

Inteligência Artificial da perspectiva da Ciência da Informação: onde estamos em termos de raciocínio computacional?

Jeanne Louize Emygdio¹

Resumo: Avanços notáveis têm sido possíveis com os algoritmos para raciocínio automático da Inteligência Artificial (IA). Tais avanços levam cientistas da informação a questionar se há sentido na classificação tradicional após a chegada dos motores de busca. A questão impacta não só a CI, mas a Biblioteconomia, a medida em que os dados do século XXI não estão mais em estantes. Não existe restrição física que exija a organização para a qual os esquemas de classificação foram originalmente concebidos. Questiona-se mesmo se um esquema prévio pode prever necessidade de um usuário. O objetivo deste artigo é discutir os limites do raciocínio automático e em que medida substitui o humano na classificação. A metodologia adotada compreende: i) revisão bibliográfica sobre fundamentos do raciocínio humano e computacional; ii) a descrição do raciocínio automático produzido por duas abordagens populares: a ontologia e o aprendizado de máquina; iii) comparação entre os tipos de raciocínio. Foram resultados encontrados: as abordagens automáticas exibem claras limitações, mas ontologias são o que há de mais próximo daquilo que pessoas podem fazer na classificação e, portanto, mais próxima da CI. Conclui-se, portanto, que a assistência da IA é profícua, mas as abordagens analisadas – raciocínio automático clássico e probabilístico – estão longe de substituir o raciocínio humano em tarefas classificatórias.

Palavras-chave: ontologias; classificação automática; aprendizado de máquina; inteligência artificial; organização do conhecimento.

ARTIFICIAL INTELLIGENCE from the perspective of INFORMATION SCIENCE: WHERE ARE WE TOWARD COMPUTATIONAL REASONING?

Abstract: Notable advances have been possible from the use of algorithms for automatic reasoning of Artificial Intelligence (AI). Such advances lead information scientists to question whether there is any sense in the traditional classification after the advent of search engines. The issue impacts not only CI, but Librarianship since the data of the 21st century are no longer on shelves. There is no physical restriction that requires the organization for which the classification schemes were originally designed. Some claim even if it is possible that a previous scheme predicts a user's need. This paper aims to discuss the limits of automatic reasoning and the extent to which it substitutes the human in the classification tasks. The methodology

¹Doutora em Gestão e Organização do Conhecimento. Pesquisadora do Laboratório Representação do Conhecimento, Ontologias e Linguagem (RECOL/UFGM). E-mail: jeanne.emygdio@gmail.com. Lattes: <http://lattes.cnpq.br/6849175974671037>. ORCID: <https://orcid.org/0000-0002-7329-4447>.

adopted was: i) bibliographic review on essentials of human and computational reasoning; ii) the description of the automatic reasoning produced by two popular approaches: ontology and machine learning; iii) comparison between types of reasoning. Findings: we found that automatic approaches have clear limitations, but this is something well-known. The best contribution is to identify ontologies as closest to what people can do in the classification and, therefore, the closest to the Information Science. Conclusions: AI assistance can be useful for human, insofar as the analyzed approaches – classic automatic reasoning and probabilistic reasoning – are far from replacing human reasoning in classificatory tasks.

Keywords: ontologies. automatic sorting; machine learning; artificial intelligence; knowledge organization.

1 INTRODUÇÃO

O que parece conectar seres humanos e máquinas a partir dos sistemas de IA é a capacidade de raciocinar ou de produzir inferências. Em geral, se alguém infere, chega a uma conclusão. O raciocínio simples é feito a partir de proposições, sentenças que as pessoas usam no dia a dia para se referir aos fatos do mundo.

Do ponto de vista filosófico, infere-se uma proposição “p” a partir de uma proposição “q” quando se afirma “p” com base em “q”. Do mesmo modo, infere-se uma proposição “p” do fato “F” quando se afirma “p” com base na ocorrência de “F”. Além disso, infere-se uma proposição “p” de uma proposição “q” quando qualquer um que afirme “q” está comprometido em afirmar “p” (SPARKES, 1991). Por exemplo, *Tom* afirma que estava em Londres na última terça-feira às 12h. *John* infere que *Mary* não poderia estar almoçando com *Tom* em *Brighton* na terça-feira, mas *Jay* afirma que viu *Mary* almoçando em *Brighton* naquele dia de terça-feira. Então, *John* sabe que alguém não está dizendo a verdade.

As pessoas usam o raciocínio também e, sobretudo, para classificar, uma atividade fundamental em Biblioteconomia e Ciência da Informação (BCI). De maneira similar, a classificação também tem sido realizada com algum sucesso pelos algoritmos de IA. A conexão entre estes campos – IA e BCI – já foi levantada na literatura há mais décadas: “*seres humanos usam categorias para pensar e falar, e os algoritmos de IA também precisam de algum tipo de esquema de categorias para funcionar adequadamente*” (SARACEVIC, 1996, p.1).

No entanto, imprecisões cometidas pela IA têm chamado a atenção, não apenas dos cientistas, mas também público em geral. Um exemplo é o caso da classificação da Declaração de Independência dos Estados Unidos como um discurso de ódio pelo algoritmo do *Facebook* (THE GUARDIAN, 2018). Este tipo de incorretude deixa claro que ainda há muito trabalho a ser feito para tornar a IA parte de rotina das pessoas. No escopo da BCI, questões diversas também têm sido levantadas sobre a interação entre a classificação realizada por humanos e por máquinas (HJØRLAND, 2012).

O presente artigo lança alguma luz sobre o que realmente pode ser feito pela IA. Para atingir os objetivos, apresentam-se fundamentos do raciocínio e da classificação humana (Seção 2); em seguida, apresentam-se fundamentos básicos da IA e explicadas as duas principais abordagens atuais para o raciocínio automático – ontologias e aprendizado de máquina – descrevendo seu funcionamento (Seção 3); finalmente, discute-se em que medida o raciocínio automático pode substituir o raciocínio humano em tarefas de classificação e afins, e conclui-se sobre a proximidade entre pessoas e BCI e a pesquisa em ontologias.

2 RACIOCÍNIO HUMANO: VISÃO

Esta seção explica conceitos básicos do raciocínio humano e as principais teorias científicas (seção 2.1), além de introduzir atividades de interesse para os objetivos aqui relacionados ao raciocínio, em particular, a classificação (seção 2.2).

