TÍTULO: Implicações na filosofia da ciência a partir da utilização de Redes Neurais Artificiais

RESUMO:

O avanço tecnológico permitiu que novos algoritmos de inteligência artificial emergissem, dentre os novos modelos, os de redes neurais artificiais se apresentam como um dos mais disseminados, devido sua acurácia na previsão. No entanto, este tipo de algoritmo é comumente chamado de "caixa-preta" por utilizar múltiplas camadas obscuras, tornando a inacessível sua lógica-matemática e gerando problemas que incidem na filosofia da ciência. O objetivo deste artigo é a apresentação de problemas epistemológicos que emergem da utilização de inteligência artificial, e mais especificamente, do método de redes neurais artificiais (RNA).

PALAVRAS CHAVE:

Redes neurais, inteligência artificial, filosofia da ciência

1. Introdução

Em março de 2016, O Alpha Go, Inteligência artificial criada pela Google foi o primeiro programa a derrotar o campeão mundial de Go. A partir daquela data foi estabelecido um marco na inteligência artificial. O jogo de Go, muito difundido na China, possui uma complexidade muito alta. Ele possui 10^{171} posições possíveis, contra 10^{50} no xadrez, tornando inviável a utilização dos algoritmos criados até então. Utilizando profundas redes neurais artificiais (RNA), o Alpha Go conseguiu este feito, utilizando apenas *reinforcement learning*, um modelo de RNA dinâmico que atua sem a interferência de dados humanos (Hubert *et al.*, 2017).

É enorme o potencial do uso desta tecnologia, assim como suas implicações epistemológicas e éticas (Bostrom, 2003; Bostrom e Yudkowsky, 2011). Problemas insolúveis por humanos, agora podem ser desvendados por esta tecnologia que não utiliza apenas dados estruturados e supervisionados para funcionar, mas também RNA (Russell *et al.*, 2016).

Alan Turing, na década de 1940, desenvolveu um teste que consistia em um operador humano interagir com a inteligência artificial (IA) de forma direta e física, onde não seria possível a distinção de o robô de um ser humano. Para isso, seria necessário a criação de um robô e de uma inteligência artificial que conseguisse ter a percepção em tempo real e interagisse (Russell *et al.*, 2016).

Acreditávamos que este teste perduraria por muito tempo, entretanto, o Google Duplex, inteligência artificial do Google já conseguiu passar por este teste em contatos telefônicos utilizando seus algoritmos de redes neurais (Kietzmann, Paschen e Treen, 2018).

Pesquisas anteriores (Horvitz, Breese e Henrion, 1988) já analisavam as implicações da IA em ciências sociais como a teoria de decisão. Entretanto a renovação de interesse por este tópico se dá quando a evolução e disseminação de alguns algoritmos, como as RNA mudam a forma como os postulados são entendidos.

Além disso, o presente artigo também apresenta sua relevância por buscar a integração entre áreas de conhecimentos distintas (Horvitz, Breese e Henrion, 1988) objetivando o avanço do assunto em âmbito acadêmico.

Nesse sentido, o objetivo deste artigo é a apresentação de problemas epistemológicos emergidos da utilização de inteligência artificial. Especificamente, implicações do método de redes neurais artificiais (RNA) sobre postulados.

Especificamente, na primeira parte do artigo será apresentado o advento da inteligência artificial e seus reflexos nos estudos de filosofia da ciência. Em seguida, serão apresentados os modelos atuais de IA, especificamente as RNA. Ademais serão apresentadas suas implicações na filosofia da ciência. Por fim serão debatidas algumas alternativas de estudos que endereçam os problemas apresentados.

