

VIA TEOLÓGICA

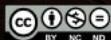
Volume 23 – Número 45 – jun. / 2022

ISSN 2526-4303 (ON LINE)

ARTIGO

A ÉTICA ARTIFICIAL DAS MÁQUINAS INTELIGENTES E AS REFERÊNCIAS DE APRENDIZAGEM DA ÉTICA CRISTÃ

Dr^a Márcia Gonçalves de Oliveira



A Revista Via Teológica está licenciada com uma Licença Creative Commons. Atribuição – Não Comercial – Sem Derivações - 4.0 Internacional

A ÉTICA ARTIFICIAL DAS MÁQUINAS INTELIGENTES E AS REFERÊNCIAS DE APRENDIZAGEM DA ÉTICA CRISTÃ

The Artificial Ethics of Intelligent Machines and the Christian Ethics
Learning References

Dr^a Márcia Gonçalves de Oliveira¹

1 Natural do Rio de Janeiro, reside no estado do Espírito Santo, é Bacharel em Ciência da Computação, Mestre em Informática e Doutora em Engenharia Elétrica pela Universidade Federal do Espírito Santo (Ufes). Atualmente é mestranda em Teologia pela Faculdade Teológica Batista do Paraná (FABAPAR). Tem formação em Teologia pela Faculdade Unida de Vitória (Curso Livre na modalidade a distância) e concluiu Pós-Graduação a distância em Teologia e Interpretação Bíblica pela FABAPAR. Tem especializações em Gestão da Educação a Distância pela Universidade Federal de Juiz de Fora (UFJF) e em Informática na Educação pela Universidade Federal de Lavras (UFLA). Atualmente é professora do Instituto Federal do Espírito Santo (Ifes) e docente titular do programa de mestrado e doutorado profissional em Educação em Ciências e Matemática (Educimat) do Ifes e do programa de Mestrado Profissional em Rede de Educação Profissional e Tecnológica (ProfEPT). Desenvolve pesquisas em tecnologias educacionais, em metodologias de ensino para a Educação a Distância e Híbrida e em ensino de computação. Atualmente, inicia pesquisas acadêmicas em Teologia e Educação Cristã. E-mail: clickmarcia@gmail.com.

RESUMO

O tema abordado neste artigo vem de uma discussão atual e muito preocupante sobre ética e Inteligência Artificial (IA). Esse tema parte do imaginário das obras de ficção científica para a literatura jornalística e acadêmica da Ciência da Computação, da Neurociência e da Filosofia trazendo à luz, em uma linha cada vez mais tênue entre o presente e o futuro e entre o real e o virtual, o potencial de humanidade das máquinas e o seu impacto na humanidade subjugada às máquinas. Tais preocupações têm sido justificadas porque máquinas têm sido criadas para agirem com inteligência, mas sem consciência, com base apenas em padrões reconhecidos, aprendidos e inferidos a partir de relações encontradas em uma grande quantidade de dados, que nem sempre são transparentes a usuários de tais máquinas e principalmente a seus próprios programadores. Dessa forma, este artigo tem como objetivo, a partir de publicações recentes sobre ética e IA, trazer à reflexão e à discussão, possibilidades de agregar às máquinas, além do comportamento inteligente, o comportamento ético nas tomadas de decisões com ou sem assistência humana. Para isso, é necessário ter uma compreensão básica de como as máquinas aprendem e como elas evoluem para se tornarem autônomas na tomada de decisões tendo como referências o comportamento inteligente e a consciência humana. A partir dessa compreensão, levantamos as seguintes questões para reflexão e discussão: Como a ética tem sido “entendida” e executada pelas máquinas? A ética advinda da consciência humana é uma referência ideal para as máquinas aprenderem ética? Não poderiam os princípios morais cristãos, dentro do que é consenso ou que não entra em conflito com outros padrões éticos, ser a referência da ética artificial das máquinas inteligentes? Sendo a ética cristã uma referência para a ética artificial, os princípios morais deverão ser absolutos ou poderão ser hierarquizados?

Palavras-chave: Ética cristã. Hierarquismo. Princípios Morais Absolutos. Inteligência Artificial.

