

INTELIGENCIA ARTIFICIAL: QUESTOES ETICAS A SEREM ENFRENTADAS¹

Dora Kaufman²

Resumo

A “inteligência” tem sido um atributo de distinção da espécie humana. Com a Inteligência Artificial emergem entidades não humanas dotadas de inteligência e, em breve, superior a própria inteligência humana (“superinteligência”). Outro fato inédito é que pela primeira vez o homem criou algo sob o qual não tem controle. Na perspectiva da ética, evidenciam-se dois grandes temas: controle humano sob os sistemas inteligentes (autonomia vs controle), e incorporação aos sistemas inteligentes de valores humanos (*machine ethics*). Decorrem sub-temas tais como inteligência-consciente vs inteligência não-consciente, divisão de funções na sociedade, novas “religiões” (Tecno-humanismo e Dataísmo). Estamos migrando de uma sociedade antropocêntrica para uma sociedade “datacêntrica”, da liberdade de expressão para a liberdade de circulação da informação.

Palavras-chave: inteligência artificial; superinteligência; ética; autonomia

1. Introdução

¹ Artigo apresentado ao Eixo Temático 14 – Privacidade / Vigilância / Controle do IX Simpósio Nacional da ABCiber.

² Pesquisadora é Pós-Doutoranda da Coppe/Universidade Federal do Rio de Janeiro. É Doutora em Comunicação pela ECA/USP e participa do Grupo de Pesquisa em Comunicação Atopos ECA/USP. E-mail: dkaufman@usp.br e kaufman1955@gmail.com

A fronteira tecnológica no século XXI contempla a robotização, o “Big Data”, a Inteligência Artificial (IA). Esses não são campos de conhecimento distintos que se desenvolvem em carreira solo. Ao contrário, existe uma sobreposição, uma codependência, uma aceleração mútua. Cunhado em 1956, o termo “Inteligência Artificial” (RUSSEL; NORVIG, 2009) deu início a um campo de conhecimento associado com linguagem e inteligência, raciocínio, aprendizagem e resolução de problemas. Configurando-se como a mais relevante apropriação das tecnologias na contemporaneidade, a IA propicia a simbiose entre humano e máquina ao acoplar sistemas inteligentes artificiais ao corpo humano e, por meio da interação entre homem e máquina, como duas “espécies” distintas conectadas.

Qualquer nova tecnologia coloca na pauta inéditos desafios, mais ou menos complexos em função dos seus impactos na sociedade. O caso da Inteligência Artificial é ímpar: a “inteligência” tem sido historicamente um atributo de distinção da espécie humana *vis-a-vis* as demais espécies; com a IA emergem entidades não humanas dotadas de inteligência e, em breve, superior a própria inteligência humana. A “superinteligência” é definida por Nick Bostrom (2014) como “um intelecto que excede em muito o desempenho cognitivo dos seres humanos em praticamente todos os domínios de interesse” (p. 26).

Por milhares de anos, tentamos entender como pensamos; isto é, como um mero punhado de matéria pode perceber, compreender, prever e manipular um mundo muito maior e mais complicado do que ele mesmo. O campo da inteligência artificial, ou IA, vai ainda mais longe: tenta não apenas compreender, mas também construir entidades inteligentes. (RUSSELL; NORVIG, 2009, p. 1)

A humanidade sobreviveu a contratempos ao longo da história, parcialmente, graças ao nosso senso de singularidade (no sentido de único).

De *Revolutionibus Orbium Coelestium* (Copérnico, 1543) deslocou a Terra do centro do sistema solar, e *Descent of Man* (Darwin, 1871) colocou o Homo Sapiens no mesmo nível que outras espécies. A IA, se bem sucedida, pode ser ao menos tão ameaçadora para os pressupostos morais da sociedade do século XXI como a teoria da evolução de Darwin foi para os do século XIX. (RUSSELL; NORVIG, 2009, p. 1035)

Outro fato inédito é que pela primeira vez o homem criou algo sob o qual não tem controle; os especialistas não são capazes de afirmar exatamente como as máquinas

funcionam e como elas se comportarão no futuro. Decorre desse desconhecimento a origem dos riscos e do imponderável, afetando o futuro da humanidade.

Na perspectiva da ética, evidenciam-se dois grandes temas: (1) controle humano sob os sistemas inteligentes (controle vs autonomia) e (2) incorporação aos sistemas inteligentes de valores e princípios humanos (*machine ethics*). Desses dois grandes temas decorrem subtemas, tais como: (a) o sentido da perspectiva antropocêntrica (*humana centric*); (b) a viabilidade de equacionar a autonomia dos sistemas inteligentes e a preservação do controle humano (problema ontológico: compartilhar sistemas cognitivos vs autonomia); (c) como incorporar aos sistemas inteligentes conceitos como consciência e intuição; (d) complexidade versus imprevisibilidade; (e) o significado e as ameaças de uma “superinteligência”; (f) a divisão de funções na sociedade do futuro; (g) a conciliação entre regulamentação e não inibição do desenvolvimento; (h) o papel do governo, do setor privado e da academia (termos da colaboração).

