação de não se validar a base com usuários finais, o que restringiu a análise de sua eficácia no uso cotidiano. Além disso, as validações internas envolvendo a orientadora podem ser expandidas para incluir

outros atores. Por fim, embora a metodologia proposta seja potencialmente replicável, sua aplicação fora do contexto específico da graduação da ECI UFMG pode demandar ajustes e adaptações.

9.3 Trabalhos futuros

Como desdobramento desta pesquisa, sugere-se a expansão e aplicação da base de conhecimento em outros cursos da UFMG. Embora o estudo de caso tenha se concentrado na ECI UFMG, a curadoria realizada contemplou normas acadêmicas gerais, como a Resolução Complementar 01/2018 (NGG) e resoluções complementares, que possuem validade institucional para todos os cursos de graduação da universidade. Dessa forma, a base construída pode ser adaptada e replicada em diferentes contextos acadêmicos, ampliando seu alcance e impacto. Além disso, futuros trabalhos podem explorar a integração da base de conhecimento com sistemas institucionais já existentes, realizar testes de usabilidade com estudantes e docentes e investigar sua aplicação também em outros níveis de ensino, como a pós-graduação.

9.4 Considerações finais

O papel do profissional da informação se mostrou central na construção da base de conhecimento, atuando como um curador qualificado entre o usuário e os conteúdos normativos, subsidiando a implementação de inovações tecnológicas. Assim, a pesquisa evidencia que, embora a tecnologia seja um recurso essencial, ela precisa ser complementada com a expertise do profissional da informação, que irá trabalhar na organização, no tratamento e na contextualização da informação. Observa-se que a automação é particularmente eficaz para lidar com demandas repetitivas, garantindo rapidez e precisão, mas não substitui dimensões humanas como empatia, julgamento ético e criatividade, indispensáveis no atendimento acadêmico. Com isso, o equilíbrio deve ser mantido para que os benefícios de *chatbots* de atendimento, como a desburocratização e a agilidade no atendimento, tragam inovação e impacto social e não gerem novas barreiras de acesso e riscos à privacidade e à proteção de dados.

REFERÊNCIAS

- ADAMOPOULOU, E.; MOUSSIADES, L. Chatbots: History, Technology, and Applications. **Machine Learning with Applications**, v. 2, 100006, 2020. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2666827020300062>. Acesso em: 19 mai. 2024.
- AGRAWAL, R. *et al.* WASABI Contextual BOT. **International Journal of Intelligent Systems**, v. 37, n. 8, p. 5234-5251, 2022. Disponível em: <https://ieeexplore.ieee.org/document/9989007>. Acesso em: 19 mai. 2024.
- AGUS SANTOSO, H. *et al.* Dinus Intelligent Assistance (DINA) Chatbot for University Admission Services. In: INTERNATIONAL CONFERENCE ON INFORMATION AND COMMUNICATIONS TECHNOLOGY, 2018. **Anais [...]**. [S.l.: s.n.], 2018. Disponível em: <https://ieeexplore.ieee.org/document/8549797>. Acesso em: 19 mai. 2024.
- ALEIXO, M. R. *et al.* O papel do gestor e curador da informação nos novos comportamentos informacionais. **Perspectivas Em Ciência Da Informação**, v. 25, n. 3, p. 49–62, 2020. Disponível em: <https://doi.org/10.1590/1981-5344/3841>. Acesso em: 25 set. 2024
- ALEKSEEV D., SHAGALOVA P., SOKOLOVA E. **Development of a Chatbot Using Machine Learning Algorithms to Automate Educational Processes**. Proceedings of the International Conference on Computer Graphics and Vision “Graphicon”, CEUR”. 2021, p. 1104-1113. Disponível em: https://library.keldysh.ru/prep_vw.asp?lg=e&pid=9361. Acesso em: 9 out. 2025
- ALLISON, D. Chatbots in the Library: Is It Time? **UNL Libraries: Faculty Publications**, 2012. Disponível em: <https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1294&context=libraryscience>. Acesso em: 9 ago. 2025.
- ALMEIDA JÚNIOR, O. F. de. Mediação da informação: ampliando o conceito de disseminação. In: VALENTIM, Marta Lígia Pomim (Org.). **Gestão da informação e do conhecimento**. São Paulo: Polis; Cultura Acadêmica, 2008. p. 41-54.
- ALMEIDA JÚNIOR, O. F. de. Mediação da informação: um conceito atualizado. In: BORTOLIN, Sueli; SANTOS NETO, João Arlindo dos; SILVA, Rovilson José da (Org.). **Mediação oral da informação e da leitura**. Londrina: ABECIN, 2015. 278p. p.9-32.
- ALQAIDI, S.; ALHARBI, W.; ALMATRAFI, O. A support system for formal college students: A case study of a community-based app augmented with a chatbot. **Educational Technology & Society**, v. 24, n. 4, p. 156-171, 2021. Disponível em: <https://ieeexplore.ieee.org/document/9759796>. Acesso em: 19 maio 2024.
- ÁLVARES, L. Organização da Informação e do conhecimento: conceitos, subsídios interdisciplinares e aplicações. **Revista Brasileira de Biblioteconomia e Documentação**, v. 8, n. 2, p. 205–206, 2013. Disponível em: <https://rbbd.febab.org.br/rbbd/article/view/245>. Acesso em: 24 jun. 2025.

ALVES, A. P. S. *et al.* Development and evaluation of a chatbot for the Regional Museum of São João del-Rei. **Centro Latinoamericano de Estudios en Informática**. 2018. Disponível em: <https://clei.org/clei2018/docs/SLIS/182920.pdf>. Acesso em: 8 ago. 2025

ASSOCIAÇÃO NACIONAL DOS DIRIGENTES DAS INSTITUIÇÕES FEDERAIS DE ENSINO SUPERIOR (Andifes). “Lei Orgânica das Universidades Públicas Federais”. 2002. Disponível em: www.ant.cfh.ufsc.br/files/2012/04/lou.pdf. Acesso em: 29 jan. 2025.

ATTIGERI, G.; AGRAWAL, A.; KOLEKAR, S. Advanced NLP Models for Technical University Information Chatbots: Development and Comparative Analysis. **IEEE Access**, 2024. Disponível em: <https://doi.org/10.1109/ACCESS.2024.3368382>. Acesso em: 19 mai. 2024.

AUSTIN, John L. **How to do Things with words**. New York: Oxford University Press, 1965.

BAEZA-YATES, R.; RIBEIRO-NETO, B. **Modern Information Retrieval**. New York: ACM Press, 1999.

BAEZA-YATES, R.; RIBEIRO-NETO, B. **Recuperação de Informação: conceitos e tecnologia das máquinas de busca**. Porto Alegre: Bookman, 2013.

BAIZAL, Z. A. *et al.* Movie recommender chatbot based on Dialogflow. **Entertainment Computing**, v. 45, p. 100-115, 2023. Disponível em: <https://ijece.iaescore.com/index.php/IJECE/article/view/26761>. Acesso em: 19 mai. 2024.

BARBOSA, C. A era da curadoria. **Museologia & Interdisciplinaridade**, [S. l.], v. 2, n. 4, 2013. Disponível em: <https://periodicos.unb.br/index.php/museologia/article/view/16370>. Acesso em: 1 ago. 2025.

BARDIN, L. **Análise de Conteúdo**. 4. ed. São Paulo: Edições 70, 2016.

BARRETO, I. *et al.* Development and evaluation of the GISSA Mother-Baby ChatBot application in promoting child health. **Ciência & Saúde Coletiva**, v. 26, n. 5, p. 1679-1690, 2021. Disponível em: <https://www.scielo.br/j/csc/a/HGwWTZGbC5zfHXsTqNFkYnM/?lang=en>. Acesso em: 19 mai. 2024.

BATES, M. J. Toward an integrated model of information seeking and searching. **New Review of Information Behaviour Research**, v. 3, p. 1-15, 2002. Disponível em: https://pages.gseis.ucla.edu/faculty/bates/articles/info_SeekSearch-i-030329.html. Acesso em: 19 mai. 2024.

BATES, M. J. Information and Knowledge: An Evolutionary Framework for Information Science. **Information Research: An International Electronic Journal**, v. 10, n. 4, 2005. Disponível em: <http://files.eric.ed.gov/fulltext/EJ1082014.pdf>. Acesso em: 19 mai. 2024.

BATES, M. J. Information behavior. **Encyclopedia of Library and Information Sciences**. [S.l]: Taylor & Francis, 2017.

BECK, K. *et al.* Manifesto ágil para o desenvolvimento de software. 2001. Disponível em: <https://agilemanifesto.org/iso/ptbr/manifesto.html>. Acesso em: 19 mai. 2024.

BELKIN, N. J.; CROFT, W. B. Information filtering and information retrieval: Two sides of the same coin? *Communications of the ACM*, v. 35, n. 12, p. 29–38, 1992. Disponível em: <https://dl.acm.org/doi/10.1145/138859.138861> Acesso em: 19 maio 2024.

BENDOTTI, H. *et al.* Introducing Quin: The Design and Development of a Prototype Chatbot to Support Smoking Cessation. **Digital Health**, v. 9, p. 1-15, 2023. Disponível em: <https://academic.oup.com/ntr/article/26/5/612/7343970>. Acesso em: 19 mai. 2024.

