



**CONSIDERAÇÕES INICIAIS ACERCA DA ATRIBUIÇÃO DE
PERSONALIDADE A AGENTES AUTÔNOMOS DE INTELIGÊNCIA
ARTIFICIAL**

**INITIAL CONSIDERATIONS ON THE PERSONIFICATION OF AUTONOMOUS
ARTIFICIAL INTELLIGENCE AGENTS**

Giovana Figueiredo Peluso Lopes¹

RESUMO: Neste trabalho, busca-se investigar se agentes autônomos de inteligência artificial poderiam ser considerados como pessoas pelo Direito. Para tanto, serão investigados os motivos que levam à atribuição de personalidade às pessoas naturais e aos entes coletivos para, posteriormente, verificar se os mesmos se aplicariam a sistemas artificiais. Será argumentado que, por não possuírem intencionalidade intrínseca, a tais agentes somente é possível atribuir a personalidade em sentido puramente formal, baseada em critérios utilitários, a exemplo do que ocorre com as pessoas jurídicas. Metodologicamente, a pesquisa é de natureza bibliográfica e caráter teórico-propositivo, valendo-se de fontes documentais diretas e indiretas.

Palavras-chave: Inteligência artificial; Personalidade; Pessoa jurídica; Intencionalidade; Consciência.

ABSTRACT: In this paper, the author seeks to investigate if autonomous agents based on artificial intelligence could be regarded as persons by Law. In order to do so, the

¹ Mestranda em Direito pela Universidade Federal de Minas Gerais (UFMG). Graduação em Direito pela Universidade Federal de Juiz de Fora (UFJF). Integrante do grupo de pesquisa em Direito, Tecnologia e Inovação (DTI UFMG) e do GENE – Grupo de Estudos em Neuroética e Neurodireito. Advogada consultiva em privacidade e proteção de dados pessoais. E-mail: giovanaplopes@gmail.com.

Artigo submetido em 20/03/20 e aprovado em 15/06/20

reasons that inform the personification of human beings and legal entities will be analyzed, as to verify if those could also be applied to artificial systems. It will be argued that since such agents do not possess intrinsic intentionality, it is only possible to grant them personality in a strictly formal sense, based in utilitarian criteria, as in the case of legal entities. Methodologically, this research is a theoretical-propositional, bibliographic study using direct and indirect documentary sources

Keywords: Artificial intelligence; Personality; Legal person; Intentionality; Consciousness.

INTRODUÇÃO

A ideia de que seres humanos conseguiriam, em determinado momento histórico, desenvolver máquinas capazes de “pensar” por si próprias e agir autonomamente é algo que permeia o imaginário coletivo desde tempos primordiais, servindo como pano de fundo para a criação de inúmeras obras literárias e cinematográficas². Entretanto, o que antes se apresentava como mera ficção científica, atualmente configura uma real possibilidade, capaz de revolucionar aspectos fundamentais da vida em sociedade de maneira ainda mais intensa do que o amplo acesso à Internet e a telefones celulares fizeram ao longo das últimas décadas.

Sistemas baseados em inteligência artificial (IA) já operam nas mais diversas áreas do cotidiano, passando muitas das vezes despercebidos por seus usuários - estando presentes em provedores de aplicação do comércio eletrônico, das principais plataformas de busca e compartilhamento de dados, e das redes sociais. Carros autônomos já circulam pelas ruas de algumas cidades do mundo, e a possibilidade de se empregar robôs como assistentes terapêuticos nos lares de indivíduos com deficiências motoras graves tem sido

² Cita-se, como exemplo, o clássico romance *Frankenstein* ou o Moderno Prometeu de Mary Shelly; o popular desenho animado *The Jetsons*, produzido pela empresa Hanna-Barbera; e a personagem *HAL 9000*, principal antagonista de *2001: Uma Odisséia no Espaço*, de Stanley Kubrick e Arthur C. Clarke.

estudada³. Ainda, por serem capazes de identificar padrões imperceptíveis ao olho humano, o uso de sistemas de IA para a execução de complexas operações financeiras tem sido cada vez mais difundido⁴.

Cumprе ressaltar que, no presente momento, por mais sofisticados e inovadores que sejam tais sistemas de inteligência artificial, pode-se afirmar que eles são, no máximo, semiautônomos. Isto porque é possível identificar, por trás de seu agir, as diretivas humanas responsáveis por determinado funcionamento ou tomada de decisão, rastreando-a diretamente de volta à programação do *software*, design da máquina, ou dados que lhe foram alimentados ao longo do processo de aprendizagem.

Na medida em que sistemas artificiais passem a serem capazes de atuar de maneira verdadeiramente autônoma, ou seja, independentemente de um *input* humano direto e baseando-se em informações que o próprio sistema adquire e analisa (VLADECK, 2014), será possível que eles executem determinadas ações não antecipadas por seus criadores, muitas delas com consequências danosas.

Surgem, a partir disso, indagações acerca de qual deverá ser o tratamento jurídico conferido a tais agentes artificiais autônomos. Por exemplo, questiona-se a respeito de como deverá ser atribuída responsabilidade, tanto moral quanto legal, na hipótese de ocorrência de um dano não diretamente rastreável a um indivíduo humano.

Quando o comportamento de uma IA se tornar imprevisível até mesmo para aquele que a programou, será tentador atribuir-lhe uma espécie de vontade ou intencionalidade. Afinal, seres humanos possuem a tendência de antropomorfizar entes cujo comportamento se assemelhe ao seu, projetando neles suas próprias qualidades, de forma a fazê-los aparentar mais humanos (DARLING, 2012).

Nesse contexto, faz-se necessário compreender até que ponto é possível comparar o agir autônomo de um sistema artificial ao de um ser humano. É imprescindível que os eventuais incursos normativos partam de premissas corretas acerca do tema, e sejam informados por aquilo que a inteligência artificial é, de fato, capaz de fazer.

³A título de exemplo, veja-se o projeto “Robots for Humanity”..

⁴ Atualmente, *softwares* de *High Frequency Trading* (HFT ou negociação de alta-frequência) já são responsáveis por mais de 60% (sessenta por cento) das operações realizadas na Bolsa de Valores de Nova York (KAPLAN, 2015, p. 61).

