

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
FACULDADE DE DIREITO

GIOVANA FIGUEIREDO PELUSO LOPES

INTELIGÊNCIA ARTIFICIAL E PESSOALIDADE

Juiz de Fora

2017

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
FACULDADE DE DIREITO

INTELIGÊNCIA ARTIFICIAL E PESSOALIDADE

Trabalho de conclusão de curso apresentado à Faculdade de Direito da Universidade Federal de Juiz de Fora como requisito parcial para obtenção do grau de Bacharel em Direito, sob orientação do Professor Dr. Denis Franco Silva.

Juiz de Fora

2017

FOLHA DE APROVAÇÃO

GIOVANA FIGUEIREDO PELUSO LOPES

INTELIGÊNCIA ARTIFICIAL E PESSOALIDADE

Trabalho de conclusão de curso apresentado à Faculdade de Direito da Universidade Federal de Juiz de Fora como requisito parcial para obtenção do grau de Bacharel em Direito, submetido à Banca Examinadora composta pelos membros:

Orientador: Prof. Dr. Denis Franco Silva

Universidade Federal de Juiz de Fora

Prof. Dr. Flávio Henrique Silva

Universidade Federal de Juiz de Fora

Prof. Dr. Sérgio de Carvalho Ávila Negri

Universidade Federal de Juiz de Fora

PARECER DA BANCA:

() APROVADO

() REPROVADO

Juiz de Fora, de de 2017

RESUMO

O presente trabalho busca investigar se sistemas autônomos baseados em inteligência artificial devem ser tratados como pessoas pelo ordenamento jurídico, seja na acepção formal ou substancial do termo, e a quem deverá ser imputada responsabilidade em casos de eventuais danos por eles causados. A fim de responder aos questionamentos propostos, será inicialmente investigado se tais agentes autônomos possuem os estados subjetivos imprescindíveis para a intencionalidade intrínseca que possibilitam a dimensão da agência, utilizando-se como referencial teórico a concepção de John Searle acerca da consciência. Em seguida, serão analisadas as razões que informam a personificação de pessoas jurídicas para avaliar se as mesmas podem se aplicar a sistemas de inteligência artificial, ou se a subjetivação de tais entes promoveria aquilo que Rodotà denomina de “expropriação da subjetividade”. Finalmente, concluir-se-á que agentes artificiais autônomos não devem ser qualificados como pessoas para o Direito, sugerindo-se que a responsabilização por eventuais danos por eles causados seja imputada àquele que disponibilizou a tecnologia ao público, de forma objetiva e agravada.

Palavras-chave: Inteligência artificial; personalidade; filosofia da mente; expropriação da subjetividade; responsabilidade objetiva.

ABSTRACT

This paper aims to investigate if Law should regard autonomous systems based on artificial intelligence algorithms as persons, whether in the formal or substantial sense of the term, and to whom liability should be attributed in case of any damage caused by them. In order to answer these questions, it will be first investigated if such autonomous agents do possess the subjective states essential for intrinsic intentionality as a condition to the dimension of agency, using John Searle's conception of consciousness as theoretical framework. Following, the reasons that inform the personification of legal entities will be analyzed in order to assess if they can also be applied to artificial intelligence systems or if it would represent a manifestation of the phenomena Rodotà named as "expropriation of subjectivity". The research concludes that autonomous artificial agents should not be qualified as persons by Law, and that the liability for any damages caused by them should lie on those who made the technology available to the public following a model of aggravated strict liability.

Keywords: Artificial intelligence; philosophy of mind; expropriation of subjectivity; strict liability.

Sumário

1. Introdução	6
2. Uma breve história da IA	7
3. Personalidade, intencionalidade e agência	11
4.O problema da consciência	14
5. A consciência segundo John Searle	16
6. As razões da pessoa jurídica e a expropriação da subjetividade.....	20
7. Conclusão: IA, pessoa ou ferramenta?	25
8. Referências bibliográficas	26

1. Introdução

Atualmente, sistemas baseados em inteligência artificial (IA) já operam nas mais diversas áreas do cotidiano, passando muitas das vezes despercebidos por seus usuários. Inúmeras realizações tecnológicas que antes pareciam pertencer apenas à ficção científica acabaram por se tornar reais e pertencentes à prática humana. Carros autônomos já vagam pelas ruas de algumas cidades do mundo, e sua inevitável chegada ao mercado de consumo poderá revolucionar o transporte rodoviário. Máquinas autônomas são capazes de executar complexas operações financeiras e de identificar, por meio de *softwares* de reconhecimento facial, potenciais terroristas. Em terreno mais rotineiro, tem-se a aplicação maciça de *machine learning*, subcampo da IA, em pesquisas, anúncios e até mesmo vídeos do *Youtube*.

Diante de sua crescente ubiquidade e enorme potencial de comercialização, a inteligência artificial tem gerado investimentos maciços por parte do setor privado, e o prognóstico que se apresenta é que seu uso torne-se cada vez mais disseminado em um futuro próximo.

Surgem, a partir disso, indagações acerca de como deverá ser atribuída responsabilidade - moral e legal - na hipótese de ocorrência de um dano envolvendo um agente artificial, principalmente diante de situações em que há renúncia de controle direto por seres humanos em processos de tomada de decisões.

O problema a ser investigado no presente trabalho, portanto, pode ser colocado da seguinte forma: deve o ordenamento jurídico considerar sistemas de inteligência artificial como centros autônomos de imputação de direitos e deveres, ou seja, como *peessoas*? E ainda, como desdobramento direto do problema principal, a quem deverá ser imputada responsabilidade pelas ações de uma IA em caso de eventuais danos?

A hipótese que se formula é de que agentes artificiais autônomos não devem ser tratados como pessoas pelo Direito, seja em sua acepção substancial ou formal. A responsabilidade por eventuais danos causados por suas condutas deverá ser atribuída, dessa forma, àquele que disponibilizou a tecnologia ao público.

A fim de se analisar não apenas a possibilidade, mas a adequação em se atribuir personalidade a tais agentes, adotar-se-á como referencial teórico a concepção de John Searle acerca da consciência, sobretudo no que diz respeito a dois pontos específicos: o reconhecimento do papel “causal” do cérebro humano na constituição da consciência, e a diferenciação entre intencionalidade intrínseca e extrínseca.

A amplitude das bases teóricas sobre as quais se funda a pesquisa proposta implica, necessariamente, em seu desenvolvimento a partir de uma perspectiva interdisciplinar que aborde, além dos aspectos estritamente jurídicos do problema, aspectos históricos, filosóficos e relativos à ciência da cognição e da informação.

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.

2. Uma breve história da IA

A melhor definição acerca do que seja inteligência artificial talvez tenha sido aquela formulada em 1955 por John McCarthy, cientista norte-americano da computação que, juntamente com seus colegas, estruturou a questão nos seguintes termos: fazer com que uma máquina se comporte de maneira que, caso se tratasse de um ser humano, seria considerada inteligente¹ (MCCARTHY et al., 1955).

De fato, ainda que inexista consenso na literatura especializada sobre o conceito de inteligência artificial, é possível afirmar, em linhas gerais, tratar-se da tentativa de reprodução, em sistemas artificiais, da cognição humana e seus mais variados componentes, como o aprendizado, a memória e o processo de tomada de decisões.

A definição de inteligência artificial aqui apresentada enquadra-se, dessa forma, naquilo que Searle (2000) denomina “IA fraca”: partindo-se do pressuposto de que programas computacionais são ferramentas úteis para o estudo da mente humana e o desenvolvimento de tecnologias, tem-se que um programa de inteligência artificial é, no máximo, a *simulação* de um processo cognitivo – e não um processo cognitivo por si só.