2.1 TEORIAS DO RACIOCÍNIO

Existem teorias das Ciências Cognitivas e da Psicologia consideradas como as abordagens mais aceitáveis para se compreender como as pessoas raciocinam. Esses campos de pesquisa acumularam um vasto volume de resultados empíricos durante as últimas décadas, considerando quatro principais abordagens ao raciocínio: i) a abordagem baseada em regras, ou baseada em regras sentenciais; ii) a abordagem de modelos mentais; iii) a abordagem probabilística; iv) a abordagem de complexidade relacional (WARREN, 2017).

A abordagem baseada em regras, que remonta ao século XIX, advoga que o raciocínio humano é baseado em sistemas formais. Tais sistemas proporcionam a capacidade

de obter conclusões a partir de sentenças através de: i) conectivos lógicos, como “e”, “se”, “ou”, “não”; ii) quantificadores, como “todos”, “alguns”.

Um subconjunto dos quantificadores dá origem à predicados para silogismos, como originalmente analisado por Aristóteles. Por exemplo, se todos sabem que João considera festas entediantes, pode-se deduzir que John não estará presente na festa desta noite (RIPS, 1983). Este exemplo incorpora um princípio lógico familiar, denominado, *modus ponens*, uma regra de inferência que pode ser resumida por “P implica Q e P é afirmado como verdadeiro, portanto, Q deve ser verdadeiro”. No entanto, nem todas as situações são tão diretas e extensas teorias têm sido concebidas pelos adeptos da abordagem baseada em regras para explicar operações mais complexas (BRAINE; O'BRIEN, 1991; STALNAKER, 1968; 1984).

Uma questão discutida na literatura é o grande número de modelos produzidos pela abordagem do raciocínio baseado em regras. Por exemplo, há 64 premissas lógicas possíveis para um simples silogismo contendo apenas duas premissas e uma conclusão. Manter tantos modelos na memória aumenta a possibilidade de esquecer ou negligenciar alguns, o que pode levar a conclusões inconsistentes ou a um viés de resposta em conclusões válidas (JOHNSON-LARD; BARA, 1984).

A abordagem baseada em regras tem sido classificada como um caso particular de outra, denominada abordagem do modelo mental, considerado o modelo mais genérico. A abordagem do modelo mental se tornou popular em Psicologia, Ciência Cognitiva e Linguística por defender que para raciocinar, as pessoas criam modelos da realidade e verificam conclusões confrontando-as com estes modelos (JOHNSON-LAIRD, 1999).

Outra abordagem, denominada abordagem probabilística, oferece evidências empíricas para demonstrar que as pessoas cometem erros substanciais e sistemáticos ao realizar tarefas de raciocínio padrão. De acordo com os preceitos de outras abordagens, as pessoas seriam caracterizadas como irracionais, embora os defensores da perspectiva probabilística discordem da interpretação. Para alcançar resultados razoáveis, a abordagem probabilística usa mais a probabilidade condicional e menos a lógica. Apesar de seu sucesso relativo, a abordagem

probabilística não provê explicação sobre a operação subjacente ao raciocínio humano (OAKSFORD; CHATER, 2001).

A abordagem de complexidade relacional foi criada para quantificar a complexidade associada com o raciocínio sobre as relações. De acordo com essa perspectiva, a complexidade de uma relação está no número de elementos associados, em outras palavras, em sua cardinalidade: unária, binária, ternária e assim por diante. Para a maioria das pessoas, o limite máximo é a cardinalidade quatro. Em outras palavras, quatro é o número de relações que as pessoas podem lidar em função das restrições de memória (HALFORD et al., 2007). Por fim, a abordagem de complexidade relacional é classificada com um tipo de abordagem de modelo mental em que os modelos são conceitualizados de forma diferente (ZIELINSKI et al., 2010).

Por último, cabe enfatizar que não há evidências suficientes para estabelecer uma teoria unificada para o raciocínio humano. As variações na forma como as pessoas raciocinam, a complexidade do cérebro humano, as abordagens distintas, as diferentes maneiras de enfrentar problemas, dentre outros fatores têm dificultado a concepção de uma teoria definitiva (STENNING; YULE, 1997; STENNING; VAN LAMBALGEN, 2008).

2.2 CLASSIFICAÇÃO E CATEGORIZAÇÃO

Mesmo assim, o raciocínio é um tema de pesquisa crucial abordado por várias teorias de campos notáveis, como já mencionado, e está também relacionado a outra investigação significativa para a Ciência da Informação: os estudos em classificação. Certamente, é impossível pensar sem categorias e atribuição de uma categoria a algo pode ter impactos na sociedade. Por exemplo, a sentença “quando um ser humano começa a existir?” (SMITH; BROGAARD, 2003, p. 1) exibida em uma análise metafísica, pode ser transcrita como “quando um organismo possui características que o qualificam a pertencer à categoria dos seres humanos?”

De forma rigorosa, “classe” não é um sinônimo exato para “categoria”, muito embora os termos sejam usados de forma análoga (JACOB, 2004). A questão aqui é observar as categorias naturais e construídas pelo homem. O primeiro caso considera tipos naturais, entidades aristotélicas que existem independentemente dos pensamentos

humanos; o último considera classes, artefatos humanos criados para propósitos específicos (JANSEN, 2008). Neste sentido, para ilustrar a diferença, pode-se falar sobre a “categoria das árvores”, que engloba todas as árvores do mundo; e, a “classe das árvores no jardim daquela universidade”, por exemplo, com o propósito de designar um jardineiro para cuidar das plantas. Deste ponto em diante, mesmo admitindo-se estas distinções, busca-se respeitar o uso de um ou outro termo de acordo com a escolha do autor.

A pesquisa envolvendo categorias e classificação é um assunto central dentro da BCI, onde se pode observar uma história de pelo menos 200 anos de pesquisa (TENNIS, 2016). No entanto, várias questões discutíveis a respeito de ambas classes e categorias veem povoando a mente dos cientistas: como classificar entidades do mundo? Qual é a relação entre as classes e o mundo? Como classificações científicas devem ser construídas? Existe uma categoria superior única ou existem várias? Como distinguir uma categoria da outra? (STUDTMANN, 2013).