Tendo isto em mente, o presente trabalho tem por objetivo investigar se sistemas autônomos baseados em inteligência artificial poderiam ser considerados como pessoas pelo Direito⁵, seja em sua acepção substancial, como a personalidade que é atualmente reconhecida a seres humanos, ou em sua acepção formal, como centros autônomos de imputação de direitos e deveres, a exemplo da personalidade jurídica.

A hipótese apresentada no trabalho é a de que agentes artificiais autônomos, por apresentarem apenas uma intencionalidade extrínseca, não possuem os estados subjetivos necessários ao seu reconhecimento como pessoas a partir de uma concepção ético-normativa do termo. Sua personificação, portanto, somente é possível se baseada em uma concepção puramente formal do instituto, justificando-se a partir de critérios utilitários e servindo, em última análise, a fins de pessoas humanas.

Metodologicamente, a pesquisa é de natureza bibliográfica e caráter teórico-propositivo, valendo-se de fontes documentais diretas e indiretas, e consiste numa abordagem qualitativa por traços de significação, por meio do procedimento da análise de conteúdo.

1. A DUPLA ACEPÇÃO DA PERSONALIDADE

A problemática acerca da personificação de agentes de inteligência artificial é passível de ser encontrada na literatura jurídica já em 1992, quando Lawrence B. Solum (1992), em trabalho intitulado “*Legal Personhood for Artificial Intelligences*”, analisa a questão através de uma série de experimentos mentais de forma a transfigurar a questão puramente teórica em termos legais – questionando-se, por exemplo, se uma IA poderia servir como um administrador fiduciário, ou se a ela poderiam ser concedidos direitos fundamentais, como o direito à liberdade de expressão.

Quase três décadas depois, a resposta à pergunta colocada por Solum – poderia uma inteligência artificial se tornar uma pessoa? - continua em aberto. Mais recentemente, o debate ganhou fôlego com a Resolução de 16 de fevereiro de 2017 do Parlamento Europeu com recomendações à Comissão de Direito Civil sobre Robótica (2015/2103-INL).

⁵ Neste trabalho, tratar-se-á o Direito como fenômeno universalizável e, por conseguinte, as questões abordadas não terão como foco nenhum ordenamento jurídico específico.

Nela, ao tratar da questão da responsabilidade, solicita-se que a Comissão, quando da avaliação do impacto de seu futuro instrumento legislativo, explore todas as soluções legais possíveis. Dentre elas, a criação de um status jurídico específico para que ao menos os robôs autônomos mais sofisticados possam ser considerados como *peças eletrônicas*, com direitos e obrigações específicas, incluindo-se o dever de reparar qualquer dano que possam causar (UNIÃO EUROPEIA, 2017).

A proposta de considerar agentes autônomos de inteligência artificial como *peças eletrônicas* visa à criação de uma nova categoria de entes expressamente dotados, ainda que de maneira limitada, de capacidade jurídica e, portanto, sujeitos de direitos e obrigações. Aproxima-se, dessa forma, da personalidade em sentido formal, compreendida como a aptidão genérica para figurar em um dos polos de uma relação jurídica, e que permite que entes abstratos e claramente desprovidos de intencionalidade intrínseca, como sociedades e associações, sejam subjetivados pelo ordenamento normativo.

Note-se que o debate acerca da atribuição de personalidade a agentes artificiais tem se desenvolvido, tradicionalmente, a partir de dois contextos distintos: o primeiro deles, de caráter pragmático, engloba argumentos que visam salvaguardar os interesses das pessoas envolvidas nas interações com tais agentes, evitando, por exemplo, a sua responsabilização direta pelas decisões e ações por eles tomadas autonomamente⁶ – como nas discussões em pauta no âmbito europeu. Seus argumentos tratam, portanto, a personalidade a partir de uma visão técnica, como atributo.

O discurso acerca da personalidade engloba, ainda, outro sentido para o conceito, traduzindo o conjunto de características e atributos da pessoa humana, considerada objeto de proteção privilegiada pelo ordenamento, bem jurídico representado pela afirmação da dignidade humana. É aquilo que se entende por personalidade como valor, ou personalidade substancial, sendo, portanto, peculiar à pessoa natural (TEPEDINO, 2001).

É justamente nesse contexto em que se encaixa a segunda categoria argumentativa comumente encontrada na literatura acerca da personificação de sistemas de IA, e que se

⁶ No que tange à responsabilidade civil por danos causados por uma IA, veja-se: ČERKA, Paulius; GRIGIENĖ, Jurgita; SIRBIKYTĖ, Gintarė. Liability for damages caused by artificial intelligence. *Computer Law & Security Review*, v. 31, n. 3, p. 376-389, 2015; PIRES, Thatiane; SILVA, Rafael. A Responsabilidade Civil pelos Atos Autônomos da Inteligência Artificial: Notas iniciais sobre a resolução do Parlamento Europeu. *Revista Brasileira de Políticas Públicas*, v. 7, n. 3, p. 239-254, 2017.

refere ao problema da consciência fabricada artificialmente e a conseqüente necessidade de se atribuir a tais entes certos direitos a princípio somente aplicáveis à pessoa humana.

Uma análise completa do tema deve ter em mente essa diferenciação entre a personalidade como atributo e a personalidade como valor, vez que as razões que determinam a personificação de seres humanos não são passíveis de equiparação àquelas consideradas para a subjetivação de entes coletivos. Isto porque, conforme será explicitado ao longo deste trabalho, enquanto a personalidade natural se apresenta como reflexo da dignidade imanente ao substrato ontológico do ser humano, tendo como base certos predicados fundamentais, a personalidade jurídica parte de uma abordagem funcionalista, segundo a qual a personificação de entes abstratos por determinado ordenamento baseia-se em critérios de utilidade e surge a partir da atribuição social.