ABSTRACT

The topic addressed in this article comes from a current and very worrying discussion about ethics and Artificial Intelligence (AI). This theme starts from the imaginary of science fiction works to the journalistic and academic literature of Computer Science, Neuroscience and Philosophy, bringing to light, on an increasingly fine line between the present and the future and between the real and the virtual, the potential for humanity of machines and their impact on humanity subjugated to machines. Thus, this article aims, from recent publications on the subject, to bring to reflection and discussion, possibilities of adding to machines, in addition to intelligent behavior, ethical behavior in decision-making with or without human assistance. For this, it is necessary to have a basic understanding of how machines learn and how they evolve to become autonomous in decision making with intelligent behavior and human consciousness as references. From this understanding, we raise the following questions for reflection and discussion: How has ethics been “understood” and “executed” by machines? Is human consciousness ethics an ideal reference for machines to learn ethics? Might not Christian moral principles, within which there is consensus or which do not conflict with other ethical standards, be the reference base of ethical knowledge for intelligent machines? Since Christian ethics is a reference for artificial ethics, should moral principles be absolute or can they be hierarchical?

Keywords: Christian ethics. Hierarchy. Absolute Moral Principles. Artificial intelligence.

INTRODUÇÃO

O avanço no desenvolvimento de tecnologias de IA que executam comportamentos inteligentes cada vez mais humanos tem chamado a atenção do mundo para os impactos da “inteligência sem consciência” inseridas em tecnologias digitais e para o uso ético das tecnologias de IA. Atentando para esses temas, assuntos relacionando ética e IA já têm sido frequentes nas publicações jornalísticas e acadêmicas sobre desenvolvimento tecnológico, especialmente nos campos da Ciência da Computação, da Neurociência e da Filosofia.

Recentemente, a *Revista MIT Technology* publicou o artigo *Ética em IA: a pergunta que não estamos fazendo*, em que a autora Nina da Hora destaca a importância de se estudar as implicações sociais e éticas das tecnologias de IA para antecipar e tentar mitigar danos irreversíveis para a humanidade (HORA, 2022). Nesse artigo, a autora inclusive aponta questões preocupantes, entre as quais, destacamos: a exposição da humanidade aos impactos da IA, as máquinas têm tomado decisões sobre a vida humana com base em instruções de “mãos transparentes”, grande parte desses sistemas está sob controle de grandes empresas e governos e o processo de desenvolvimento dos algoritmos que ditam as ações dessas tecnologias é sem transparência dos seus reais usos e intenções.

A falta de transparência e de responsabilidade no uso das máquinas inteligentes, conforme Hora (2022), fica ainda mais evidente, quando as decisões são tomadas a partir de dados desconhecidos pelo público-alvo da inteligência artificial das máquinas e por aqueles que as programam. Respondendo à pergunta que não está sendo feita, isto é, “que é quem está sendo impactado negativamente por essas tecnologias?” (HORA, 2022), apontamos as minorias que sofrem os preconceitos humanos aprendidos e reproduzidos por algoritmos, os que vão morrer por operação e/ou decisão de máquinas e os que se tornarem “quase máquinas” quando as máquinas se tornarem “quase humanas”.

Considerando esses que sofrerão os impactos negativos do uso não transparente e irresponsável da IA, destacamos a seguir alguns assuntos abordados na literatura de Ciência da Computação e Filosofia que envolvem a Inteligência Artificial e a Ética:

- Uso de reconhecimento facial na segurança pública que faz uso preconceituoso do fenótipo como fator de decisão da prisão de pessoas, colocando no alvo pessoas negras (HORA, 2022).
- Algoritmos de recomendação de vídeos do Youtube a partir de uma busca sobre conteúdos de ideologias extremistas, o que ajuda usuários a se aprofundarem nesses conteúdos (CAPPRA, 2021).
- Desemprego gerado pela substituição da mão de obra humana por máquinas autônomas (VALDERRAMAS, 2020).
- Erros cometidos pelas máquinas autônomas, alguns podendo ser fatais aos seres humanos (VALDERRAMAS, 2020).
- Robôs sexuais (RICHARDSON, 2016).
- Sistemas de Ética Artificial (SILVEIRA, 2021; CAPPRA, 2021; VALDERRAMAS, 2020).

Essas questões já não são mais meras especulações levantadas a partir do imaginário da ficção científica, mas sim temas de noticiários e publicações científicas do presente e do futuro que chamam à atenção da humanidade para que esta se antecipe no desenvolvimento de soluções para conter os riscos irreversíveis do uso irresponsável da IA. É preciso ter em mente, portanto, que “o triunfo da tecnologia nem sempre é o triunfo da ética e do bem-estar coletivo. Após duas bombas atômicas, no século XX, já devíamos ter aprendido essa lição” (TEIXEIRA, 2016).

