

INTELIGÊNCIA ARTIFICIAL: ANÁLISE BIBLIOMÉTRICA DE PESQUISAS ACADÊMICAS, CURRÍCULOS LATTES E GRUPOS DE PESQUISA DO CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO

ARTIFICIAL INTELLIGENCE: BIBLIOMETRIC ANALYSIS OF ACADEMIC RESEARCH, LATTES CURRICULUM AND RESEARCH GROUPS FROM NATIONAL COUNCIL FOR SCIENTIFIC AND TECHNOLOGICAL DEVELOPMENT DATABASE

Priscila Ramos-Carvalho^a
Fabio Castro Gouveia^b
Marcos Gonçalves Ramos^c

RESUMO

Introdução: A inteligência artificial possibilita uma diversidade de aplicações em diferentes áreas do conhecimento. O estudo apresenta um retrato do desempenho do campo no contexto brasileiro. **Objetivo:** Mapear as pesquisas acadêmicas no âmbito das dissertações e teses realizadas em programas de pós-graduação, a produção científica quanto aos aspectos de distribuição geográfica, instituições, áreas de conhecimento, canal de comunicação, rede de coautoria e temas relacionados à questão da inteligência artificial. **Metodologia:** A pesquisa de caráter exploratório e descritivo foi baseada em procedimentos metodológicos quantitativos por meio da bibliometria e da análise de conteúdo. As fontes de informação utilizadas foram os Grupos de Pesquisa e a Plataforma Lattes do Conselho Nacional de Desenvolvimento Científico e Tecnológico, bem como a Biblioteca Digital Brasileira de Teses e Dissertações, do Instituto Brasileiro de Informação em Ciência e Tecnologia. As ferramentas para extração, análise e visualização de dados escolhidas foram o scripLattes, Gephi e IRaMuTeQ. As coletas de dados foram realizadas em outubro de

^a Doutoranda no Programa de Pós-Graduação em Ciência da Informação no Instituto Brasileiro de Ciência e Tecnologia em parceria com a Universidade Federal do Rio de Janeiro (PPGCI/IBICT-UFRJ), Rio de Janeiro, Brasil. E-mail: carvalho.priscila@gmail.com

^b Doutor em Química Biológica pela Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brasil. Tecnologista em Saúde Pública da Fundação Oswaldo Cruz. E-mail: fgouveia@gmail.com

^c Doutorando no Programa de Pós-Graduação em Ciência da Informação no Instituto Brasileiro de Ciência e Tecnologia em parceria com a Universidade Federal do Rio de Janeiro (PPGCI/IBICT-UFRJ), Rio de Janeiro, Brasil. E-mail: marcosgon54@hotmail.com

2022, tendo no Lattes a definição do período de 30 anos. **Resultados:** Foram identificados 759 Grupos de Pesquisa, 20.400 currículos Lattes e 3.073 registros de teses e dissertações. As áreas de conhecimento mais frequentes identificadas nas três bases que abrangem a questão foram. Ciências Exatas e da Terra, Engenharias e Ciências Sociais Aplicadas. As regiões mais relevantes em duas das bases foram Sudeste e Sul. As instituições que se destacaram foram a Universidade de São Paulo e Universidade Federal de Santa Catarina, as quais reúnem a maior produção científica sobre inteligência artificial no Brasil. As publicações científicas em sua maioria se concentram em artigos completos publicados em periódicos e os trabalhos completos em conferências. Alguns temas identificados tinham relação com setores de petróleo e gás, agricultura, educação e saúde, bem como discussões éticas e regulamentação. **Conclusões:** Os pesquisadores e grupos de pesquisa sobre inteligência artificial no Brasil têm aparentemente experimentado um período de crescimento. Embora as aplicações atravessem diferentes áreas de conhecimento, a rede de coautoria de produção científica demonstrou ser pouco multidisciplinar.

Descritores: Inteligência Artificial. Plataforma Lattes. Grupos de Pesquisa. Biblioteca Digital de Teses e Dissertações.

1 INTRODUÇÃO

As metrias da informação se inserem no âmbito da aplicação de técnicas estatísticas e matemáticas para descrição de aspectos da informação. O estudo métrico intitulado “Bibliometria” de Otlet em 1934, no “*Traité de Documentation*”, termo originalmente conhecido como “bibliografia estatística” por meio de Hulme em 1923 (ARAUJO, 2006, p. 12), ganha espaço no mundo acadêmico ao longo do século XX (GONTIJO, ARAÚJO, OLIVEIRA, 2019) por possibilitar meios de avaliação da produção científica de um país.

Os indicadores e análises da atividade de pesquisadores são importantes, no que se refere às investigações e a produção científica, para a gestão de recursos do governo brasileiro, na esfera federal ou estadual, e definição de diretrizes, bem como políticas de incentivo ao desenvolvimento da ciência, inovação e tecnologia (OLIVEIRA, ARAÚJO, 2020).

As universidades públicas são o principal local de produção de pesquisas científicas, isto é, centros de geração de conhecimento e tecnologias essenciais para o desenvolvimento econômico e competitividade do Brasil (CARVALHO, RAMOS, PIMENTA, 2020). Portanto, as teses e dissertações defendidas nos programas de pós-graduação em instituições brasileiras são resultados da produção científica nacional em diversas áreas de conhecimento.

O governo brasileiro aderiu à iniciativa internacional do Governo Aberto (em inglês *Open Government Partnership*), lançada em 2011 por oito países fundadores (África do Sul, Brasil, Estados Unidos, Filipinas, Indonésia, México, Noruega e Reino Unido) com objetivo de incentivar a implementação de práticas voltadas à transparência do governo, ao acesso à informação pública e à participação social.

O Brasil publicou o 5º Plano de Ação Nacional do Governo Aberto em 2021, composto por doze compromissos e dentre estes o Compromisso 8, tem como objetivo construir uma proposta de modelo de avaliação que fomente a ciência aberta, sob responsabilidade do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), para a transparência da ciência através de novos mecanismos de avaliação.

O Compromisso 8 é formado por 11 marcos que visam a concretização e implementação, a saber: 1) levantamento dos critérios nacionais e internacionais de avaliação de pesquisadores e instituições de pesquisa, com enfoque nas práticas de ciência aberta. 2) proposição de critérios de qualificação de repositórios de dados e repositórios de publicações. 3) proposição de diretrizes para concessão de fomento e desenvolvimento de projetos de pesquisa. 4) proposição de indicadores de métricas alternativas para medição do impacto de pesquisas científicas. 5) proposição de indicadores para a ciência cidadã. 6) proposta de qualificação das revistas científicas e critérios alternativos nas dimensões de ciência aberta para enriquecer a estratificação do Qualis. 7) levantamento de percepções e promoção da sensibilização sobre os produtos do compromisso. 8) proposição de indicadores de avaliação que contemplem a ciência aberta. 9) implantação de observatório para o progresso de práticas de ciência aberta no Brasil. 10) estudo comparativo do impacto esperado versus impacto observado na produção dos programas de pós-graduação. 11) proposição de critérios de qualificação de produtos técnicos, tecnológicos e artísticos que favoreçam a ciência aberta.

Dentre as contribuições para o acesso aberto aos dados de pesquisa científicas produzidas no Brasil, vale destacar duas bases de dados: o Diretório de Grupo de Pesquisa e a Plataforma Lattes, ambas iniciativas pertencentes ao

Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), além da Biblioteca Digital de Teses e Dissertações (BDTD) sob gestão do IBICT.

A BDTD faz parte do programa brasileiro de acesso aberto à informação científica e disponibiliza o acesso livre aos conteúdos de teses e dissertações via internet, contribuindo com a difusão de informações, a visibilidade da produção científica e o incentivo à ciência aberta e cidadã.

O Diretório de Grupos de Pesquisa é uma base de dados capaz de mostrar um retrato das áreas de conhecimento onde se desenvolvem os estudos científicos no Brasil. A falta de informação sobre os grupos e a produção científica pode comprometer os incentivos à pesquisa, pois o volume de publicações é um critério de avaliação para distribuição de recursos no país (MOREIRA, MÜLLER, VILAN FILHO, 2020).

A Plataforma Lattes tornou-se uma ferramenta de apoio à gestão de ciência ao possibilitar o processo de cadastramento, armazenamento e consulta de dados curriculares de acadêmicos em um mesmo ambiente amigável (PAULA, 2004; BERMEJO *et al.*, 2005). O conhecimento advindo do Lattes está alinhado com métricas quantitativas da informação como a bibliometria e cienciometria no sentido de permitir a exploração, identificação ou validação de padrões das atividades científicas como produção e redes de coautores em um determinado grupo de interesses (NICHOLSON, 2006).

No contexto da produção científica nacional, esta investigação teve como objetivo mapear a produção acadêmica sobre inteligência artificial considerando aspectos, tais como: áreas de conhecimento, instituições, localização geográfica, rede de coautoria acadêmica e temas relacionados à questão no Brasil. Para tanto, a bibliometria e análise de conteúdo foram escolhidas como referências de procedimentos quantitativos metodológicos e os dados para análise foram coletados no Diretório de Pesquisa e na Plataforma Lattes como também na Biblioteca Digital de Teses e Dissertações.