2. IA E PERSONALIDADE COMO VALOR

Conforme ressalta Spaemann (2007), as pessoas ocupam uma posição particular dentre tudo aquilo que existe. Elas não constituem, em sua totalidade, uma espécie natural, sendo necessário, dessa forma, compreender o significado de tal nomenclatura para que seja possível diferenciar *algo* de *alguém*. Assim, o autor se indaga o que exatamente se pretende dizer quando se fala em alguém, portanto uma pessoa, atribuindo-lhe, por conseguinte, o direito a um *status* particular.

Nesse contexto, e a partir da consagrada definição de Boécio (2005, p. 168), segundo a qual a pessoa é uma “substância individual de natureza racional”, a filosofia buscou distinguir as características com base nas quais se atribui a determinados seres tal *status* específico. Tais tentativas podem ser sumarizadas em duas correntes: a descritiva e a axiológica (BROZEK, 2017).

A primeira delas objetiva precisar justamente aquilo que na concepção de Boécio vem definido como “racional”, enumerando uma série de predicados empíricos (e, notadamente, mentais) a partir dos quais as pessoas devem ser definidas. A corrente descritiva faz-se presente com maior intensidade no pensamento anglo-saxão, em especial na filosofia analítica, sendo possível remontar suas origens às ideias de Locke (SPAEMANN, 2007, p. 3).

A segunda corrente, por outro lado, compreende o conceito de pessoa como um *nomen dignitatis*, atribuindo-lhe, portanto, conotações axiológicas. Ela se faz presente no pensamento thomista e, principalmente, na filosofia kantiana e neokantiana, servindo como um conceito basilar à fundamentação dos direitos humanos (SPAEMANN, 2007, p. 4).

No ordenamento jurídico brasileiro, a noção de pessoa (envolvendo a pessoa natural e a pessoa jurídica), possui um caráter técnico, definindo-a como um centro autônomo de imputação de direitos e deveres, e não se encontra conectada concretamente a uma concepção filosófica, seja ela descritiva ou axiológica. Entretanto, nota-se uma clara associação entre ser *pessoa* e ser *humano*, uma vez que se reconhece que a personalidade, enquanto atributo técnico, não pode se encontrar desvinculada da personalidade enquanto valor, em se tratando de pessoas humanas. Tal ênfase conferida à humanidade nos discursos acerca da personalidade traduz a necessidade de se atentar para o substrato material personificado, ou seja, para as bases deontológicas de inclusão na categoria (SILVA, 2009). Isto porque

tendo em vista as bases do reconhecimento da pessoalidade, existe uma ampla área de sobreposição lógica entre ser humano e ser pessoa para o Direito, apesar de não ser esta absoluta, bem como existe determinado conteúdo ou substância no conceito de pessoa, somente por analogia aplicável este último aos entes ideais ou pessoas coletivas e apenas em seu aspecto formal (...) (SILVA, 2009, p. 33)

De fato, não é possível se falar em *pessoa* substancial sem fazer menção a este “conteúdo” ou “substância”, ao substrato material da personalidade, não através da mera descrição do conjunto de entes personificados por determinado ordenamento, mas da análise das características essenciais ao seu reconhecimento como tal – através de um conceito normativo de pessoa (SILVA, 2009, p. 34).

Nesse sentido, e a partir de uma análise histórico-filosófica do conceito de pessoa, Stancioli identifica três predicados fundamentais, ou dimensões incontornáveis, necessários à identificação de determinado ente como pessoa. A primeira delas consiste na autonomia moral, no sentido de que a pessoa humana é um ente com uma identidade prática autodeterminada, capaz de agir por razões morais que passam pelo crivo de seu próprio juízo racional (STANCIOLI, 2010).

Em Stancioli, a autonomia encontra-se conectada à capacidade que um indivíduo possui de eleger os valores constitutivos daquilo que ele mesmo considera como uma vida boa, passando, para tanto, por um processo de autoconhecimento. Isto porque em argumentos que se fundam em expectativas de vida pós-religiosas – no sentido de que o discurso público não se pode legitimar em instâncias teológico-metafísicas – a construção de uma vida boa ou digna está a cargo das próprias pessoas humanas, capazes de fala e de ação. Garante-se a todos, dessa forma, a mesma liberdade para desenvolver uma autocompreensão ética, a fim de formular uma concepção pessoal de vida boa de acordo com capacidades e critérios próprios (STANCIOLI, 2010, p. 116).

O segundo predicado fundamental da personalidade, indissociável do primeiro, consiste na alteridade, isto é, no estar em relação ao outro, que revela o caráter essencialmente intersubjetivo do conceito – somente é possível se constituir e se afirmar como pessoa perante o outro e dependentemente do outro (STANCIOLI, 2010, p. 84).

Conforme explica Taylor, não se pode ser um *self* por si só, mas apenas em relação a certos interlocutores: de um lado, os parceiros de conversação que foram essenciais ao alcance de uma autodefinição própria; de outro, aqueles que hoje são cruciais para a continuidade da apreensão de linguagens de autocompreensão. Em suma, o conceito de pessoa envolve, necessariamente, alguma referência a uma comunidade definitória ou a “redes de interlocução” (TAYLOR, 2005, p. 55).

Por fim, a terceira dimensão incontornável do conceito de pessoa diz respeito à dignidade, como consequência da constatação de que as ações pessoais são axiologicamente carregadas por valores, articulados quando da formação da identidade prática das pessoas. A dignidade consiste, dessa forma, no reconhecimento ético-normativo de que agir por razões que remetem a valores que os próprios agentes tomam como seus, como constitutivos da sua própria autocompreensão prática, constituindo o traço definidor que diferencia a pessoa de qualquer outro tipo de ente.

Tais valores são imprescindíveis para a compreensão da pessoa, constituindo aquilo que Taylor denomina de *hiperbens*: “bens que não apenas são incomparavelmente mais importantes que os outros como proporcionam uma perspectiva a partir da qual esses outros devem ser pesados, julgados e decididos” (TAYLOR, 2005, p. 90). Eles são constitutivos da personalidade, justamente por definirem os marcos para aquilo que se julga ser a direção de uma vida boa, possibilitando um entendimento social recíproco

acerca de quais seres seriam merecedores de dignidade e de que forma ela poderia ser concretizada.

