DOI http://dx.doi.org/10.1590/1981-5344/61756

CONSTRUÇÃO E ANÁLISE DE PROTÓTIPO DE *CHATBOT* GENERATIVO COM *RETRIEVAL-AUGMENTED GENERATION* (RAG) A PARTIR DE UM VOCABULÁRIO CONTROLADO PARA APLICAÇÃO NO DOMÍNIO DA PROPRIEDADE INTELECTUAL (PI)

Paula Cotrim de Abrantes
Universidade Federal do Estado do Rio de Janeiro (UNIRIO)
Rio de Janeiro, Rio de Janeiro, Brasil

Vania Lisboa da Silveira Guedes
Universidade Federal do Rio de Janeiro (UFRJ)
Rio de Janeiro, Rio de Janeiro, Brasil

RESUMO
A partir da criação de um protótipo de vocabulário controlado, no domínio da Propriedade Intelectual (PI), para ser o contexto de um protótipo de *chatbot*, objetiva-se analisar se o protótipo de *chatbot* está fornecendo respostas precisas, com definições e garantia literária, e se não está fornecendo respostas inventadas. A partir do protótipo de tesauro do domínio de PI, de Pereira (2022), foi elaborada a estruturação de um protótipo de vocabulário controlado no mesmo domínio que alicerçou o protótipo de *chatbot*. Foram usadas ferramentas de Inteligência Artificial (IA): *LangChain*, *FAISS Vector Store*, modelo GPT-4o mini, da *OpenAI* e RAG. O protótipo de vocabulário controlado foi usado como o contexto para o *bot*. A partir de testes empíricos nesta pesquisa, o *bot* visou responder perguntas com garantia literária. No entanto, suas configurações não impediram algumas imprecisões. Observa-se, desta forma, que é necessária a continuação da investigação científica em sistemas de inteligência artificial, atuando em conjunto com vocabulários controlados. A partir de resultados exploratórios, verificou-se que o RAG auxilia *chatbots* na identificação semântica das perguntas para trazerem respostas corretas, mas são necessários ajustes finos na programação de *chatbots* futuros para ampliar a capacidade de precisão semântica.

Palavras-chave: Inteligência Artificial. Propriedade Intelectual. Vocabulário controlado. Chatbot. LangChain.

CONSTRUCTION AND ANALYSIS OF A GENERATIVE CHATBOT PROTOTYPE WITH RETRIEVAL-AUGMENTED GENERATION (RAG) BASED ON A CONTROLLED VOCABULARY FOR APPLICATION IN THE FIELD OF INTELLECTUAL PROPERTY (IP)

ABSTRACT
Based on the creation of a controlled vocabulary prototype in the field of Intellectual Property (IP), intended to provide the context for a chatbot prototype, this study aims to analyze whether the chatbot prototype provides accurate answers supported by definitions and literary warrant and whether it avoids generating fabricated responses. Based on Pereira's (2022) thesaurus prototype for the IP domain, a controlled vocabulary prototype was structured within the same domain, which served as the foundation for the chatbot prototype. Artificial Intelligence (AI) tools were employed, including LangChain, FAISS Vector Store, OpenAI's GPT-4o mini model, and Retrieval-Augmented Generation (RAG). The controlled vocabulary prototype was used as the context for the bot. Based on empirical testing conducted in this research, the bot was designed to answer questions with literary warrant. However, its configurations did not prevent some inaccuracies. Thus, the findings indicate the need for continued scientific investigation into artificial intelligence systems operating in conjunction with controlled vocabularies. Based on the exploratory results, it was found that RAG assists chatbots in the semantic identification of questions, enabling them to provide accurate answers; however, further fine-tuning of the programming of future chatbots is necessary to enhance their semantic accuracy.

Keywords: Artificial Intelligence. Intellectual Property. Controlled vocabulary. Chatbot. LangChain.

DOI http://dx.doi.org/10.1590/1981-5344/61756

DECLARAÇÃO DE DISPONIBILIDADE DE DADOS
data-available – Os dados de pesquisa estão disponíveis em repositório.
https://github.com/pcotrim/ChatbotPI

Recebido em: 15/09/2025
Aceito em: 21/09/2026
Editora Chefe: Lorena Tavares de Paula

1 INTRODUÇÃO

O campo da Inteligência Artificial (IA) teve sua origem em 1956, em um *workshop*, no *Dartmouth College*, nos Estados Unidos. Pesquisadores como McCarthy, Minsk e Shannon ministraram o curso que tinha como foco compreender como máquinas poderiam mimetizar o ser humano nos mais amplos aspectos (Coeckelbergh, 2024). No entanto, estudos e objetos relacionados a imitar o ser humano vêm desde a Antiguidade. Aristóteles buscava criar um sistema de raciocínio dedutivo. No século VIII antes de Cristo (a.C.), no Egito, uma estátua do deus Amon foi construída para mover-se e "falar", compondo um imaginário que pode ser relacionado a uma representação inicial da IA (Barone; Boesing, 2015). Entre brinquedos mecânicos da Idade Média e a melhor compreensão do neurônio no século XX, a IA avançou até chegar ao seu nível de evolução do século XXI.

A partir desses avanços, com o decorrer do tempo, surgiram ferramentas como os *chatbots*. Em 1966, surgiu o *chatbot* Eliza, que imitava um psicanalista rogeriano e emulava uma conversa simples. Desde então, houve grandes evoluções nos sistemas computacionais, sua capacidade de processamento aumentou e, em linha com o que se previa a Lei de Moore, expressivos volumes de dados conseguiram ser rapidamente processados, o que contribuiu para que se desenvolvessem tecnologias relacionadas ao *Big Data*.

Os *chatbots* dos anos 2020 conseguem processar e responder em Linguagem Natural (LN), mesmo em um conjunto de dados massivos. Isso decorre por conta do avanço do Processamento de Linguagem Natural (PLN) e dos bilhões de tokens utilizados no treinamento dos seus modelos de linguagem. No entanto, eles podem também entregar informações imprecisas por não entenderem bem a pergunta. *Chatbots* tradicionais que possuem vocabulários controlados em sua base de conhecimento podem minimizar esse problema (Abrantes, 2025b; Silva, 2020), um teste empírico válido é testar se, em *chatbots* generativos, essas ferramentas também poderiam oferecer boa precisão em suas respostas.

Dessa forma, a partir do protótipo de tesauro em Propriedade Intelectual (PI) de Pereira (2022), essa pesquisa elaborou uma estruturação de um protótipo de vocabulário controlado, no mesmo domínio, para servir de

DOI http://dx.doi.org/10.1590/1981-5344/61756

contexto para um protótipo de *chatbot* criado também nesta pesquisa. Esses protótipos tiveram como base os princípios teóricos da Organização do Conhecimento (OC) e da IA. Quanto a isso, vale destacar que a OC está presente na sociedade desde a época aristotélica.

Para Lima e Maculan (2024, p. 3), a OC é o campo que "[...] trata da organização do conhecimento registrado [...]" e, nesse sentido, pode abranger a forma "[...] de como o conhecimento pode ser entendido, organizado, descrito e representado, de tal forma que possa ser acessado adequadamente e disponibilizado para qualquer um que o busque". Nessa direção, é importante frisar que a pesquisa de Pereira (2022) foi elaborada e validada por professores do Programa de Pós-Graduação em Propriedade Intelectual (PPGPI) do Instituto Nacional da Propriedade Industrial (INPI).

A PI no mundo é normatizada pela *World Intellectual Property Organization* (WIPO) e possui centenas de países membros (WIPO [2024]) que, por conta dos mais diversos acordos internacionais, facilitam a própria organização da PI de modo global. No entanto, é um campo em que ainda há muitas dúvidas, o que pode prejudicar a capacidade de inovação de um país, pois desconhecendo as regras, pessoas podem perder direitos que não saberiam que poderiam ter. Dessa forma, um protótipo de *chatbot* que responda perguntas sobre PI, com garantia literária, pode agregar relevância à recuperação da informação no campo da inovação. No entanto, embora Raj (2019) explique que a partir do auxílio em diversas questões laborais os *chatbots* possam contribuir com a ampliação da produtividade, as respostas desses *bots* precisam ser verdadeiras, críveis e confiáveis.

