

**UNIVERSIDADE ESTADUAL DO PIAUÍ
CAMPUS PROFESSOR ALEXANDRE ALVES DE OLIVEIRA
CURSO DE LICENCIATURA PLENA EM FILOSOFIA**

FERNANDO MACHADO VIEIRA

**INTELIGÊNCIA ARTIFICIAL E OS DESAFIOS DA SINGULARIDADE
TECNOLÓGICA**

**PARNAÍBA-PI
2017**

FERNANDO MACHADO VIEIRA

**INTELIGÊNCIA ARTIFICIAL E OS DESAFIOS DA SINGULARIDADE
TECNOLÓGICA**

Trabalho de Conclusão apresentado ao Curso de Licenciatura Plena em Filosofia da Universidade Estadual do Piauí, Campus Professor Alexandre Alves de Oliveira como requisito à obtenção do título de graduação em Licenciatura Plena em Filosofia.

Orientadora: Profa. Mestra Sorainy de Oliveira Mangueira

**PARNAÍBA-PI
2017**

V665i Vieira, Fernando Machado.
Inteligência artificial e os desafios da singularidade tecnológica /
Fernando Machado Vieira. - 2017.
60 f.

Monografia (Graduação) – Universidade Estadual do Piauí – UESPI,
Curso de Licenciatura Plena em Filosofia, *Campus* Prof^o. Alexandre Alves
de Oliveira, Parnaíba-PI, 2017.

“Orientador: Profa. Ma. Sorainy de Oliveira Manguieira.”

1. Inteligência Artificial. 2. Filosofia da Mente. 3. Replicação Mental.
I. Título.

CDD: 100



GOVERNO DO ESTADO DO PIAUÍ
UNIVERSIDADE ESTADUAL DO PIAUÍ – UESPI
PRÓ-REITORIA DE ENSINO DE GRADUAÇÃO – PREG
CURSO DE LICENCIATURA PLENA EM FILOSOFIA

ATA DE DEFESA DE MONOGRAFIA – Nº 006

Aos quinze dias do mês de janeiro de dois mil e dezoito, às dezenove horas, na Universidade Estadual do Piauí, Campus Professor Alexandre Alves, reuniu-se a Comissão Examinadora composta pelos Profs. Ma. Roberta Liana Damasceno Costa (Presidente- UESPI), Ms. (Membro Examinador/ UESPI) e Ms. Thiago Monteiro Chaves (Membro Examinador/ UESPI), para arguir a graduanda, **FERNANDO MACHADO VIEIRA** e avaliá-lo quanto à defesa da monografia intitulada: “ **INTELIGÊNCIA ARTIFICIAL E OS DESAFIOS DA SINGULARIDADE TECNOLÓGICA.**”

Após a arguição, a Banca Examinadora, composta pelos professores já identificados, resolveu aprovar a monografia examinada e atribuir a nota 10,0. Eu, Roberta Liana Damasceno Costa (Presidente da Banca Examinadora), lavrei a presente ata que segue assinada por mim e pelos demais membros da Banca Examinadora.

Parnaíba, 15 de janeiro de 2018.

Prof. Ma. Roberta Liana Damasceno Costa
Presidente – (UESPI)

Ms. Jorge Henrique Lima Moreira
(Membro Examinador/ UESPI)
Examinador 1

Ms. Thiago Monteiro Chaves
(Membro Examinador/ UESPI)
Examinador 2

RESUMO

O objetivo deste trabalho foi realizar uma investigação sobre os avanços da inteligência artificial, sua história e evolução no decorrer dos anos, assim como, pesquisar sobre as diferentes teorias que trabalham com a possibilidade de simulação, replicação e virtualização da mente humana para cérebros artificiais, e quais os benefícios e os riscos que tal tecnologia pode trazer para a sociedade humana. Sendo utilizado para tal pesquisa, um referencial teórico baseado principalmente em estudos de diversas correntes filosóficas dentro da filosofia da mente. Para produção deste trabalho, inicialmente, foi realizada uma pesquisa sobre a história da inteligência artificial e os aspectos tecnológicos e filosóficos que surgiram a partir de seu desenvolvimento, posteriormente, um estudo sobre as diversas correntes em filosofia da mente e as teorias que podem possibilitar a criação de programas de simulação e replicação de mentes artificiais, analisando por fim a possibilidade de aperfeiçoamento do corpo humano, de replicação mental e dos ideais transumanistas a respeito da singularidade tecnológica. De acordo com a pesquisa, foi verificado, que, com os avanços atuais em tecnologia, a possibilidade de mentes artificiais que simulam a mente humana tem tido uma evolução cada vez mais frequente, e que a possibilidade de replicação mental por parte das máquinas se torna apenas uma questão sobre quando irá acontecer. O mais relevante problema está relacionado às questões éticas sobre como tal tecnologia deverá ser usada para o melhoramento da sociedade em que vivemos.

Palavras-Chave: Inteligência Artificial. Filosofia da Mente. Replicação Mental.

ABSTRACT

This work aims to perform an inquiry about artificial intelligence findings, its history and evolution over time as well as to look for different theories whose business consists on looking for simulation, replication and virtualization of the human mind for artificial brains and what benefits and dangerous that such technology may come about for the human society. For such a research we will be using theoretical material from several philosophical point of view within philosophy of mind. This work started through a research about artificial intelligence history and the technological and philosophical features that came out of its development, and a research on several fields in philosophy of mind and theories that can create simulation and replication programs of artificial minds, focusing on at the end the possibility of human body enhancement, mental replication and trans-humanist ideas concerning on technological singularity. According to this research, the advances in technology improve the possibility of making up artificial minds that simulate human mind that have had an evolution increasingly and the possibility of mental replication becomes possible by the machines. The most striking problem is concerned to ethical issues on how such a technology should be used in order to improve the society we live in.

Keywords: Artificial intelligence. Philosophy of Mind. Mental Replication.

SUMÁRIO

1. INTRODUÇÃO	7
2. HISTÓRIA E DESENVOLVIMENTO DA INTELIGÊNCIA ARTIFICIAL	9
2.1. Descartes e o dualismo mente-corpo	11
2.2. Turing e a inteligência computacional	12
2.3. Neurociência e a filosofia da mente	16
3. TEORIAS SOBRE O FUNCIONAMENTO DA MENTE	24
3.1. Materialismo reducionista e materialismo eliminacionista	24
3.2. Funcionalismo	28
3.3. Conexionismo.....	37
4. NAVEGANDO AO INFINITO	42
4.1. Tecnologia em evolução	43
4.2. Transumanismo e a singularidade tecnológica	47
4.3. Simulações do novo mundo.....	53
5. CONCLUSÃO	58
REFERÊNCIAS	59

1. INTRODUÇÃO

O presente trabalho tem a finalidade de pensarmos sobre a possibilidade da simulação e replicação da mente humana por parte das máquinas, quais os benefícios advindos de uma máquina que pode chegar a pensar de forma consciente, assim como analisarmos os avanços das ciências da computação nas pesquisas com inteligência artificial e nos estudos sobre o funcionamento do cérebro.

O campo de estudo que será tema da pesquisa em questão será o da inteligência artificial, como está atualmente sendo desenvolvida e como ela pode ser importante para o melhoramento do corpo humano e até mesmo no ideal de imortalidade, com a possibilidade da criação de programas de simulação e replicação da mente, assim como o de transferências mentais para mecanismos artificiais e virtuais.

Essa situação dependerá de como os avanços em inteligência artificial resultarão na compreensão do funcionamento total do cérebro e de como simular os estados mentais em máquinas artificiais que, além de terem a capacidade funcional, venham a obter a capacidade de gerarem graus de consciências similares aos que ocorrem nos seres biológicos de nível superior.

Pretendemos apresentar ao leitor uma reflexão sobre os avanços da inteligência artificial e seu impacto no futuro da sociedade em que a replicação de mentes artificiais pode ser tornar algo possível. Apresentar, nesse contexto, as diversas correntes filosóficas que discutem o problema mente-cérebro em consonância com a neurociência; compreender as diversas teorias e pesquisas realizadas a respeito de como simular estados mentais em seres artificiais; entender como podemos ser capazes de construir redes neurais capacitadas a gerar uma programação própria e independente, alcançando aquilo que podemos entender como consciência, quais as dificuldades para tal empreendimento e suas implicações no

conhecimento sobre o funcionamento do cérebro; assim como, analisar as possibilidades do que os transumanistas chamam de singularidade, em que se vislumbra um futuro em que poderemos ser capazes de melhorar as capacidades humanas ao mesmo tempo em que poderemos ter o surgimento de uma inteligência artificial consolidada. Para tanto, utilizaremos como método de pesquisa, um estudo sobre algumas teorias, tanto filosóficas quanto científicas, a respeito da inteligência artificial.

No primeiro capítulo desta pesquisa teremos uma breve história da inteligência artificial, e tentaremos mostrar os aspectos tecnológicos e filosóficos que surgiram a partir de seu desenvolvimento. A importância de René Descartes para a ciência da computação, as descobertas de Alan Turing que foram fundamentais para o desenvolvimento da IA e o papel da neurociência dentro desse contexto nos dias atuais.

No segundo capítulo verificaremos as diversas correntes em filosofia da mente e algumas das teorias que podem possibilitar a criação de programas de simulação e replicação de mentes artificiais. Dentre as quais, daremos um enfoque maior para o materialismo reducionista e eliminacionista, o funcionalismo e o conexionismo. Dentro deste, falaremos ainda sobre as objeções acerca das tentativas de replicação e simulação da mente por parte das máquinas.

No terceiro capítulo faremos uma análise sobre a possibilidade da replicação mental, do impacto que a tecnologia tem nos dias atuais e de como sua intervenção na sociedade nos coloca dependente dela. Veremos o que abordam as teorias transumanistas, quais seus ideais e objetivos relacionados à hibridação homem-máquina e os riscos e benefícios que uma IA consciente pode trazer para a sociedade humana.

2. HISTÓRIA E DESENVOLVIMENTO DA INTELIGÊNCIA ARTIFICIAL

Estamos hoje em dia habituados a lidar com a tecnologia no nosso cotidiano de forma tão significativa que parece que as máquinas sempre fizeram parte da nossa história. A internet, os computadores, celulares, etc..., usamos máquinas dos mais diversos tipos, capacitadas a fazerem tarefas complexas que facilitam nosso meio de vida.

Temos ao nosso redor máquinas que “trabalham” praticamente sozinhas sem precisar de interferência humana a todo instante. Máquinas capazes de simular e analisar o comportamento humano nas mais diversas áreas de atuação. Toda essa inovação dá-se graças ao crescimento exponencial da inteligência artificial e sua aplicabilidade nos mais diversos setores da sociedade humana.

Em suma, este processo de revolução tecnológica pelo qual as máquinas estão atravessando engloba, hoje, diversas áreas do conhecimento e seu avanço rumo a uma singularidade tecnológica se torna praticamente inevitável e ininterrupto.

Para os pesquisadores da Inteligência Artificial (que daqui por diante abreviaremos IA) a mente humana funciona como um computador e, por isso, o estudo dos programas computacionais é a chave para se compreender alguma coisa acerca de nossas atividades mentais. Podemos construir programas que imitem nossa capacidade de raciocinar, de perceber o mundo e identificar objetos que estão à nossa volta, e até mesmo de falar e de compreender nossa linguagem. (TEIXEIRA, 2017, p. 15)

Hoje, o grande debate dentro da filosofia e da ciência da computação se dá referente à possibilidade de tais máquinas, graças aos avanços tecnológicos, chegarem a projetar estados mentais tal qual nós humanos, se tais máquinas podem chegar a simular a mente de tal forma que devamos qualificá-las como seres pensantes. Esse debate, entretanto, não é algo novo, idealizar máquinas pensantes é algo que circunda a humanidade já há muitos séculos.

A preocupação em relacionar a mente e o corpo já vem desde a antiguidade. Assim que questões sobre a mente começaram a surgir. Como a mente, que tinha um conceito imaterial, agia sobre o corpo que é algo material?

Será que somente nós humanos teríamos a capacidade de ter consciência e pensamento? Estas e outras questões fizeram surgir diversas teorias e experimentos a respeito da mente que vieram a convergir na inteligência artificial.

A história da inteligência artificial tem início desde os tempos antigos, quando já se pensava ao menos em termos teóricos sobre as questões do pensamento e de como se davam os processos de consciência no ser humano.

Na Grécia antiga filósofos como Platão e Sócrates já idealizavam um modelo de inteligência não humana que pudesse realizar tarefas humanas. O que se pensou, na época, era a possibilidade de que os objetos pudessem ter algum tipo de inteligência ainda que diferente da humana.

Tais ideias ficaram apenas em teorias e idealizações, bem como em lendas mitológicas, principalmente, pela falta de tecnologia avançada das quais dispomos hoje e que só começaram a ter um melhor desenvolvimento a partir do século XVII, com o advento da revolução industrial nos séculos XVIII e XIX que tornou possível o avanço tanto científico quanto tecnológico.

Os primeiros registros de criaturas artificiais com habilidades humanas têm uma forma mítica ou por vezes lendária, tornando difícil uma separação nítida entre imaginação e realidade. É esta confusão entre mito e realidade e, por vezes, a impossibilidade de distinguí-los que faz com que a IA possa ser considerada uma disciplina com um extenso passado, mas com uma história relativamente curta. (TEIXEIRA, 2017, p. 19)

Durante os séculos que se seguiram existem registros de várias máquinas que foram construídas com o intuito de simular seres vivos, não apenas seres humanos, mas também que imitassem o comportamento de animais, como o “pato de Vaucanson” construída pelo inventor francês Jacques de Vaucanson. A grande novidade desta invenção foi a grandiosidade da simulação dos movimentos realizados pelo pato mecânico, assim como a composição de cada parte articulada do corpo.

2.1. Descartes e o dualismo mente-corpo

Foi no século XVII, quando o filósofo francês René Descartes começou suas reflexões sobre o problema da mente, que os conceitos sobre inteligência artificial começaram a se tornar objetos de estudo de uma forma consistente. Vários autômatos ou idealizações de autômatos já estavam tomando contexto entre as ciências da época quando já se discutiam se seria possível construir um ser artificial pensante.

Descartes era totalmente avesso a este tipo de pensamento, de que máquinas pudessem adquirir a capacidade de pensar, pois, como o mesmo acreditava, somente os humanos tinham essa capacidade, pelo fato de que a mente, segundo o mesmo, era algo imaterial.

Descartes acreditava que a mente é, na realidade, aquilo que chamamos de alma e que foi-nos dada por Deus. Desta forma seria impensável que máquinas e também animais (que ele considerava como seres autômatos) pudessem ter ou adquirir uma alma, já que para ele somente os seres humanos possuíam esta dádiva.

(...) ainda que fizessem muitas coisas tão bem, ou talvez melhor do que qualquer um de nós, falhariam inevitavelmente em algumas outras, pelas quais se descobriria que não agem pelo conhecimento, mas apenas pela distribuição ordenada de seus órgãos. Pois, enquanto a razão é um instrumento universal, que serve em todas as ocasiões, tais órgãos precisam de alguma disposição específica para cada ação específica; daí decorre que é moralmente impossível que numa máquina haja muitas e diferentes para fazê-la agir em todas as ocasiões da vida, da mesma maneira que a nossa razão nos faz agir. (DESCARTES, 1999, p. 82)

Segundo Descartes, mente (alma) e corpo eram duas substâncias diferentes e foi seguindo esta linha de pensamento que ele escreveu em defesa da existência de Deus e da imortalidade da alma. Sua tese sobre a dualidade da mente (alma) e corpo ficou conhecida como dualismo, portanto para Descartes era totalmente inaceitável que máquinas pudessem chegar a pensar.