Deve-se acrescentar, ainda, que a existência pessoal é sempre mediada por uma base física, traduzida, no caso da pessoa humana, pelo corpo. Disso resulta que

não se pode compreender totalmente o fenômeno da pessoa ao se desconsiderar as dimensões espaciais e temporais em que esta se insere através da corporeidade, visto que do corpo nascem e se propagam as significações que fundamentam a existência tanto individual quanto coletiva, sendo o lugar e o tempo nos quais a existência toma forma através da fisionomia singular de um ator (SILVA, 2009, p. 12).

A capacidade que as pessoas possuem de eleger, em um processo dialógico, os valores que definirão aquilo que elas próprias consideram como uma vida boa ou digna, encontra-se altamente dependente de determinadas faculdades mentais que, uma vez que se rechaça qualquer espécie de dualismo de origem cartesiana, estão necessariamente vinculadas ao seu substrato material, conforme se verá adiante.

Uma vez estabelecidas as características essenciais a um conceito ético-normativo de pessoa, cumpre investigar a possibilidade de enquadramento de sistemas autônomos de inteligência artificial na categoria, analisando até que ponto tais sistemas são capazes de reproduzir os estados subjetivos de consciência que os seres humanos possuem e que possibilitam a sua compreensão como pessoas. Buscar-se-á, nesse ponto, refutar a ideia de que a mente está para o cérebro (ou, de maneira geral, para o corpo), tal como o programa está para o *hardware* de um computador – visão tradicionalmente conhecida como computacionismo.

2.1. O cérebro humano como sistema adaptativo complexo

O computacionismo consiste na visão de que o comportamento inteligente pode ser causalmente explicado pelas computações realizadas pelo sistema cognitivo (ou cérebro) de um agente (PICCININI, 2009). Em sua versão mais radical, denominada de funcionalismo computacional, afirma-se que estados mentais são, pura e simplesmente, estados computacionais, disso resultando a hipótese de que o cérebro de qualquer animal, poderia ser reduzido a um algoritmo e simulado em um computador digital.

O computacionismo está por trás do propósito basilar da inteligência artificial: o de projetar máquinas computacionais capazes de executar tarefas cognitivas, ou, conforme colocado por McCarthy (1955), o de fazer com que uma máquina se comporte de maneira que, caso se tratasse de um ser humano, seria considerada inteligente.

Quando levada ao limite, tem-se que uma filosofia computacionista não apenas prevê a possibilidade de que todo o espectro de experiências humanas possa ser digitalmente reproduzido por um computador, mas também que, em virtude do crescimento exponencial do poder computacional vivenciado nas últimas décadas, máquinas serão capazes de suplantar, em um futuro próximo, as capacidades mentais humanas em sua totalidade – naquilo que se convencionou chamar de singularidade⁷.

Acredita-se, dessa forma, ser apenas uma questão de tempo até que seja possível emular a consciência humana em sistemas artificiais, sendo o único impedimento para tanto o atual estágio da tecnologia, ou seja, o fato de ainda não terem sido criadas as bases e os programas necessários para que tal marco seja alcançado. Muito pouco se questiona, contudo, acerca da premissa básica por trás da teoria computacional: seria mesmo possível simular as funções superiores de um cérebro humano em uma máquina digital?

A crença em um poder computacional ilimitado tornou-se popular graças à clássica teoria de Church-Turing segundo a qual todo processo efetivo – isto é, para o qual exista um algoritmo ou um processo mecânico de computação - pode ser efetuado por meio de uma máquina universal de Turing (COPELAND, 2017). Embora a hipótese em questão tenha sido idealizada originalmente apenas para sistemas formais, ela acabou sendo interpretada por inúmeros autores como definindo um limite computacional para todos os fenômenos naturais, resultando na crença de que nenhum sistema computacional físico poderia exceder a capacidade de uma máquina de Turing (NICOLELIS; CICUREL, 2015).

Ignora-se, dessa forma, que a definição de computabilidade proposta por Turing é aplicável somente a questões relativas ao formalismo matemático, assumindo, por conseguinte, certas suposições que limitam a sua aplicabilidade a sistemas biológicos. Ela assume, por exemplo, que a representação informacional é abstrata, formal e sintética, quando, na verdade, em sistemas biológicos complexos como o cérebro humano, a

⁷ Cf. KURZWEIL, Ray. *The Singularity is Near: When humans transcend biology*. Nova York: Penguin Books, 2005.

informação encontra-se intimamente ligada à estrutura física do organismo e possui uma rica estrutura semântica (NICOLELIS; CICUREL, 2015, p. 40).

Uma interpretação “física” e não puramente formal da tese de Church-Turing não leva em consideração que o sistema nervoso central de animais constitui um sistema complexo, no sentido de que suas funções superiores derivam de propriedades emergentes – propriedades possuídas pelo sistema como um todo, mas não por seus componentes individualmente considerados. Qualquer tentativa realmente precisa de simulação deveria, necessariamente, incluir em sua definição de complexidade todas as interações do sistema com entidades externas (como o meio ambiente e cérebros de outros indivíduos), uma vez que todos esses fatores interagem e modificam o sistema complexo em análise (NICOLELIS; CICUREL, 2015, p. 41).

O sistema nervoso central apresenta-se, ainda, como um sistema adaptativo, visto que o cérebro possui a habilidade de se reorganizar constantemente – tanto a nível funcional como estrutural – como resultado de suas experiências presentes e passadas. A informação nele processada é utilizada, portanto, para reconfigurar a sua função e a sua morfologia, criando uma integração recursiva perpétua entre informação e tecido neural (NICOLELIS; CICUREL, 2015, p. 41-42). Trata-se do princípio da plasticidade neural, em virtude do qual a representação interna de mundo mantida pelo cérebro, incluindo-se o próprio senso de si, ou *self*, permanece em contínua transformação ao longo da vida de um indivíduo.