2. O advento da Inteligência artificial

Existem várias definições de Inteligência artificial na literatura (Faceli *et al.*, 2011), desde as mais orientadas a racionalidade como "o estudo das faculdades mentais através do uso de modelos computacionais"(Charniak e McDermott, 1985) até as mais orientadas sob a ótica humanista como "A arte de criar máquinas que executam funções que exigem inteligência quando executadas por pessoas"(Kurzweil *et al.*, 1990). Para este trabalho será adota a premissa da tomada de ação baseada em um agente racional:

> Um agente é apenas algo que age (agente vem do latim *agere*, fazer). É claro que todos os programas de computador fazem algo, mas espera-se que os agentes de computação façam mais: operar de forma autônoma, perceber seu ambiente, persistir por um período de tempo prolongado, adaptar-se à mudança e criar e perseguir metas. Um agente racional é aquele que age de forma a alcançar o melhor resultado ou, quando há incerteza, o melhor resultado esperado. (RUSSELL. et al., 2016, p. 4, tradução nossa)

No verão de 1956, um dos grandes marcos para a história da inteligência artificial aconteceu em Dartmouth. Um grupo de pesquisadores, liderados por John McCarty realizou um workshop de dois meses sobre inteligência artificial. Dentre os pesquisadores estavam Minsky, Shannon, Rochester, Arthur Samuel, Ray Solomonoff, Oliver Selfridge, Allen Newell e Herbert Simon. As principais descobertas e avanços na área não aconteceram naquele ano, mas a criação deste grupo fez com que o desenvolvimento sobre o tema ganhasse consistência e progredisse (Russell *et al.*, 2016).

No decorrer da década de 1960, apesar de muita expectativa gerada pouco progresso foi feito. Por apenas realizarem manipulações sintéticas simples, a maioria dos programas não conseguia corresponder às expectativas quando confrontadas com a realidade (Russell *et al.*, 2016).

Apenas nos anos 1970 é que ocorreu uma proliferação das aplicações da AI em solução de problemas reais (Faceli *et al.*, 2011). Nesse sentido, os pesquisadores mudaram a abordagem, e em detrimento a um olhar raso sobre o problema e um profundo trabalho sobre o algoritmo. Os pesquisadores se aprofundaram em ambos os casos, o que resultou na utilização efetiva da IA (Russell *et al.*, 2016). Depois de ultrapassarem esta barreira, nos anos 80 a IA evoluiu não apenas na área acadêmica, mas agora também para a indústria, onde está presente até os dias de hoje (Russell *et al.*, 2016).

Com o avanço da disponibilidade de grandes bases de dados, no início de 2001 (Russell *et al.*, 2016), além dos avanços significativos em infraestrutura e poder computacional (Faceli *et al.*, 2011) a utilização da IA foi amplamente aceita na indústria.

Uma das áreas de pesquisa da computação que mais tem crescido atualmente é o *machine learning* (Faceli *et al.*, 2011). Sua utilização é aplicável em domínios como ciências, negócios, governos (Lecun, Bengio e Hinton, 2015), que é difícil definir quais são seus campos de atuação (Russell *et al.*, 2016).

Apesar de inúmeros trabalhos e desenvolvimentos na área acadêmica desde 1956, a maior parte das pesquisas se refere a parte heurística em tecnologia da informação (TI) da IA (McCarthy e Hayes, 1969), que é o mecanismo que, com base na informação, resolve o problema e decide o que fazer. Para este trabalho, será dedicado o entendimento sobre a epistemologia em TI ou a representação do mundo de tal forma que a solução dos problemas decorre dos fatos expressos em sua representação.