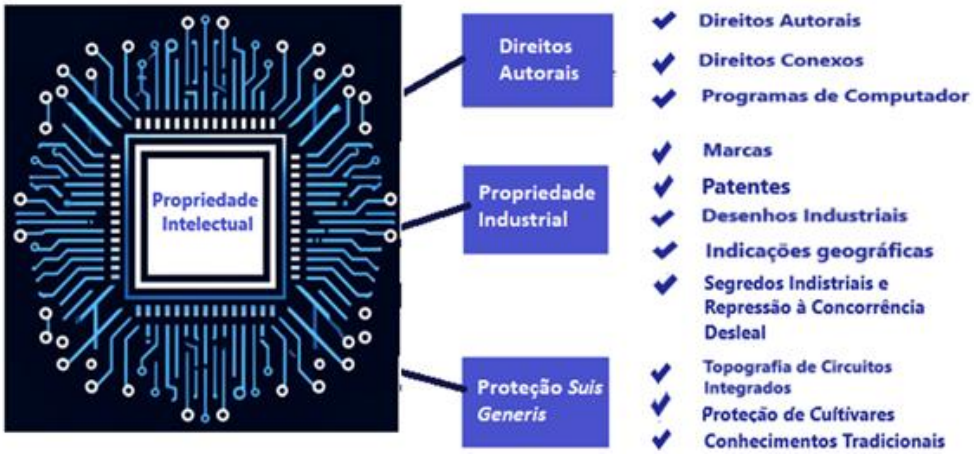
Nessa perspectiva, esta pesquisa tem o objetivo de, a partir da criação de dois protótipos no domínio de PI, analisar se o protótipo de *chatbot* está realmente contribuindo para esse domínio, se está fornecendo respostas precisas, com definições que possuem garantia literária, e não está inventando respostas que não existem. Para isso, na seção dois, o domínio da PI foi melhor explicado; na seção três, foram abordados os Sistemas de Organização de Conhecimento (SOCs), com foco nos vocabulários controlados e sua importância; na seção quatro, discorreu-se sobre IA; na seção cinco, foi estruturada uma metodologia bem detalhada; na seção seis, estão as análises e a discussão dos resultados desta pesquisa e,

posteriormente, vêm suas conclusões e referências. A seção seguinte aborda o tema da PI, na medida em que os protótipos são sobre esse domínio do conhecimento.

2 PROPRIEDADE INTELECTUAL

No decorrer do tempo, a humanidade mostrou-se capaz de inovar, de criar coisas novas e aperfeiçoar as existentes. Isso tornou o progresso científico e tecnológico possível, no entanto, para garantir os direitos sobre essa atividade inovativa, existe a PI. Ela irá garantir a proteção das criações imateriais, as quais formam um dos mais importantes ativos econômicos no mundo contemporâneo, no entanto, seu papel não se limita aos aspectos econômicos, mas abarca também o conceito de fomento e proteção dos bens culturais característicos da identidade de um povo (Abrantes, 2022). Sua melhor visualização taxonômica está disposta na Figura 1, a seguir:

Figura 1 – Taxonomia principal da Propriedade Intelectual



Fonte: Adaptada de Abrantes (2025b, p. 161).

A PI tem suas origens na demarcação de bens individuais e na busca do novo. Ela concede, por normativas legais, direitos sobre algo que uma pessoa criou e, conforme a Figura 1, sejam eles pertencentes ao direito autoral (exemplo: livros, programas de computador, peças de teatro), à propriedade industrial (exemplo: patente, marca, desenho industrial, indicação geográfica e segredo industrial) e à proteção *sui generis* (exemplo: topografia de circuitos integrados, cultivares – criação de nova variedade de plantas, conhecimentos tradicionais – tradições folclóricas e costumes e saberes de um povo) (Organização Mundial de Propriedade Intelectual; Instituto Nacional da

DOI http://dx.doi.org/10.1590/1981-5344/61756

Propriedade Industrial, 2018). A garantia desses direitos, proporciona mais segurança no uso dos ativos da PI, isso permite que as pessoas possam desenvolver suas atividades criativas com maior amparo jurídico.

Quando se pensa no desenvolvimento industrial de um país, onde milhões de reais são investidos para gerar inovação, ter patentes, por exemplo, com uma boa aplicação mercadológica, significa que investimentos poderão ter retorno, mesmo que não seja algo garantido, o fato é que o invento está protegido legalmente e pode trazer retorno financeiro. Todo esse movimento cíclico favorece a inovação e pode contribuir com o avanço tecnológico de um país.

Tecnologias desenvolvidas com aplicações de IA, por exemplo, têm direito à PI e podem auxiliar nesse processo de inovação na medida que a informação pode ser acessada e compartilhada mais facilmente (Abrantes, 2025b). No entanto, a mesma autora afirma que esse conhecimento precisa ser verdadeiro, comprovado, possuir informações com garantia literária que darão o respaldo para as pessoas terem acesso à informação de forma rápida e eficiente. Quanto a isso, vocabulários controlados podem ser muito úteis, conforme explica a seção seguinte.

3 VOCABULÁRIOS CONTROLADOS NA PERSPECTIVA DO CAMPO DA ORGANIZAÇÃO DO CONHECIMENTO

A procura por alguma forma de viver, com mais comodidade, é inerente à natureza do ser humano e, na medida em que procura meios para isso acontecer, ele estuda, pesquisa, trabalha e gera informações, que precisam ser devidamente organizadas e representadas. No decorrer do tempo, essa problemática informacional se torna mais imperiosa e, no final do século XIX, os meios para se organizar o conhecimento mostraram-se, cada vez mais, mandatórios. Nesse período, na Europa, emergiu o Movimento Bibliográfico (Mendes, 2016).

A partir desse movimento, foram surgindo obras e autores que fizeram parte da construção do campo da OC. No final do século XIX, podem-se citar a Classificação Decimal Universal (CDU), criada por Paul Otlet e Henri La Fontaine, e a Classificação Decimal de Dewey (CDD), por Melvil Dewey; na década de 1930, a Classificação Bibliográfica de Bliss e a Teoria da

Classificação Facetada, por Ranganathan (Piedade, 1977). Segundo a *International Society for Knowledge Organization* (ISKO) (Ohly, 2018), Bliss e Dahlberg estão entre os pesquisadores que criaram o termo OC.

Nesse contexto, com a dinâmica da 1ª Guerra Mundial e posteriormente com a crise de 1929, houve mudanças econômicas e sociais e se intensificou a produção científica e tecnológica, ocasionando o aumento da produção da informação (Pinheiro; Loureiro, 1995). Por conta disso, muitos pesquisadores voltaram-se para o estudo de formas melhores de organização do conhecimento, especialmente devido à intensificação das preocupações, no contexto da 2ª Guerra Mundial, da questão da explosão da informação. Esse movimento propiciou a percepção de que era fundamental a formalização da OC como campo científico. Em 1989, foi criada a ISKO, uma sociedade científica, que possui interdisciplinaridade com a Ciência da Informação (CI), Linguística, Ciência da Computação, entre outras (ISKO, [2023?]). A OC é entendida por Dahlberg (1993, p. 211, tradução nossa) 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.

Os conceitos, unidades do conhecimento, devem estar devidamente representados por termos que integram SOCs, que consistem em sistemas e são entendidos por Mazzocchi (2018, p. 54, tradução nossa) como "um termo genérico usado para se referir a uma ampla gama de itens (por exemplo, cabeçalhos de assunto, tesauros, esquemas de classificação e ontologias)". Carlan e Medeiros (2011, p. 56) afirmam que "são ferramentas semânticas com vocabulários estruturados e formalizados, usadas para o tratamento e a recuperação da informação". Os SOCs desempenham a função de representar um domínio específico. Nesse quesito, Smiraglia (2012, p. 114, tradução nossa), citado por Hjørland (2017), define domínio como "uma unidade de análise para a construção de um SOC". Assim, o autor entende que existe uma base ontológica de um determinado conhecimento consolidado, a partir de hipóteses comuns, que concordam semanticamente.

Quanto a esse entendimento, os vocabulários controlados são um tipo de SOC e, para Campos (2021, p. 80), "é uma ferramenta terminológica que possibilita determinar o melhor termo que deve ser usado na representação de

DOI http://dx.doi.org/10.1590/1981-5344/61756

informações". Além disso, a mesma autora alerta que é preciso observar a natureza terminológica que existe em cada domínio do conhecimento. Nesse sentido, Sena e Martins (2022) ensinam que vocabulários controlados têm a função de fazer um controle terminológico a partir da semântica de um determinado domínio. Portanto, na medida em que "expressar a intenção da consulta usando poucas palavras restringe a semântica do que pode ser expresso" (Baeza-Yates; Ribeiro-Neto, 2013, p. 21), é fundamental que vocabulários controlados possuam boa padronização terminológica para que consigam ser assertivos na recuperação da informação desejada pelo usuário.

Conforme Campos (2021), os vocabulários controlados precisam representar bem seu conteúdo conceitual em determinado contexto; para isso, existem as definições que vão dar sentido ao domínio representado. Sendo assim, nos vocabulários controlados, estão os termos que irão designar os conceitos que integram o sistema de conceitos de um dado universo. Esses conceitos, expressos primeiramente de forma abstrata, assumem forma verbal e são entendidos a partir de suas características que fazem parte da sua definição (Dahlberg, 1978).

Sob essa perspectiva, há de se ressaltar que a LN apresenta muitos sinônimos de palavras e os sistemas de recuperação da informação precisam também identificar essas situações. Esses sinônimos de palavras, quando tratados no contexto da OC, são controlados por remissivas e apenas um termo é selecionado como descritor para designar determinado conceito. Dessa forma, termos equivalentes, diferentes regionalismos ou expressões idiomáticas com mesmo significado, podem igualmente ser recuperados com eficiência. Nesse ponto, Harpring (2016) explica que anéis de sinônimos são muito úteis em vocabulários controlados, pois eles expandem a recuperação da informação. Segundo a autora, isso acontece porque esses anéis podem conter também quase-sinônimos, que possuem significado próximo a um determinado conceito.

