POIESIS
REVISTA DE FILOSOFIA
ISSN:2448-3095

A CONSCIÊNCIA EM ROUSSEAU E A INTELIGÊNCIA ARTIFICIAL

CONSCIOUSNESS IN ROUSSEAU AND ARTIFICIAL INTELLIGENCE

Sérgio Bernardo de Almeida[1]

Resumo: O presente artigo tem como objetivo trazer para o debate contemporâneo a concepção de consciência em Rousseau para poder discutir com as investigações contemporâneas sobre a IA. Pretende-se investigar a relação entre a noção de consciência em Rousseau e as possibilidades e desafios da inteligência artificial (IA). Partindo de uma perspectiva histórica e filosófica, cogita-se analisar como Rousseau concebeu a consciência como um sentimento interior que orienta o homem na busca da verdade, da moralidade e da felicidade, e como essa ideia se relaciona com as teorias e práticas da IA, que visa simular ou reproduzir aspectos da cognição humana por meio de algoritmos e máquinas. A hipótese central é que a consciência em Rousseau não pode ser reduzida a um mecanismo ou a um conjunto de regras, mas envolve uma dimensão afetiva, criativa e transcendente que desafia os limites e os riscos da IA.

Palavras-Chave: Consciência. Inteligência Artificial. Rousseau. Filosofia.

Abstract: This article aims to bring Rousseau's conception of consciousness into the contemporary debate in order to discuss it with contemporary investigations into AI. The aim is to investigate the relationship between Rousseau's notion of consciousness and the possibilities and challenges of artificial intelligence (AI). Starting from a historical and philosophical perspective, we intend to analyze how Rousseau conceived consciousness as an inner feeling that guides man in the search for truth, morality and happiness, and how this idea relates to the theories and practices of AI, which aims to simulate or reproduce aspects of human cognition through algorithms and machines. The central hypothesis is that consciousness in Rousseau cannot be reduced to a mechanism or a set of rules, but involves an affective, creative and transcendent dimension that challenges the limits and risks of AI.

Keywords: Consciousness. Artificial Intelligence. Rousseau. Philosophy.

[1] Mestre em Filosofia pela Universidade Federal da Bahia. Email: bernardofm@hotmail.com

Introdução

O objetivo deste artigo é trazer para o debate contemporâneo a concepção de consciência em Rousseau para discutir com as investigações contemporâneas sobre a Inteligência Artificial (IA). Pretende-se investigar a relação entre a noção de consciência em Rousseau e as possibilidades e desafios da IA. Partindo de uma perspectiva histórica e filosófica, pretende-se analisar como Rousseau concebeu a consciência como um sentimento interior que orienta o homem na busca da verdade, da moralidade e da felicidade, e como essa ideia se relaciona com as teorias e práticas da IA, que visa simular ou reproduzir aspectos da cognição humana por meio de algoritmos e máquinas. A hipótese central é que a consciência em Rousseau não pode ser reduzida a um mecanismo ou a um conjunto de regras, mas envolve uma dimensão afetiva, criativa e transcendente que desafia os limites e os riscos da IA.

Diante da importância do tema, se faz necessário, uma análise conceitual e histórica da noção de consciência em Rousseau, considerando seus aspectos epistemológicos, morais, políticos e religiosos, suas fontes e influências filosóficas. Bem como, uma comparação crítica entre a consciência em Rousseau e a IA, abordando questões como: o que significa ter consciência? O que é a consciência para Rousseau? Como surge a consciência? A consciência é uma propriedade exclusiva dos humanos ou pode ser atribuída a outros seres ou entidades? É possível simular a consciência em uma máquina?

Este artigo pretende contribuir para o debate interdisciplinar sobre as implicações filosóficas e sociais da IA, bem como para o resgate e a atualização do pensamento de Rousseau. Além do mais, pretende-se contribuir para o entendimento da consciência humana a partir da filosofia de Jean-Jacques Rousseau e sua relevância para as discussões contemporâneas sobre IA. Contudo, busca fornecer *insights* críticos sobre os desafios éticos, sociais e filosóficos da criação de sistemas conscientes de IA, promovendo uma reflexão aprofundada dessas tecnologias na sociedade e na natureza humana.

1. Consciência das Máquinas

Desde o fim da segunda guerra mundial (1945) teóricos e pesquisadores das Inteligências Artificiais (IA) acreditam que a consciência exibida por sistemas biológicos poderia ser replicável em instâncias não biológicas, ou seja, artificiais. Essa é uma questão

polêmica que já se apresentava entre os pensadores desde o século XVII, como é o caso de René Descartes e Thomas Hobbes, por exemplo. Descartes acredita que enquanto corpo orgânico, o homem é animal, o que significa descrevê-lo como uma máquina, mais complexa certamente e diferente dos outros sistemas materiais, e tudo quanto ocorre nesta máquina deve ser fisicamente explicado. O começo do *Tratado do Homem* explica claramente essa sua concepção.

