

14º Concurso de Monografia 'Levy & Salomão Advogados'

**INTELIGÊNCIA ARTIFICIAL E PERSONALIDADE: AGÊNCIA,
IMPUTAÇÃO E RESPONSABILIDADE**

Anna Flávia Aguilhar Santos de Oliveira
Faculdade de Direito da Universidade Federal de Juiz de Fora (UFJF)
4º ano

INTELIGÊNCIA ARTIFICIAL E PERSONALIDADE: AGÊNCIA, IMPUTAÇÃO E RESPONSABILIDADE

RESUMO

O trabalho discute a Inteligência Artificial e a questão em torno da responsabilização por eventuais danos que os agentes dotados de IA venham a causar a terceiros. Toma-se como marco teórico o estudo de John Searle acerca da consciência, elemento fundamental ao trabalho para refutar a hipótese de que esses entes poderiam ser dotados de personalidade sendo, portanto, meras ferramentas. Ademais, sustenta-se que a responsabilização por atos decorrentes da conduta de IAs deve ser direcionada aos desenvolvedores ou àqueles que disponibilizam essas tecnologias no mercado, numa perspectiva de responsabilidade civil.

ABSTRACT

The paper discusses Artificial Intelligence and the question about accountability for possible damages that agents with AI will cause to third parties. The theoretical framework is John Searle's study of consciousness, a fundamental element of the work to refute the hypothesis that these beings could be endowed with personality and, therefore, mere tools. In addition, it is argued that liability for acts arising from the conduct of AIs should be directed to the developers or those who make these technologies available in the market, from an civil liability perspective.

1. INTRODUÇÃO

O presente trabalho trata das implicações do uso de tecnologias de inteligência artificial (IA), especialmente no que concerne a possíveis danos decorrentes de sua agência sem a interferência de humanos.

O problema investigado, portanto, pode ser enunciado da seguinte forma: agentes artificiais (IA's) devem ser tratados como centros autônomos de imputação de direitos e deveres, ou seja, como pessoas para o direito? Quem deve ser responsabilizado por danos decorrentes do comportamento de sistemas baseados em tecnologias de inteligência artificial?

A hipótese que se quer tratar aqui é a de que agentes de inteligência artificial (IA's) não devem ser qualificados como pessoa para o direito. Portanto, a pessoa responsável pela disponibilização da tecnologia ao público (através do mercado ou diretamente, como no caso de sistemas governamentais) deve ser responsabilizada por eventuais danos causados de forma objetiva.

Como marco teórico utiliza-se a distinção feita por John Searle entre intencionalidade intrínseca e extrínseca. Partindo dessa premissa objetiva-se demonstrar como a consciência não se encontra replicada em agentes artificiais e, a partir disso, não se pode conceber tais agentes como dotados de personalidade, como centros de imputação de direitos e deveres, de forma análoga a pessoas naturais. Uma IA nada mais é do que uma ferramenta.

A perspectiva, por alguns adotada, acerca da personificação abstrata de agentes artificiais (como pessoa jurídica) também será refutada.

Atentando-se aos propósitos da investigação, parte-se de uma revisão bibliográfica jurídico-compreensiva, a fim de, através do método dedutivo, expor os pontos concernentes à hipótese que se coloca e reafirmá-la através da refutação dos pontos contrapostos a ela.

2. O QUE SERIA UMA “INTELIGÊNCIA ARTIFICIAL”?

O conceito de Inteligência Artificial - IA surge no contexto da ciência cognitiva, tendo como marco histórico o Simpósio Hixon, em 1948, onde pesquisadores de diversas áreas do conhecimento, como matemática, psicologia, linguística e antropologia reuniram-se em torno do objetivo de começar a delinear o objeto de tal campo do conhecimento (GARDNER, 2003).

Dentro de tal campo (ciência cognitiva) encontra-se a ideia de IA (Inteligência Artificial), ligada à engenharia de sistemas, cujo objetivo é utilizar o conhecimento acerca do

funcionamento de processos cognitivos para gerar, artificialmente, cognição, ou seja, replicar processos humanos de aprendizagem, memorização, tomada de decisão e outros.

A proposta de Inteligência Artificial surge, portanto, da tentativa inicial de replicar a cognição humana através de processos artificiais.