BENGIO, Y. *et al.* A Neural Probabilistic Language Model. **Journal of Machine Learning Research**, v. 3, p.1137–1155, 2003. Disponível em: <http://www.jmlr.org/papers/volume3/bengio03a/bengio03a.pdf>. Acesso em: 19 mai. 2024.

BENOTTI, L.; MARTINEZ, M. C.; SCHAPACHNIK, F. A tool for introducing computer science with automatic formative assessment. **IEEE Transactions on Learning Technologies**, v. 11, n. 2, p. 179-192, 2017. Disponível em: <https://eric.ed.gov/?id=EJ1183012>. Acesso em: 19 mai. 2024.

BOMMASANI, R. *et al.* On the Opportunities and Risks of Foundation Models. arXiv preprint arXiv:2108.07258, 2022. Disponível em: <https://arxiv.org/abs/2108.07258>. Acesso em: 25 set. 2024.

BORKO, H. Information science: what is it? **American Documentation**, v. 19, n. 1, p. 3-5, 1968. Disponível em: <https://bit.ly/2DLQfkL>. Acesso em: 19 mai. 2024.

BRANDTZÆG, P. B., FØLSTAD, M. **Chatbots**: Changing user needs and motivations. **Interactions**, v. 24, n. 5, p.38-43, 2017. Disponível em: <https://dl.acm.org/doi/10.1145/3236669>. Acesso em: 19 mai. 2024.

BOTPRESS. **Plataforma de criação de chatbots e agentes de IA**. Disponível em: <https://botpress.com/pt>. Acesso em: 14 jun. 2025.

BRASCHER, M.; CAFÉ, L. Organização da informação ou organização do conhecimento? In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 9, 2008, São Paulo. **Anais [...]**. São Paulo: ANCIB, USP, 2008. p. 1-14. Disponível em: <https://skat.ihmc.us/rid=1KR7TM7S9-S3HDKP-5STP/BRASCHERCAF%C3%89>. Acesso em: 19 mai. 2024.

BUCKLAND, M. K. Information as thing. **Journal of the American Society for Information Science**, v. 42, n. 5, p. 351–360, 1991. Disponível em:

[https://asistdl.onlinelibrary.wiley.com/doi/10.1002/\(SICI\)1097-4571\(199106\)42:5%3C351::AID-ASI5%3E3.0.CO;2-3](https://asistdl.onlinelibrary.wiley.com/doi/10.1002/(SICI)1097-4571(199106)42:5%3C351::AID-ASI5%3E3.0.CO;2-3). Acesso em: 19 mai. 2024.

CAMPOS, R. **Vocabulário Controlado**: Teoria e Prática. São Paulo: Perspectiva, 2001.

CAPURRO, R.; HJORLAND, B. **The concept of information**. *Annual Review of Information Science and Technology*, v. 37, p. 343–411, 2003. Disponível em: <https://www.scielo.br/j/pci/a/j7936SHkZJkpHGH5ZNYQXnC/?format=pdf&lang=pt.1440370120>. Acesso em: 6 jun. 2025.

CARVALHO, R. C. de. **Chatbot aplicado à recuperação de informação**: um modelo orientado a metadados. Tese (doutorado) 2022. 214 f. Universidade Estadual Paulista (Unesp), Faculdade de Filosofia e Ciências, Marília. Disponível em: <https://repositorio.unesp.br/items/4aa1e4a4-aaa1-4651-ae4e-77c174cb448a>. Acesso em: 6 jun. 2025.

CASE, D. O.; GIVEN, L. M. **Looking for Information**: A Survey of Research on Information Seeking, Needs, and Behavior. 4. ed. Bingley, UK: Emerald Group Publishing, 2016.

CASELI, H. M.; NUNES, M. G. V. (org.). **Processamento de Linguagem Natural**: Conceitos, Técnicas e Aplicações em Português. 2 ed. São Carlos: BPLN, 2024. Disponível em: <https://brasileiraspln.com/livro-pln/2a-edicao>. Acesso em: 6 jun. 2025.

CASTELLS, M. O espaço de fluxos. *In*: CASTELLS, M. **A sociedade em rede**. São Paulo: Paz e Terra, 2000. p. 467-521.

CHAN, C. H. *et al.* Developing a Chatbot for College Student Programme Advisement. *In*: INTERNATIONAL CONFERENCE ON COMPUTER SCIENCE AND EDUCATION, 2018. **Anais [...]**. [S.l.: s.n.], 2018. Disponível em: <https://ieeexplore.ieee.org/document/8456189> Acesso em: 6 jun. 2025.

CHASE, M. Academic Libraries Can Develop AI Chatbots for Virtual Reference Services with Minimal Technical Knowledge and Limited Resources. **Evidence Based Library and Information Practice**, [S. l.], v. 19, n. 2, p. 136–138, 2024. Disponível em: <https://journals.library.ualberta.ca/ebliip/index.php/EBLIP/article/view/30523>. Acesso em: 6 jun. 2025.

CHEN, L.; CHEN, P.; LIN, Z. Artificial intelligence in education: A review. **IEEE Access**, v. 8, p. 75264-75278, 2020. Disponível em: <https://www.scirp.org/reference/referencespapers?referenceid=3599963>. Acesso em: 6 jun. 2025.

CHOW, J. C. L.; SANDERS, L.; LI, K. Design of an Educational Chatbot Using Artificial Intelligence in Radiotherapy. **AI**, v. 4, n. 1, p. 319-332, 2023. Disponível em: <https://doi.org/10.3390/ai4010015>. Acesso em: 9 ago. 2025.

CHUMKAEW, S. The development of chatbot provided registration information services for students in distance learning. **Distance Education Technology**

Journal, v. 15, n. 2, p. 112-128, 2023. Disponível em: <https://assumptionjournal.au.edu/index.php/abacjournal/article/view/7730/3858>. Acesso em: 6 jun. 2025.

CLARIZIA, F. *et al.* Chatbot: An education support system for student. In: **International Symposium on Cyberspace Safety and Security**. Springer, 2018.

CLEVELAND, D.; CLEVELAND, A. **Introduction to indexing and abstracting**. 3. ed. Colorado: A division of Greenwood Publishing Group, 2001.

COMARELLA, R. L.; CAFÉ, L. M. Chatterbot: conceito, características, tipologia e construção. **Informação & Sociedade**, v. 18, n. 2, 2008. Disponível em: <https://www.proquest.com/scholarly-journals/chatterbot-conceito-caracteristicas-tipologia-e/docview/1493901212/se-2>. Acesso em: 6 jun. 2025.

COMBELLES, A.; EBERT, C.; LUCENA, P. Design Thinking. **IEEE Software**, v. 37, n. 2, p. 21-24, 2020. Disponível em: <https://doi.org/10.1109/MS.2019.2959328>. Acesso em: 6 jun. 2025.

CONEGLIAN, C. S.; SEGUNDO, J. E. S. Inteligência artificial e ferramentas da web semântica aplicadas à recuperação da informação: um modelo conceitual com foco na linguagem natural. **Informação & Informação**, v. 27, n. 1, p. 625–651, 2022. Disponível em: <https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/44729>. Acesso em: 6 jun. 2025.

CONEGLIAN, C. S.; GONÇALEZ, P. R.; SANTARÉM SEGUNDO, J. E. O Profissional da Informação na Era do Big Data. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis/SC, Brasil, v. 22, n. 50, p. 128–143, 2017. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2017v22n50p128>. Acesso em: 6 jun. 2025

CRESWELL, J. W. Research design: qualitative, quantitative, and mixed methods approaches. 4. ed. **Thousand Oaks**: SAGE Publications, 2013.

CURRÁS, E. **Ontologias, taxonomia e tesouros em teoria de sistemas e sistemática**. Tradução de Jaime Robredo. Brasília: Thesaurus, 2010.

CUNHA, M. B. da; CAVALCANTI, C. R. de O. **Dicionário de biblioteconomia e arquivologia**. Brasília: Briquet de Lemos, 2008.

DAHLBERG, I. A referent-oriented, analytical concept theory for interconcept. **International Classification**, v. 5, n. 3, p. 142-151, 1978. Disponível em: <https://www.imrpress.com/journal/ko/5/3/10.5771/0943-7444-1978-3-142>. Acesso em: 26 mai. 2025.

DAHLBERG, L. Knowledge organization: its scope and possibilities. **Knowledge Organization** v. 20, n. 4, 1993, p. 211-222.

DAHLBERG, I. Knowledge organization: a new science? **Knowledge Organization**, Frankfurt, v. 33, n. 1, p. 11-19, 2006. Disponível em: <https://article.imrpess.com/journal/KO/33/1/10.5771/0943-7444-2006-1-11/8c20111318736b459116aea37c06f56a.pdf>. Acesso em: 26 mai. 2025.

DANIEL, G. *et al.* Multi-platform Chatbot Modeling and Deployment with the Jarvis Framework. **International Conference on Advanced Information Systems Engineering**, 2019. Disponível em: <https://modeling-languages.com/wp-content/uploads/2019/03/JarvisCaise2019.pdf>. Acesso em: 26 mai. 2025.