2 ACERCA DA INTELIGÊNCIA ARTIFICIAL

O fascínio pelas máquinas e relatos sobre autômatos tem sido comum ao longo da história como, por exemplo, o caso do cientista e inventor Leonardo Da

Vinci que desenhou um modelo de robô na forma de um cavaleiro medieval em 1495. Séculos depois, o matemático e filósofo Thomas Hobbes no livro “Leviatã” de 1651 pode ter previsto a inteligência artificial:

Vendo que a vida não é mais que um movimento dos membros, cujo início ocorre em alguma parte principal interna, por que não poderíamos dizer que todos os autômatos (máquinas que se movem a si mesmas por meio de molas, tal como um relógio) possuem uma vida artificial (HOBBS, tradução de MONTEIRO, SILVA, 2019, p. 9).

O primeiro artigo científico sobre o assunto foi publicado em 1943 no *Bulletin of Mathematical Biophysics*, com o título *A logical calculus of the ideas immanent in nervous activity*, do fisiologista McCulloch e do lógico Pitts, onde propuseram modelos simples de neurônios artificiais e mostraram que qualquer operação computacional poderia ser calculada através de uma rede de neurônios interconectados (RUSSEL; NORVIG, 2004; NILSSON, 2009).

Turing em 1948 propôs investigar a possibilidade de uma máquina mostrar um comportamento inteligente no relatório *Intelligent Machinery*. O cientista publicou em 1950 o artigo *Computing Machinery and Intelligence* em que sugeriu a pergunta: “As máquinas podem pensar?” (TURING, 1950, p. 433). Neste documento, o autor discorre sobre uma forma de avaliar a questão através do jogo da imitação, chamado de Teste de Turing, no qual poderia distinguir entre uma máquina ou humano através de um método de perguntas e respostas.

Outro cientista importante foi McCarthy, pioneiro ao incentivar a área de inteligência artificial através da organização de um curso em 1956 nos Estados Unidos da América que reuniu pesquisadores como Marvin Minsky, Claude Shannon, Nathaniel Rochester, Allen Newell, Herbert Simon e Alex Bernstein (RUSSEL, NORVIG, 2004; NILSSON, 2009; MITCHELL, 2019).

Os pesquisadores motivados começaram a explorar diversos caminhos a fim de mecanizar a inteligência e se dividiram nas seguintes vertentes metodológicas: a) sistemas simbólicos ou especialistas: abordagem teve sua origem na lógica matemática, na década de 1930, tendo como objetivo ensinar a máquina a pensar codificando uma série de regras lógicas (*if, then*), o que funcionou em sistemas para tomada de decisão e jogos. b) conexionista ou redes neurais: abordagem com base no modelo cerebral de McCulloch e Pitts, na

década de 1940, com intuito de instruir o computador regras dominadas por humanos, através da imitação da arquitetura subjacente do cérebro, constituindo camadas de neurônios artificiais para transmitir e receber informações em uma estrutura semelhante à rede de neurônios biológicos. c) acionistas: abordagem surgiu na cibernética, na década de 1980, em busca de realizar a inteligência por meio de características comportamentais advindas da interação entre sujeito e o ambiente externo (LEE, 2019; SUN, YANG, 2019).

Com base na indagação “Que significado psicológico e filosófico devemos anexar aos esforços recentes de simulações em computador de capacidades cognitivas de humanos?”, Searle mostrou duas abordagens para Inteligência Artificial (IA) em 1980:

De acordo com a IA Fraca, o principal valor do computador no estudo da mente é que isso nos dá uma ferramenta muito poderosa. Por exemplo, nos permite formular e testar hipóteses de uma forma mais rigorosa e forma precisa. Mas de acordo com a IA Forte, o computador não é apenas uma ferramenta no estudo da mente; em vez disso, o computador devidamente programado é realmente uma mente, no sentido que os computadores com os programas certos podem ser literalmente ditos como capazes de compreender e ter outros estados cognitivos. Na IA Forte, porque o computador programado tem estados cognitivos, os programas não são meras ferramentas que nos permitem testar explicações psicológicas; em vez disso, os programas são eles próprios as explicações (SEARLE, 1980, p. 417)¹.

Nessa linha, uma série de definições e aplicações foram propostas em diferentes áreas do conhecimento. Nilsson explica que a inteligência artificial é “aquela atividade dedicada a tornar as máquinas inteligentes e inteligência é a qualidade que permite a uma entidade funcionar de maneira adequada e com visão de futuro em seu ambiente” (NILSSON, 2009, p. 13)².

¹ Tradução nossa: *According to weak AI, the principal value of the computer in the study of the mind is that it gives us a very powerful tool. For example, it enables us to formulate and test hypotheses in a more rigorous and precise fashion. But according to strong AI, the computer is not merely a tool in the study of the mind; rather, the appropriately programmed computer really is a mind, in the sense that computers given the right programs can be literally said to understand and have other cognitive states. In strong AI, because the programmed computer has cognitive states, the programs are not mere tools that enable us to test psychological explanations; rather, the programs are themselves the explanations.*

² Tradução nossa: *For me, artificial intelligence is that activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment. According to that definition, lots of things - humans, animals,*

Para Lee, a inteligência artificial seria um passo para o homem se compreender: “a elucidação do processo de aprendizagem humana, a quantificação do processo de pensamento humano, a explicação do comportamento humano e a compreensão do que torna a inteligência possível” (LEE, 2019, p. 17). Russell e Norvig afirmam que a inteligência artificial “se preocupa não apenas com a compreensão, mas também com a construção de entidades inteligentes - máquinas que podem calcular como agir eficazmente e de forma segura em uma ampla variedade de novas situações” (RUSSELL, NORVIG, 2021, p. 19).³

Domingos esclarece que a inteligência artificial tem como objetivo ensinar aos computadores como executar o que os humanos fazem melhor, buscando padrões e aprendendo com a experiência. Enquanto o aprendizado de máquina, subcampo da inteligência artificial, permite reconhecer padrões em grandes conjuntos de dados e aplicar estas descobertas em novos dados (DOMINGOS, 2017). O aprendizado profundo, subcampo do aprendizado de máquina, é baseado em um conjunto de algoritmos que tentam modelar altos níveis de abstrações em dados utilizáveis, imitando o cérebro humano (CARVALHO, GOUVEIA, RAMOS, 2022).

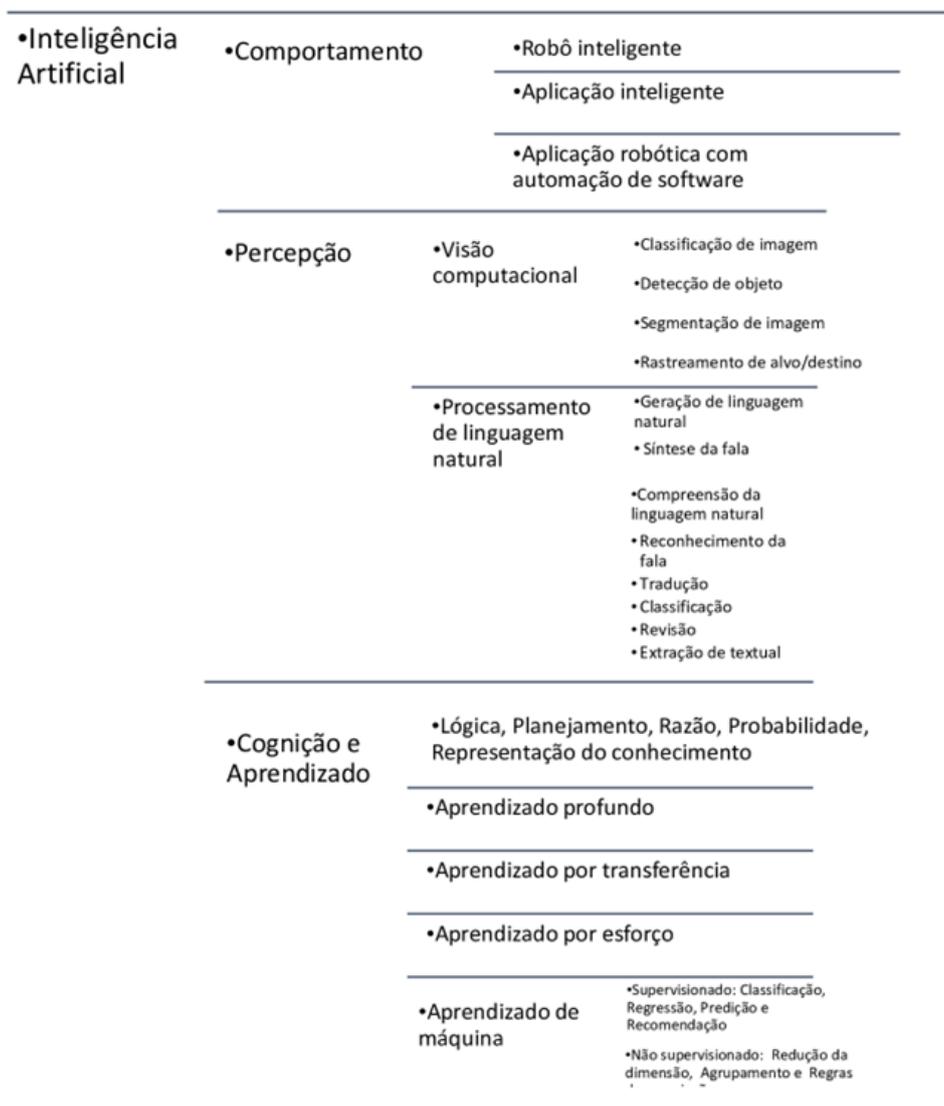
A partir das discussões sobre definições, teorias e metodologias, Wang *et al.* (2021) explicam três possíveis categorias de inteligência artificial: 1) IA geral: configura um estado considerado avançado por se equiparar à inteligência humana, no que tange à capacidade de realizar tarefas intelectuais genéricas. 2) IA super: exprime uma idealização da evolução da inteligência artificial ao ponto de superar a inteligência humana, em relação a lidar com tarefas complexas para o ser-humano. 3) IA estreita ou especializada: abarca sistemas artificiais que executam uma única tarefa específica e suas aplicações podem ser divididas em comportamento, percepção, e cognição e aprendizado. A