Imaginaremos, diz o autor, "homens em tudo a nós semelhantes, mas consideraremos, inicialmente, neles apenas uma máquina sem alma, sendo esta, como se sabe, realmente distinta do corpo" (Descartes, 1973, p.21). Contrariando a visão da escolástica, para a qual toda organização se reduz a alma, Descartes pretende explicar a fisiologia animal, a partir da circulação do corpo, das diferentes modalidades da matéria. Os espíritos animais nada mais são do que as partes mais delicadas do sangue que passam do coração ao cérebro, e a seguir do cérebro aos músculos, que eles movem à maneira de nossos comandos hidráulicos.

No *Discurso do Método* Descartes aceita a possibilidade de máquinas que imitem animais, como um macaco, por exemplo, ou de outros animais sem razão. Essas máquinas, segundo ele, não "agem pelo conhecimento, mas somente pela disposição de seus órgãos" (1973, p.68). Pois, ao passo que a razão é um instrumento universal, que pode servir em todas as espécies de circunstâncias, "tais órgãos necessitam de alguma disposição particular para cada ação particular" (1973, p.68). E, embora fizessem muitas coisas tão bem, ou talvez melhor do que qualquer de nós, falhariam infalivelmente em algumas outras, pelas quais se descobriria que não agem pelo conhecimento, mas somente pela disposição de seus órgãos, afirma Descartes (1973, p.68). Para ele a razão necessitaria de outras disposições que as máquinas não podem possuir. Não é de se estranhar que algumas concepções de Descartes apareçam em Rousseau, pois, como afirma Gouhier (2005), a postura filosófica de Rousseau se assemelha à de Descartes no *Discurso do Método* e nas *Meditações*.

Desse modo, Descartes afirma que é moralmente impossível que uma máquina tenha ações como nós, pois, não tem razão, "para fazê-la agir em todas as ocorrências da vida, tal como a nossa razão nos faz agir" (1973, p.68). Nesse sentido, acredita numa máquina que possa proferir palavras, e admite que possa proferir até mesmo algumas palavras em resposta às ações físicas que causam uma mudança em seus órgãos: por exemplo, se alguém tocou em um lugar particular, perguntaria o que se queria dizer ou se fosse tocado em algum outro lugar, ele gritaria que estava sendo ferido e assim por diante. Mas não poderia colocar

palavras diferentes para responder ao significado de tudo o que é dito em sua presença, como até mesmo os seres humanos menos inteligentes podem fazer.

A primeira dificuldade que surge quando tentamos conceber constructos de IA inteligentes e conscientes reside no fato de nós mesmos, os construtores, não sabermos definir a faculdade - propriedade *Consciência* com precisão. Se admitirmos que a consciência se apresente como a relação entre a mente e o mundo e que esta se estrutura por significados e sentidos, como representá-los algoritmicamente? O que significa estar vivo? O que significa existir? E o mais significativo, como definir sentimentos, estados e sensações complexos como dor, crença, amor, ódio, honra, piedade, desejo, saudade e até mesmo a consciência e a inteligência? Essas são perguntas que ainda não podemos admitir que as máquinas possam representar como nós.

Em certa medida a consciência não é a execução de um código, por isso mesmo, um problema considerado bem difícil. Mas isso, não é um empecilho para uma boa parte de cientistas e filósofos, entre eles os funcionalistas. O funcionalismo, na filosofia da mente, é a doutrina, segundo a qual, acredita o que torna algo um estado mental de um tipo particular não depende da constituição interna, mas da maneira como funciona, ou do papel que desempenha no sistema do qual é parte. Essa doutrina está enraizada na concepção de alma de Aristóteles (2011) e tem ressonâncias na concepção de Hobbes (1992, 2000) a respeito da mente como uma "máquina de calcular", mas se tornou totalmente articulada (e popularmente endossada) apenas no último terço do século XX, afirma Leyser (2020). Os pensadores dessa corrente acreditam que a inteligência humana é um sistema de regras que processam dados e tem a finalidade de resolver determinados problemas, este sistema de dados pode ser instalado em diferentes *hardwares.*

Para Markus Gabriel, o funcionalismo, assim como o materialismo, é uma forma de religião, pelo menos no sentido de que ele nunca pode ser provado ou refutado por meios de evidências empíricas. Desse modo, materialismo, naturalismo e funcionalismo não são, como um todo, suposições que não podem ser provadas por meio das ciências naturais, mas sim "interpretações metafísicas da realidade", afirma Markus Gabriel (2021, p. 126).

