



# XXI ENANCIB

Encontro Nacional de Pesquisa em Ciência da Informação

50 anos de Ciência da Informação no Brasil:  
diversidade, saberes e transformação social

Rio de Janeiro • 25 a 29 de outubro de 2021

## XXI Encontro Nacional de Pesquisa em Ciência da Informação – XXI ENANCIB

### GT-8 – Informação e Tecnologia

#### PERSPECTIVAS DO USO DO APRENDIZADO DE MÁQUINA EM BIBLIOTECAS: reflexões iniciais de uma pesquisa em andamento

#### *PERSPECTIVES OF THE USE OF MACHINE LEARNING IN LIBRARIES: initial reflections of an ongoing research*

Rafaella Carine Montereis – Universidade de Brasília (UnB)

Dalton Martins Lopes – Universidade de Brasília (UnB)

#### Modalidade: Resumo Expandido

**Resumo:** Trata-se de um relato de pesquisa em andamento acerca das aplicações do *machine learning*, em bibliotecas, cujo objetivo principal é mapear os benefícios e os impactos que o aprendizado de máquina pode oferecer no desenvolvimento de produtos e serviços. Deste modo, recorreu-se à revisão sistemática de literatura, que resultou em um *corpus* inicial de 416 documentos, o qual foram selecionados 51 documentos após a execução final do protocolo de pesquisa. Por fim, conclui-se que este estudo proporciona ao pesquisador refletir acerca da necessidade de desenvolvimento de novas competências profissionais frente a um cenário exponencial de produção de dados.

**Palavras-Chave:** inteligência artificial; aprendizado de máquina; biblioteca; serviços de informação; produtos de informação.

**Abstract:** This is an ongoing research report on machine learning applications in libraries, whose main objective is to map the benefits and impacts that machine learning can offer in the development of products and services. Thus, we resorted to a systematic literature review, which resulted in an initial *corpus* of 418 documents, from which 51 documents were selected after the final execution of the research protocol. Finally, it is concluded that this study allows the researcher to reflect on the need to develop new professional skills in an exponential scenario of data production.

**Keywords:** artificial intelligence; machine learning; library; information services; information products.

## 1 INTRODUÇÃO: UM BREVE CENÁRIO DA INTELIGÊNCIA ARTIFICIAL

A inteligência artificial é um elemento com capacidade transformadora em quase todos os aspectos de nossa vida, e desta forma se destaca também no âmbito profissional e científico. Neste cenário alçaram o *status* de ciência, que se manifestou em diferentes áreas

do conhecimento, inclusive na Ciência da Informação desde a sua gênese, por meio de suas relações interdisciplinares.

Com relação à sua origem, o matemático e cientista da computação inglês Alan Turing publicou em 1950 um histórico artigo, denominado *Computing machinery and intelligence*, considerado um marco no campo da IA. Neste artigo, Turing propôs um teste com o objetivo de verificar se a máquina consegue representar o papel de um humano no jogo da imitação, “enganando” o interrogador, de maneira que ele não consiga fazer distinção entre o humano e a máquina. Já a expressão “Inteligência Artificial” é atribuída a John McCarthy, datada de 1956. McCarthy (2007, p. 2) define a área como “a ciência e engenharia de máquinas inteligentes, especialmente programas inteligentes de computador”. Em suma, ela se baseia na capacidade das máquinas simularem ações de maneira análoga ao raciocínio humano, cujo objetivo é resolver problemas, simular situações ou tomar decisões de maneira inteligente.