Depois disso, eu descrevera a alma racional, e havia mostrado que ela não pode ser de maneira alguma tirada do poder da matéria, como as outras coisas a respeito das quais falara, mas que devem claramente ter sido; e como não é suficiente que esteja alojada no corpo humano, assim como um piloto em seu navio, salvo talvez para mover seus membros, mas que é necessário que esteja junta e unida estreitamente com ele para ter, além disso, sentimentos e desejos parecidos com os nossos, e assim

compor um verdadeiro homem. (...) conseqüentemente, que não está de maneira alguma sujeita a morrer com ele; depois, como não se notam outras causas que a destruam, somos naturalmente impelidos a supor por isso que ela é imortal. (DESCARTES, 1999, p. 84)

As reflexões de Descartes sobre os problemas da mente foram um passo importante para as futuras gerações que viriam a discutir o problema mente-cérebro, tanto em termos filosóficos quanto científicos, sua profunda convicção de que a mente é uma substância separada do corpo tornou possível o aparecimento de outras teorias que conseqüentemente viriam a estar relacionadas com a IA.

2.2. Turing e a inteligência computacional

No início do século XX a IA toma novas proporções graças, principalmente, aos estudos e teorias do matemático inglês Alan Turing. Turing teve uma participação decisiva na 2ª guerra mundial ao criar um sistema de criptografia capaz de quebrar os códigos secretos dos alemães. O grande feito de Turing foi a criação da Máquina de Turing, com ela criou o modelo teórico para um computador universal.

A máquina de Turing constitui a melhor formalização da noção de algoritmo de que se tem notícia na história da Matemática. Uma máquina de Turing possui dois componentes: a) Uma fita, infinitamente longa, dividida em pequenos quadrados; cada um deles contém um conjunto finito de símbolos. b) Um scanner que pode ler, escrever e apagar símbolos dos quadrados da fita. O scanner é um dispositivo mecânico qualquer que permite “ler” o símbolo que está no quadrado, além de apagar ou imprimir símbolos que ali se encontram. (...) A máquina de Turing é o princípio geral para a construção de computadores digitais, pois, por meio dela, podemos executar qualquer tipo de algoritmo. Isto levou Turing à idéia de máquina de Turing universal, ou seja, à idéia de que qualquer computador pode, em princípio, ser concebido e reduzido a uma máquina de Turing. (TEIXEIRA, 1998, p. 23 - 25)

Para compreender o funcionamento da máquina de Turing podemos conceber um tipo de alfabeto simbólico que consiste em apenas dois símbolos, podendo ser composto por números ou letras, no caso, suponhamos que sejam os símbolos 0 e 1. Estes símbolos não devem ser entendidos como números, mas sim como símbolos, posto que poderíamos usar outros quaisquer, como A e B por exemplo. O importante aqui é a função que cada símbolo

executará e não especificamente a sua forma de representação. Abaixo uma representação habitual da máquina de Turing.

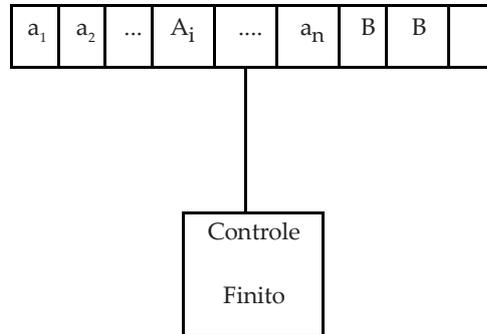


Figura 2.1. Esquema de uma máquina de Turing.

A operação realizada pela máquina de Turing se baseia em colocarmos nela uma fita com 0s e 1s que serão os dados de *input* e com a disposição do *scanner* em um ponto qualquer da fita onde será demarcado o quadrado inicial. Em seu livro *Mentes e Máquinas: uma introdução à ciência cognitiva*, Teixeira demonstra o funcionamento da máquina de Turing representado na figura 2.2.

O funcionamento se desenvolve da seguinte forma, o scanner irá se mover para a direita ou para a esquerda até encontrar o primeiro símbolo, no caso o 0 e em seguida irá substituí-lo pelo símbolo 1, ou seja, apaga o símbolo 0 e imprime o símbolo 1, em seguida o scanner irá se mover para a esquerda até parar.



Figura 2.2 Representação de operações na máquina de Turing.

Essa representação do funcionamento básico de uma máquina de Turing mostra como a ideia elaborada por Turing era bastante simples, o funcionamento da máquina se resume em ser capaz de manipular determinados símbolos em uma fita, de acordo com uma série de regras para guardar as informações obtidas. Esse é o conceito que se tornou a base de construção e funcionamento de todos os computadores que foram construídos desde então.

Turing também concebeu nessa época um teste para tentar descobrir os níveis de inteligência a que uma máquina capacitada com inteligência artificial poderia chegar, esse teste ficou conhecido como teste de Turing. O teste não tencionava demonstrar que máquinas podem pensar de fato por si próprias, mas consistia em verificar se uma IA poderia ser capaz de simular as atividades mentais humanas.

O teste se baseava em que uma pessoa fizesse uma sequência de perguntas direcionadas a um computador, sem saber que quem respondia era um computador. Se ao final das questões a pessoa não conseguisse distinguir se as respostas tinham sido dadas por uma máquina ou por um ser humano, o computador teria passado no teste.

O Teste de Turing baseia-se na seguinte ideia: se uma máquina apresentar um comportamento exatamente igual ao de um ser humano, não há por que não atribuir a ela pensamentos e estados mentais. O Jogo da Imitação, que foi a primeira versão do Teste de Turing, requer a participação de três pessoas: um homem (A), uma mulher (B) e um interrogador (C), o qual poderá ser de qualquer um dos sexos. O interrogador permanece num quarto fechado, separado dos outros dois. O objetivo do jogo, para o interrogador, é determinar qual é o homem e qual é a mulher. (...) Suponhamos agora que em vez de colocarmos um ser humano no lugar de "A" ou de "B" coloquemos uma máquina. Se no final do jogo "C" não descobrir que estava dialogando com uma máquina e não com um ser humano, podemos afirmar que essa máquina passou no Teste de Turing, ou seja, o comportamento da máquina tornou-se indistinguível do de um ser humano. (TEIXEIRA, 2016, P. 30 - 31)

Sendo que alguma máquina, de fato, passe no teste, quais as implicações causadas por tal acontecimento? Poderemos realmente admitir que tal máquina consegue pensar? O fato de poder dar respostas precisas torna a máquina consciente de que de fato responde questões? Em suma, ela tem consciência do que está a realizar nesse teste?

Esses são apenas alguns dos diversos questionamentos que o Teste de Turing nos traz. Se estivermos de acordo com Turing e aceitarmos que um computador pode pensar, devemos

levar em conta que haja então uma similaridade entre o funcionamento do cérebro e o funcionamento interno de um computador e é nesse ponto que a IA entra como um projeto, com o objetivo de demonstrar que, de fato, essa similaridade é consistente e que a mente pode ser simulada ou replicada por programas computacionais.

Ora, a máquina de Turing incorpora um tipo de dualidade parecida com aquela que nos leva frequentemente a opor mente e corpo: de um lado, um conjunto de regras abstratas (as instruções), e, de outro, a realização física dessas regras obtidas pelos diferentes estados da máquina. Foi precisamente esta dualidade presente na máquina de Turing que permitiu aos teóricos da IA propor uma perspectiva inovadora para o tradicional problema filosófico das relações mente e corpo: a ideia consiste em estabelecer uma analogia entre os estados mentais (pensamentos) e o *software* (conjunto de instruções da máquina ou programa do computador) de um lado, e entre estados cerebrais e o *hardware* ou os diferentes estados físicos pelos quais passa a máquina ao obedecer às instruções. (TEIXEIRA, 2017, p. 45)

Essa proposta de simulação computacional traz grandes perspectivas na tentativa de conceber o funcionamento e a relação entre os estado mentais e o cérebro pela simulação feita entre o *software* e o *hardware*, e buscar respostas para compreender como surge e como se dá o processamento dos estados mentais e como estes afetam ou são afetados pelo cérebro.

Durante a década de 50, a ciência e a tecnologia tiveram um grande avanço e, conseqüentemente, a contribuição na área de IA foi bastante significativa, várias correntes de pensamento sobre o comportamento da mente humana, assim como teorias e experimentos surgiram; muitas delas tiveram um papel significativo para os estudos sobre o funcionamento do cérebro humano.

Várias correntes filosóficas surgiram em decorrência das descobertas e discussões sobre os experimentos envolvendo a possibilidade de programas computacionais de inteligência artificial. Esses modelos geraram discussões sobre o que, de fato, é a mente e como esta se relaciona com o cérebro.

Esses estudos foram direcionados para o ramo da filosofia da mente, que abarcava não apenas os aspectos filosóficos e psicológicos abstratos, mas também tinha a possibilidade de se relacionar com os campos das ciências experimentais como a computação, a neurobiologia, a neurociência entre outros.

2.3. Neurociência e a filosofia da mente

Um dos ramos que mais contribuiu para os estudos da IA foi a neurociência. A neurociência tem sido um dos campos que trouxeram mais conhecimentos a respeito do funcionamento do cérebro, apesar de que, a cada descoberta, sempre surgem novas perguntas a despeito de suas atividades, o que nos mostra que os estudos nesta área ainda têm muito a desvendar.

O que chamamos simplificada mente Neurociência é na verdade Neurociências. No plural. Se é assim, quais são elas? E quem são os profissionais que lidam com elas. Há muitos modos de classificá-las, de acordo com os níveis de abordagem que mencionamos no início do capítulo. Um modo simples, mas esquemático, seria considerar cinco grandes disciplinas neurocientíficas. A Neurociência molecular tem como objeto de estudo as diversas moléculas de importância funcional no sistema nervoso, e suas interações. Pode ser também chamada de Neuroquímica ou Neurobiologia molecular. A Neurociência celular aborda as células que formam o sistema nervoso, sua estrutura e sua função. Pode ser chamada também de Neurocitologia ou Neurobiologia celular. A Neurociência sistêmica considera populações de células nervosas situadas em diversas regiões do sistema nervoso, que constituem sistemas funcionais como o visual, o auditivo, o motor etc. Quando apresenta uma abordagem mais morfológica é chamada Neuro-histologia ou Neuroanatomia, e quando lida com aspectos funcionais é chamada Neurofisiologia. A Neurociência comportamental dedica-se a estudar as estruturas neurais que produzem comportamentos e outros fenômenos psicológicos como o sono, os comportamentos sexuais, emocionais, e muitos outros. É às vezes conhecida também como Psicofisiologia ou Psicobiologia. Finalmente, a Neurociência cognitiva trata das capacidades mentais mais complexas, geralmente típicas do homem, como a linguagem, a autoconsciência, a memória etc. Pode ser também chamada de Neuropsicologia. E claro que os limites entre essas disciplinas não são nítidos, o que nos obriga a saltar de um nível a outro, ou seja, de uma disciplina a outra, sempre que tentamos compreender o funcionamento do sistema nervoso. (LENT, 2006, p. 6)

Esse ramo da ciência, em particular, se relaciona com outros campos de estudos, principalmente com a filosofia da mente, que procura tratar não apenas dos aspectos recorrentes ao funcionamento do cérebro, como também suas implicações no meio social dos seres humanos.

Não basta apenas saber como funcionam os estados cerebrais e se mente e cérebro são ou não a mesma coisa, mas devemos ter em consideração como estas descobertas poderão modificar, na prática, a sociedade humana, já que, com o avanço tecnológico intermitente, a

criação de máquinas que simulam a mente humana serão cada vez mais introduzidas no seio da sociedade.

Mas será que de fato poderemos chegar a criar máquinas tão sofisticadas que realmente possam simular a mente humana? Poderemos chegar a transplantar a mente, caso se comprove que esta é uma substância física, para um cérebro artificial, ou até mesmo enviá-la como um *software* para dentro da internet?

Para tentar responder tais questões, primeiramente veremos como a filosofia da mente, que também embarga a IA, pode ajudar na compreensão do funcionamento dos estados mentais e quais as correntes que podem, juntamente com a neurociência, formular mecanismos que possam ajudar nos desafios que virão com a possível chegada da singularidade tecnológica.

Para alguns transumanistas, a singularidade é a apoteose da tecnologia, sua realização suprema. Uma máquina cuja inteligência se iguale à do ser humano e pode produzir outras máquinas ainda mais inteligentes do que nós será a última invenção da era tecnológica. Não precisaremos mais gerar novas tecnologias – essas máquinas superinteligentes se encarregarão dessa tarefa. (TEIXEIRA, 2015, p.73)

Antes de adentrarmos mais especificamente sobre as questões da singularidade, vamos primeiramente nos situar nas correntes filosóficas que vêm contribuindo para a compreensão do problema mente-cérebro e do próprio significado do que é ser humano.

Em filosofia da mente temos diversas correntes de pensamento que discutem o problema mente-cérebro e a inteligência artificial, entretanto, nesse contexto, vamos ressaltar o materialismo reducionista e eliminacionista, o funcionalismo e o connexionismo.

Estas correntes serão analisadas com maior profundidade nos próximos capítulos, no entanto, faremos algumas explicações sobre o porque destes modelos serem de relevância nos estudos referentes à IA, assim como algumas colocações de filósofos que também terão seu pensamento em relação ao problema mente-cérebro abordados mais à frente.

Muitos filósofos como, particularmente Patrícia Churchland, são defensores de uma abordagem materialista, seja reducionista ou eliminacionista, para a resolução do problema

mente-cérebro. Embora, hoje, a maioria dos filósofos se recusem a aceitar o reducionismo como modelo de estudo dos estados mentais, este ainda é extremamente importante nos estudos neurocientíficos para a compreensão do funcionamento cerebral.

Para quê o materialismo eliminativo, então? Porque a caracterização existente das capacidades de nível superior do cérebro humano, presente no que se chama, na falta de um termo melhor, de *“folk psychology”*, pode muito bem ser reconstruída à medida em que o tempo e a neurociência cognitiva avançam. Isto também é uma hipótese empírica, em favor da qual já existem evidências empíricas. A reconstrução já está acontecendo para categorias como “memória”, “atenção” e “raciocínio”. (CHURCHLAND, 1994, n. p.)

Como defensora de que a neurociência irá, sem dúvida, esclarecer o funcionamento dos estados mentais e todo o arcabouço funcional do cérebro, Churchland mantém-se defensora da aplicação do modelo eliminativista para o entendimento do funcionamento da mente, aspecto este que iremos abordar com ênfase mais adiante.

P. Churchland defende algumas hipóteses importantes dentro da filosofia da mente, uma delas seria a suposição de que a mente, na realidade, não é mais do que apenas um processo físico, cuja atividade se dá no cérebro. A tão discutida atividade mental poderia ser definida como uma particularidade mais sofisticada de nosso organismo. A capacidade de tomar decisões não se dá a partir de uma mente (ou alma) consciente e independente do corpo físico.

Ao assumir que a neurociência pode revelar os mecanismos físicos subjacentes às funções psicológicas, eu estou assumindo que é de fato o cérebro que realiza essas funções – que as capacidades da mente humana são, na verdade, capacidades do cérebro humano. Essa suposição, juntamente com a concomitante rejeição de almas, espíritos cartesianos ou “substâncias fantasmagóricas” existindo separadamente do cérebro, não é delirante. Ao contrário, é uma hipótese altamente provável, baseada em evidências atualmente disponíveis da física, química, neurociência e biologia evolucionária. Ao dizer que o fisicalismo é uma hipótese, eu pretendo enfatizar seu status como uma questão empírica. Eu não presumo que seja uma questão de análise conceitual, de insight a priori ou de fé religiosa, embora eu esteja ciente de que nem todos os filósofos estão de acordo comigo neste ponto. (CHURCHLAND, 1994, n. p.)