Nos tempos atuais, o conceito de ser humano está em jogo. A era digital acredita que as máquinas podem resolver problemas melhores que os seres humanos. De fato, há muito tempo as máquinas já superam a inteligência humana em vários domínios. Vernor Vinge (1993) declarou que em 30 anos teríamos uma tecnologia significativa para criar uma inteligência sobre–humana e logo em seguida a era humana terminará dando início a uma era

pós–humana. As IAs já "superaram a maior parte dos seres humanos em inteligência em situações simuladas" (Gabriel, 2021, p. 21). Não é de agora que programas de Xadrez vencem e já venceram os melhores jogadores do mundo. Seja o GPS, as ferramentas de busca na internet, as redes sociais, os aparelhos auditivos com algoritmo que filtra os ruídos, sistemas de apoios às decisões médicas que em alguns casos detectam até câncer de mama e oferecem o melhor tratamento, robôs de estimação, de limpeza, cortadores de grama, de resgate, cirurgiões, entre outros, já utilizam a IA para os mais variados serviços. Mas seria possível a IA adquirir consciência, ou singularidade como preveem muitos cientistas e filósofos?

A questão da consciência das máquinas não é algo de momento. Um dos primeiros a questionarem essa possibilidade foi Alan Turing no seu artigo intitulado *Computing Machinery and Intelligence* (1950) e posteriormente os trabalhos de Vernor Vinge (1993) e Kurzueil (2005), estes dois últimos foram os responsáveis pela popularização do termo singularidade tecnológica, eles acreditam numa explosão de inteligência, ou seja, a ideia de uma superinteligência das máquinas.

Alan Turing é o responsável pela famosa pergunta: "Podem as máquinas pensar?" Em seu artigo, ele deixa claro que o ser humano pode construir uma máquina para imitar e reproduzir o trabalho que ele faria manualmente. Turing imaginou que se as máquinas pudessem reproduzir o trabalho manual através de seus algoritmos, então um computador eletrônico com portas lógicas e lógicas algoritmos conseguiria reproduzir o pensamento humano em um jogo de decisão. Ou seja, trata-se da possibilidade de efetuar tomadas de decisões que permitissem participar do, assim chamado, jogo da imitação. Turing ressaltou que o seu objetivo não é provar que as máquinas pensam, mas utilizar a IA de maneira tão bem-sucedida que a questão seja substituída por "E se as máquinas desempenhassem tão bem o papel a ponto de o interrogador não perceber que o seu interlocutor é uma máquina, e não um ser humano?". Precisamos concordar que ao copiar o modo de falar de um ser humano não torna uma máquina igual a nós. Mas os questionamentos sobre a IA se apresentam como questões filosóficas que ainda não temos respostas. Portanto, a pergunta de Turing sobre se as máquinas podem pensar, continua mais atual do que nunca.

Diante dos progressos científicos, cientistas, futurólogos, filósofos e políticos acreditam que não demorará muito até que as máquinas atinjam de uma vez por todas a consciência e se tornem livres. Nick Bostrom (2018) apresenta um futuro sombrio para a humanidade, ele acredita numa perspectiva que no futuro poderemos entender o cérebro humano de tal maneira, que poderíamos replicá-lo em máquinas, em sistemas inorgânicos. Portanto, nesse

caso, seria possível imaginar uma superinteligência capaz de criar processos internos dotados de status moral.

2. O problema da consciência das máquinas

Até meados dos anos 1980 não existia uma diferença para os conceitos de “mente’’, “estados mentais” e “experiência consciente”, designavam quase a mesma coisa. A diferença entre "mente" e "consciência" não era uma preocupação para os cientistas das IAs. Até a metade dos anos de 1980, estabelecer essa distinção não parecia ser uma preocupação visível entre os filósofos da mente. Havia grande entusiasmo com as perspectivas abertas pela IA e com a possibilidade de simulação mecânica das atividades mentais humanas através da construção de mentes artificiais. Pouco importava se uma máquina de jogar xadrez sabia ou não o que estava fazendo. O tema "consciência" não fazia parte da agenda dos funcionalistas, pois, para estes, processamento de informação e experiência consciente eram dissociáveis. Mas seria possível simular a cognição humana sem simular, ao mesmo tempo, seu aspecto consciente? Não seria essa uma diferença essencial entre mentes artificiais e humanas?

Esses foram os problemas que começaram a ser formuladas no final da década de 1980. Tudo se passava como se a simulação da atividade mental humana fosse uma tarefa perfeitamente exequível, dependendo apenas de avanços tecnológicos. Restaria apenas saber o que tornaria um estado mental algo consciente, e, para isso, seria necessário responder algumas questões que não deixavam de causar alvoroço: O que é consciência? Qual papel exerce na explicação da cognição humana? Existe cognição sem consciência? Terá a consciência um papel causal na produção da cognição e do comportamento? Podemos tratar a questão da consciência como um problema científico, isto é, como um problema empírico?