and some machines - are intelligent. Machines, such as “smart cameras”, and many animals are at the primitive end of the extended continuum along which entities with various degrees of intelligence are arrayed.

³ Tradução nossa: *Artificial intelligence is concerned with not just understanding but also building intelligent entities—machines that can compute how to act effectively and safely in a wide variety of novel situations.*

seguir, a Figura 1 exibe uma versão traduzida e adaptada da classificação dos autores.

Figura 1 – Classificação da inteligência artificial estreita



Fonte: Criação baseada em Wang *et al.* (2021) e Sun & Yang (2019).

A IA geral configura um modelo de sistemas computacionais inteligentes vistos em filmes de ficção científica. Até o momento, as máquinas não alcançaram um sistema integrado capaz de realizar qualquer tarefa intelectual humana, como fazer conexões e expressar ideias criativas de modo a demonstrar a capacidade de pensamento abstrato e a habilidade de planejar e aprender com o passado. A IA super se encaixa nas reflexões sobre a singularidade e autoaperfeiçoamento das máquinas. O debate gira em torno das

consequências positivas e negativas na civilização humana, entretanto os impactos a longo prazo são imprevisíveis (FRANA, KLEIN, 2021).

A IA estreita é a mais comum no nosso dia a dia. Esta categoria é considerada IA fraca devido à incapacidade de executar tarefas diferentes das especificadas, ou seja, o algoritmo é limitado a resolver um determinado problema (SUN, YANG, 2019; FRANA, KLEIN, 2021). No que diz respeito ao comportamento, os programas artificiais buscam integrar processos de detecção, *feedback* e comunicação, como no caso da automação por *software* ou robôs (WANG *et al.*, 2021). O aspirador de pó autônomo iRobot Roomba de Rodney Brooks é um exemplo de robótica baseada em comportamento e inteligência artificial (FRANA, KLEIN, 2021).

A percepção das máquinas corresponde a capacidade de interagir e processar informações do mundo exterior através de dispositivos inteligentes e conectados à internet como câmeras, microfones, sensores, *smartphones* e etiquetas de identificação por radiofrequência, que capturam dados estruturados (tamanho, temperatura e pressão) e não-estruturados (som, vídeo e imagens). Nessa direção, a visão computacional busca extrair informações e conhecimento dos dados brutos gerados, por exemplo, pela infraestrutura de cidades inteligentes. O processamento de linguagem natural visa a interpretação do significado da linguagem natural por máquinas, incluindo a compreensão da linguagem no caso de voz humana e compreensão semântica, assim como a geração de linguagem natural como a produção de texto com base em regras gramaticais (WANG *et al.*, 2021; FRANA, KLEIN, 2021).

A cognição e aprendizado abrangem sistemas artificiais para aprender, analisar, avaliar, julgar e tomar decisões com base nas informações como, por exemplo, o Watson da IBM. O processamento de linguagem natural possibilita ao Watson a habilidade de entender frases com nuances e atribuir significados a termos e conceitos. O aprendizado de máquina oferece a competência de aprender com a experiência e a interação com os dados para desenvolver soluções inteligentes a partir do histórico de experimentação (JARRAHI, 2018; WANG *et al.*, 2021).

Uma infinidade de aplicações do Watson IBM⁴ são disponibilizadas como caso da área financeira em que a GM Financial utiliza o Watson Assistant na automação de respostas às consultas dos clientes. Na parte de viagens, a Lufthansa aderiu ao IBM Watson Advertising e Watson-powered Natural Language Understanding para oferecer viagens através de conversas personalizadas com os clientes. Na saúde, o Memorial Sloan Kettering Cancer Center em parceria com IBM utilizou Watson Health como ferramenta de suporte à decisão clínica para auxiliar os médicos na escolha de terapias para pacientes com câncer (JARRAHI, 2018).

As categorias abordadas por Wang *et al.* (2021) se alinham com o enfoque de Lee (2019) sobre as fases do desenvolvimento da inteligência artificial ser dividido em quatro ondas: 1) onda da internet: envolve a expansão das empresas chamadas de ponto.com. 2) onda de negócios: engloba adoção do uso de algoritmos para resolver questões do dia a dia como diagnósticos de doenças, análise de ações na bolsa e avaliação de desempenho de professores. 3) onda da percepção: abrange a digitalização do mundo físico, onde os algoritmos aprendem a realizar o reconhecimento facial e a ver o que há ao seu redor. 4) onda autônoma: abarca desde os carros e drones autônomos, robôs inteligentes em fábricas, a mecanização da agricultura até automatização de uma loja de *fast food* (LEE, 2019).

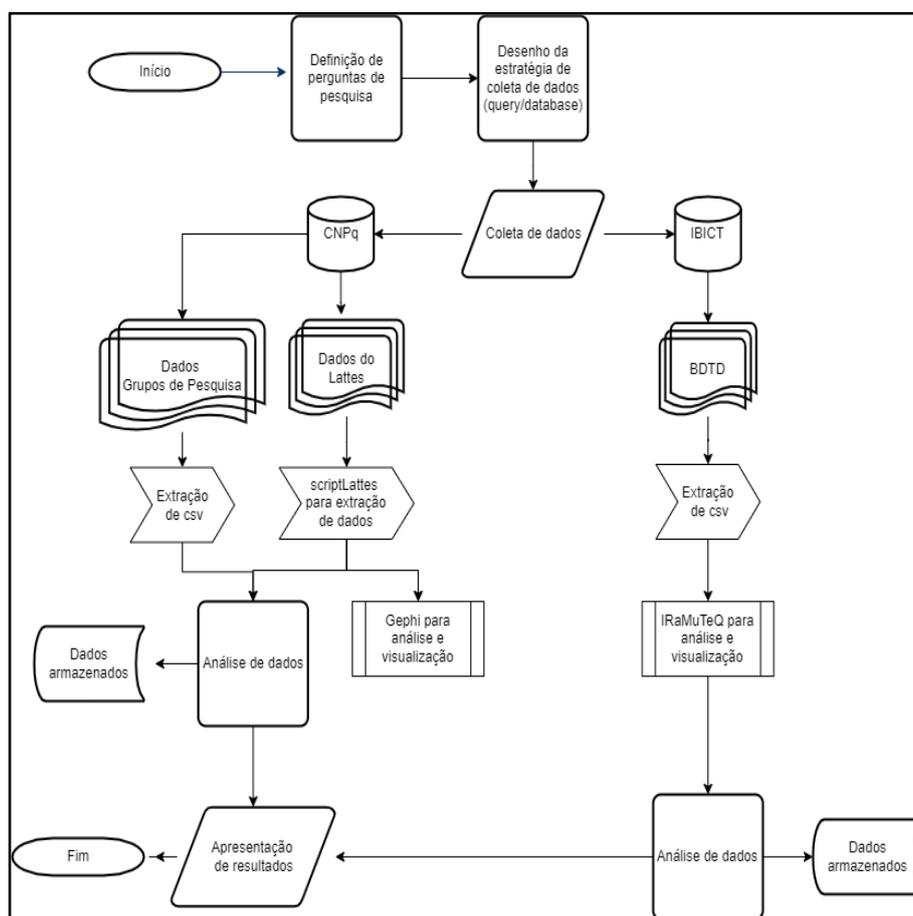
A área de inteligência artificial tem se delineado em meio a períodos de aquecimento em que há um aumento do investimento, desenvolvimento e euforia de ideias, como também passado por estágios de inverno em que ocorre o oposto, em que não há interesse em investir na questão, devido ao não resultado previsto ou idealizado por alguns (LEE, 2019). Aparentemente, estamos em uma fase de expansão da inteligência artificial, com lançamento do Chat GPT, estimulada pela concorrência tecnológica entre Estados Unidos da América e China.

⁴ Watson IBM. Disponível em: <https://www.ibm.com/watson>. Acesso em: 23 dez. 2022.