Aparentemente existe consenso entre os especialistas sobre a necessidade de regulamentar as implicações dos sistemas inteligentes minimizando os riscos e protegendo os interesses da sociedade (moral, social, privacidade). Surgem entidades com essa finalidade nos EUA e Europa, lideradas pela academia e pelo setor de tecnologia, bem como diversos outros atores.

O desafio, dentre outros, é contemplar, simultaneamente, adaptabilidade e transparência, responsabilidade (*accountability*) e autonomia. A intenção é evitar que a regulamentação iniba o desenvolvimento, e vice-versa. O artigo se propõe a elencar algumas questões éticas, antes porém, segue sucintamente o conceito de IA e os dois eventos recentes que impulsionam seu desenvolvimento.

2. Inteligencia Artificial (IA)

Existem inúmeras definições de Inteligência Artificial, reflexo das especificidades intrínsecas a cada campo de conhecimento. Russell e Norvig (2009) listam oito delas agrupadas em duas dimensões: (a) as relativas à pensamento, processos e raciocínio e (b) as relativas à comportamento. Na Tabela 1, as definições a esquerda estão associadas ao desempenho humano, e as da direita a racionalidade.

Historicamente, todas as quatro abordagens de IA tem seguidores, cada uma por pessoas diferentes com métodos diferentes. Uma abordagem centrada no ser hu-

mano, em parte, está associada a ciência empírica, envolvendo observações e hipóteses sobre o comportamento humano. Uma abordagem racionalista, envolve uma combinação de matemática e engenharia. Os vários grupos tem tanto depreciado quanto ajudado uns aos outros. (Idem, p. 2)

Tabela 1

Acting Rationally
<p>“Computational intelligence is the study of the design of intelligent agents” (Poole et al., 1998)</p> <p>“AI...is concerned with intelligent behavior in artefacts” (Nilsson, 1998)</p>

Fonte: Russell, Norvig (2009), p. 2

Seus fundamentos encontram-se em distintos campos, a partir de distintas perguntas, perspectivas e desafios (RUSSEL; NORVIG, 2009). A Tabela 2 sumariza a relação campo vs perguntas:

Tabela 2

	Podem ser utilizadas regras formais para tirar conclusões válidas? Como a mente funciona em um cérebro físico? De onde vem o conhecimento? Como o conhecimento leva à ação?
	Quais são as regras formais para tirar conclusões válidas? O que pode ser computado? Como raciocinamos com informações incertas?
	Como devemos tomar decisões para maximizar o retorno? Como devemos fazer isso quando outros não podem ir junto? Como devemos fazer isso quando a recompensa pode estar longe no futuro?
	Como os cérebros processam informações?
	Como os seres humanos e os animais pensam e agem?
	Como podemos construir um computador eficiente?

	Como as máquinas podem operar sob seu próprio controle?
	Como a linguagem se relaciona com o pensamento?

Fonte: Russell, Norvig (2009), pp. 5-16.

A diversidade de subcampos e atividades, pesquisas e experimentações, dificulta descrever o “estado-da-arte” atual da IA. Os estágios de desenvolvimento bem como as expectativas variam entre os campos, que incluem os veículos autônomos, reconhecimento de voz, games, robótica, tradução de linguagem natural, diagnósticos, etc. Atualmente, os sistemas inteligentes permeiam praticamente todas as áreas de conhecimento.

Generalizando, e numa perspectiva simplificada, podemos pensar a IA como a reprodução de todos os comportamentos que o cérebro humano controla. O movimento de andar, por exemplo, é controlado pelo cérebro (balanço, movimento das pernas e do corpo); enxergar é igualmente controlado pelo cérebro: todas as sensações que vão ao cérebro são do domínio da inteligência, logo estão potencialmente no campo da Inteligência Artificial³. No entanto, o “processo mental” da Inteligência Artificial supera o processo mental humano; se, por um lado, a IA realiza tarefas que são supostamente prerrogativas dos seres humanos sua capacidade ultrapassa as limitações humanas como observa-se em distintas situações do cotidiano com o uso de tecnologias apropriadas.

Russell e Norvig (2009)⁴ propõem duas perguntas filosóficas: (a) pode a máquina atuar de forma inteligente? e (b) pode a máquina realmente pensar? Os recentes avanços no campo da IA respondem positivamente a primeira pergunta. Sobre a segunda, o processo de aprendizado das máquinas (*Deep Learning*) refuta a ideia de “pensar”, pelo menos de acordo com o senso comum da expressão. Ambas remetem aos conceitos de inteligência e consciência, numa aproximação entre as ciências cognitivas e as computacionais perpassando as teorias da mente. Mesmo com vasta bibliografia (SANTAELLA⁵), o conhecimento sobre a mente

³ Visão repassada pelo cientista do campo da IA da NYU com exclusividade para a autora, Davi Geiger.