Quanto a respostas a estas perguntas, houve uma divisão de teorias entre grupos de filósofos, os mais conhecidos foram os chamados naturalistas, que acreditavam poder explicar a natureza da consciência através de teorias computacionais ou através do estudo do funcionamento cerebral. Filósofos e cientistas cognitivos como Churchland (1986, 1995) e Daniel Dennett (1991), por exemplo, seguem essa tendência, apostando no triunfo de teorias materialistas e aliando-as, às vezes, ao darwinismo. Estes naturalistas, os quais foram denominados anteriormente de funcionalistas, afirmam sobre o ser humano e o nosso pensar, em sua forma – padrão, “que somos descritíveis completamente aos moldes das ciências naturais e por isso também somos em princípio, replicáveis” (Gabriel, 2021, p.124).

O naturalismo deu um duro golpe na filosofia ao rejeitar a estratégia filosófica baseada na análise conceitual, acentuando que, em vez de ficarmos perguntando o que é a consciência - a busca por uma resposta para o *hard problem* de que nos fala Chalmers – temos de tratar essa questão como um problema científico, isto é, como um problema empírico. Em vez de definir consciência, decidiram estudar suas manifestações. Churchland, por exemplo, propunha uma estratégia do tipo "dividir para dominar": em vez de tentarmos elaborar uma teoria geral da consciência, temos de elaborar teorias específicas de processos mentais nos quais essa se manifesta, ou seja, teorias acerca da natureza da atenção, da memória, dos processos cerebrais subjacentes à produção do sono e da vigília etc. Quando desvendássemos todos esses aspectos da nossa vida mental, encontrando seus correlatos neurais estaríamos de posse de uma teoria da consciência.

A razão pela qual o funcionalismo é tão difundido atualmente é “porque ele permite acreditar indiretamente no naturalismo e no materialismo, sem ter de fornecer verdadeiras evidências ou argumentos filosóficos para ele” (Gabriel, 2021, p. 125). Essa concepção levou a uma separação entre filosofia da mente e neurociência, presente nos tempos atuais. Aqui entra a importância da teoria da Consciência de Rousseau, para justamente contrapor a esta convicção funcionalista de que poderíamos replicar em máquinas algo parecido com a nossa consciência.

Se as máquinas fossem apenas robôs, como uma máquina a vapor ou um vídeo game, seriam apenas meros equipamentos inconscientes, não seria necessário nenhum questionamento de nossa parte. O fato é que chegamos a um estágio de desenvolvimento tal, caso aconteça dessas máquinas adquirirem consciência, ou se forem construídas de modo que tenham mentes conscientes ou que “estejam associadas a experiências conscientes” ou, por alguma razão tenham dado a elas “status moral”. Então, seria importante “considerar de que maneira o resultado final afetaria essas mentes de máquinas” (Bostrom, 2008, p.305), uma vez que exista a possibilidade delas se tornarem numericamente dominantes.

E se algum dia, pudéssemos construir cérebros artificiais capazes de superar o cérebro humano? Ao questionar essa possibilidade, Bostrom (2008) defende que essa superinteligência poderia se tornar muito poderosa e nosso destino dependeria da superinteligência das máquinas. Mesmo com essa possibilidade Bostrom demonstra que estamos em vantagem, pois, nós que construímos as máquinas. “Poderíamos construir uma superinteligência que protegesse os valores humanos” (Bostrom, 2008, p.24).

Não é só Bostrom que afirma sobre a possibilidade das máquinas pensarem, entre os teóricos defensores de uma consciência nas máquinas podemos destacar além de Bostrom (2018), Chalmers (1996) e Harari (2016). Este último considera pode haver vida em meio inorgânico (*inverbis*) e a vida "eclodirá do reino inorgânico" (Harari, 2016, p.53). Personalidades como Bill Gates e Stephen Hawking (2014), alertaram para uma explosão de inteligência, que Nick Bostrom chamou de Superinteligência. Mas, felizmente, do outro lado, temos os que defendem o contrário como Searle (1980), Dreyfus (1975) e o brasileiro Miguel Nicolelis (2019), entre outros.

Contrariando os funcionalistas que acreditam em máquinas superinteligentes, Searle (1980), pelo contrário, afirma que a probabilidade das máquinas pensarem é de quase 0%. Searle em seu texto, *Mentes, cérebros e programas* (1997), utiliza o argumento do quarto chinês para provar que as máquinas não podem ter compreensão do que fazem. Ele se imagina em um quarto onde lhes dão umas folhas com um texto em chinês, além dessas folhas lhes dão um roteiro com regras em inglês, para correlacionar o primeiro texto com o segundo. Searle compreende o inglês, mas não sabe nada em chinês. Em um terceiro momento lhes dão blocos de papel com símbolos em chinês com algumas instruções, para correlacioná-los com os demais documentos. Ao correlacionar os símbolos com as regras, ele consegue responder às perguntas feitas em chinês, mesmo sem saber chinês.