Neste cenário, a inteligência artificial produziu um vasto conjunto de técnicas, dentre as quais se destaca o aprendizado de máquina (AM), conhecida também por sua expressão em inglês *machine learning* (ML). No que diz respeito à esta expressão, o seu termo foi cunhado pela primeira vez por Arthur Samuel, em 1959. Ele o definiu como o campo de estudo que permite que computadores aprendam sem que sejam programados explicitamente. Ademais, a ML apresenta duas importantes virtudes que são a possibilidade de resolver tarefas de processamento de informação e a capacidade de operar em ambientes dinâmicos (CASTRO; FERRARI, 2016, p. 15). Em resumo, o aprendizado de máquina tem por objetivo criar sistemas que aprendam com base em experiências a partir de um processo dinâmico de adaptação com o ambiente.

Neste cenário, a aplicação e o desenvolvimento destas tecnologias em unidades de informação podem ocasionar profundas transformações em seu ambiente. E, à medida em que as tecnologias baseadas em IA e ML se desenvolvem, vários aplicativos podem ser capazes de aprimorar e produzir novos produtos e serviços para as bibliotecas. Por exemplo, essas tecnologias podem permitir que as bibliotecas coletem metadados automaticamente de suas coleções digitais ou melhorem o reconhecimento óptico de caracteres de textos, ampliando a sua capacidade de processamento de informações.

Assim, neste cenário promissor, as discussões acerca da temática em questão encontram-se ainda em fase inicial como um tema emergente na CI brasileira. Um exemplo

destas investigações podemos encontrar em Peres (2017), o qual buscou mapear em sua dissertação, os trabalhos produzidos sobre Inteligência Artificial à luz da Ciência da Informação em eventos da Associação Nacional de Pesquisa e Pós-Graduação em Ciência da Informação (ANCIB), e concluiu que a literatura sobre o tema no Brasil ainda é escassa. Outros autores que enveredaram pesquisas sobre o tema foram Silva e Nathansohn (2018, p. 121), o qual em trabalho publicado no Encontro Nacional de Pesquisa e Pós-graduação em Ciência da Informação (ENANCIB), afirmaram que o campo da Inteligência Artificial pode ser muito mais explorado pela área da Ciência da Informação, considerando a sua interdisciplinaridade entre os campos da Ciência da Computação e a Ciência Cognitiva. Outra contribuição brasileira foi dada por Neves (2020, p. 186), que observou que as pesquisas que tratam acerca da Inteligência Artificial e da computação cognitiva em unidades de informação são predominantemente estrangeiras.

Em relação ao âmbito internacional, a produção teórica e prática em Inteligência Artificial aplicada às bibliotecas, possui uma situação muito distinta da brasileira. A literatura publicada na área é substancial, e desta forma, podemos destacar algumas iniciativas. O Instituto de Tecnologia de Massachusetts (MIT), desenvolveu um sistema de aprendizagem de máquina alimentada por redes neurais denominado Hamlet (*How about Machine Learning Enhanced Theses*), cujo objetivo é explorar e enriquecer as coleções de teses do MIT. Outra iniciativa de destaque foi a produção de uma declaração sobre bibliotecas e inteligência artificial pela Federação Internacional de Associação de Bibliotecas e Instituições (IFLA, 2020), que visa delinear as principais considerações sobre o uso de tecnologias de IA e ML em bibliotecas e sugerir os papéis que elas devem se esforçar para assumir em uma sociedade com integração crescente em IA.

Estas iniciativas demonstram o grande potencial que as pesquisas em IA aplicadas a unidades de informação apresentam no desenvolvimento de novos produtos e serviços, dado que essas novas tecnologias auxiliam no tratamento, na análise e na recuperação de grandes volumes de dados, ao buscarem padrões e simplificarem as rotinas de trabalho, permitindo ao bibliotecário centrar os seus esforços no cumprimento de tarefas mais complexas. E, diante do cenário brasileiro *sui generis* com aplicações quase inexistentes e poucas discussões, e dado a relevância desta investigação, em uma perspectiva instigante e revolucionária, o profissional da informação necessita desenvolver novas competências de modo a se adaptar a este novo cenário. E, tendo em vista a natureza disruptiva destas

aplicações em unidades de informação e o cenário exponencial da produção de dados em meio eletrônico, este breve relato, fruto de uma pesquisa de pós-graduação em andamento, se propõe a investigar como o desenvolvimento de produtos e serviços por meio da aprendizagem de máquina podem impactar e beneficiar as unidades de informação.