DE-LA-TORRE-UGARTE-GUANILO, M. C.; TAKAHASHI, R. F.; BERTOLOZZI, M. R. Revisão sistemática: noções gerais. **Revista da Escola de Enfermagem da USP**, v. 45, n. 5, p. 1260–1266, out. 2011. Disponível em: <https://www.scielo.br/j/reeusp/a/CRjvBKKvRRGL7vGsZLQ8bQj/abstract/?lang=pt>. Acesso em: 26 mai. 2025.

DEL SOLE, D. *et al.* Papel da curadoria de conteúdo na implantação de chatbots em serviços públicos. **Revista Gestão Do Conhecimento e Tecnologia da Informação**, v. 8, n. 1, p. 31-46, 2025. Disponível em: <https://doi.org/10.31501/rgcti.v8i1.15087>. Acesso em: 26 maio. 2025.

DEWEY, M. **Classification and Subject Index for Cataloguing and Arranging the Books and Pamphlets of a Library**. [S.l.: s.n.], 1876.

DUBLIN CORE METADATA INITIATIVE. **Dublin Core Metadata Element Set**, Version 1.1. 2020. Disponível em: <https://www.dublincore.org/specifications/dublin-core/dces/>. Acesso em: 26 mai. 2025.

ELSAYED, H. S. *et al.* Chatbot as a Virtual Assistant to Retrieve Information from Datasheets Using Memory Controllers Domain Knowledge. **Computer Systems Science and Engineering**, v. 46, n. 3, p. 3245-3260, 2023. Disponível em: <https://ieeexplore.ieee.org/document/10382771>. Acesso em: 9 mai. 2024

EREKATA, O. *et al.* Development of a Text and Speech Enabled Conversational Agent for Students' Activities Planning Using Dialog Flow. **Educational Technology Research and Development**, v. 69, n. 3, p. 1453-1470, 2021. https://link.springer.com/chapter/10.1007/978-981-16-1480-4_44. Acesso em: 9 mai. 2024.

FEITOSA, W. R. *et al.* Proposal for a Chatbot Prototype in the Plant Health Department of Brazilian Ministry of Agriculture. **Computers and Electronics in Agriculture**, v. 175, p. 105-118, 2020. Disponível em: <https://ieeexplore.ieee.org/document/9311048>. Acesso em: 9 maio 2024

FEITOZA, F. R. **Cecílio**: um chatbot para automação do atendimento aos usuários nas Instituições Federais de Ensino Superior. Dissertação (Mestrado em Ciência da Computação) - Universidade Federal da Paraíba, João Pessoa, 2021. Disponível

em: https://repositorio.ufpb.br/jspui/handle/123456789/21943?locale=pt_BR. Acesso em: 12 out. 2025.

FERNEDA, E. **Recuperação de Informação**: análise sobre a contribuição da Ciência da Computação para a Ciência da Informação. Tese. Universidade de São Paulo. Escola de Comunicação e Artes, São Paulo, 2003.

FERNEDA, E. **Introdução aos modelos computacionais de recuperação de informação**. Rio de Janeiro: Ciência Moderna, 2012.

FERNEDA, E. Recuperação de Informação: a representação do subjetivo. *In*: JORENTE, M *et al.* (Orgs.) **Curadoria Digital e Gênero na Ciência da Informação**. Marília: Oficina Universitária; São Paulo: Cultura Acadêmica, 2021. p. 193-209. Disponível em: https://ebooks.marilia.unesp.br/index.php/lab_editorial/catalog/view/273/2849/4956. Acesso em: 12 out. 2025.

FLOWISE. **FlowiseAI**: Open source low-code tool for developers to build customized LLM orchestration flow & AI agents. Disponível em: <https://flowiseai.com/>. Acesso em: 14 jun. 2025.

FOOSHERIAN, M. *et al.* Enhancing Pipeline-Based Conversational Agents with Large Language Models. **AI Communications**, v. 36, n. 2, p. 89-108, 2023. Disponível em: <https://arxiv.org/abs/2309.03748> Acesso em: 9 maio 2024

FOSKETT, A. C. **The subject approach to information**. 5. ed. London: Library Association Publishing, 1996.

FOWLER JR., Floyd J. **Pesquisa de levantamento**: métodos práticos para coleta de dados de pessoas. 4. ed. São Paulo: Bookman, 2011.

FUJITA, M. S. **A política de indexação para representação e recuperação da informação**. *In*: LEIVA, I. G., FUJITA, M. S. L., eds. Política de indexação [online]. São Paulo: Cultura Acadêmica; Marília: Oficina Universitária, 2012, p. 17-28. Disponível em: <https://books.scielo.org/id/dk5kd/pdf/leiva-9786559541300-02.pdf>. Acesso em: 9 mai. 2024.

GILLILAND, A. J. **Setting the stage**. *In*: Baca, M. Introduction to metadata 3rd. ed. Los Angeles: Getty Publications, 2016. Disponível em: <http://www.getty.edu/publications/intrometadata/setting-the-stage/>.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002.

GIL, A. C. **Como elaborar projetos de pesquisa**. 6. ed. São Paulo: Atlas, 2017.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. Massachusetts: MIT Press, 2016. Disponível em: <https://books.google.com.br/books?id=Np9SDQAAQBAJ&printsec=copyright&hl=pt-BR#v=onepage&q&f=false>. Acesso em: 8 set. 2025.

GOOGLE. **Google Assistant**. 2016. Disponível em: <https://assistant.google.com/>. Acesso em: 22 abr. 2025.

GOWDA, C. P. *et al.* Development of Information Technology Telecom Chatbot: An Artificial Intelligence and Machine Learning Approach. **2021 2nd International Conference on Intelligent Engineering and Management (ICIEM)**, London, United Kingdom, 2021, p. 216-221, doi: 10.1109/ICIEM51511.2021.9445354. Disponível em: <https://ieeexplore.ieee.org/document/9445354?denied=>. Acesso em: 8 ago. 2025.

GRICE, H. P. Logic and conversation. In: COLE, Peter; MORGAN, Jerry L. (Ed.). **Syntax and semantics 3: speech acts**. New York: Academic Press, 1975. p. 41-58.

GRIOL, D. *et al.* A Statistical Approach to Spoken Dialog Systems Design and Evaluation. **Speech Communication**, v. 50, n. 8-9, p. 666, 2008. Disponível em: <https://hal.science/hal-00499213/document> Acesso em: 8 ago. 2024.

GUIMARÃES, L. J., ROCHA, E. C. de F. Práticas informacionais e design thinking Abordando usuários 3.0 na Ciência da Informação. **RDBCI: Revista Digital De Biblioteconomia E Ciência Da Informação**, v. 19, e021029, 2021. Disponível em: <https://doi.org/10.20396/rdbci.v19i00.8666871>. Acesso em: 8 ago. 2024.

GUIMARÃES, L. J. **Chatbot em contexto**: design de experiência do usuário aplicado à recuperação da informação no catálogo de teses e dissertações da CAPES. Tese de Doutorado, 323 f. 2022. Universidade Federal de Minas Gerais, Escola de Ciência da Informação, Belo Horizonte.

GUERRA GONZÁLEZ, J. T. El bibliotecario académico universitario como curador de contenidos digitales: precisiones conceptuales y prácticas. **Biblioteca Universitaria**, Distrito Federal, México, vol. 20, núm. 2, julio-diciembre, 2017, p. 94-107. Disponível em: <https://www.redalyc.org/articulo.oa?id=28553811003>. Acesso em: 20 nov. 2024.

GUNAWAN *et al.* **Development of Intelligent Telegram Chatbot Using Natural Language Processing**. 1-5. 2021. Disponível em: <https://ieeexplore.ieee.org/document/9678471>. Acesso em: 12 out. 2025.

HIRSCHBERG, J.; MANNING, C. D. Advances in natural language processing. **Science**, v. 349, n. 6245, p. 261-266, 2015. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/26185244/>. Acesso em: 8 ago. 2024.

HJORLAND, B. Terminology and knowledge organization. **Knowledge Organization**, v. 44, n. 1, p. 1-16, 2016. Disponível em: https://www.isko.org/cyclo/knowledge_organization. Acesso em: 8 ago. 2024.

HODGE, G. **Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files**. Washington: Council on Library and Information Resources, 2000. Disponível em: <https://www.clir.org/pubs/reports/pub91/>. Acesso em: 8 ago. 2024.

HNATUSHENKO, V.; OSTROVSKA, K., NOSOV, V. Development and Research of a Chatbot Using the Linguistic Core of Amazon Lex V2. **Conference Intelligent**

Systems Workshop at CoLInS, 2024. Disponível em: <https://ceur-ws.org/Vol-3688/paper5.pdf>. Acesso em: 8 ago. 2025.

HUANG, M.; ZHU, X.; GAO, J. Challenges in building intelligent open-domain dialog systems. **ACM Transactions on Information Systems**, v. 38, n. 3, p. 1-32, 2020. Disponível em: <https://dl.acm.org/doi/10.1145/3383123>. Acesso em: 8 ago. 2024.

HYRMET, M.; ARBANA, K. The Impact of Chatbots in Driving Digital Transformation. **International Journal of E-services and Mobile Applications**, v.13, n.4, p. 88-104, 2021. Disponível em: <https://ideas.repec.org/a/igg/jesma0/v13y2021i4p88-104.html>. Acesso em: 30 set. 2024.