> No que diz respeito ao chinês, eu simplesmente me comportei como um computador; executei operações computacionais com base em elementos formalmente especificados. Para os propósitos do idioma chinês, eu sou simplesmente uma instanciação de um programa de computador (Searle, 1997, p.4).

Para Searle o computador não entende nada, assim como ele não entende nada de chinês, não existe da parte dele (no caso do idioma chinês) e do computador nenhuma compreensão. O computador apenas faz o que lhe foi determinado pelo programador. Nesta perspectiva, o computador não entende nada, assim como uma máquina de somar e um carro também não compreendem nada. Retomando a pergunta de Turing, se as máquinas podem pensar, Searle afirma que sim, "nós somos precisamente essa máquina" (Searle, 1997, p.12), ou seja, só os seres humanos podem pensar.

Searle também acrescenta que o cérebro é um computador digital. O ponto é que a capacidade causal do cérebro para produzir intencionalidade não pode consistir na instanciação de um programa de computador, pois para cada programa, sempre é possível que haja algo que o instancie e, contudo, não tenha estados mentais. "Seja lá o que o Cérebro faça

para produzir intencionalidade, esta não pode consistir na instanciação de um programa, pois nenhum programa é por si só suficiente para produzir a intencionalidade" (Searle, 1997, p.15).

Embora os pensadores reflitam sobra a IA, é necessário também fazer um debate sobre a consciência com certa urgência para se puderem prever quais entidades inteligentes terão experiências subjetivas. A questão de saber se máquinas inteligentes devem ter alguns direitos depende crucialmente de serem conscientes e sofrerem ou sentirem alegria. Assim como "vida" e "inteligência", não há definição correta indiscutível da palavra "consciência". "Na verdade, existem muitos concorrentes, incluindo senciência, vigília, autoconsciência, acesso a informações sensoriais e capacidade de fundir informações em uma narrativa" (Tegmark, 2020, p.289).

Até aqui, apresentamos as diversas opiniões e possibilidades de algum dia as máquinas pudessem atingir níveis conscientes, também expomos, de modo mais tímido, os pensadores que defendem a impossibilidade das máquinas pensarem. A partir de agora, trataremos da concepção de Rousseau sobre a consciência e em seguida faremos uma relação com as IAs de nosso tempo. Tratamos da questão de um modo sem conceituar a consciência. Para tratar desta questão pormenorizada, partiremos da compreensão de Rousseau.

O termo consciência em sua concepção filosófica tem muito pouca relação com o significado atribuído ao senso comum. No senso comum a consciência é entendida como um estado de percepção do homem sobre o universo que o cerca. Por exemplo, fala-se de ter consciência ou de estar consciente quando a pessoa estiver apta a apreender aquilo que está a sua volta. Deste modo, o sono é um exemplo de falta de consciência. Para Rousseau, falar da consciência requer muito mais que uma breve análise do mundo exterior. É importante, pensar não somente na consciência enquanto faculdade de apreensão do mundo, mas também como ela se forma, e sua relação com este mundo exterior.

3. A consciência em Rousseau

A noção de consciência é um dos temas mais instigantes e provocantes da filosofia. A consciência é aquilo que há de mais íntimo no ser humano, estando relacionada tanto com a existência de um conteúdo psíquico, como também a possibilidade de retorno a si mesmo. Muitos filósofos entendem este conceito com diferentes acepções, dentre eles Platão, que como afirma Derathé (1948), influenciou a visão dualista do pensamento Rousseauniano e atribui à consciência o sentido de uma relação da alma com ela mesma, como uma lembrança, uma opinião, uma memória. Na obra *Filebo*, Platão utiliza tal termo da seguinte forma,

> Será que os seres vivos sempre têm consciência do que se passa com eles, não se processando nenhum crescimento sem que o percebamos, nem qualquer outra alteração da mesma natureza, ou acontecerá precisamente o contrário? (Platão, 2009, p. 233).

No fragmento em destaque, percebe-se que Platão utiliza o termo consciência como sinônimo de conhecimento, de um saber interior do ser, uma lembrança ou raciocínio, que tem como mediadora a linguagem. Assim, a consciência só estará presente no homem a partir do momento em que a linguagem se encontre desenvolvida.

Em Rousseau não é fácil chegar a um consenso sobre o conceito de consciência, ele ora demonstra a consciência como um sentimento, ora uma faculdade ou, em outros momentos, a própria alma do homem. A consciência como as demais faculdades, encontra-se na natureza do ser primitivo potencialmente, pois, o homem no estado de natureza possui somente dois sentimentos ativos como o amor de si e a piedade. Sendo o amor de si o sentimento que conduz o ser humano a sua autoconservação e a piedade, o sentimento que leva o homem ao sentir repulsa ao ver o sofrimento do outro.