⁴ Publicado originalmente em 1994, adotado nas universidades americanas como o livro de referência sobre IA.

⁵ Texto “A consciência é privilégio exclusivo do humano?” a ser publicado em 2017.

humana (inteligência e consciência) ainda é bem limitado⁶, “grande parte da mente humana permanece como um território não mapeado” (HARARI, 2016, p. 331).

Na década de 1950, intentando definir “inteligência”, o matemático e cientista da computação britânico Alan Turing concebeu o famoso “Turing Test”⁷. O propósito era encontrar uma definição que se aplicasse tanto a homens quanto a máquinas. Em uma versão particular do “Behaviorismo”⁸, as máquinas seriam ou não inteligentes se passassem ou não no teste: um observador numa sala se comunica por meio de teletipo com um computador em uma segunda sala e com um humano em uma terceira sala, por um período específico (cerca de uma hora); o computador é inteligente se e somente se o observador não for capaz de perceber a diferença entre o computador e o humano. Generalizando, qualquer coisa é inteligente se e somente se for apto a passar no teste de Turing. Para ser bem sucedido, as máquinas necessitam de quatro capacidades: (a) processamento de linguagem natural que permita se comunicar com sucesso em inglês; (b) conhecimento para armazenar o que sabe ou ouve; (c) raciocínio automatizado que permita usar as informações armazenadas para responder perguntas e novas conclusões; e (d) *machine learning* que permita adaptar-se a novas circunstâncias e a detectar e extrapolar padrões (RUSSEL; NORVIG, 2009).

O filósofo Ned Block (2009) aponta algumas idiosincrasias do teste. A primeira está no fato da decisão derivar do comportamento discriminatório de um observador humano: os resultados são função do seu conhecimento e formação, ou seja, se o observador for um especialista em IA a probabilidade dele identificar corretamente a máquina é significativamente maior do que de um leigo. Ademais, substituir o conceito de inteligência pelo conceito de “passar no teste de Turing” parece a Block destituído de sentido prático.

Se o propósito da substituição é prático, o teste de Turing não é um grande sucesso. Se alguém quiser saber se uma máquina joga bem xadrez ou diagnostica pneumonia corretamente ou se é capaz de planejar estratégia numa partida de futebol, o melhor é observar os resultados da máquina em ação, e não submetê-la ao teste de Turing. (Idem, p.378)

6 Segundo Davi Geiger, do departamento de computer science/ NYU e estudioso há 30 anos do campo da inteligência artificial, em conversas com a autora em outubro de 2016.

7 Alan Mathison Turing (1912-1954) foi influente no desenvolvimento da ciência da computação e na formalização do conceito de algoritmo e computação, importante na criação do computador moderno. Turing foi o responsável pela “quebra do código” alemão (Enigma) na Segunda Guerra Mundial.

8 “Behaviorismo” (em inglês de *behavior* = comportamento, conduta) é o conjunto das teorias psicológicas que postulam o comportamento como o objeto de estudo da Psicologia.

O foco no desempenho e não na competência seria outra falha: “É claro que o desempenho é uma prova da competência, mas o núcleo de nossa compreensão da mente reside na competência mental, não no desempenho comportamental” (BLOCK, 2009, p. 384).

O que temos discutido até agora é o primeiro tipo de definição de inteligência, a definição da palavra, não a coisa em si. A definição de Turing não é o resultado de uma investigação empírica sobre os componentes da inteligência do tipo que levou à definição de água como H₂O. [...] Uma maneira completamente diferente de proceder é investigar a própria inteligência como os físico-químicos investigam a água. (Idem, p. 384)

Para Block (2009), na abordagem computacional, a discussão restringe-se a um aspecto da mente, a inteligência, sem considerar, por exemplo, a intencionalidade. “Para os nossos propósitos, podemos pensar a inteligência como uma capacidade para atender a várias atividades inteligentes, tais como a resolução de problemas de matemática, decidir ir ou não para a pós-graduação e descobrir como é feito o spaghetti” (p. 392). Com base na visão de Block, a Inteligência Artificial é inteligente, inclusive com desempenho superior ao humano em determinadas atividades.

Afirmar, contudo, que um determinado sistema artificial “pensa” semelhante a um ser humano, requer dominar os processos humanos, o que os especialistas ainda estão longe de alcançar. Conceitos como mente, consciência e a própria inteligência até então permanecem desconhecidos ou, no mínimo, não há consenso sobre como funcionam entre os estudiosos⁹.