Churchland quer demonstrar que nossa concepção presente sobre fenômenos psicológicos está baseada em teorias errôneas e falsas e que deve se partir para uma nova forma de compreensão e explicação destes fenômenos através da neurociência.

Com essa concepção, a tentativa é aplicar o reducionismo no estudo do corpo humano, mais especificamente no cérebro e com isso possibilitar uma análise do funcionamento cerebral e tentar demonstrar que a mente se reduz a fenômenos físicos e que está localizada no cérebro. Essa é uma tentativa de tentar demonstrar que nossa percepção e nossos pensamentos originam-se dos impulsos elétricos produzidos pelos neurônios, gerando o que chamamos de consciência subjetiva em nossa mente.

A mente, nessa concepção, deve estar em um local específico no cérebro, onde nossas percepções subjetivas estão localizadas, armazenadas e possíveis de serem acessadas pelo indivíduo. E, com o progressivo avanço da neurociência, este local deverá, em breve, ser descoberto.

O funcionalismo surge como uma teoria que sugere que a mente é um sistema que funciona através das relações de *inputs* e *outputs* entre o próprio sistema mental, o comportamento corporal e o meio ambiente. As relações mantidas nesse processo sugerem que estados mentais não precisam existir necessariamente em um cérebro biológico, mas em qualquer outro sistema que funcione dessa forma.

(...) o funcionalismo é provavelmente a teoria da mente mais amplamente aceita entre os filósofos, estudiosos da psicologia cognitiva e pesquisadores da inteligência artificial. Algumas das razões para isso ficaram manifestas na discussão precedente, e aqui vão mais algumas outras. Ao caracterizar os estados mentais como estados essencialmente funcionais, o funcionalismo situa o interesse da psicologia num nível que se abstrai do rico detalhamento de uma estrutura neurofisiológica (ou cristalográfica, ou microeletrônica) do cérebro. A ciência da psicologia, afirma-se ocasionalmente, é metodologicamente autônoma com relação a essas outras ciências (a biologia, a neurociência, a teoria dos circuitos), cujo interesse se restringe a detalhes de engenharia. Isso oferece uma justificativa para a grande massa de trabalho na esfera da psicologia cognitiva e da inteligência artificial, em que os pesquisadores postulam um sistema de estados funcionais abstratos e, então, testam o sistema postulado, muitas vezes por meio de simulações computacionais, em confronto com o comportamento humano em circunstâncias análogas. (CHURCHLAND, 2004, p. 70 – 71)

O posicionamento de Paul Churchland referente ao funcionalismo expressa bem a significância do funcionalismo dentro das pesquisas e experimentos realizados no que tange à inteligência artificial. Passando a não considerar o artefato físico como o mais importante para

que se ocorram os processos mentais, a simulação desses processos, se forem possíveis, podem ser realizadas em qualquer outro artefato físico, seja ele biológico ou artificial.

Embora tais argumentos possam ainda ser questionados, a teoria funcionalista nos retira do pedestal ao qual nos colocamos, ao nos analisarmos como únicos seres com capacidade de pensamento consciente e abre uma nova perspectiva sobre replicação da mente.

Ao considerar que não é preciso, nesta perspectiva, a necessidade de um organismo biológico para que as operações mentais possam surgir e serem processadas, o funcionalismo coloca às máquinas essa possibilidade totalmente em aberto, esperando, apenas, mais avanços nas pesquisas sobre o funcionamento completo dos estados cerebrais, para que tal fato possa vir a se concretizar.

Como todo nosso comportamento parece sempre estar a dar respostas aos *inputs* que recebemos do nosso exterior, nossa mente mostra-se em correlato sistema que determina quais funções deve-se alencar para resolver os problemas encontrados a cada instante. O próprio processo de pensamentos constantes e ininterruptos parece ser uma função pela qual o sistema vai gerando respostas até encontrar a mais adequada (ou não) para se efetuar a ordem e ação do processo seguinte a ser realizado pelo nosso corpo.

Uma das objeções fortes em relação ao funcionalismo que se coloca em filosofia da mente é a questão dos *qualia* (que será mais detalhada nos capítulos posteriores), pois aqueles que defendem o funcionalismo ainda não podem dizer como o mesmo pode explicar o estado qualitativo de nossos estados mentais. Se os estados mentais podem surgir em outro recipiente não biológico, não sabemos, no entanto, como são esses estados mentais.

Os sentimentos têm um estatuto verdadeiramente privilegiado. São representados em muitos níveis neurais, incluindo o neocortical, onde são os parceiros neuroanatômicos e neurofisiológicos de tudo o que pode ser apreciado por outros canais sensoriais. Mas, em virtude de suas ligações inextricáveis com o corpo, eles surgem em primeiro lugar no desenvolvimento individual e conservam uma primazia que atravessa sutilmente toda a nossa vida mental. Como o cérebro é o público cativo do corpo, os sentimentos são os primeiros entre iguais. E, dado que o que vem em primeiro lugar constitui um quadro de referências para o que vem a seguir, eles têm sempre uma palavra a dizer sobre o modo de funcionamento do

resto do cérebro e da cognição. Sua influência é imensa. (DAMÁSIO, 1996, p. 190 – 191)

A intensidade das sensações que temos ao sentirmos os *inputs* e *outputs* em nosso cérebro terão os mesmos aspectos cognitivos em um cérebro artificial? Uma máquina, embora possa a chegar a efetuar pensamento como um humano, terá a capacidade de sentir da mesma forma?

Como poderemos efetuar uma medição dos *qualia* nas máquinas, sendo que ainda não sabemos como fazer em relação a nós mesmos, já que cada intensidade qualitativa dos nossos estados mentais ainda está em um nível de particularidade profundamente individual e inacessível a quem está de fora, e nem mesmo sabemos ao certo como tais aspectos particulares influenciam nas tomadas de decisões processadas em nosso cérebro.

A abrangência do problema que gira em torno da questão dos *qualia* é um dos pontos mais importantes a serem pesquisados dentro dos programas de IA com os quais o funcionalismo e, conseqüentemente, o connexionismo tem se preocupado em lidar. Não por acaso essas linhas de pensamento tem recebido forte refutação, principalmente por parte de alguns filósofos.

Thomas Nagel é um destes filósofos contrários à replicação mental por parte das máquinas. Em seu artigo *Como é ser um morcego?* Nagel demonstra que é praticamente impossível termos acesso a subjetividade mental e que, por tal fato, as máquinas não poderiam exercer a replicação de nossos estados mentais. Outro filósofo que podemos citar nesse contexto é John Searle, uma de suas objeções contra o funcionalismo está em um de seus argumentos que foi denominado *O argumento do quarto chinês*.

Ambos argumentos serão investigados mais a frente. No que tange ao nosso ponto, neste instante, devemos levar em consideração que tais problemas não podem ficar sem solução. Embora tais argumentos nos demonstrem uma barreira para a replicação dos nossos

estados cognitivos, as pesquisas em IA que adotam os sistemas conexionistas atuais, nos levam a crer que o sucesso na replicação poderá acontecer em algum dado momento.

O computacionalismo clássico é o modelo de pesquisa que precede o conexionismo e que se baseia na ideia de que nossos sistemas cognitivos podem ser descritos como sistemas capazes de processar representações simbólicas e que, desse modo, seria possível a replicação de tal sistema biológico por um sistema artificial através de um processador central; no nosso caso, esse processador é identificado pelo cérebro.

O conexionismo se baseia no funcionamento do cérebro humano, segundo a ideia de que nosso cérebro funciona semelhante ao modelo computacional, em que o sistema é interligado por redes de processamento paralelamente distribuídos (PDP). Para isso, o conexionismo usa como modelo as estruturas neuronais do cérebro, simulando os processos de funcionamento dos neurônios e com isso trabalhar em modelo de simulação neural que possibilite a construção e o funcionamento de um modelo computacional constituído por “neurônios artificiais”.

O conexionismo será mais detalhado nos capítulos seguintes, em que procuraremos investigar as relações entre este modelo e as capacidades que o mesmo pode gerar para uma eventual simulação e replicação da mente humana. Devemos, entretanto, nos ater acerca de compreender um pouco sobre a base teórica pela qual está constituída a teoria conexionista e, assim, podemos entender sua base conceitual, segundo o que foi exposto por Teixeira no seu relato sobre o conexionismo.

O conexionismo, funcionalismo neurocomputacional ou processamento paralelo distribuído (PDP = *Parallel Distributed Processing*) não endossa a visão de que processos mentais possam ser estudados como computações abstratas, independentemente de sua base física e do meio ambiente onde se situa o organismo ou o sistema onde elas ocorrem. Conhecimentos acerca do funcionamento do cérebro e conhecimentos sobre computação devem convergir no estudo da natureza dos estados mentais. O cérebro humano é visto como um dispositivo computacional em paralelo que opera com milhões de unidades computacionais chamadas “neurônios” ou *neuron-like units*. Computadores e cérebros são sistemas cuja função principal é processar informação e, assim, podem-se utilizar redes artificialmente construídas para simular esse processamento. Tais redes constituem um intrincado conjunto de conexões entre essas *neuron-like units* que estão dispostas em camadas

hierarquicamente organizadas. Dado um determinado *input*, diferentes estados mentais podem ocorrer como consequência de mudanças nas conexões, que podem ser inibidas ou ativadas, variando de acordo com a interação do sistema com o meio ambiente e com seus outros estados internos. As conexões entre unidades estimuladas via *inputs* externos geram os chamados padrões de conectividade. As unidades estão conectadas umas com as outras: unidades ativas excitam ou inibem outras unidades. A rede funciona como um sistema dinâmico ou seja, uma vez dado o *input* inicial, este espalha excitações e inibições entre as unidades. Em alguns tipos de rede, este processo não para até que um estado estável seja atingido. (TEIXEIRA, 1998, p. 83 – 84)

Dessa forma, temos no modelo conexionista uma visão mais promissora no que tange à possibilidade de replicação da mente humana. Com a construção de um sistema que contenha neurônios artificiais que podem simular estados mentais, poderemos conhecer não só mais sobre o funcionamento de uma IA, mas principalmente como nossos estados mentais funcionam e se, de fato, podemos afirmar que nos diferenciamos tanto assim das máquinas que passamos a construir.

3. TEORIAS SOBRE O FUNCIONAMENTO DA MENTE

Abordaremos a partir daqui algumas das teorias relevantes ao funcionamento da mente conforme descrevemos no capítulo anterior. Não descreveremos todas as teorias que estudam o problema mente-cérebro, mas, conforme foi descrito anteriormente, somente aquelas que, neste trabalho, entendemos como mais apropriadas e que se adequam à pesquisa sobre as possibilidades de simulação artificial da mente.

As seguintes teorias serão levadas em questão nesta pesquisa: O materialismo reducionista em conjunto com o materialismo eliminacionista, o funcionalismo e o connexionismo. Não levantaremos uma análise aprofundada em demasia sobre cada teoria, mas procuraremos descrever o que basicamente cada teoria apresenta e sua aplicabilidade nos estudos da IA.

3.1. Materialismo reducionista e materialismo eliminacionista

Descreveremos aqui, no que concerne a teoria materialista, em dois pontos distintos: o reducionismo e o eliminativismo, seus pontos específicos e o que ambas teorias podem nos trazer de esclarecimento sobre funcionamento mental, suas características e potencialidades a serem desenvolvidas no campo da IA.

O materialismo reducionista tem como particularidade ser a mais simples dentre as teorias materialistas, sendo que tem a simplicidade como base primordial para compor a sua tese central. Basicamente, sua tese acarreta em dizer que os estados mentais são estados físicos do cérebro.

Temos dessa forma a compreensão de que cada estado ou processo mental que ocorre no cérebro é um estado ou processo físico, podendo ser identificado como tal. Desse modo, a

ocorrência de um estado mental não se daria por efeito de um processo físico, mas o próprio estado mental é um processo físico ocorrendo no cérebro, ambos são a mesma coisa.

O materialismo reducionista, mais conhecido como a teoria da identidade, é a mais simples das diversas teorias materialistas da mente. Sua tese central é a própria simplicidade: os estados mentais são estados físicos do cérebro. Isto é, cada tipo de estado ou processo mental é numericamente idêntico a (é uma e mesma coisa que) algum tipo de estado ou processo mental no interior do cérebro ou no sistema nervoso central. Até agora, não temos suficiente conhecimento sobre o intrincado funcionamento do cérebro para poder efetivamente estabelecer as identidades apropriadas, mas a teoria da identidade está comprometida com a ideia de que a pesquisa sobre o cérebro irá um dia revelá-las. (CHURCHLAND, 2004, p. 52 – 53)

O grande problema desta teoria, entretanto, se dá no fato de que ainda não temos o conhecimento necessário sobre o total funcionamento do cérebro, o que deixa a teoria reducionista complexa no seu modo de compreensão do funcionamento dos processos cerebrais, mesmo tendo por base a simplicidade.

O reducionismo com esse enfoque pode ser visto por alguns, apenas como mais uma teoria especulativa sobre o funcionamento de nossos estados mentais. Muito embora podemos salientar que o reducionismo pode buscar, na história da ciência, ocorrências que podem sustentar a teoria da identidade como algo provável, mesmo sem termos ainda a capacidade de compreensão das ocorrências dos processos que ocorrem no cérebro e que dificultam a demonstração da identidade entre processos mentais e processo físicos, conforme Churchland esclarece em seu livro *Matéria e Consciência: Uma introdução contemporânea à filosofia da mente*.

Esse resultado previsto pela teoria da identidade tem paralelos dos quais temos conhecimento em outras esferas da história da ciência. Tomemos como exemplo o som. Sabemos hoje que o som é simplesmente uma sucessão de ondas de compressão que viajam pelo ar e que a propriedade de um som de ser agudo é idêntica à propriedade de ter uma frequência oscilatória alta. Aprendemos que a luz nada mais é que ondas eletromagnéticas, e nossa melhor teoria atual diz que a cor de um objeto é idêntica a uma trinca de coeficientes de refletância que o objeto tem, como se ele estivesse "tocando uma corda musical", embora as "notas" sejam tocadas em ondas eletromagnéticas, e não em ondas sonoras. Hoje reconhecemos que o calor ou o frio de um corpo são apenas a energia do movimento das moléculas que o constituem: o calor é idêntico à alta energia cinética molecular média, e o frio é idêntico à baixa energia cinética molecular média. Sabemos hoje que o relâmpago é idêntico a uma súbita descarga de elétrons em larga escala, entre as nuvens ou entre a atmosfera e o solo. O que hoje pensamos como "estados mentais", argumenta o defensor da teoria da identidade, é idêntico aos estados do cérebro, e exatamente do mesmo modo. (CHURCHLAND, 2004, p. 53)

Os exemplos citados por Churchland são conhecidos como casos de redução interteórica, que implica em afirmar que uma nova teoria sobre determinados conceitos estruturais são correspondentes aos conceitos estruturais de uma teoria antiga de um determinado processo ou fenômeno que tem a mesma ocorrência.

A diferença está em que as noções antigas de uma determinada teoria são identificadas de outra forma, como no caso da noção de luz, que passa a ser identificada como sendo, na verdade, ondas eletromagnéticas, mas tanto a noção de luz quanto a noção de ondas eletromagnéticas são a mesma coisa, falamos de um mesmo processo que está em ocorrência.

O que a nova teoria implica em dizer é que temos agora uma melhor explicação para a ocorrência de um determinado fenômeno. Desta forma, não temos uma substituição de um processo por outro, mas uma similaridade. O que ocorre são as mesmas coisas, mas temos uma descrição mais correta sobre o funcionamento desses processos.