Nesse contexto, em uma relação interdisciplinar com o campo da IA, Silva (2020) afirma que o controle de sinônimos é um trabalho importante a ser feito na base de conhecimento de *chatbots* tradicionais, pois isso permite que essa aplicação de IA identifique melhor palavras semanticamente idênticas. Corroborando com Silva (2020), Abrantes (2025b) salienta que *chatbots*

tradicionais ou híbridos (com regras e com *machine learning*/*deep learning*), que possuem bases de conhecimento com termos padronizados, possibilitam minimizar a ambiguidade que existe na LN. Assim, quando se busca recuperar uma determinada informação em um sistema, isso ocorre com muito mais precisão. Dessa forma, no contexto de Abrantes (2025b), seja a dúvida do usuário "em quais **naturezas as patentes** se enquadram?" ou "quais **modalidades de patentes** existem?" ou "quais são os **tipos de patentes**", por conta do uso do vocabulário controlado dentro do sistema de IA, ele poderá ter a capacidade de identificar palavras-chave dessas questões, interligadas a termos tratados, com suas devidas definições e sinonímias, o que o permite perceber que elas têm o mesmo significado.

Entende-se que, mesmo com palavras diferentes, o *chatbot* passará a reconhecer seu conceito. Nesse sentido, Dodebei (1997, p. 73) explica que:

> [...] do ponto de vista epistemológico, o conceito pode ser analisado segundo ações mentais, ou seja, a habilidade da mente em comparar algo novo com conhecimento adquirido.

Embora essa seja uma ação humana, sistemas de IA procuram imitar essas ações, inclusive reconhecendo o mesmo conceito em diferentes palavras. Quando se trata de aplicações de IA generativas, essa identificação do significado pode ocorrer por conta das *embeddings* e outras aplicações que atuam em sistemas de IA, explicadas na seção seguinte. Como será observado, o campo da IA também se preocupa em criar e usar ferramentas que representem bem a informação e colaborem para que seus sistemas tenham uma boa compreensão da LN.

4 INTELIGÊNCIA ARTIFICIAL

Desde sua criação, o campo da IA, que pertence à Ciência da Computação, passou por momentos de grandes e baixos investimentos, a depender do seu retorno financeiro no mercado (Rahman, 2022). Nos anos 2020, observa-se que altos aportes financeiros voltaram à tona, em 2022, segundo a Associação Brasileira de Empresas de *Softwares* (ABES, 2023), a IA generativa obteve ainda mais recursos e, empresas como *International Business Machines Corporation* (IBM), *Google*, *Microsoft*, alocaram recursos expressivos em IA, tenha a vista a quantidade de pedidos de patente dessas empresas (APDC..., 2024).

DOI http://dx.doi.org/10.1590/1981-5344/61756

De acordo com Cao e colaboradores (2018), os modelos probabilísticos dos anos 1950 e 1960 foram evoluindo até o ponto que possibilitou o surgimento das IAs generativas, que são capazes de criar coisas novas e, a partir de suas bases de conhecimento, elas reúnem condições de fazer PLN, ter visão computacional e criar imagens, vídeos, músicas, escrever um livro, apenas com um comando humano. Elas usam algoritmos de *machine learning* e *deep learning* que simulam as redes neurais humanas para chegar nesses resultados. Esses algoritmos usam o modelo de LN *transformers* que conseguem fazer PLN com rapidez e eficiência (IBM, 2024). No entanto, *chatbots* generativos também "alucinam" – termo técnico (Conselho Nacional de Ética para as Ciências da Vida, 2024); e podem trazer respostas fora da realidade e imprecisas, questão que vocabulários controlados podem resolver. Algoritmos de IA e vocabulários controlados, de forma integrada, podem trazer respostas com qualidade, o que é uma perspectiva importante para esta pesquisa.

Chatbots como o *ChatGPT* da *OpenAI*, usam tecnologias de IAs generativas e podem fazer uso de técnicas como a *Retrieval-Augmented Generation* (RAG), que admite o emprego do *chunking*. O RAG faz com que o *bot* consiga localizar, em documentos, a informação que o usuário busca de modo mais preciso (Rivera Berrío, 2024). Essa técnica também possui um hiperparâmetro de temperatura e, com valor zero, faz com que as respostas do *bot* sejam menos criativas e associadas aos próprios documentos a serem analisados (Rivera Berrío, 2024; IBICT, 2024). No que se refere às limitações técnicas do RAG, salienta-se a possibilidade de haver recuperação semântica para conteúdos semelhantes, mas que não obrigatoriamente tragam precisão à pergunta do usuário. Por exemplo, na base de contexto fornecida, existem, conceitos sobre direitos autorais e PI. Como são conceitos com significado semântico próximo, um *chatbot* poderia entender que o conceito requisitado se referiria ao direito autoral, mas, na verdade, ele diria respeito ao conceito de PI. O trecho recuperado não seria algo aleatório, pois teria relação com a pergunta do usuário, no entanto, poderia não ter a precisão necessária para responder corretamente à pergunta.

Quanto ao *chunking*, ele possibilita que o *corpus* textual fornecido ao *chatbot* tenha a possibilidade de reconhecer de forma mais precisa a resposta pretendida pelo usuário, assim, ele recupera melhor a informação

(Finardi *et al*., 2024). Ele divide os textos em *chunks*, que são pedaços dos textos. Porém, o *chunking* pode trazer problemas, porque talvez a resposta certa possa estar dividida entre dois pedaços do *corpus* textual, o que fará a análise de *embeddings* relativa à proximidade semântica das palavras ter dificuldades nos resultados e o *bot* possa apresentar conflito (Qu; Tu; Bao, 2024). Uma forma de contornar esse problema de descontinuidade da formação dos *chunks* está no uso de marcadores que delimitam exatamente o conteúdo de cada *chunk*, desse modo, descontinuidades seriam evitadas. Esta solução é viável para arquivos menores, onde a inserção manual de tais marcadores é possível. Portanto, sua utilidade ao aperfeiçoar o resultado da busca do usuário ainda está em investigação.

O *Facebook AI Similarity Search* (FAISS) também é uma ferramenta útil quando se busca que o *chatbot* traga respostas corretas. "É uma biblioteca que permite que desenvolvedores pesquisem rapidamente por *embeddings* de documentos multimídia que sejam semelhantes entre si" (Meta, 2024, tradução nossa). Os *embeddings* são "modelos que representam dados, como texto ou imagens, em um espaço vetorial" (LangChain…, 2024a, tradução nossa). Ou seja, as palavras em um texto se transformam em números e, quanto mais as palavras têm semânticas parecidas, mais esses números ficam próximos em um espaço vetorial. É necessário fazer essa transformação em números para uma melhor compreensão computacional.

Quanto ao *LangChain*, ele é um *framework* em *Python* que consegue se conectar a várias *Large Language Models* (LLMs) por meio de uma única interface e integrar diferentes LLMs em um mesmo projeto (LangChain…, 2024b). Isso facilita bastante testes de desenvolvedores nos mais diversos modelos de linguagem de sistemas de IA, pois passa a não ser necessária uma programação específica para cada um desses modelos.

Todas essas ferramentas são importantes quando se buscam respostas corretas baseadas em garantia literária. Além disso, perguntas dos usuários nem sempre são bem elaboradas, e essas ferramentas e técnicas têm potencial de conseguir compreender realmente qual informação o usuário procura. Nesse sentido, a seção seguinte descreve como o protótipo de *chatbot* criado para esta pesquisa foi elaborado e analisado.

DOI http://dx.doi.org/10.1590/1981-5344/61756

5 METODOLOGIA

A partir do protótipo de tesauro do domínio da PI de Pereira (2022), disponível para acesso *online* na Universidade Federal do Rio Grande do Sul (UFRGS, 2022), uma pesquisadora da área da CI elaborou uma estruturação do protótipo de vocabulário controlado no mesmo domínio e disponibilizou em seu *GitHub*[2]. O tesauro de Pereira (2022) foi estruturado a partir do *corpus* textual de uma tese de doutorado defendida em 2022, no PPGPI/INPI. Para criar o tesauro, Pereira (2022) fez análise bibliométrica, no contexto da Teoria Zipfiana, e análise de conteúdo na linha de Laurence Bardin. Esse SOC foi validado no curso da pesquisa de doutorado no PPGPI/INPI.

Para a estruturação do protótipo de vocabulário controlado, foi necessário fazer a exportação no formato .pdf do tesauro de Pereira (2022). Posteriormente, foi criado um documento no formato .txt, com os termos desse tesauro, suas definições e sinonímias foram mantidas; todos os grafos e relacionamentos sintagmáticos do tesauro de Pereira (2022) foram excluídos.