3 PROCEDIMENTOS E MÉTODOS

A pesquisa exploratória e descritiva teve o intuito de mapear a produção acadêmica brasileira sobre inteligência artificial por meio do detalhamento das informações coletadas nas bases de dados utilizadas como universo de estudo (GIL, 2002). A Figura 2 apresenta o processo da investigação visando responder às questões da pesquisa.

Figura 2 – Diagrama do processo da pesquisa



Fonte: Elaborado no programa Draw.io (2022).

As questões do estudo foram as seguintes: a) Quais as áreas de conhecimento? b) Quais as principais instituições? c) Qual a distribuição geográfica dos pesquisadores? d) Qual o volume de produção científica? e) Qual a rede de coautoria identificada por grande área de pesquisa? f) Quais os termos identificados na produção científica? g) Quais os temas de teses e dissertações relacionados à inteligência artificial?

A partir do ponto de vista da Ciência da Informação, a bibliometria foi escolhida como metodologia de pesquisa focada nos estudos quantitativos sobre a produção, a disseminação e uso da informação científica (MACIAS-CHAPULA, 1998). A bibliometria inicialmente era “voltada para a medida de livros” no que tange a quantidade de edições, exemplares, palavras contidas nos livros, espaço ocupado pelos livros nas bibliotecas e estatísticas relativas à indústria do livro (ARAUJO, 2006, p. 13).

Ao longo do tempo, este método passou a examinar as relações entre diferentes variáveis da informação como os recursos humanos, artigos, periódicos e capítulos de livro, como também a produção e o consumo destes documentos e sua distribuição, no que se refere, por exemplo, às citações e ao número de autores por artigo (BRAGA, 1974).

Na sequência, a análise de conteúdo foi agregada como metodologia por conter um conjunto de procedimentos sistemáticos e técnicas de descrição do conteúdo, com indicadores que possibilitam inferir conhecimentos relativos ao processo de codificação dos textos (BARDIN, 1977). Para tanto, o software de análise textual IRaMuTeQ foi selecionado para dar suporte a análise quantitativa, enumerando de forma automática as ocorrências e identificando as unidades lexicais no *corpus* textual, a fim de transformar resultados brutos em informações significativas.

As fontes de informações exploradas no estudo foram: a) Diretório de Grupos de Pesquisa com análise de planilha disponibilizada pela base do CNPq. b) Plataforma Lattes com as ferramentas scriptLattes e Gephi para extração, análise e visualização de dados. c) Biblioteca Digital Brasileira de Teses e Dissertações por meio da análise do programa IRaMuTeQ.

A primeira etapa da pesquisa foi realizar a coleta dos dados no Diretório de Grupos de Pesquisa no website do CNPq por meio da busca do termo em português inteligência artificial com a geração e download de planilha na extensão .csv (texto) com os resultados para análise dos dados.

A segunda etapa foi utilizar o scriptLattes por meio das fases: a) coleta de dados na Plataforma Lattes através da expressão de busca em português, inteligência artificial, como termo exato, pelo período de 30 anos (1992 até 2021),

na busca avançada por assunto. b) exploração dos dados consolidados pelo scriptLattes v.8.11r onde foi gerado um relatório com projetos de pesquisa, produções bibliográficas, técnicas e artísticas e um arquivo de grafo de coautorias. c) utilização do Gephi para análise de coautoria e termos identificados na produção científica.

A terceira etapa foi a análise com IRaMuTeQ, sendo necessário realizar os seguintes passos: a) coleta de dados na BDTD através da busca pelo termo inteligência artificial e download de planilha na extensão .csv (texto) com resultados. b) análise geral dos dados em planilha. c) tratamento dos dados dos títulos das teses e dissertações para geração de um *corpus* textual como, por exemplo, a limpeza de caracteres de acentos e sinalização de arquivo na extensão .txt (**** *biblioteca). d) verificação e análise de dados nas opções disponibilizadas pelo programa IRaMuTeQ para seleção da melhor opção. A seguir apresentamos as fontes de informações e os programas utilizados durante o processo de pesquisa.

3.1 DIRETÓRIO DE GRUPOS DE PESQUISA

O Conselho Nacional de Desenvolvimento Científico e Tecnológico desenvolveu o projeto de construção de uma base de dados que comportasse as informações sobre os grupos de pesquisa em atividade no Brasil em 1992. A base compreende um conjunto de dados organizados hierarquicamente de pessoas envolvidas em atividades de pesquisa de diversos domínios de conhecimento, por meio de linhas de pesquisa comuns aos membros do grupo que estão localizados geograficamente em universidades, instituições de ensino superior, institutos de pesquisa científica, institutos tecnológicos, laboratórios de pesquisa e desenvolvimento de empresas estatais ou privadas, assim como organizações não governamentais com atuação em pesquisa na área de ciência e tecnologia (PAULA, 2004).

O Diretório de Grupos de Pesquisa divide as informações em dois módulos: a) base corrente: base de dados relacional e operacional onde constam informações referentes aos grupos de pesquisa cadastrados e atualizados pelos pesquisadores responsáveis. b) base censitária: base de dados para consultas

por meio de indexação textual, bem como análise, avaliação e consolidação de informações dos censos bianuais realizados até 2016.

Esta base de dados é importante, pois a dinâmica de produção científica de grupos de pesquisa brasileiros ajuda na compreensão de como um determinado campo científico tem evoluído e contribui para a promoção de políticas públicas de fomento à pesquisa e a distribuição de verbas para programas de pós-graduação (MOREIRA, MÜLLER, VILAN FILHO, 2020).

3.2 PLATAFORMA LATTES

Em 16 de agosto de 1999, o CNPq lançou com cerca de 35 mil currículos cadastrados a Plataforma Lattes, uma base de dados constituída por uma estrutura relacional onde estão armazenados dados providos no cadastramento ou na atualização do currículo Lattes pelos usuários, sendo possível a realização de consultas textuais de informações não estruturadas originárias da base operacional (PAULA, 2004).

A Plataforma Lattes ganhou o prêmio excelência em Governo Eletrônico (e-Gov) de melhor projeto em 2004, da Associação Brasileira de Entidades Estaduais de Tecnologia da Informação e Comunicação (ABEP-TIC). Uma das vantagens da base Lattes é que as informações dos currículos, grupos de pesquisa e projetos de pesquisa são públicas, permitindo o conhecimento sobre os atores ligados ao sistema nacional de inovação do país (BERMEJO *et al.*, 2005; BALANCIERI *et al.*, 2005).

Outro diferencial da base Lattes é a característica curricular das informações submetidas pelos cientistas brasileiros, que contempla tanto os dados pessoais e profissionais, como também a produção científica durante a carreira acadêmica, tais como: artigos, patentes, participação em congressos, orientações de alunos, entre outros (MAGALHÃES *et al.*, 2014).

Cabe sinalizar que embora as informações da Plataforma Lattes sejam de livre acesso, a modalidade de busca, visualização e extração de dados são de certa forma limitadas como, por exemplo, a pesquisa por palavras-chave não possibilita a identificação de onde se encontra no currículo do pesquisador (MAGALHÃES *et al.*, 2014). Nesse sentido, o sistema pode acabar dificultando

o acesso ao conhecimento contido em sua base.

3.3 BIBLIOTECA DIGITAL BRASILEIRA DE TESES E DISSERTAÇÕES

O projeto de criação da biblioteca nacional de teses e dissertações eletrônicas foi aprovado em 2022, tendo como objetivos: a) construção de um sistema nacional de biblioteca digital para integrar iniciativas locais, nacionais e internacionais. b) definição de tecnologias e de um padrão nacional de metadados para teses e dissertações eletrônicas. c) desenvolvimento de ferramentas para treinamento e implementação de padrões das bibliotecas digitais locais em universidades e em outros centros de pesquisa (SOUTHWICK, 2006).

A condução do projeto foi realizada por um comitê técnico-consultivo constituído por representantes de instituições, a saber: Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Secretaria de Educação Superior (SESu), Universidade de São Paulo (USP), Pontifícia Universidade Católica do Rio de Janeiro (PUC-RIO) e Universidade Federal de Santa Catarina (UFSC). Além disso, o projeto contou com o apoio da Financiadora de Estudos e Pesquisas (FINEP).

Nas últimas décadas, a Biblioteca Digital Brasileira de Teses e Dissertações⁵ passou por muitos marcos e processos desde a sua concepção até a difusão do padrão de metadados: a) em 2003, aprovação do projeto de estruturação do sistema da biblioteca digital de teses e dissertações. b) em 2006, lançamento do portal da biblioteca digital (*website*). c) em 2012, revisão do padrão brasileiro de metadados para descrição de teses e dissertações por um grupo de trabalho coordenado pelo IBICT. d) em 2013, atualização do Sistema de Publicação Eletrônica de Teses e Dissertações (TEDE) e do portal da biblioteca digital. e) em 2014, lançamento do novo padrão de metadados, do TEDE e do portal. f) entre 2016 e 2017, atualização e apresentação do novo

⁵ BDTD. Disponível em: <https://bddd.ibict.br/vufind/Content/history>. Acesso em: 06 nov. 2022.

portal e coletor de metadados.