Muito se discutiu acerca de qual seria o procedimento para a identificação de uma IA dotada de consciência, até o surgimento de um importante marco: o teste de Turing. O objetivo deste teste é verificar a capacidade de uma máquina apresentar comportamento inteligente, tal qual um ser humano. Tal, supostamente, verificar-se-ia se, em uma interlocução com uma IA, suas respostas fossem indistinguíveis daquelas ofertadas por uma pessoa humana em todo seu grau de liberdade e independentemente de parametrização (TURING, 1950).

Surge, a partir daí, a distinção entre dois tipos de IA.

Uma IA forte seria resultado de um processo artificial que replique a consciência e cognição humana. Uma IA fraca seria resultado de um processo artificial que simule alguns aspectos do comportamento humano.

Sistemas baseados em IA já operam, atualmente, em diferentes espaços e, muitas vezes, nem mesmo são notados por usuários. Como exemplo tem-se os cookies e informações geradas por sites da internet que “memorizam” os dados consultados a fim de serem retornados em outras páginas, fazendo com que informações sensíveis aos usuários cheguem até eles em outros momentos.

Note-se que o desenvolvimento e a disseminação do uso de IA's apresenta não apenas um prognóstico positivo, no que toca à eficiência na execução de tarefas e processamento de dados. Há que se mencionar a face negativa do surgimento de tais agentes supostamente dotados de inteligência (aqui tomada como sendo a capacidade de executar funções complexas, atribuída, geralmente, somente a seres humanos).

Além de preocupações acerca do impacto econômico e no mercado de trabalho em virtude da substituição de humanos por IA's na execução de uma série de tarefas, deve-se atentar, também, para os riscos que podem ser gerados ou potencializados por tais agentes, notadamente se eventualmente tratados como centros autônomos de imputação, qual seja, como pessoas. Danos causados por veículos autônomos podem apresentar magnitude maior do que os verificados atualmente, visto que tais sistemas de navegação são desenvolvidos a partir de codependência e coordenação entre IA's distintas. O tratamento de IA's como

centros de imputação, neste caso, serviria apenas à externalização de custos e riscos pelo desenvolvedor em desfavor de pedestres ou ocupantes do veículo, por exemplo.

3. INTELIGÊNCIA ARTIFICIAL, INTENCIONALIDADE E AGÊNCIA

A noção de agência refere-se a processos de determinação de causalidade normativa, e não meramente físico-causal. Pressupõe, portanto, intencionalidade. Por exemplo: se uma pessoa, ao ser empurrada, choca-se com um objeto de arte causando sua queda e destruição, não se diz ser essa pessoa o agente que causou o resultado, mas sim quem a empurrou.

Agência pressupõe intencionalidade na causa.

A intencionalidade quanto ao resultado, por sua vez, não se refere à questão de se estabelecer uma relação de causalidade normativa, mas sim à problemática da imputabilidade. Justamente por isso é possível identificar ações que sejam normativamente causa adequada a um resultado (o ente executa uma ação intencional que é causa do resultado) sem que, todavia, impute-se normativamente o resultado ao agente, pela ausência de intencionalidade direta quanto ao resultado (ausência de dolo), ou ausência de intencionalidade indireta (previsibilidade) quanto ao resultado (ausência de culpa). Assim é também possível a imputação de responsabilidade quando ausente intencionalidade quanto ao resultado, desde que presente intencionalidade quanto à causa (responsabilidade pelo risco, ou objetiva).

Poderia se objetar a tal afirmação a partir de situações como a responsabilidade civil imputada a pessoas jurídicas ou hipóteses de responsabilidade por ato de terceiro. Todavia, seria um equívoco, visto que, em tais hipóteses, o ente responde pela ação de outro. A imputação de responsabilidade, neste caso, é feita não ao agente, mas a terceiro e independe de agência.

John Searle (1998) ao analisar a problemática da consciência e a possibilidade de sua replicação por agentes artificiais deixa claro que a visão comum acerca de IA's, difundida pela cultura pop e alimentada por seus desenvolvedores, baseia-se em uma suposta analogia entre computadores e o cérebro humano. Na verdade, uma inversão epistemológica na qual busca-se explicar o cérebro e o fenômeno da consciência a partir do funcionamento de computadores.