Por outro lado, adeptos de uma “IA forte” acreditam que “a mente está para o cérebro tal como o programa está para o *hardware* do computador” (SEARLE, 2000, p. 36). Sendo assim, qualquer sistema físico que tivesse um programa correto (isto é, com *inputs* e *outputs* corretos) possuiria uma mente, no mesmo sentido que seres humanos possuem mentes. O único impedimento, portanto, seria o fato de ainda não terem sido criadas as bases e

¹ O termo foi apresentado à Universidade de Dartmouth por John McCarthy, Marvin Minsky, Nathaniel Rochester e Claude Shannon como proposta de estudos para o verão de 1956 acerca de questões relacionadas à inteligência artificial. O *workshop* que sucedeu à pesquisa, nos meses de julho e agosto, acabou por se tornar o marco oficial do nascimento da IA. Para mais, veja-se: <<http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>>.

programas necessários para tanto, tratando-se apenas de uma questão de tempo até que seja possível *replicar* a consciência humana.

Com efeito, dotar computadores de uma inteligência similar àquela possuída por seres humanos tem sido um sonho de especialistas em informática desde o surgimento da computação eletrônica. Já em 1950, no célebre trabalho intitulado *Computing Machinery and Intelligence*, Alan Turing questionou acerca da possibilidade de máquinas pensarem, concluindo, posteriormente, tratar-se de uma questão muito sem sentido para merecer discussão (TURING, 1950, p. 8). Ao invés, direcionou seu foco para a potencialidade de computadores digitais reproduzirem não propriamente processos mentais humanos, mas as manifestações externas desses processos – propondo, como método de aferição, o Teste de Turing, utilizado até os dias atuais.

Em sua versão contemporânea, coloca-se um ser humano para interagir, apenas por meio de texto e por não mais que cinco minutos, com um computador ou com outro ser humano, devendo adivinhar, ao final do tempo, com qual dos dois estava interagindo. Se um computador conseguir consistentemente enganar os participantes, considera-se que o mesmo é inteligente². Note-se, portanto, que o Teste de Turing representa uma abordagem de IA fraca para o problema de se comprovar a IA forte.

Nas décadas subsequentes, o campo da inteligência artificial passou por momentos de expansão e retração, na medida em que a resolução de alguns dos problemas inicialmente propostos revelou-se mais difícil do que o previsto inicialmente, e outros, simplesmente insuperáveis com as tecnologias disponíveis à época. Foi apenas no final dos anos de 1990 que as inovações em IA passaram a ocorrer em ritmo acelerado, com os pesquisadores da área direcionando seu foco à aplicação da tecnologia a problemas do mundo real, como reconhecimento de imagem e diagnóstico médico (ESTADOS UNIDOS, 2016).

Nesse contexto, é possível destacar alguns importantes marcos na história da inteligência artificial, como a vitória, em 1997, do computador *Deep Blue*, desenvolvido pela IBM, sobre o campeão mundial de xadrez Garry Kasparov³. Outros avanços significativos incluem o desenvolvimento, pela agência governamental norte-americana DARPA (*Defense Advanced Research Projects Agency*), de um agente cognitivo que aprende e organiza –

² Para comentários acerca do Prêmio Loebner, competição anual de inteligência artificial baseada no Teste de Turing, veja-se: FLORIDI, Luciano; TADDEO, Maria Rosaria; TURILLI, Matteo. Turing's Imitation Game: Still an Impossible Challenge for All Machines and Some Judges – An Evaluation of the 2008 Loebner Contest. *Minds and Machines*, v. 19, n. 1, 2008, p. 145-150.

³LATSON, Jennifer. Did Deep Blue Beat Kasparov Because of a System Glitch? *Time*, 2015. Disponível em: <<http://time.com/3705316/deep-blue-kasparov/>>. Acesso em: 14 ago. 2017.

CALO, na sigla inglesa -, e que levou à criação do assistente pessoal *Siri* pela Apple⁴; e a vitória do computador *Watson*, desenvolvido também pela IBM, no jogo televisivo de perguntas e respostas *Jeopardy!*, em 2011⁵.

O atual entusiasmo no que diz respeito à pesquisa e desenvolvimento de tecnologias de inteligência artificial, entretanto, teve início aproximadamente em 2010, e foi movido pelos seguintes fatores: i) a existência de métodos estatísticos e probabilísticos cada vez mais sofisticados; ii) a disponibilidade de um número expansivo de dados; iii) a acessibilidade a um enorme poder computacional a baixo custo; iv) e a crescente adequação de ambientes às novas tecnologias, como a automação residencial e a criação de cidades ‘inteligentes’ (FLORIDI et al., 2017). Tais fatores, que se retroalimentam, possibilitaram o crescimento exponencial da criação e aperfeiçoamento de sistemas de IA nos últimos anos, não aparentando ser uma tendência passageira.

Para além das previsões mais otimistas, surgem indagações acerca do impacto da utilização de tecnologias de IA sobre aspectos fundamentais da vida em sociedade. Nesse sentido, podem ser apontadas as preocupações relativas ao mercado de trabalho, visto que, na medida em que agentes artificiais forem substituindo seres humanos na realização de determinadas funções - em especial aquelas que exigem uma menor qualificação -, corre-se o risco de que se acentue o desemprego e, conseqüentemente, a desigualdade econômica⁶.

Ao problema do desemprego soma-se, ainda, outro: trabalhadores pagam impostos, máquinas não. Se, por um lado, os trabalhos “robotizados” requerem maiores recursos financeiros para amortizar o desemprego, por outro, geram menos recursos fiscais. A solução pode estar na taxação, direta ou indiretamente, de IAs, conforme sugere o relatório feito pelo Comitê de Assuntos Jurídicos do Parlamento Europeu sobre “*Civil Law Rules on Robotics*”(UNIÃO EUROPEIA, 2016).

Uma forma de se justificar a imposição seria considerar tais agentes como entes subjetivos pelo Fisco. Como não se enquadram nem na categoria de pessoas naturais nem na

⁴ULANOFF, Lance. Apple is Turning Siri into a Next-level Artificial Intelligence. *Mashable*. Disponível em: <<http://mashable.com/2016/06/13/siri-sirikit-wwdc2016-analysis/#vPOglN1eWiqn>>. Acesso em: 14 ago. 2017.

⁵ MARKOFF, John. Computer Wins on ‘Jeopardy!’: Trivial, it’s not. *The New York Times*. Disponível em: <<http://www.nytimes.com/2011/02/17/science/17jeopardy-watson.html?pagewanted=all&mcubz=3>>. Acesso em: 14 ago. 2017.

⁶ O futuro impacto da inteligência artificial sobre o mercado de trabalho é considerado tão significativo que, em dezembro de 2016, foi publicado um relatório pela Casa Branca intitulado “*Artificial Intelligence, Automation and the Economy*”, que apresenta propostas políticas específicas sobre como lidar com as mudanças acarretadas. São elas: investir e desenvolver IAs pelos seus inúmeros benefícios; educar e treinar os americanos para os trabalhos do futuro; e auxiliar e “empoderar” (*empower*) os trabalhadores de forma a garantir um amplo crescimento compartilhado (ESTADOS UNIDOS, 2016).

de pessoas jurídicas, uma das sugestões do relatório é que se lhes atribua o status de “pessoas eletrônicas”⁷ - no sentido técnico, é claro, de entidades às quais o ordenamento confere capacidade jurídica, tornando-se, assim, sujeitos de direito (FLORIDI, 2017).

O uso expansivo de tecnologias de IA gera, ainda, as já mencionadas preocupações acerca de como se dará a responsabilização quando da ocorrência de um dano envolvendo um agente artificial, sobretudo em casos em que o de controle direto por pessoas humanas em processos de tomada de decisão for renunciado.

Questiona-se, a título de exemplo, quem – ou o que – será responsabilizado quando um veículo autônomo causar um acidente, ou até que ponto médicos poderão delegar a tarefa de diagnosticar as condições de saúde de um paciente a programas de escaneamento inteligente, sem se exporem ao risco de serem responsabilizados por negligência quando um erro for cometido pelo sistema.