Desde 2022, a BDTD tem sido mantida pelo IBICT, formada por 141 instituições participantes e cerca de 765.802 registros. Convém acrescentar que uma vantagem da biblioteca digital é a disponibilização de dados abertos, sendo possível exportar o resultado de uma busca através de arquivo de extensão de texto .csv (*comma-separated values*) ou json (*javascript object notation*). Por sua vez, uma desvantagem da base é que nem todas as dissertações e teses produzidas em instituições acadêmicas brasileiras constam na biblioteca digital.

3.4 IRAMuTEQ

Pierre Ratinaud desenvolveu em 2009 um software no laboratório de pesquisa em Ciências Sociais da Universidade de Toulouse com base no Alceste (*Analyse Lexicale par Contexte d'un Ensemble de Segments de Texte*), um programa de análise textual criado por Max Reinert na década de 1990, no laboratório de Psicologia Social da *École des Hautes Études en Sciences* em Paris (CAMARGO, JUSTO, 2013).

O IRaMuTeQ (*Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires*) é uma ferramenta de código aberto e gratuita, fundamentada nos módulos estatísticos das linguagens de programação R e Python, a fim de quantificar as estruturas significativas em um determinado texto de modo a identificar a informação essencial contida neste *corpus* textual. O programa disponibiliza análises textuais, tais como: análise de similitude, análise fatorial por correspondência e análise de classificação hierárquica descendente (CARVALHO, 2020).

3.5 SCRIPTLATTES

O scriptLattes⁶ é um software livre desenvolvido na linguagem Python em 2005, para o sistema operacional Linux, pioneiro na pesquisa de conjuntos de dados acadêmicos oriundos de currículos da Plataforma Lattes. A ferramenta

⁶ ScriptLattes. Disponível em: <https://scriptlattes.sourceforge.net/>. Acesso em: 06 nov. 2022.

possui os seguintes módulos: (a) coleta de dados: seleciona e baixa os currículos Lattes no formato HTML. (b) pré-processamento de dados: busca as informações relacionadas ao currículo desde o endereço profissional, lista de produções acadêmicas até a lista de supervisões em andamento e concluídas. (c) tratamento de redundância: detecção e eliminação das produções duplicadas após o pré-processamento dos currículos. (d) rede de colaboração: criação de uma representação gráfica da rede de colaboração entre os membros do grupo com base em suas produções científicas mais relevantes. (e) mapas de pesquisa: geração de mapa de geolocalização dos pesquisadores listados a partir do endereço informado na plataforma. (f) relatórios: elaboração de relatórios referente à produção bibliográfica, técnica e artística, mais orientações acadêmicas (MENA-CHALCO, JUNIOR, 2009; FERRAZ, QUONIAM, 2013).

3.6 GEPHI

Em 2008, um grupo formado por Mathieu Bastian, Sébastien Heymann, Mathieu Jacomy e outros cientistas da computação da *Université Technologique de Compiègne* desenvolveram o Gephi⁷ com objetivo de fornecer métodos de análise de rede sem a necessidade do aprendizado da teoria dos grafos para cientistas sociais (JACOMY *et al.*, 2014). A primeira versão 0.6 do Gephi foi lançada em 2009 na modalidade de código aberto com as opções de importar, visualizar, filtrar, manipular e exportar uma infinidade de tipos de redes (BASTIAN, HEYMANN, JACOMY, 2009).

Nos últimos anos, o programa passou por diversas fases, a saber: (a) em 2010, a versão 0.7 foi o resultado da revisão da interface gráfica com a inserção do framework *opengl* (*open graphics library*) para aceleração gráfica e renderização de gráficos 3D. (b) em 2011, a versão 0.8 trouxe gráficos dinâmicos e *plugins*. (c) em 2015, versão 0.9 introduziu um mecanismo para lidar com gráficos maiores através de otimização da memória e resolveu os problemas de compatibilidade com sistema operacional Mac OS. (d) em 2016, a versão 0.9.1 ampliou o acesso do programa para outras línguas além do inglês, tais como:

⁷ Gephi. Disponível em: <https://gephi.org/>. Acesso: 05 nov. 2022.

francês, espanhol, japonês, russo, polonês, português, chinês, tcheco e alemão. (e) em 2017, a versão 0.9.2 disponibilizou uma API (*Application Programming Interface*) e adicionou o suporte para Microsoft SQL Server. (f) em 2021, a versão 0.9.3 apresentou correções de bugs e resolveu a necessidade de instalação de Java. (g) Nesta pesquisa utilizamos a versão 0.9.7 com atualizações de desempenho a partir da versão 0.9.3.

4 RESULTADOS E DISCUSSÃO

O resultado da recuperação da informação nas três bases de dados pelo termo de busca inteligência artificial foi: a) Grupos de Pesquisa: 759. b) currículos Lattes: 20.400. c) BDTD: 3.073 registros de teses (894) e dissertações (2.179).

A primeira análise do estudo foi sobre as áreas de conhecimento que têm realizado pesquisas sobre inteligência artificial. A Tabela 1 mostra as cinco áreas específicas com maior volume de trabalhos identificados por base de dados.

Tabela 1 – Áreas de conhecimento específicas (CNPq, Lattes e BDTD)

Grupos de Pesquisa	Qtd. (%)	Currículo Lattes	Qtd. (%)	BDTD	Qtd. (%)
Ciência da Computação	281 (37%)	Ciência da Computação	5.684 (28%)	Ciência da Computação	251 (35%)
Engenharia Elétrica	113 (15%)	Direito	2.598 (13%)	Engenharia Elétrica	209 (30%)
Direito	63 (8%)	Engenharia Elétrica	1.473 (7%)	Direito	39 (6%)
Administração	24 (3%)	Administração	914 (4%)	Engenharia Produção	24 (3%)
Medicina	22 (3%)	Educação	622 (3%)	Administração	17 (2%)

Fonte: Dados de pesquisa (2022).

Quanto às grandes áreas de conhecimento identificadas nas bases exploradas, notamos três áreas principais em representatividade, mas em diferentes posições, a saber: a) Grupos de Pesquisa: 42% Ciências Exatas e da Terra, 25% Engenharias e 15% Ciências Sociais Aplicadas. b) BDTD: 44% Engenharias, 41% Ciências Exatas e da Terra, e 9% Ciências Sociais Aplicadas.

c) Lattes: 32% Ciências Exatas e da Terra, 22% Ciências Sociais Aplicadas e 14% Engenharia.

Um traço interessante percebido no resultado sobre as áreas de conhecimento é que o Direito tem avançado nos estudos sobre inteligência artificial, demonstrando a preocupação com a regulamentação da questão no Brasil, além das áreas de Administração, Medicina e Educação. As diferenças observadas nos percentuais podem ter relação com o tempo ou período em que cada uma dessas bases pode revelar. Grupos de Pesquisa sofrem atualizações anuais e refletem a consolidação de pesquisadores em alto nível de envolvimento acadêmico. As teses e dissertações representam as pesquisas correntes em andamento vinculadas aos programas de pós-graduação. Por último, a base Lattes tem o acumulado de pesquisadores que atuaram no campo e os acadêmicos que mais recentemente se direcionaram ao tema, mesmo que por intermédio de um trabalho pontual.

Convém também ponderar que cerca de 10% da amostra dos currículo Lattes (2.051) não possuía dados sobre área de conhecimento, o que de certa forma pode gerar inconsistência em análises do scriptLattes, pois o preenchimento e atualização do currículo Lattes deve ser realizado por cada pesquisador ou profissional.

A segunda análise foi sobre a abrangência geográfica, ou seja, onde os pesquisadores e pesquisas acadêmicas sobre inteligência artificial têm sido desenvolvidas no país. A Tabela 2 revela os cinco estados mais relevantes quantitativamente nas bases do Diretório de Grupos de Pesquisa (DGP) e BDTD. Os dados do Lattes foram de difícil determinação tanto da unidade da federação quanto da instituição devido à ausência de padrão para inserção na base.

Tabela 2 – Estados brasileiros (DGP-CNPq e BDTD)

Grupos de Pesquisa	Qtd. (%)	BDTD	Qtd. (%)
São Paulo	131 (17%)	São Paulo	1.156 (38%)
Minas Gerais	90 (12%)	Rio Grande do Sul	319 (11%)
Rio Grande do Sul	67 (9%)	Santa Catarina	274 (9%)
Rio de Janeiro	50 (7%)	Minas Gerais	211 (7%)

Santa Catarina	50 (7%)	Pernambuco	198 (7%)
----------------	------------	------------	-------------

Fonte: Dados de pesquisa (2022).

O resultado das regiões brasileiras reflete o volume identificado de pesquisas nos estados. Em relação aos Grupos de Pesquisas, destacaram-se em quantitativamente o Sudeste com 38%, Nordeste com 25% e Sul com 22%. Na BDTD, as principais regiões foram Sudeste com 49%, Sul com 26% e Nordeste com 17%.

Nos resultados sobre a localização das pesquisas brasileiras sobre inteligência artificial, vale destacar em volume o estado de Pernambuco na região Nordeste, fora do eixo Sudeste e Sul, onde a maioria dos estudos está em andamento.