Ou seja, busca-se explicar o fenômeno da consciência, de maior nível de complexidade, a partir de uma analogia com um processo de menor grau de complexidade, a computação. Isto se daria porque, de forma resumida, pensar é processar informação e o processamento de informação é justamente um processo de manipulação de símbolos através

de regras. Como os computadores fazem manipulação de símbolos através de regras, a melhor maneira de se compreender o pensamento ou a cognição seria estudar programas computacionais de manipulação de símbolos que existam em computadores ou cérebros (SEARLE, 1998).

Todavia, tal analogia não se sustenta. De fato, no sentido em que os seres humanos seguem regras, de modo algum os computadores as seguem, mas apenas atuam de acordo com certos procedimentos formais, desprovidos de intencionalidade intrínseca (SEARLE, s.d.). A sistemas formais de processamento de informação, como aos computadores, só se pode atribuir intencionalidade no sentido de uma intencionalidade “como-se”, ou seja, extrínseca (SEARLE, 2006). Trata-se, na verdade, de adotar uma postura intencional diante do sistema. Entretanto, a distinção entre intencionalidade intrínseca e intencionalidade extrínseca, ou “como-se”, não pode ser desprezada. Como afirma John Searle:

Se você rejeita a distinção entre intencionalidade intrínseca e extrínseca, resulta que tudo no universo tem intencionalidade. Tudo no universo segue leis da natureza e, por esta razão, tudo se comporta dentro de um determinado grau de regularidade, e por esta razão tudo se comporta como-se estivesse seguindo uma regra, tentando executar determinado projeto, atuando de conformidade com determinados desejos etc. Por exemplo, suponha que eu deixe cair uma pedra. A pedra tenta chegar ao centro da Terra, porque quer chegar ao centro da Terra, e fazendo assim segue a regra $S=1/2gt^2$. Em resumo, o preço de rejeitar a distinção entre intencionalidade intrínseca e como se é absurdo, porque torna mental tudo no universo (SEARLE, 2006, P. 121-122).

O que tal passagem demonstra é que muitas das ações humanas (que são desencadeadas por processos mentais) baseiam-se em representações mentais, e não apenas em processos sintáticos, ou seja, em procedimentos formais de obediência a regras. São dotadas, portanto, de conteúdo semântico. Isto leva a um raciocínio simples, de apenas três passos, que desconstrói a analogia entre cérebros e computadores:

1. Programas são totalmente sintáticos.
2. As mentes têm uma capacidade semântica.
3. A sintaxe não é a mesma coisa que a semântica, nem é, por si só, suficiente para garantir um conteúdo semântico. Consequentemente, programas não são mentes (SEARLE, 1998, p. 38).

Esta postura de identificação entre mentes e programas decorre ainda de um equívoco metodológico acerca do fenômeno da consciência. Isto porque o que se pode definir como consciência é um conjunto de estados subjetivos, aos quais não se pode ter acesso a partir da perspectiva do observador ou de terceira pessoa. Pode-se pressupor, a partir do comportamento, que outros entes que não o observador possuam consciência, mas não se pode experimentar a consciência de outrem. Do fato de não se poder experimentar estados

subjetivos de outros, todavia, não se pode inferir que todo e qualquer sistema que se comporte como se fosse portador de consciência seja também portador de crenças, desejos e outros estados mentais.

Esta tese baseia-se, portanto, no equívoco de desconsiderar que nem toda realidade é objetiva, mas parte dela é subjetiva, não sendo acessível por outros os peculiares estados de consciência de alguém (SEARLE, 2006). Na verdade, é uma tentativa de se superar o dualismo corpo/mente a partir de uma perspectiva reducionista que incorre em um erro filosófico quanto ao que seja propriamente essa redução.