São justamente esses atributos que definem um sistema complexo adaptativo que fazem com que as tentativas de simular de maneira acurada as principais funções cerebrais de um organismo biológico estejam fadadas ao fracasso:

[...] a rica semântica dinâmica que caracteriza as funções cerebrais não pode ser reduzida à sintaxe limitada dos algoritmos usada por computadores digitais. Isso ocorre porque as propriedades emergentes que surgem simultaneamente dos diferentes níveis organizacionais físicos do cérebro... não são efetivamente “computáveis” por uma máquina de Turing. Ao invés, essas propriedades podem ser apenas aproximadas temporariamente por uma simulação digital (NICOLELIS; CICUREL, 2015, p. 42-43).

Tem-se, dessa forma, que o fluxo de informação que ocorre dentro de um organismo, em especial dentro do cérebro animal, por envolver múltiplos níveis organizacionais e por continuamente modificar sua matéria orgânica, encontra-se

intimamente atrelado ao próprio substrato físico que ele modifica. Em outras palavras, em sistemas adaptativos complexos, como o sistema nervoso central de um ser humano, a informação é altamente dependente do substrato orgânico, não se aplicando a eles a dicotomia *software/hardware* para descrever de maneira distinta o modo de processamento integrado de informação feito pelo sistema, como ocorre em computadores digitais.

Conforme previamente colocado, no cérebro humano a informação encontra-se intimamente ligada à estrutura física do organismo e possui uma rica estrutura semântica. Justamente por isso, ele é o único objeto natural conhecido capaz de processar dois tipos diferentes de informação: a simbólica, também conhecida como informação de Shannon e Turing, e a informação de tipo Gödeliana.

A primeira é expressa através de uma sintaxe rígida gerada por bits e bytes, e o recipiente de uma mensagem desse tipo tem que decodificá-la para extrair o seu significado, conhecendo previamente, para tanto, um código externo. A informação de Gödel, por sua vez, como é gerada diretamente por um sistema integrado – no caso, o cérebro –, possui um rico espectro semântico que amplifica o significado da linguagem do sujeito: a forma pela qual ele comunica os próprios pensamentos e sentimentos (NICOLELIS; CICUREL, 2015, p. 58). Dessa forma, não se faz necessário nenhum código para o seu processamento, visto que seu significado é reconhecido instantaneamente pelo sistema.

Esta capacidade demonstrada pelo cérebro de processar a informação de Gödel é explicada por Nicolelis e Cicurel (2015) pelo fato de que, para além de executar computações de caráter puramente digital, ele também computa utilizando componentes analógicos, algo totalmente diverso da maneira pela qual computadores digitais operam. Note-se que as experiências subjetivas da consciência humana, como a sensação de dor e o senso de si, são expressas por meio da informação de tipo Gödeliana.

2.2. Intencionalidade intrínseca e processos mentais

Embora informações Gödelianas não sejam diretamente acessíveis por terceiros, sua existência é independente de um observador, exibindo não apenas efetividade causal, ou seja, a capacidade de desencadear alterações no cérebro, mas também se encontrando vinculadas à história integrada de cada cérebro individual. Elas descrevem, essencialmente, o ponto de vista único do enredo de vida de cada indivíduo, influenciando profundamente suas crenças, comportamentos, e pensamentos (NICOLELIS; CIGUREL, 2015, p. 61).

Afirmar que algo existe independentemente de um observador (ou de intencionalidade) significa dizer que a sua existência não está condicionada às atitudes, pensamentos, ou intencionalidade de seus observadores, usuários, criadores, compradores ou agentes conscientes intencionais, no geral. Por outro lado, um objeto ou característica que é dependente de um observador, como o próprio nome revela, existe somente na medida em que agentes conscientes intencionais consideram-lhes como tal (é o caso do dinheiro, do casamento, da propriedade, dentre outros) (SEARLE, 1998).

Seres humanos possuem intencionalidade intrínseca, no sentido de que seus processos mentais existem na natureza independentemente de um observador. Conforme ressaltado por Searle, esses estados mentais humanos “são acerca de, ou se referem a, ou são objetos e estados de coisas no mundo diferentes deles mesmos” (SEARLE, 1998, p. 21). Significa dizer, portanto, que possuem conteúdo mental, além de quaisquer estruturas formais que possam ter: a sua existência requer mais do que a mera manipulação de símbolos formais, que por si só são desprovidas de qualquer significado. Ela necessita de uma significação associada a esses símbolos, ou seja, uma semântica – e a sintaxe, por si só, não é capaz de garanti-la⁸.

Por outro lado, programas computacionais operam, em suma, a partir de uma estrutura puramente sintática, desprovidos, portanto, de qualquer conteúdo semântico, ou

⁸ Com o intuito de demonstrar que a semântica não é intrínseca à sintática, Searle propôs o experimento mental que ficou conhecido como o “Experimento do Quarto Chinês”, definido nos seguintes termos: “Imagine que você execute as etapas de um programa elaborado para responder as perguntas em um idioma que você não compreende. Eu não entendo chinês, então imagino que estou trancado em uma sala cheia de caixas com símbolos chineses (a base de dados), recebo uma pequena quantidade de símbolos chineses (perguntas em chinês), e, então, procuro em um manual (o programa) o que deveria fazer. Realizo algumas operações com os símbolos de acordo com as regras (i.e., executo as etapas do programa) e entrego uma pequena quantidade de símbolos (respostas às perguntas) aos que se encontram fora do quarto. Eu sou um computador executando um programa para responder perguntas em chinês, mas ao mesmo tempo não compreendo uma palavra em chinês” (SEARLE, 1998, p. 38).

intencionalidade. Eles mimetizam, apenas, o comportamento intencional humano através do processamento de símbolos (tradicionalmente definidos em sequências de 0's e 1's), ou seja, através de parâmetros pré-estabelecidos de *inputs e outputs*.

Além disso, no que diz respeito à computação, tem-se que, à exceção dos poucos casos em que agentes conscientes de fato computam (somando $2 + 2$ para obter 4, por exemplo), ela é dependente do olhar de um observador, não sendo intrínseca à física do sistema. Ao contrário dos processos mentais, que existem independentemente do significado que outras pessoas a eles atribuem, a computação existe apenas quando relacionada a uma interpretação computacional do fenômeno físico subjacente:

O que dizer sobre os computadores comerciais? Que realidade da física ou da química transforma estas pulsações elétricas em símbolos computacionais? Nenhuma. As palavras “símbolo”, “sintaxe” e “computação” não nomeiam propriedades intrínsecas da natureza como “placa tectônica”, “elétron” ou “consciência”. Os impulsos elétricos são independentes do observador; mas, a interpretação computacional é relativa aos observadores, usuários, programadores etc. (SEARLE, 1998, p. 42-43).