A Filosofia ofereceu numerosas respostas com ênfase em três escolas principais: i) essencialismo, que classifica as coisas de acordo com sua essência natural, um método originado em Aristóteles; ii) análise de conjuntos, que divide entidades em grupos cujos membros compartilham características similares, embora nenhuma delas seja a essencial, como defendido por exemplo, por *Wittgenstein*; e iii) classificação histórica, que classifica entidades de acordo com relações causais ao invés de características qualitativas, por exemplo, a sugestão de *Darwin* para classificar organismos (ERESHEFSKY, 2000).

A despeito da escola escolhida, um bom ponto de partida para refletir quão desafiador é lidar com categorias, é lembrar que algo pode ser membro de mais de uma. Por exemplo, uma “flor” pode ser observada, e então categorizada como “planta”, um “tipo de decoração”, um “item para venda”, e assim por diante. A atribuição de uma variedade de categorias para uma única entidade vem do uso de hierarquias: uma hierarquia permite considerar que “flor”, é um “objeto físico”, junto com “carro” e “pedra”; mas também algo membro de categorias como “planta”, “ser vivo”, dentre outros (GORMAN; SANFORD, 2004).

Portanto, hierarquias não resolvem tudo, pois pode-se relacionar categorias de várias outras formas. Por exemplo, a flor é uma “planta” e também “um item para venda”. Por outro lado, um “carro” pode ser “algo para venda”, mas não é uma “planta”. Nem tudo “à venda” é uma “planta”. Assim, as categorias se sobrepõem gerando obstáculos à organização hierárquica. Uma forma de evitar tal sobreposição é adotar uma distinção simples previamente mencionada: uma flor é, de acordo com sua própria natureza, uma espécie de planta, uma entidade natural; por outro lado, é um “item para venda” não é natural, é uma classe que existe porque alguém tomou a decisão de vendê-la em uma loja. Se forem consideradas apenas categorias naturais, não haverá sobreposição e a ordem hierárquica é preservada (GORMAN, 2004).

Porém, sabe-se que os seres humanos criam divisões do mundo e, portanto, categorias. Uma “faixa de pedestre” desenhada no piso da rua é um exemplo dessa divisão; um “código postal” é outro exemplo. De fato, as pessoas dividem o mundo por atividades humanas de demarcação. Então, a questão neste caso é discutir se as categorias criadas pelas pessoas são mapeadas para o mundo ou se impõem divisões ao mundo. Para ilustrar o primeiro caso, considerando-se um projeto arquitetônico, a realidade deve corresponder ao projeto através da construção de algo; da mesma forma, para ilustrar o segundo caso, uma corporação deve corresponder ao estatuto da corporação que estabelece obrigações aos seus membros e limites às suas ações (BITTNER; SMITH, 2008).

A realidade pode ser vista de diferentes maneiras e cada uma destas visões envolve um esquema particular de categorias, ou, uma perspectiva, a partir do ângulo do observador. Por exemplo, engenheiros veem “carros” de forma diferente dos clientes que compram o próprio carro. Este fato leva a outras questões, por exemplo, se há mais de um esquema, um para cada propósito específico, e como todos poderiam ser correlacionados? Por um lado, os problemas podem se resolver a partir de um ponto de vista prático, criando um esquema para cada necessidade; por outro lado, ao criar vários esquemas, o cientista vê-se na posse de vários arranjos sem um bom parâmetro para determinar qual deles é melhor.

Assim, como não há uma forma objetiva de decidir qual é o melhor ou o mais básico, cabe a decisão humana sobre qual esquema usar em cada situação. É importante recordar que todas estas questões levantadas são ontológicas por natureza, de forma que as diferentes perspectivas não modificam a realidade bruta em si. Além disso, elas não cobrem várias outras situações que envolvem moralidade, valores, direitos, para mencionar alguns, mas provém uma boa forma de lidar com a complexidade das categorias.

3 RACIOCÍNIO COMPUTACIONAL

Na Ciência da Computação, em particular na IA, a palavra “inferência” é associada com inferência lógica. Porém, existe outro tipo de inferência adequada problemas em situações em que a lógica clássica não é adequada. No domínio do conhecimento incompleto, da incerteza, as inferências são ditas “probabilísticas”. Ambos os tipos de inferência mencionados se dizem capazes de realizar processos de classificação de forma autônoma. Na presente seção, avalia-se o sentido do termo “classificação” no escopo dessas duas abordagens – a lógica e a probabilística – as quais são representadas, respectivamente, por ontologias e aprendizado de máquina (AM).

3.1 INFERÊNCIAS LÓGICAS COM ONTOLOGIAS

Adentrar o reino da IA, um campo complexo que combina Ciência da Computação, Ciência Cognitiva, Linguística, para mencionar alguns, exige estar informado sobre princípios fundamentais e diferentes termos técnicos. Na presente seção, apresentam-se tais princípios para então explicar qual classificação automática pode ser obtida com o uso de ontologias.

No âmbito da IA, há linguagens de representação compostas por conjuntos de sentenças lógicas, conjuntos esses denominados “bases de conhecimento” (BC). As BCs são capazes de representar um domínio do conhecimento para propósitos computacionais. A palavra “sentença” tem um sentido técnico: é uma assertiva, uma sentença que declara fatos sobre o mundo. Tais sentenças tem sintaxe e semântica: enquanto a primeira especifica as sentenças bem estruturadas, por exemplo, “ $x + y = 4$ ” é bem estruturada e “ $x \ 2 \ y + =$ ” não é; a segunda define a verdade a respeito das sentenças em um modelo, por exemplo, a semântica da aritmética especifica que

uma sentença “ $x + y = 4$ ” é verdadeira em um modelo para o qual $x = 2$ e $y = 2$, mas é falsa em um modelo para o qual $x = 1$ e $y = 1$.

Assim, modelos expressam a noção de verdade ao englobar um conjunto de sentenças verdadeiras. Uma inferência lógica é realizada por algoritmos para verificação de modelos, os quais enumeram os possíveis modelos e verificam se certa sentença é verdadeira em todos os modelos que compõem a BC (RUSSELL; NORVIG, 2007). Neste contexto, inferências lógicas envolvem implicações lógicas, por exemplo, se a sentença B é verdadeira em todos os modelos em que a sentença A é verdadeira diz-se que “A implica B”. As implicações lógicas geram conclusões, por exemplo, as sentenças “Sócrates é um homem” e “Todos os homens são mortais” permite inferir que “Sócrates é mortal”. Além das implicações lógicas, outras noções são importantes para o entendimento das inferências no âmbito da IA moderna, a saber: *consistência*, *confiabilidade*, *completude*, *equivalência*, *validade*, *satisfatibilidade* (GENESERETH; NILSSON, 1988).