A terceira análise foi sobre as instituições onde os pesquisadores e as investigações ocorrem sobre inteligência artificial. A Tabela 3 exhibe as cinco principais instituições no quesito quantitativo para as bases DGP e BDTD.

Tabela 3 – Instituições (DGP-CNPq e BDTD)

Grupos de Pesquisa	Qtd. (%)	BDTD	Qtd. (%)
Instituto Federal de São Paulo (IFSP)	29 (4%)	Universidade de São Paulo (USP)	347 (11%)
Universidade de São Paulo (USP)	24 (3%)	Universidade Estadual de Campinas (UNICAMP)	318 (10%)
Universidade Federal de Santa Catarina (UFSC)	20 (3%)	Universidade Federal de Santa Catarina (UFSC)	273 (9%)
Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP)	19 (3%)	Universidade Federal do Rio Grande do Sul (UFRGS)	234 (8%)
Universidade Tecnológica Federal do Paraná (UTFPR)	18 (3%)	Universidade Federal de Pernambuco (UFPE)	190 (6%)

Fonte: Dados de pesquisa (2022).

O resultado das regiões onde se concentram as pesquisas reflete claramente a identificação das principais instituições acadêmicas. Cabe sinalizar a importância das universidades públicas, aonde grande parte das pesquisas brasileiras são desempenhadas no Brasil, sendo um total de 69 universidades

federais distribuídas: 11 Norte, 20 Nordeste, 8 Centro-Oeste, 19 Sudeste e 11 Sul. A USP e Unicamp são instituições estaduais localizadas em São Paulo onde se concentram a maioria das empresas e riqueza do país.

A quarta análise foi sobre a produção científica dos pesquisadores identificadas na amostra por meio da utilização do scriptLattes, exibida na Tabela 4 a seguir.

Tabela 4 – Produção científica geral (Lattes)

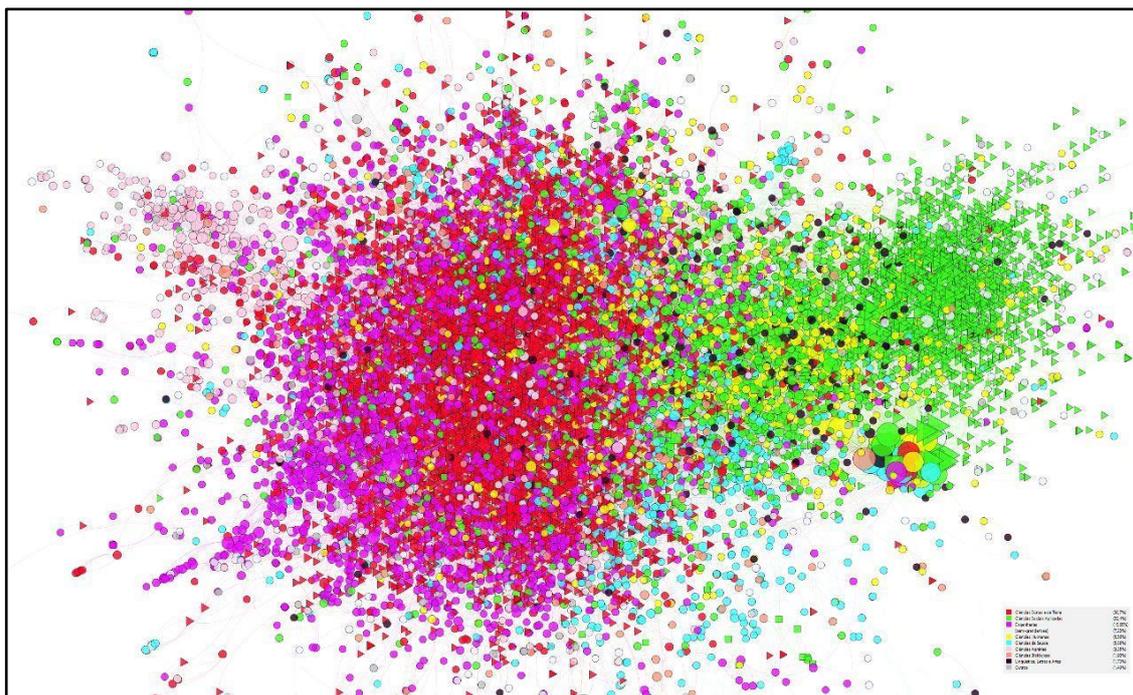
Descrição	Quantidade
Artigos completos publicados em periódicos	252.915
Livros publicados/organizados ou edições	35.329
Capítulos de livros publicados	91.235
Trabalhos completos publicados em anais de congressos	252.264
Resumos expandidos publicados em anais de congressos	56.653
Resumos publicados em anais de congressos	167.322
Total de produção bibliográfica	855.718

Fonte: Dados de pesquisa (2022).

Notou-se que o maior quantitativo de produção bibliográfica consolidada a partir dos 20.400 currículos selecionados está nos artigos completos em semelhante volume para os trabalhos completos em anais de congressos. Sabe-se que cada grande área de pesquisa tem foco maior ou menor entre estas duas tipologias, assim como a de capítulos de livros e livros publicados. O equilíbrio quantitativo dessas duas tipologias está alinhado com as áreas de Ciências da Computação e Direito em destaque na Tabela 1.

A quinta análise de resultados foi sobre a rede de coautoria a partir da produção científica. A Figura 3 expõe o grafo da rede gerada pelo Gephi para o componente gigante usando o layout ForceAtlas 2.

Figura 3 – Rede de coautoria acadêmica (Lattes)



Fonte: Dados de pesquisa (2022).

As atribuições de cores na legenda referem-se às grandes áreas as quais os pesquisadores se destacaram, sendo a primeira listada em seu Lattes. Apesar do grafo retratar toda a produção dos pesquisadores em inteligência artificial, observou-se uma divisão no grafo entre o campo das Ciências Sociais Aplicadas (em verde) e o das Ciências Exatas e da Terra (em vermelho). Destacamos também no grafo as duas áreas mais frequentes, Ciência da Computação e Direito marcados como triângulos e Ciência da Informação com quadrados.

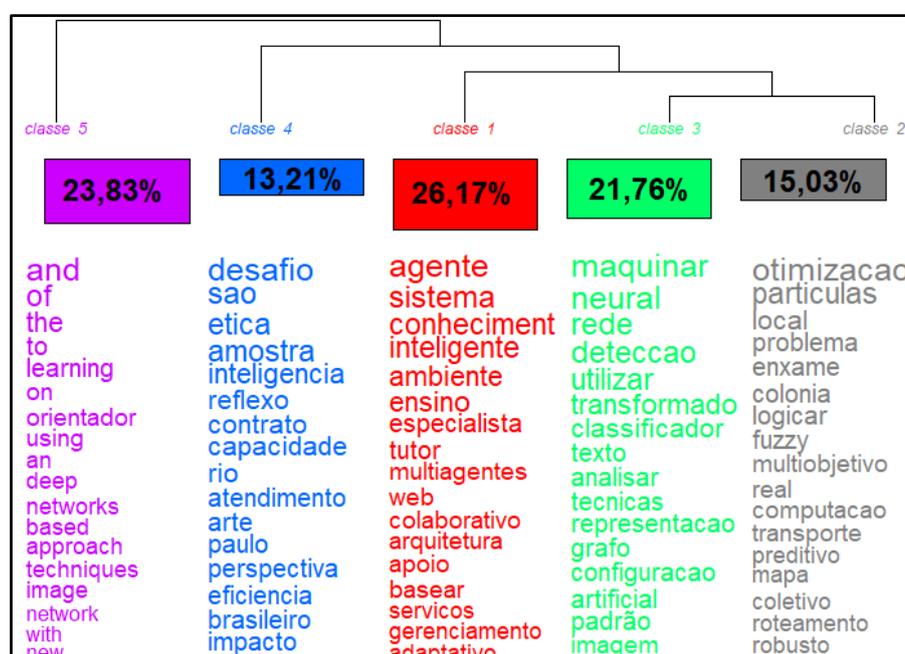
A área de Ciências da Saúde (em rosa-claro) forma um agrupamento à esquerda e as Engenharias ficam margeando a área de Ciências Exatas e da Terra. As Ciências Humanas se encontram distribuídas entre as Ciências Sociais e Aplicadas. Um agrupamento no canto inferior direito apresenta os pesquisadores com maior centralidade de grau e onde se percebe uma multidisciplinaridade de pesquisadores em colaboração.

A sexta análise de resultados foi sobre a rede de termos a partir da produção científica, realizada no Gephi usando o layout ForceAtlas 2 para uma melhor visualização e interpretação dos dados, exibida na Figura 4.

uma correlação das palavras dos Segmentos de Textos (ST) formados com a média três frases, a lematização dos termos para um radical do termo, a divisão com base na frequência de palavras e a produção de um esquema hierárquico de classes chamado dendrograma (SALEM,1986).