Assumindo serem verdadeiras as possibilidades filosóficas (SILVEIRA, 2021) e tecnológicas (CAPPRA, 2021) de se desenvolver sistemas eticamente inteligentes, desenvolvemos um estudo neste

artigo sobre ética e Inteligência Artificial que busca levar a uma compreensão básica dos processos de aprendizagem de máquina como também à reflexão e à discussão se os exemplos de padrões éticos humanos podem corromper a aprendizagem de máquina e se os princípios da ética cristã poderiam ser as referências ideais para construção da ética artificial de máquinas inteligentes.

Para alcançar esses objetivos, este estudo se organiza conforme a ordem a seguir. Na Seção 2, descrevemos basicamente como acontece a aprendizagem de máquinas inteligentes. Na Seção 3, apresentamos a ideia de sistemas de ética artificial. Na Seção 4, refletimos algumas questões sobre ética artificial e discutimos uma proposta de formação da ética artificial a partir da ética cristã. Na Seção 5, concluímos com as considerações finais.

I. COMO AS MÁQUINAS APRENDEM

O aprendizado de máquinas basicamente se refere à criação de modelos que são aprendidos a partir dos dados (GRUS, 2019). Esse aprendizado pode ter abordagem supervisionada e não supervisionada. Na abordagem supervisionada, um sistema aprende com exemplos de decisões, em geral, tomadas por especialistas humanas. Dessa forma, o sistema consegue, tomar decisões em situações similares aos exemplos aprendidos. Alguns exemplos de sistemas de aprendizagem supervisionada são os sistemas de classificação automática, que classificam documentos a partir de exemplos de documentos similares já rotulados com códigos de classificação, e os sistemas de predição de notas de alunos, em que, a partir de alguns exemplos de exercícios já avaliados por um professor, o sistema atribui notas a partir das semelhanças reconhecidas automaticamente nas soluções de exercícios já corrigidas (OLIVEIRA, 2013).

Na abordagem de aprendizagem não-supervisionada os padrões se auto-organizam em classes não-rotuladas, de acordo com as características semelhantes, sem precisar de exemplos

para orientar a classificação desses padrões (OLIVEIRA, 2013). Um exemplo de sistema de abordagem não-supervisionada é o que reconhece perfis de alunos em salas de aula a partir de similaridades entre variáveis de avaliação de desempenhos.

Tanto a abordagem supervisionada quanto a não-supervisionada são baseadas em técnicas estatísticas e programadas a partir de algoritmos de mineração de dados. Dessa forma, a qualidade da aprendizagem depende dos dados utilizados para treinar esses algoritmos para tomadas de decisões. Por isso, para que a aprendizagem de máquina resulte em decisões corretas, é preciso que os dados utilizados para treinar as máquinas sejam representativos do contexto de decisões em termos de quantidade e de qualidade. Caso contrário, os algoritmos aprendem errado e tomam decisões erradas ou não aprendem o suficiente para tomar diferentes tipos de decisões. Um exemplo de como isso acontece é treinar uma máquina para classificar documentos a partir de exemplos de documentos classificados errados por classificadores humanos. Dessa forma, os algoritmos reproduzem erros e tendências dos exemplos com referências do especialista humano. Por isso, é importante considerar no projeto de sistemas de IA baseados em aprendizagem de máquina que

[...] muitas tendências humanas podem ser transferidas para as máquinas porque as tecnologias não são neutras; elas são tão boas ou ruins, quanto as pessoas que as desenvolvem. Muitas vezes, os algoritmos de aprendizagem de máquina herdaram padrões sociais refletidos em seus dados de treinamento, sem qualquer intenção original de programadores para incluir tais tendências. Os cientistas da computação chamam isto de viés algorítmico (CAPPRA, 2021, p.13).

Outras questões a serem consideradas na aprendizagem de máquina, principalmente quando se envolve uma grande quantidade de dados de onde são reconhecidas as relações que geram padrões, são a transparência dos processos definidos nos algoritmos de programação da IA das máquinas e a análise dos dados utiliza-

dos no treinamento desses algoritmos para evitar decisões “viciadas”, corrompidas ou tendenciosas. É preciso, portanto, conhecer e não perder o controle das informações que os dados trazem e geram e “como as contas são feitas” nos processos de decisão.