A *consistência* descreve um conjunto de sentenças onde nenhuma contradiz qualquer outra. A consistência, contudo, não é um processo e sim uma condição. Quando o algoritmo produz apenas sentenças “implicadas” – conclusões lógicas obtidas a partir de premissas lógicas via implicação – diz-se que se mantém a verdade e o algoritmo é confiável.

Outra propriedade relevante é a *completude*: a capacidade do algoritmo em derivar quaisquer sentenças implicadas. A *equivalência* é obtida quando duas sentenças são verdadeiras para o mesmo conjunto de modelos. Uma sentença possui *validade* se for verdadeira em todos os modelos. Finalmente, a *satisfatibilidade* descreve uma sentença que é verdadeira em algum modelo, ou pelo menos é satisfeita em algum modelo da BC (LEVESQUE, 1989).

Aplicações contemporâneas, como a Web Semântica, contém dispositivos – os “motores de inferência” ou “raciocinadores automáticos” – capazes de realizar inferências automáticas no escopo das ontologias como artefatos. Motores de inferência são artefatos de software projetados para organizar e raciocinar sobre categorias da ontologia e operam sobre a Lógica de Descritiva (BAADER et al., 1992). A Lógica Descritiva consiste de uma família de lógicas que provê linguagens de

representação capazes de criar definições para categorias, além de relacionamentos entre essas elas.

Um dos usos mais comuns dos motores de inferência é a classificação. O termo “classificação” nomeia o processo de verificar se uma instância pertence a certa classe. Além da classificação, motores de inferência também podem realizar tarefas adicionais, por exemplo: subsunção, que verifica se uma classe é um subconjunto de outras classes pela comparação de suas definições; e consistência, que verifica se os critérios de associação a uma classe são logicamente satisfatórios. Uma linguagem baseada na web e composta por perfis de Lógica Descritiva é a *Web Ontology Language* (OWL), um fragmento da Lógica de Primeira Ordem (FOL) que serve como linguagem de representação padrão na Web Semântica (W3C, 2019).

Para compreender o funcionamento de um motor de inferência, considere-se um conjunto de arquivos OWL contendo “O”, uma ontologia que inclui vários axiomas envolvendo classes. O axioma é uma sentença lógica considerada verdadeira. Neste contexto, uma vez acionado, o motor de inferência executa três testes: de consistência, de satisfatibilidade e de subsunção (SATTLER; STEVENS; LORD, 2014).

Primeiro, o motor de inferência executa o teste de consistência verificando se existe um modelo de O. Isso significa perguntar se existe uma estrutura de classes e relações para os quais todos os axiomas em O sejam verdadeiros. Por exemplo, uma ontologia composta por cinco axiomas (Quadro 1) falha neste teste pois “Bob” não pode ser instância de classes disjuntas, a saber “humano” e “esponja”. Classes disjuntas são justamente aquelas que não podem ter membros em comum, e portanto, nenhuma subclasse em comum. Se apenas um axioma for violado, a O falha no teste de consistência.

Quadro 1: Alguns axiomas da ontologia O

VERSÃO LÓGICA	VERSÃO EM LINGUAGEM NATURAL
1. Classe: Humano	1. Há uma classe chamada Humano
2. Classe: Esponja	2. Há uma classe chamada Esponja

3. Instância: Bob tipo: esponja	3. Bob é uma instância da classe Esponja
4. Instância: Bob tipo: humano	4. Bob é uma instância da classe Humano
5. Disjuntas: Humano, Esponja	5. Instância da classe Humano não pode ser instância da classe Esponja

Fonte: Elaborado pelos autores (2021).

Segundo, o motor de inferência executa o teste de satisfatibilidade verificando se existe um modelo de O contendo a instância “ x ” da classe “ A ”. Isso significa verificar a existência de uma estrutura de classes e relações que satisfaça a todos os axiomas em O , além de verificar se A tem a instância x .

Quadro 2: Axiomas e relações na ontologia O

VERSÃO LÓGICA	VERSÃO EM LINGUAGEM NATURAL
1. <i>Célula Eucariótica SubClassOf: Célula AND (temParte some Núcleo)</i>	1. Célula eucariótica é um tipo de célula que tem núcleo como parte
2. <i>Hemácias SubClassOf: Célula Eucariótica AND (temParte ONLY (nãoNúcleo))</i>	2. Hemácia é um tipo de célula eucariótica que não tem núcleo como parte
3. <i>Sangue SubClassOf: (temParte some Hemácia)</i>	3. Sangue é um tipo de entidade que tem Hemácias como parte

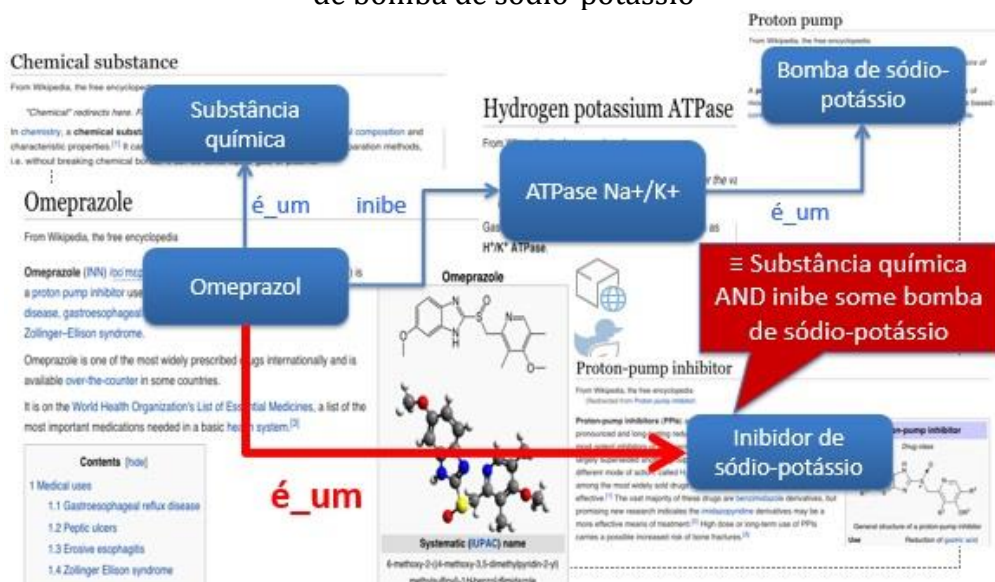
Fonte: Elaborado pelos autores (2021).