IBM. **Seis tipos de chatbots e como escolher o ideal para sua empresa**. Disponível em: <https://www.ibm.com/br-pt/think/topics/chatbot-types>. Acesso em: 30 jul. 2025.

INGWERSEN, P.; JARVELIN, K. **The Turn: Integration of Information Seeking and Retrieval in Context**. New York: Springer, 2005.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Thesaurus Brasileiro da Educação (Brased)**. Brasília: INEP, 2025. Disponível em: <https://vocabularyserver.com/brased/>. Acesso em: 8 ago. 2025.

IPARRAGUIRRE-VILLANUEVA, O. *et al.* Agente inteligente para la gestión de incidencias. **RISTI**, Porto, n. 51, p. 99-115, set. 2023. Disponível em: http://scielo.pt/scielo.php?script=sci_arttext&pid=S1646-98952023000300099&lng=pt&nrm=iso. Acesso em: 11 jul. 2025.

ISMAIL, M.; ADE-IBIJOLA, A. Lecturer's apprentice: A chatbot for assisting novice programmers. In: **International Multidisciplinary Information Technology and Engineering Conference (IMITEC)**, 2019, p. 1-8. Disponível em: https://www.researchgate.net/publication/339592677_Lecturer's_Apprentice_A_Chatbot_for_Assisting_Novice_Programmers. Acesso em: 8 ago. 2024.

JIAO, A. R. An Intelligent Chatbot System Based on Entity Extraction Using RASA NLU and Neural Network. **Journal of Physics: Conference Series**, v. 1631, n. 1, p. 012-025, 2020. Disponível em: <https://iopscience.iop.org/article/10.1088/1742-6596/1487/1/012014>

JORENTE, M. J. V., PADUA, M. C., NAKANO, N. (2019). O design da informação como recurso interdisciplinar da curadoria digital em contextos complexos da Ciência da Informação. **Perspectivas em Ciência da Informação**, v. 24, n. 3, p.35–58, 2019. Disponível em: <https://doi.org/10.1590/1981-5344/3127>. Acesso em: 6 jun. 2025.

JURAFSKY, D.; MARTIN, J. H. **Speech and Language Processing**. 3. ed. Cambridge, MA: MIT press, 2023. Disponível em: <https://web.stanford.edu/~jurafsky/slp3/>. Acesso em: 8 ago. 2024.

JURAFSKY, D.; MARTIN, J. H. **Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition**. 2.

ed. Upper Saddle River: Prentice Hall, 2009. Disponível em: <https://web.stanford.edu/~jurafsky/slp3/>. Acesso em: 24 set. 2024.

KARRI, S. P. R.; KUMAR, B. S. Deep Learning Techniques for Implementation of Chatbots. In: **IEEE INTERNATIONAL CONFERENCE ON COMPUTING, COMMUNICATIONS AND NETWORKING TECHNOLOGIES**, 2020. Proceedings [...]. [S.l.]: IEEE, 2020. Disponível em: <https://ieeexplore.ieee.org/document/9104143>. Acesso em: 8 ago. 2024.

KINGCHANG, T.; CHATWATTANA, P.; WANNAPIROON, P. Artificial Intelligence Chatbot Platform: AI Chatbot Platform for Educational Recommendations in Higher Education. **Educational Technology International**, v. 25, n. 3, p. 234-251, 2024. Disponível em: <https://www.ijiet.org/vol14/IJiet-V14N1-2021.pdf>. Acesso em: 8 ago. 2024.

KITCHENHAM, B. Procedures for performing systematic reviews. **Keele University Joint Technical Report**, Staffordshire, v. 33, n. 2004, p. 1-26, 2004. Disponível em: <https://www.inf.ufsc.br/~aldo.vw/kitchenham.pdf>. Acesso em: 15 mar. 2024.

KUHLTHAU, C. C. The Information Search Process Revisited: Is the Model Still Useful? **Information Research**, v. 13, n. 4, 2004. Disponível em: <https://informationr.net/ir/13-4/paper355.html>. Acesso em: 8 ago. 2024.

LABADZE, L.; GRIGOLIA, M.; MACHAIDZE, L. Role of AI chatbots in education: systematic literature review. **Int J Educ Technol High Educ.**, v. 20, n. 56, 2023. Disponível em: <https://doi.org/10.1186/s41239-023-00426-1>.

LAKATOS, E. M.; MARCONI, M. A. de. **Fundamentos de metodologia científica**. 5. ed. São Paulo: Atlas, 2003.

LANCASTER, F. Wilfrid. **Indexação e resumo: teoria e prática**. Brasília: Brique de Lemos, 2004.

LANCASTER, F. W. **Information Retrieval Systems: Characteristics, Testing, and Evaluation**. 2. ed. Berkeley: Wiley, 1979.

LANDBOT. **Landbot**: No-code chatbot builder. Disponível em: <https://landbot.io/>. Acesso em: jun. 2025.

LANGCHAIN. **LangChain**: Build context-aware reasoning applications. Disponível em: <https://python.langchain.com/>. Acesso em: 14 jun. 2025.

LE COADIC, Y. F. **A Ciência da Informação**. 2. ed. Brasília: Brique de Lemos, 2004.

LEE, K. *et al.* Can chatbots help reduce the workload of administrative officers? -implementing and deploying faq chatbot service in a university. In: **International Conference on Human-Computer Interaction**. Springer, 2019.

LEMOS, D. L. S.; MARTINS, D. L.; SOUZA, R. R. Organização e representação da informação e do conhecimento em contextos informacionais: uma proposta de um

modelo teórico-conceitual para a qualidade de objetos culturais digitais. **Fronteiras da Representação do Conhecimento**, Belo Horizonte, v. 3, n. 2, p. 86-124, 2023. Disponível em: <https://cip.brapci.inf.br/download/228265>. Acesso em: 12 jan. 2024.

LEWIS, P. *et al.* Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. **Advances in Neural Information Processing Systems**, v. 33, 2020. Disponível em: <https://arxiv.org/abs/2005.11401>. Acesso em: 18 set. 2025.

LIMA, G. A. de. (Org.). **Organização e representação do conhecimento e da informação para web: práticas metodológicas**. Rio de Janeiro: Interciência, 2020.

LIMA, G. A. de. Organização e representação do conhecimento e da informação na web: teorias e técnicas. **Perspectivas em Ciência da Informação**, v. 25, número especial, p. 57-97, fev/2020. Disponível em: <https://repositorio.ufmg.br/handle/1843/41621>. Acesso em: 8 ago. 2024.

LIMA, G. A.; FUJITA, M. S.; REDIGOLO, F. M. A importância do contexto para a indexação. **Ponto de Acesso**, Salvador, v. 15, n. 3, p. 283-302, 2021. Disponível em: <https://doi.org/10.9771/rpa.v15i3.47469>. Acesso em: 8 ago. 2024.

LÓPEZ, A. *et al.* From Process Models to Chatbots. **Information Systems**, v. 81, p. 108-123, 2019. Disponível em: https://upcommons.upc.edu/bitstream/handle/2117/178761/process_models_chatbot_s-7.pdf?sequence=3&isAllowed=y. Acesso em: 8 ago. 2024.

LYNNYK, Y.; KRESTYANPOL, L.; ROZVOD, E. Development of a Natural Language Chatbot Interface for Website Users. **Web Development Journal**, v. 12, n. 1, p. 45-62, 2024. Disponível em: <https://journals.uran.ua/eejet/article/view/299200>. Acesso em: 8 ago. 2024.

MABUNDA, K.; ADE-IBIJOLA, A. Pathbot: An intelligent chatbot for guiding visitors and locating venues. In: **6th International Conference on Soft Computing & Machine Intelligence (ISCMi)**, 2019, p. 160-168. Disponível em: <https://ieeexplore.ieee.org/document/9004411>. Acesso em: 8 ago. 2024.

MACEDO, N. D. **Iniciação à pesquisa bibliográfica: guia do estudante para a fundamentação do trabalho de pesquisa**. São Paulo, SP: Loyola, 1994.

MAIA, M. da S.; SANTOS NETO, J. A. dos. La curación de contenidos informativos como posibilidad de mediación implícita de la información en bibliotecas híbridas. **Revista EDICIC**, [S. l.], v. 4, n. 1, p. 1–18, 2024. Disponível em: <https://ojs.edicic.org/revistaedicic/article/view/294>. Acesso em: 30 jul. 2025.

MANNING, C. D.; SCHÜTZE, H. **Foundations of Statistical Natural Language Processing**. Cambridge: MIT Press, 1999.

MANNING, C. D.; RAGHAVAN, P.; SCHÜTZE, H. **Introduction to Information Retrieval**. Cambridge: Cambridge University Press, 2009. Disponível em:

<https://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf> Acesso em: 9 maio 2024.

MANYCHAT. **ManyChat**: Chat marketing made easy. Disponível em: <https://manychat.com/>. Acesso em: 14 jun. 2025.

MATALONI NETO, D. **Chatbots na administração pública**: A inteligência artificial à luz do princípio da eficiência administrativa, do direito de acesso à informação e da sustentabilidade (Dissertação de Mestrado), Universidade do Vale do Itajaí, 2023. Disponível em: <https://biblioteca.univali.br/pergamumweb/vinculos/pdf/David%20Matalon%20Neto%202023.pdf>. Acesso em: 12 maio 2024.