Esse protótipo de vocabulário controlado fez parte da estruturação do protótipo de *chatbot* em PI, denominado, "ChatbotPI", que foi criado e avaliado nesta pesquisa. Quanto ao protótipo de *chatbot*, disponível em Abrantes (2025a), ele foi construído em linguagem *Python*, utiliza o *framework LangChain*, a biblioteca *FAISS* (age como um *Vector Store*), usou o modelo GPT-4o mini, da *OpenAI*, e técnicas do RAG com a abordagem *chunking*. Para hospedagem na internet de forma gratuita do *chatbot* criado, foi necessário abrir uma conta na plataforma Streamlit.io (2024). Além disso, foi preciso possuir um perfil no *GitHub*, onde foram colocados os arquivos que posteriormente foram transferidos de forma automática para o perfil no Streamlit.io. No entanto, para o *bot* ser alimentado por uma LLM, uma das pesquisadoras deste estudo criou um perfil pago na *OpenAI* para utilizar o modelo de linguagem do *ChatGPT*. A estrutura de programação do *bot* foi construída e cedida pelo pesquisador Dr. Antonio Carlos Souza de Abrantes.

Dentro dessa programação, uma pesquisadora da área da CI incluiu e ajustou os dados necessários para que o *bot* pudesse responder às perguntas sobre o domínio de PI no espectro epistemológico da OC e disponibilizou em

[2] https://github.com/pcotrim/ChatbotPI.

seu *GitHub*[3] (Abrantes, 2024). Houve separação entre as definições, com suas devidas garantias literárias e sinonímias no *corpus* textual (contexto), com o carácter # (tralha). Na programação do *bot* foi realizado o seguinte *prompt*:

> Você é um assistente virtual que busca responder perguntas dos usuários. Responda às perguntas de acordo com as notas de definição dos termos disponíveis no contexto. Se a você forem solicitadas as referências bibliográficas das definições, responda de acordo com a garantia literária que é informada no contexto. Nele estão disponíveis também sinônimos de alguns termos, assuma que esses termos que possuem sinônimos, têm as mesmas definições e referências bibliográficas, portanto, entregue respostas sem ambiguidades. Somente entregue respostas que constem nesse contexto (Abrantes, 2024)[4].

As perguntas que este *chatbot* deveria responder não poderiam ter alucinações de nenhuma hipótese, seu hiperparâmetro de temperatura é zero. As perguntas, disponíveis em Abrantes (2024), foram obtidas a partir de buscas no *Google*, no dia 03/12/2024, onde foi realizada a estratégia de busca "perguntas frequentes" AND "propriedade intelectual". Nesta pesquisa optou-se por escolher perguntas frequentes apenas de instituições federais de ensino brasileiras para trazer à tona as dúvidas mais usuais de pessoas envolvidas com a inovação no setor da educação pública no Brasil. Dessa forma, foram escolhidas cinco perguntas que estavam no campo semântico que o *bot* teria capacidade para responder, sabendo-se previamente que ele deveria trazer respostas que trouxessem as definições dos termos inseridos em sua programação. Destarte, algumas das perguntas frequentes, da Universidade Federal do Amazonas (UFAM, 2023) e da Universidade Federal da Integração Latino-Americana (UNILA, 2021) participaram dessa pesquisa com as perguntas disponíveis em Abrantes (2024). A seguir, o Quadro 1 resume as informações desta seção.

Quadro 1 – Etapas de construção dos protótipos de vocabulário controlado/ *chatbot*

Etapa	Descrição
Proveniência dos dados iniciais	Tesauro criado por Pereira (2022), a partir de tese do PPGPI/INPI (2020). Baseado em análise bibliométrica (Teoria Zipfiana) e análise de conteúdo (Bardin). Validado durante o doutorado.
Conversão dos dados iniciais	Tesauro exportado em .pdf e convertido em um protótipo de vocabulário controlado em .txt. Ele possui termos, definições e sinonímias. Grafos e relações sintagmáticas foram excluídos.

[3] https://github.com/pcotrim/ChatbotPI.
[4] Arquivo disponível em Abrantes (2024) com definições, referências e sinônimos.

DOI http://dx.doi.org/10.1590/1981-5344/61756

Etapa	Descrição
Vocabulário controlado	Estruturação feita por pesquisadora da CI e disponibilizada em *GitHub*[5].
Integração com *ChatbotPI*	Protótipo de *chatbot* criado em *Python*. *Framework LangChain*, FAISS (*Vector Store*), modelo GPT-4o mini (*OpenAI*) e técnicas RAG com *chunking*.
Hospedagem	Conta gratuita no Streamlit.io, integração automática com *GitHub*. Para uso da LLM, é necessário perfil pago na *OpenAI*.
Programação base	Estrutura de programação cedida pelo Dr. Antonio Carlos Souza de Abrantes.
Modelo de *embeddings* utilizado	embeddings = HuggingFaceEmbeddings(model_name='sentence-transformers/all-MiniLM-L6-v2')
Parâmetros do FAISS	persisted_vectorstore = FAISS.load_local("faiss_index_datamodel_law", embeddings, allow_dangerous_deserialization=True)
Customização	Pesquisadora da CI inseriu dados (definições, sinonímias, garantias literárias) em formato textual com separação por caractere # como contexto para o *chatbot*, e formulou o *prompt* que ele deveria responder.
Prompt resumido do *bot*	Responder apenas com definições disponíveis no contexto, incluindo referências e sinônimos, sem ambiguidades.
Coleta de perguntas	Buscas em 03/12/2024 no *Google* com "perguntas frequentes" AND "propriedade intelectual". Seleção de FAQs de instituições federais de ensino brasileiras.

Fonte: Elaborado pelas autoras (2024).

O Quadro 1 resume o percurso metodológico deste artigo, esclarece informações importantes sobre o tesauro de Pereira (2022), sua transformação em um protótipo de vocabulário controlado e sua posterior integração em um protótipo de *chatbot*. Os sites e *softwares* que fizeram parte da construção dessas ferramentas, o *prompt* e informações sobre a coleta dos dados são também pontos-chave da metodologia deste artigo, por isso fizeram parte do Quadro 1.

É válido ressaltar que o modelo de *embeddings* utilizado, "HuggingFaceEmbeddings" "mapeia frases e parágrafos para um espaço vetorial denso de 384 dimensões e pode ser usado para tarefas como agrupamento ou busca semântica" (Hugging Face, 2026, tradução nossa). Ele é rápido e simples, qualidades que facilitam sua implementação num protótipo. No que se refere à utilização do caracter #, ele favorece a estrutura

[5] https://github.com/pcotrim/ChatbotPI.

lógica do conteúdo e evita descontinuidades no texto de cada *chunk*. Quanto à configuração da recuperação (*top-k*), para cada pergunta do usuário, o FAISS retorna apenas o um *chunk* mais semelhante. Quanto à arquitetura mais detalhada do *pipeline* RAG, ele possui a etapa da indexação, da consulta e da geração da resposta, de modo resumido, é estruturado da seguinte forma: pergunta -> *embedding* da pergunta -> FAISS -> *similarity search top-k* = 1 -> *chunk* + *score* -> *system instruction*: pergunta + histórico + contexto recuperado -> GPT-4o mini -> resposta -> *Streamlit* -> usuário.

Com relação às métricas, neste estudo não foram usadas métricas quantitativas específicas, a avaliação foi feita com base na análise da resposta apresentada pelo *chatbot* e comparada com os conceitos dos termos no vocabulário controlado. Para tal, entende-se que os *chatbots* utilizam na busca a tokenização (frequência de uso das palavras) que alinha-se aos modelos semânticos zipfinianos da linguística cognitiva e da bibliometria na CI. As respostas sofreram influência das *embeddings* e do modelo utilizado. Dessa forma, o FAISS + *embeddings* fazem a recuperação dos dados, enquanto o GPT-4o mini, em sua maior parte, reconhece os dados e escreve a resposta a partir do contexto. Esse modelo foi escolhido pela pesquisa, por ser em 2024, segundo a *OpenAI* (2024), um modelo de baixo custo, mas com 128 mil *tokens*, até 16 mil t*okens* de saída, ser bem avaliado em vários *benchmarks*, possuir inteligência textual e possibilidade de desenvolvimento de aplicativos. Quanto à visualização da integração, entre um SOC e uma aplicação de IA, isso será mostrado e discutido na seção seguinte.