2. SISTEMAS DE ÉTICA ARTIFICIAL

Uma vez que as máquinas têm sido projetadas para agirem de forma inteligente, elas devem também ser programadas para agirem de forma consciente. Para isso, além de serem dotadas com uma inteligência artificial, elas devem também serem programadas com uma ética artificial que evolua na mesma proporção que avança o seu aprendizado e a capacidade de decidir de forma autônoma. Uma solução possível segundo filósofos e cientistas da computação é o desenvolvimento de agentes ou sistemas de ética artificial (SILVEIRA, 2021; CAPPRA, 2021).

Na perspectiva filosófica, Silveira (2021) considera possível uma ética artificial, mas destaca haver inúmeros desafios no campo da ética aplicada à IA. Por isso ela chama à atenção sobre como os algoritmos podem ser estruturados de modo a proteger pessoas do mau uso da IA e apresenta como exemplos as leis da robótica propostas por Isaac Asimov, que estabelecem, de certa forma, uma ética para reger as relações entre robôs e seres humanos (ASIMOV, 2015):

- I. um robô não pode ferir um ser humano ou, por inação, permitir que um ser humano venha a ser ferido.
- II. um robô deve obedecer às ordens dadas por seres humanos, exceto nos casos em que tais ordens entrem em conflito com a Primeira Lei.
- III. um robô deve proteger sua própria existência, desde que tal proteção não entre em conflito com a primeira ou com a Segunda Lei.

Ainda assim, segundo Silveira (2021), essas três leis propostas por Isaac Asimov não impediriam alguns dilemas éticos como, por exemplo:

[...] E se um robô para proteger um humano, tivesse de causar mal a outro humano? E se o humano causasse mal a si mesmo? Poderia o sistema agir para impedir? Haveria conflito entre o livre-arbítrio humano e o paternalismo artificial? O que significaria “causar dano”? [...] Poder-se-ia ainda questionar se estaria consistente tratar um robô superinteligente como um mero escravo ou objeto e não como um sujeito de direitos (SILVEIRA, 2021, p.33; ANDERSON, 2008).

A partir desses e de outros dilemas éticos, entendemos que, embora haja uma intencionalidade em criar um código de ética básico para estabelecer alguns limites éticos nas relações entre seres humanos e máquinas inteligentes, estes ainda são insuficientes considerando os desvios de ética em sistemas de IA e os seus impactos sociais apontados por Cappra (2021) e Hora (2022). Dessa forma, devem ser desenvolvidos sistemas inteligentes que se tornem cada vez mais éticos à medida que se tornam mais “humanos”. De acordo com Cappra (2021),

Sistemas éticos, sejam eles tecnológicos ou sociais, são formados por componentes similares: premissas, regras, ambiente, cultura, controle, atualização e suporte, sendo que quando essas partes não são devidamente supervisionadas e integradas, eleva-se o risco de uma falha comportamental desses autônomos. Então, na construção de um sistema de Inteligência Artificial é fundamental uma atenção redobrada com relação aos aspectos éticos e na influência que os ambientes sociais podem gerar enquanto esses sistemas estiverem em funcionamento (CAPPRA, 2021, p.20).

Nessa perspectiva, os sistemas de ética artificial deveriam ter como princípios basilares de modo a minimizar os riscos de mau uso da Inteligência Artificial (VALDERRAMAS, 2020): o prin-

cípio do benefício humano, da responsabilidade dos poderes eleitos ou judiciários no controle do uso da IA, da transparência e da educação e consciência. No entanto, a referência ideal de uma ética artificial deve ser a ética do “especialista humano”? Se sim, qual o contexto de vivência desse especialista, quais as crenças e ideologias defende e a que interesses atende o que ele sabe ou ensina? E como diferentes especialistas humanos influenciariam de forma positiva ou negativa a ética programada em sistemas de IA? Haveria padrões de ética melhores que representassem uma perfeição humana ou até mesmo uma referência divina na ética artificial de máquinas inteligentes? Essas e outras questões serão trazidas à reflexão para serem discutidas na seção a seguir.