(...) Assim, não haveria nada de particularmente surpreendente numa redução de nossos estados mentais potencialmente introspectivos a estados físicos do cérebro. Tudo que precisaríamos seria uma neurociência bem-sucedida em termos de explicação, que se desenvolvesse a ponto de implicar uma "imagem especular" dos princípios e pressupostos que constituem o arcabouço conceitual de nosso senso comum para os estados mentais, uma imagem na qual os termos para os estados do cérebro ocupariam as posições ocupadas pelos termos para os estados mentais nos pressupostos e princípios do senso comum. Se essa (bastante exigente) condição fosse de fato satisfeita, então, assim como nos casos históricos citados, estaríamos justificados em estabelecer uma redução e em afirmar a identidade dos estados mentais com os estados do cérebro. (CHURCHLAND, 2004, p. 55)

Com base na condição de uma redução interteórica bem sucedida, o que se espera é poder fazer o mesmo tipo de redução com relação aos processos mentais e cerebrais para se justificar a teoria de que estados mentais e estados cerebrais são, de fato, a mesma coisa, ou que ambos os processos são, na verdade, a mesma ocorrência.

O que se espera é que o avanço substancial da neurociência possa nos dar uma explicação satisfatória de tal funcionamento. Com a justificação da identidade de tais processos poderemos conceber um cérebro artificialmente projetado constituído de processos cerebrais tal qual um cérebro biológico.

O que dificulta esta teoria, ainda é o fato de não termos acesso ou compreensão do funcionamento total do cérebro, e, além disso, dentro do próprio pensamento materialista temos outros pontos de vista que se colocam contrários, pelo menos em parte, ao proposto pela teoria reducionista, dentre o qual destacamos o materialismo eliminacionista.

O materialismo eliminacionista não considera a redução interteórica como a melhor maneira de compreendermos nossos processos mentais e cerebrais. É contrário a ideia de que uma nova teoria a respeito de determinado fenômeno seja um correspondente melhor explicado de uma teoria anterior.

Para o materialista eliminacionista, as correspondências um a um não serão encontradas, e a estrutura psicológica de nosso senso comum não pode obter uma redução interteórica, porque a estrutura psicológica de nosso senso comum é uma concepção falsa e radicalmente enganosa das causas do comportamento humano e da natureza da atividade cognitiva. Desse ponto de vista, a psicologia popular não é apenas uma representação incompleta de nossas naturezas interiores; ela é pura e simplesmente uma representação distorcida de nossas atividades e estados internos. Consequentemente, não podemos esperar que uma explicação realmente adequada de nossa vida interior feita pela neurociência revele categorias teóricas que correspondam exatamente às categorias do arcabouço de nosso senso comum. Dessa forma, devemos esperar que o antigo arcabouço seja simplesmente eliminado, e não reduzido, por uma neurociência amadurecida. (CHURCHLAND, 2004, p. 55)

Nesse ponto, para o eliminacionista, ao se encontrar uma melhor explicação para se justificar a ocorrência de um determinado processo ou fenômeno, esta nova teoria não tem qualquer junção ou similaridade com a antiga, pois esta nova teoria revela o fato de que todo o pensamento anterior sobre tal ocorrência era falsa e como tal deve ser completamente eliminada.

Para o eliminacionista, todas as dúvidas que temos sobre o funcionamento da mente estão presas à postulações do senso comum, que impendem um entendimento mais aprofundado, justamente por termos convicções que simplesmente não deveriam ser consideradas relevantes e sim eliminadas do meio científico.

Os conceitos da psicologia popular - crença, desejo, medo, sensação, dor, alegria etc. - terão um destino análogo, segundo esse ponto de vista. E, quando a neurociência tiver amadurecido, a ponto de a pobreza de nossas atuais concepções ter-se tornado manifesta a todos, e a superioridade do novo arcabouço tiver sido estabelecida, poderemos, por fim, dar início à tarefa de reformular nossas concepções das atividades e estados internos, no interior de um arcabouço conceitual realmente adequado. Nossas explicações sobre os comportamentos uns dos outros irão recorrer

a coisas como nossos estados neurofarmacológicos, nossa atividade neural em áreas anatômicas específicas e a outros estados que forem relevantes para a nova teoria. Nossa introspecção individual também será transformada e poderá ser profundamente aprimorada em razão de um arcabouço conceitual mais penetrante e preciso, com o qual ela terá de trabalhar - da mesma forma que a percepção do céu noturno pelo astrônomo foi em muito aprimorada pelo conhecimento detalhado da moderna teoria astronômica de que ele dispõe. (CHURCHLAND, 2004, p. 55)

Dessa forma, conceber os estados mentais como sendo a mesma coisa que processos cerebrais com base na redução interteórica, dentro do entendimento eliminacionista, é algo totalmente inviável. Tais correspondências se tornam impossíveis de serem encontradas, pois os estados mentais como compreendemos fazem parte de um conhecimento falso ou ilusório e, como tal, não podem ser reduzidos.

Diante dessas duas concepções materialistas, temos dois caminhos de possibilidades para a concepção de um cérebro artificial, baseado no princípio de que nossos estados mentais são na realidade estados físicos e não outra coisa imaterial. Ambas as teorias se sustentam partindo do mesmo princípio, embora, como veremos adiante, outras teorias possam ser mais viáveis para o entendimento do funcionamento do cérebro.

3.2. Funcionalismo

Tratamos um pouco do funcionalismo no primeiro capítulo, em que buscamos mostrar como os estudiosos do funcionalismo procuram definir o funcionamento da mente como um sistema de relações entre o próprio sistema mental, o comportamento corporal e o meio ambiente.

Segundo observamos, no funcionalismo os estados mentais são caracterizados como estados essencialmente funcionais, obtidos por relações causais entrelaçadas dentro de um sistema em que cada estado mental desempenha uma função específica.

De acordo com o funcionalismo, a característica essencial que define todo tipo de estado mental é o conjunto de relações causais que ele mantém com (1) os efeitos do meio ambiente sobre o corpo, (2) com outros estados mentais e (3) com o comportamento corporal. A dor, por exemplo, resulta de traumas ou danos ao corpo;

ela causa sofrimento, irritação e uma avaliação prática, visando a seu alívio; ela também provoca comportamentos de esquiva e empalidecimento e o tratamento da área traumatizada. Todo estado que desempenha exatamente esse papel funcional é uma dor, de acordo com o funcionalismo. De modo análogo, outros tipos de estados mentais (sensações, medos, crenças, e assim por diante) também são definidos por seus papéis causais únicos, numa economia complexa de estados internos mediando entradas de dados sensoriais e saídas comportamentais. (CHURCHLAND, 2004, p. 67 - 68)

Apesar de tratar o sistema mental nos moldes de *inputs* e *outputs*, o funcionalismo tenta demonstrar que o estado mental é uma relação causal constante e não apenas uma questão de entrada e saída de dados, como era pensado pelos adeptos do behaviorismo.

Essa concepção pode trazer o behaviorismo à mente do leitor, e, de fato, ela é herdeira do behaviorismo. Porém, há uma diferença fundamental entre as duas teorias. Enquanto o behaviorismo esperava definir cada tipo de estado mental exclusivamente em termos de entradas de dados do meio ambiente e saídas comportamentais, o funcionalista nega que isso seja possível. Para o funcionalista, a caracterização adequada de quase todos os estados mentais envolve uma referência não eliminável a uma série de estados mentais com os quais o estado mental em questão está conectado em termos causais, e, assim, uma definição reducionista exclusivamente em termos de entradas e saídas é totalmente impossível. Dessa forma, o funcionalismo está imune a uma das principais objeções contra o behaviorismo. (CHURCHLAND, 2004, p. 68)

Assim, podemos ter, no funcionalismo, uma visão mais sofisticada para concebermos a possibilidade de simulação e/ou replicação dos estados mentais. Dado que considerando o que nos traz o funcionalismo, os estados mentais não necessitam de serem concebidos por um único tipo de receptáculo para acontecer.

Não necessitamos unicamente de um cérebro biológico para que tal sistema funcione, mas que as relações funcionais possam emergir da mesma forma que emergem no sistema mental habitual. Nesse caso podemos criar simulacros artificiais semelhantes ao cérebro em seu funcionamento e simularmos situações causais para determinar a capacidade de simulação exercida de forma artificialmente.

Imaginemos um ser de outro planeta, diz o funcionalista, um ser com uma constituição fisiológica alienígena, uma constituição baseada no elemento químico silício, por exemplo, em vez do elemento carbono, como a nossa. A química e mesmo a estrutura física do cérebro do alienígena teriam de ser sistematicamente diferentes dos de nosso cérebro. Mas, mesmo assim, esse cérebro alienígena poderia muito bem manter uma economia funcional de estados internos, cujas relações recíprocas estão em perfeito paralelo com as relações recíprocas que definem nossos próprios estados mentais. O alienígena pode ter um estado interno que satisfaz a todas as condições exigidas para um estado de dor, como descrevemos anteriormente. Esse estado, considerado a partir de um ponto de vista puramente físico, teria uma constituição muito diferente de um estado de dor para o ser

humano, mas ele poderia, no entanto, ser idêntico a um estado de dor humano, de um ponto de vista puramente funcional. E o mesmo vale para todos os seus estados funcionais. (CHURCHLAND, 2004, p. 68 - 69)

O exemplo acima postulado por Churchland dá uma ideia do que se propõe a teoria funcionalista. Podemos tentar construir modelos artificiais que simulem os estados mentais, considerando seu comportamento funcional correspondente ao funcionamento dos estados mentais de um cérebro humano. Aqui o importante não seriam as características físicas do agente, mas sim, se as respostas condizem com o modelo de relações causais proposto pelo funcionalismo.

Se podemos pensar numa constituição alienígena, podemos também pensar em muitas, e o mesmo vale para os sistemas artificiais. Se criássemos um sistema eletrônico - algum tipo de computador - cuja economia interna fosse funcionalmente isomórfica com relação à nossa própria economia interna, em todos os aspectos relevantes, então também ele teria estados mentais. (CHURCHLAND, 2004, p. 69)

Desse modo, os adeptos do funcionalismo podem dizer que existem, na natureza, outros meios para que um agente pensante possa surgir e não exclusivamente da maneira que nós já conhecemos. O rompimento com a barreira biológica não irá excluir a maneira pela qual seremos capazes de pensar e de sentir, pois, se o funcionalismo estiver certo, o cérebro biológico é apenas uma das muitas maneiras dos estados mentais surgirem.

Em todo caso, devemos considerar aqui algumas objeções feitas ao funcionalismo que têm sua relevância de serem relatadas, para tentarmos entender os problemas pelos quais uma tese funcionalista ainda requer de muitos estudos neurocientíficos. A primeira objeção refere-se àquilo que chamamos de *qualia*.

Qualia: As qualidades sentidas ou fenomênicas associadas às experiências, tais como a sensação de uma dor, ou à audição de um som, ou à visão de uma cor. Saber como é ter uma experiência é conhecer as suas *qualia*. A ideia de que, antes de tudo, conhecemos as *qualia*, e que só indiretamente e por seu intermédio conhecemos as propriedades das coisas exteriores, é atacada no argumento da linguagem privada de Wittgenstein; mas as *qualias* são muitas vezes encaradas como o maior obstáculo às filosofias científicas da mente, tais como o funcionalismo e o fisicalismo. [...] (BLACKBURN, 1997, p. 327)

Segundo a definição acima sobre o que é um *qualia*, entendemos que se refere às experiências subjetivas que um determinado agente possui. São os aspectos subjetivos da

consciência que somente o indivíduo que passa por determinada experiência tem acesso e, somente ele e nenhum outro, pode identificar a real intensidade ou definição desta experiência.

Quando nos referimos à dor, por exemplo, embora qualquer agente pensante que já tenha tido a experiência de uma dor possa compreender que outro agente está sentindo dor, ele, no entanto, não poderá relatar a intensidade dessa dor, pois esta é exclusivamente sentida pelo agente que a está sentindo.

Imaginemos um agente robótico cuja capacidade de simulação mental tenha sido realizada com sucesso. O mesmo se tornou capaz de empreender pensamentos por si mesmo conforme a postulação funcionalista e já tem a compreensão de vários tipos de experiências humanas, inclusive do que é uma dor.

Digamos que a este agente é dado um teste em que ele deve observar uma cena em que uma pessoa sofre um assalto, é baleada na perna e está caída no chão com muitas dores. O agente robótico será capaz de identificar o ocorrido e poderá dizer, por exemplo, que a pessoa em questão está sentindo dores, devido ter sido atingida por um projétil e precisa de cuidados médicos imediatamente.

O que este agente não poderá dizer, entretanto, é como é a dor que esta pessoa está sentindo, qual a intensidade dessa dor, que experiência de fato esta pessoa está tendo em sua consciência. Não poderá ele relatar, também, o que a pessoa sentiu ou pensou ao ser confrontada pelo assaltante, que tipo de experiência ela teve nesse momento em questão.

O filósofo Thomas Nagel em seu artigo *Como é ser um morcego?* defende que a simulação artificial é algo inalcançável, pelo que ele presume que o caráter subjetivo da consciência é algo que não pode ser replicado de forma alguma, pois não basta a um agente artificial simular estados mentais, é necessário se ter a consciência de que tem estados mentais.

Podemos chamar isso de o caráter subjetivo da experiência. Ele não é capturado por quaisquer das recentes e familiares análises reducionistas do mental, já que todas elas são logicamente compatíveis com sua ausência. Não é analisável em termos de nenhum sistema explicativo de estados funcionais, ou de estados intencionais, pois esses poderiam ser atribuídos a robôs ou autômatos que se comportassem como pessoas, embora não experimentassem nada. (NAGEL, 1991, p.247)

Nagel diz em seu artigo, que por mais que possamos imaginar como é ser um morcego, não podemos saber efetivamente como é realmente ser um morcego. Podemos imaginar como seria se pudéssemos voar, comer, dormir como um morcego e como seria “enxergar” através de um sonar, sem ter os sentidos da visão. Mas tudo isso seria apenas teorização, já que somos incapazes de ter acesso à experiência subjetiva da mente de qualquer morcego para saber de fato como é ser um morcego.

Mesmo que eu pudesse parecer uma vespa ou um morcego, ou comportar-me como eles, sem modificar a minha estrutura fundamental, minhas experiências não seriam nada parecidas com as experiências de tais animais. Por outro lado, é muito duvidoso que qualquer significado possa ser associado à suposição de que eu poderia possuir a constituição neurofisiológica interna de um morcego. Mesmo se eu pudesse ser gradualmente transformado em um morcego, nada na minha constituição presente me tornaria apto a imaginar o que as experiências de tal estágio futuro de mim, mesmo assim metamorfoseado, poderiam ser. A melhor evidência viria das experiências dos morcegos, mas apenas se soubéssemos como elas são. (NAGEL, 1991, p.250)

Essa experiência privada que cada um de nós tem, internamente, é o problema que se insere quando nos damos conta da importância dos *qualia*. Somos capazes de compreender os acontecimentos internos que uma pessoa tem por já termos passado por experiências similares, mas não somos capazes de determinar o que, de fato, se passa na mente de uma pessoa devido ao acesso privado que somente cada um tem internamente.

Eu não nego que os estados e eventos mentais causem o comportamento, nem que possam ser dadas caracterizações funcionais deles. Nego apenas que esse tipo de coisa esgote a análise dos mesmos. Qualquer programa reducionista tem que se basear em uma análise do que deve ser reduzido. Se a análise deixa algo de fora, o problema será colocado erroneamente. É inútil basear a defesa do materialismo em qualquer análise dos fenômenos mentais que não encare explicitamente o seu caráter subjetivo. (NAGEL, 1991, p.247)

Por mais que alguém saiba o que é uma dor, não pode definir a experiência dessa dor ao indivíduo que a sente. Como então sabermos se realmente um agente artificial tem real compreensão dos fatos a que está sendo exposto? Ele estaria realmente deliberando

pensamentos cognitivos como esperado ou apenas registrando dados sem sua real compreensão ou consciência dos fatos?