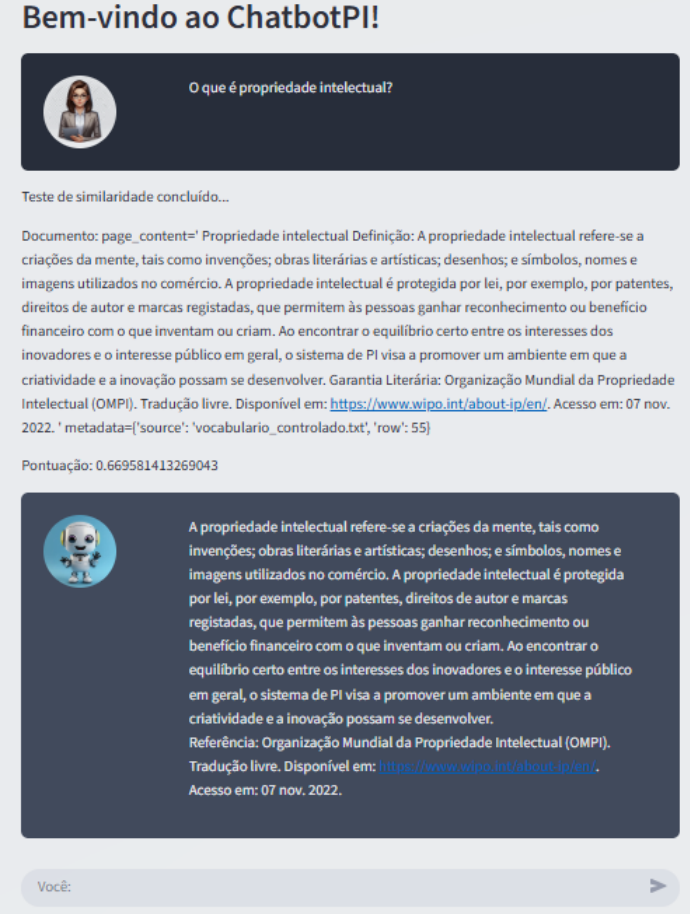
6 ANÁLISE E DISCUSSÃO DOS RESULTADOS

Após todos os ajustes no protótipo de tesauro de Pereira (2022), foi criada a estrutura do protótipo de vocabulário controlado, ele possui 106 termos, sendo que 39 são sinônimos. Vale destacar que o protótipo de *chatbot*, "ChatbotPI", mostra ao usuário em sua interface em que posição do contexto estava a definição e a garantia literária da resposta apresentada, como também apresenta o resultado final da resposta ao usuário já com a atuação dos algoritmos de LLM, conforme mostrado em Abrantes (2025a).

DOI http://dx.doi.org/10.1590/1981-5344/61756

Dito isso, a primeira pergunta foi "o que é propriedade intelectual?". Vista na Figura 2.

Figura 2 – O que é propriedade intelectual?



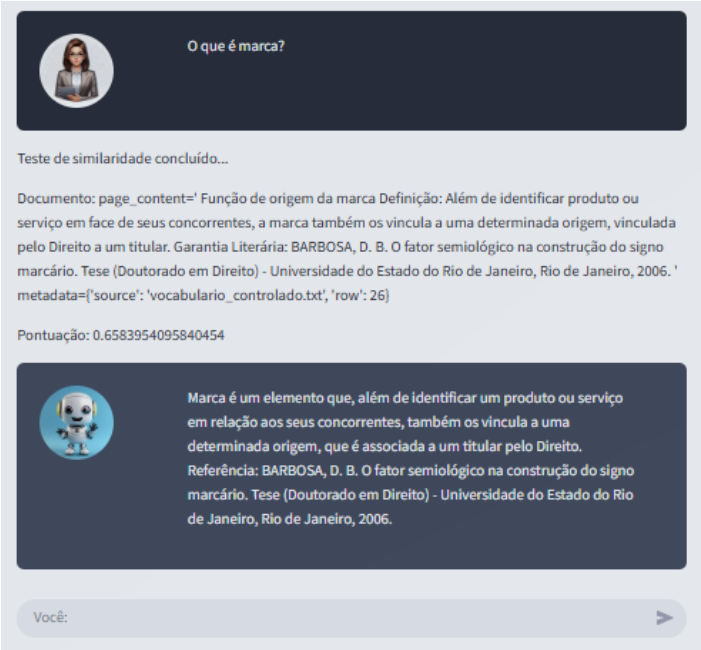
Fonte: ChatbotPI (Abrantes, 2025a).

A Figura 2 indica que o *ChatbotPI* trouxe exatamente a mesma definição e a garantia literária do protótipo do vocabulário controlado. Sua pontuação de similaridade após análise vetorial foi de: 0.669581413269043. Isso mostra que o *ChatbotPI* teve um bom reconhecimento da pergunta, a classificou bem, e conseguiu localizar no contexto onde estaria a melhor resposta de acordo com *prompt* inserido internamente em sua programação. O sistema de IA, com um protótipo de vocabulário controlado em seu contexto, propiciou uma boa estrutura relacional entre os dados. Isso resultou em precisão e eficácia na recuperação da informação, elementos que necessariamente precisam estar presentes na comunicação entre usuários e sistemas de informação, de acordo com Kobashi (2008). Na perspectiva da IA generativa, o programa calculou os *embeddings* de cada *chunk* em que o arquivo .txt foi dividido e comparou com os *embeddings* da pergunta

formulada pelo usuário. Esse cálculo de similaridade foi realizado internamente pelo próprio *FAISS Vector Store*, ainda antes do programa fazer o acesso à LLM da *OpenAI* propriamente dita, que atuou na composição da resposta final.

A segunda pergunta foi: "O que é marca?". O *ChatbotPI* se equivocou e trouxe a definição e a garantia literária erradas. Sua pontuação de similaridade após análise vetorial foi de: 0.6583954095840454. No entanto, a resposta não está correta segundo indica o contexto inserido na programação do protótipo. Houve imprecisão do *chatbot*, como mostra a Figura 3.

Figura 3 – O que é marca?



Fonte: ChatbotPI (Abrantes, 2025a).

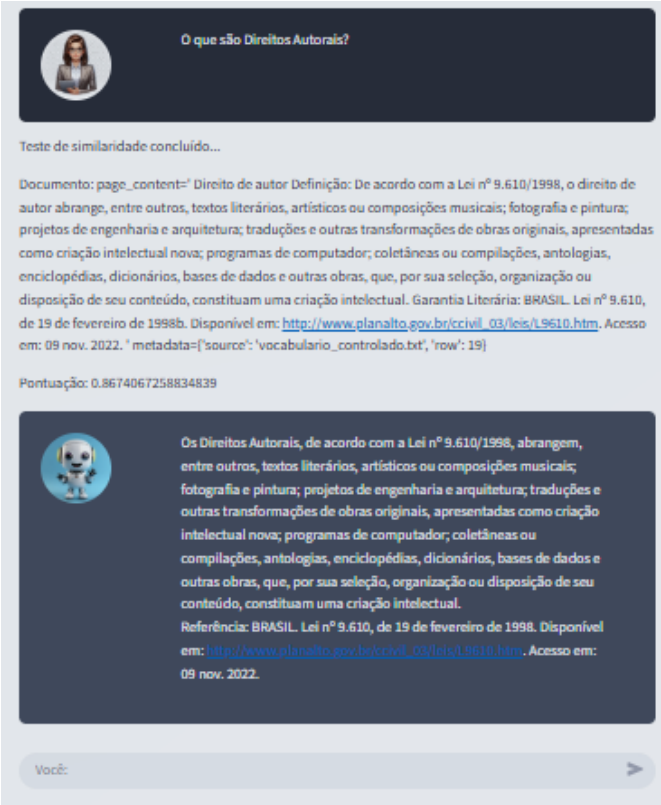
Conforme Figura 3, a resposta do *ChatbotPI* foi "Marca é um elemento que, além de identificar [...]". A resposta da definição do protótipo do vocabulário controlado é: "Marca é um sinal distintivo cujas funções principais são identificar a origem e distinguir produtos ou serviços de outros idênticos, semelhantes ou afins de origem diversa [...]" (Abrantes, 2024); ou seja, errou a definição. Após o teste de similaridade realizado pelo *FAISS Vector Store* nos *chunks*, o algoritmo selecionou um único bloco a partir do teste de similaridade, isso proporcionou a entrega de resultado que o *bot* mais identificou com a pergunta que constava no *corpus* textual do vocabulário controlado, mas houve equívoco na identificação da melhor resposta. Uma vez detectado o *chunk* que o *bot* reconheceu como mais relevante, foi

DOI http://dx.doi.org/10.1590/1981-5344/61756

montado um *prompt* interno de maneira automática pelo *bot*, esse *prompt* continha: as instruções próprias que um *chatbot* usa em sua programação, as perguntas de seu histórico, a pergunta atual e o *corpus* textual do *chunk* selecionado pelo *bot*. Com esse *prompt*, o *ChatbotPI* fez então a consulta ao modelo de linguagem da *OpenAI* para fornecer a resposta em LN, que nesse caso reconheceu mal a pergunta.

A terceira pergunta foi: "O que são Direitos Autorais?", conforme a Figura 4.

Figura 4 – O que são Direitos Autorais?