A questão se torna especialmente problemática a partir da possibilidade de criação de armas autônomas letais (no inglês, *lethal autonomous weapons*, ou LAWs), definindo-se a autonomia, aqui, como a capacidade de um sistema para operar livre de direção, monitoramento ou controle humano. (SCHERER, 2016). Diferenciam-se, portanto, de armamentos militares já existentes que, embora façam uso de tecnologias de inteligência artificial, ainda não são capazes, por exemplo, de selecionar alvos ou de disparar, sem um *input* humano direto⁸.

Soma-se a todo o exposto até aqui o fato de que as tecnologias de inteligência artificial desenvolveram-se, até muito recentemente, em uma espécie de vácuo normativo, a exceção de algumas legislações específicas que regulam a utilização de veículos autônomos, drones e programas-robôs no comércio eletrônico. Conforme assinala Scherer (2017), são poucas as leis ou regulamentos existentes que abordam especificamente os desafios únicos gerados pela

⁷ Ao tratar da questão da responsabilidade, o relatório solicita que a Comissão, ao avaliar o 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 pessoas eletrônicas, com direitos e obrigações específicas, incluindo-se o dever de reparar qualquer dano que possam causar; aplicando-se a personalidade eletrônica a casos em que robôs tomem decisões inteligentes e autônomas, ou interajam com terceiros independentemente (UNIÃO EUROPEIA, 2016).

⁸Os proponentes do uso de armas autônomas letais costumam apontar duas supostas consequências vantajosas para tanto. A primeira seria que o caráter autônomo dessas máquinas tornaria desnecessária a presença de humanos na linha de fogo, salvando, assim, inúmeras vidas. A segunda, que, se apropriadamente programadas, tais máquinas seriam mais éticas que o ser humano médio – não estariam tentadas, por exemplo, a cometerem crimes de guerra (MICHALCZAK, 2017). A possibilidade de se programar tais armas em conformidade com as normas de Direito Internacional Humanitário, porém, é altamente questionável. Nesse sentido, veja-se: SCHERER, Matthew. Who's to Blame (Part 2): Could Autonomous Weapon Systems Navigate the Law of Armed Conflict? *Law and AI*, 2017. Disponível em: <<http://www.lawandai.com/2016/02/17/whos-to-blame-part-3-could-autonomous-weapon-systems-navigate-the-law-of-armed-conflict/>>. Acesso em: 4 set. 2017.

IA, e a discussão sobre sua possível personificação e responsabilização mostra-se incipiente também no âmbito acadêmico.

Justamente pela inteligência artificial levantar questões fundamentais relativas aos seus impactos éticos, sociais e econômicos, em outubro de 2016, foram lançados relatórios iniciais pelo *White House Office of Science and Technology Policy* pelo *House of Commons' Science and Technology Committee*, para além do já mencionado relatório formulado pelo Parlamento Europeu, abordando as questões mais prementes relativas à temática, bem como precauções a serem adotadas em conjunto pelos diversos setores da sociedade, de forma a se preparar para o futuro da IA.

Apesar de suas diferenças, é possível identificar, dentre os três relatórios, valores comuns a serem reforçados, em especial no que diz respeito à transparência, à responsabilidade (*accountability*) e ao impacto positivo de tais tecnologias sobre a economia e a sociedade. Todos eles identificam, ainda, as questões mais prementes relativas à inteligência artificial a serem tratadas, como suas possíveis consequências sobre a educação, o mercado de trabalho e a segurança nacional.

Ainda que preparados independentemente uns dos outros, o fato de os três relatórios terem sido lançados à mesma época revela como a questão tratada clama, urgentemente, por atenção e normatização – a começar pelo questionamento acerca do possível tratamento de sistemas artificiais autônomos como *pessoas*.

2. Personalidade, intencionalidade e agência

A fim de responder aos questionamentos levantados, faz-se necessário, em um primeiro momento, delimitar o que exatamente se quer dizer quando da utilização do termo “pessoa”.

A rigor, há dois sentidos técnicos para o conceito de personalidade, para o Direito. O primeiro associa-se à qualidade para ser sujeito de direito, sendo aplicável, assim, tanto às pessoas físicas quanto às jurídicas. Trata-se da personalidade enquanto atributo, ou personalidade formal. O segundo traduz 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 enquanto valor, ou personalidade substancial, sendo, portanto, peculiar à pessoa natural (TEPEDINO, 2008).

Inicialmente, quando do início da busca por cientificidade no âmbito jurídico-normativo, definiu-se o conceito de pessoa de forma qualitativa e apenas em seu sentido puramente formal, como a aptidão genérica para figurar em um dos polos de uma relação jurídica, titularizando direitos e deveres.

Conforme aponta Mota Pinto (1996, p. 84-85), neste sentido técnico-jurídico inexistência necessária entre a noção de pessoa, ou sujeito de direito, e a noção de ser humano. Entretanto, se o Direito visa, em última instância, a disciplina de interesses humanos, é logicamente forçoso que ao menos alguns homens sejam dotados de personalidade jurídica.

Atualmente, todos os seres humanos são considerados pessoas pelo ordenamento, isto é, a eles deve-se conferir o atributo técnico da personalidade. Tal compreensão desenvolveu-se justamente através da rejeição a um conceito meramente formal de pessoa, visto que, ao admitir a exclusão de seres humanos do universo personificado, possibilita, conseqüentemente, a sua exclusão do universo de portadores de direitos e deveres, igualando-os a entes meramente instrumentais, como as pessoas jurídicas (SILVA, 2009).

Afirmar que todo ser humano é pessoa para o Direito, note-se, é considerar que a personalidade, enquanto atributo técnico, não pode se encontrar desvinculada da personalidade enquanto valor, em se tratando de pessoas humanas. A ênfase conferida à humanidade nos discursos acerca da personalidade, aponta Silva (2009), não significa a adesão a um pensamento jusnaturalista, mas traduz a necessidade de se atentar para o substrato material personificado, ou seja, para as bases deontológicas de inclusão na categoria. Isto porque

tendo em vista as bases do reconhecimento da personalidade, 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* enquanto valor sem fazer menção a este “conteúdo” ou “substância”, ao substrato material da personalidade⁹, não através da mera descrição do conjunto de entes personificados pelo ordenamento, mas da análise das características essenciais ao seu reconhecimento como tal – um conceito normativo de pessoa.

⁹ Similarmente, Maria Celina Bodin de Moraes (2006, p. 118-119), ao fazer referência ao substrato material da dignidade, identifica-o com a existência de um sujeito moral, que reconhece a existência dos outros como sujeitos iguais a ele, merecedores do mesmo respeito à integridade psicofísica, dotado de vontade livre e integrante do grupo social.

Ao se questionar acerca do que é ser pessoa, como um *self* dentro do contexto da identidade moderna, Silva (2009, p. 42) propõe que se deve considerar como tal todo ente que apresente racionalidade autônoma, recorte psicofísico individual e que se encontre em relação aos outros. Como sujeito racional e eminentemente livre, tem-se que a autonomia encontra-se no centro do conceito contemporâneo de pessoa, sendo esta constituída por sua capacidade de eleger fins, formular desejos, possuir interesses e, eventualmente, revisá-los e até mesmo abandoná-los.

O estar em relação ao outro, por sua vez, revela o caráter essencialmente intersubjetivo do conceito, somente sendo possível se constituir e se afirmar como pessoa perante o outro e dependente do outro. Conforme explica Taylor (1997, p. 55), 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 definitiva ou a “redes de interlocução”.

Finalmente, ser pessoa compreende a ideia de individualidade, ou seja, não apenas afirmar a si mesmo como um agente racional capaz de eleger fins, possuir desejos e crenças, mas também compreender que estes são distintos dos fins, desejos e crenças de outros. Pressupõe-se, assim, a existência de uma fronteira identitária entre o “eu” e o outro, uma extensão espaço-temporal representada de forma privilegiada pelo corpo (SILVA, 2009).