MATTHEWS, P. **Knowledge Organisation Systems for Chatbots and Conversational Agents**: A Review of Approaches and an Evaluation of Relative Value-Added for the User. *In*: International Society for Knowledge Organization (ISKO). Knowledge Organization at the Interface: Proceedings of the Sixteenth International ISKO Conference, 2020. p. 274-283. Disponível em: 10.5771/9783956507762-274. Acesso em: 15 mar. 2024.

MENDOZA, S., *et al.* A Model to Develop Chatbots for Assisting the Teaching and Learning Process. **Sensors (Basel)**, v. 22, n.15, p. 5532, 2022. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/35898035/>. Acesso em: 9 maio 2024.

MESHAM, S. *et al.* (2021). College Enquiry Chatbot using Rasa Framework. **2021 Asian Conference on Innovation in Technology (ASIANCON)**, 1-8. Disponível em: <https://api.semanticscholar.org/CorpusID:238415802> Acesso em: 8 ago. 2024.

MITCHELL, T. M. **Aprendizado de máquina**. New York: McGraw-Hill Education, 2019.

MISHRA, K. *et al.* Advancements in Conversational AI: Building Mental Health Chatbot with BERT Model. **Computational Intelligence and Neuroscience**, v. 2023, p. 1-12, 2023. Disponível em: <https://ieeexplore.ieee.org/document/10489148>. Acesso em: 8 ago. 2024

MORAES NETO, A. J. *et al.* Uma abordagem baseada em dados abertos conectados e chatbot para disponibilizar o catálogo de cursos da Rede Federal de Educação Profissional, Científica e Tecnológica. *In*: Simpósio Brasileiro de Informática na Educação (SBIE), 31., 2020, Online. **Anais [...]**. Porto Alegre: Sociedade Brasileira de Computação, 2020. p. 1263-1272. Disponível em: <https://doi.org/10.5753/cbie.sbie.2020.1263>. Acesso em: 12 fev. 2024.

MOORE, K. *et al.* A comprehensive solution to retrieval-based chatbot construction. **Information Sciences**, v. 625, p. 45-68, 2023. Disponível em: <https://arxiv.org/abs/2106.06139>. Acesso em: 8 ago. 2024.

MÚNERA TORRES, M. T.; SALAZAR ÁLVAREZ, L. M.; OSORIO, A. S. Estudio inicial de un chatbot para estudiantes de la modalidad virtual de la Escuela Interamericana

de Bibliotecología. **Investig. bibli**, Ciudad de México , v. 36, n. 90, 2022. Disponível em: <https://dialnet.unirioja.es/servlet/articulo?codigo=8472293>. Acesso em: 09 ago. 2025.

MUHYIDIN, A.; SETIAWAN, M. A. F. Developing UNYSA Chatbot as Information Services about Yogyakarta State University. **Indonesian Journal of Information Systems**, v. 4, n. 1, p. 67-82, 2021. Disponível em: <https://iopscience.iop.org/article/10.1088/1742-6596/1737/1/012038>. Acesso em: 8 ago. 2024

NADKARNI, P. M.; OHNO-MACHADO, L.; CHAPMAN, W. W. Natural language processing: an introduction. **Journal of the American Medical Informatics Association**, v. 18, n. 5, p. 544-551, 2011. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/21846786/>. Acesso em: 9 mai. 2024.

NASHARUDDIN, N. A. *et al.* Designing an Educational Chatbot: A Case Study of CikguAIBot. **Malaysian Journal of Educational Technology**, v. 18, n. 2, p. 89-104, 2021. Disponível em: <https://ieeexplore.ieee.org/document/9498011>. Acesso em: 9 mai. 2024.

NASCIMENTO SILVA, P. Agilidade na Ciência da Informação: um relato de ensino de métodos ágeis para o profissional da informação. **Páginas a&b: arquivos e bibliotecas**, Porto, Portugal, p. 16–31, 2023a. Disponível em: <https://ojs.letras.up.pt/index.php/paginasueb/article/view/13603>. Acesso em: 14 out. 2025.

NASCIMENTO SILVA, P. Recuperação de informação na Ciência da Informação: produção acadêmico-científica brasileira (2012-2021), 2023b. **Transinformação**, v. 35, e237336, 2023. Disponível em: <https://doi.org/10.1590/2318-0889202335e237336>. Acesso em: 15 maio 2025.

NASCIMENTO SILVA, P.; SOARES, J. R. Recuperação de informações acadêmicas na UFMG. 2025. **Mendeley Data**, v. 1. Disponível em: <https://doi.org/10.17632/j5x5jz7cfr.1>. Acesso em: 16 de dez. 2025.

NIELSEN, J. **Usabilidade**: projetando interfaces com o usuário em mente. Tradução de Maria Cecília de Souza Minayo. Rio de Janeiro: Campus, 2000.

NOVELLINO, M. S. Instrumentos e metodologias de representação da informação. **Informação & Informação**, [S. l.], v. 1, n. 2, p. 37–45, 1996. DOI: 10.5433/1981-8920.1996v1n2p37. Disponível em: <https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/1603>. Acesso em: 9 jul. 2025.

OKONKWO, C. W., ADE-IBIJOLA, A. Chatbots applications in education: A systematic review, **Computers and Education: Artificial Intelligence**, v. 2, 2021. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2666920X21000278?via%3Dihub>. Acesso em: 5 set. 2024.

OLIVEIRA, P. F.; MATOS, P. Introducing a Chatbot to the Web Portal of a Higher Education Institution to Enhance Student Interaction. **International Journal of Educational Technology in Higher Education**, v. 20, n. 1, p. 1-18, 2023. Disponível em: <https://www.mdpi.com/2673-4591/56/1/128>. Acesso em: 5 set. 2024.

OPENAI. **GPT-4 Technical Report**. 2025. Disponível em: <https://openai.com/research/gpt-4>. Acesso em: 08 set. 2025.

PANDA, S.; CHAKRAVARTY, R. Adapting intelligent information services in libraries: a case of smart AI chatbots. **Library Hi Tech**, v. 40, n. 2, p. 456-473, 2022. Disponível em: <https://www.emerald.com/insight/content/doi/10.1108/lhtn-11-2021-0081/full/html>. Acesso em: 12 maio 2025.

PATEL, N. P. *et al.* AI and Web-Based Human-Like Interactive University Chatbot (UNIBOT). **International Journal of Engineering and Advanced Technology**, v. 8, n. 6, p. 2456-2463, 2019. Disponível em: <https://ieeexplore.ieee.org/document/8822176>. Acesso em: 6 jun. 2024.

PEREIRA, J. V., CARVALHO, A. V. A curadoria de conteúdo e as possibilidades de atuação do bibliotecário como curador. **Revista ACB: Biblioteconomia em Santa Catarina, Florianópolis**, v. 28, n. 1, p. 1-17, jan./dez., 2023. Disponível em: <https://dialnet.unirioja.es/descarga/articulo/9080531.pdf>. Acesso em: 9 maio 2024.

PIMENTA, R. M.; ALMEIDA JÚNIOR, O. F. de. Mediação da informação e múltiplas linguagens. **Tendências da Pesquisa Brasileira em Ciência da Informação**, [S. l.], v. 2, n. 1, 2009. Disponível em: <https://revistas.ancib.org/tpbci/article/view/170>. Acesso em: 7 jul. 2025.

PRAKASAM, S. *et al.* Design and Development of AI-Powered Healthcare WhatsApp Chatbot. **2023 2nd International Conference on Vision Towards Emerging Trends in Communication and Networking Technologies (ViTECoN)**, Vellore, India, 2023, p. 1-6, doi: 10.1109/ViTECoN58111.2023.10157423. Disponível em: <https://ieeexplore.ieee.org/document/10157423>. Acesso em: 8 ago. 2025.

PRODANOV, C. C.; FREITAS, E. C. **Metodologia do Trabalho Científico: métodos e técnica da pesquisa e do trabalho acadêmico**. 2. ed. – Novo Hamburgo: Feevale, 2013.

RANOLIYA, B. R.; RAGHUWASHI, N.; SINGH, S. Chatbot for university related faqs. In: **International Conference on Advances in Computing, Communications, and Informatics (ICACCI)**, 2017, p. 1525-1530. Disponível em: <https://ieeexplore.ieee.org/document/8126057>. Acesso em: 21 jun. 2025.

RAMALAKSHMI, K. *et al.* Using Artificial Intelligence Methods to Create a Chatbot for University Questions and Answers. **Journal of Educational Computing Research**, v. 61, n. 4, p. 789-812, 2023. Disponível em: <https://www.mdpi.com/2673-4591/59/1/16> Acesso em: 24 jun. 2025.

REINA, D. R.; CRUZ, C. Curadoria de chatbots: conceptualização, estratégias e indicadores de desempenho. **RISTI**, v. 40, n.2, 2020. Disponível em: <https://scielo.pt/pdf/rist/n40/1646-9895-rist-40-2.pdf>. Acesso em: 24 jun. 2025.

REYES, A., *et al.* Gestor e Curador da Informação: tendências, perfis e estratégias de reconhecimento. **Páginas a&b: Arquivos E Bibliotecas**, n.7, p. 3–15, 2017. Disponível em: <https://brapci.inf.br/v/67432>. Acesso em: 24 jun. 2025.