Quando se fala em redução em termos filosóficos, pode-se estar diante de uma das seguintes hipóteses: a) se demonstra que A é na verdade B, ou seja, $A=B$ (redução identitária); b) se demonstra que A é causado por B, de forma que os fenômenos de fundo são responsáveis pelos fenômenos de superfície (redução causal). Na analogia computacional da mente, trata-se de buscar uma redução em termos de identidade entre mente e processos formais de processamento de informação, algo como o que ocorre, por exemplo, quando se analisa o pôr-do-sol. Ora, sabe-se que o sol não se põe, mas sim que existe uma ilusão de que ele se põe por parte de observadores estáticos em virtude do movimento de rotação da terra. Sendo assim, o pôr-do-sol é, na verdade, o movimento de rotação da terra. Daí decorre o materialismo de que a mente é o cérebro, constituindo-se como sinapses e outros fenômenos físico-químicos.

Tal postura (redução identitária entre consciência e processos computacionais) encontra-se em voga visto que se aceitar uma redução causal pode levar, aparentemente, a um dualismo de substâncias, sendo que os processos físico-químicos do cérebro seriam a causa de outra entidade: a mente. Todavia, isto somente ocorre caso atenha-se a um modelo de causalidade que opera com eventos distintos, como, por exemplo, “o disparo da arma causou a morte da vítima”. O risco de tal dualismo é justamente situar em um plano metafísico o conhecimento dos processos e fenômenos mentais e, portanto, não científico.

Isto deriva, contudo, de outro equívoco. Muitos fenômenos físicos causais funcionam de acordo com o modelo logo acima descrito de causalidade, mas nem todos. Cite-se, como exemplo, a solidez de uma mesa. Sabemos que ela é causada pelo arranjo e nível de energia das moléculas (reducionismo identitário), nem mesmo algo distinto da mesa. A solidez é uma propriedade, que emerge a partir da estrutura físico-química da mesa.

Para Searle, portanto, consciência e intencionalidade revelam “um aspecto biológico e natural do ser, tal qual a digestão, o crescimento ou a fotossíntese” (SEARLE, 1998, p.25).

Algoritmos computacionais, por tratarem-se meramente de operadores sintáticos, portanto, não são capazes de gerar o surgimento de fenômenos semânticos. Dessa maneira, ao se tratar de sistemas baseados em IA, a eles se aplicaria, tão somente, a ideia de IA fraca, ou seja, uma simulação de comportamento consciente ou intencional.

Há quem entenda ser possível replicar a consciência humana em uma máquina, já que para essas pessoas, aquela não passa de um software adaptável a qualquer hardware em que for instalada. A concepção de Searle fundamenta-se no que ele chama de qualia, que seria a própria consciência, como uma série de estados qualitativos mentais. Portanto, experiências subjetivas.

Para Daniel Dennett (2017) inexitem os chamados qualia,. Nossas mentes funcionariam, portanto, a partir de inputs, de estímulos, e possuímos disposições reativas que podem andar em conjunto já que nosso cérebro seria uma espécie de hardware e a consciência, um software.

Dennett, portanto, pressupondo ser a mente como programa e o cérebro como uma máquina, acredita ser possível a criação de agentes artificiais, ou seja, na possibilidade de uma IA forte. Todavia, admite que os processos sintáticos baseados em códigos binários como os utilizados em sistemas digitais simplesmente são incapazes de replicar a maneira pela qual o cérebro humano processa informações e permite o surgimento de fenômenos como a intencionalidade (DENNET, 2017). Não oferece aparentemente, portanto, uma objeção à teoria de Searle, apenas uma hipótese ou um argumento modal¹.

Neste ponto, conclui-se, a partir da visão de Searle, que sistemas de IA, independentemente do grau de sofisticação, carecem de estados subjetivos e, portanto, de intencionalidade. Não podem, assim, ser caracterizados como agentes.

4. INTELIGÊNCIA ARTIFICIAL E PERSONALIDADE

¹ Além da polarização entre as visões de Searle e Dennett, apresentam-se outras concepções do que seria a consciência e de como uma IA seria capaz de tratá-la. Roger Penrose, por exemplo, entende os dois modelos de IA como falsos, visto que há aspectos da consciência e da mente que jamais serão passíveis de replicação. Penrose faz essa assertiva baseado na análise do Teorema da Incompletude de Godel – que diz que uma teoria é incompleta quando apresenta uma sentença indecidível, ou seja, que não prova nem sua veracidade nem sua falsidade. Logo, se há partes da mente não computáveis, não é possível tomá-las como simuláveis pela IA (RIBEIRO, 2003).