As características apresentadas dependem também de certa estabilidade ao longo do tempo – estabilidade esta que constitui a identidade pessoal de um sujeito, e é comumente definida em termos de autoconsciência reflexiva e continuidade de estados psicológicos. Em virtude do caráter eminentemente relacional do conceito de pessoa, dependente de interação social, não é possível afirmar que esta identidade pessoal seja estática, pelo contrário, ela é essencialmente mutável, sem o que não se poderia cogitar em liberdade ou autodeterminação (SILVA, 2009).

Para os fins intentados pelo presente trabalho cumpre, ainda, fazer uma breve menção ao conceito de agência, intimamente conectado com o conceito de pessoa, bem como à sua relação com as noções de intencionalidade e imputabilidade. Agência diz respeito a processos de determinação de causalidade normativa, e não meramente físico-causal. Se, por exemplo, um indivíduo, ao ser empurrado contra um objeto, causa a sua destruição, não se pode afirmar que foi ele quem causou o resultado, mas sim aquele que o empurrou. A agência pressupõe, dessa forma, intencionalidade na causa.

Difere, assim, da problemática da imputabilidade, que não se refere ao estabelecimento de uma relação de causalidade normativa, mas sim diz respeito à intencionalidade quanto ao resultado. É justamente em razão disso que é possível identificar determinadas ações que, embora sejam causa adequada a um resultado, não resultam na imputação normativa desse resultado ao agente que o causou, seja pela ausência de intencionalidade direta ou indireta quanto a ele (ausência de dolo e culpa, respectivamente).

Da mesma forma, é possível identificar situações em que, ainda que ausente a intencionalidade quanto ao resultado, imputa-se responsabilidade ao agente quando presente a intencionalidade na causa. Cita-se, nesse sentido, a modalidade agravada de responsabilidade civil objetiva, prevista no parágrafo único do artigo 927 do Código Civil brasileiro.

Embora tenham sido aqui tratadas de maneira resumida, é possível notar como a problemática da pessoalidade, agência e responsabilidade são altamente dependentes das relações que se estabelecem entre mente e corpo. Racionalidade, capacidade para se autodeterminar e para agir intencionalmente são atributos que dependem, em última instância, da consciência, fazendo-se necessário, portanto, analisar exatamente em que consiste a mesma, e se ela se faz presente não apenas em indivíduos humanos, mas também em agentes artificiais autônomos.

4. O problema da consciência

Por se tratar da tentativa de replicar a cognição humana através de processos artificiais, a inteligência artificial possui uma relação direta de interdependência com o estudo da consciência. Ocorre que, conforme aponta Searle (1998), inexiste um consenso, inclusive dentre a comunidade científica, acerca do que venha a ser a consciência, e sobre como exatamente os processos neurobiológicos no cérebro são capazes de causá-la.

O problema, ressalta, encontra-se no fato de que ainda não há um princípio teórico unificador da consciência, a exemplo de uma teoria da seleção natural, no campo da evolução, ou de uma teoria das placas tectônicas, na geologia – não existe uma teoria única de como o cérebro funciona (SEARLE, 1998). Resta a consciência, dessa forma, como *mistério*:

Pelo que sabemos, os processos relevantes ocorrem nos níveis micro das sinapses, neurônios, feixes de neurônios e complexos celulares. Toda nossa vida consciente é causada por esses processos de nível inferior, mas só temos uma vaga ideia de como eles funcionam (SEARLE, 1998, p. 31).

Disso resultam inúmeras teorias que tentam explicar – e até mesmo refutar – o imemorial problema da mente-corpo, em torno do qual giram questões como: corpo e mente são duas entidades separadas? Se sim, como se mantêm conjuntamente em um único organismo vivo, e como funcionam em sincronia? Se, por outro lado, não se trata de entidades distintas, a mente é corpórea? O que é a mente, e qual é a sua relação com o corpo? (BUNGE, 1980). A maior parte das soluções apresentadas para o problema, contudo, acaba por negar a existência ou por minimizar o estatuto de um ou de outro.

A própria formulação do problema em termos de “mente-corpo” já revela a existência, ao menos em nível linguístico, de uma solução preconcebida de herança cartesiana, isto é, de um dualismo psicofísico, ou a concepção de que se tratam de entidades separadas. Bunge (1980) ressalta, por exemplo, que não se concebe um problema do “movimento-corpo” na mecânica, ou da “digestão-trato digestivo” na fisiologia, mas sim do movimento dos corpos e da função digestiva do trato digestivo. Não se reificam propriedades, estados ou eventos, exceto quando estes são pertencentes ao sistema nervoso.

O dualismo de origem cartesiana de fato exerce influência, até os dias atuais, sobre muitos teóricos da consciência, que resistem em tratá-la como um fenômeno biológico semelhante a qualquer outro. Assim, situam a consciência e outros fenômenos mentais fora do mundo físico ordinário e, por conseguinte, fora do domínio da ciência natural, que acreditam dever tratar apenas de fenômenos objetivamente observáveis (SEARLE, 2000).

Dentre os contemporâneos adeptos do dualismo, pode-se citar Roger Penrose, e David Chalmers. O primeiro enquadra-se na categoria por não acreditar na existência de um universo unificado, mas, antes, que há um universo mental separado, que se encontraria “preso” ao universo físico. O autor concebe, ainda, um terceiro universo, para além do físico e do mental, que seria aquele dos objetos abstratos, tais como os números e outras entidades matemáticas (SEARLE, 1998).

David Chalmers (1999), por sua vez, partindo da teoria do duplo aspecto da informação, afirma que a natureza da consciência é essencialmente informacional. Ao considerar que toda e qualquer informação possui dos aspectos básicos, um físico e outro fenomenológico, o autor acredita ser possível explicar a emergência da experiência consciente de processos físicos, sem, contudo, reduzi-la a estes. As experiências surgiriam, assim, em virtude de estados informacionais, enquanto que o aspecto objetivo seria encontrado no próprio processamento físico.

De outro lado, as tentativas de combate ao dualismo acabam por incorrer em um diferente tipo de erro, qual seja, o de supor que a consciência possa ser reduzida a processos

físicos, ou até mesmo eliminada por completo. Trata-se daquilo que foi denominado pela filosofia como materialismo, e que conta, atualmente, com um grande número de adeptos, podendo-se destacar Daniel Dennett.

Diferentemente dos dualistas, Dennett (1991) não acredita que a mente seja algo distinto do corpo, ou que se sobreponha a ele como um centro de comando de ações. Pelo contrário, afirma ser a consciência o resultado de ações não conscientes realizadas pelas células e, conseqüentemente, passíveis de observação científica. Justamente por isso, propõe como forma de estudo da mente a “heterofenomenologia”, sob uma perspectiva de terceira pessoa, em oposição a uma ontologia subjetiva.

Essa abordagem acerca do problema da consciência parte da aceitação de que não existem conteúdos intrínsecos e exclusivos da mente. Ou seja, o autor nega que todos os fenômenos conscientes sejam experiências subjetivas qualitativas – aquilo que Searle (1998, p. 36) denomina de *qualia*. Para ele, existe um nível intermediário entre o fenomenológico e o neurofisiológico dos conteúdos representacionais, e a mente compreenderia este nível em que há processamento de informações. Tem-se, assim, que o agir inteligente exige, necessariamente, a manipulação de informação (DENNETT *apud* SILVEIRA, 2013).

Justamente por isso o autor acredita na possibilidade de uma IA forte. Se os seres humanos funcionam a partir de estímulos (*inputs*), e possuem disposições reativas que podem andar conjuntamente, tem-se que o cérebro é como um computador digital (*hardware*), e a mente consciente é como um programa de computador (*software*), resultando desta interação o agir inteligente, traduzido na capacidade de processar informações.