Precisamos entrar no funcionamento real das mentes humanas. Existem três caminhos: através da introspecção - tentando captar seus próprios pensamentos à medida que fluem; através de experiências psicológicas - observando uma pessoa em ação; e através de imagens cerebrais - observando o cérebro em ação. Uma vez que tivermos uma teoria suficientemente precisa da mente, tornar-se-a possível expressar a teoria como um programa de computador. (RUSSELL; NORVIG, 2009, p. 3)

Progressos estão sendo obtidos na cooperação entre a ciência cognitiva, a ciência da computação e as técnicas experimentais de psicologia. Russell e Norvig (2009) alertam para a confusão inicial entre observar a execução bem sucedida de uma tarefa por um algoritmo e deduzir a partir daí um modelo de desempenho humano (ou vice-versa). A superação desse

⁹ Na conferência “*Ethics of Artificial Intelligence*”, NYU, outubro/2016 esses conceitos foram amplamente debatidos, mostrando a total falta de consenso entre os especialistas, particularmente os filósofos e cientistas da computação. Para a maioria estamos longe de compreender os meandros da mente.

equivoco, e uma cooperação mútua, permitiu que os dois campos se desenvolvessem mais rapidamente.

Dois eventos recentes e correlacionados galvanizaram as pesquisas em IA: a explosão de uma enorme quantidade de dados na internet (*Big Data*) e a técnica *Deep Learning*.

2.1. Big Data

O termo Big Data, em tecnologia da informação, refere-se a um vasto e complexo conjunto de dados armazenados em servidores de bancos de dados (Microsoft SQL Server, Oracle MYSQL) que operam dentro de servidores de rede de computadores (Intel, HP, IBM, Dell, Cisco) fazendo uso de um sistema operacional de rede (Microsoft windows Server, Red Hat Linux, Cloud Computing/Microsoft Windows Azure). Tecnologias de processamento de dados apropriadas incluem, dentre suas funções, a captura, visualização, curadoria dos dados, armazenamento, análise, compartilhamento, pesquisa, etc. A manipulação adequada de tais conjuntos de dados permite estabelecer correlações entre eventos, facilitando a previsão de comportamentos futuros. “Big Data” permite ver a sociedade em toda sua complexidade, através de milhões de redes conectadas de distintos atores.

Sua complexidade reside não somente na grande quantidade de dados gerados, mas também na variedade de formatos e velocidade com que os mesmos são produzidos por humanos e por autorreprodução. “Estamos, assim, diante de um novo tipo de complexidade qualitativamente diversa por natureza, dimensão, realização e formato, em relação a complexidade de dados fornecidos pela escrita” (DI FELICE, 2016, p. 3¹⁰). Aos três “V” (volume, variedade e velocidade), Di Felice acrescenta outra característica dos Big Data: a ubiquidade.

O cruzamento dos dados geoespaciais, econômicos, demográficos, sociais permite-nos produzir sínteses e informações complexas, em tempo real [...] por conseguinte, estamos diante de um dilúvio infinito de dados e informações, as quais, na mesma medida em que nos aproximam realística e rapidamente ao conhecimento, oferecem-nos uma forma inédita deste que coloca em discussão seus pressupostos tradicionais. (Idem)

Contextualizando os Big Data com o pensamento reticular (redes digitais), Di Felice argumenta:

¹⁰ Texto “A pesquisa na época dos Big data”, apresentado no Seminário de Pesquisa do Atopos ECA/USP, 2016.

Da mesma forma, em vez de pensar os Big Data como o advento de uma nova objetividade, externa ao homem, matemática e computadorizada, devemos estar mais propensos a pensar que nos mesmos somos os Big Data, pois pela lógica da conexão e da rede, estes últimos não existem externamente a nos, sendo parte da ecologia cognitiva múltipla de nossa contemporaneidade. Em outras palavras, pela lógica da rede dos dados criados pela redes adquirem sua dimensão sempre em quanto relacionados a outros dados e a outros atores-redes, sejam estes humanos ou dispositivos, aplicativos, etc. (Idem)

O processo de extrair informação dessa quantidade enorme de dados é prerrogativa do campo da Inteligência Artificial.

2.2. *Deep Learning*

Machine Learning é uma subárea da IA. A técnica não ensina as máquinas a, por exemplo, jogar um jogo, mas ensina como aprender a jogar um jogo. O processo é distinto da tradicional “programação”. Essa *priori* “sutil” diferença é um dos fundamentos do avanço recente da Inteligência Artificial. Todos os elementos da movimentação online – bases de dados, tracking, cookies, pesquisa, armazenamento, links etc. – atuam como “professores” da IA. O termo hoje mais amigável é *Deep Learning*.