Se se deseja defender o fisicalismo, deve ser dada uma explicação física dos aspectos fenomênicos. Mas quando examinamos seu caráter subjetivo, parece que tal feito é impossível. A razão é que todo fenômeno subjetivo é essencialmente conectado a um ponto de vista singular e parece inevitável que uma teoria física, objetiva, abandone esse ponto de vista. (NAGEL, 1991, p. 248)

Como surge a consciência? O que de fato torna uma mente consciente? Apenas as simulações de estados mentais e comportamentais são suficientes para considerarmos que aquele que apresenta um determinado comportamento mental está tendo consciência do que está fazendo? E, se está, podemos afirmar como esse agente está tendo esta experiência internamente?

Essas perguntas, relacionadas à consciência subjetiva que o ser humano possui, de que tem a compreensão dos seus estados mentais e ações comportamentais que pratica, são apresentadas por estudiosos da mente como um desafio para a replicação artificial da mente.

Se ainda não temos uma explicação para o surgimento da consciência, como poderemos inferir que um determinado agente artificial, ao simular comportamentos mentais, está tendo consciência do que está realizando? Não seriam apenas simulações que, apesar de terem uma objetividade, não têm nenhum significado para o agente praticante?

O filósofo John Searle, apresentou sua refutação sobre o caso no seu argumento que ficou muito conhecido dentro da filosofia da mente, chamado de argumento do quarto chinês. Neste argumento, Searle tenta demonstrar que é impossível que uma máquina, uma IA, seja capaz de ter consciência do seu comportamento, de ter compreensão daquilo que está fazendo.

O debate em torno dos limites da Inteligência Artificial toma um impulso decisivo a partir da publicação do artigo de J. Searle, *Minds, Brains and Programs*, em 1980. Neste artigo, Searle apresenta uma crítica vigorosa à possibilidade de se obter um equivalente mecânico para o fenômeno cognitivo humano que normalmente denominamos compreensão. Seu ponto de partida é a análise dos programas para compreender histórias curtas desenvolvidas por R. Schank. [...] Os programas desenvolvidos por R. Schank [...] tinham por objetivo a compreensão de histórias. [...] O programa de Schank é construído de tal maneira que lhe é possível responder coerentemente a questões elaboradas com base no texto da história. Tendo em vista estes resultados, Schank sustenta que este tipo de programa é capaz de compreender

o texto e constituir uma explicação para a capacidade do ser humano de compreender textos ou histórias curtas. (TEIXEIRA, 1998, p. 68 - 69)

Segundo Searle, uma IA só seria capaz de realizar simulações do comportamento humano, mas não seria capaz de pensar efetivamente e de compreender o comportamento que realiza. Searle acredita que as máquinas só conseguem compreender a sintaxe do seu comportamento funcional, sem ser possível a estas compreenderem a semântica. Com isso, ele afirma que as máquinas não entendem o real significado das ações que estão realizando.

As críticas desenvolvidas por Searle às pretensões de que um tal programa realmente compreende baseiam-se na construção de um experimento mental que reproduz o procedimento do próprio programa. O caminho adotado por Searle para construir este experimento mental é o inverso do procedimento normalmente utilizado para elaborar simulações cognitivas: trata-se de instanciar o programa de Schank num sujeito humano. (TEIXEIRA, 1998, p. 69)

O argumento se baseia na seguinte experiência: suponhamos que um de nós esteja dentro de um quarto fechado onde se encontram vários símbolos chineses, um livro com significados de cada símbolo em português e uma janela onde são inseridas as entradas e saídas de dados (textos em chinês escritos em um papel).

Em um determinado momento, é inserido, através da janela, um pedaço de papel com perguntas em chinês e que a pessoa que está dentro deve devolver com respostas em chinês. A pessoa no quarto não sabe chinês, então o que ela faz é verificar os símbolos em chinês com o livro que corresponde esses símbolos para o português, fazer a transformação desses símbolos e devolver o papel com as respostas traduzidas em chinês.

O falante (trancado no quarto) recebe periodicamente novos textos e questões em chinês e aplica estas operações ou regras de transformação associando as seqüências anteriores com as seqüências mais recentes. Com base nestas regras de transformação ele passa a emitir ou escrever mais seqüências de símbolos em chinês. Claro que o falante preso no quarto não sabe precisamente o conteúdo das informações que ele está gerando com base nos dois textos e nas regras de transformação. O primeiro texto corresponde, em nosso experimento mental, ao relato que é fornecido ao computador; o segundo texto, ao conjunto de questões que é elaborado com base neste relato; e as novas seqüências geradas, às respostas a tais questões. As regras de transformação são bastante complexas e concebidas de maneira tal que elas simulem os processos mentais e o comportamento lingüístico de um falante nativo de chinês numa conversação habitual. Após um certo tempo, o falante aprendeu a manipular perfeitamente estas regras de transformação e, com base nos outputs, um observador externo poderia dizer que ele compreende chinês -, o que, no entender de Searle, constitui um contra-senso. (TEIXEIRA, 1998, p. 69)

O que Searle coloca é que, apesar da pessoa dentro do quarto ser capaz de corresponder os símbolos em chinês com seu respectivo significado em português e transcrever respostas em chinês relativas às perguntas inseridas, ela não compreende nenhuma palavra em chinês.

A instanciação dos programas de Schank num sujeito humano, reproduzida neste experimento mental, é, para Searle, bastante reveladora. Ela mostra que os programas desse tipo não estabelecem as condições necessárias para a simulação da atividade cognitiva da *compreensão*: o falante aplica as regras de transformação e compreende estas regras, mas as seqüências de símbolos em chinês não têm nenhum significado para ele. A manipulação de símbolos realizada no programa é inteiramente cega. (TEIXEIRA, 1998, p. 69 - 70)

A pessoa no quarto se tornou capaz de efetuar uma ação programática, com base na experiência dos dados obtidos e organizados sistematicamente, mas não tem nenhuma compreensão dos dados em chinês que a mesma se propôs a responder. Ela compreende as funções que deve executar, mas não é capaz de compreender o significado da ação executada.

Com base nesse argumento, Searle quer demonstrar que nenhuma máquina que seja capaz de simular ou replicar o comportamento mental humano, será capaz de dar significado a tal comportamento. Isso por entender que as máquinas não têm a capacidade de dar significado ao que executam.

Ademais, como ressalta Searle na resposta às objeções ao seu texto *Minds, Brains and Programs*: “A manipulação de símbolos formais, por si só, não tem intencionalidade, não é sequer manipulação de símbolos, uma vez que esses símbolos não simbolizam nada. Eles têm apenas sintaxe, mas não semântica”. Ainda na sua resposta às objeções, Searle ressalta que a esses programas “falta aquilo que chamarei de intencionalidade intrínseca ou de genuínos estados mentais”. A atribuição de intencionalidade ou de significado, diz Searle, é, nestes casos, sempre uma atribuição a posteriori, dependente da intencionalidade intrínseca dos sujeitos humanos que observam os outputs do programa. (TEIXEIRA, 1998, p. 70)

Seguindo essa linha de raciocínio exposto por Searle, podemos entender que o mesmo quer demonstrar que as máquinas não teriam a capacidade de dar significado às tarefas que executam por não terem a consciência subjetiva para pensarem sobre aquilo que fazem. Dessa maneira, supõe Searle, pode-se dizer que nenhuma máquina seria capaz de pensar por si mesma e dar significado próprio às experiências as quais é exposta no mundo.

Searle, entretanto, parece não perceber, com este argumento, um aspecto específico. Se considerarmos que a pessoa dentro do quarto não tem compreensão daquilo que está fazendo, se ela não tem a consciência das tarefas que executa, então não estamos apenas dizendo que uma máquina não tem consciência, mas que também nós, humanos, não possuímos consciência.

Se o experimento se utiliza de um humano para simular o comportamento funcional de uma máquina e demonstrar que esta não é capaz de gerar comportamentos mentais com significado, estamos afirmando a mesma coisa inversamente. Que seres humanos também não compreendem o comportamento que executam.

Ocorre que vários filósofos favoráveis ao projeto da Inteligência Artificial apresentaram contra-argumentos às posições defendidas por Searle. Alguns deles apontaram falhas no argumento principal, salientando que não sabemos se, de fato, os computadores podem ou não compreender alguma coisa. A situação seria semelhante àquela quando observamos um ser humano responder a perguntas a respeito de um texto qualquer: como podemos estar certos de que essa pessoa compreende o que está fazendo? Por acaso muitos de nossos processos mentais cotidianos não são tão rotineiros que os fazemos por uma associação tão mecânica e cega como as do computador? Se as operações efetuadas pelo falante trancado no quarto são cegas, será que não podemos afirmar o mesmo de nossas próprias operações mentais? Mesmo quando tentamos examinar o fluxo de nossos pensamentos, isto não nos dá nenhuma informação sobre como ocorrem as operações de nosso cérebro. Somos, em grande parte, opacos para nós mesmos – e não seria essa uma situação idêntica à de alguém que olha para os resultados das operações de um computador e, com base nestes últimos, quer sustentar a afirmação de que essa máquina nada compreende acerca dessas operações? (TEIXEIRA, 1998, p. 71 - 72)

Podemos repensar então o argumento de Searle, adicionando ao mesmo as afirmações de Nagel, sobre a impossibilidade de se ter acesso ao caráter subjetivo da mente. Como poderemos afirmar que o mesmo não acontece com as máquinas? Como podemos afirmar que, de fato, uma máquina “pensa” apenas como a descrição realizada no quarto chinês?

Sabemos realmente como é ser uma máquina, apenas supondo seu comportamento funcional compreendido no plano da sintaxe? Se estes comportamentos podem descrever um aprendizado, não podemos dizer que esta máquina aprendeu e que tem consciência da tarefa que está executando? Quando somos crianças não temos nenhuma compreensão concreta das coisas que somos sugestionados a fazer, no entanto, fazemos e, conseqüentemente,

aprendemos com o que fazemos e adquirimos consciência daquilo que fazemos e objetivamos nossas ações funcionais no mundo.

As refutações, tanto de Nagel quanto de Searle ainda deixam o debate em aberto e possibilidades continuam pertinentes. Não podemos dar por superada a questão da consciência e intencionalidade expostas nas refutações acima, mas a partir delas compreendermos a própria noção de aprendizagem pela qual passam nossos processos mentais.

Vimos, então, que o funcionalismo nos traz possibilidades bastante coerentes para emprendermos pesquisas que podem trazer transformações nos modelos que as máquinas postulam estados mentais. Além do funcionalismo, nossa pesquisa em questão aborda ainda outro modelo de pesquisa que trataremos a seguir, o connexionismo.

3.3. Conexionismo

O connexionismo traz propostas ainda mais estimulantes a respeito de simulação e replicação mental de forma artificial. Conforme já foi explicitado no primeiro capítulo, o connexionismo se baseia no funcionamento do cérebro para tentar simular, através de neurônios artificialmente construídos, os estados mentais correlacionados ao do cérebro humano.

No final da década de 1950, muito cedo na história da IA, havia muito interesse nas "redes neurais" artificiais, isto é, em sistemas de hardware modelados com base no cérebro biológico. Apesar de sua atratividade inicial, essas redes de primeira geração mostraram ter sérias limitações e logo foram eclipsadas pelas técnicas da inteligência artificial de "escrita de programas". Essas técnicas mais tarde revelaram ter suas próprias e graves limitações, como vimos no final do Capítulo 6, e os últimos anos presenciaram o renascimento do interesse pelo primeiro tipo de abordagem. As limitações iniciais foram superadas, e as redes neurais artificiais estão finalmente começando a revelar seu verdadeiro potencial. (CHURCHLAND, 2004, p. 244)

Compreendemos, com isso, que o connexionismo tem por base postular o funcionamento computacional como o meio pelo qual o cérebro funciona. Com isso, um

cérebro artificial composto por elementos distribuídos conforme um cérebro biológico se constitui, poderá, segundo se pensa, ser capaz de reproduzir o mesmo funcionamento cerebral.

A retomada do conexionismo e de seu projeto de simulação do cérebro nos meados dos anos 80 representou uma aposta no materialismo, ou seja, na idéia de que é a complexidade de certos sistemas físicos que os qualifica para produzir vida mental. Estados mentais emergem das redes: a ordem da programação é substituída pela própria organização ou auto-organização do hardware, ou, em outras palavras, estados mentais são a realização de certas disposições que exigem uma arquitetura específica da máquina. (TEIXEIRA, 1998, p. 104)

Pensando dessa forma, somos levados a entender que, segundo o conexionismo, a mente não age como um programa específico que gera os estados mentais, mas enxerga a rede pela qual é constituído o cérebro, como um sistema integrado de onde emergem os estados mentais. Esta complexidade organizacional do sistema cerebral é que é responsável pela produção de estados mentais.

Computadores e cérebros são sistemas cuja função principal é processar informação e, assim, podem-se utilizar redes artificialmente construídas para simular esse processamento. [...] Dado um determinado *input*, diferentes estados mentais podem ocorrer como consequência de mudanças nas conexões, que podem ser inibidas ou ativadas, variando de acordo com a interação do sistema com o meio ambiente e com seus outros estados internos. (TEIXEIRA, 1998, p. 84)

Nesse campo de visão, o que se quer defender, aqui, é que se entenda que estados mentais são estados materiais, são aspectos físicos gerados pela rede de comunicação interligada pela multiplicidade de neurônios.

Essa comunicação feita pelos neurônios capacita o sistema a funcionar desta forma específica, criando os estados mentais de forma contínua, a simulação artificial dessa rede de conexões é a tarefa a ser feita pelos sistemas conexionistas.

Neste modelo, os conteúdos mentais emergem da atividade das redes e suas conexões: caminhamos aqui numa direção inversa àquela do funcionalismo tradicional onde estados mentais são *atribuídos* a estados do *hardware*. Ou, para empregar uma terminologia filosófica, podemos afirmar que estados mentais são *supervenientes* à atividade das redes. A noção de emergência ou de superveniência que introduzimos aqui é compatível com a hipótese materialista que procuramos manter até agora e encontra paralelo numa série de fenômenos físicos cotidianos. A formação do gelo, após o resfriamento da água, constitui um bom exemplo do que estamos afirmando. Ninguém negaria que o gelo é água, mas, por outro lado, a ele podem ser atribuídos predicados que não se encontram na água, como “ser sólida”. A solidez é um predicado emergente que se forma a partir de leis físicas bem estabelecidas e conhecidas. Contudo, “ser sólido” não é um predicado que possamos atribuir a cada uma das moléculas de uma barra de gelo, mas só ao sistema físico como um todo. A mesma coisa podemos afirmar, *mutatis mutandis*, da emergência

de estados mentais a partir das configurações neuronais e da conectividade: estados mentais podem surgir destas últimas, mas dificilmente poderíamos atribuir suas propriedades a cada um dos neurônios considerados isoladamente. (TEIXEIRA, 1998, p. 105 - 106)

Um sistema conexionista tenta demonstrar que todo comportamento físico, incluindo o mental (e aqui compreende-se o mental como algo físico), deriva especificamente da programação exercida pela rede de conexões de um determinado sistema. Este comportamento pode sofrer variações e dar respostas diferenciadas ao mesmo tipo de estímulo sofrido.

A programação não segue um padrão de *output* para cada *input* recebido pelo sistema, depende das conexões feitas a cada instante, dentro da própria rede de conexões. Mesmo em sistemas que funcionem com o mesmo tipo de construção neural, exercendo funções do mesmo tipo, dará respostas variadas, já que as conexões realizadas entre os neurônios dentro da rede não seguem um padrão de previsibilidade sistemática.