5. A consciência segundo John Searle

Searle (1998) aponta como motivo por trás da ampla aderência ao materialismo o temor de que, caso se aceite a existência e a irreducibilidade da consciência, necessariamente deverá se aceitar também algum tipo de ontologia dualista. Oferece, dessa forma, uma solução ao problema mente-corpo que evita tanto o dualismo quanto o materialismo, ao menos na maneira em que são tradicionalmente concebidos: a análise de que os processos cerebrais causam a consciência, mas que a consciência, propriamente dita, é uma “propriedade do cérebro”. Explica-se.

Todos os fenômenos mentais, conscientes ou inconscientes, são causados por processos que têm lugar no cérebro. Tome-se como exemplo as sensações de dor: elas são

causadas por uma série de eventos que se iniciam nas terminações nervosas livres e terminam no tálamo e em outras regiões do cérebro. Mas, no tocante às sensações efetivas, os acontecimentos interiores ao sistema nervoso central são suficientes para causar a dor - como aquelas provocadas por membros fantasmas ou mediante estimulação artificial de partes do cérebro. O mesmo é verdade para os demais fenômenos mentais, isto é, todos eles são causados por processos interiores ao cérebro (SEARLE, 2000) ¹⁰.

Soma-se a isso o fato de que todos esses “fenômenos mentais são justamente características do cérebro (e, talvez, do resto do sistema nervoso central)” (SEARLE, 2000, p. 25). A dificuldade em se compreender como cérebros podem causar mentes e, ao mesmo tempo, serem as mentes características do cérebro fundamenta-se, segundo o autor, em uma má compreensão acerca da causalidade. Tradicionalmente, quando se afirma que A causa B, supõe-se a existência de dois acontecimentos distintos, um identificado como a causa, o outro como efeito – crê-se, desse modo, em uma concepção dualista de causalidade, em que os eventos que ocorrem no mundo físico causam acontecimentos no mundo mental (SEARLE, 2000).

Propõe, portanto, um conceito mais sofisticado de causalidade, baseado na distinção entre as micro e macropropriedades de um sistema:

Cada objeto é composto de micropartículas [que] têm características, ao nível das moléculas e dos átomos, como também ao nível mais baixo das partículas subatômicas. Mas, cada objeto tem também certas propriedades como a solidez da mesa, o caráter líquido da água e a transparência do vidro, que são características superficiais ou globais dos sistemas físicos. Muitas destas propriedades de superfície ou globais podem explicar-se facilmente por meio do comportamento dos elementos a micronível. Por exemplo (...) o caráter líquido da água explica-se pela natureza das interações entre as moléculas de H₂O. Estas macrocaracterísticas são causalmente explicadas pelo comportamento dos elementos ao micronível (...) No caso do caráter líquido, da solidez e da transparência, não temos nenhuma dificuldade em supor que as características de superfície são *causadas pelo* comportamento dos elementos ao micronível e, ao mesmo tempo, aceitamos que os fenômenos de superfícies *são justamente* características dos sistemas em questão (SEARLE, 2000, p. 26-27).

Tem-se, portanto, que nesse mesmo sentido de “causado por” e “realizado em”, pode-se afirmar que os fenômenos mentais são causados por processos que ocorrem no cérebro, ao nível neuronal ou modular, e, simultaneamente, realizam-se no próprio sistema que consiste em neurônios (SEARLE, 2000, p. 28).

¹⁰Para Searle, tudo aquilo que importa para a vida mental humana é causado por processos interiores ao cérebro. No que se refere aos estados mentais causantes, crucial é o que ocorre no interior do cérebro, e não o estímulo externo ou periférico. Se os acontecimentos fora do sistema nervoso central aconteceram, mas nada ocorreu no cérebro, não haverá acontecimentos mentais: “é o princípio sobre cuja base funciona a anestesia cirúrgica: o estímulo exterior é impedido de ter os efeitos relevantes no sistema nervoso central” (SEARLE, 2000, p. 24).

Justamente por isso, não procede a afirmação de que a mente está para o cérebro assim como o programa está para o *hardware* do computador – e que resulta na consequência de que não existe nada de essencialmente biológico acerca da consciência humana.

A natureza dessa refutação, ao contrário do que pensam os entusiastas da IA forte, não possui relação alguma com o estágio atual da tecnologia dos computadores, mas sim com a sua própria definição: programas de computador operam a partir de uma estrutura puramente sintática, desprovidas, portanto, de qualquer conteúdo semântico ou *intencionalidade*. Eles mimetizam, apenas, o comportamento intencional através do processamento de símbolos (tradicionalmente sequências de 0 e 1), ou seja, através de parâmetros pré-estabelecidos de *inputs* e *outputs*

Os estados mentais humanos, por outro lado, e conforme ressaltado por Searle (2000, p. 21), “são acerca de, ou se referem a, ou são objetos e estados de coisas no mundo diferentes deles mesmos”. Significa dizer, portanto, que possuem conteúdo mental, além de quaisquer estruturas formais que possam ter: ainda que ocorram em séries de símbolos, conclui, deve haver algo além no pensamento do que meras séries abstratas, vez que estas, por si só, são desprovidas de qualquer significado. Para testar tal hipótese, propõe aquilo que ficou conhecido como o experimento do quarto chinês, formulado 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).

Com base na situação descrita, resulta que não é possível aprender chinês pela simples manipulação de símbolos formais. Se a efetivação do programa não é o bastante para que o ser humano do experimento obtenha uma compreensão do chinês, igualmente não basta para que outro computador digital a obtenha. Possuir estados mentais, como entender uma língua, implica mais do que a manipulação de símbolos formais. Implica, na verdade, uma significação associada a esses símbolos, ou seja, uma *semântica* - e a sintaxe, por si só, não é capaz de garanti-la (SEARLE, 2000).

Ao Argumento do Quarto Chinês, que demonstra que a semântica não é intrínseca à sintaxe, Searle (1998) acrescenta outro: a sintaxe não é intrínseca à física. Para compreender a natureza de tal argumento, deve-se distinguir, inicialmente, fenômenos que existem

independentemente de um observador daqueles cuja existência é relativa à observação de terceiros. Certas propriedades da natureza, como montanhas e moléculas, existem independentemente de um observador no sentido de que sua existência não depende de qualquer pensamento que se tenha sobre elas. Outras, por sua vez, como o dinheiro e o casamento, existem apenas quando relacionadas a algum agente que as compreenda como tal.

No que diz respeito à computação, tem-se que, a 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).

Disso resulta a possibilidade de se conferir uma interpretação computacional a qualquer fenômeno, caso se aceite a definição proposta por Turing (1950) de que computadores são equipamentos que manipulam símbolos formais, tradicionalmente definidos em 0's e 1's. A título de exemplo, poder-se-ia afirmar que uma janela consiste em um computador bastante simplificado, atribuindo a ela o número 1 e o número 0 para simbolizar, respectivamente, quando ela se encontra aberta e fechada (SEARLE, 1998).

Tem-se, portanto, que a analogia cérebro-máquina não se sustenta, uma vez que seres humanos não seguem regras do mesmo modo como os computadores a seguem, qual seja, em conformidade a meros procedimentos formais. Seres humanos possuem intencionalidade intrínseca, no sentido de que seus processos mentais existem na natureza independentemente de um observador. A sistemas de inteligência artificial somente é possível atribuir intencionalidade extrínseca, visto que atuam “como se” possuíssem crenças, desejos ou estados mentais próprios. O seu comportamento é intencional apenas na medida em que são interpretados como tal por um agente (SEARLE, 1998).

Rejeitar a distinção entre intencionalidade intrínseca e extrínseca resulta no fato de que tudo no universo age de maneira intencional, por seguir as leis da natureza e,

consequentemente, atuar dentro de um determinado grau de regularidade, “como se” estivesse tentando executar algum projeto ou realizar algum desejo¹¹.