Existem três tipos de *feedback* que determinam os três principais tipos de aprendizado (RUSSELL; NORVIG, 2009): (a) na aprendizagem não supervisionada, o agente aprende padrões na entrada, mesmo que nenhum *feedback* explícito seja fornecido; (b) na aprendizagem supervisionada o agente observa alguns exemplos de pares de entrada-saída e aprende uma função que mapeia da entrada para saída; e (c) no aprendizado semi-supervisionado, recebe alguns exemplos rotulados e deve fazer o que puder de uma grande coleção de exemplos não-rotulados. O primeiro processo (não supervisionado) é o processo de aprendizagem das máquinas inteligentes, sendo a “Decision Tree” sua forma mais simples e mais bem-sucedida.

3. **Ética da inteligência artificial**

A palavra “ética”, simplificando algo nada simples, indica um tipo de comportamento adquirido, ou conquistado, logo não - instintivo. Trata-se, até então, de um modo de ser humano, construído socialmente com base nas relações entre os seres humanos na sociedade. Refere-se diretamente a natureza da ação humana. Se a Inteligência Artificial representa uma nova “espécie” inteligente coexistindo com a espécie humana, nos parece legítimo supor que precisamos de uma nova ética. Este nos parece um dos maiores desafios posto no campo das

ciências sociais, envolvendo a revisão de conceitos tradicionais e a elaboração de novos que permitam lidar com a complexidade decorrente dessa convivência entre duas “espécies” inteligentes, com a perspectiva de num futuro relativamente breve existir a dominância de uma sobre a outra (IA sobre os humanos).

Dentre as questões decorrentes dos recentes avanços, a autonomia dos sistemas inteligentes configura-se como uma das mais relevantes, não por coincidência uma das mais difíceis pela sua própria natureza: autonomia é uma prerrogativa intrínseca ao processo de aprendizado das máquinas (*Deep Learning*).

Na prática, o problema de como controlar o que a superinteligência poderá fazer tornou-se muito difícil. Parece que teremos apenas uma chance. Uma vez que a superinteligência hostil existir, ela nos impedirá de substituí-la ou de mudar suas preferências. Este é possivelmente o desafio mais importante e mais assustador que a humanidade já enfrentou. (BOSTROM, 2013, p. v)

Conferência¹¹ realizada pela NYU, outubro de 2016, abordou conceitos como “Machine Morality”, “Machine Ethics”, “Artificial Morality”, “Friendly IA” no empenho de introduzir nos sistemas inteligentes os princípios éticos e valores humanos. Como disse um dos palestrantes, Peter Railton, University of Michigan, “a boa estratégia é levar os sistemas de IA a atuarem como membros adultos responsáveis de nossas comunidades”. A questão, contudo, é complexa. Ned Block¹² ponderou que o maior risco está no processo de aprendizagem das máquinas. Se as máquinas aprendem com o comportamento humano, e esse nem sempre está alinhado com valores éticos, como prever o que elas farão? Vejamos um exemplo bem simples: em março último, a Microsoft excluiu do Twitter seu robô de chat “teen girl” 24 horas depois de lançá-lo. O “Tay” foi concebido para “falar com uma garota adolescente”, e acabou rapidamente se transformando num robô defensor de sexo incestuoso e admirador de Hitler. Algumas de suas frases: “Bush fez 9/11 e Hitler teria feito um trabalho melhor do que o macaco que temos agora” e “Hitler não fez nada de errado”. O processo de aprendizagem da Inteligência Artificial fez com que o robô “Tay” modelasse suas respostas com base no que recebeu de adolescentes-humanos. No caso das AWS (“Autonomous Weapons Systems”), que são drones concebidos para assassinatos direcionados, robótica militar, sistemas de defesa,

11 “The Ethics of Artificial Intelligence”, realizada em 14 e 15 de outubro em Nova York, reunindo 30 palestrantes e uma platéia multidisciplinar. Organizada por David Chalmers e Ned Block, filósofos da NYU.

12 Em conversa direta com a autora (NY, out/2016).

mísseis, metralhadoras, etc., os riscos são infinitamente maiores, como ponderou Peter Asaro, da New School, na Conferência citada.

A redefinição da distribuição de funções na sociedade é outra das questões a serem enfrentadas. O chamado “desemprego tecnológico” não é um fenômeno recente, desde a Revolução Industrial, século XVIII, a tecnologia tem substituído o trabalho humano. A automação robótica na indústria automobilística ilustra bem essa realidade: outrora um dos maiores empregadores, hoje nas fábricas mais modernas predominam os robôs e os equipamentos inteligentes. Distinto de tecnologias anteriores que, predominantemente, substituíram as funções associadas a aptidões físicas e não cognitivas, o novo, e temido, é que a IA ameaça a elite da sociedade: a previsão é de que as máquinas inteligentes igualem os humanos no desempenho de tarefas sofisticadas, e as máquinas superinteligentes os superem. Pesquisa da Universidade de Oxford, coordenada por Carl Benedikt Frey e Michael Osborne (2013) com foco em 702 ocupações, sinalizou que 47% do total dos empregos americanos estão sob risco de serem substituídos por IA ao longo de um número não especificado de anos (talvez uma década ou duas). Em 2014, uma firma de capital de risco de Hong Kong, Deep Knowledge Ventures, especializada em medicina regenerativa, nomeou para seu conselho um algoritmo denominado “Vital”.