Aos sistemas de inteligência artificial somente é possível atribuir intencionalidade extrínseca, visto que atuam “como se” possuíssem crenças, desejos, ou estados mentais próprios (SEARLE, 1998, p. 45). O seu comportamento é intencional apenas na medida em que é interpretado como tal por um agente – no caso, por um ser humano.

A rejeição da diferença entre intencionalidade intrínseca e extrínseca, quando levada ao extremo, tem como resultado o fato de que tudo no universo age de maneira intencional. Uma vez que todos os objetos e fenômenos universais estão sujeitos às leis da natureza, atuando, conseqüentemente, dentro de um determinado grau de regularidade, poder-se-ia interpretar o seu comportamento “como se” estivesse tentando executar algum projeto ou realizar algum desejo⁹.

Dessa forma, quando um sistema autônomo de inteligência artificial apresenta um comportamento imprevisível do ponto de vista de um observador externo, ou até mesmo para os seus programadores, não é possível interpretá-lo como fruto de uma vontade, ou de um agir intencional por parte dele. Na verdade, como tais sistemas de IA não estão

⁹ Nesse sentido, Searle fornece o exemplo de uma pedra que, ao cair, tenta chegar ao centro da Terra, por possuir tal desejo. Ao fazê-lo, segue a regra $S=1/2gt^2$. Busca-se demonstrar, dessa forma, o absurdo de se ignorar a diferença entre ambos os tipos de intencionalidade, que torna mental tudo no universo (SEARLE, 1998, p. 47).

inerentemente limitados por noções preconcebidas ou determinadas limitações cognitivas do cérebro humano, eles são capazes de conceber soluções que podem não ter sido consideradas, ou até mesmo rejeitadas em favor de opções mais intuitivamente atrativas, por seres humanos.

Conforme os argumentos previamente elencados, tem-se que a analogia entre agentes artificiais autônomos e seres humanos não se sustenta. A interpretação dos atos de uma IA como sendo intencionais reflete uma natural propensão humana de antropomorfizar determinados entes cujo comportamento se assemelhe ao seu (DARLING, 2012).

Conclui-se, portanto, que agentes artificiais autônomos, ao contrário de seres humanos, não possuem, até o momento, os estados subjetivos de consciência imprescindíveis a um comportamento intrinsecamente intencional, não sendo possível, dessa forma, ser-lhes reconhecida a condição de pessoa substancial. Afinal, são pessoas humanas, enquanto agentes conscientes intencionais, que atribuem intencionalidade ao comportamento de uma inteligência artificial.

3. IA E PERSONALIDADE COMO ATRIBUTO

Ocorre que, em grande parte das ordens jurídicas contemporâneas, atribui-se a entes abstratos e claramente desprovidos de intencionalidade intrínseca a personalidade em sentido formal. Ser humano não é uma condição necessária à concessão da personalidade jurídica, sendo o exemplar mais óbvio de tal asserção as corporações modernas. O que, então, justifica a atribuição de personalidade a um determinado ente não humano?

Em sua análise acerca da natureza da pessoa jurídica, Ferrara adota como ponto de partida a distinção entre o sentido filosófico do conceito de pessoa, que no plano da antropologia filosófica designa a natureza do homem, como na já mencionada formulação de Boécio (2005, p. 168) segundo a qual a pessoa é uma “substância individual de natureza racional”, e o sentido jurídico do termo, em que a pessoa goza de um sentido técnico que se identifica com sujeito de Direito (FERRARA apud GONÇALVES, 2016).

Sustenta, então, que a pessoa possui uma realidade distinta e autônoma da realidade material, consistindo em um centro ideal ao qual se imputa a capacidade e a

titularidade de situações jurídicas. A pessoa, dessa forma, afigura-se sempre como uma criação normativa. Até mesmo a subjetividade jurídica do homem é, nesse sentido, uma criação do Direito, fruto de uma atribuição do Estado, razão pela qual o autor justifica a possibilidade de se atribuir igualmente subjetividade jurídica a outras realidades que não a pessoa humana – os entes morais (FERRARA apud GONÇALVES, 2016).

Não obstante, deve-se ter em mente que a personificação de entidades abstratas se apresenta, invariavelmente, como um modo de realização mediata de interesses humanos, ou seja, como uma veste jurídica para a sua realização. Nas palavras de Ferrara, “todo instituto jurídico deve sempre ter como causa imediata ou remota o interesse dos homens, e a este princípio o instituto da personalidade não é exceção” (FERRARA apud GONÇALVES, 2016, p. 31).

Nesse sentido, tem-se que a assimetria existente entre o tratamento conferido às pessoas naturais e aquele dispensado às pessoas jurídicas justifica-se, conforme indicado por Matta Machado, em razão da própria natureza da última. Utilizando-se do método da analogia para análise do tema, o autor afirma que entre a pessoa natural e a pessoa jurídica há uma analogia de proporcionalidade *imprópria*: toma-se por base um termo unívoco, que se transfere a outros sujeitos em virtude das semelhanças que estes guardam entre si (MATTA MACHADO, 1954). Assim,

o termo pessoa, unívoco, designa tanto homem quanto entes coletivos, que, nessa condição, gozam igualmente de existência real. Todavia, essas duas realidades não se assemelham na ordem entitativa, *i.e.* no plano das essências, e sim no plano jurídico da operatividade, *i.e.*, quanto à sua maneira de proceder no campo do Direito (EBERLE, 2006, p. 38).

Por outro lado, entre homem e pessoa natural, firma-se uma analogia de proporcionalidade *própria*, em que o conceito em causa é, ele mesmo, análogo, por identificar sujeitos que se assemelham em caracteres essenciais. Homem e pessoa assemelham-se, portanto, no plano da essência, de maneira tão fundamental que ao legislador não é dado promover a sua dissociação (MATTA MACHADO, 1954, p. 71-72).