Neste caso, uma “Hemácia” tem um “núcleo” porque é uma “célula eucariótica” (Quadro 2, linha 1) e, ela não tem “núcleo”, pois é declarada como célula sem um “núcleo” (Quadro 2, linha 2). Assim, um modelo de O não pode ter uma instância de Hemácia, pois tal instância teria como características “núcleo” e “sem núcleo”. Diz-se que O contém classes não-satisfatórias, à saber, “Hemácias” e “Sangue”.

Terceiro, para duas classes “ A ” e “ B ” que pertencem à O , o motor de inferência faz o teste de subsunção, ou seja, testa se “ A é subclasse de B ”. Em outras palavras, o teste verifica se, em cada modelo de O , cada instância de A é também instância de B . Há mais de uma maneira de testar a subsunção, por exemplo, pode-se construir um modelo com uma instância de A e “não B ” e, se tal modelo falhar, então A e “não B ” não podem ter instâncias em nenhum modelo de O . Portanto, A é menos genérico que B , ou dito de outra forma, A é subsumido por B .

Um exemplo real na área Biomédica ilustrar o teste de subsunção, também chamado “classificação” (Figura 1). Uma BC possui a classe “omeprazol” como subclasse da categoria “substância química”; há uma outra relação estabelecendo que “omeprazol” inibe “ATPase Na⁺/K⁺”, um tipo de “bomba de sódio-potássio”. Nesse caso, a BC infere que “omeprazol” é também um “inibidor de sódio-potássio” (seta vermelha na Figura 1).

Figura 1: Inferência lógica - Omeprazol é um inibidor de bomba de sódio-potássio



Fonte: Brochhausen (2018).

3.2 INFERÊNCIAS PROBABILÍSTICAS NO APRENDIZADO DE MÁQUINA

Historicamente, dois tipos de aprendizado se destacam na pesquisa em IA. O primeiro é ilustrado por computadores jogando xadrez: o sistema armazena os movimentos que resultam em xeque-mate e tentam aplicá-los novamente. O segundo tipo de aprendizado é a generalização: o sistema tenta aplicar aprendizado prévio em situações análogas, mas não idênticas. Na verdade, o AM atual consiste de algoritmos que manipulam dados para tomar decisões de acordo com probabilidades. Funciona de forma similar ao cérebro humano, baseando-se em probabilidades para decidir se sabe algo. Por exemplo, uma pessoa olha para um gato persa e se baseia em sua experiência prévia (dados) para identificar o gato como gato e, especificamente, como gato Persa (MCLAY, 2018).

Apesar do sucesso atual, a AM é um campo originado na década de 1950. Computadores das últimas décadas eram lentos e de memória insuficiente para lidar com grandes conjuntos de dados exigidos e, assim, apenas recentemente o AM provou sua utilidade. Enquanto os algoritmos tradicionais precisam de regras estabelecidas para controlar seu uso, um algoritmo de AM é capaz de aprender a partir de conjuntos de dados para então descobrir padrões e expandir sua capacidade de predição (APTÉ, DAMERAU e WEISS, 1994).

O termo “predizer”, em geral, se refere à habilidade de afirmar o futuro, mas em AM diz respeito a capacidade de atribuir um valor a uma variável, nem sempre envolvendo um aspecto temporal. Se existe uma variável de saída que se deseja predizer – quantitativa ou não – um algoritmo de AM pode ser aplicado, a partir de um conjunto de características, de acordo com as seguintes etapas (KELLEHER; MAC; D'ARCY, 2015):

- Existe um conjunto de dados chamado *dados de treinamento*, que já contém resultados desejáveis e apresenta mensurações típicas para um conjunto de objetos conhecidos;
- A partir do conjunto de dados de treinamento, o algoritmo pode construir um *modelo de predição*, tornando possível prever uma saída para outros objetos até então desconhecidos.

De maneira semelhante, o processo de construção de modelos para predições baseadas em dados conhecidos se baseia em séries históricas de dados similares e é denominado *análise preditiva de dados* (KELLEHER; MAC; D'ARCY, 2015). Neste processo, existem duas abordagens centrais:

- *Aprendizado supervisionado*: o nome “supervisionado” se refere a variável de resultado que orienta o processo de aprendizado;
- *Aprendizado não supervisionado*: não há variável de resultado e o objetivo é saber como dados são agrupados e identificar padrões.

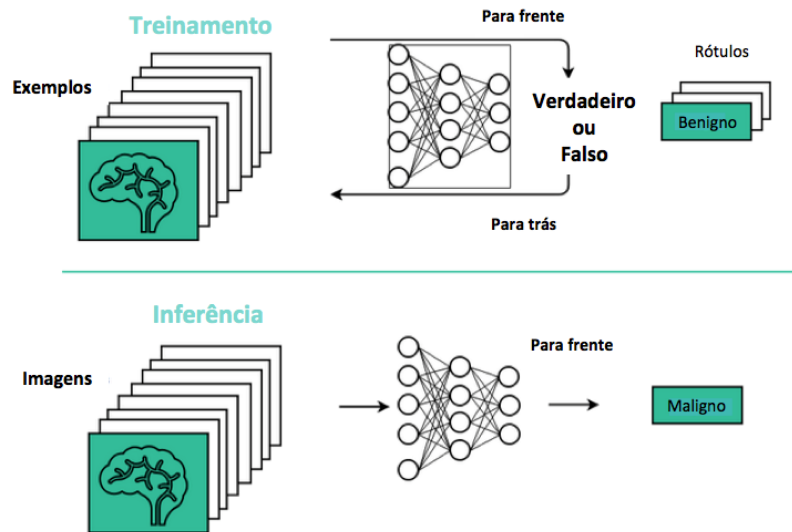
Na abordagem supervisionada, duas aplicações bem conhecidas são: i) a classificação, em que se classifica o tipo de uma variável, além de prover uma saída

quantitativa e; ii) a regressão, em que as medidas de saída são qualitativas. Exemplos de problemas atuais tratados com sucesso pelas abordagens supervisionadas são (HASTIE; TIBSHIRANI; FRIEDMAN, 2008):

- Predição de *spams*, é um problema de classificação: os dados de treinamento são em milhões de e-mails já classificados por pessoas, além das frequências relativas das palavras que ocorrem nas mensagens de e-mail;
- Identificação de câncer, é um problema de regressão por conta do resultado qualitativo: há uma série de correlações entre o nível de antígenos específicos da próstata (PSA) e outras medições clínicas em um grupo de homens prestes a sofrer prostatectomia. O objetivo foi predizer o “log de PSA” (lpsa) – uma escala logarítmica de tempo versus valores marcadores do tumor de próstata – e mais dez outras medições para variáveis que impactam no tratamento;
- Identificação manuscrita, é um problema de classificação: uma amostra de dados extraída de códigos postais manuscritos em envelopes de correio e a tarefa é predizer, a partir de uma matriz de intensidades de *pixels*, a identidade de cada imagem (0,1, ..., 9), ou seja, decidir se uma figura é um número.