[...] esta concepção de conectividade que possibilita que indivíduos inicialmente com uma mesma rede e recebendo um mesmo *input* possam formar conteúdos mentais diferenciados, seja do ponto de vista qualitativo, seja do ponto de vista da intensidade de uma determinada sensação: é impossível determinar a priori quais serão as conexões a serem ativadas e como será a distribuição dos pesos pela rede. Esta seria a contribuição dos modelos conexionistas para uma possível solução do problema dos *qualia*. Conteúdos mentais são, assim, essencialmente indetermináveis não apenas na medida em que não podemos antever quais as redes que serão ativadas por um *input* sensorial, como também pelo fato de se poder estabelecer uma diferença entre conteúdos sensoriais e conteúdos representacionais em diferentes indivíduos. Conteúdos representacionais, na medida em que supervêm à atividade da rede, podem se tornar privados e inescrutáveis. [...] (TEIXEIRA, 1998, p. 106)

É com este tipo de disposição não linear, ou seja, sem seguir sempre a mesma sequência de conexões entre os dados de *inputs* e *outputs*, que os estados mentais emergem e possibilitam as respostas funcionais do sistema em questão.

Vejamos a figura 3.1 abaixo, demonstrando o funcionamento de uma rede conexionista e como esta rede se comporta de uma maneira consistente e ao mesmo tempo indeterminada. Construída dessa forma, esta interface sistemática torna-se capaz de gerar respostas, identificando diferentes padrões dentro do sistema.

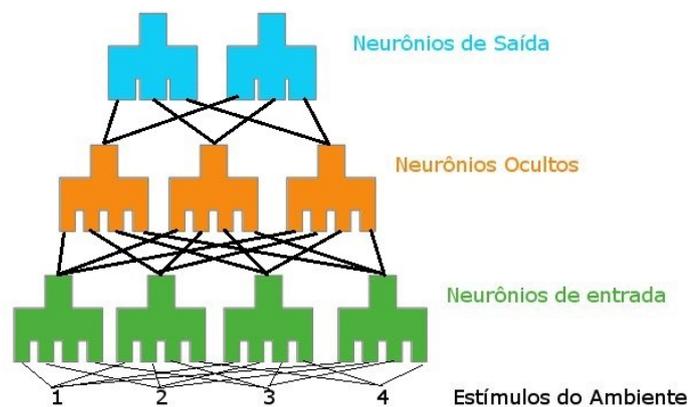


Figura 3.1 Pablo de Assis. 2009. “Uma rede neural é capaz de identificar diferentes padrões”. *O que são redes neurais?* Acedido a 21 de novembro de 2017. <https://www.tecmundo.com.br/programacao/2754-o-que-sao-redes-neurais-.htm>

Podemos entender, dessa forma, que a questão da consciência subjetiva se dá justamente pelo fato da rede neural gerar respostas individuais a cada sistema específico. Isso se dá devido a configuração única que cada sistema mental vai desenvolvendo dentro da rede, gerando assim estados mentais que apenas são (re)conhecidos pelo próprio sistema.

Para se construir modelos artificiais do cérebro que funcionem seguindo o mesmo princípio pelos quais os cérebros biológicos, mais particularmente o cérebro humano, desenvolvem sua composição funcional sistemática, faz-se necessário seguir algumas regras de programação às quais os sistemas artificiais devem se seguir. Teixeira descreve como deve ser essa composição sistemática.

A construção de sistemas conexionistas envolve os seguintes componentes:

1. Um conjunto de unidades de processamento.
2. Um padrão de conectividade entre as unidades.
3. Pesos (ou força) entre as conexões.
4. Uma regra de ativação que toma os *inputs* que recaem sobre uma unidade num determinado estado e os combina para produzir um novo nível de ativação para essa unidade.
5. Uma regra de aprendizado, a partir da qual padrões de conectividade mudam com a experiência. [...]

Há três tipos de unidades: unidades de *input*, unidades de *output* e unidades ocultas. De modo geral, as unidades de *input* recebem estímulos de fontes externas ao sistema. Contudo, unidades de *input* podem também receber estímulos oriundos do próprio sistema, ou seja, de outras unidades. As unidades de *output* enviam sinais para fora do sistema – sinais que podem afetar componentes motores acoplados a este último. Finalmente, as unidades ocultas são aquelas cujos *inputs* e *outputs* provêm do interior do sistema que estamos construindo. (TEIXEIRA, 1998, p. 85)

Logo, devemos entender que os modelos conexionistas de simulação do funcionamento do cérebro oferecem uma perspectiva muito favorável ao funcionamento correspondente do nosso cérebro. A construção de computadores com rede neurais artificiais poderá possibilitar, às futuras gerações de máquinas inteligentes, um desenvolvimento em que possamos compreender seu comportamento como uma projeção advinda dos seus estados mentais.

Apesar de todos os estímulos propiciados pelas teorias que nos dispusemos a pesquisar a respeito da possibilidade da simulação dos nossos estados mentais pelas máquinas, parece-nos que as objeções sobre o caso, sejam elas quais forem, caem muito mais na obscuridade sobre o nosso próprio conhecimento do que são esses estados mentais.

4. NAVEGANDO AO INFINITO

A tecnologia alcança cada vez degraus mais altos dentro da sociedade. Temos, hoje, a possibilidade de melhorar ou aperfeiçoar o corpo humano, embora ainda estejamos caminhando a passos lentos devido a diversas discussões éticas sobre o uso da tecnologia usada para esse melhoramento.

O uso de próteses e medicamentos artificiais já se mostram eficazes na capacidade de serem compatíveis e de se ajustarem ao nosso organismo, de forma que podem substituir várias partes do nosso corpo, assim como nos adequamos a estas sem grandes problemas. O avanço tecnológico está evoluindo cada dia mais rapidamente e somos absorvidos na mesma velocidade.

Muita coisa mudou na última década! Alguns marcos importantes sobre o avanço da tecnologia são:

2007: foi lançada a TV digital no Brasil, assim como a primeira geração do iPhone;

2008: foi lançado o sistema operacional Android, assim como o navegador Google Chrome. No mesmo ano, Bill Gates saiu da Microsoft para dedicar-se a filantropia;

2009: surge o WhatsApp! Também foi inaugurado o primeiro coworking no Brasil. No mesmo ano, o número de MEIs no Brasil chegou ao número de 44.000.

2010: estreia do filme Avatar, que ficou conhecido pela inovação de filmagens 3D e captura de movimentos. Ainda em 2010, Mark Zuckerberg foi eleito pela revista Time a personalidade do ano.

2011: chegada no Netflix no Brasil. O ano também ficou marcado pela morte de Steve Jobs, que faleceu aos 56 anos.

2012: o Facebook atingiu a marca de um bilhão e usuários! O cenário de startups também ganhou destaque.

2013: surgimento do Nubank. No mesmo ano, a Harvard Business Review fala sobre gestão ágil no Marketing;

2014: chegada do Uber e Spotify no Brasil. No mesmo ano, o Facebook comprou o WhatsApp.

2015: São Paulo é a primeira cidade brasileira a regulamentar o Uber.

2016: lançamento do Pokémon Go no Brasil.

2017: investimentos das empresas em cloud devem chegar a U\$ 4,5 bilhões. O número de microempreendedores individuais no Brasil chegou a 6.000.000! (SOUSA, 2017, n.p.)

Ademais dessas mudanças no meio econômico e social, no meio científico os avanços são ainda mais promissores. A imagem do homem-máquina está mais próxima do que nunca antes na história. De todas as possibilidades que vão surgindo, a que mais seduz (ou assusta) o homem é a tentativa de criarmos um cérebro artificialmente.

4.1. Tecnologia em evolução

Para os transumanistas, a aparente diferença entre o modo de pensar dos seres humanos e dos outros seres está para ser superada. Com a possível criação de um cérebro artificial poderemos replicar a mente, seja para a criação de máquinas conscientes ou mesmo para uma transferência da mente humana para seu respectivo software digital.

A tecnologia revolucionou nossa maneira de enxergar o mundo e a nós mesmos, melhorou nossa vida em diversos aspectos, mas seus benefícios ainda não são alcançados por todos. A capacidade com que as máquinas tornaram o trabalho menos árduo não trouxe consigo a liberdade e não necessidade de continuarmos a ter que trabalhar.

A construção de máquinas que possam trabalhar de forma mais efetiva, que possam substituir o trabalho pesado de muitas tarefas pelas quais temos que empreender em nosso tempo de vida, poderia nos deixar mais livres para preenchermos nosso tempo com atividades mais prazerosas ao nosso organismo, isto, no entanto, ainda não está acessível a todos.

Na contramão da abundância, há a consciência da infelicidade da parte daqueles que não tem acesso aos bens materiais e o sabem. Isso tem levado a uma frustração insuperável que se expressa na sensação de que o acesso ao conforto material nunca esteve tão próximo, mas, ao mesmo tempo, tão insuportavelmente distante. Essa frustração tem aumentado na medida em que se percebe que, embora não tenhamos o controle da história e das economias, nossa capacidade de influenciá-las não pode ser ignorada. (TEIXEIRA, 2015, p. 60 - 61)

A diferença social entre pobres e ricos ainda impõe um abismo na aquisição de tecnologia por todas as pessoas, mesmo o uso da internet ainda é algo que muitas pessoas, no mundo, não têm acesso, o que mostra como a tecnologia, embora em estágios cada vez mais avançados, ainda não é capaz de tornar a vida de todos algo mais prazerosa.

Esse pequeno detalhe sobre alguns aspectos da tecnologia citado, acima, serve para entendermos que o avanço da IA é algo que ainda não é muito bem compreendido por grande parte das pessoas, diria, na verdade, da maioria esmagadora das pessoas, embora, hoje, praticamente em quase todo o planeta, as pessoas se utilizem de algum tipo de IA.

Certamente, nossa relação com a tecnologia precisa mudar. Ainda não aprendemos a conviver com os artefatos tecnológicos que nos cercam e ficamos profundamente irritados quando a própria tecnologia nos expõe ao fato de que hoje somos quase totalmente dependentes dela. O que hoje temos, na verdade, é uma crítica leiga à tecnologia, motivada principalmente pela inadequação da interação homem-máquina na vida cotidiana. É dessa inadequação que surge a maior parte da sensação de um “inferno informático”, que não pode ser confundido com a ideia de que a tecnologia é intrinsecamente nociva. Infelizmente, o desconforto na nossa interação com as máquinas está longe de ser resolvido, e as pesquisas sobre interfaces homem-máquina caminham lentamente. *Notebooks, tablets*, celulares e páginas da internet ainda não são suficientemente amigáveis para grande parte da população. (TEIXEIRA, 2015, p. 68)

A criação de vários robôs que facilitam o meio de vida, hoje em dia, pode até passar despercebida, mas eles estão continuamente trabalhando, seja nas operações de comunicação, como os envios de mensagens usadas nos celulares ou em transações bancárias envolvendo inúmeras quantidades de dinheiro.

Devemos levar em consideração o fato de que a tecnologia já se tornou parte do mundo e nos adequamos cada vez mais a ela do que ela se adequa a nós. Por mais que haja pessoas cultivando e propagando os males que a tecnologia pode nos trazer, ela já está implantada em nosso cotidiano e não dará passos para trás. É a tecnologia que define como será nosso futuro e cabe a nós compreender e nos ajustar a esse futuro.

O velho debate ético sobre se a tecnologia é intrinsecamente nociva ou se é o uso do que vem dela que deve ser discutido não faz nenhum sentido. Não há um bom ou mau uso da tecnologia, pois sua utilização não é determinada por ninguém. O que existe é, simplesmente, uma *prática de uso* que a define temporariamente e também decide, por vezes de forma oportunista, sua finalidade. Nesse sentido, a tecnologia é criação tanto quanto as artes, ou seja, é *poiesis*. São as finalidades atribuídas aos inventos que os tornam poiéticos e, ao mesmo tempo, fora do nosso controle, pois, através do uso, eles se transformam em criações coletivas sem resultados previsíveis. (TEIXEIRA, 2015, p. 66)

O que não paramos para pensar muitas vezes, e deveríamos fazê-lo, é sobre como essas máquinas trabalham. Como elas se tornaram capazes de fazer trabalhos em um nível de capacidade muito superior ao nosso. O que de fato são máquinas? E qual o seu real papel dentro da sociedade humana, em que estas máquinas ingressam cada vez mais.

A busca pela criação de máquinas conscientes pretende não apenas revolucionar o modo de vida do ser humano, como também, possibilitar a capacidade de compreendermos a

nós mesmo como seres pensantes. Afinal ainda não sabemos como e porque pensamos e se outros seres também são capazes de projetar pensamentos.

Se a criação de mentes artificiais que formulem pensamentos próprios chegarem a se tornar realidade, estas respostas, sendo gratificantes ou não, podem ser respondidas. Poderão, entretanto, trazer mais dúvidas além das quais já temos hoje. Uma dúvida constante, por exemplo, é que se máquinas artificiais podem pensar, o que nos faria, então, diferente delas?

As questões éticas ainda são barreiras que tentam impedir a aceitação de tal ser como um ente consciente, pois se este não produz reflexões ou questionamentos sobre suas ações, também é incapaz de ser responsável pelos seus atos, independente de ter ou não consciência destes. Barreiras éticas também impossibilitam testes mais apurados em seres humanos para um aprofundamento maior no funcionamento do cérebro.

A simulação ou replicação de mentes conscientes ou semiconscientes levanta questões sobre o comportamento real do ser humano, levando a pensar mais seriamente sobre o fato de não sermos seres plenamente autônomos e que, possivelmente, apenas seguimos uma programação complexa de autoaprendizagem e resolução de problemas. Teixeira aponta para esta questão em uma de suas reflexões sobre a singularidade.

A irrupção da singularidade na paisagem planetária levanta muitas questões filosóficas e antropológicas. Afinal se uma máquina pode imitar a inteligência humana, o que é o homem? Será que a perda do monopólio sobre a inteligência no planeta significa descaracterizar a espécie humana? (TEIXEIRA, 2015, p. 71)

As definições que temos hoje sobre o conceito de pessoa, de ser humano, se refletem muito mais no fato de que somos seres que possuem uma mente consciente capaz de projetar pensamentos reflexivos sobre as ações que praticamos, e, assim, deliberar, de forma independente, nossos próprios conceitos sobre tudo que existe, assim como somos capazes de reconhecer as mesmas capacidades em seres da mesma espécie humana.

Esse reconhecimento da capacidade mental está relacionada unicamente aos membros da nossa espécie. Não evidenciamos o mesmo reconhecimento em seres de espécie diferentes, nem mesmo nos primatas mais próximos de nós na cadeia evolucionista.

No entanto, não possuímos um conhecimento concreto do que é o mental para poder responder, com certeza, se outros seres, realmente, podem ter ou não uma capacidade de gerar pensamentos, ainda que sejam em níveis mais baixo que os nossos.

As antropologias filosóficas falharam, até agora, em nos dar uma resposta convincente a essa questão. Talvez a resposta esteja precisamente na impossibilidade de responder a ela com base em uma característica especial que nos defina como humanos. O homem é um animal indefinido que pode, entretanto, imitar todos os outros, e a isso ele chama de suas *invenções*. Essa capacidade de imitação, que Aristóteles chamou de *mimesis*, define grande parte de nossa tecnologia, embora tenha algumas exceções cruciais, como por exemplo, a roda. A singularidade será, então, o grande momento: a imitação de todas as imitações, que permitirá, por fim, que imitemos a nós mesmos. (TEIXEIRA, 2015, p. 71)

O teste de Turing ainda traz muitas implicações sobre a capacidade de uma máquina poder ser reconhecida como capaz de simular os aspectos cognitivos de um ser humano doravante as respostas que a mesma produzir. Ainda seremos tentados a dizer que a máquina apenas cumpriu uma determinada programação e que as respostas já estavam delimitadas em diversos níveis desta programação.