No homem natural ou primitivo de Rousseau, encontram-se duas faculdades que o distingue dos outros animais. O livre arbítrio que é a capacidade de frear seus impulsos, diferentemente da natureza instintiva encontrada nos outros animais e a faculdade da perfectibilidade que consiste na capacidade que o indivíduo tem de apropriar-se das coisas em benefício de sua sobrevivência. A perfectibilidade também viabiliza associar-se a outros indivíduos, adaptar-se a ambientes e buscar melhores alternativas. Além do mais, “o homem natural não sabe o que é morrer, a morte para ele é um acontecimento natural, não a teme, apenas vive” (Rousseau, 1999. p.173). A consciência da morte só será possível ao homem social.

Rousseau inaugura assim, uma teoria dos sentimentos, contrariamente à filosofia de seu entorno que dava à razão um status mais elevado. A consciência para Rousseau está alicerçada pela questão do sentimento. No *Discurso sobre a desigualdade* (1999), é possível perceber três tipos de sentimentos: os sentimentos primários, próximos da natureza física, os sentimentos de ternura e pacificação, que favorecem os laços sociais e os sentimentos puramente sociais, nomeados de paixões. Assim, a consciência se encontra adormecida no homem primitivo e só poderá despertar e desenvolver-se a partir do momento em que o homem primitivo for submetido a alguma situação que lhe torne indispensável o despertar dessa faculdade.

Aparentemente pode-se perceber na obra rousseauniana três sentidos para a palavra consciência: o primeiro, no sentido de guia do homem; o segundo, com a responsabilidade de ser uma faculdade do julgamento, e o terceiro como sentimento, como já foi citado. No *Emílio ou da Educação* Rousseau afirma que a consciência "é o verdadeiro guia do homem; ela está para a alma assim como o instinto está para o corpo" (2014, p.405). Podemos afirmar que na sua antropologia a consciência é apresentada como elemento decisivo na constituição moral e espiritual do homem. Esta é uma noção fundamental, pois, "dizer consciência é dizer tudo do homem". Ela é a voz da alma, enquanto as paixões são a voz do corpo. "Ela é expressão do ser, é dom de Deus, mas é também fundamento da moral e princípio do conhecimento" (Silva, 2004, p.141).

No livro I do *Emílio,* Rousseau indica que a razão e a consciência são duas faculdades correlativas: a razão ensina a conhecer o bem e o mal e a consciência faz amar a um e a odiar o outro. Embora independente da razão, a consciência não pode desenvolver-se sem ela. No *Emílio,* Rousseau define a consciência da seguinte forma,

> Consciência! Consciência! Instinto divino, imortal e celeste voz; guia seguro de um ser inteligente e limitado, mas inteligente e livre; juiz infalível do bem e do mal que tornas os homens semelhantes a Deus (Rousseau, 2014, p. 412).

Quando Rousseau diz que a consciência é "juiz infalível do bem e do mal que tornas os homens semelhantes a Deus" ele está dizendo de maneira mais clara que existe no fundo das almas "um princípio inato de justiça e de virtude a partir do qual, apesar de nossas próprias máximas, julgamos nossas ações e as de outrem como boas ou más, e é esse princípio que dou o nome de consciência" (Rousseau, 2014, p.409). Para Jean Lacroix (1978, p.81), "a consciência segundo Rousseau expressa o ser, a existência no estado puro, pois existir é ter consciência de si". Consequentemente, é pela consciência, que clareia sua inteligência e seu julgamento, que o homem se eleva sobre as outras criaturas.

Os atos da consciência não são juízos em operações da razão, mas sentimentos, pois nossas ideias vêm de fora, restando em nós os sentimentos que as apreciam. A consciência por meio das sensações nos faz conhecer o bem. Só ao ser humano cabe determinar o que é bom ou mau. Assim, não posso me enganar, pois "a consciência não engana jamais, ela é o verdadeiro guia do homem" (Rousseau, 2004, p.405). Tal consciência não pode ser simplesmente o produto da experiência, ela é uma capacidade inata que requer tempo e maturidade para despertá-la.

A voz da consciência é uma expressão, uma percepção das exigências da própria ordem interior do homem, a qual constitui sua verdadeira necessidade e seu verdadeiro bem,

"que o orienta a agir corretamente, de acordo com os desígnios providenciais de Deus para todos os seres humanos". Ela é em nós uma faculdade inata, autônoma, "uma potência de intimação permanente, um eco da harmonia que existe nas profundezas da personalidade humana" (Silva, 2004, p.146) e que, se for ouvido, inspirará o homem a proceder por suas próprias ações, de acordo com a ordem Divina.