Como consequência do debate ocorrido nas últimas décadas acerca dessa possibilidade de dissociação, que se limita às pessoas jurídicas através do instituto da desconsideração, nota-se, mais recentemente, a adoção de uma perspectiva funcional da personalidade jurídica. Significa dizer que, para além de uma análise do que é a pessoa

jurídica, passou-se a investigar para que ela serve, qual a sua função dentro da ordem jurídica (FERRARA apud GONÇALVES, 2016, p. 51).

3.1. As razões da pessoa jurídica

Tendo em vista tal escopo funcional da personificação, Negri (2016, p. 8) aponta para o fato de que as razões pelas quais se atribui personalidade também a entes coletivos não se baseiam em uma espécie de ontologia moral correspondente a estes, mas revelam, na realidade, uma função heurística: “o termo pessoa jurídica, ao simplificar um complexo de relações jurídicas, facilita a descrição do próprio fenômeno jurídico”.

Nada impede que todos os integrantes de uma sociedade ou associação, por exemplo, relacionem-se conjuntamente com terceiros. Entretanto, a interposição de uma nova subjetividade jurídica permite que esse mesmo processo possa ser descrito de forma unitária, sem a necessidade de menção, a cada ato praticado, de todos os sócios que se mostram retratados pelo termo pessoa jurídica (NEGRI, 2016).

Para além da simplificação de relações jurídicas, o recurso à personificação constitui um importante instrumento para a articulação patrimonial, fazendo com que a separação de um patrimônio em virtude de uma determinada finalidade possa ser melhor compreendida a partir da criação de um novo sujeito. A viabilidade de uma estrutura patrimonial autônoma do ente abstrato, que não se confunde com aquela de seus membros, revela-se, assim, dependente do reconhecimento de uma nova subjetividade jurídica, reduzindo-se desse modo a temática da separação patrimonial à análise das consequências do processo de personificação (NEGRI, 2016, p. 9).

Finalmente, o autor aponta também como razão por trás da pessoa jurídica a constituição de um sistema de imputação direta pelos atos praticados por seus órgãos, através da mediação orgânica, desenvolvida em contraposição à representação. Enquanto nesta representante e representado são vistos como sujeitos distintos, tem-se que naquela o órgão é concebido como parte do próprio ente coletivo, possibilitando, ao menos em tese, um processo de imputação mais amplo. Não se atribui à pessoa jurídica, assim, apenas os efeitos correlatos ao ato praticado, mas o ordenamento a reconhece como verdadeira autora do mesmo (NEGRI, 2016).

Verifica-se, dessa forma, que a personalidade jurídica, ao contrário do que ocorre com a pessoa natural, não se fundamenta na existência de determinada ontologia moral subjacente ao ente personificado. Com efeito, a partir de uma análise funcional do instituto, é possível afirmar que as pessoas jurídicas, em última análise, servem a fins de pessoas humanas, e que o ordenamento normativo concede a personalidade a tais entes com base em critérios de utilidade, atendendo a demandas sociais de indivíduos que já os compreendem e com eles interagem como entidades distintas – seja de seus membros, ou seja de seus controladores ou programadores.

É possível concluir, portanto, que a impossibilidade de se replicar em agentes artificiais estados subjetivos de consciência, como intencionalidade, não representa, por si só, um óbice à personificação formal dos mesmos, com as respectivas implicações decorrentes no que concerne à delimitação de sua capacidade de direito, em virtude da existência da categoria das pessoas jurídicas.

Entretanto, uma eventual atribuição de personalidade a esses agentes deverá ser pautada, necessariamente, em uma análise funcional do instituto, que leve em consideração os interesses dos indivíduos humanos que com eles se relacionarão para que estes, como agentes conscientes intencionais, não tenham eventualmente seus direitos minimizados em prol de sistemas artificiais¹⁰.

CONSIDERAÇÕES FINAIS

No presente trabalho, buscou-se investigar se agentes autônomos de inteligência artificial poderiam ser considerados como pessoas pelo Direito. Para fins didáticos, optou-se por uma abordagem que trouxesse consigo as diferenças entre aquilo que se convencionou denominar personalidade substancial (ou personalidade como valor) e personalidade formal (ou personalidade como atributo).

A partir das dimensões incontornáveis do conceito ético-normativo de pessoa (autonomia moral, alteridade e dignidade), analisou-se, inicialmente, a possibilidade de enquadramento de agentes de inteligência artificial autônomos nessa categoria,

¹⁰ Não faz parte do escopo do presente trabalho apresentar os argumentos favoráveis e contrários à atribuição de personalidade a sistemas de inteligência artificial, mas apenas investigar a possibilidade ou não de sua personificação.

verificando até que ponto estes seriam capazes de reproduzir os estados subjetivos internos que os seres humanos possuem e que permitem a sua compreensão como pessoas.

Com base na teoria do cérebro relativístico proposta por Miguel Nicolelis e Ronald Cicurel, foi argumentado que os atributos que definem o cérebro humano como um sistema complexo adaptativo fazem com que as tentativas de simular de maneira acurada, utilizando-se um ponto de vista computacional ou puramente matemático, estejam fadadas ao fracasso.

Nesse tipo de sistema, a informação encontra-se altamente dependente do substrato orgânico e, portanto, inexistente a dicotomia *software/hardware* para descrever de maneira distinta o modo de processamento integrado de informação que é feito pelo sistema.

Desse modo, viu-se que seres humanos possuem intencionalidade intrínseca, por possuírem um conteúdo mental cuja existência independe do olhar externo de um observador. A agentes artificiais autônomos, por outro lado, por operarem a partir de uma estrutura puramente sintática, somente é possível atribuir intencionalidade extrínseca, visto que atuam “como se” possuíssem crenças, desejos, ou estados mentais próprios.

Concluiu-se, assim, que o comportamento de sistemas de inteligência artificial, ainda que imprevisível, apenas é intencional na medida em que é interpretado como tal por um agente – no caso, por um ser humano. Desse modo, eles não possuem os estados subjetivos de consciência imprescindíveis a um comportamento intencional intrínseco, não sendo possível reconhecê-los a condição de pessoa substancial.