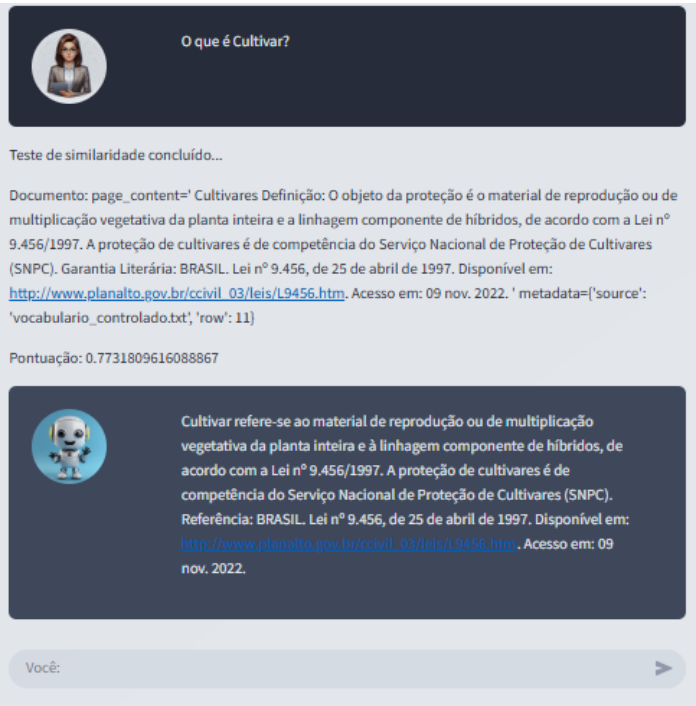
Fonte: ChatbotPI (Abrantes, 2025a).

A Figura 4 indica que a pontuação de similaridade após análise vetorial do *bot* foi de: 0.8674067258834839. No entanto, o protótipo confundiu a definição de direito de autor com direito autoral. Por serem palavras semanticamente muito próximas, o *ChatbotPI* não conseguiu trazer a resposta correta segundo indicava o contexto inserido na sua programação, dificuldade apontada por Qu, Tu e Bao (2024) que o uso do RAG pode trazer. A resposta esperada seria "de acordo com a Lei nº 9.610/1998, o direito autoral

abrange o direito de autor e os direitos conexos. O programa de computador também é considerado direito de autor mas sofre efeito [...]" (Abrantes, 2024).

Ao analisar um sistema de IA, a partir da experimentação, se colhe resultados positivos e negativos, pois a construção do conhecimento acontece assim, com acertos e erros e de modo verificável. Dessa forma, após testes, é possível construir o conhecimento necessário para que erros sejam identificados e posteriormente mais estudados para serem corrigidos. Na vetorização dos *chunks* do *corpus* textual, foi entregue uma resposta que seria a resposta a oferecer ao usuário. A configuração do *ChatbotPI* também o impossibilitou de fornecer mais de uma opção de resposta a partir de outras definições. Conforme os arquivos no *GitHub*[6] de Abrantes (2024), o *bot* foi programado com uma variável denominada "k" onde só se permite trazer um bloco de definição e garantia literária na resposta ao usuário. Para esta pesquisa, optou-se que ele fosse programado para trazer a definição exata ou a mais próxima disso. Nesse ponto, o *ChatbotPI* trouxe a mais próxima, pois trouxe uma definição alinhada com a de direito autoral, mas que não seria a 100% correta tendo como base seu contexto. A quarta pergunta foi "O que é Cultivar?", conforme a Figura 5.

Figura 5 – O que é Cultivar?



Fonte: ChatbotPI (Abrantes, 2025a).

[6] https://github.com/pcotrim/ChatbotPI.

DOI http://dx.doi.org/10.1590/1981-5344/61756

De acordo com a Figura 5, a pontuação de similaridade foi 0.7731809616088867. A resposta está correta e atenderia perfeitamente ao usuário em sua dúvida, não houve alucinação ou imprecisão por parte do *bot* que reconheceu bem a pergunta e a enquadrou corretamente no *corpus* textual referente ao protótipo de vocabulário controlado, trazendo praticamente a mesma definição. Para Campos e Gomes (2022, p. 123),

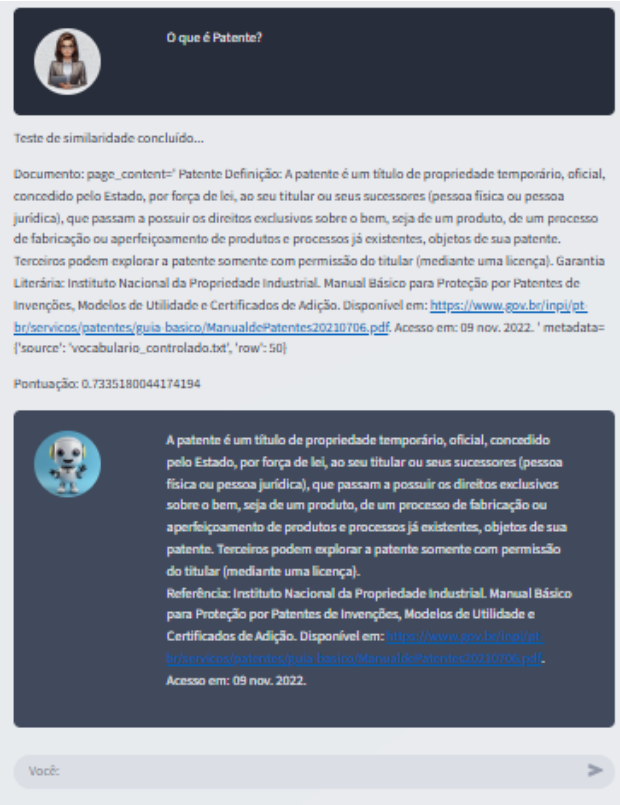
> […] as definições se apresentam como um elemento fundamental para a construção de instrumentos terminológicos, pois elas permitem posicionar o conceito em um dado universo.

Nesse sentido, o conceito de "cultivar" conseguiu ser bem representado na avaliação da similaridade semântica que as *embeddings* proporcionam a um *chatbot* generativo, nesse caso ao *ChatbotPI*.

A quinta pergunta foi "O que é Patente?". Para essa pergunta consta uma definição exata no contexto do protótipo de vocabulário controlado. O *bot* teria que fazer a identificação semântica a partir do uso das *embeddings* e verificar uma definição que mais abrangesse semanticamente a pergunta. No resultado, o *bot* classificou bem a pergunta e entregou a resposta correta, o que atenderia ao usuário, conforme mostra a Figura 6.

Figura 6 – O que é Patente?

CC BY



Fonte: ChatbotPI (Abrantes, 2025a).

Com pontuação de similaridade 0.7335180044174194, a definição que o protótipo de *chatbot* trouxe abrange a resposta que está no conceito do termo "patente", conforme documento disponibilizado em Abrantes (2024). Percebe-se que o *ChatbotPI* fez uma boa análise de proximidade de similaridade semântica e identificou, a partir de seu contexto, que a palavra patente possuía em sua definição características que poderiam responder à pergunta do usuário. Aliás, definição próxima ao estilo defendido por Dahlberg (1978) e dentro da perspectiva de domínio entendida por Hjørland (2017) quando cita Smiraglia (2012). Sendo assim, pessoas que se interessem sobre a propriedade industrial, acadêmicos ou quem tiver dúvidas sobre termos da PI, poderiam ter acesso a esse tipo de informação de forma rápida e segura por meio de um *chatbot*. Com um vocabulário controlado no seu contexto, os termos já estariam tratados, ou seja, possuiriam definição e relacionamentos de sinonímias, o que traz a todo o sistema um rigor científico, teórico e prático que o campo da OC possui.

Compreende-se, portanto, com base nos resultados apresentados nesta seção e nos testes empíricos realizados, que houve sucesso parcial na experimentação. Das cinco respostas analisadas, duas apresentaram

DOI http://dx.doi.org/10.1590/1981-5344/61756

imprecisões. Entretanto, esses resultados não permitem atribuir tais erros exclusivamente à etapa de geração da LLM. Em uma arquitetura RAG, a imprecisão da resposta pode ter origem em diferentes componentes do *pipeline*, incluindo a segmentação do *corpus*, a representação semântica produzida pelo modelo de *embeddings*, a recuperação e o ranqueamento realizados a partir da busca vetorial e, posteriormente, a geração da resposta pela LLM. No caso do *ChatbotPI*, a utilização de k=1 também constitui um fator a ser considerado, uma vez que somente o primeiro *chunk* ranqueado é fornecido ao modelo generativo.

Dessa forma, a dificuldade observada na distinção de conceitos com semântica próxima, pode estar associada tanto à representação vetorial desses conceitos, quanto à etapa que faz sua recuperação e ranqueamento. Como o experimento realizado não teve como objetivo metodológico isolar sistematicamente cada um desses componentes, não há certeza sobre a origem de cada erro observado.

Observa-se, desta forma, que é necessária a continuação da investigação científica, em sistemas de IA atuando em conjunto com vocabulários controlados, para que se chegue a melhores resultados para os usuários. Nesse sentido, sugerem-se pesquisas futuras sobre essas temáticas. Quanto à seção seguinte, ela apresenta as conclusões desta investigação científica.

7 CONSIDERAÇÕES FINAIS

A partir de uma base teórica interdisciplinar que englobou os campos da PI, da OC e da IA, procurou-se chegar a um protótipo de *chatbot* que trouxesse definição e garantia literária em suas respostas. Para isso, foi criado um protótipo de vocabulário com termos e definições, garantia literária e controle das sinonímias de termos que compõem o domínio da PI, a partir de um protótipo de tesauro do mesmo domínio, nos moldes explicados por Pereira (2022).