A seguir, serão apresentados os procedimentos metodológicos empregados no desenvolvimento da pesquisa.

## 2 METODOLOGIA

Esta investigação é derivada de uma pesquisa de mestrado em andamento o qual trata-se de uma investigação de caráter qualitativo e quantitativo, com a abordagem exploratória, de natureza pura e com o uso da pesquisa bibliográfica, por meio da revisão sistemática da literatura (RSL), para a coleta de dados. Deste modo, a finalidade deste trabalho é apresentar um breve recorte acerca do estado da arte da Inteligência Artificial, com ênfase na aprendizagem de máquina, no contexto da Ciência da Informação.

E, para a construção da pesquisa bibliográfica foi utilizado a metodologia de revisão sistemática da literatura (RSL) com base nas diretrizes propostas por Galvão e Ricarte (2019) para o campo da Ciência da Informação, complementados pelos estudos produzidos por Kitchenham (2004) e Felizardo et al (2017) para o campo da Ciência da Computação, com o objetivo de mapear as pesquisas acerca da aplicação da AM em bibliotecas.

A seguir serão apresentados os procedimentos metodológicos empregados no trabalho.

### 2.1 Procedimentos metodológicos

Na fase de planejamento foram produzidas diretrizes para a condução da pesquisa e produção de um protocolo com o propósito de apoiar o desenvolvimento da RSL. Estas diretrizes se baseiam nos objetivos da pesquisa de modo a auxiliar na identificação, seleção e avaliação de estudos.

À vista disso, as palavras-chaves foram selecionadas por meio de consulta prévia à literatura e posterior “tradução” em termos de busca auxiliados pelo Tesouro brasileiro de Ciência da Informação e pelo tesouro da base de dados LISTA (*Library, Information Science Abstracts*). Assim, compõem a *string* de busca da pesquisa: (“aprendizado de máquina” OR

“*machine learning*” OR “*aprendizaje automático*”) AND (biblioteca OR *library*). O recorte idiomático compreende os idiomas português, inglês e espanhol.

Quanto ao corte temporal para a extração dos dados de pesquisa, serão analisadas publicações entre dezembro de 2010 a janeiro de 2021, além disso não haverá restrições quanto à cobertura geográfica ou quanto ao suporte informacional.

Com relação às fontes de informação, foram selecionadas as seguintes bases<sup>1</sup> de dados, a saber: BRACIP<sup>2</sup> (Base de Dados Referencial de Artigos em Ciência da Informação), BDTD<sup>3</sup> (Biblioteca Digital de Teses e Dissertações; LISTA<sup>4</sup> (*Library, Information Science Abstracts*); ISTA<sup>5</sup> (*Information Science & Technology Abstracts*); e, *Web of Science*<sup>6</sup>.

No que tange à inclusão de documentos na RSL, a publicação deverá atender a pelo menos 1 (um) dos critérios estabelecidos para inclusão de documentos na RSL, elencados no quadro abaixo.

---

<sup>1</sup> A base LISA (Library and Information Science Abstracts) não estava disponível para consulta no Portal de Periódicos da CAPES e na Biblioteca Central da Universidade de Brasília em fevereiro de 2021.