ROBREDO, J.; BRÄSCHER, M. (Orgs.). **Passeios pelo bosque da informação: estudos sobre a representação e organização da informação e do conhecimento – erocio**. Brasília DF: IBICT, 2010. Disponível em: <https://www2.senado.leg.br/bdsf/item/id/189812>. Acesso em: 24 jun. 2025.

RODRIGUEZ, S.; MUNE, C. Uncoding library chatbots: deploying a new virtual reference tool at the San Jose State University library. **Reference Services Review**, v. 50, n. 1, p. 34-47, 2022. Disponível em: https://scholarworks.sjsu.edu/faculty_rsca/3506/. Acesso em: 24 jun. 2025.

ROBREDO, J. **Da Ciência da Informação revisitada aos sistemas humanos de informação**. Brasília: Thesaurus, 2003. 262 p.

ROHRIG, C.; HEB, D. Omniman: A mobile assistive robot for intralogistics applications. **Engineering Letters**, v. 27, n. 4, 2019. Disponível em: <https://www.fh-dortmund.de/publikationen/fis.import.pub.de.4170.php?loc=en>. Acesso em: 24 jun. 2025.

ROSENFELD, L.; MORVILLE, P.; ARANGO, J. **Information Architecture: For the Web and Beyond**. Sebastopol, CA: O'Reilly Media, 2015. ISBN 978-1491911686.

RUDKOWSKY, E. *et al.* More than Bags of Words: Sentiment Analysis with Word Embeddings. **Communication Methods and Measures**, v. 12, n. (2–3), p. 140–157, 2018. Disponível em: <https://doi.org/10.1080/19312458.2018.1455817>. Acesso em: 24 jun. 2025.

RUSSELL, S. J.; NORVIG, P. **Artificial intelligence: a modern approach**. 4. ed. Harlow, UK: Pearson Education, 2021.

RUKHIRAN, M.; NETINANT, P. Automated information retrieval and services of graduate school using chatbot system. **International Journal of Information Management**, v. 62, p. 102-118, 2022. Disponível em: <https://ijece.iaescore.com/index.php/IJECE/article/view/27818>. Acesso em: 24 jun. 2025.

SAFI, Z. *et al.* Technical Aspects of Developing Chatbots for Medical Applications: Scoping Review. **Journal of medical Internet research**, v. 22, n. 12, e19127, 2020. Disponível em: <https://doi.org/10.2196/19127>. Acesso em: 8 ago. 2025.

SALTON, G. **Automatic information organization and retrieval**. New York: McGraw-Hill, 1968.

SANJAYA, A.; SANTOSO, H. Virtual Assistant Prototype for Property Sales. **International Journal of Interactive Mobile Technologies**, v. 15, n. 12, p. 123-138, 2021. Disponível em: https://www.researchgate.net/publication/357358151_Virtual_Assistant_Prototype_for_Property_Sales. Acesso em: 21 maio 2025.

SANTANA, R. C. Ciclo de vida dos dados: uma perspectiva a partir da ciência da informação. **Inf. Inf.**, Londrina, v. 21, n. 2, p. 116 – 142, maio/ago., 2016. Disponível em: <https://ojs.uel.br/revistas/uel/index.php/informacao/article/view/27940>. Acesso em: 24 jun. 2025.

SANTANA, R. *et al.* A Chatbot to Support Basic Students Questions. In: BRAZILIAN SYMPOSIUM ON COMPUTERS IN EDUCATION, 32., 2021. **Anais [...]**. [S.l.: s.n.], 2021. Disponível em: <https://ceur-ws.org/Vol-3059/paper6.pdf>. Acesso em: 24 jun. 2025.

SARACEVIC, T. Ciência da informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 1, n. 1, p. 41-62, 1996.

SARACEVIC, T. Relevance: A Review of the Literature and a Framework for Thinking on the Notion in Information Science. **Journal Of The American Society , For Information Science And Technology**, v. 58, n. 13, p. 2126–2144, 2007. Disponível em: https://ils.unc.edu/courses/2017_fall/inls509_002/papers/Saracevic07Part3.pdf. Acesso em: 24 jun. 2025.

SAVOLAINEN, R. Everyday life information seeking: Approaching information seeking in the context of “way of life”. **Library & Information Science Research**, v. 17, n.3, p. 259-294, 1995. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/0740818895900489#preview-section-cited-by>. Acesso em: 8 ago. 2025
l: O'Reilly, 2015.

SAVOLAINEN, R. Information behavior and information practice: reviewing the “umbrella concepts” of information-seeking studies. **The Library Quarterly**, v. 77, n. 2, p. 109-132, 2007. Disponível em: <https://www.jstor.org/stable/10.1086/517840>
Acesso em: 24 jun. 2025.

SCHWABER, K., SUTHERLAND, J. **Guia do Scrum: O Guia Definitivo para o Scrum: As Regras do Jogo**. 2020. Disponível em: <https://scrumguides.org/docs/scrumguide/v2020/2020-Scrum-Guide-PortugueseBR-3.0.pdf>. Acesso em: 24 jun. 2025.

SEJNOWSKI, T. J. Large Language Models and the Reverse Turing Test. **Neural Computation**, v. 35, n. 3, p.309-342, 2023. Disponível em: https://doi.org/10.1162/neco_a_01563. Acesso em: 24 jun. 2025.

SEVERINO, A. J. **Metodologia do trabalho científico**. 23. ed. São Paulo: Cortez, 2013. Disponível em: https://presencial.moodle.ufsc.br/pluginfile.php/632598/mod_resource/content/1/Meto

dologia_do_Trabalho_Cient%C3%ADfico_-_1%C2%AA_Edi%C3%A7%C3%A3o_-_Antonio_Joaquim_Severino_-_2014.pdf. Acesso em: 12 out. 2025.

SHAWAR, S., ATWELI, I. Chatbots: are they really useful?. In: Proceedings of the 2007 international conference on artificial intelligence (**ICAI'07**), p. 441-447, 2007. Disponível em: <https://www.periodicos.capes.gov.br/index.php/acervo/buscador.html?task=detalhes&id=W1496253462>. Acesso em: 24 jun. 2025.

SHNEIDERMAN, B. **Human-Centered AI**. New York: Oxford University Press, 2020. Disponível em: https://books.google.com.br/books/about/Human_centered_AI.html?id=YS9VEAAAQBAJ&redir_esc=y. Acesso em: 24 jun. 2025.

SIEBRA, S. A.; BORBA, V. R.; MIRANDA, M. K. F. O. Curadoria digital: um termo interdisciplinar. **Informação & Tecnologia**, v. 3, n. 2, 2016. Disponível em: <https://brapci.inf.br/v/41848>. Acesso em: 24 jun. 2025.

SILVEIRA, P. E. da. **Fundamentos da POO** (Programação Orientada por Objectos). 2015. Disponível em: http://repositorio.ulusiada.pt/bitstream/11067/1720/1/LEE_19_10.pdf. Acesso em: 10 out. 2025.

SINGLA, A. *et al.* Developing a Chatbot to Support Individuals With Neurodevelopmental Disorders: Tutorial. **JMIR Medical Informatics**, v. 12, e45123, 2024. Disponível em: <https://pmc.ncbi.nlm.nih.gov/articles/PMC11220430/>. Acesso em: 24 jun. 2025.

SOARES, J. R.; NASCIMENTO SILVA, P. Chatbots no Brasil: panorama da produção acadêmica. V Fórum de Pesquisas Discentes do Programa de Pós-Graduação em Gestão e Organização do Conhecimento. **Anais...V FORPED PPGGOC**, Belo Horizonte, 2024. Disponível em: <https://forped.eci.ufmg.br/revista/forped/article/view/142/87>. Acesso em: 5 set. 2024.

SOERGEL, D. The rise of ontologies or the reinvention of classification. **Journal of the American Society for Information Science**, v. 50, n. 12, p. 1119–1120, 1999. Disponível em: <https://www.dsoergel.com/cv/B70.pdf>. Acesso em: 20 maio 2024.

SOMMERVILLE, I. **Software Engineering**. 10. ed. Boston: Pearson, 2016. SOUZA, R. R. Sistemas de recuperação de informações e mecanismos de busca na web: panorama atual e tendências. **Perspectivas em Ciência da Informação**, v. 11, n. 2, p. 161–173, maio 2006. Disponível em: <https://www.scielo.br/j/pci/a/7tt9ykG8xTGBWsyYnDhmghr/>. Acesso em: 24 jun. 2025.

SRIVASTAVA, S.; PRABHAKAR, T. Características desejáveis de uma plataforma de construção de chatbot. **IEEE International Conference on Humanized Computing and Communication with Artificial Intelligence (HCCAI)**, p.61-64, 2020. Disponível em: <https://doi.org/10.1109/HCCAI49649.2020.00016>. Acesso em: 20 maio 2024.

STEPANOV, M. S. *et al.* The Automation of Client Servicing in University and College Admission Office. **Educational Technology & Society**, v. 26, n. 2, p. 134-149, 2023. Disponível em: <https://ieeexplore.ieee.org/document/10092103>. Acesso em: 26 Jul. 2020.