Rousseau entende a consciência como elemento decisivo na constituição moral e espiritual do homem. Esta é uma assertiva decisiva, pois "dizer consciência é dizer tudo do homem" (Silva, 2004, p.141). Enquanto a consciência é a voz da alma, as paixões são a voz do corpo. Os atos da consciência, não são juízos nem operação da razão, mas sentimentos. Ela é um sentimento interior que nos faz agir de acordo com o que temos de mais íntimo. Por isso, não posso me enganar, porque "a consciência não engana jamais, ela é o verdadeiro guia do homem" (Rousseau, 2014, p.405). A voz da consciência, é, portanto, uma percepção das exigências da própria ordem interior, a qual o orienta a agir bem, de acordo com os desígnios de Deus para toda a humanidade.

Ao homem que deixa de ouvir sua consciência, "pode ter dificuldades para identificar se o que ouve são as instruções do seu próprio ser inato, de acordo com os desígnios de Deus" (Silva, 2004, p. 146). Pois as paixões podem confundir essa voz interior, portanto é preciso saber reconhecê-la e segui-la, para Rousseau todo ser humano tem a capacidade de sentir as instruções da consciência e se guiar por ela.

4. Considerações finais

O campo dos estudos sobre as IAs ainda é muito recente, por isso temos mais perguntas do que respostas. Ao questionar como Rousseau reagiria a uma tecnologia que tenta simular o pensamento e o comportamento humano, é uma questão que implica uma difícil resposta, pois ele viveu no século XVIII, muito antes do surgimento da IA.

No entanto, podemos tentar imaginar como ele aplicaria seus princípios filosóficos a esse tema. Rousseau poderia interpretar a IA de dois modos distintos, primeiro, uma possível interpretação é que seria crítico da IA, pois via com desconfiança os avanços científicos e tecnológicos que podiam aumentar as desigualdades sociais. Poderia também argumentar que a IA é uma forma de alienação do ser humano, que perde sua autonomia e sua identidade ao se submeter às máquinas.

Outra possível interpretação é que Rousseau seria favorável à IA, desde que ela fosse usada para o bem comum e para o aperfeiçoamento da humanidade. Ele poderia reconhecer

que a IA tem potencial para resolver problemas complexos, ampliar o conhecimento e melhorar a qualidade de vida das pessoas. Poderia defender que a IA deve ser regulada por um contrato social, que estabeleça os limites éticos e legais de sua aplicação. Essas são apenas algumas hipóteses baseadas nas ideias de Rousseau e nos avanços da IA. Não há uma resposta definitiva ou única para essa questão, pois ela envolve aspectos filosóficos, históricos e sociais.

Quanto à questão da consciência, Rousseau a entende como um sentimento interior que orienta o homem na busca da verdade. Nesse sentido seria difícil reduzi-la a um conjunto de regras ou produzi-la em um programa de computador, pois Rousseau entende a consciência como uma faculdade inata. É a voz da consciência em nós que nos orienta a seguir nossos sentimentos mais puros e autênticos. Portanto, a consciência é uma faculdade humana, jamais poderia ser reproduzida por algum algoritmo. A consciência para Rousseau é uma característica fundamentalmente humana, uma faculdade inata que guia nossas ações e nos distingui dos animais.

Referências

I- Obras de Rousseau

ROUSSEAU, Jean. Jaques. **Confissões.** Tradução Raquel Queiroz e José Benedicto Pinto. Bauru- SP: Edipro, 2008.

ROUSSEAU, Jean – Jaques. **Discurso sobre a origem e os fundamentos da desigualdade entre os homens.** Coleção os Pensadores. Vol. I e II. Abril Cultural. São Paulo, 1999.

ROUSSEAU, Jean Jaques**. Do Contrato Social ou Princípios do Direito Político.** Tradução Lourdes Santos Machado. Coleção os Pensadores. São Paulo, Abril Cultural, 2005.

ROUSSEAU, Jean Jaques. **Emílio ou da Educação**. Trad. Roberto Leal Ferreira. São Paulo, Martins Fontes, 2014.

II- Bibliografia Geral

ARISTÓTELES. **Da alma** – De anima. Bauru: Edipro, 2011.

BOSTROM, Nick. **Superinteligência:** caminhos, perigos e estratégias para um novo mundo. Rio de Janeiro: DarkSide Books, 2018.

BOSTROM, Nick; CIRKOVI´C, Milan M. **Global Catastrophic Risks.** New York, OXFORD University Press, 2008.

CHALMERS, David. **The Conscious Mind:** in Search of a fundamental theory. New York, OXFORD University Press, 1996.

CHALMERS, David. **La mente consciente:** en busca de una teoría fundamental. Barcelona, Gedisa, 1999.