Buscou-se analisar se o uso do RAG, com aplicação do *chunking* no *corpus* textual e o *FAISS Vector Store*, poderia ser considerado em *chatbots* para uma melhor compreensão semântica do *bot* criado para responder às perguntas dos usuários, a partir de um protótipo de vocabulário controlado;

mas em 5 (cinco) perguntas (60%), 3 (três) (40%) estavam corretas. Os resultados deste estudo são exploratórios e retratam as especificidades do experimento a partir de uma amostra reduzida, não sendo generalizáveis para outros sistemas ou configurações RAG. Embora seja possível aplicar testes estatísticos de significância a dados binários (exemplo: resposta correta/incorreta), o número limitado de respostas avaliadas diminui o poder inferencial. Ainda assim, os achados evidenciam pontos positivos e limitações do *ChatbotPI* e servem de subsídio para pesquisas futuras.

Nesse sentido, recomenda-se avaliar separadamente as diferentes etapas do pipeline RAG. Na etapa de representação semântica, podem ser comparados diferentes modelos de *embeddings*; na recuperação e no ranqueamento, podem ser testados diferentes valores de *top-k*; na segmentação, podem ser comparadas diferentes estratégias de *chunking*; e, na etapa de geração, podem ser avaliados diferentes modelos de linguagem, configurações de *prompt* e parâmetros de geração. Essa análise permitiria identificar com maior precisão em qual etapa do *pipeline* são introduzidas as imprecisões observadas.

Para tanto, conclui-se que os ajustes finos na programação dos *chatbots* precisam ser realizados com mais testes, pois é necessário que ele tenha a capacidade de reconhecer a diferença entre conceitos semanticamente parecidos, como também de realizar abstrações semânticas. Os *chatbots* precisam conseguir fazer a análise de vetorização de similaridade no documento todo e identificar corretamente cada conceito. O RAG corta o *corpus* textual em *chunks* e, por conta disso, a análise de similaridade pode ser prejudicada em algumas situações. Esses cortes podem deixar o *bot* "confuso" e fazer com que ele não consiga entregar uma resposta correta. O *prompt* também é um fator de grande importância para o *chatbot*. É preciso detalhar bem o que se quer e fazer vários testes para verificação dos resultados, e mesmo bons resultados podem ser melhorados.

Como contribuição social e acadêmica, é importante ressaltar que o *ChatbotPI* está disponível na internet em Abrantes (2025a), assim como seus códigos (Abrantes, 2024); e podem ser usados para testes utilizando o modelo GPT-54 mini, pois, por conta da rápida obsolescência tecnológica, o modelo GPT -4o mini tornou-se incompatível com a configuração do protótipo desta

DOI http://dx.doi.org/10.1590/1981-5344/61756

pesquisa. A atualização do modelo pode contribuir para respostas mais precisas, mas seus efeitos não foram objeto do experimento original. Finalmente, espera-se que este estudo contribua com todos os campos do conhecimento mencionados e incite outros estudos alinhados ao tema discutido.

REFERÊNCIAS

ABRANTES, Antonio Carlos Souza de. **Introdução ao sistema de patentes**: aspectos técnicos, institucionais e econômicos. 2. ed. Rio de Janeiro: Lumen Juris, 2022. 416 p.

ABRANTES, Paula Cotrim de. ChatbotPI. *In*: GITHUB. [*S. l.*]: Github, 2024. Disponível em: https://github.com/pcotrim/ChatbotPI. Acesso em: 04 dez. 2024.

ABRANTES, Paula Cotrim de. Bem-vindo ao ChatbotPI. *In*: STREAMLIT.APP. [*S. l.*]: streamlit, 2025a. Disponível em: https://chatbotpi.streamlit.app/. Acesso em: 8 out. 2026.

ABRANTES, Paula Cotrim de. **Inter-relação da Organização do Conhecimento com a Inteligência Artificial em Sistemas de Organização do Conhecimento**: análise empírica em protótipo de chatbot e análise em patentes no período de 2000 a 2022. 2025. Tese (Doutorado em Ciência da Informação) – Instituto Brasileiro de Informação em Ciência e Tecnologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2025b. Disponível em: http://ridi.ibict.br/handle/123456789/1426, Acesso em: 07 out. 2026.

APDC – Digital Business Comunity. IBM, Google e Microsoft lideram patentes de IA generativa. **APDC – Digital Business Comunity**, [*S. l.*], 09 fev. 2024. Atualidade internacional. Disponível em: https://www.apdc.pt/noticias/atualidade-internacional/ibm-google-e-microsoft-lideram-patentes-de-ia-generativa. Acesso em: 02 dez. 2024.

ASSOCIAÇÃO BRASILEIRA DE EMPRESAS DE SOFTWARES [ABES]. ChatGPT provoca aumento no investimento em Inteligência Artificial para 45% dos executivos entrevistados pelo Gartner. **ABES**, São Paulo, 06 jun. 2023. Noticias. Disponível em: https://abes.com.br/chatgpt-provoca-aumento-no-investimento-em-inteligencia-artificial-para-45-dos-executivos-entrevistados-pelo-gartner/. Acesso em: 30 jul. 2024.

BAEZA-YATES, Ricardo; RIBEIRO-NETO, Berthier. **Recuperação de Informação**: conceitos e tecnologia das máquinas de busca. [*S. l.*]: Bookman Editora, 2013. 590 p.

BARONE, Dante Augusto Couto; BOESING, Ivan Jorge. **Inteligência artificial**: diálogos entre mentes e máquinas. Porto Alegre: AGE : EVANGRAF, 2015. 309 p.

CAO, Yihan *et al*. A comprehensive survey of ai-generated content (aigc): A history of generative ai from gan to chatgpt **J. ACM**, v. 37, n. 4, p. 1-44, 2018. Disponível em: https://arxiv.org/pdf/2303.04226. Acesso em: 09 jun. 2024.

CAMPOS, Maria Luiza de Almeida. Tesauros: principais aspectos teóricos e metodológicos para a sua elaboração. *In*: FUJITA, Mariângela Spotti Lopes; MOREIRA, Walter (org.). **Manual do planejamento, construção e manutenção do Tesauro Unesp para bibliotecas**: do conceitual a práxis. São Paulo: Cultura Acadêmica, 2021. p. 79-93.

CAMPOS, Maria Luiza de Almeida.; GOMES, Hagar Espanha. Uma perspectiva metodológica para a elaboração de instrumentos terminológicos: a relevância das definições. **Fronteiras da Representação do Conhecimento**, v. 1, n. 2, 2022.

CARLAN, Eliana; MEDEIROS, Marisa Bräscher Basílio. Sistemas de Organização do Conhecimento na visão da Ciência da Informação. **Revista Ibero-americana de Ciência da Informação**, Brasília (DF), v. 4, n. 2, p. 53-73, ago./dez. 2011. DOI https://doi.org/10.26512/rici.v4.n2.2011.1675.

COECKELBERGH, Mark. **Ética na inteligência artificial**. Tradução Clarisse de Souza; Edgar Lyra; Matheus Ferreira; Waldyr Delgado. Rio de Janeiro: Editora UBU : Editora PUC-Rio, 2024. 192 p. (Coleção Exit, v. 15).
CONSELHO NACIONAL DE ÉTICA PARA AS CIÊNCIAS DA VIDA. **Inteligência Artificial (IA)**: inquietações sociais, propostas éticas e orientações políticas. Lisboa: Conselho Nacional de Ética para as Ciências da Vida, 2024. 61 p. (Livro Branco).

DAHLBERG, Ingetraut. A referent-oriented, analytical concept theory for INTERCONCEPT. **KO KNOWLEDGE ORGANIZATION**, v. 5, n. 3, p. 142-151, 1978.

DAHLBERG, Ingetraut. Knowledge organization: its scope and possibilities. **KO KNOWLEDGE ORGANIZATION**, v. 20, n. 4, p. 211-222, 1993.

DODEBEI, Vera Lúcia Doyle Louzada de Mattos. **O Sentido e o Significado do Documento para a Memória Social**. 1997. Tese (Doutorado em Comunicação e Cultura) – Faculdade de Comunicação, Instituto Brasileiro de Informação em Ciência e Tecnologia, Universidade Federal do Rio de Janeiro; Rio de Janeiro, 1997.

FINARDI, Paulo *et al*. The Chronicles of RAG: the Retriever, the Chunk and the Generator. **arXiv preprint**, p. 1-16, 2024. Disponível em: https://arxiv.org/pdf/2401.07883. Acesso em: 29 nov. 2024.

HARPRING, Patrícia. **Introdução aos vocabulários controlados**: terminologia para arte, arquitetura e outras obras culturais. Tradução Christina Maria Müller. São Paulo: Secretaria da Cultura do Estado: Pinacoteca de São Paulo: ACAM Portinari, 2016. 288 p.

DOI http://dx.doi.org/10.1590/1981-5344/61756

HJØRLAND, Birger. **Domain Analysis. Knowledge Organization**. v. 44, n. 6, p. 436-464. 2017. Disponível em: https://www.isko.org/cyclo/domain_analysis#2. Acesso em: 09 ago. 2025.