SUTSKEVER, I., VINYALS, O., LE, Q. V. Sequence to Sequence Learning with Neural Networks. **NeurIPS**. 2014. Disponível em: <https://arxiv.org/abs/1409.3215>. Acesso em: 26 Jul. 2020.

TANWAR, P. *et al.* AI Based Chatbot for Healthcare using Machine Learning. **International Journal of Healthcare Management**, v. 16, n. 3, p. 256-271, 2023. Disponível em: <https://ieeexplore.ieee.org/document/10465728/>. Acesso em: 12 mai. 2025.

TASKIN, Z.; AL, U. Natural language processing applications in library and information science. **Online Information Review**, v. 43, n. 4, p. 676-690, 12 Aug. 2019. Disponível em: Disponível em: <https://doi.org/10.1108/OIR-07-2018-0217>. Acesso em: 26 Jul. 2020.

THAIPRASERT, B. *et al.* Development of a Class Materials Search System using LINE Chatbot. **Educational Technology Research and Development**, v. 71, n. 4, p. 1023-1040, 2023. Disponível em: <https://ieeexplore.ieee.org/document/10153145>. Acesso em: 12 abr. 2025.

TIDIO. **Free live chat and chatbot software for your website**. Disponível em: <https://www.tidio.com/>. Acesso em: 14 jun. 2025.

TOMMY, L.; KIRANA, C.; RISKI, L. The Combination of Natural Language Processing and Entity Extraction for Academic Chatbot. **Journal of Information Technology Education**, v. 19, p. 567-584, 2020. Disponível em: <https://ieeexplore.ieee.org/document/9268851>. Acesso em: 30 set. 2025.

TOUIMI, B Y. *et al.* Intelligent Chatbot-LDA Recommender System. **International Journal of Emerging Technologies in Learning (iJET)**, [S. l.], v. 15, n. 20, p. 4–20, 2020. Disponível em: <https://online-journals.org/index.php/i-jet/article/view/15657>. Acesso em: 30 set. 2024.

TRINDADE, A. S da .; OLIVEIRA, P. C de . Inteligência artificial (ia) generativa e competência em informação: habilidades informacionais necessárias ao uso de ferramentas de ia generativa em demandas informacionais de natureza acadêmica-científica . **Perspectivas em Ciência da Informação**, v. 29, p. e–47485, 2024. Disponível em: <https://www.scielo.br/j/pci/a/GVCW7KbcRjGVhLSrmy3PCng/?format=html&lang=pt>. Acesso em 19 set. 2025.

UNIVERSIDADE FEDERAL DE MINAS GERAIS. **Resolução Complementar Nº 01/2018**, de 20 de Fevereiro De 2018. Aprova as Normas Gerais de Graduação da Universidade Federal de Minas Gerais (UFMG). Belo Horizonte, 2018.

UNIVERSIDADE FEDERAL DE MINAS GERAIS. **Escritório de Governança de Dados Institucionais. UFMG em números**. 2021. Disponível em: <https://www.ufmg.br/egdi/>. Acesso em: 4 fev. 2025.

VASILEVICH, A. *et al.* Language Knowledge Graphs for Smarter Multilingual Chatbots. **Language Resources and Evaluation**, v. 56, n. 4, p. 1234-1251, 2022. Disponível em: <https://ceur-ws.org/Vol-3235/paper2.pdf>. Acesso em: 20 mai. 2024.

VASWANI, A. *et al.* Attention is all you need. In: **Advances in Neural Information Processing Systems**, 2017, p. 5998-608. Disponível em: <https://arxiv.org/abs/1706.03762>. Acesso em: 20 maio 2024.

VILLANUEVA, D. P. P.; AGUILAR-ALONSO, I. A Chatbot as a Support System for Educational Institutions. **Educational Technology & Society**, v. 24, n. 3, p. 78-92, 2021. Disponível em: <https://ieeexplore.ieee.org/document/9615271/>. Acesso em: 30 set. 2024.

WANG, Y.; LIU, S.; SONG, L. Designing an Educational Chatbot with Joint Intent Classification and Slot Filling. **Computers & Education**, v. 189, p. 104-119, 2022. Disponível em: <https://ieeexplore.ieee.org/document/10148291>. Acesso em: 20 maio 2024.

WEISER, M. The computer for the 21st century. **Scientific American**, New York, v. 265, n. 3, p. 94-104, set. 1991. Disponível em: <https://www.lri.fr/~mbl/Stanford/CS477/papers/Weiser-SciAm.pdf>. Acesso em: 20 maio 2024.

WEIZENBAUM, J. ELIZA—a computer program for the study of natural language communication between man and machine. **Communications of the ACM**, v. 9, n. 1, p. 36-45, 1966. Disponível em: <https://dl.acm.org/doi/10.1145/365153.365168>. Acesso em: 22 abr. 2025.

WILSON, T. D. Models in information behaviour research. **Journal of Documentation**, v. 55, n. 3, p. 249–270, 1999. Disponível em: <https://doi.org/10.1108/EUM000000007145>. Acesso em: 22 abr. 2025.

WHITTLE, D.; HALL, L. Applying cognitive computing to legal services. **Artificial Intelligence and Law**, v. 30, n. 3, p. 387-404, 2022. Disponível em: <https://www.scienceopen.com/hosted-document?doi=10.14236/ewic/HCI2022.35>. Acesso em: 20 maio 2024.

YANG, X. *et al.* Deep Learning-Based Intelligent Chatbot in a Digital Campus, 2023 2nd International Conference on Data Analytics, **Computing and Artificial Intelligence (ICDACAI)**, Zakopane, Poland, 2023, p. 196–199, Disponível em: <https://ieeexplore.ieee.org/document/10361205>. Acesso em: 20 maio 2024.

YOSSY, E. H.; BUDIHARTO, W. Knowledge-based chatbot for humanoid robot In restaurant for question and answering system. **ICIC Express Letters**, [S.l.], v.13, n. 3, 2022. Disponível em: <http://www.icicelb.org/ellb/contents/2022/3/elb-13-03-11.pdf>. Acesso em: 8 ago. 2025.

ZAMBIASI, S. P.; RABELO, R. J. **Arisa Nest** – A Cloud-Based Platform for Development of Virtual Assistant. Core. Instituto de Informática, Universidade Federal do Rio Grande do Sul, 2020. Disponível em:
<https://core.ac.uk/works/128783049/?t=1cddb3fb39b474895731b6af158578a7-128783049>.
Acesso em: 8 ago. 2025.

ANEXO A

Quadro 14 - Recorte do questionário da pesquisa sobre o comportamento de busca dos estudantes de graduação e pós-graduação da UFMG

Pergunta	Objetivo	Tipo (aberta/ fechada)/ escala likert	Opções	Fundamento/justificativa das opções
Seção 2:				
Em quais momentos você acessa informações acadêmicas (normas, regimentos ou resoluções) da UFMG? (Considere a opção mais frequente).	Identificar em quais momentos os discentes acessam as informações acadêmicas da UFMG.	Fechada	<ul style="list-style-type: none"> • No início do semestre. • No final do semestre. • No período de matrícula. • Durante todo o semestre. • Quando tenho um problema ou ocorre uma situação específica. • Nunca acessei. 	Permite identificar os períodos do semestre em que há maior demanda por informações acadêmicas.
Como você acessa informações acadêmicas (normas, regimentos ou resoluções) da UFMG? (Marque a opção mais frequente).	Identificar como (comportamento de busca) os discentes acessam as informações acadêmicas da UFMG.	Fechada	<ul style="list-style-type: none"> • Em ambientes digitais, utilizando computador (notebook ou desktop) ou smartphone. • Por contato telefônico (celular ou telefone fixo). • Pessoalmente, na Unidade Acadêmica. • Nunca acessei. 	Identifica os principais meios utilizados pelos discentes para acessar informações, como canais digitais, telefone ou presencialmente.
Quais locais/ambientes você utiliza para recuperar informações acadêmicas (normas, regimentos ou resoluções) da UFMG? (Marque a opção mais frequente).	Identificar onde os discentes acessam as informações acadêmicas da UFMG (ambientes e fontes de informação).	Fechada	<ul style="list-style-type: none"> • <i>Site</i> oficial da UFMG. • <i>Site</i> do curso ou do programa. • Sistemas acadêmicos (SIGA, Diário e Moodle). • Consulto o meu e-mail institucional (mensagens já encaminhadas). • Pesquisa em motores de busca, na <i>internet</i> (Google, Bing ou outro). 	Identifica a quais alternativas os discentes recorrem quando não encontram informações, revelando desafios na comunicação institucional.

			<ul style="list-style-type: none"> • Pesquisa em modelos de IA (ChatGPT, Gemini, Copilot ou outro). • Pesquisa (pergunto) em redes sociais ou grupos de mensagens administrados por colegas da turma ou do curso (WhatsApp, Telegram ou outro). • Não recuperei nenhuma informação acadêmica. • Outros 	
Seção 2				
Quando você encontra dificuldades para localizar uma informação acadêmica (normas, regimentos ou resoluções), mesmo em outras fontes não oficiais, qual é o seu procedimento habitual?	Identificar como (comportamento de busca) os discentes acessam as informações acadêmicas da UFMG.	Fechada	<ul style="list-style-type: none"> • Vou pessoalmente na Secretaria ou na Central de Atendimento. • Envio e-mail para a secretaria do curso (para encaminhar ao Colegiado), à Central de Atendimento ou ao órgão responsável. • Pergunto para os professores. • Pergunto para colegas da UFMG. • Telefone para a secretaria do curso ou para a Central de Atendimento. • Desisto da busca. • Não tive dificuldades. • Outros. 	Identifica a quais alternativas os discentes recorrem quando não encontram informações, revelando os desafios na comunicação institucional.