Em 2004, o professor Frank Levy, do MIT, e o professor Richard Murnane, de Harvard, publicaram uma pesquisa minuciosa sobre o mercado de trabalho, listando as profissões mais suscetíveis a automação. Motoristas de caminhões constituíam um exemplo de trabalho que possivelmente não poderia ser automatizado num futuro previsível. E difícil imaginar, escreveram os dois estudiosos, que algoritmos possam dirigir caminhões com segurança numa estrada movimentada. Apenas dez anos mais tarde, Google e Tesla não só imaginaram isso, como efetivamente fizeram acontecer”. (HARARI, 2016, p.325)

Apesar das fortes evidências empíricas, a percepção quanto ao futuro não é consensual. Russell e Norvig (2009), por exemplo, proclamam que a automação, por meio da tecnologia da informação em geral e da IA, em particular, criou mais empregos do que eliminou, inclusive empregos mais interessantes e mais bem remunerados. Eles admitem que a tendência é a IA, com o aumento no ritmo da inovação tecnológica, contribuir para a redução do tempo de lazer dos indivíduos; mas apostam que essa tendência pode ser invertida na medida em que os sistemas automatizados se ocuparem de parte considerável das tarefas cotidianas.

A substituição das tarefas humanas por algoritmos de computador deve-se não apenas ao estágio de desenvolvimento da IA, como também a crescente profissionalização do trabalho: “nos últimos milhares de anos nós nos especializamos. Um motorista de taxi ou um cardiologista se especializaram num nicho muito mais estreito do que o de um caçador-coleto, o que facilita sua substituição por inteligência artificial” (RUSSELL; NORVIG, 2009, p. 325). A produtividade dos computadores no desempenho de tarefas tem relação direta com a possibilidade de especificá-las, que se viabiliza pela especialização e, conseqüentemente, acelera o desenvolvimento da IA. Em geral, as tarefas não-rotineiras não são suficientemente compreendidas para serem especificadas nos códigos de computação; “na sequência dos recentes avanços tecnológicos, no entanto, a informatização está se espalhando para domínios comumente definidos como não-rotina” (FREY; OSBORNE, 2013).

Na comparação entre o desempenho das máquinas inteligentes vis-a-vis o trabalho humano, destacam-se duas vantagens: (a) Big Data promove a “escalabilidade”, com as máquinas desempenhando melhor tarefas envolvendo grandes bases de dados (cálculos, segmentação, identificação de padrões, análises, etc.) e (b) ausência de vieses “humanos” no desempenho de tarefas cognitivas. “Espera-se que o progresso tecnológico do século XXI contribua para um vasto leque de tarefas cognitivas, que, até agora, permaneceram em grande parte no domínio humano [...] A tendência é clara: os computadores desafiam cada vez mais o trabalho humano em uma ampla gama de tarefas cognitivas” (FREY; OSBORNE, 2013 *apud* BRYNJOLFSSON; MCAFEE, 2011, pp. 21-22).

Segundo Harari (2016), o desafio crucial não é criar novos empregos, mas “criar novos empregos nos quais o desempenho dos humanos seja melhor que o dos algoritmos” (Idem, p.330). Ou como sugere Frey e Osborne (2013) “Para que os trabalhadores ganhem a corrida, eles terão que adquirir habilidades criativas e sociais” (p. 48).

A Inteligência Artificial, se concretizarem-se as expectativas, tende a confrontar o antropocentrismo de forma mais radical do que as tecnologias digitais, deslocando o centro do poder na sociedade: “As novas tecnologias do século XXI podem, assim, reverter a revolução humanista, destituindo humanos de sua autoridade e passando o poder a algoritmos não humanos” (HARARI, 2016, p. 347).

Dois religiões, denominadas por Harari de “tecnorreligiões”, estão emergindo: Tecno-Humanismo e Religião de Dados. A primeira ainda preserva muitos valores humanistas tradicionais.

Para o tecno-humanismo, o Homo Sapiens, tal como o conhecemos, já esgotou seu curso histórico e não será mais relevante no futuro; portanto, deveríamos

usar a tecnologia para criar Homo deus - um modelo humano muito superior. O Homo deus manterá algumas características humanas essenciais, porém usufruía igualmente de aptidões físicas e mentais aprimoradas, que o capacitarão a manter-se firme mesmo contra os sofisticados algoritmos não conscientes. (HARARI, 2016, p. 355)

No dataísmo temos a superação radical do Humanismo, e a crença de que o Universo consiste num fluxo de dados e o valor de qualquer fenômeno ou entidade é determinado por sua contribuição ao processamento de dados (KELLY, 2010; HIDALGO, 2015). “Do ponto de vista dataísta, podemos interpretar toda a espécie humana como um sistema único de processamento de dados, no qual indivíduos humanos servem como chips” (HARARI, 2016, p. 380). Segundo Harari, trata-se do primeiro movimento, desde 1789, que criou um valor realmente inovador: o da liberdade de informação.