Um exemplo de abordagem não supervisionada pode ser encontrado no domínio das expressões de DNA (HASTIE; TIBSHIRANI; FRIEDMAN, 2008). O problema tem duas variáveis principais – genes e amostras – para agrupar as amostras do experimento. A inferência com AM foi obtida com dados da radiologia (Figura 2) para detectar células cancerosas.

Figura 2: Inferências probabilísticas: depois de treinado, o algoritmo pode detectar células malignas



Fonte: Njenga (2016).

4 DISCUSSÃO

Apresentaram-se uma visão geral sobre o raciocínio humano e a classificação, alguns conceitos seminais de IA e mais mecanismos de inferências por ontologias e por AM. A partir daí, foi possível discutir tópicos polêmicos, a partir de uma breve comparação entre o raciocínio humano (seção 2) e o raciocínio autônomo (seção 3).

Como é fácil perceber, as teorias subjacentes ao raciocínio humano são, de alguma forma, refletidas no reino do raciocínio autônomo. A abordagem baseada em regras humanas corresponde às inferências em ontologias; e a abordagem probabilística humana corresponde a certas modalidades de algoritmos de AM. A abordagem do modelo mental humano é imitada por modelos que devem ser satisfeitos na abordagem baseada em regras, na qual, em todos os modelos possíveis, o conjunto de sentenças deve ser verdadeiro.

Tais conclusões não são originais, pois o esforço de aprofundar nestes assuntos complexos não é uma tarefa trivial. Cabe uma revisão mais ampla e publicação de artigos no futuro para incluir mais detalhes. Após tal ressalva, o restante da presente seção descreve falhas e limitações do AM e das ontologias. Essa descrição conduz à uma reflexão sobre os relatos encontrados hoje facilmente em revistas e mesmo em

artigos científicos, permitindo ao leitor verificar o que é sensacionalismo e o que é situação real.

Vários relatos de fontes populares têm apresentado a utilidade do AM em uma gama de situações. Markoff (2012), por exemplo, ressalta que o *deep learning* – uma modalidade de AM – já está presente na vida diária, em serviços como a assistente pessoal *Siri* da *Apple* e no *Street View* da *Google*. Lewis-Kraus (2016) apresenta sua visão sobre como o AM está transformando o tradutor do *Google* a partir de um trabalho contínuo, mesmo quando as pessoas estão dormindo.

Por outro lado, também existem estudos apontando empecilhos e falhas na utilização de algoritmos de AM. Marcus (2018) afirma que o AM é apenas uma técnica estatística, e tais técnicas incluem desvios em suas próprias conclusões e suposições. Apresenta uma lista de dez desafios enfrentados por aplicações de AM (MARCUS, 2018). As aplicações de ML:

- Precisam de cada vez mais dados, continuamente;
- São superficiais e de capacidade limitada para reuso de resultados, envolvendo cenários apenas ligeiramente modificados;
- Não são capazes de lidar com estruturas hierárquicas;
- Não são capazes de lidar com inferências de mundo-aberto;
- Não são transparentes, e por isso chamados de “caixas-pretas”;
- Não exibem integração com conhecimento especializado;
- Não podem distinguir causalidade de correlação;
- Supõem um mundo amplamente estável;
- Operam como aproximações de respostas nem sempre confiáveis;
- São difíceis de projetar.

O paradigma atual de AM é dominado pelo que se chama de “AM agnóstico profundo”, no qual os pares entrada-saída são empregados para gerar modelos probabilísticos em um processo chamado treinamento (vide seção 3.2). Tais modelos são: i) estocásticos, porque eles manipulam dados apenas em sentido probabilístico, por exemplo, entradas são conectadas às saídas probabilisticamente e só; ii) agnósticos, porque não se baseiam em conhecimento especializado relacionado à tarefa em que estão envolvidos; iii) profundo, no sentido de que a arquitetura tem múltiplas camadas de unidades computacionais distribuídas (LANDGREBE; SMITH, 2019).

Os modelos são agnósticos, ou seja, não possuem conhecimento de linguística. Várias condições devem ser atendidas para que modelos assim operem adequadamente, condições estas que parecem estar ao alcance apenas de governos e de gigantes corporativos do mundo da tecnologia da informação (LANDGREBE; SMITH, 2019):

- Exigem um vasto volume de dados de treinamento, tuplas de entrada-saída conectadas pelo melhor desempenho humano;
- Dados de entrada-saída devem ser semelhantes porque AM requer padrões resultantes de processos recorrentes;
- Dados de entrada também devem ser abundantes, pois milhões de registros são necessários para representar uma variância completa.

Brochhausen (2018) apresenta um exemplo simples da falta de conhecimento ou de semântica para explicar a forma como algoritmos de ML gerenciam representações de objetos (Figura 3). Ao se considerar três objetos comuns e distintos, por exemplo, uma casa, uma mesa e um carro, a ausência de qualquer semântica leva os algoritmos de AM a identificá-los apenas como “coisas azuis”.

Figura 3: Três diferentes objetos identificados como coisas azuis



Fonte: Brochhausen (2018).