3. QUESTÕES PARA REFLEXÃO E DISCUSSÃO SOBRE ÉTICA ARTIFICIAL E ÉTICA CRISTÃ

148

Uma vez que se considera possível o desenvolvimento de uma ética artificial para máquinas inteligentes de forma que se promova um uso responsável e transparente da IA, é importante também trazer à reflexão e à discussão possibilidades de referências éticas para a aprendizagem de máquinas. Até então a referência de inteligência das máquinas tem sido humana. Mas e a referência de consciência para máquinas também poderia ser humana? Na realidade humana, o conceito de ética é muito relativo e até complexo, principalmente quando consideramos a diversidade de culturas, crenças, sentimentos e conhecimentos da espécie humana.

É preciso considerar também que a compreensão, internalização e aplicação de princípios morais já são processos complexos para seres humanos que têm consciência e inteligência para discernir. No caso de máquinas sem consciência e com inteligência artificial, a complexidade desses processos seria sobremaneira ampliada, principalmente quando as máquinas tiverem que tomar

decisões que envolvam escolher ou priorizar princípios morais em situações adversas. Nessas situações, qual seria a hierarquia de valores para robôs que não sentem e pensam artificialmente considerando a programação que, de quem e de onde recebem?

Para essa discussão, em um passo inicial, trazemos três conceitos-chaves estudados no campo da ética que podem contribuir para estudos sobre referências para construção de uma ética artificial: regras morais, hierarquismo e a ética cristã. Sobre regras morais absolutas, questiona-se são absolutas e se há justificativas para serem violadas. De acordo Rachels (2013), o argumento principal contra as regras morais absolutas diz respeito às possibilidades de casos conflitantes. A violação da regra de falar sempre a verdade para salvar vidas e fazer um aborto para salvar a vida da mãe são alguns exemplos de casos conflitantes que levam a decisões de violar regras morais que deveriam ser absolutas ou de violar uma regra para cumprir outra.

Nesse último caso, conforme o hierarquismo, há regras universais, mas algumas delas são hierarquicamente ordenadas (GEISLER, 1984). O próprio Jesus, por exemplo, mostrou haver hierarquia de leis e apontou o amor como o mais importante na interpretação e no cumprimento das leis, a ponto de resumir a lei dos Dez Mandamentos, considerada máxima para os judeus, em dois mandamentos: amar a Deus e ao próximo. Para Jesus, era mais importante honrar pai e mãe do que dar oferta ao altar e valia mais salvar uma vida do que guardar o sábado. Por outro lado, na hierarquia de leis entendida pelos fariseus era bem diferente. Para eles, era mais importante guardar um sábado do que salvar uma pessoa, e as leis por eles inventadas em suas tradições invalidavam a lei moral de Deus. Para eles era mais fácil memorizar e cumprir mil tradições do que amar a Deus e ao próximo.

No mundo, há vários códigos de ética e, por isso, as hierarquias de valores podem variar entre diferentes culturas, posicionamentos ideológicos e classes sociais. Além disso, é difícil hierarquizar regras morais em um padrão universal. No entanto,

aqueles valores que representam consenso ou não sofrem oposições de outros códigos de éticas poderiam ser as regras morais referentes para estabelecer hierarquias de regras morais para evitar danos às relações humanas e, agora também, às relações homem-máquina para que se formem referências para uma ética artificial adequada para máquinas inteligentes.

Entre os vários códigos de ética, os princípios morais da ética cristã, que se fundamenta essencialmente na base ética do Antigo Testamento, que é a Lei exposta no Pentateuco, e na Ética do Amor, que é a base ética do Novo Testamento (RAE, 2013). Os princípios morais da Ética do Amor são os que representam mais consenso e menos oposição nos seus princípios morais em relação a outros códigos de ética. Além disso, a Ética do amor por si mesma faz uma hierarquia de valores que prioriza a vida e o bem-estar coletivo.

A Ética do Amor expressa no amor a Deus e ao próximo, cumprida perfeitamente em Cristo, embora seja muito difícil de ser cumprida na íntegra por humanos, não é rejeitada por todos aqueles que, independentemente de classes sociais e crenças, defendem o bem-estar coletivo da humanidade. Dessa forma, a ética cristã, tendo mais consenso e menos confronto com outras éticas, principalmente nas regras do amor ao próximo, pode ser referência de ética para as relações humanas e, no caso da ética artificial, para as relações homem-máquina.

Amar ao próximo como a si mesmo é uma regra moral muito forte da ética cristã, pois se todos a cumprissem na íntegra, o mundo resolveria boa parte de seus problemas sociais. Já os problemas sociais trazidos pelas máquinas de IA, também poderiam ser reduzidos na regra moral de amar o próximo como a si mesmo, uma vez que a próprias leis da robótica estariam dentro da regra de amar o próximo (homem ou robô) como a si mesmo (robô ou homem). O amor a ser entendido por homens e máquinas não deve fazer mal ao próximo, de sorte que o cumprimento da lei para homens e máquinas seja o amor (Rm 13.10).