HUGGING FACE. **Sentence-transformers/all-MiniLM-L6-v2**. [*S. l.*]: Hugging Face, 2026. Disponível em: https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2/blob/main/README.md. Acesso em: 14 set. 2026.

INTERNATIONAL BUSINESS MACHINES CORPORATION [IBM]. **What is a transformer model?** [*S. l.*]: IBM, [2024]. Disponível em: https://www.ibm.com/topics/transformer-model. Acesso em: 04 dez. 2024.

INTERNATIONAL SOCIETY FOR KNOWLEDGE ORGANIZATION (ISKO). **ISKO Internacional**. [*S. l.*]: ISKO, [2023?] Disponível em: https://portal.isko.org/. Acesso em: 22 abr. 2023.

INSTITUTO BRASILEIRO DE INFORMAÇÃO EM CIÊNCIA E TECNOLOGIA [IBICT]. **Workshop de Aplicações de Inteligência Artificial**. [*S. l.*], 2024 YouTube: @IBICTbr. Disponível em: https://www.youtube.com/watch?app=desktop&v=YOG48oafNsU. Acesso em: 29 nov. 2024.

KOBASHI, Nair Yumiko. **Vocabulário controlado**: estrutura e utilização. Brasília, DF: Escola Nacional de Administração Pública, 2008.

LANGCHAIN Docs. Conceptual guide. *In*: LANGCHAIN Docs. **Concepts**. [*S. l.*]: LangChain Docs, 2024a. Disponível em: https://python.langchain.com/docs/concepts/. Acesso em: 04 dez. 2024.

LANGCHAIN Docs. Why LangChain? *In*: LANGCHAIN Docs. **Concepts**. [*S. l.*]: LangChain Docs,2024b. Disponível em: https://python.langchain.com/docs/concepts/why_langchain/. Acesso em: 04 dez. 2024.

LIMA, Gercina Ângela de; MACULAN, Benildes Coura Moreira dos Santos. Universo do conhecimento: classificação e categorização sob o prisma da organização do conhecimento. **RDBCI**: Revista Digital de Biblioteconomia e Ciência da Informação, Campinas, v. 22, p. 1-25, 2024. DOI https://doi.org/10.20396/rdbci.v22i00.8675419.

MAZZOCCHI, Fulvio. Knowledge organization system (KOS). **Knowledge Organization**. v. 45, n. 1, p. 54-78, 2018. Disponível em: https://www.isko.org/cyclo/kos. Acesso em: 07 out. 2026.

MENDES, Luciana Corts. O Movimento Bibliográfico: organização do conhecimento no contexto da modernidade. **InCID: Revista de Ciência da Informação e Documentação**, v. 7, n. [esp.], p. 134-151, ago. 2016. DOI https://doi.org/10.11606/issn.2178-2075.v7iespp134-151.

META. **Tools**. [*S. l.*]: Meta, 2024. Disponível em: https://ai.meta.com/tools/faiss/. Acesso em: 04 dez. 2024.

OHLY, H. Peter. The field of knowledge organization. *In:* HJØRLAND, Birger; GNOLI, Claudio. **Encyclopedia of Knowledge Organization**. Ohio: ISKO, 2018. Disponível em: https://www.isko.org/cyclo/dahlberg. Acesso em: 13 set. 2025.

OPENAI. **GPT-4o mini**: inteligência com bom custo-benefício. [*S. l.*]: OpenAI, 2024. Disponível em: https://openai.com/pt-BR/index/gpt-4o-mini-advancing-cost-efficient-intelligence/. Acesso em: 14 set. 2026.

ORGANIZAÇÃO MUNDIAL DE PROPRIEDADE INTELECTUAL; INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL. **Curso geral de Propriedade Intelectual a distância**. [*S. l.*]: [*s. n.*], 2018a. 14 p. (Módulo 2: Introdução à Propriedade Intelectual, v. 3).

PEREIRA, Vina Stuart. **Gestão do Conhecimento Científico no Programa de Pós-Graduação em Propriedade Intelectual e Inovação do INPI**: proposta de metodologia de construção de tesauro para organização do conhecimento. 2022. Tese (Doutorado em Propriedade Intelectual e Inovação) – Instituto Nacional de Propriedade Industrial, Rio de Janeiro, 2022.

PIEDADE, Maria Antonietta. **Introdução à teoria da classificação**. Rio de Janeiro: Interciência, 1977. 185 p.

PINHEIRO, Lena Vania Ribeiro; LOUREIRO, José Mauro Matheus. Traçados e limites da ciência da informação. **Ciência da informação**, v. 24, n. 1, 1995.

QU, Renyi; TU, Ruixuan; BAO, Forrest. Is Semantic Chunking Worth the Computational Cost? **arXiv preprint**, p. 1-21, 2024. Disponível em: https://arxiv.org/pdf/2410.13070. Acesso em: 29 nov. 2024.

RAHMAN, Was. **Inteligência artificial e aprendizado de máquina**. Tradução Lana Lim; Anna Lim. São Paulo: Editora Senac, 2022. 173 p.

RAJ, Sumit. **Construindo Chatbots com Python**: usando Natural Language Processing e Machine Learning. São Paulo: Editora Novatec, 2019. 183 p.

RIVERA BERRÍO, Juan Guillermo. **Inteligencias artificiales generativas**. Córdoba: Red Educativa Digital Descartes, 2024. 469 p. Disponível em: https://prometeo.matem.unam.mx/recursos/VariosNiveles/iCartesiLibri/recursos/Inteligencias_Artificiales_Generativas_2024/. Acesso em: 29 nov. 2024.

SENA, Priscila Machado Borges; MARTINS, Maralyza Pinheiro. **Profissionais da informação no contexto de inovações tecnológicas**. Florianópolis: Rocha Gráfica e Editora, 2022. 210 p. (Selo Nyota).

SILVA, Rafael João. **Batendo um papo com a informação**: o uso dos chatbots para a recuperação da informação e a contribuição da Ciência da Informação nesse processo. 2020. Dissertação (Mestrado em Ciência da Informação) – Universidade de São Paulo, São Paulo, 2020. DOI 10.11606/D.27.2020.tde-10032021-013140.

DOI http://dx.doi.org/10.1590/1981-5344/61756

SMIRAGLIA, Richard P. Epistemology of domain analysis. *In*: SMIRAGLIA, Richard P.; LEE, Hur-Li (ed.). **Cultural frames of knowledge**. Würzburg: Ergon Verlag, 2012. p. 111-124.

STREAMLIT.IO. **A faster way to build and share data apps**. Disponível em: https://streamlit.io/. [*S. l.*]: Snowflake, 2024. Acesso em: 01 dez. 2024.

UNIVERSIDADE FEDERAL DO AMAZONAS [UFAM]. **Perguntas Frequentes**. [Manaus]: UFAM, 2023. Disponível em: https://protec.ufam.edu.br/perguntas-frequentes.html. Acesso em: 03 dez. 2024.

UNIVERSIDADE FEDERAL DA INTEGRAÇÃO LATINO-AMERICANA [UNILA]. **Perguntas Frequentes**. Foz do Iguaçu: UNILA, 2021. Disponível em: https://portal.unila.edu.br/prppg/inovacao/inovacao-tecnologica/perguntas-frequentes. Acesso em: 03 dez. 2024.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL [UFRGS]. **Tesauro do Programa de Pós-Graduação em Propriedade Intelectual e Inovação do INPI (extrato)**. [Porto Alegre]: UFRGS, 2022. Disponível em: https://www.ufrgs.br/tesauros/index.php/thesa/terms/380. Acesso em: 28 ago. 2024.

WORLD INTELLECTUAL PROPERTY ORGANIZATION [WIPO]. **Member States**. [*S. l.*]: WIPO, [2024]. Disponível em: https://www.wipo.int/members/en/. Acesso em: 29 nov. 2024.

CONFLITOS DE INTERESSE

As autoras declaram não haver conflitos de interesse relacionados a este trabalho.

CONTRIBUIÇÕES DAS AUTORIAS

Informa-se nesta seção as funções de cada autoria, de acordo com a taxonomia CRediT, conforme orientado na página da revista PCI:

Função	Definição
Conceituação	Paula Cotrim de Abrantes.
Curadoria de dados	Paula Cotrim de Abrantes.
Análise Formal	Paula Cotrim de Abrantes.
Obtenção de financiamento	—
Investigação	Paula Cotrim de Abrantes.
Metodologia	Paula Cotrim de Abrantes.
Administração do projeto	—
Recursos	—
Software	Paula Cotrim de Abrantes.
Supervisão	—
Validação	Paula Cotrim de Abrantes.
Visualização [de dados (infográfico, fluxograma, tabela, gráfico)]	Paula Cotrim de Abrantes.
Escrita – primeira redação	Paula Cotrim de Abrantes.
Escrita – revisão e edição	Paula Cotrim de Abrantes; Vania Lisboa da Silveira Guedes.