Personalidade, para o Direito, inicialmente refere-se à aptidão genérica para a titularidade de direitos e deveres como um dos polos de uma relação jurídica. Neste sentido, uma pessoa é um centro de imputação.

Todavia, para além da mera qualificação formal enquanto centro de imputação, pode-se afirmar que, atualmente, opera-se com dois distintos conceitos de personalidade:

- a) personalidade enquanto atributo, ou personalidade formal ou abstrata e;
- b) personalidade enquanto valor, ou personalidade substancial.

4.1 Personalidade formal e personalidade substancial

O conceito de pessoa foi redefinido, quando do início da busca por cientificidade no âmbito jurídico-normativo, como um conceito qualitativo e puramente formal, tradicionalmente definindo-se personalidade como a aptidão genérica para figurar no polo ativo ou passivo de relações jurídicas, ou seja, para ser titular de direitos e deveres. Tal concepção puramente formal pode bem ser exemplificada pela passagem abaixo:

Num sentido puramente técnico ser pessoa é precisamente ter aptidão para ser sujeito de direitos e obrigações; é ser um centro de imputação de poderes e deveres jurídicos, ser um centro de uma esfera jurídica. Neste sentido técnico-jurídico não há coincidência entre a noção de pessoa ou sujeito de direito e a noção de ser humano. Os seres humanos não são necessariamente, do ponto de vista lógico, pessoas em sentido jurídico. (...)

O conceito técnico-jurídico de pessoa não coincide, portanto, necessariamente com o de homem ou de ser humano. Se o direito tem, todavia, em vista a disciplina de interesses humanos, se todo o direito é constituído por causa e para serviço dos homens, é logicamente forçoso que, pelo menos alguns homens sejam dotados de personalidade jurídica (MOTA PINTO, 1996, p. 84-85)

Tal visão espelha nitidamente um paradigma positivista normativista da ciência do direito no qual a juridicidade de uma norma depende exclusivamente de sua inserção em um sistema de validação lógico-formal, sendo seu conteúdo definido a partir de fatores reais de poder determinantes do contexto político.

Historicamente, a rejeição a um conceito meramente formal de pessoa (que admite a inclusão ou exclusão de seres humanos do universo personificado para o direito e, conseqüentemente, do universo de portadores de direitos e deveres, igualando-os, ainda, a entes meramente instrumentais, como as pessoas jurídicas) levou à busca por uma conexão intrínseca entre personalidade jurídica e elementos fáticos relacionados ao substrato ontológico do ente personificado. Ou seja, uma concepção substancial de pessoa, determinante de uma capacidade de direito voltada a preservar, por meio normativo, as

características manifestas pelo ente no plano fático (NINO, 1989;TEPEDINO, 2001; BODIN DE MORAES, 2003).

Esta concepção substancial de pessoa ligada, principalmente, a características normalmente presentes no plano fático (embora não exclusivamente) em indivíduos da espécie humana - como consciência, intencionalidade e, portanto, aptidão para a agência, - é determinante da titularidade de um conjunto de direitos, historicamente identificados como direitos humanos. Neste contexto surgiu a suposta relação de identidade entre os conceitos de ser humano e pessoa.

4.2 IA e personalidade enquanto valor ou personalidade substancial

Ausentes em uma IA todos os demais aspectos apresentados no plano fático por uma pessoa natural, como uma pessoa humana, o único traço remanescente que poderia autorizar o reconhecimento de personalidade a tais entes, a condição de pessoa substancial, seria a presença de estados subjetivos, ou seja, a presença de consciência e, portanto, intencionalidade intrínseca que permitiria a dimensão da agência.

Sistemas de IA, todavia, são desprovidos de intencionalidade, mimetizando, apenas, o comportamento intencional a partir de parâmetros pré-estabelecidos de output diante de elementos de input pré-selecionados. Operam, portanto, a partir de uma estrutura puramente sintática, da qual é exemplificativa o chamado “experimento do quarto chinês”:

Imagine que você execute as etapas de um programa elaborado para responder perguntas em um idioma que você não compreende. Eu não entendo chinês, então imagino que estou trancado numa 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).

Por se caracterizarem como sistemas de IA fraca, portanto, indiscutível que não se lhes reconheça personalidade enquanto valor ou personalidade substancial.