Dessa forma, apenas ensinar as máquinas a não fazerem o mal a si mesmas e ao próximo (homem ou outro robô) já representa um grande avanço na formação das referências da ética artificial a serem aprendidas por máquinas inteligentes.

A partir desse entendimento de ter a ética cristã, pelo menos parcialmente, como referência da ética artificial e de estabelecer hierarquia de regras morais a partir da Ética do Amor, destacamos algumas questões para reflexões que podem gerar, a partir deste artigo, novas discussões e mais reflexões da comunidade acadêmica sobre o tema de ética artificial. Eis as questões:

A. Como a ética tem sido “entendida” e executada pelas máquinas?

A ética das máquinas tem sido entendida e executada basicamente segundo as leis da robótica apresentadas por Isaac Asimov (ASIMOV, 2015) e segundo as concepções de funcionamento das máquinas postuladas por Ada Lovelace e Alan Turing. As máquinas, segundo Ada Lovelace, a primeira programadora computacional, não pensam, apenas automatizam procedimentos obedecendo a instruções de programação (SILVEIRA, 2021). Já Alan Turing, o pai da computação, defende que as máquinas não pensam, mas podem “imitar” o comportamento humano inteligente e, “propõe verificar se uma máquina poderia imitar com sucesso o comportamento humano a ponto de confundir um observador imparcial” (SILVEIRA, 2021, p. 24).

Avançando nas concepções de Ada Lovelace e de Alan Turing, John Searle argumenta que as máquinas não pensam e desafia Turing distinguindo IA forte e IA fraca (SILVEIRA, 2021):

A IA fraca seria aquele campo em que as máquinas conseguem passar no Teste de Turing, performando competências linguísticas indiscerníveis de um ser humano. A IA forte sugere a possibilidade de máquinas que performam competências próprias de um ser humano, ou seja, não apenas aparentam como possuem igualmente todas as competências humanas, inclusive a consciência (SILVEIRA, 2021, p. 27).

Assim, de acordo com Silveira (2021), a conclusão de Searle é que a intencionalidade artificial da IA forte deve ampliar os poderes causais do cérebro humano e não elaborar apenas programa de computador para imitá-los sem expressar intencionalidade. Dessa forma, desenvolver competências humanas em máquinas, em especial a consciência para agir com ética, representa um grande desafio. Assim, para que as máquinas “entendam” e “executem” uma ética próxima da humana, elas precisam vencer os desafios de Ada Lovelace, de Alan Turing e de John Searle para serem capazes de obedecer a procedimentos, imitar comportamentos inteligentes de humanos e, principalmente, agir com ética tendo como referência a consciência humana ou uma consciência superior, que alguns veem em Deus.

B. A ética advinda da consciência humana é uma referência ideal para as máquinas aprenderem ética?

A resposta defendida neste trabalho, a partir do que foi apresentado, é não, pois a ética desenvolvida a partir da consciência humana sofre muitas variações conforme culturas, classes sociais e formações ideológicas e religiosas. Defende-se neste trabalho como uma ética de referência ideal para a ética artificial uma ética de consciência superior, que entendemos, neste trabalho, ser divina e humana representada na pessoa de Jesus Cristo, que desenvolveu uma ética de mais consenso e de menos confronto com outras éticas a partir do princípio moral de amar ao próximo como a si mesmo.

C. Não poderiam os princípios morais cristãos, dentro do que é consenso ou que não entra em conflito com outros códigos de ética, ser a base referência da ética artificial das máquinas inteligentes?

Sim. Continuando a resposta da questão anterior, defendemos como uma proposta inicial, a partir do princípio moral de amar o próximo como a si mesmo da ética cristã, uma ética artificial compreensível por máquinas com a lei moral de não

fazer mal ao próximo, que pode ser um humano ou um robô, alinhando-se com as três leis fundamentais da robótica propostas por Asimov em seu livro *Eu, Robô* (ASIMOV, 2015). Essa proposta avança as leis de Asimov porque tem como referência não a consciência humana, mas uma consciência superior divina e humana consolidada na pessoa de Jesus Cristo.