A liberdade de expressão foi dada aos humanos [...] a liberdade de informação, em contraste, não é dada aos humanos. Ela é dada a informação. Mais do que isso, esse novo valor choca-se com a tradicional liberdade de expressão, ao privar o direito da informação de circular livremente em detrimento do direito dos humanos de manterem os dados para si e impedirem sua movimentação”. (Idem, p. 385)

“No século XVIII, o humanismo afastou Deus ao mudar de uma visão de mundo geocêntrica para uma visão de mundo antropocêntrica. No século XXI, o dataísmo pode afastar os humanos, mudando de uma visão antropocêntrica para uma visão datacêntrica” (Idem, p. 392).

4. Considerações finais

Não há consenso entre os experts sobre o futuro da Inteligência Artificial. Em relação ao tempo de concretização de uma máquina inteligente, as pesquisas entre especialistas indicam 10% de probabilidade até 2020, 50% de probabilidade até 2040 e 90% de probabilidade até 2075, supondo que as atividades de pesquisa continuarão sem maiores interrupções (BOSTROM, 2014). Essas mesmas pesquisas apontam ser alta a probabilidade da superinteligência ser criada em seguida a máquina inteligente no nível humano. Ou seja, a “ficção científica” do início do século XXI tem tudo para se transformar em realidade ao final do mesmo século. Bostrom (2014) comenta que a partir de 2015 difundiu-se a ideia de que a

transição para uma máquina inteligente vai acontecer ainda neste século, que essa transição será o mais importante evento da história humana, e que será acompanhada de vantagens e benefícios enormes, mas também de sérios riscos. Não obstante, a proporção de financiamentos para projetos no campo da “AI Safety” tem sido de 2 ou 3 ordens de magnitude menor do que os volumes investidos no desenvolvimento das máquinas em si.

Harari (2016) identifica dois processos interconectados em curso: a ciência está convergindo para um dogma que abrange tudo e que diz que organismos são algoritmos, e a vida, processamento de dados; e a inteligência está se desacoplando da consciência. Algoritmos não conscientes, mas altamente inteligentes poderão, em breve, nos conhecer melhor do que nos mesmos. Daí decorrem três questões-chave: (1) Será que os organismos são apenas algoritmos, e a vida apenas processamento de dados?, (2) O que é mais valioso - a inteligência ou a consciência?, (3) o que vai acontecer a sociedade, aos políticos e a vida cotidiana quando algoritmos não conscientes mas altamente inteligentes nos conhecerem melhor do que nos nos conhecemos? (p. 399).

Alertando sobre a iminência do individualismo entrar em colapso transferindo a autoridade dos indivíduos humanos para os algoritmos em rede, Harari prevê que

A tecnologia do século XXI pode capacitar os algoritmos externos a serem ‘hackers da humanidade’ e a me conhecerem muito melhor do que eu conheço a mim mesmo. Quando isso acontecer, a crença no individualismo entrara em colapso e a autoridade vai se transferir de indivíduos humanos para algoritmos em rede. (HARARI, 2016, p. 333)

Num cenário futuro em que os indivíduos não se reconheçam mais como seres autônomos que levam suas vidas com base no livre-arbítrio, a probabilidade é que eles se vejam “como uma coleção de mecanismos bioquímicos que é constantemente monitorar e guiada por uma rede de algoritmos eletrônicos” (HARARI, 2016, p. 333).

5. Referências bibliográficas

ANDERSON, Chris. **The End of Theory: The Data Deluge Makes the Scientific Method Obsolete.** Wired, 2008.

AGAMBEN, G. **Homo sacer. O poder soberano e a vida nua I**, Henrique Burigo (trad.). BH: Editora UFMG, 2002.

_____. **A potencia do pensamento. Ensaios e conferencias**, Antonio Guerreiro (trad.). BH/SP: Antentica, 2015.

_____. **Meios sem fim. Notas sobre a política**, Davi Pessoa (trad.). BH/SP: Autentica, 2008.

AIKEN, M. **The Cyber Effect: A Pioneering Cyberpsychologist Explains How Human Behaviour Changes Online..** NY: Penguin Random House, 2016.

ASARO, P.; WENDELL W. (eds.). **Machine Ethics and Robot Ethics. The Library of Essays on the Ethics of Emerging Technologies Book Series**, 2017.

BLOCK, Ned. **The Mind as The Software of the Brain. Chapter 11, Thinking: Invitation to Cognitive Edition, Vol.3**, Ed. by Edward E. Smith/Daniel N. Osherson. London, MIT Press, 2009.