Dando prosseguimento à pesquisa, e considerando-se que grande parte dos ordenamentos jurídicos contemporâneos atribui a entes abstratos claramente desprovidos de intencionalidade intrínseca o status formal de pessoas, passou-se à análise da personalidade como atributo.

A partir da análise funcional do, notou-se que, no que tange à pessoa jurídica, inexistente qualquer espécie de ontologia moral que fundamente a sua personificação. Pelo contrário, foi elencada uma série de razões que informam a personalidade jurídica, como a articulação patrimonial, a constituição de um sistema de imputação direta pelos atos praticados pelos seus órgãos e a simplificação de relações jurídicas em um único centro de imputação. Em suma, foi possível verificar que a concessão de personalidade a tais entes é comumente fundamentada a partir de critérios de utilidade.

Na presente pesquisa, concluiu-se que a ausência de intencionalidade intrínseca em agentes artificiais autônomos não representa, por si só, um óbice à sua personificação formal. Embora a eles não possa ser reconhecido o *status* de pessoa natural, é possível a sua subjetivação através do instituto da personalidade jurídica. Esta, por sua vez, deverá necessariamente ser pautada em um escopo funcional da personalidade, levando sempre em consideração os interesses de pessoas humanas que com elas estabelecerão relações comunicativas relevantes.

REFERÊNCIAS

BOÉCIO, Anício Severino. *Escritos (Opuscula Sacra)*. [Trad.] Juvenal Savian Filho. São Paulo: Martins Fontes, 2005.

BROZEK, Bartosz. The Troublesome ‘Person’. In: KURKI, Visa A. J.; PIETRZYKOWSKI, Tomasz (Eds.). *Legal Personhood: Animals, artificial intelligence and the unborn*. Cham: Springer, 2017, p. 3-14.

ČERKA, Paulius; GRIGIENĖ, Jurgita; SIRBIKYTĖ, Gintarė. Liability for damages caused by artificial intelligence. *Computer Law & Security Review*, v. 31, n. 3, 2015, p. 376-389.

COPELAND, B. Jack. The Church-Turing Thesis. In: ZALTA, Edward N. *The Stanford Encyclopedia of Philosophy*, 2017. Disponível em: <https://plato.stanford.edu/archives/win2017/entries/church-turing/>. Acesso em: 15 mar. 2020.

DARLING, Kate. *Extending Legal Rights to Social Robots*. We Robot Conference, Miami, 2012, p. 1-18. Disponível em: http://robots.law.miami.edu/wp-content/uploads/2012/04/Darling_Extending-Legal-Rights-to-Social-Robots-v2.pdf. Acesso em: 12 mar. 2020.

EBERLE, Simone. *A Capacidade entre o Fato e o Direito*. Porto Alegre: Sergio Antonio Fabris, 2006.

GONÇALVES, Diogo Costa. Contributo para o Estudo da Pessoa Jurídica no Direito Civil Brasileiro. *Civilística*, a. 5, n. 1, 2016, p. 1-57.

KAPLAN, Jerry. *Humans Need Not Apply: A guide to wealth and work in the age of artificial intelligence*. New Haven: Yale University Press, 2015.

KURZWEIL, Ray. *The Singularity is Near: When humans transcend biology*. Nova York: Penguin Books, 2005.

MATTA MACHADO, Edgar de Godoi da. Conceito Analógico de Pessoa Aplicado à Personalidade Jurídica. *Revista da Faculdade de Direito da Universidade Federal de Minas Gerais*, v. 6, 1954, p. 55-78.

MCCARTHY et al. *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, 1955. Disponível em: <<http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>>. Acesso em: 10 mar. 2020.

NICOLELIS, Miguel A. L.; CICUREL, Ronald. *O Cérebro Relativístico: Como ele funciona e por que ele não pode ser simulado por uma máquina de Turing*. São Paulo: Kios Press, 2015.

PICCININI, Gualtiero. Computationalism in the Philosophy of Mind. *Philosophy Compass*, 2009, p. 1-21.

PIRES, Thatiane; SILVA, Rafael. A Responsabilidade Civil pelos Atos Autônomos da Inteligência Artificial: Notas iniciais sobre a resolução do Parlamento Europeu. *Revista Brasileira de Políticas Públicas*, v. 7, n. 3, 2017, p. 239-254.

SEARLE, John. *O Mistério da Consciência*. [Trad.] André Pinheiro e Vladimir Safatle. São Paulo: Paz e Terra, 1998.

SILVA, Denis Franco. *Do humano ao pós-humano: Pessoa e autonomia privada no contexto do aperfeiçoamento biônico*. 2009. Tese (Doutorado). Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2009.

SOLUM, Lawrence B. Legal Personhood for Artificial Intelligences. *North Carolina Law Review*, v. 70, n. 4, 1992, p. 1231- 1287.

SPAEMANN, Robert. *Personae: Sulla differenza tra “qualcosa” e “qualcuno”*. [Trad.] Leonardo Allodi. Bari: Editori Laterza e Figli, 2007.

STANCIOLI, Brunello. *Renúncia ao Exercício de Direitos da Personalidade: Ou como alguém se torna o que quisier*. Belo Horizonte: Del Rey, 2010.

TAYLOR, Charles. *As fontes do self: A construção da identidade moderna*. [Trad.] Adail Ubirajara Sobral e Dinah de Abreu Azevedo. São Paulo: Loyola, 2005.

TEPEDINO, Gustavo. A tutela da Personalidade no Ordenamento Civil Constitucional. In: TEPEDINO, Gustavo (Org.). *Temas de Direito Civil*. 2 ed. Rio de Janeiro: Renovar, 2001, p. 23-54.

UNIÃO EUROPEIA, *European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103-INL)*. Estrasburgo: Parlamento Europeu, 2017.

VLADECK, David C. *Machines Without Principles: Liability Rules and Artificial Intelligence*. *Washington Law Review*, v. 89, n. 01, 2014, p. 117-150.