CHURCHLAND, Paul. **Matéria e consciência:** uma introdução contemporânea à filosofia da mente. São Paulo: UNESP, 2004.

CHURCHLAND, Paul. **Neurophilosophy:** toward a unified science of the mind/brain. Cambridge: MIT Press, 1986.

DESCARTES, René. Coleção os Pensadores. **Discurso do Método. Meditações.** Abril Cultural, 1973.

DENNET, Daniel. **Consciousness explained**. London: Penguin Books, 1991.

DREYFUS, Humberto. **O que os computadores não podem fazer**. Rio de Janeiro, Eldorado, 1975.

GABRIEL, Markus. **O sentido do pensar:** a filosofia desafia a inteligência artificial. Petrópolis – RJ, Vozes, 2021.

GOUHIER, Henri. **Les méditations métaphysiques de Jean-Jacques Rousseau**. Paris, Vrin: 2005.

HARARI, Yuval Noah. **De animales a dioses**. Buenos Aires: Debate, (2013) 2016a.

HARARI, Yuval Noah. **Homo deus:** uma breve história do amanhã. São Paulo: Companhia das Letras, (2015) 2016b.

HAWKING, Stephen. **Transcedence looks at the implications of artificial intelligence – bust are we taking AI seriously enough?.** Carta enviada ao jornal The independente**.** 2014, Disponível em: https://www.independent.co.uk/news/science/stephen-hawking-transcendence-looks-at-the-implications-of-artificial-intelligence-but-are-we-taking-ai-seriously-enough-9313474.html. Acessado em 03/08/2024.

HOBBES, Thomas. **Leviatã**. São Paulo: Martins Fontes, 2003.

HOBBES, Thomas. **Tratado sobre el cuerpo – De Corpore**. Madrid: Ed. Trota, 2000.

HOBBES, Thomas. **Natureza humana. Lisboa**: Casa da Moeda, 1992.

KURZWEIL. Rai. **A singularidade está próxima.** Porto Alegre, Iluminuras, 2018.

LACROIX, Jean. La conscience selon Rousseau. *In:* **Jean-Jacques Rousseau et la crise contemporaine de la conscience** - Colloque international du deuxième centenaire de la mort de J.-J. Rousseau, Paris: Beauchesne, 1978.

LEYSER, Kevin Daniel dos Santos. **Filosofia da Mente.**São Paulo, Uniasselvi, 2020.

NICOLELIS, Miguel. **O verdadeiro criador de tudo:** como o cérebro humano esculpiu o universo como nós o conhecemos. São Paulo, Planeta, 2020.

NICOLELIS, Miguel; CICUREL, Ronald. **O Cérebro Relativismo**: como ele funciona e por que ele não pode ser simulado por uma máquina de Turing. Independently Published, 2019.

PLATÃO. **Filebo**. Tradução Edson Bini. 1ª Edição. São Paulo: Edipro, 2009. IV volume. (Coleção Diálogos – Platão).

SEARLE, John. **Minds, Brains, and Programs. Behavioral and Brains Sciences 3:417-457.** Cambridge, University press, 1980. Disponível em: http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=5322AF285B05EC8E53E67AB168D086AD?doi=10.1.1.83.5248&rep=rep1&type=pdf&ei=N3CgTY2zOcaltweWgdn0Ag&usg=AFQjCNEPvV8Ag3nymDPLcOX5wcpyID9BJA. Acesso em: 03/08/2024.

SEARLE, John. **A redescoberta da mente**. 2ª ed. São Paulo, Martins Fontes, 2006.

SEARLE, John. *In*: TEIXEIRA, J.F, **Mentes, Máquinas e Consciência:** uma introdução à filosofia da mente, Editora UFSCar, São Carlos, pp. 61-94, 1997.

SILVA, Genildo Ferreira. **Rousseau e a fundamentação da moral:** entre razão e religião. 2004. 256 p. Tese (Doutorado em Filosofia) – Programa de Pós-Graduação em Filosofia, Instituto de Filosofia e Ciências Humanas, Universidade Estadual de Campinas, Campinas, São Paulo, 2004.

TURING, Alan. Computação e Inteligência. *In.* TEIXEIRA, J. F. **Cérebros, máquinas e consciência** - Uma introdução a filosofia da mente. São Carlos: UFSCar. 1996, p.19-60.

TEGMARK, Max. **Vida 3.0:** o ser humano na era da inteligência artificial. São Paulo, Benvirá, 2020.

VINGE, Vernor. The Coming Technological Singularity**:** How to Survive in the Post-Human Era. *In*: **Vision-21:** Interdisciplinary Science and Engineering in the Era of Cyberspace, 11-22. nasa Conference Publication 10129. nasa Lewis Research Center, 1993.