Sistemas baseados em inteligência artificial não possuem, portanto, os estados subjetivos imprescindíveis para a intencionalidade intrínseca que possibilitam a dimensão da agência. Sendo neles ausentes, ainda, todos os demais aspectos apresentados no plano fático por uma pessoa natural, não é possível, dessa forma, ser-lhes reconhecida a condição de pessoa substancial.

6. As razões da pessoa jurídica e a expropriação da subjetividade

Considerando-se, porém, que grande parte dos ordenamentos jurídicos contemporâneos atribui a entes abstratos e desprovidos de intencionalidade intrínseca a personalidade em sentido formal, necessário investigar as razões que informam a personificação da pessoa jurídica para avaliar se o mesmo poderia ser feito em relação a IAs.

Embora tanto homens quanto entes coletivos sejam nomeados pessoas pelo ordenamento, deve-se ter em mente que diferentes motivos operam para que ambos, enquanto sujeitos de direitos e deveres, sejam igualmente nomeados pessoas. Se o querer legislativo atua sobremaneira no que tange à personificação dos entes coletivos (autorizando, indiretamente, até mesmo a desconsideração da sua personalidade), o mesmo não ocorre em relação à pessoa natural, em razão do substrato ontológico e dignidade imanente do homem (EBERLE, 2006).

A assimetria existente entre o tratamento conferido às pessoas naturais e aquele dispensado às pessoas jurídicas justifica-se, conforme indicado por Matta Machado (1954) 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. 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 entitiva, *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).

¹¹A fim de ilustrar seu argumento, Searle (1998) 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 $S=1/2gt^2$. O exemplo fornecido demonstra o absurdo de se ignorar a diferença entre intencionalidade intrínseca e “como se”, que torna mental tudo no universo.

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).

Justamente por encerrar uma polissemia, entretanto, o conceito de personalidade acaba, muitas das vezes, utilizado a um só tempo como valor e como aptidão para ser sujeito de direito, resultando, assim, em duas frequentes conclusões equivocadas: a atribuição do valor jurídico representado pela personalidade indistintamente a pessoas naturais e jurídicas, e a atribuição de personalidade a todos os entes a quem o ordenamento confere a qualidade de ser sujeito de direito (BARBOZA; BODIN DE MORAES; TEPEDINO, 2007).

As razões pelas quais se atribui personalidade também a entes coletivos, como empresas e associações, 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” (NEGRI, 2016, p. 8).

Conforme ressalta Negri (2016), nada impede que todos os integrantes de uma sociedade 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.

O autor destaca, ainda, para além da já mencionada, outras razões que conformam a personalidade jurídica, a saber: a articulação patrimonial, vez que o reconhecimento de um novo sujeito implica, por consequência, na afirmação de uma estrutura patrimonial autônoma; a constituição de um sistema de imputação direta pelos atos praticados por órgãos da pessoa jurídica; e a estabilização do processo de coordenação de ações funcionalmente integradas, ou seja, a organização (NEGRI, 2016, p. 9-10).

Perder de vista as razões específicas que informam as pessoas jurídicas, equiparando-as indistintamente às pessoas naturais, pode conduzir ao processo denominado por Stefano Rodotà (2007) de “expropriação da subjetividade”. Tal processo consistiria na usurpação, no plano concreto, de direitos inerentes ao ser humano, sob o pretexto de proteção do sujeito abstrato, manifestando-se, por exemplo, nos discursos acerca da titularidade de direitos fundamentais ou da personalidade por pessoas jurídicas de direito privado (NEGRI, 2016).

Exemplar da ocorrência de tal fenômeno é a decisão proferida pela Suprema Corte Americana no caso *Burwell v. Hobby Lobby* (ESTADOS UNIDOS, 2014), na qual se reconheceu, pela primeira vez na história do referido tribunal, a reivindicação de crença religiosa por parte de uma pessoa jurídica com fins lucrativos. Assim, sob a escusa de exercício dessa liberdade religiosa, permitiu-se que o ente coletivo em questão deixasse de ofertar a seus funcionários seguro de saúde gratuito, vez que qualquer cobertura mínima disponível no mercado abrange a oferta de “pílulas do dia seguinte” a seus segurados.

Cita-se, ainda, no ordenamento pátrio, a aplicação muitas vezes indiscriminada dos direitos da personalidade às pessoas jurídicas, com fundamento no art. 52 do Código Civil brasileiro e, mais especificamente, na Súmula 227 do Superior Tribunal de Justiça, que reconhece a possibilidade de que uma pessoa jurídica sofra danos morais.

É importante ressaltar que, na prática, direitos como o nome e a honra são invocados não por revelarem aspectos fundamentais da personalidade de entes coletivos, mas sim para garantir o livre exercício da função por eles exercida, havendo, portanto, uma diferença substancial no que tange aos interesses tutelados quando da afirmação dos direitos da personalidade da pessoa natural. Assim,

sem se atentar para a diferença entre as razões que determinam a proteção do ser humano e as que determinam a proteção da pessoa jurídica, as normas aplicáveis aos direitos da personalidade são simplesmente transpostas para a proteção da atividade empresarial e para a tutela de interesses de natureza exclusivamente econômica (NEGRI, 2016, p. 13).

Portanto, fundamental ter em mente, ao se cogitar a personificação de sistemas de inteligência artificial, que as razões que informam o reconhecimento da personalidade a seres humanos não se equiparam àquelas que levaram à sua atribuição às pessoas jurídicas. Resta investigar, portanto, se tais motivos específicos se fazem presentes também no caso de IAs.

Inicialmente, cumpre ressaltar que a noção de organização, através do processo de coordenação de ações funcionalmente integradas, embora seja uma característica inerente ao próprio fenômeno associativo, não constitui monopólio da pessoa jurídica (NEGRI, 2016). Sistemas baseados em inteligência artificial são, por si só, ferramentas organizacionais, desenvolvidas com o intuito de desempenhar tarefas de maneira coordenada através do processamento simbólico de *inputs* e *outputs* pré-estabelecidos. Além disso, não abrangem uma pluralidade de agentes cujas complexas relações demandem a sua simplificação em um único ente personificado, a exemplo do que ocorre com sociedades e associações.

No que diz respeito à autonomia patrimonial, sabe-se que a mesma se justifica em virtude da necessidade de exteriorizar de uma parcela dos custos derivados da atividade desenvolvida, diminuindo-se, assim, o chamado “risco empresa”. Em outras palavras, atribui-se personalidade ao ente abstrato para que, mediante a afirmação de uma estrutura patrimonial autônoma, haja limitação dos custos da atividade em favor dos sócios e em desfavor da coletividade, possibilitando, assim, o desenvolvimento da própria atividade econômica.

A existência de um patrimônio autônomo da pessoa jurídica relaciona-se diretamente, portanto, com a possibilidade de imputação direta ao ente por eventuais danos causados por seus agentes. Sendo a imputabilidade normativa passível de ser estabelecida independentemente de intencionalidade, inexistem, a princípio, impedimentos para que uma IA possa ser considerada como um centro autônomo de imputação de relações jurídicas.

Bastaria, para tanto, a existência de um patrimônio autônomo a ela associado que responderia pela reparação ou compensação da vítima na hipótese de ocorrência de um dano, a exemplo do que já ocorre para fins de securitização. Em tais casos, tem-se um patrimônio independente associado à seguradora, que se apresenta como centro de imputação, e que responde por eventuais danos que ocorrerem ao segurado.