Em outro caso no âmbito jurídico, o algoritmo de ML COMPAS², utilizado para avaliação do grau de periculosidade de um réu e determinação de sua pena, apresentou um viés ao atribuir maior probabilidade de reincidência criminal a um grupo de pessoas, sem correspondência com a realidade (YONG, 2018). O caso gera reflexões sobre três dos desafios para AM apontados por Marcus (2018), que são relacionados a característica, já mencionada, da caixa preta: i) a falta de transparência sobre o cálculo do índice de risco que dificulta o entendimento das justificativas para as penalidades; ii) a utilização de resultados probabilísticos para justificar penas, sem análise das variáveis, comprovando que respostas por aproximação não são plenamente confiáveis e, por fim; iii) a prática de aprendizado por generalização se aplica apenas a situações análogas, mas não idênticas (MCLAY, 2018). Estes três aspectos sinalizam a necessidade da intervenção humana para acompanhamento e análise de resultados probabilísticos de AM, identificação das causas de distorções e a busca por correções.

Algumas críticas sugerem que o custo destes processos automáticos, usados na AM, são proibitivos, caso se espere resultados precisos. Após treinar um grande volume de dados, as abordagens convencionais de AM ficam estagnadas no tempo, porque, diferentemente dos humanos, não podem continuar a aprender a todo momento. Tal fato provoca a necessidade de recursos ininterruptos para treinamento dos algoritmos (KANAN, 2019).

A questão dos custos parece também ser a maior reclamação em relação ao raciocínio a partir das ontologias. A necessidade de especialistas para a construção

² *Correctional Offender Management Profiling for Alternative Sanctions*, software proprietário americano. Disponível em: <http://www.northpointeinc.com/files/downloads/Risk-Needs-Assessment.pdf>. Acesso em: 21/01/2021.

de ontologias, às quais, apenas depois de completamente axiomatizadas e populadas podem oferecer alguma inferência, tem onerado os projetos, a despeito de notáveis exceções (ASHBURNER et al., 2000; BARD; RHEE; ASHBURNER, 2005; ROSSE; MEJINO, 2003; SCHEUERMAN; CEUSTERS; SMITH, 2009). Outra restrição bem conhecida é o limite da capacidade de representação do mundo em lógica e o equilíbrio entre expressividade e computabilidade (RECTOR et al., 2019). Também não é claro se inferências em ontologias podem ser usadas em aplicações diárias, por exemplo, nos conhecidos sistemas de suporte à decisão. Sugere-se que raciocínio silogístico fornecido por ontologias é mais útil para os desenvolvedores que as constroem, do que para um leigo ou um especialista. Dificuldades conhecidas para inferências em ontologias são (RECTOR et al., 2019):

- Ontologias fazem uso de conjuntos não estruturados de regras, e regras desse tipo já provaram dificuldades em escalabilidade;
- A forma de raciocínio, ou, o processo denominado “justificação”, para lidar com inferências em OWL é difícil de operar por inspeção manual.

Há tentativas de modificar e adicionar constructos à linguagem OWL, para torná-la mais acessível à especialistas (HORRIDGE et al., 2011). Pelo menos, cabe destacar, ontologias não sofrem da falta de semântica observada nos algoritmos de AM.

5 CONSIDERAÇÕES FINAIS

Neste artigo introduziram-se os processos de raciocínio humano e de raciocínio computacional para propósitos de classificação. Alguns conceitos seminais da IA e do raciocínio humano foram discutidos, na tentativa de compreender realizações dos computadores que executam algoritmos de IA. Após essa elucidação, discutiram-se limitações das abordagens de IA, observando que muito ainda o que fazer. Observa-se que há uma diversidade de artigos apresentando vantagens e desafios das tecnologias que tentam substituir tarefas humanas, mas não há muitas iniciativas na CI. Esse artigo é uma iniciativa nesse sentido. Espera-se que os progressos da IA possam ser úteis e não repitam o ocorrido no XX, quando as tão anunciadas novidades não eram factíveis, frustrando as expectativas da sociedade quanto ao futuro da área.

A respeito dos aspectos da classificação, em função de tudo o que foi apresentado, acredita-se que IA pode funcionar como assistente para as pessoas em um contexto onde o grande volume de dados suplanta a capacidade biológica do cérebro humano. É realmente difícil imaginar que, com a tecnologia atual, seja possível falar em inteligência nas máquinas: trata-se na verdade de uma abordagem “força-bruta” obtida pelos avanços na memória e na capacidade de processamento.

Como Searle (1980) já provou há mais de 40 anos com o seu caso do Quarto Chinês, computadores estão distantes da inteligência humana:

“Imagine um inglês nativo que não conhece chinês trancado em um quarto cheio de caixas de símbolos chineses (um banco de dados) junto com um livro de instruções para manipulação destes símbolos (o programa). Imagine que as pessoas fora do quarto, enviem outros símbolos chineses, que, desconhecidos pela pessoa no quarto, são perguntas em chinês (entrada). E imagine que seguindo as instruções do programa o homem no quarto é capaz de enviar símbolos chineses que são respostas corretas para as questões (saída). O programa habilita a pessoa no quarto a passar pelo teste de Turing para compreender chinês, mas ela não entende uma palavra em chinês”.

REFERÊNCIAS

- APTÉ, C., DAMERAU, F., WEISS, S. M. Automated learning of decision rules for text categorization. *ACM Transactions on Information Systems*, v. 12, n. 3, p. 233–251, 1994.
- ASHBURNER, M. et al.. Gene Ontology: tool for the unification of biology. *Nat Genet.* v. 25, n.1, p. 25-29, 2000.
- BAAEDER, F. et al. (1992). *Terminological Knowledge Representation: A Proposal for a Terminological Logic*. Retrieved March 2, 2015 from <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.41.8116>
- BARD J., RHEE SY, ASHBURNER M. An ontology for cell types. *Genome Biol.*, v. 6, n. 2, 2005.
- BITTNER, T., SMITH, B.. A Theory of Granular Partitions. *In: Katharine Munn, Barry Smith (eds.). Applied Ontology: an introduction*, Ontos-Verlag, Berlin. 2008.
- BRAINE, M. D., O'BRIEN, D. P. A theory of if: A lexical entry, reasoning program, and pragmatic principles. *Psychological Review*, Vol. 98, No.2, 182, 1991.
- BROCHHAUSEN, M. (2018). *Lecture Notes*. University of Arkansas For Medical Science. <https://humanities.uams.edu/faculty/mathias-brochhausen-ph-d/>

ERESHEFSKY, M. *The Poverty of the Linnaean Hierarchy: A Philosophical Study of Biological Taxonomy*. Cambridge University Press, Cambridge. 2004.

GENESERETH, M.R., NILSSON, J. *Logical foundations of artificial intelligence*. Morgan Kaufmann, San Francisco. 1988.