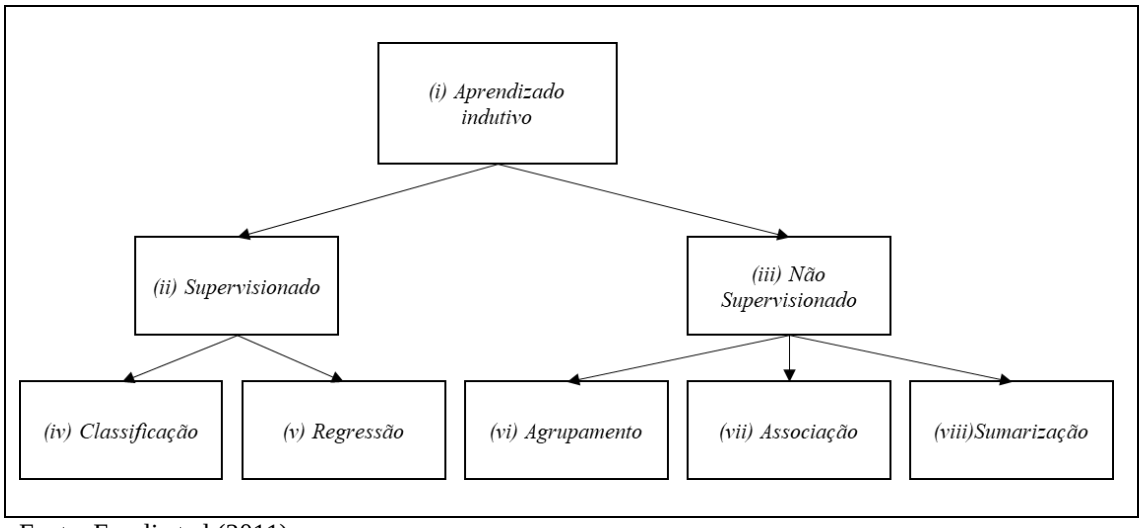
Para o desenvolvimento de algoritmos de IA, é necessário um grande comprometimento com o conhecimento e como ele é obtido, e este é um dos maiores problemas filosóficos da IA (McCarthy e Hayes, 1969).

3. Modelos atuais de AI

Existem vários modelos e técnicas de IA utilizados atualmente para solucionar diferentes tipos de problemas (Faceli *et al.*, 2011; Russell *et al.*, 2016). Na figura 1, Faceli et al.(2011) sumariza a hierarquia de aprendizado de máquina de acordo com a tarefa de aprendizado. Inicialmente apresentando como ponto de partida o (i) *aprendizado indutivo* por ser o processo pelo qual é possível realizar generalizações a partir dos dados. Na sequência, o (ii) *aprendizado supervisionad*o acontece quando há interferência humana no aprendizado, gerando uma resolução preditiva. Já a (iii) *não supervisionada*, não tem interferência humana em sua resolução, apresentando uma solução descritiva (Faceli *et al.*, 2011).

Em um nível inferior estão apresentadas as tarefas de (iv) *classificação*, onde a variável dependente é discreta e (v) *regressão*, onde a variável dependente é contínua. Em outra ramificação, o (vi) agrupamento lida com formação de grupos de acordo com suas similaridades, a (vii) *associação*, que busca encontrar associações de atributos em um conjunto de dados e a (viii*) sumarização*, que busca encontrar uma descrição simples do conjunto de dados (Faceli *et al.*, 2011).

Figura1. Hierarquia de aprendizado de máquina



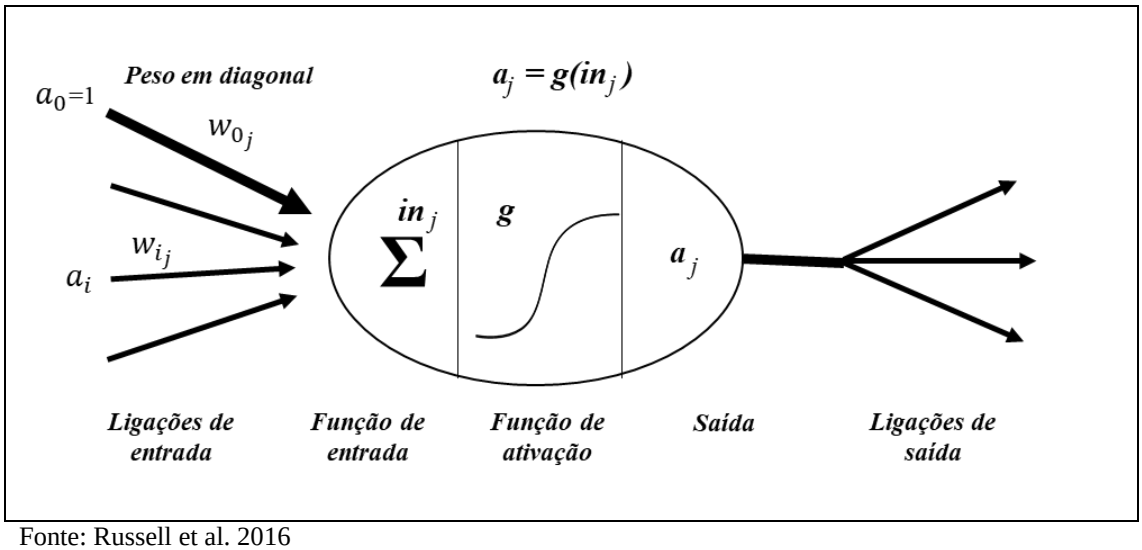
Fonte: Faceli et al.(2011)