4.3 IA e personalidade atributo ou formal

Considerando que ordenamentos jurídicos contemporâneos já dotam entes abstratos e desprovidos de intencionalidade intrínseca de personalidade, necessário investigar acerca da possibilidade de se reconhecer a sistemas de IA personalidade jurídica na modalidade de personalidade atributo ou formal.

Concebida inicialmente por Savigny como uma ficção, advinda da extensão que se fez da noção de sujeito, foi o direito levado a personificar determinados entes em razão de necessidades práticas (NEGRI, 2016).

Kelsen sustenta, nesse sentido, que a pessoa jurídica deveria funcionar como expressão figurativa, instrumento linguístico que serviria para dar unidade a um conjunto de normas que regula uma pluralidade de indivíduos (2008). O que ocorre portanto, é uma simplificação através do termo “pessoa jurídica” de situações complexas que advêm de condutas de indivíduos (pessoas naturais).

Não de maneira simplista deu-se a definição da natureza da pessoa jurídica, sendo que para tanto foram propostas inúmeras teorias até se obter entendimento que fosse mais conveniente e majoritariamente aceito. Muitas foram as teorias engendradas com intuito de interpretar a natureza da pessoa jurídica. Partindo das chamadas “teorias da ficção”, que as entendiam como sendo meras criações da lei e da doutrina, passando pela ideia de instituição com finalidade socialmente útil e das chamadas “teorias orgânicas” ou da realidade objetiva chegou-se à definição atualmente aceita de sua natureza enquanto realidade técnica.

Inobstante as teorias formuladas e atual concepção da natureza jurídica dos entes abstratos dotados de personalidade jurídica, revela-se imprescindível a análise da função a ser cumprida dentro do ordenamento por sua personificação. Isto tendo-se em mente, sempre, o conhecido brocardo romano: *Ius hominum causa constitutum*.

Neste aspecto, nota-se mesmo que a ausência de uma análise funcionalista do instituto pessoa jurídica é uma das razões da frequente confusão relativa aos atributos da pessoa jurídica e os da pessoa natural. É o que ocorre, por exemplo, ao se falar em danos morais sofridos pela pessoa jurídica ou de seus direitos de personalidade.

Isto o que denomina Rodotà de “expropriação da subjetividade”, processo no qual sob pretexto de se proteger o ente abstrato, direitos inerentes ao ser humano são usurpados (RODOTÁ, 2007).

Exemplar de tal fenômeno é a decisão da Suprema Corte Americana no caso *Burwell v. Hobby Lobby*, na qual se autorizou a uma pessoa jurídica de fins lucrativos a, sob a escusa do exercício da liberdade religiosa, deixar de cumprir determinação legal de oferta gratuita de seguro-saúde a seus funcionários. Isto porque a cobertura minimamente compreensiva, em qualquer modalidade ofertada no mercado, incluía a disponibilização de “pílulas do dia seguinte” a segurados.

Como bem observa Negri, deve-se entender o processo de hipostasiação da pessoa jurídica para de fato se aferir qual o papel do termo “pessoa jurídica” no discurso do Direito (NEGRI, 2016).

Note-se que o termo pessoa jurídica desempenha, a rigor, um papel heurístico, de simplificação (NEGRI, 2016), a partir de determinadas razões que podem ser descritas de forma sucinta:

Constata-se, em resumo, no caso da pessoa jurídica, a presença das seguintes razões: i) simplificação de situações jurídicas complexas com a constituição de um centro unitário de imputação; ii) a articulação patrimonial, na medida em que o reconhecimento do novo sujeito implica também a afirmação de uma estrutura patrimonial autônoma; iii) constituição de um sistema de imputação direta dos atos praticados pelos órgãos da pessoa jurídica; iv) estabilização do processo de coordenação de ações funcionalmente integradas, representado pela noção de organização (NEGRI, 2016, p.9).

Desconsiderar as razões de personificação formal acima elencadas é o que dá origem ao problema da “expropriação da subjetividade” ou “naturalização” da pessoa jurídica, na terminologia cunhada por Negri e que consiste, basicamente, na ilusão de que o processo de imputação de direitos e deveres da pessoa jurídica representa um processo completo, à semelhança daquele previsto para a pessoa natural (NEGRI, 2016, p. 10).