<p>Em sua opinião, qual é a confiabilidade das informações acadêmicas disponibilizadas em canais/fontes não oficiais da UFMG?</p>	<p>Identificar a percepção dos discentes (avaliação do acesso, qualidade e confiabilidade conforme a respectiva fonte) quanto ao acesso às informações acadêmicas da UFMG.</p>	<p>Escala Likert</p>	<ul style="list-style-type: none"> • Muito baixa, pois sempre há inconsistências e divergências. • Baixa, pois nem sempre estão corretas. • Regular, pois algumas informações estão corretas. • Alta, pois grande parte das informações está correta. • Muito alta, pois sempre estão corretas. • Nunca acessei. 	<p>Avalia o grau de confiança dos discentes nas informações obtidas fora dos canais oficiais.</p>
<p>Como você avalia o acesso e a recuperação de informações acadêmicas na UFMG?</p>	<p>Identificar a percepção dos discentes (avaliação do acesso, qualidade e confiabilidade conforme a respectiva fonte) quanto ao acesso às informações acadêmicas da UFMG.</p>	<p>Escala Likert</p>	<ul style="list-style-type: none"> • Muito difícil, porque as informações estão dispersas ou não sei onde procurar. • Difícil, porque levo muito tempo para encontrar o que eu procuro ou preciso recorrer a outras pessoas. • Regular, porque nem sempre as informações estão completas ou claras. • Fácil, porque os canais oficiais são organizados e intuitivos ou não tenho dificuldade em encontrar o que preciso. • Muito fácil porque sempre encontro as informações rapidamente e sem esforço ou os canais atendem plenamente às minhas necessidades. • Nunca acessei. 	<p>Mede a percepção dos alunos sobre a dificuldade ou facilidade de encontrar informações acadêmicas.</p>
<p>Seção 3</p>				
<p>Qual é a sua dúvida mais recorrente sobre</p>	<p>Identificar as principais dúvidas dos</p>	<p>Aberta</p>	<p>Permite mapear os principais questionamentos e lacunas no acesso à informação.</p>	

as informações acadêmicas ou a última que se recorde?	discentes sobre informações acadêmicas.		
Na sua opinião, as informações acadêmicas são acessíveis e bem organizadas? Por qual motivo?	Identificar a percepção dos discentes (avaliação do acesso, da qualidade e da confiabilidade conforme a respectiva fonte) quanto ao acesso às informações acadêmicas da UFMG.	Aberta	Coleta percepções qualitativas sobre a estrutura e clareza das informações disponíveis.
Você tem alguma sugestão para aprimorar o acesso e a recuperação de informações acadêmicas da UFMG?	Coletar sugestões para aprimorar o acesso às informações acadêmicas.	Aberta	Reúne propostas de melhoria diretamente dos usuários do sistema de informação acadêmica.

Fonte: Nascimento Silva e Soares (2025).

APÊNDICE A

Descrição das ferramentas de desenvolvimento de *chatbot*

Landbot

Categoria: *No-code*

Foco: Interface visual e automação de conversas

URL: <https://landbot.io/>

O Landbot (LANDBOT, 2025) caracteriza-se como uma plataforma *no-code* voltada à criação de interfaces conversacionais visuais, priorizando a experiência do usuário a partir de elementos gráficos interativos. Sua arquitetura permite a construção de fluxos de diálogo por meio de um editor visual *drag-and-drop*, dispensando conhecimentos de programação. A ferramenta destaca-se pela capacidade de criar experiências conversacionais que transcendem o formato tradicional de *chat*, incorporando elementos como botões, carrosséis de imagens e formulários integrados. No contexto acadêmico, o Landbot oferece potencial para a criação de interfaces mais atrativas e intuitivas, especialmente adequadas para processos que envolvem a coleta de informações estruturadas, como formulários de matrícula ou cadastros estudantis. Sua integração com APIs externas possibilita a incorporação de funcionalidades de IA, embora não de forma nativa. A plataforma opera sob modelo *freemium*, oferecendo funcionalidades básicas gratuitas, com limitações de interações mensais.

ManyChat

Categoria: *No-code*

Foco: Marketing conversacional e redes sociais

URL: <https://manychat.com/>

O ManyChat (MANYCHAT, 2025) posiciona-se como uma solução especializada em marketing conversacional, com forte orientação para redes sociais e aplicações comerciais. Sua arquitetura privilegia a automação de campanhas e o engajamento por meio de plataformas como WhatsApp, Instagram e Facebook Messenger. A ferramenta oferece recursos avançados de segmentação de público, automação de respostas baseadas em gatilhos comportamentais e integração com sistemas de

CRM. Para aplicações institucionais, o ManyChat apresenta limitações relacionadas ao seu direcionamento comercial, embora possa ser adaptado para comunicação institucional via WhatsApp. Sua interface *no-code* facilita a implementação por equipes sem conhecimento técnico, e os recursos de automação podem ser aproveitados para a distribuição de informações acadêmicas. A plataforma não possui integração nativa com IA generativa, dependendo de APIs externas para um PLN mais sofisticado.

Tidio

Categoria: *No-code*

Foco: Atendimento ao cliente e *live chat*

URL: <https://www.tidio.com/>

O Tidio (TIDIO, 2025) combina funcionalidades de *chatbot* com recursos de *live chat*, posicionando-se como uma solução híbrida para atendimento. Sua arquitetura permite a transição fluida entre atendimento automatizado e humano, incorporando IA básica para a compreensão de intenções simples. A plataforma destaca-se pela facilidade de implementação em *websites* a partir de *widgets* personalizáveis e pela interface intuitiva de gerenciamento de conversas. No ambiente acadêmico, o Tidio oferece potencial para complementar sistemas de atendimento estudantil, permitindo o direcionamento automático de dúvidas básicas e o encaminhamento de questões complexas para atendentes humanos. Sua IA integrada, embora limitada, possibilita o reconhecimento de padrões conversacionais simples. A ferramenta opera sob modelo *freemium*, com funcionalidades básicas gratuitas adequadas para implementações piloto.

Flowise

Categoria: *Low-code open source*

Foco: Orquestração de IA e flexibilidade técnica

URL: <https://flowiseai.com/>

O Flowise (FLOWISE, 2025) representa uma abordagem inovadora ao desenvolvimento de *chatbots*, oferecendo uma interface visual para a orquestração de LLMs. Como solução *open source*, essa ferramenta proporciona flexibilidade total de customização e integração, permitindo a construção de *pipelines* complexos de

PLN por meio de componentes modulares. Sua arquitetura, baseada em nós visuais, facilita a configuração de fluxos de IA sem programação extensiva, embora exija conhecimento técnico para configuração e manutenção. No contexto acadêmico, o Flowise oferece potencial para o desenvolvimento de soluções altamente especializadas, com capacidade de integração com bases de conhecimento institucionais e modelos de IA personalizados. A natureza *open source* elimina custos de licenciamento, transferindo os investimentos para infraestrutura e manutenção técnica.

Botpress

Categoria: *Low-code*

Foco: Desenvolvimento empresarial e escalabilidade

URL: <https://botpress.com/pt>

O Botpress (BOTPRESS, 2025) consolida-se como uma plataforma empresarial para o desenvolvimento de *chatbots*, equilibrando facilidade de uso com robustez técnica. Sua arquitetura modular permite a integração nativa com diversos modelos de IA, oferecendo recursos avançados de PLN, gerenciamento de contexto conversacional e análise de performance. A plataforma destaca-se pela sua capacidade de suportar múltiplos canais de comunicação simultaneamente, incluindo *web*, WhatsApp e APIs customizadas. Seu editor visual permite o desenvolvimento de fluxos conversacionais complexos com pouco código enquanto mantém a flexibilidade para customizações avançadas. No ambiente acadêmico, o Botpress oferece infraestrutura adequada para soluções escaláveis, com recursos de *analytics* e gerenciamento de usuários, essenciais para implementações institucionais.

LangChain

Categoria: *Framework open source*

Foco: Desenvolvimento avançado e integração de IA

URL: <https://python.langchain.com/>

O LangChain (LANGCHAIN, 2025) constitui um *framework* de desenvolvimento em Python especificamente projetado para aplicações de IA generativa e PLN. Sua arquitetura oferece abstrações de alto nível para integração com diversos modelos de linguagem, bases de dados vetoriais e fontes de conhecimento externas,

permitindo a construção de sistemas conversacionais altamente sofisticados. Como *framework* de código aberto, o LangChain proporciona controle total sobre a implementação, possibilitando a criação de soluções personalizadas para necessidades específicas. No contexto acadêmico, essa ferramenta oferece potencial para o desenvolvimento de sistemas de RI avançados, com capacidade de processamento de documentos institucionais complexos e geração de respostas contextualizadas. Sua implementação exige conhecimento avançado de programação Python e arquitetura de sistemas, representando a opção mais técnica entre as alternativas analisadas.