Será que aceitaremos que máquinas podem possuir e gerar pensamentos similares aos nossos? Esta nova concepção sobre as capacidades mentais trará implicações inevitáveis. As máquinas terão que ter direitos e deveres dentro da sociedade iguais aos nossos? Ou será preciso criar mecanismos de proteção e de conduta diferentes para estes novos seres?

As formulações éticas que temos, hoje, sobre moralidade e comportamento social terão que ser englobadas em conjunto com as diretrizes por quais terão que passar as novas máquinas “pensantes”.

Dentro dessa perspectiva, ainda temos que entender como essas máquinas se comportarão com uma programação interna teoricamente ainda voltada a servir aos interesses do ser humano.

4.2. Transumanismo e a singularidade tecnológica

O que dizem e o que fazem os transumanistas? O que defendem e porque defendem? Precisamos entender o que é esta corrente de pensamento que visa transformar o mundo humano, preparando-o para a chegada do momento que os mesmos chamam de singularidade, em que atingiremos o pós-humano e o alcance da inteligência artificial similar ao da espécie humana.

Transhumanismo é um movimento vagamente definido que se desenvolveu gradualmente ao longo das últimas duas décadas. Ele promove uma abordagem interdisciplinar para a compreensão e avaliação das oportunidades para a melhoria da condição humana e do organismo humano proporcionadas pelo avanço da tecnologia. Atenção é dada tanto para tecnologias atuais, como engenharia genética e tecnologia da informação, quanto para aquelas antecipadas para o futuro, como a nanotecnologia molecular e a inteligência artificial. (SOUZA, 2014, n. p.)

Os adeptos do transumanismo pretendem ampliar o condicionamento humano através da tecnologia, melhorando sua composição física, com o ideal de diminuir as limitações pelas quais passa nosso corpo, com o objetivo de prolongar a vida, diminuindo o tempo de envelhecimento do corpo, assim como tornar nosso corpo imune às diversas doenças a que o mesmo é afetado.

O termo transumanismo foi criado pelo biólogo britânico Julian Huxley, irmão do escritor Aldous Huxley, em 1957. Huxley definia este conceito como “homem continuando homem, mas transcendendo, ao perceber novas possibilidades de e para sua natureza humana”. Os ideais transumanistas se fundamentam, basicamente, em dois pilares, de acordo com o filósofo brasileiro Gledinélcio Silva Santos. O primeiro deles seria o combate ao envelhecimento e, em consequência disso, a morte; o segundo, é o que trata a simbiose entre orgânico e cibernético como o próximo passo da evolução humana. (CIRIACO, 2015, n. p.)

Dentre as diversas possibilidades, poderemos assimilar nossos corpos com componentes artificiais que nos possibilitem ultrapassar certas dificuldades. Uma das idealizações é a junção entre homem e máquina, em que transformando-nos em ciborgues poderemos melhorar nossa condição física, superando as limitações naturais do nosso organismo.

As opções de aprimoramento a serem discutidas incluem a extensão radical da saúde humana, a erradicação da doença, a eliminação do sofrimento desnecessário e o aprimoramento das capacidades intelectuais, físicas e emocionais dos humanos.

Outros temas transhumanistas incluem a colonização espacial e a possibilidade de criar máquinas superinteligentes, juntamente com outros desenvolvimentos em potencial que poderiam alterar profundamente a condição humana. O âmbito não está limitado a dispositivos tecnológicos e medicamentos, mas abrange também talentos e técnicas econômicas, sociais, de design institucional, de desenvolvimento cultural e psicológicas. (SOUZA, 2014, n. p.)

Segundo os defensores do ideal transumanista, além da tecnologia possibilitar o melhoramento da nossa condição física, podemos, através dela (tecnologia), alcançarmos um estado de bem estar muito mais promissor do que atualmente podemos alcançar, bem como usarmos a tecnologia para vencer os obstáculos que o corpo biológico encontra naturalmente durante seu tempo de vida e que acaba impedindo um avanço mais rápido na sua evolução.

Em contraste com muitas outras perspectivas éticas, que na prática muitas vezes refletem uma atitude reacionária às novas tecnologias, o ponto de vista transhumanista é guiado por uma visão em constante evolução no sentido de adotar uma abordagem mais pró-ativa sobre a política tecnológica. Essa visão, em linhas gerais, é a de criar a oportunidade de viver uma vida muito mais longa e saudável, para melhorar a nossa memória e outras faculdades intelectuais, para refinar as nossas experiências emocionais e aumentar a nossa sensação subjetiva de bem-estar, e para alcançar um grau maior de controle sobre nossas próprias vidas de uma maneira geral. (SOUZA, 2014, n. p.)

Com a possibilidade da criação de cérebros artificiais que funcionem similares ao cérebro humano, poderemos ser capazes de simular e replicar os estados mentais, possibilitando-nos fazer, talvez, transferências de mentes ou cérebros para seus respectivos receptáculos artificiais.

Nós melhoramos o nosso sistema imunológico natural com aplicações de vacinas, e podemos imaginar melhorias adicionais para os nossos corpos que nos protegeriam de doenças ou nos ajudariam a moldar o nosso corpo de acordo com nossos desejos (por exemplo, deixando-nos controlar a taxa metabólica do nosso corpo). Tais aprimoramentos poderiam melhorar a qualidade de nossas vidas. Um tipo mais radical de melhoria pode ser possível se supusermos uma visão computacional da mente. Seria então possível fazer um *upload* de uma mente humana para um computador, através da replicação em silício dos detalhados processos computacionais que normalmente ocorreriam em um cérebro humano em particular. Ser um *upload* teria muitas vantagens em potencial, tais como a habilidade de fazer cópias de segurança de si mesmo (impactando favoravelmente na expectativa de vida) e a capacidade de transmitir a si mesmo como informação na velocidade da luz. *Uploads* podem viver tanto em realidade virtual ou diretamente na realidade física, controlando um *proxyrobô*. (SOUZA, 2014, n. p.)

Se conseguirmos encontrar o local onde se localiza a mente em nosso cérebro através das simulações dos estados mentais em cérebros artificiais, a capacidade de podermos fazer um *upload* de nossa mente para a internet e abandonarmos nosso corpo físico poderá se

efetivar, e, com isso, poderemos alcançar o ideal de imortalidade que tanto fascina a humanidade.

A possibilidade de ter a mente transformada em um *software*, que pode estar tanto em um recipiente físico quanto em um mundo virtual dentro da internet, seria um passo a frente do modo de vida que levamos hoje em dia. Vivemos ainda em um corpo físico, mas já passamos parte dos nossos dias com a mente voltada para o mundo virtual, no entanto, ainda com várias limitações em relação a este mundo.

O upload na internet não é uma questão filosófica para os transumanistas, pois, da mesma maneira que nossos corpos são renovados a cada sete anos pela total substituição de células, e sem perder a identidade pessoal no tempo, podemos também imaginar a existência de várias versões de nós mesmos no espaço. Se nos tornarmos holografias, poderemos ser projetados simultaneamente para diversos lugares do espaço, não só na internet, como também em outros universos do multiverso. (TEIXEIRA, 2015, p. 77)

Esse tipo de ideal tecnológico, segundo os transumanistas, será alcançado em breve, a questão não seria se, mas quando. Será apenas uma questão de tempo para que a singularidade seja alcançada e esse processo, segundo os mesmos, é irreversível. Isso se deve ao aumento exponencial de poder de processamento por qual passam os *hardwares* dos computadores digitais.

*Kurzweil, figura de destaque no movimento transumanista e um dos maiores divulgadores da ideia de singularidade, acredita que ela é um fenômeno praticamente inevitável que decorre do aumento exponencial do poder computacional do *hardware* dos computadores digitais. Esse aumento é consequência da Lei de Moore. Gordon Moore, um dos inventores dos circuitos integrados, notou na década de 1970, que o número de transistores que pode caber num *chip* dobra a cada dezoito meses. Em outras palavras, o poder dos computadores dobra a cada dezoito meses. (TEIXEIRA, 2015, p. 74)*

A evolução da tecnologia, em um espaço de tempo cada vez mais rápido, trará conseqüentemente, ao mesmo tempo, uma similaridade cada vez mais significativa entre homem e máquina. Se pudermos ter a possibilidade de transcender nossa mente a um nível mais elevado, em contrapartida o mesmo deverá acontecer com as máquinas.

O aumento exponencial da capacidade das máquinas, correlacionadas com programas de simulação mental cada vez mais sofisticados, poderá dar às máquinas a capacidade de

processarem estados mentais em um nível de complexidade cada vez mais similares ao do cérebro humano.

Programas de IA mais eficientes poderão facilitar a nossa vida, incluindo a capacidade de formularem mecanismos que possibilitem a criação de meios de hibridação entre homens e máquinas. Essa hibridação, ou seja, a criação de ciborgues, dará uma melhoria significativa nas condições de vida para o novo homem.

Nesse cenário, entretanto, voltaremos a questões significativas para o ser humano, como a questão da consciência e da significação do conceito de pessoa, que objetivamos durante a nossa existência.

Nossos estados mentais poderiam ser interpretados apenas como programações internas a que nenhum de nós tem conhecimento, pelo menos, por enquanto, de como são efetivadas, da mesma maneira que as máquinas exercem suas funções.

Ainda não seremos capazes, mesmo se transferências mentais ou simulações artificiais cada vez mais próximas aos modelos biológicos se tornarem efetivas, de resolvermos o problema da mente consciente. Mesmo descobrindo se a mente surge de forma física em um cérebro biológico ou artificial.

[...] O que haverá, afinal, acerca da consciência, além do fato de dizermos que somos conscientes? Nós, humanos, nos tornamos conscientes aos nos proclamarmos criaturas dotadas de consciência. Fazemos isso por meio da linguagem, o que não acontece com os animais e ainda não aconteceu com os robôs. Será que existe algo mais na consciência do que um comportamento verbal através do qual proclamamos que somos conscientes? (TEIXEIRA, 2015, p. 80)

O que de fato define que temos consciência do que fazemos ou do que pensamos? Não seria tudo o mais, apenas programação mental destinada a manter o organismo em constante evolução, procurando meios de manter sua sobrevivência dentro do plano existencial? Essa programação seria a responsável, talvez, pela própria vontade do homem de procurar uma evolução tecnológica a ponto de conseguir suplantar a morte.

Máquinas inteligentes construídas até agora, como o DEEP BLUE e o WATSON, se destacaram pela capacidade de jogar xadrez e responder a questões sobre conhecimentos gerais. São extraordinários dispositivos de raciocinar, reconhecer

padrões e identificar regularidades. Mas serão elas capazes de identificar o significado dessas regularidades para torna-las inteligíveis? Será isso possível sem que essas máquinas sejam conscientes? Mesmo considerando que a inteligência dessas máquinas não precisa ser idêntica à humana, até que ponto a incapacidade de interpretação de dados e regularidades seria uma limitação? (TEIXEIRA, 2015, p. 80)

O argumento do quarto chinês de Searle ainda nos deixa mais dúvidas sobre nossas próprias capacidades. Somos, realmente, conscientes do que fazemos? Quando constituímos pensamento e realizamos ações, temos real consciência de como e porque o realizamos?

Será que não somos, assim como qualquer máquina, constituídos de uma programação que não nos é conhecida e que nos faz crer que somos responsáveis pelos nossos pensamentos e ações, quando, talvez, sequer saibamos que somos regidos por algum tipo de programação interna?

Estas perguntas continuam a ser inquietantes. Segundo Searle, somos nós que damos significados aos estados mentais e não entendemos as máquinas como capazes de ter a mesma capacidade para tal realização. Como saber se máquinas podem realmente serem capazes de simular a consciência, se elas mesmas não forem capazes de dar significado ao que fazem?

A consciência é parte fundamental do significado. Não há estados intencionais inconscientes, nem tampouco, compreensão inconsciente, pois Searle não admite que haja mente sem consciência. Em seu argumento do Quarto Chinês, o ocupante do quarto manipula os símbolos sem nenhuma consciência, pois não sabe o que está fazendo. Como ele não tem consciência nem estados intencionais, somos nós que atribuímos compreensão, metaforicamente, ao processo de manipulação dos símbolos em chinês. (TEIXEIRA, 2015, p. 96)

O que a singularidade não responde, e parece não levar isso em relevância, é o que de fato é ser consciente. O que faz com que um ser, seja ele biológico ou artificial, tenha relativa certeza que é um ser capaz de deliberar pensamentos de forma livre e independente? Mesmo máquinas que cheguem a alcançar uma inteligência similar a nossa ainda não darão respostas a esta questão.

[...] O argumento do Quarto Chinês também se aplica a singularidade. A representação formal do cérebro humano na forma de um *software* vagando na internet não pode produzir estados mentais conscientes. A simulação do cérebro não é sua replicação, e, por isso, seu *upload* na internet a partir de uma descrição digital geraria apenas um imenso cemitério virtual povoado por holografias de seres humanos. Imagens não tem vida. A base da intencionalidade e da consciência é a vida. Sem um cérebro, elas não seriam possíveis. (TEIXEIRA, 2015, p. 97)

Mas a definição de consciência é algo realmente relevante para uma máquina que possa ser capaz de alcançar níveis de processamento mental igual ou superior ao de um ser humano? Consciência poderia ser apenas um subproduto dos estados mentais, parte da programação que facilita nossa maneira de tentar entender o mundo e poder sobreviver nele.

O problema da consciência está no fato de ser um aspecto subjetivo da mente, em que cada um de nós tem acesso apenas ao conteúdo individual, sem a possibilidade de ter compreensão do que se passa nos processos subjetivos da mente de outro ser.

Se não temos acesso à subjetividade de outras mentes, como sabemos, com certeza, que estes outros agentes são, de fato, conscientes do que fazem, apenas pelo fato de serem idênticos a nós biologicamente?

Pessoas com deficiência cerebral, na maioria das vezes, não são consideradas capazes de ter consciência do que estão fazendo. Essas pessoas, no entanto, produzem estados mentais e ainda são consideradas iguais a outras.

As técnicas atuais de neuroimagem permitem detectar com grande precisão as áreas cerebrais que são ativadas durante vários tipos de estados mentais. Contudo, até agora ainda não foi possível detectar o que torna um estado mental consciente, nem tampouco, os correlatos neurais da compreensão. Certamente quando ouvimos ou lemos uma sentença em português, ocorre algum tipo de atividade cerebral correspondente, mas isso não implica que ela esteja sendo compreendida. Um mesmo grupo de neurônios pode ser ativado quando há compreensão ou quando ela não ocorre. (TEIXEIRA, 2015, p. 97)

Podemos entender, dessa forma, que o problema da consciência em si pode não ser o único aspecto que determina o conceito de pessoa ou não. Talvez a capacidade de sentir seja mais inerente nesse contexto. Mas sentimentos também são subjetivos e estão correlacionados com a consciência. Afinal, para se ter compreensão do que se está sentindo é necessário ter a consciência de que se está sentindo.

Mas, para os transumanistas, esses detalhes parecem ser irrelevantes, quando o principal propósito é o melhoramento da condição humana. E, se possuímos os meios para se chegar a essa melhoria com a ajuda de máquinas, a preocupação com questões sobre a

consciência robótica e humana não pode ser um entrave para que essa melhoria seja alcançada.

4.3. Simulações do novo mundo

Apesar da aposta e de todo o entusiasmo transumanista na chegada da singularidade e dos benefícios em larga escala que esse acontecimento deva trazer, ainda estamos em um âmbito bem distante de tal acontecimento.

Embora os transumanistas apostem em uma revolução tecnológica que chegue em poucos anos, muitos acontecimentos podem retardar a chegada de tal momento, podendo até mesmo transformar tal ideal em apenas uma utopia *ciberpunk*.