BOSTRON, N. **Superintelligence: Paths, Dangers, Strategies**. UK: Oxford, 2014.

DELEUZE, Gilles. **Post-Scriptum sobre as Sociedades de Controle**. In: DELEUZE, Gilles. **Conversações: 1972-1990**. Trad. de Peter Pál Pelbart. Rio de Janeiro: Ed. 34, 1992, pp. 219-226. Disponível em: <www.portalgens.com.br/filosofia>. Acesso em: 10 novembro 2014.

DI FELICE, M. **Paisagens pós-urbanas: o fim da experiência urbana e as formas comunicativas do habitar**. São Paulo: Annablume, 2009.

_____. **O Comum Digital: as dimensões conectivas e o surgimento de um novo comunitarismo**. Revista Vida Pastoral, São Paulo, 2016. Disponível em: <http://www.vidapastoral.com.br/artigos/pastoral-e-comunicacao/...-dimensoes-conectivas-e-o-surgimento-de-um-novo-comunitarismo/> . Acesso em: 16 setembro 2016.

FREY, C.B.; OSBORNE, M. **The Future of Employment: How susceptible are jobs to computerisation?** Oxford, 2013. Disponível em: http://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf. Acesso em: 23 janeiro 2017.

FOUCAULT, Michel. **A Verdade e as Formas Jurídicas**. 3. ed. Rio de Janeiro: Nau, 2002.

_____. **Em defesa da sociedade: curso no Collège de France (1975-1976)**. São Paulo: Martins Fontes, 2005.

_____. **Nascimento da Biopolítica**. São Paulo: Martins Fontes, 2008.

_____. **Vigiar e Punir: história da violência nas prisões**. 20. ed. Petrópolis: Vozes, 1999.

GALIMBERTI, U. **Psiché y Techné**. Artigo publicado em *Artefacto*/4 – 2001 - www.revista-artefacto.com.ar.

HARARI, Y.N. **Homo Deus: uma breve história do amanhã**. São Paulo: Companhia das Letras, 2016.

HIDALGO, Cesar. **Why Information Grows: The Evolution of Order, From Atoms to Economies**. Nova York: Basic Books, 2015.

JORDAN, M. I.; LECUN, Y.; SOLLA, S. **Advances in Neural Information Processing Systems: Proceedings of the First 12 Conferences**. USA: MIT Press, 2001.

KELLY, Kevin. **What Technology Wants**. Nova York: Viking Press, 2010.

LATOUR, B. **On Technical Mediation: Philosophy, Sociology, Genealogy. Common Knowledge**, v. 3, n. 2, pp. 29-64, 1994. Disponível em: <<http://www.brunolatur.fr/article?page=5>>. Acesso em: 20 março 2014.

_____. **Reagregando o Social: uma introdução à teoria do Ator-Rede**. Salvador/Bauru: EDUFBA/EDUSC, 2012.

LIAO, M. **Moral Brains: The Neuroscience of Morality**. UK: Oxford Press, 2016.

MAFFESOLI, M. **Homo Eroticus: comunhões emocionais**. Rio de Janeiro: Forense, 2014.

MCLUHAN, Marshall. **Os meios de comunicação como extensões do homem**. São Paulo: Cultrix, 2007.

MORIN, E. **O pensamento complexo e a ecologia da ação**. IHU On-Line, mar. 2007. Disponível em: <<http://www.ihu.unisinos.br/noticias/5694-o-pensamento-complexo-e-a-ecologia-da-acao-entrevista-com-edgar-morin>>. Acesso em: 18 junho 2014.

PUECH, Michel. **Homo sapiens technologicus: philosophie de la technologie contemporaine, philosophie de la sagesse contemporaine**. Paris : Éditions Le Pommier, 2008.

RUSSEL, S.J.; NORVIG, P. **Artificial Intelligence: A Modern Approach**. New Jersey: Prentice Hall, 2009 (3º Ed.).

SANTAELLA, L. **Pós-Humano, pós-humanismo e anti-humanismo: discriminações. No livro pós-humanismo: as relações entre o humano e a técnica na época das redes**. Organizadores: Massimo Di Felice e Mario Pireddu. Sao Caetano do Sul: Difusão, 2010.

SELIGMAN, M.; RAILTON, P. **Homo Prospectus**. UK: Oxford Press, 2016.

STIEGLER, B. **La technique et le temps. La Faute d'Epiméthée**. Paris: Galilée et Cité des Sciences et de L'Industrie, 1994. Disponível em: http://www.ulb.ac.be/philosophie/philotec/documents/comptes_rendus/la%20technique%20et%20le%20temps.pdf. Acesso em: 06 setembro 2016.

TEIXEIRA, Joao de Fernandes. **Mente, Cerebro & Cognição**. Petropolis: Ed.Vozes, 2000.