Há de se ter em vista, entretanto, a quem favorece a exteriorização dos custos representados pela criação de uma nova subjetividade com autonomia patrimonial, e em desfavor de quem os riscos representados pela atividade desenvolvida são limitados. Tem-se, dessa forma, que o controle de riscos em favor dos sócios e em desfavor de terceiros não é admitido quando confrontarem os interesses de pessoas naturais em flagrante posição de vulnerabilidade na relação estabelecida com o ente personificado. É o que ocorre, por exemplo, quando do inadimplemento de verbas de natureza laboral: diante da insuficiência do patrimônio da pessoa jurídica, os seus sócios são chamados a responder pelo adimplemento com seu patrimônio particular.

Em se tratando de sistemas de inteligência artificial, não se pode olvidar que sua criação e disponibilização ao mercado são feitas, tradicionalmente, por pessoas jurídicas. Significa dizer que os custos associados à atividade exercida (pesquisa e desenvolvimento de tais tecnologias) já foram devidamente redistribuídos à comunidade por meio da imputação direta ao ente jurídico abstrato.

A personificação de agentes artificiais autônomos representaria, portanto, nova exteriorização de custos, agora em favor da pessoa jurídica e em prejuízo das pessoas naturais que, possuidoras de estados subjetivos internos, são aptas a possuir desejos e crenças, formular objetivos e também vê-los frustrados, constituem o polo mais vulnerável de qualquer

relação envolvendo uma IA. Dificultar-se-ia ainda mais, portanto, a reparação integral dos danos por elas suportados, privilegiando única e exclusivamente os interesses de seus desenvolvedores.

Pelas razões expostas, conclui-se que um ordenamento fundado no princípio da dignidade da pessoa humana e comprometido com a solidariedade social não deve atribuir a sistemas de inteligência artificial o *status* de pessoa, ainda que enquanto atributo, visto que a sua personificação inauguraria uma nova etapa no processo de expropriação da subjetividade identificado por Rodotà.

Sugere-se, dessa forma, que a responsabilidade por eventuais danos causados por uma IA recaia sobre aquele que disponibilizou a tecnologia ao público, de forma objetiva e agravada, com fundamento no disposto no artigo 927, parágrafo único, do Código Civil brasileiro¹².

Conforme destaca Fernando Noronha (1993), a responsabilidade agravada consiste em nova categoria da responsabilidade civil, a par da subjetiva, fundada na culpa, e da objetiva, que exige nexo de causalidade entre a atividade do responsável e o dano ocorrido. Aqui, para que haja a imputação, exige-se que o dano possua estreita conexão com a atividade exercida pelo responsável, devendo este ser considerado inerente, característico ou típico daquela.

Para além da estreita relação com a atividade normalmente exercida, tal modalidade exige, ainda, que o dano ocorrido afete a integridade física ou psíquica de uma pessoa, não bastando, para que haja o agravamento, que o evento danoso tenha atingido apenas objetos ou bens.

A responsabilidade agravada justifica-se, segundo Noronha (1993) pelo objetivo econômico da atividade normalmente desenvolvida – aquele que se beneficia de uma atividade lícita e que seja potencialmente perigosa adquire o ônus de arcar com as eventuais consequências danosas inerentes ao processo produtivo ou distributivo, inclusive aquelas causadas por empregados, prepostos e, acrescente-se aqui, agentes artificiais autônomos.

¹²Art. 927. Aquele que, por ato ilícito (arts. 186 e 187), causar dano a outrem, fica obrigado a repará-lo. Parágrafo único. Haverá obrigação de reparar o dano, independentemente de culpa, nos casos especificados em lei, ou quando a atividade normalmente desenvolvida pelo autor do dano implicar, por sua natureza, risco para os direitos de outrem (BRASIL, 2002).

6. Conclusão: IA, pessoa ou ferramenta?

Neste trabalho, buscou-se investigar se sistemas baseados em inteligência artificial, em virtude de seu crescente potencial para atuação autônoma, deveriam ser tratados como centros de imputação de direitos e deveres pelo ordenamento jurídico, ou seja, como pessoas. Tal questionamento revelou-se necessário na medida em que, com o desenvolvimento de sistemas cada vez mais sofisticados, surge a possibilidade de renúncia de controle humano direto em processos de tomada de decisões envolvendo uma IA, restando a dúvida de quem deve ser responsabilizado na eventual hipótese de ocorrência de um dano – a própria máquina, seu proprietário, ou aquele que a criou?

Para responder aos problemas levantados, fez-se necessário, em um primeiro momento, investigar as razões que justificam a atribuição de personalidade a seres humanos (personalidade enquanto valor) e a entes coletivos (personalidade enquanto atributo), para analisar se as mesmas se fazem presentes no caso de agentes artificiais autônomos.

No que tange à personalidade substancial, verificou-se que estão ausentes em sistemas de IA as bases deontológicas essenciais à sua inclusão na categoria, quais sejam, racionalidade autônoma, recorte psicofísico individual e alteridade. Por sua natureza incorpórea, tais sistemas somente poderiam ser considerados como tal na presença de intencionalidade intrínseca, requisito para a noção de agência.

Os argumentos apresentados por Searle, entretanto, demonstram que agentes artificiais ‘inteligentes’ operam a partir de uma estrutura puramente sintática, desprovida de qualquer conteúdo semântico. Eles apenas mimetizam o comportamento intencional através de parâmetros pré-estabelecidos de *inputs* e *outputs* - justamente por isso, possuem intencionalidade extrínseca, ou “como se”.

Seres humanos, por outro lado, possuem estados mentais qualitativos subjetivos, no sentido de que são acerca de ou referentes a alguma coisa, e que só existem quando experimentados por um ser humano ou outro tipo de “sujeito”. Possuem, portanto, conteúdo semântico e, conseqüentemente, a intencionalidade intrínseca necessária à agência.

Não sendo possível o reconhecimento do *status* substancial de pessoa a sistemas de IA, prosseguiu-se à investigação das razões que informam a subjetivação da pessoa jurídica, de forma a analisar se a tais sistemas poderia ser atribuída a personalidade formal.

Constatou-se que a personificação de entes coletivos, como empresas e associações, não tem como fundamento uma espécie de ontologia moral correspondente a estes, mas sim desempenha uma função heurística, na medida em que simplifica, em uma única

subjetividade, um complexo de relações jurídicas, facilitando a compreensão do próprio fenômeno. Para além da referida função, foram elencadas outras razões que informam a pessoa jurídica, como a articulação patrimonial, a mediação orgânica e a organização.

Ao analisá-las dentro do contexto ora investigado, entretanto, verificou-se que sistemas baseados em IA são, por si só, ferramentas organizacionais, desenvolvidas com o intuito de desempenhar tarefas de forma coordenada através de parâmetros sintáticos pré-estabelecidos. Desnecessária, portanto, a atribuição de personalidade formal e eles para fins de simplificação ou organização.

Por outro lado, a sua personificação para fins de imputação direta através do reconhecimento de um patrimônio autônomo, a exemplo do que ocorre em casos de securitização, embora possível, não se revelou adequada, tendo em visto que representaria um mecanismo de custos e controle de riscos em favor dos entes coletivos que tradicionalmente desenvolvem essas tecnologias, e em prejuízo da parte mais vulnerável da relação – a pessoa natural.

Concluiu-se, dessa forma, que a atribuição de personalidade a sistemas baseados em inteligência artificial inauguraria uma nova etapa do processo denominado por Rodotà de “expropriação da subjetividade”: usurpar-se-iam, no plano concreto, direitos inerentes ao ser humano, *pessoa* substancialmente considerada, sob o pretexto de proteção do sujeito abstrato, *mera ferramenta*. Justamente por isso, sugeriu-se que a responsabilidade por eventuais danos causados por uma IA recaia sobre aquele que disponibilizou a tecnologia ao público, de forma objetiva e agravada.

7. Referências bibliográficas

BODIN DE MORAES, Maria Celina. O Conceito de Dignidade Humana: Substrato axiológico e conteúdo normativo. In: SARLET, Ingo W. [Org.]. *Constituição, Direitos Fundamentais e Direito Privado*. 2 ed. Porto Alegre: Livraria do Advogado Editora, 2006, p. 118-119.