Com o advento do exponencial aumento da quantidade de dados gerados por dispositivos de localização, registros de internet, dispositivos inteligentes, entre outros, modelos de IA convencionais se tornaram limitados no processamento dessa informação (Lecun, Bengio e Hinton, 2015) e novos modelos baseados em otimização começaram a ser proeminentes no campo de IA, principalmente em tarefas de classificação, figurando entre as técnicas mais utilizadas em problemas de alta precisão (Faceli *et al.*, 2011). Os métodos baseados na otimização, consistem em encontrar os mínimos e máximos de uma função objetivo de variáveis (Haykin, 1994), ou seja, recorrer à otimização desta função (Faceli *et al.*, 2011).

Uma das técnicas adotadas em IA que recorrem a otimização é a de redes neurais artificiais (RNA). Buscando a evolução da IA, naturalmente se toma como base o cérebro humano(Faceli *et al.*, 2011; Haykin, 1994; Russell, Dewey e Tegmark, 2016). Apesar de existirem algumas pesquisas desde a década de 40 (McCulloch e Pitts, 1943; Minsky e Papert, 1988), a área de RNA começou seu pleno desenvolvimento apenas após os anos 80 com o desenvolvimento de *hardwares* mais rápidos e novas arquiteturas de RNA com maior capacidade de representação e de algoritmos de aprendizado mais sofisticados (Faceli *et al.*, 2011)

As RNAs são sistemas computacionais distribuídos compostos de unidades de processamento densamente interconectados (Faceli *et al.*, 2011). Essas unidades que computam funções matemáticas podem ser chamadas de neurônios artificiais(Faceli *et al.*, 2011), nós ou unidades (Russell *et al.*, 2016).

Figura 2. Neurônio típico



Fonte: Russell et al. 2016

A Figura 2 demonstra o funcionamento de um neurônio típico. Os neurônios são conectados por ligações direcionadas. A ligação de *i* para a unidade *j* serve para a propagar a ativação a_i de *i* para *j*. Onde, cada ligação possui um peso *w*. Assim, a saída expressa em função matemática pode ser apresentada na equação abaixo (Russell *et al.*, 2016).