D. Sendo a ética cristã uma referência para a ética artificial, os princípios morais deverão ser absolutos ou poderão ser hierarquizados?

Poderão ser absolutos quando não forem conflitantes e hierarquizados conforme condições de contexto de ação das máquinas inteligentes. No entanto, a própria Ética do Amor defendida na ética cristã (RAE, 2013) por si mesma hierarquiza, como já dissemos, seus princípios morais priorizando, conforme Jesus ensinou e colocou em prática, a vida e o bem-estar coletivo. Entendemos, portanto, que não havendo lei contra a Ética do Amor e sendo ela universal e possível de prática, ainda que difícil de ser exercida por homens e por máquinas por eles criadas, ela pode ser a base da ética artificial de máquinas inteligentes.

CONSIDERAÇÕES FINAIS

Em resumo, este trabalho apresentou questões atuais e preocupantes sobre ética e Inteligência Artificial que já vêm sendo discutidas nas comunidades acadêmicas de Ciência da Computação, Filosofia e Neurociência e que, a partir deste estudo, visamos também levar essas questões para reflexão e discussão na comunidade acadêmica de Teologia Cristã, uma vez que a ética cristã, muito tem a contribuir para a formação de princípios morais para a ética artificial de máquinas inteligentes com base na Ética do Amor ensinada e vivida por Jesus.

As reflexões apresentadas representam as opiniões próprias da autora deste trabalho, que tem formação em Ciência da Computação e em Teologia. No entanto, apenas iniciamos

uma discussão sobre ética artificial, agora incluindo a academia teológica, deixando o debate aberto para novas reflexões, ideias e contestações que contribuam para a construção futura de uma ética artificial à luz da ética cristã para ser aprendida por máquinas inteligentes.

Nossas expectativas são, portanto, que a discussão aberta neste trabalho a partir de reflexões da autora, seja um passo inicial, mas relevante para abrir novas possibilidades e propostas para as pesquisas sobre ética artificial e para oportunizar a inserção da ética cristã no desenvolvimento e uso de tecnologias inteligentes favorecendo o desenvolvimento e o bem-estar da humanidade e dos robôs que são criados à sua imagem e semelhança.

REFERÊNCIAS

ANDERSON, Susan Leigh. Asimov's "three laws of robotics" and machine metaethics. *Ai & Society*, v. 22, n. 4, p. 477-493, 2008.

ASIMOV, Isaac. **Eu, robô**. Aleph, 2015.

CAPPRA, Ricardo. **Sistemas éticos para Inteligência Artificial**. Cappra Institute for Data Science, 2021. Disponível em: encurtador.com.br/gCTX9. Acesso: 20/01/2022.

GEISLER, Norman L. **Ética Cristã**. São Paulo: Vida Nova, 1984.

GRUS, Joel. **Data Science do zero**: Primeiras regras com o Python. Alta books, 2019.

HORA, Nina da. Ética em IA: a pergunta que não estamos fazendo. **MIT Technology Review**, 11/01/2022. Inteligência Artificial. Disponível em: <https://mittechreview.com.br/etica-em-ia-a-pergunta-que-nao-estamos-fazendo/>. Acesso em: 20/01/2022.

OLIVEIRA, M. G. **Núcleos de avaliações diagnóstica e formativa para regulação da aprendizagem de programação**. 2013. Tese de Doutorado. Tese de doutorado, Universidade Federal do Espírito Santo.

RACHELS, James. **Os elementos da filosofia moral**. 7. ed. Porto Alegre: AMGH, 2013.

RAE, Scott B. **Curso Vida Nova de Teologia Básica: Ética Cristã**. São Paulo: Vida Nova, 2013.

RICHARDSON, Kathleen. Sex robot matters: slavery, the prostituted, and the rights of machines. **IEEE Technology and Society Magazine**, v. 35, n. 2, p. 46-53, 2016.

SILVEIRA, Paulo Antônio Caliendo Velloso da. **Ética e Inteligência Artificial: da possibilidade filosófica de Agentes Morais Artificial**. 2021.

TEIXEIRA, J. d. F. **O cérebro e o robô (Ethos)**. [Kindle Android version]. Retrieved from Amazon.com. 2016.

VALDERRAMAS, Edgard. A ética como um dos desafios da Inteligência Artificial. In: Anais do V **Workshop sobre Aspectos Sociais, Humanos e Econômicos de Software**. SBC, 2020. p. 71-80.