CHALMERS, David. *The conscious mind*. Oxford: Oxford University Press, 1996.

_____. Facing up to the problem of consciousness. Em: SHEAR, Jonathan. (Org.). *Explaining consciousness – the ‘hard problem’*. Cambridge: The MIT Press, 1999.

DENNET, Daniel. *Consciousness Explained*. Nova Iorque: Black Bay Books, 1991.

_____. The part of cognitive science that is Philosophy. *Topics in Cognitive Science*, v. 01, n. 02, 2009, p. 231-336. Disponível em:

<<http://onlinelibrary.wiley.com/doi/10.1111/j.1756-8765.2009.01015.x/full>>. Acesso em: 27 ago. 2017.

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

European Parliament Committee on Legal Affairs. *Civil law rules on robotics (2015/2103 (INL))*. Brussels, Belgium: European Parliament, 2016. Disponível em: <<http://www.europarl.europa.eu/sides/getDoc.do?pubRef=//EP//NONSGML%2BCOMPARL%2BPE-582.443%2B01%2BDOC%2BPDF%2BV0//EN>>. Acesso em: 18 ago. 2017.

Executive Office of the President National Science and Technology Council Committee on Technology. *Preparing for the future of artificial intelligence*. Washington, USA, 2016. Disponível em: <https://www.whitehouse.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf>. Acesso em: 18 ago. 2017.

FLORIDI, Luciano; CATH, Carine; MITTELSTADT, Brent; WACHTER, Sandra; TADDEO, Mariarosaria. Artificial Intelligence and the ‘Good Society’: the US, EU, and UK approach. *Science and Engineering Ethics*. Springer, 2017, p. 1-24.

FLORIDI, Luciano; TADDEO, Mariarosaria; TURILLI, Matteo. Turing’s Imitation Game: Still an Impossible Challenge for All Machines and Some Judges – An Evaluation of the 2008 Loebner Contest. *Minds and Machines*, v. 19, n. 1, 2008, p. 145-150. Disponível em: <<https://www.academia.edu/3491630>>. Acesso em: 2 ago. 2017.

FLORIDI, Luciano. L’Ultima Legge della Robotica. *La Repubblica*, fev. 2017. Disponível em: <<http://ricerca.repubblica.it/repubblica/archivio/repubblica/2017/02/12/lultima-legge-della-robotica37.html>>. Acesso em: 31. jul. 2017.

_____. A Fallacy That Will Hinder Advances in Artificial Intelligence. *Financial Times*, jun.2017. Disponível em: <https://www.academia.edu/33441632/A_fallacy_that_will_hinder_advances_in_artificial_intelligence>. Acesso em: 18 jul. 2017.

House of Commons Science and Technology Committee. *Robotics and artificial intelligence(No. Fifth Report of Session 2016-17)*. London, UK 2016. Disponível em: <<http://www.publications.parliament.uk/pa/cm201617/cmselect/cmsctech/145/145.pdf>>. Acesso em: 18 ago. 2017.

HUBER, Peter. Safety and the Second Best: The hazards of public risk management in the courts. *Columbia Law Review*, v. 85, n. 2, 1985, p. 277-337. Disponível em: <<https://www.jstor.org/stable/1122439>>. Acesso em: 29 ago. 2017.

LATSON, Jennifer. Did Deep Blue Beat Kasparov Because of a System Glitch? *Time*, 2015. Disponível em: <<http://time.com/3705316/deep-blue-kasparov/>>. Acesso em: 14 ago. 2017.

MARGARET ROUSE. Machine Learning. *Tech Target*, 2017. Disponível em: <<http://whatis.techtarget.com/definition/machine-learning>>. Acesso em: 4 set. 2017.

MARKOFF, John. Computer Wins on ‘Jeopardy!’: Trivial, it’s not. *The New York Times*. Disponível em: <<http://www.nytimes.com/2011/02/17/science/17jeopardy-watson.html?pagewanted=all&mcubz=3>>. Acesso em: 14 ago. 2017.

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. Disponível em: <<https://www.direito.ufmg.br/revista/index.php/revista/article/view/555/522>>. Acesso em: 4 set. 2017.

MICHALCZAK, Rafał. Animal’s Race Against the Machines. In: KURKI, Visa A. J.; PIETRZYKOWSKI, Tomasz (Eds.). *Legal Personhood: Animals, artificial intelligence and the unborn*. Cham (ZG): Springer, 2017, p. 91-101.

MOTA PINTO, Carlos Alberto da. *Teoria Geral do Direito Civil*. 3 ed., Coimbra: Coimbra Editora, 1996.

MORAES, João A; CANAL, Rodrigo. Chalmers e Searle nos Estudos da Consciência: Algum avanço? *Ciências e Cognição*, v. 14, n. 2, 2009, p. 262-275. Disponível em: <<http://www.cienciasecognicao.org/revista/index.php/cec/article/view/136/90>>. Acesso em: 2 set. 2017.

NEGRI, Sérgio Ávila. *A (Des)Naturalização Teórica da Pessoa Jurídica*. XXIV Encontro Nacional do Conpedi, Belo Horizonte, 2015.

_____. As Razões da Pessoa Jurídica e a Expropriação da Subjetividade. *Civilistica*, a. 5, n. 2, 2016, p. 1-18. Disponível em: <<http://civilistica.com/as-razoes-da-pessoa-juridica/>>. Acesso em: 6 set. 2017.

NORONHA, Fernando. Responsabilidade Civil: Uma tentativa de ressystematização. *Revista de Direito Civil, Imobiliário, Agrário e Empresarial*. São Paulo, ano 17, n. 64, abr./jun. 1993, p. 7-47.

RODOTÀ, Stefano. *La vita e le regole. Tra diritto e non diritto*. Milano: Fetrinelli Editore, 2007.

SCHERER, Matthew U. Who’s to Blame (Part 2): What is an “autonomous” weapon? *Law and AI*, 2017. Disponível em: <<http://www.lawandai.com/2016/02/10/what-is-an-autonomous-weapon/>>. Acesso em: 4 set. 2017.

_____. Regulating Artificial Intelligence Systems: Risks, challenges, competencies, and strategies. *Harvard Journal of Law and Technology*, v. 29, n. 2, 2016. Disponível em: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2609777>. Acesso em: 31 jul. 2017.

SEARLE, John. *O Mistério da Consciência*. Tradução de André Pinheiro e Vladmir Safatle. São Paulo: Paz e Terra, 1998.

_____. *Mente, Cérebro e Ciência*. Tradução de Artur Morão. Lisboa: Biblioteca de Filosofia Contemporânea, 2000.

_____. *A redescoberta da mente*. São Paulo: Martins Fontes, 1997.

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.

SILVEIRA, José Atílio. Searle e Dennett: Duas perspectivas de estudo da mente. *Problemata: Revista Internacional de Filosofia*, v. 04, n. 02, 2013, p. 238-258. Disponível em: <<http://dx.doi.org/10.7443/problemata.v4i2.16753>>. Acesso em: 30 jul. 2017.

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

TEPEDINO, Gustavo; BODIN DE MORAES, Maria Celina; BARBOZA, Heloisa Helena; *et aliter*. *Código Civil Interpretado Conforme a Constituição da República*. Vol. 1, 2 ed. São Paulo: Renovar, 2007.

TURING, Alan. Computing Machinery and Intelligence. *Mind*, v. 49, n. 236, 1950, p. 433-460. Disponível em: <<http://phil415.pbworks.com/f/TuringComputing.pdf>>. Acesso em: 16 ago. 2017.

ULANOFF, Lance. Apple is Turning Siri into a Next-level Artificial Intelligence. *Mashable*. Disponível em: <<http://mashable.com/2016/06/13/siri-sirikit-wwdc2016-analysis/#vPOglN1eWiqn>>. Acesso em: 14 ago. 2017.