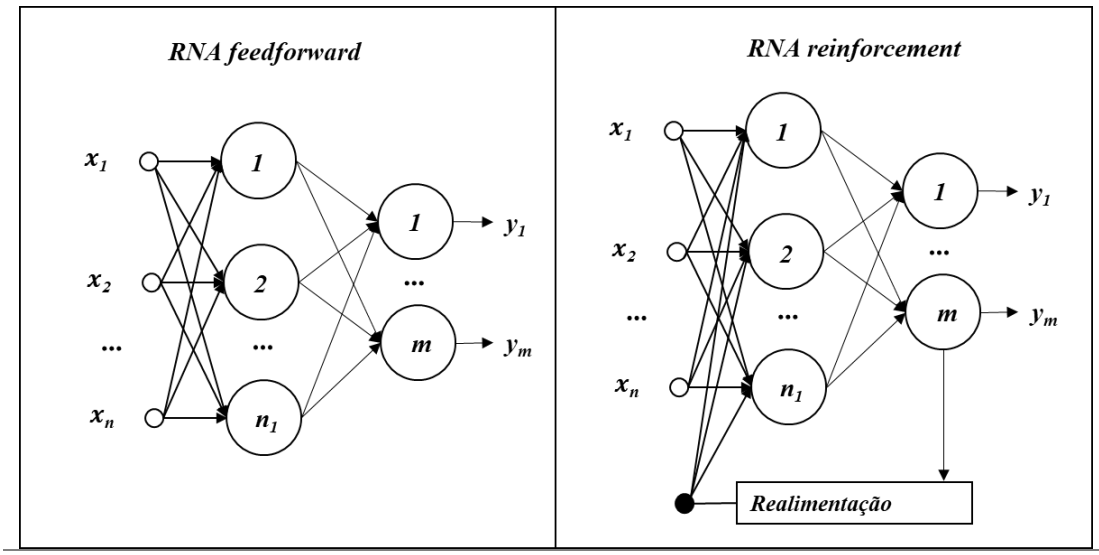
$$a_j = g(in_j) = g(\sum_{i=0}^{n} w_{i,j}\, a_i\ )$$

Estas estão dispostas em camadas com múltiplas conexões com pesos associados, que ponderam a importância de cada neurônio artificial na rede. Esses pesos vão se adequando à medida em que o processo de aprendizado ocorre, buscando a otimização (Faceli *et al.*, 2011)

A formação da "rede" pode ser realizada por duas formas distintas. A primeira é *a rede feedforward*, que pode ser definida como uma rede com conexões que trabalham em apenas uma única direção. Dessa maneira, cada nó recebe a entrada de nós predecessores e passa para nós sucessores, sem a criação de laços. A segunda é a rede *reinforcement*, que alimenta as saídas de volta para as entradas formando um sistema dinâmico (Russell *et al.*, 2016).

As RNA se apresentam como um dos métodos mais utilizados atualmente para robótica e visão computacional (Faceli *et al.*, 2011). Sua precisão e tolerância para erros a tornaram amplamente utilizada. Por outro lado, uma de suas críticas mais constantes está na compreensão de como a RNA toma a decisão. Por utilizar múltiplos hiperparâmetros, a RNA desenvolve fórmulas matemáticas muito complexas para compreensão e comumente são chamadas de "*black-box*" (Faceli *et al.*, 2011).

Figura 3. Exemplo de RNA multicamadas *feedforward* e *reinforcement*



Fonte: Autor

4. Problemas do advento da AI

O crescente desenvolvimento das pesquisas em IA, não foi sucedido por pesquisas igualmente importantes na Filosofia da Ciência sobre o tema (Bostrom e Yudkowsky, 2011). A criação de

algoritmos de RNA profundas tende a acelerar ainda mais este processo (Lecun, Bengio e Hinton, 2015) e, por consequência, esta desagregação.

Pesquisas sobre a Filosofia da Ciência em IA tendem a serem direcionados à ética quanto a criação de um ser que possui consciência de sua existência (Allen, Wallach e Smit, 2006; Bostrom e Yudkowsky, 2011; Russell, Dewey e Tegmark, 2016) ou a capacidade da máquina agir e interagir como um humano(McCarthy e Hayes, 1969).

Um dos grandes trabalhos que apontou problemas da IA foi o de McCarthy e Hayes (1969), que ao analisar o avanço da IA até aquela data, apresentaram o problema do *frame*[1]. O problema surge quando, ao representar o mundo dinâmico, além de explicar todos os objetos de estudo, precisa-se também considerar o resto como *ceteris paribus*, ou seja, outras coisas sendo iguais ou assumindo que nada mais interfere, mas de alguma forma, esteve presente no evento estudado (Kincaid, 2008).