<sup>2</sup> Site: <https://brapci.inf.br/>

<sup>3</sup> Site: <https://bdtd.ibict.br/vufind/>

<sup>4</sup> Site: <https://www.ebsco.com/products/research-databases/information-science-technology-abstracts>

<sup>5</sup> Site: <https://www.ebsco.com/products/research-databases/library-information-science-and-technology-abstracts>

<sup>6</sup> Site: <https://www.webofknowledge.com>

Quadro 1: Critérios de seleção de documentos

CRITÉRIOS DE SELEÇÃO DE DOCUMENTOS
INCLUSÃO
(I-1) Documentos que abordem as aplicações, os produtos e serviços, os impactos, os benefícios e as dificuldades para a implantação da AM em bibliotecas;
(I-2) Documentos que tratem acerca das competências do profissional da informação em bibliotecas frente ao uso de novas tecnologias com foco na AM;
(I-3) Documentos que reflitam acerca das questões éticas geradas a partir do uso de novas tecnologias em bibliotecas com foco em AM.
EXCLUSÃO
(E-1) Documentos que não foram publicados na última década;
(E-2) Documentos que não abordem as aplicações, os produtos e serviços, os impactos, os benefícios e as dificuldades para a implantação da AM em bibliotecas;
(E-3) Documentos que não tratem acerca das competências do profissional da informação frente ao uso de novas tecnologias com foco na AM;
(E-4) Documentos que não reflitam acerca das questões éticas geradas a partir do uso de novas tecnologias em bibliotecas com foco em AM;
(E-5) Documentos que não estejam no domínio da Ciência da Informação;
(E-6) Documentos que não forem publicados nos idiomas: Português, Inglês e ou Espanhol;
(E-7) Documentos que não estejam disponíveis para consulta em texto integral no formato online ou físico;
(E-8) Documentos pré-textuais e/ou pós-textuais;
(E-9) Documentos duplicados;
(E-10) Documentos plagiados.

**Fonte:** Elaborado pelos autores.

Com base nos critérios e diretrizes desenvolvidas, os dados foram recuperados, selecionados e organizados para a última fase que corresponde a avaliação e análise quantitativa do *corpus* documental. Nesta etapa foram analisados a quantidade de documentos por base (aprovados e rejeitados), a distribuição da pesquisa por ano, a quantidade de publicações por país, informações acerca da autoria e palavras-chaves utilizadas para representar os documentos e, por fim a distribuição das pesquisas por periódicos.

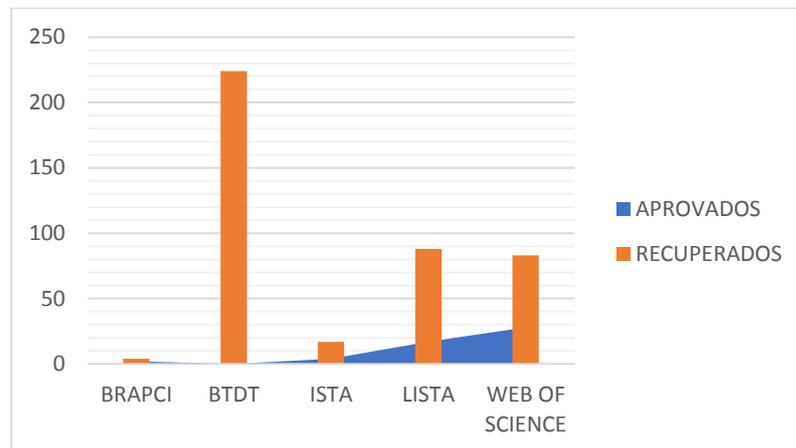
A próxima seção apresentará a análise e a discussão os dados coletados após a execução do protocolo de pesquisa.

### 3 RESULTADOS PRELIMINARES

Após a execução da *string* de busca, entre os dias 15 e 28 de fevereiro de 2021, a pesquisa retornou 4 (quatro) resultados da base BRAPCI, 224 (duzentos e vinte e quatro) da base BDTD, 17 (dezessete) da base ISTA, 88 (oitenta e oito) da base LISTA e, 83 (oitenta e

três) da base *Web of Science*, o que resultou em um *corpus* de 416 (quatrocentos e dezesseis) documentos. Em seguida, ao aplicar os critérios de inclusão e exclusão foram selecionadas 51 (cinquenta e uma) publicações, conforme representação do gráfico 1.