A Figura 5 exibe o dendrograma formado por 5 classes geradas pelos 772 segmentos classificados dos 1.002 ST, a partir de 41.077 ocorrências e 6.077 formas que compõem o *corpus* da amostra da Biblioteca Digital de Teses e Dissertações.

Figura 5 – Dendrograma de temas das teses e dissertações (BDTD)



Fonte: IRaMuTeQ (2022).

Na amostra, alguns resumos de teses e dissertações estavam em inglês, os quais foram reunidos automaticamente pelo programa na classe 5 em roxo. Esta opção foi selecionada porque o IRaMuTeQ disponibiliza os arquivos dos segmentos de texto de cada classe para análise e a possibilidade de percepção dos temas contidos no *corpus* textual.

A classe 1 em vermelho (26,17% do total de ST) encontrou aplicações de IA como, por exemplo, engenharia rodoviária, gerenciamento de centrais de informações de fretes, avaliação do comportamento de vigas de aço sujeitas a cargas, modelagem de um ambiente inteligente de educação, previsão de condições operacionais de altos-fornos na siderurgia, gerenciamento de

workflows científicos em bioinformática, e metodologia de representação do conhecimento para construção de ontologias.

A classe 2 em cinza (15,03%) revelou estudos sobre IA relacionados, por exemplo, com sistemas inteligentes para elevação de poços e controle de processos petrolíferos, monitoramento de impactos ambientais, diagnóstico em fisioterapia neuro pediátrica, análise de traumatismo cranioencefálico, análise de crédito utilizando inteligência artificial validação com dados do cartão BNDES, futebol baseado em inteligência coletiva por enxame, e sistemas multiagentes em mercados de energia elétrica.

A classe 3 em verde (21,76%) apontou propostas de uso da IA, como caso de sistema inteligente híbrido para diagnóstico de falhas em motores, modelo de veículos aéreos não tripulados, diagnóstico de câncer de mama pelas redes bayesianas, detecção de padrões e previsão de acidentes em rodovias, predição de perdas de propagação em redes de comunicações LTE (padrão 4G), grafos como uma técnica integrada de raciocínio qualitativo para diagnóstico de falhas, e classificação supervisionada de imagens.

A classe 4 em azul (13,21%) mostrou assuntos referentes a IA, tais como: discussão ética do agrupamento automático de documentos jurídicos, obras artísticas geradas pela inteligência artificial, controvérsias de aplicação em jogos de guerra navais, como as tecnologias estão moldando o novo mercado da música, detecções na espectrometria gama em amostras ambientais, e *chatbots* no processo de atendimento ao cliente.

A classe 5 em roxo (23,83%) apresentou temas no que concerne a IA como redes de sensores aplicadas ao monitoramento da saúde, aplicações em interatividade homem máquina, modelo de agente inteligente para otimização de planejamento em linhas de produção, métodos estatísticos e fuzzy aplicados na avaliação de impactos ambientais, e classificação de tábuas de madeira usando processamento de imagens digitais e aprendizado de máquina, e análise dos mecanismos de *deep reinforcement learning*.

Os temas das pesquisas acadêmicas permeiam a otimização de sistemas em empresas e setores econômicos como petróleo e gás, agricultura, educação e saúde, como também questões sobre o controle de recursos naturais. Além

disso, foi possível observar discussões sobre efeitos de aplicações, ética e regulamentação da inteligência artificial.

5 CONCLUSÃO

Os resultados apontaram para o avanço das aplicações de inteligência artificial em diversas áreas do conhecimento para além das Ciências Exatas e da Terra e Engenharia. Entretanto, embora a questão atravessasse diferentes áreas de conhecimento, a rede de coautoria de produção científica demonstrou ser pouco multidisciplinar. Além disso, foi possível notar um movimento de expansão da inteligência artificial no Direito, área de Ciências Sociais Aplicadas.

As instituições e abrangência geográfica estão diretamente relacionadas no resultado, com a proeminência de volume de produção científica no eixo Sul e Sudeste, apesar do destaque do estado de Pernambuco, o qual no Nordeste possui relevância no investimento em iniciativas de tecnologia da informação, com um grande parque tecnológico de inovação em Recife.

O estudo espera colaborar para uma visão da situação atual das pesquisas sobre inteligência artificial no Brasil, mesmo sendo uma fotografia da produção científica dos últimos 30 anos, bem como ser uma referência no delineamento metodológico quantitativo e aplicação de instrumentos de coleta e análise de dados de maneira a contribuir com futuros trabalhos nas metrias da informação. Esperamos que fiquem abertas a partir deste estudo outras possibilidades de análises quantitativas de produtividade científica sobre inteligência artificial a serem exploradas.

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) – Código de Financiamento 001 e Conselho Nacional de Desenvolvimento Científico e Tecnológico, Processo 430982/2018-6 e 315521/2020-1.

REFERÊNCIAS

- ARAÚJO, C. A. Á. Bibliometria: evolução histórica e questões atuais. **Em Questão**, Porto Alegre, v. 12, n. 1, p. 11-32, jan./jun. 2006.
- BALANCIERI, R.; BOVO, A. B.; KERN, V. M.; PACHECO, R. C. dos S.; BARCIA, R. M. A análise de redes de colaboração científica sob as novas tecnologias de informação e comunicação: um estudo na Plataforma Lattes. **Ciência da Informação**, Brasília, v. 34, n. 1, p. 64-77, 2005.
- BARDIN, L. **Análise de conteúdo**. Lisboa: Edição 70, 1977.
- BASTIAN, M.; HEYMANN, S.; JACOMY, M. Gephi: An Open-Source Software for Exploring and Manipulating Networks. *In: INTERNATIONAL AAAI CONFERENCE ON WEBLOGS AND SOCIAL MEDIA*, 3., 2009, San Jose. **Proceedings [...]**. San Jose, CA: 2009. p. 361-362.
- BERMEJO, P. H. S.; TOLENTINO, M.; SALM JR., J. F.; TODESCO, J. L.; PACHECO, R. C. dos S. O papel da comunidade virtual no processo de construção de ontologias: caso da Plataforma de Governo Eletrônico Lattes. *In: SIMPÓSIO BRASILEIRO DE SISTEMAS DE INFORMAÇÃO (SBSSI)*, 2., 2005, Florianópolis. **Anais [...]**. Florianópolis: SBC, 2005. p. 192-199.
- BRAGA, G. M. Informação, ciência, política Científica: o pensamento de Derek de Solla Price. **Ciência da Informação**, Brasília, v. 3, n. 2, 1974. DOI: 10.18225/ci.inf.v3i2.50.
- CAMARGO, B. V.; JUSTO, A. M. IRAMUTEQ: um software gratuito para análise de dados textuais. **Associação Brasileira de Psicologia**, Ribeirão Preto, v. 21, n. 2, p. 513- 518, 2013.
- CARVALHO, P. R. **Conexão informacional entre as campanhas eleitorais: análise de estratégias e postagens no Facebook**. 2020. 163 f. Dissertação (Mestrado em Ciência da Informação) – Instituto Brasileiro de Informação em Ciência e Tecnologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2020.
- CARVALHO, P. R.; RAMOS, M. G.; PIMENTA, R. M. Curadoria Digital: mapa de bibliotecas nas universidades federais brasileiras. *In: CONGRESSO INTERNACIONAL EM TECNOLOGIA E ORGANIZAÇÃO DA INFORMAÇÃO*, 6., 2020, São Paulo. **Anais [...]**. São Paulo: USP, 2020.
- CARVALHO, P. R.; GOUVEIA, F. C; RAMOS, M. G. Pesquisa exploratória sobre inteligência artificial: análise de áreas de conhecimento, instituições e temas de grupos de pesquisa do CNPq e das pesquisas acadêmicas no BDTD. *In: ENCONTRO BRASILEIRO DE BIBLIOMETRIA E CIENTOMETRIA*, 8., 2022. Maceió. **Anais [...]**. Maceió: UFAL, 2022, p. 100-107.

DOMINGOS, P. **O algoritmo mestre**: como a busca pela máquina de aprendizado definitiva vai refazer nosso mundo. São Paulo: Novatec, 2017.

FERRAZ, R. R. N.; QUONIAM, L. M. A utilização da ferramenta computacional scriptlattes para avaliação das competências em pesquisa no brasil. **Prisma.com**, Portugal, n. 21, p. 222-234, 2013.

FRANA, P. L.; KLEIN, M. J. **Encyclopedia of artificial intelligence**: the past, present, and future of AI. California: ABC-CLIO, LLC, 2021.

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

GONTIJO, M. C. A.; ARAÚJO, R. F.; OLIVEIRA, M. A produção científica sobre inteligência artificial e seus impactos: análise de indicadores bibliométricos e altimétricos. **Pesquisa Brasileira em Ciência da Informação e Biblioteconomia**, João Pessoa, v. 14, n. 3, 2019.

HOBBS, T. **Leviatã ou Matéria, forma e poder de um Estado eclesiástico e civil**. Tradução de João Paulo Monteiro e Maria Betraiz Nizza de Silva. 4. ed. São Paulo: Martins Fontes, 2019.