McCarthy e Hayes (1969) elaboraram seu trabalho objetivando, de fato, o ponto de inflexão onde a máquina conseguisse agir e tomar decisões como humanos. Entretanto, o problema do frame também nos permite questionar a IA atual, tanto no que tange *ceteris paribus*, como até mesmo os artefatos da ação causal, ou seja, os próprios agentes que se está analisando.

Ceteris paribus tampouco são incontestáveis na Filosofia. Kincaid (2008), expõe que asserções *ceteris paribus* em ciências sociais são um problema, pois levantam dúvidas sobre as alegações que tentam explicar. O autor alega que, se duvidarmos dos mecanismos *ceteris paribus*, até mesmo teorias econômicas que se utilizam desta premissa podem ter suas explicações colocadas em contestação, e suas previsões, apenas acidentes do acaso[2].

Explicações causais também tangenciam este problema. Dentre as alternativas de abordagem que buscam explicar a "causação", as que utilizam um contrafactual, tidas por alguns pesquisadores como as mais promissoras (Woodward, 2008), possuem uma grande dependência do *ceteris paribus*.

Modelos causais são a base da lógica por trás dos modelos de RNA, pois utilizam um raciocínio semelhante para, através dos dados selecionados, gerar previsões (Haykin, 1994). Particularmente nas RNA, os contextos *ceteris paribus*, podem ser incompreensíveis, posto que não necessariamente todos os parâmetros que deverão ser mantidos constantes para a realização da previsão, são conhecidos.

Dessa maneira é apresentada a primeira problemática da IA, particularmente, das RNAs na filosofia da ciência: ***(i) Ao atribuir explicação causal e previsão em modelos de RNA, como compreender e incorporar os contextos ceteris paribus?***

As RNAs também podem gerar outro problema ainda mais crítico na filosofia da ciência. Até o avanço tecnológico dos anos 90, a utilização de algoritmos de regressão e classificação funcionavam com uma estrutura lógico-matemática acessível para os pesquisadores, onde era possível, de forma trivial, elencar as variáveis importantes para o modelo assim como sua respectiva ponderação.

[1] Do original "frame problem". Algumas traduções também o referenciam como problema do quadro.
[2] No texto original colocado como "lucky accidents". Como "acidentes de sorte" não é um termo muito utilizado na língua portuguesa, foi traduzido como "acidentes do acaso".

Com o retorno das RNAs, em conjunto com o aumento do poder computacional, progressivamente novos algoritmos com um crescente número de variáveis e camadas ainda mais profundas foram surgindo (Lecun, Bengio e Hinton, 2015), gerando modelos muito complexos e incompreensíveis.

Segundo Faceli *et al.(*2011), o conhecimento gerado para a explicação e previsão de RNAs está armazenado em uma grande quantidade de parâmetros dispostas em inacessíveis equações matemáticas. Por isso são comumente referenciados como caixas-pretas.

Este evento repercute na filosofia da ciência, uma vez que além dos contextos *ceteris paribus*, nem mesmo as próprias variáveis sujeitas da ação são bem conhecidas e tampouco sua contribuição individual para o acontecimento do fenômeno. Como resultado, podemos identificar a segunda problemática: ***(ii) Ao utilizar modelos de RNA, como compreender os parâmetros intrínsecos destes contextos de alta complexidade para conseguir atribuir, de forma incontestável, explicação causal e previsão?***

Em Maio de 2018, o parlamento europeu instituiu o *General Data Protection Regulation* (GDPR), e, dentre os vários aspectos desta regulação, estão inclusas cláusulas sobre a tomada de decisão automatizada, que incide nas RNA, onde está presente o direito individual de uma explicação sobre a lógica envolvida na tomada de decisão (Guidotti *et al.*, 2018). Dessa forma, se torna emergente a busca por correspondência destas questões, não apenas em âmbito acadêmico, mas também para fins práticos.