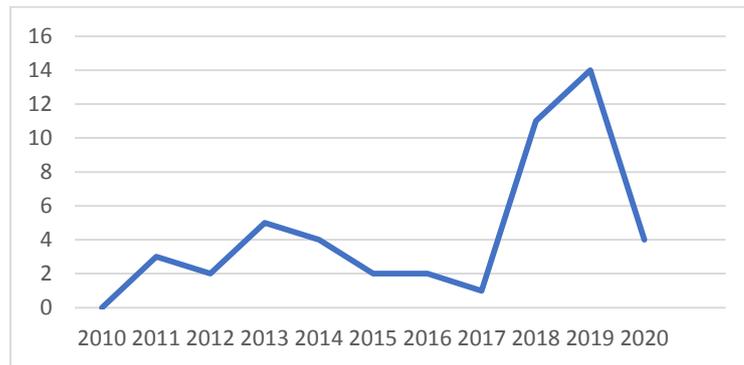
**Gráfico 1: Relação entre documentos x bases de dados**



**Fonte:** Elaborado pelos autores.

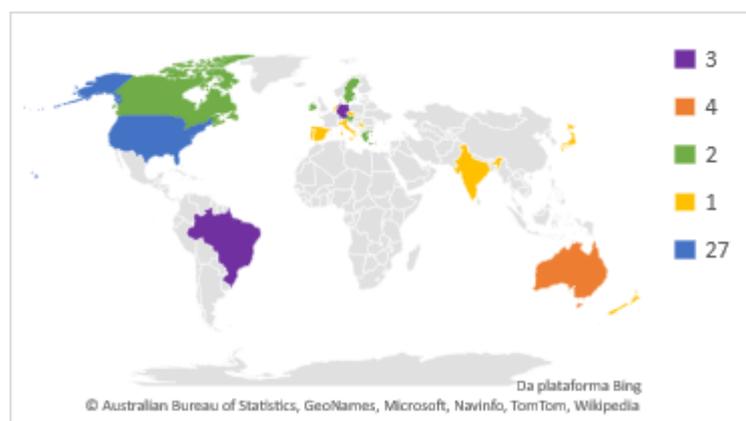
É interessante notar que a base de dados BDTD obteve um expressivo conjunto de documentos recuperados, no entanto, nenhum documento recuperado foi selecionado após a aplicação dos critérios de seleção, o que demonstra o ineditismo da pesquisa.

Com relação à distribuição de pesquisas aprovadas por ano (gráfico 2), verifica-se uma curva ascendente entre os anos de 2017 a 2019 e, logo em seguida uma brusca queda no ano de 2020. O recuo no ano de 2020 pode ser atribuído ao período de pandemia que afetou a produção científica não relacionada ao tema COVID-19, além deste contexto histórico, outro motivo para a queda justifica-se pelo fato de que as pesquisas aprovadas no ano de 2020 podem ainda não ter sido finalizadas e publicadas nos repositórios institucionais das universidades.

**Gráfico 2: Distribuição das pesquisas aprovadas por ano**

**Fonte:** Elaborado pelos autores.

No que se refere às instituições que mais publicaram sobre a temática em questão, identificaram-se com três publicações: a Universidade Federal de Minas Gerais e a Universidade de Limerick. Em seguida com duas publicações: a Universidade de Linnaeus, a Evenly distributed (instituição privada), a Universidade de Illinois, e a Universidade de Rhode Island. No entanto, se analisarmos as publicações por origem, nota-se que o país com o maior número de publicações são os Estados Unidos com 27 publicações sobre o tema, embora as duas instituições com maior número de publicações pertençam ao Brasil e a Irlanda.