JACOMY, M.; VENTURINI, T.; HEYMANN, S.; BASTIAN, M. ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. **PLOS ONE**, [S. l.], v. 9, n. 6, 2014.

JARRAHI, M. H. Artificial Intelligence and the Future of Work: Human-AI Symbiosis in Organizational Decision Making. **Business Horizons**, [S. l.], v. 61, n. 4, 2018.

LEE, K. **AI Superpowers**: China, Silicon Valley, and the New World Order. Tradução Marcelo Brandão. Rio de Janeiro: Globo Livros, 2019.

MACIAS-CHAPULA, C. O papel da informetria e da cienciometria e sua perspectiva nacional e internacional. **Ciência da Informação**, Brasília, v. 27, n. 2, p. 134-140, maio/ago.1998.

MAGALHÃES, J. L.; QUONIAM, L. M.; MENA-CHALCO, J. P.; SANTOS, A. Extração e tratamento de dados na base lattes para identificação de core competencies em dengue. **Informação & Informação**, Londrina v. 19, n. 3, p. 30-54, 2014.

MENA-CHALCO, J. P.; JUNIOR, R. M. C. ScriptLattes: an open-source knowledge extraction system from the Lattes platform. **Journal of the Brazilian Computer Society**, [S. l.], v. 15, p. 31-39, 2009.

MITCHELL, M. **Artificial Intelligence**: a guide for thinking humans. Nova York: Farrar, Straus and Giroux, 2019.

MOREIRA, J. R.; MUELLER, S.; VILAN FILHO, J. L. Produção científica dos membros dos grupos de pesquisa das áreas de informação no Brasil. **Informação & Informação**, Londrina, v. 25, n. 1, p. 1-20, 2020.

NICHOLSON, S. The basis for bibliomining: frameworks for bringing together usage-based data mining and bibliometrics through data warehousing in digital library services. **Informations Processing and Management**, v. 42, n. 3, p. 785-804, 2006.

NILSSON, N. J. **The quest for artificial intelligence: a history of ideas and achievements**. Cambridge University Press, 2009.

OLIVEIRA, D. A. de; ARAÚJO, R. F. de. A contribuição das métricas para o campo da ciência da informação. **Perspectivas em Ciência da Informação**, Belo Horizonte, p. 300-318, 2020.

PAULA, M. V. de. **Explorando o potencial da plataforma Lattes como fonte de conhecimento organizacional em ciência e tecnologia**. 2004. 148 f. Dissertação (Mestrado em Informática) - Universidade Católica de Brasília, Brasília, 2004.

RUSSELL, S. J.; NORVIG, P. **Inteligência Artificial: Um Enfoque Moderno**. 2. ed. Madrid: Pearson Education, 2004.

RUSSELL, S. J.; NORVIG, P. **Artificial Intelligence: A Modern Approach**. 4th ed. London: Pearson Education, 2021.

SALEM, A. Segments répétées et analyse statistique des données textuelles. **Histoire & Mesure**, PER-SEE Program, v.1, n. 2, p. 5-28, 1986.

SEARLE, J. R. Minds, brains, and programs. **The Behavioral and Brain Sciences**, [S. l.], v. 3, n. 3, p. 417-457, 1980.

SOUTHWICK, S. B. The Brazilian electronic theses and dissertations digital library: providing open access for scholarly information. **Ciência da Informação**, Brasília, v. 35, n. 2, p. 103-110, maio/ago. 2006.

SUN, Q.; YANG, L. From independence to interconnection: A review of AI technology applied in energy systems. **Journal of Power and Energy Systems**, [S. l.], v. 5, n. 1, p. 21-34, Mar. 2019.

TURING, A. M. Intelligent Machinery. In: INCE, D. C. (ed.). **Mechanical intelligence**. Amsterdam: North Holland, 1948. p. 107-127.

TURING, A. M. Computing Machinery, and Intelligence. **Mind**, [S. l.], v. 59, n. 236, p. 433-460, Oct. 1950.

WANG, L.; LIU, Z.; LIU, A.; TAO, F. Artificial intelligence in product lifecycle management. **The International Journal of Advanced Manufacturing Technology**, [S. l.], v. 114, p. 771-796, 2021.

ARTIFICIAL INTELLIGENCE: BIBLIOMETRIC ANALYSIS OF ACADEMIC RESEARCH, LATTES CURRICULUM AND RESEARCH GROUPS FROM NATIONAL COUNCIL FOR SCIENTIFIC AND TECHNOLOGICAL DEVELOPMENT DATABASE

ABSTRACT

Introduction: Artificial intelligence enables a diversity of applications in different areas of knowledge. The study presents a portrait of the field's performance in the Brazilian context. **Objective:** To map academic research in the scope of dissertations and theses carried out in postgraduate programs, as well as the scientific production regarding the geographical distribution, institutions, areas of knowledge, type of communication channel, coauthor network, and themes related to artificial intelligence. **Methodology:** The exploratory and descriptive research was based on quantitative methodological procedures through Bibliometrics and Content Analysis. The sources of information used were the Research Groups and the Lattes Platform of the National Council for Scientific and Technological Development, as well as the Brazilian Digital Library of Theses and Dissertations, of the Brazilian Institute of Information in Science and Technology. The tools chosen for data extraction, analysis, and visualization were scripLattes, Gephi, and IRaMuTeQ. Data collection was carried out in October 2022, with the Lattes defining the 30-year period. **Results:** 759 Research Groups, 20,400 Lattes curricula, and 3,073 thesis and dissertation records were identified. The most frequent areas of knowledge identified in the three bases covering the issue were the Exact and Earth Sciences, Engineering, and Applied Social Sciences. The most relevant regions in two of the bases were the Southeast and South. The institutions that stood out were the University of São Paulo and the Federal University of Santa Catarina, which bring together the largest scientific production on artificial intelligence in Brazil. Scientific publications are mostly concentrated in full articles published in journals and full papers in conferences. Some identified themes were related to the oil and gas, agriculture, education, and health sectors, as well as ethical and regulatory discussions. **Conclusions:** Researchers and research groups on artificial intelligence in Brazil seem to have experienced a period of growth. Although applications cross different areas of knowledge, the coauthor network of scientific production has shown to be little multidisciplinary.

Descriptors: Artificial Intelligence. Lattes Platform. Research Groups. Digital Library of Theses and Dissertations.

INTELIGENCIA ARTIFICIAL: ANÁLISIS BIBLIOMÉTRICO DE LA INVESTIGACIÓN ACADÉMICA, CURRÍCULOS LATTES Y GRUPOS DE INVESTIGACIÓN DEL CONSEJO NACIONAL DE DESARROLLO CIENTÍFICO Y TECNOLÓGICO

RESUMEN

Introducción: La inteligencia artificial posibilita una diversidad de aplicaciones en diferentes áreas del conocimiento. El estudio presenta un retrato del desempeño del campo en el contexto brasileño. **Objetivo:** Mapear la investigación académica en el ámbito de las tesis y disertaciones realizadas en programas de posgrado, así como la producción científica en cuanto a la distribución geográfica, instituciones, áreas de conocimiento, tipo de canal de comunicación, red de coautoría y temas relacionados con la inteligencia artificial. **Metodología:** La investigación exploratoria y descriptiva se basó en procedimientos metodológicos cuantitativos a través de la Bibliometría y el Análisis de Contenidos. Las fuentes de información utilizadas fueron los Grupos de Investigación y la Plataforma Lattes del Consejo Nacional de Desarrollo Científico y Tecnológico, así como la Biblioteca Digital Brasileña de Tesis y Disertaciones, del Instituto Brasileño de Información en Ciencia y Tecnología. Las herramientas elegidas para la extracción, análisis y visualización de datos fueron scripLattes, Gephi y IRaMuTeQ. La recopilación de datos se llevó a cabo en octubre de 2022, con el Lattes definiendo el período de 30 años. **Resultados:** Se identificaron 759 Grupos de Investigación, 20.400 currícula Lattes y 3.073 registros de tesis y disertaciones. Las áreas de conocimiento más frecuentes identificadas en las tres bases que abarcan el tema fueron las Ciencias Exactas y de la Tierra, Ingenierías y Ciencias Sociales Aplicadas. Las regiones más relevantes en dos de las bases fueron el Sureste y el Sur. Las instituciones que destacaron fueron la Universidad de São Paulo y la Universidad Federal de Santa Catarina, que reúnen la mayor producción científica sobre inteligencia artificial en Brasil. Las publicaciones científicas se concentran en su mayoría en artículos completos publicados en revistas y trabajos completos en conferencias. Algunos temas identificados estaban relacionados con los sectores de petróleo y gas, agricultura, educación y salud, así como discusiones éticas y reguladoras. **Conclusiones:** Los investigadores y grupos de investigación en inteligencia artificial en Brasil parecen haber experimentado un período de crecimiento. Aunque las aplicaciones abarcan diferentes áreas de conocimiento, la red de coautoría de la producción científica ha demostrado ser poco multidisciplinar.

Descriptores: Inteligencia Artificial. Plataforma Lattes. Grupos de Investigación. Biblioteca Digital de Tesis y Disertaciones.

Recebido em: 28.12.2022

Aceito em: 22.03.2023