GORMAN, M., SANFORD, J.J. Aristotelian Tradition. In: Michael Gorman and Jonathan J. Sanford. (Eds.). *Categories: Historical and Systematic Essays*. The Catholic Univ of America Press, Washington D.C.. 2004.

GORMAN, M.. Categories and Normativity. In: Michael Gorman and Jonathan J. Sanford. (eds.). *Categories: Historical and Systematic Essays*. The Catholic Univ of America Press, Washington D.C.. 2004.

HASTIE, T., TIBSHIRANI, R., FRIEDMAN, J.. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, Springer. 2008.

HJØRLAND, B. (2012). *Is classification necessary after Google?* Retrieved March 30, 2019 from http://www.iskoiberico.org/wp-content/uploads/2014/09/19-30_Hjorland.pdf.

HORRIDGE *et al.* (2011). *The Cognitive Complexity of OWL Justifications*. Retrieved August 13, 2012, from <http://ceur-ws.org/Vol-745/>

JACOB, E. K. (2004). *Classification and Categorization: A Difference that Makes a Difference*. Retrieved March 3, 2018, from <https://www.ideals.illinois.edu/handle/2142/1686>

JANSEN, L. Classifications. In: Katharine Munn, Barry Smith (eds.). *Applied Ontology: an introduction*, Ontos-Verlag, Berlin .8. ed. 2008. v. 9, p. 342.

JOHNSON-LARD, P.N., BARA, B. G.. *Cognition, Syllogistic inference*, v.16, p. 1-61, 1984.

JOHNSON-LAIRD, P.N. Deductive Reasoning. *Annual Review of Psychology*, v. 50, p. 109–35, 1999.

KANAN, C.. *Incremental Learning in Deep Neural Networks using Memory Replay*. Cognitive Science Colloquium. State University at Buffalo. 2019.

KELLEHER, J. D., MAC NAMEE, B., D'ARCY, A. *Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies*. Cambridge, The MIT Press. 2015.

LEVESQUE, H. J.. Philosophical Logic and Artificial Intelligence. In Richmond H. Thomason (ed.). *Logic and the Complexity of Reasoning*, Springer, Dordrecht. 1989.

LEWIS-KRAUS, G.(2016). *The Great A.I. Awakening*. Retrieved December 22, 2018, from <https://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html>

- MARCUS, G. (2018). *Deep Learning: A Critical Appraisal*. Retrieved December 22, 2018, from <https://arxiv.org>
- MARKOFF, J. (2012). *Scientists See Promise in Deep-Learning Programs*. Retrieved April 13, 2016, from <https://www.nytimes.com/2012/11/24/science/scientists-see-advances-in-deep-learning-a-part-of-artificial-intelligence.html>
- MCLAY, R. (2018). *Managing the rise of Artificial Intelligence*. Retrieved November 10, 2018, from <https://tech.humanrights.gov.au/sites/default/files/inline-files/100%20-%20Ron%20McLay.pdf>.
- NJENGA, P. W. (2016). *Adapting Deep Learning to Medicine*. Retrieved April 26, 2016, from <https://aws.amazon.com/blogs/startups/adapting-deep-learning-to-medicine-with-behold-ai/>
- RECTOR, A. et al. On beyond Gruber: “Ontologies” in today’s biomedical information systems and the limits of OWL. *Journal of Biomedical Informatics: X*, v. 2, p. 100002, 1 jun. 2019.
- RUSSELL, S., NORVIG, P. *Artificial Intelligence: modern approach*. New York Prentice Hall. 2017.
- OAKSFORD, M., CHATER, N. The probabilistic approach to human reasoning. *Trends in Cognitive Sciences*, v. 5 n. 8, p. 349-357, 2001.
- ROSSE, C.; MEJINO, J. L. V. A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *Journal of Biomedical Informatics, Unified Medical Language System*. v. 36, n. 6, 2003.
- SARACEVIC, T. (1996). *Interdisciplinary Nature of Information Science*. Retrieved August 30, 2017 from http://www.brapci.inf.br/repositorio/2010/03/pdf_dd085d2c4b_0008887.pdf
- SATTLER, U., STEVENS, R., LORD, P. (2014). *How Does a Reasoner Work?* Retrieved August 12, 2016 from <http://ontogenesis.knowledgeblog.org/1486>
- SCHEUERMANN RH, CEUSTERS W, SMITH B. Toward an ontological treatment of disease and diagnosis. *Summit on Translat Bioinforma*, 2009.
- SEARLE, J.. Minds, Brains and Programs. *Behavioral and Brain Sciences*, v. 3, p. 417–57, 1980.
- SMITH, B., BROGAARD, B. Sixteen Days. *The Journal of Medicine and Philosophy*, v. 28, p. 45–78. 2003.
- STALNAKER, R. A theory of conditionals. In: N. Rescher (ed.), *Studies in Logical Theory*. Blackwell, Oxford.1968.
- STENNING, K., VAN LAMBALGEN, M. *Human reasoning and cognitive science*. MIT Press. 2008.

STENNING, K., YULE, P.. Image and language in human reasoning: a syllogistic illustration. *Cognitive Psychology*, v. 34, p. 109–159, 1997.

STUDTMANN, P. (2013). *Aristotle's Categories*. Retrieved March 30, 2019 from <https://plato.stanford.edu/entries/aristotle-categories/>

TENNIS, J. T.. Structure of Classification Theory: On Foundational and the Higher Layers of Classification Theory. *Proceedings of thr 13th Annual Conference on Knowledge Organization*. Rio de Janeiro, Brazil. 2016.

THE GUARDIAN (2018). *Facebook labels declaration of independence as 'hate speech'*. Retrieved March 29, 2019 from <https://www.theguardian.com/world/2018/jul/05/facebook-declaration-of-independence-hate-speech>

ZIELINSKI, T. A., GOODWIN, G. P., HALFORD, G. S. Complexity of categorical syllogisms: an integration of two metrics. *European Journal of Cognitive Psychology*, Vol. 22, No 3, 391–421. 2010.

W3C CONSORTIUM. *Web Ontology Language*. Retrieved January 3, 2019 from <https://www.w3.org/OWL/>

YONG, E. A Popular Algorithm Is No Better at Predicting Crimes Than Random People. *The Atlantic*. **Section** Technology. Disponível em: <https://www.theatlantic.com/technology/archive/2018/01/equivant-compass-algorithm/550646/>. Acesso em: 22 jan. 2021.