5. Alternativas para a elucidação das problemáticas da AI

Uma forma de análise mais adequada da causação se apresenta como um dos itens mais importantes na agenda de trabalhos futuros sobre explicação causal (Woodward, 2008). Depois de apresentadas, as problemáticas de AI que incidem na filosofia da ciência, e que residem na complexidade de interpretação muitas vezes em tempo real dos algoritmos de RNA, serão apresentadas algumas linhas alternativas de buscas por solução dos problemas.

Um objetivo comum a ambas as perguntas é a busca por maior transparência em suas previsões. Tanto o entendimento do *ceteris paribus* de uma determinada análise, como os pesos e importâncias de variáveis em outras, são determinantes para a validade dos estudos.

Para a primeira problemática, nos modelos de RNAs, as instâncias *ceteris paribus* são tratadas como erros do modelo, ou seja, outros fatos que influenciam o modelo, mas que o algoritmo não consegue interpretar com os dados que possui. Dessa maneira é possível saber que há outras coisas influenciando o fenômeno. Sabemos o quanto a somatórias destas instâncias estão influenciando o modelo, mas não sabemos quais são estas variáveis, nem sua ponderação individual.

Nesta primeira problemática, pouco estudo foi desenvolvido para sua interpretação. Nenhuma alternativa de análise do *ceteris paribus* foi proposta. Uma forma de melhorar esta abordagem surge com o advento de métodos que interpretam dados não estruturados captados por *web crawling*, ou dispositivos que "varrem" a internet buscando dados, mas esta abordagem para este fim ainda é incipiente.

Para a segunda problemática, algumas alternativas se apresentam buscando a interpretação dos dados. O estudo apresentado por Guidotti et al. (2018), realizou uma ampla pesquisa sobre métodos de análise de modelos "caixa-preta". Os autores encontraram três métodos de interpretação. A Extração dos parâmetros por um conjunto de regras, por uma árvore de decisão ou por criação de um protótipo.

No entanto, apesar da extensa análise, as pesquisas sobre a interpretação de dados de RNAs ainda são apresentadas como problemas em aberto (Guidotti *et al.*, 2018; Mahendran e Vedaldi, 2015) e não há um acordo sobre o que é uma explicação (GUIDOTTI *et al.*, 2018).

> É evidente que a atividade de pesquisa neste campo não está fornecendo ainda um nível suficiente de importância no estudo de um formalismo geral e comum para definir uma explicação, identificando quais são as propriedades que uma explicação deve garantir, por exemplo, solidez, integridade, compacidade e compreensibilidade. Em relação a esta última propriedade, não há trabalho que aborde seriamente o problema de quantificar o grau de compreensão de uma explicação para os seres humanos, embora seja de importância fundamental. (GUIDOTTI et al. 2018, p. 36, tradução nossa)

Portanto, a necessidade de análises mais complexas que enderecem estas questões se apresenta de forma evidente, e a filosofia da ciência, de forma proeminente pode ajudar a edificar as bases de compreensão das problemáticas das RNA.

6. Considerações Finais

No presente artigo, foram apresentadas as problemáticas que o advento das RNAs repercute na filosofia da ciência. Inicialmente, expomos o contexto do surgimento da IA e sua evolução ao longo das décadas. Na sequência, focamos em um dos tipos de IA, as redes neurais artificiais, e a consequência de seu desenvolvimento, gerando novos impactos na filosofia da ciência. Por fim, algumas alternativas de solução para os problemas denotados foram oferecidas apesar de seu incipiente desenvolvimento.

Como apresentado no artigo, nenhuma solução foi desenvolvida para a primeira problemática "*(i) Ao atribuir explicação causal e previsão em modelos de RNA, como compreender e incorporar os contextos ceteris paribus?"*. Já para a segunda, *"(ii) Ao utilizar modelos de RNA, como compreender os parâmetros intrínsecos destes contextos de alta complexidade para conseguir atribuir, de forma incontestável, explicação causal e previsão?",* foram apresentadas algumas alternativas de decodificação, mas nenhuma representação definitiva de solução.