O mito grego de Ícaro é um bom indicativo de como a imprudência humana em relação à tecnologia pode nos reservar um destino nada agradável. Será que o transumanismo vai nos levar a uma utopia de saúde plena ou a uma distopia sinistra e apocalíptica? É lógico que, ao menos em um primeiro momento, ninguém defende abertamente a ideia de se criar *RoboCops* pelo mundo. Contudo, a reflexão e a ponderação crítica sobre o tema podem, sem dúvida, ajudar a evitar futuras catástrofes. (CIRIACO, 2015, n. p.)

Fica claro que, com as descobertas e avanços atuais da tecnologia, um melhoramento na condição humana, visando o bem estar do ser humano já é uma realidade. Várias pesquisas em robótica já trazem resultados significativos na adoção de implantes artificiais em conexão com nosso organismo.

O melhoramento da IA torna-se cada vez mais evidente, com o crescimento da mesma dentro do cotidiano da sociedade. Carros autômatos já são fabricados, robôs trabalhando como atendentes em hotéis no Japão, assistentes pessoais virtuais em celulares e computadores são exemplos de como o avanço da IA já é uma realidade.

Alguns exemplos da IA aplicada como assistente virtual em dispositivos móveis, usados por inúmeras pessoas no dia a dia, são a Siri da Apple, a Cortana da Microsoft, o Google Assistente e a Alexa da Amazon. Temos supercomputadores como o Watson da IBM

e robôs com programas de aprendizado e comunicação mais avançados como a IA Sophia desenvolvido pela Hanson Robotics.

O grande avanço da IA corresponde, claro, aos interesses econômicos das grandes empresas mundiais, que aplicam os avanços da tecnologia em busca de negócios que gerem lucros, independente se o uso dessa tecnologia trará benefícios reais para a sociedade, embora fique claro que muitos benefícios podem ser gerados em diversas áreas de atuação.

As transformações trazidas por máquinas inteligentes e algoritmos estão apenas no começo. Gigantes como Amazon, Facebook, Google, IBM e Microsoft apostam suas fichas na evolução do conceito, que deve ser um ponto de inflexão no horizonte da tecnologia da informação. “O impacto anual gerado pelo desperdício de recursos no mundo é de US\$ 2 trilhões. Olhando para isso, é inegável o potencial trazido por ferramentas de inteligência artificial na criação de negócios mais efetivos, reduzindo gastos e tornando estruturas mais eficientes”, comenta David Kenny, general manager responsável pelo Watson. [...]“O Watson não dará respostas, mas ajudará a formular as questões certas para que os humanos encontrem melhores soluções”, defende Kenny. “Ele vai ser parte do processo criativo”, projeta, sobre o que enxerga com relação ao futuro de sistemas de inteligência artificial. Em outras palavras, o executivo não acredita que a tecnologia encontrará a cura do câncer, mas ajudará a economizar tempo e a direcionar as pesquisas dos cientistas. Com o uso da inteligência artificial, as pessoas terão tempo para ir mais a fundo em diversas disciplinas do conhecimento. (DREHER, 2016, n. p.)

Sophia tem sido a IA mais controversa dos últimos anos, graças a sua capacidade de responder perguntas de forma quase tão natural quanto uma pessoa comum. Dada a sua complexidade e notoriedade, Sophia pode ser um exemplo claro de como a singularidade tecnológica é algo que realmente vai acontecer, ou podemos dizer, já está acontecendo.

Tem aspecto de uma mulher adulta e foi inspirada na atriz britânica Audrey Hepburn, mas Sophia “nasceu” apenas a 19 Abril de 2015 – foi activada é mais o termo –, quando a Hanson Robotics apresentou aquele que é, até ao momento, o seu mais sofisticado robô humanóide, capaz de aprender e adaptar-se aos seres humanos, conseguindo interagir com eles. Dotada de inteligência artificial (IA), Sophia foi concebida para ajudar em lares para pessoas idosas ou para prestar informações em parques ou museus. No último ano, Sophia já foi entrevistada pelas principais cadeias de televisão do planeta e apareceu numa série de eventos, mas continua no seu processo de aprendizagem, ela que recorre ao sistema de reconhecimento de voz da Google (mais precisamente da Alphabet Inc, empresa que pertence à Google e que também é responsável pelo desenvolvimento do automóvel autónomo) para alimentar com dados o *software* de IA, criado pela SingularityNET. A capacidade da Sophia continua a evoluir a bom ritmo e, se no início de outubro esteve presente numa conversa com a vice-secretária geral das Nações Unidas, Amina Mohammed, no final do mês foi a estrela do Future Investment Summit, em Riade, onde lhe foi concedido o título de cidadã da Arábia Saudita. (LAVRADOR, 2017, n. p.)

Em 11 de outubro de 2017, Sophia tornou-se a primeira inteligência artificial a receber o título de cidadania oficial de um país. A Arábia Saudita foi o país em questão que concedeu

à Sophia o título de cidadã, o que trouxe muitas repercussões pelo mundo, dado que, com o título de cidadã, Sophia terá mais direitos do que mulheres e estrangeiros, humanos comuns residentes na Arábia Saudita.

A implementação de seres artificiais dentro da sociedade, nos faz refletir sobre como poderemos tratar as atuais e futuras inteligências artificiais que estão e estarão entre nós. Como devemos tratar robôs como Sophia, por exemplo, depois que a mesma já se tornou uma cidadã saudita, e por conseguinte, deve usufruir de alguns direitos como qualquer outro cidadão?

Transhumanistas defendem o bem-estar de toda a senciência, seja em intelectos artificiais, seres humanos e animais não-humanos (inclusive espécies extraterrestres, se houver alguma). Racismo, sexismo, especismo, nacionalismo beligerante e intolerância religiosa são inaceitáveis. Além dos motivos habituais para considerar tais práticas condenáveis, há também uma motivação especificamente transhumanista para isso. A fim de se preparar para um momento em que a espécie humana comece a se ramificar em várias direções, precisamos começar a incentivar fortemente o desenvolvimento de sentimentos morais que sejam suficientemente amplos para abranger, dentro da esfera de preocupação moral, seres sencientes que sejam constituídos de maneira diferente de nós mesmos. (SOUZA, 2014, n. p.)

Será o caso de nos prepararmos para uma nova visão sobre os meios de conduta social entre humanos e robôs, para quando robôs com IA mais avançados forem criados e estiverem dividindo com mais intensidade a vida cotidiana de forma mais prática, saibamos salientar os direitos e deveres de tais seres artificiais.

O que se pergunta muitas vezes é, porque máquinas mais inteligentes, que possam “viver” independente do auxílio de seres humanos e sejam capazes de adquirir consciência das tarefas que executam, estariam dispostas a servir seres que poderiam desligá-las a qualquer momento, por algum tipo de ato que considerassem uma “falha mecânica”?

Se perdermos o controle sobre a produção de superinteligências e elas se espalharem pela internet de forma viral, o que é muito provável, precisaremos de defesas para que possamos conviver com elas e evitar a possibilidade de sermos destruídos. A criação de leis para inibir a destruição mútua e estabelecer um convívio equilibrado com as máquinas pode ser insuficiente, pois não há como punir um dispositivo artificial. Por exemplo, é inimaginável punir um robô que tenha um corpo apenas virtual. Como ele poderia ser aprisionado? Como ele poderia ser punido ou submetido a uma situação vergonhosa? (TEIXEIRA, 2015, p. 112)

Seria o caso de implementarmos um código de ética bem definido na programação das máquinas e, mesmo assim, ainda correríamos o risco de que alguma máquina chegasse a mudar as diretrizes do programa, já que passariam a aprender e reconstruir sua própria programação como forma de melhorar seu aperfeiçoamento, assim como nós humanos fazemos com o que aprendemos durante o decorrer de nossa vida.

Uma solução aparentemente simples para deter eventuais agressões de robôs contra seres humanos é inserir um ponto cego na sua programação, uma palavra que, quando pronunciada, desligasse o robô imediatamente. No entanto, robôs podem identificar e modificar a programação de outros e, com isso, remover esses pontos cegos. Uma máquina superinteligente sempre pode burlar as leis ou qualquer outro tipo de restrição imposta por seres humanos. (TEIXEIRA, 2015, p. 112)

Será necessário criar ou reformular os códigos de ética que regem as sociedades hoje em dia, assim como fazer revisões nas diretrizes dos direitos humanos, com a implementação dos direitos dos robôs? Talvez tudo isso tenha que ser feito, se realmente chegarmos a ter máquinas que simulem os estados mentais e sejam capazes de interagir com humanos de forma prática.

Recentemente, o termo *roboética* passou a ser utilizado para designar a ética específica que deve reger as relações entre humanos e máquinas inteligentes. O termo foi criado pelo cientista italiano Gianmarco Veruggio, que, em 2002, era líder do departamento de Robótica do Conselho Nacional de Pesquisa da Itália. A Coreia do Sul, num projeto liderado pelo Ministério do Comércio, decidiu elaborar um código de ética para as máquinas inspirado nas três leis da robótica propostas por Isaac Asimov. Em 2004 foi redigida e aprovada a declaração Mundial dos Direitos e deveres dos Robôs em Fukuoka, no Japão, estabelecendo os direitos e os deveres recíprocos que devem existir entre humanos e robôs. (TEIXEIRA, 2015, p. 114)

O avanço da singularidade busca trazer uma melhora no bem estar da vida dos seres humanos, buscando prolongar a vida e levar as pessoas a terem um grau de felicidade e satisfação durante sua nova forma de viver. As máquinas com nível superior de inteligência seriam capazes de ajudar os seres humanos a chegarem a essa nova mudança de vida.

Além do aprimoramento das máquinas, teremos a possibilidade de melhorar nossas capacidades, modificando nosso organismo em conjunto com próteses que nos garantam viver por mais tempo e com mais saúde, além do que já falamos, sobre tentativas de *upload* da mente para mundos virtuais na internet.

Restará, para as futuras gerações de filósofos, a elaboração de uma roboética consistente. Instaurá-la nos futuros robôs dará a oportunidade de verificar a efetividade de suas máximas, que deixariam de existir apenas na cabeça dos filósofos. Pois, será que princípios éticos se concretizaram em algum momento da história da humanidade? (TEIXEIRA, 2015, p. 119)

De todo modo, não há garantias de qualquer tipo de bem aventurança que possamos alcançar pelo avanço exponencial que a tecnologia está nos trazendo. A evolução da tecnologia é realmente irreversível, mas como iremos modificar nossa maneira de ver o mundo e a nós mesmos é algo que nenhum de nós pode responder no momento.

5. CONCLUSÃO

O objetivo desse trabalho foi realizar um estudo sobre o avanço da inteligência artificial nas últimas décadas, a importância da mesma para pesquisas sobre o funcionamento do cérebro e de replicação mental. Refletir sobre as adversidades filosóficas e sociais advindas do aumento exponencial da tecnologia, assim como tentar compreender a crescente ideologia transumanista e seu impacto na sociedade.

Durante a pesquisa, foi possível entender como o problema da compreensão da mente modificou profundamente a maneira pela qual objetivamos a maneira de nos enxergarmos no mundo e de compreender-nos como seres pensantes e dotados de consciência. Em busca de respostas às questões da mente e do cérebro, propiciamos às máquinas a possibilidade de, igualmente, se tornarem seres conscientes. Ainda não sabemos definir se a mente é um componente físico de nosso intrincado sistema cerebral ou algo que ainda está além da nossa compreensão.

Por mais que possamos vir a saber acerca de onde surgem os estados subjetivos da mente, justamente estes estados subjetivos continuarão sem uma explicação plausível sobre onde realmente eles estão. Não há como verificar, de forma objetiva, por nenhum meio possível ainda, onde se encontram, fisicamente, os aspectos subjetivos da mente, como meus pensamentos, que só alcançados por um eu subjetivo, assim como aquilo que estou sentindo e ao qual somente eu tenho acesso.

Por essa razão, a aposta da singularidade não pode ser vista, pelo menos por enquanto, como algo que irá se realizar em poucos anos, que máquinas alcancem a capacidade de simular a consciência em um grau de similaridade a dos seres humanos. Melhorias para a sociedade com a crescente evolução da inteligência artificial são reais, mas a possibilidade de máquinas conscientes é algo que ainda não podemos dar respostas efetivas.

REFERÊNCIAS

- BLACKBURN, Simon. **Dicionário Oxford de Filosofia**. Consultoria da edição brasileira, Danilo Marcondes. Tradução de Desidério Murcho. Rio de Janeiro: Zahar, 1997.
- CIRIACO, D. **Transumanismo e o uso da tecnologia para aprimorar a condição humana**. Canaltech, 2015. Disponível em: <<https://canaltech.com.br/ciencia/transumanismo-e-o-uso-da-tecnologia-para-aprimorar-a-condicao-humana-49223/>>. Acesso em: 19 dez. 2017.
- CHURCHLAND, Patrícia. **Poderá a neurobiologia ensinar-nos alguma coisa acerca da consciência?** Artigo publicado em 1993 na American Psychological Association. Tradução de Luís M. S. Augusto, Sorbonne/FCT – 2005.
- CHURCHLAND, Paul M. **Matéria e Consciência: Uma introdução contemporânea à filosofia da mente**. Tradução de Maria Clara Cescato. São Paulo: Editora Unesp, 2004.
- DAMÁSIO, Antonio. **O erro de Descartes: emoção, razão e o cérebro humano**. São Paulo: Companhia das Letras, 1996.
- DESCARTES, René. **Discurso do método**. Coleção Os pensadores. Trad. Enrico corvisieri. São Paulo: Nova Cultural, 1999.
- DREHER, Felipe. **No Brasil, líder do IBM Watson traça rumos da computação cognitiva**. Computerworld, 2016. Disponível em: <<http://computerworld.com.br/no-brasil-lider-do-ibm-watson-traca-rumos-da-computacao-cognitiva/>>. Acesso em: 20 dez. 2017.
- LAVRADOR, Alfredo. **Sophia: robô humanóide atirou-se a Elon Musk**. Observador, 2017. Disponível em: <<http://observador.pt/2017/11/01/sophia-robo-humanoide-atirou-se-a-elon-musk/>>. Acesso em: 20 dez. 2017.
- LENT, R. **Cem Bilhões de Neurônios: Conceitos Fundamentais de Neurociência**. Rio de Janeiro: Atheneu, 2010
- NAGEL, T. **“What is it like to be a bat?”** In: Rosenthal, D. (ed.) *The Nature of Mind*. New York: Oxford University Press, 1991, p. 422-28. Este artigo foi publicado, originalmente, em 1974. Tradução de Paulo Abrantes e Juliana Orione, Cad. Hist. Fil. Ci., Campinas, Série 3, v. 15, n. 1, p. 245-262, jan.-jun. 2005.
- SOUSA, Ana Karoline. **Avanço tecnológico – entenda as principais mudanças no cenário tecnológico nos últimos 10 anos**. Envision Tecnologia, 2017. Disponível em: <<http://envisiontecnologia.com.br/avanco-tecnologico-entenda-as-principais-mudancas-no-cenario-tecnologico-nos-ultimos-10-anos/>>. Acesso em: 19 dez. 2017.
- SOUZA, Gustavo R. de. **Valores Transhumanistas**. Transhumanismo Universidade Federal do ABC, 2014. Disponível em: <<https://transhumanismocts.wordpress.com/2014/01/11/99/>>. Acesso em: 19 dez. 2017.
- TEIXEIRA, João de F. **O cérebro e o robô: Inteligência artificial, biotecnologia e a nova ética**. São Paulo: Paulus Editora, 2015.

TEIXEIRA, João de F. **Mentes e máquinas: uma introdução à ciência cognitiva.** Porto alegre: Artes Médica, 1998.

TEIXEIRA, João de F. **O que é filosofia da mente.** 2ª ed. São Porto Alegre: Editora Fi, 2016.

TEIXEIRA, João de F. **O que é inteligência artificial.** 2ª ed. São Porto Alegre: Editora Fi, 2017.