**Gráfico 2: Quantidade de publicações por país**

**Fonte:** Elaborado pelos autores.

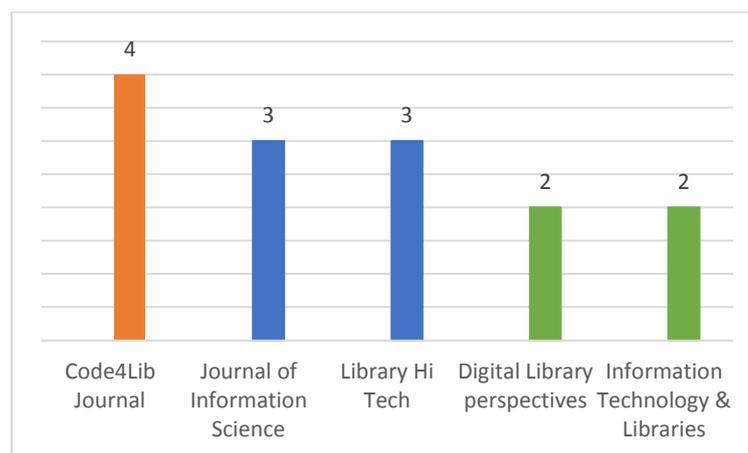
Já em relação à avaliação dos dados de autoria das pesquisas, destacam-se os pesquisadores: Arash Joorabchi, Abdulhussain E. Mahdi e Alberto H. F. Laender, filados respectivamente os dois primeiros à Irlanda e o último ao Brasil, com três publicações cada. Bohyun Kim, Jason Griffey, Anderson Ferreira, Alan Filipe Santana e Koraljka Golub, com duas publicações cada, filiados a instituições dos Estados Unidos, Brasil e Suécia. Todavia,

apesar de os Estados Unidos se configurarem como o principal centro de publicações sobre o tema no mundo, outros países se destacam em número absoluto de publicações por pesquisadores, respectivamente o Brasil e a Irlanda com três publicações cada.

No que se refere ao conjunto de palavras-chaves adotadas pelos pesquisadores, a expressão *machine learning* lidera com 26 menções, seguida de *artificial intelligence* com 11 menções, *metadada* e *libraries* com 7 menções cada, e por fim *digital library* com 6 menções. As palavras-chaves utilizadas revelam uma tendência das pesquisas em IA e ML serem aplicadas ao contexto das bibliotecas digitais.

Por fim, o gráfico 3 apresenta a relação dos periódicos com maior número de publicações realizadas sobre ML em bibliotecas.

**Gráfico 3: Distribuição das pesquisas aprovadas por periódico**



**Fonte:** Elaborado pelos autores.

Constata-se que o periódico com mais publicações é o *Code4Lib Journal*, uma revista de acesso aberto americana, produzida por voluntários voltada para o público interessado em tecnologias da informação e inovações no contexto das bibliotecas. Em seguida, com três publicações respectivamente: *Journal of Information Science*, periódico acadêmico pertencente ao grupo Sage Journals, que cobre as áreas da Ciência da Informação e da gestão do conhecimento, e o periódico *Library Hi Tech*, publicação do grupo Emerald Insight, cujo objetivo é publicar artigos, relatórios de conferências e estudos de caso sobre tecnologias em bibliotecas. Em seguida, com duas publicações, aparece novamente um periódico do grupo Emerald Insight denominado *Digital library perspectives* que publica pesquisas relacionadas à curadoria e objetos digitais, e o periódico *Information Technology & Libraries* de responsabilidade da *American Library Association (ALA)* que publica material

relacionado a todos os aspectos da tecnologia da informação em bibliotecas. Em síntese, observa-se um grande número de publicações periódicas especializadas no segmento da tecnologia em bibliotecas.