Contudo, desenvolver algoritmos de IA, tanto de RNA como outros, requer compromissos sobre o que é conhecimento e como ele é obtido (McCarthy e Hayes, 1969). Esta necessidade

de transparência, não recai apenas em âmbito acadêmico, mas mudanças na regulamentação de proteção de dados (Guidotti *et al.*, 2018), também reforça a necessidade de mais estudo e desenvolvimento neste campo.

Portanto, a área de inteligência artificial se apresenta como um frutífero campo para o desenvolvimento de pesquisas epistemológicas, seja por seu desenvolvimento disruptivo que cria a necessidade de novos olhares sobre temas antes bem definidos, ou com a necessidade de uma abordagem humana que deve definir as diretrizes para o desenvolvimento da inteligência artificial.

7. Referências

ALLEN, C.; WALLACH, W.; SMIT, I. IEEE INTELLIGENT SYSTEMS Why Machine Ethics? **IEEE Intelligent Systems**, n. July/August, p. 12–17, 2006.

BOSTROM, N. Ethical Issues in Advanced AI. **Cognitive, emotive and ethical aspects of decision making in humans and in artificial intelligence**, v. 2, p. 12–17, 2003.

BOSTROM, N.; YUDKOWSKY, E. The ethics of artificial intelligence. **Cambridge University Press**, 2011.

CHARNIAK, E.; MCDERMOTT, D. **Introduction to Artificial Intelligence Programming**. [s.l: s.n.].

FACELI, K. *et al.* Inteligência Artificial: Uma abordagem de aprendizado de máquina. 2011.

GUIDOTTI, R. *et al.* A Survey Of Methods For Explaining Black Box Models. v. 51, n. 5, 2018.

HAYKIN, S. **Neural networks: a comprehensive foundation**. [s.l.] Prentice Hall PTR, 1994.

HORVITZ, E. J.; BREESE, J. S.; HENRION, M. Decision theory in expert systems and {AI}. **Intl. J. Approximate Reasoning**, v. 2, n. 3, p. 247–302, 1988.

HUBERT, T. *et al.* Mastering the game of Go without human knowledge. **Nature**, v. 550, n. 7676, p. 354–359, 2017.

KIETZMANN, J.; PASCHEN, J.; TREEN, E. Artificial intelligence in advertising: How marketers can leverage artificial intelligence along the consumer journey. **Journal of Advertising Research**, v. 58, n. 3, p. 263–267, 2018.

KINKAID; H. Social Sciences. In: SILBERSTEIN, M. **The Blackwell Guide to the Philosophy of Science**, 2008. Chapter 14, p. 290-311

KURZWEIL, R. *et al.* **The age of intelligent machines**. [s.l.] MIT press Cambridge, MA, 1990. v. 579

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, v. 521, n. 7553, p. 436–444, 2015.

MAHENDRAN, A.; VEDALDI, A. **Understanding deep image representations by**

inverting them Proceedings of the IEEE conference on computer vision and pattern recognition. **Anais**...2015Disponível em: <http://www.cv-foundation.org/openaccess/content_cvpr_2015/papers/Mahendran_Understanding_Deep_Image_2015_CVPR_paper.pdf>

MCCARTHY, J.; HAYES, P. J. Some Philosophical Problems from the Standpoint of Artificial Intelligence. *In*: **Readings in Artificial Intelligence**. [s.l.] Elsevier, 1969. p. 431–450.

MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity: The bulletin of mathematical biophysics. **Bulletin of Mathematical Biophysics**, v. 5, n. 4, p. 115–133, 1943.

MINSKY, M. L.; PAPERT, S. A. Perceptrons. 1988.

RUSSELL, S.; DEWEY, D.; TEGMARK, M. Research Priorities for Robust and Beneficial Artificial Intelligence. p. 105–114, 2016.

RUSSELL, S. J. *et al.* **Artificial Intelligence A Modern Approach Third Edition**.

WOODWARD; J. Explanation. In: SILBERSTEIN, M. **The Blackwell Guide to the Philosophy of Science**, 2008. Chapter 3, p. 37-54