#### 4 CONCLUSÃO

Este ensaio propôs uma breve discussão teórica acerca dos conceitos basilares da Inteligência Artificial, com ênfase em *machine learning*, aplicada às bibliotecas sem a presunção de esgotar o tema em questão. No entanto, é importante ressaltar que explorar estes conceitos no contexto das bibliotecas, permite ao pesquisador refletir sobre os novos desafios do profissional da informação frente a inclusão de novas tecnologias no campo informacional, tendo em vista que as aplicações em AM já são uma realidade em muitos países conforme os estudos selecionados apontaram. Ademais, é notável que para garantir o desenvolvimento tecnológico em unidades de informação é necessário e pertinente que o bibliotecário busque por conhecimento de modo a adquirir habilidades e competências necessárias para aplicar a AM. Todavia, as publicações sobre o tema ainda são pouco expressivas no cenário brasileiro, se concentrando em instituições pertencentes aos Estados Unidos e, por conseguinte, majoritariamente em língua inglesa.

Desta forma, este trabalho apresentado em formato de síntese, com resultados em estágios iniciais, se propõe a avançar em novas discussões sobre a aplicações da aprendizagem de máquina em produtos e serviços de bibliotecas relatados na literatura; além das novas competências do profissional da informação e das questões éticas geradas neste novo contexto, e, por último, produzir um conjunto de recomendações para a aplicação de técnicas em aprendizagem de máquina para bibliotecas, que serão publicados em sua totalidade na dissertação.

Por fim, à medida em que novas tecnologias são incorporadas às bibliotecas, revela-se de suma importância analisar este novo cenário, de forma a assegurar o seu uso de maneira segura, ética e sustentável de modo a promover a missão social do bibliotecário.

#### REFERÊNCIAS

ANDROMEDA, Yelton. Hamlet: neural-net-powered prototypes for library discovery. In: GRIFFEY, J. (org.). Artificial intelligence and machine learning in libraries. **Library Technology Reports**, v. 55, n. 1, jan. 2019, p. 10-15.

ARTHUR, Samuel. **Computing bit by bit, or Digital computers made easy**. In: Proceedings of the Institute of Radio Engineers, 41, 19 p. 1223-1230.

CASTRO, Leandro Nunes de; FERRARI, Daniel Gomes. **Introdução à mineração de dados: conceitos básicos, algoritmos e aplicações**. São Paulo: Saraiva, 2016.

FEDERAÇÃO INTERNACIONAL DE ASSOCIAÇÃO DE BIBLIOTECAS E INSTITUIÇÕES (IFLA). **IFLA statement on libraries and artificial intelligence**. IFLA: Haia, 17 set. 2020.

FELIZARDO, Kátia Romero et. **Revisão sistemática da literatura em engenharia de software: teoria e prática**. Rio de Janeiro: Elsevier, 2017.

GALVÃO, Maria Cristine Barbosa; RICARTE, Ivan Luiz Marques. Revisão sistemática da literatura: conceituação, produção e publicação. **Logeion: Filosofia da Informação**, v. 6, n. 1, p. 57-73, 15 set. 2019.

KITCHENHAM, Barbara. **Procedures for Performing Systematic Reviews**. Keele: Keele University, 2018.

MCCARTHY, John. **What is artificial intelligence?** Stanford: Universidade de Stanford, 2017.

NEVES, Barbara Coelho. **Inteligência artificial e computação cognitiva em unidades de informação: conceitos e experiências**. Logeion: Filosofia da Informação, v. 7, n. 1, p. 186-205, 13 set. 2020.

PERES, Fernando. Rodrigues. **O estudo da inteligência artificial no âmbito da Ciência da Informação**. 2017. 64 p. Dissertação (Mestrado em Ciência da Informação) – Universidade de Londrina, Londrina, 2017.

SILVA, Narjara; NATHANSOHN, Bruno. Análise da produção científica em Inteligência Artificial na área da Ciência da Informação no Brasil. In: ENANCIB, 19., 2018, Marília. **Anais [...]**. Marília: ANCIP, 2018. p. 111-216.

TURING, Alan Mathison. Computing machinery and intelligence. **Mind**, Inglaterra, v. 59, n. 236, oct. 1950, p. 433-460.