Ao se cogitar personificar sistemas de IA de maneira análoga à que se personifica entes abstratos, como sociedades ou fundações, portanto, deve-se verificar se as mesmas razões se apresentam ou, ao menos, que razões se apresentam para tanto.

Assim, ao contrário do que ocorre com entes abstratos personificados atualmente, qualquer necessidade de simplificar situações complexas envolvendo diversos agentes ou de se estabilizar um processo de coordenação de ações funcionalmente integradas representado pela noção de organização se encontra ausente ao se tratar de IA's. Desprovidas de intencionalidade, não se lhes aplica a noção de agência. Como sistemas informacionais de caráter sintático, desempenham por si as tarefas de coordenação típicas da noção de organização a partir de parâmetros pré-estabelecidos. A rigor, sistemas computacionais foram desenvolvidos como ferramentas de organização.

A afirmação de uma estrutura patrimonial autônoma, por sua vez, apresenta relação de codependência com a de imputação direta. Busca-se a imputação direta através da personificação do ente abstrato pessoa jurídica justamente para que, mediante a separação patrimonial, seja possível a externalização de uma parcela de custos que derivem de riscos associados a determinadas formas de agência. Inevitáveis os riscos de dano associados à agência, a imputação direta de responsabilidade, combinada à autonomia patrimonial, permite

limitar os custos efetivos gerados pela eventual concretização do dano, diminuindo, assim, o chamado risco empresa.

Como a imputabilidade normativa pode ser estabelecida independentemente de intencionalidade do agente, ou mesmo de forma independente do desempenho da ação que deu causa ao resultado, como ocorre em casos de responsabilidade objetiva agravada (que dispensa até mesmo o nexo de causalidade adequada), a ausência de intencionalidade em sistemas de IA não se revela, teoricamente, como um obstáculo a que sejam tratados como centros autônomos de imputação.

Exemplifica-se. No caso de veículos autônomos, guiados por uma IA a partir de outputs pré-determinados diante da ocorrência de um input pré-selecionado em seu algoritmo, pode-se dizer que a agência, a rigor, deve ser atribuída àquele que estabelece a correlação entre input e output. Eventual dano que se concretize a partir da ação, todavia, poderia ser imputado ao sistema de IA, desde que a imputação direta esteja associada à afirmação de patrimônio autônomo. Assim, em caso de um acidente, o patrimônio associado à IA responderia pela indenização à vítima. Uma hipótese análoga ao que ocorre nos casos de securitização, onde um patrimônio autônomo associado a um centro de imputação (a seguradora) responde pelos danos causados pela agência do condutor. Um contrato de seguro nada mais é, a rigor, que um mecanismo de externalização de parcela dos custos resultantes da eventual concretização de um dano. O risco é controlado limitando-se o custo da ação ao valor do prêmio.

Não se admite, todavia, através da personificação formal de entes abstratos, a externalização ou limitação de custos associados a toda e qualquer forma de agência ou mecanismo de coordenação de agência. Pecunia non olet não é um princípio aplicável nesta hipótese. Observa-se, além da natureza do dano gerador do custo, a quem favorece a externalização e em desfavor de quem o risco é controlado.

A autonomia patrimonial e imputação direta de responsabilidade a sociedades empresárias, por exemplo, permite, a princípio, externalizar custos em favor dos sócios controlando riscos decorrentes do exercício da atividade ou de sua forma de organização em desfavor de terceiro. Se o dano, todavia, foi causado a um empregado (inadimplemento de verbas laborais ou acidente de trabalho), diante de insuficiência do patrimônio autônomo atribuído à pessoa jurídica, os sócios são chamados a responder. Não se permite a externalização de custos, nesse caso, diante de pessoas naturais em posição de vulnerabilidade.

O mesmo ocorre em relação a consumidores quando se verifica a adoção da chamada “teoria menor da desconsideração” pelo código de defesa do consumidor em seu art. 28.

Considerando que sistemas de IA são desenvolvidos e colocados no mercado por pessoas jurídicas, que, por sua vez, são desprovidas também de intencionalidade, isto significa que os custos associados ao desempenho e forma de organização da atividade (pesquisa, desenvolvimento e comercialização de sistemas de IA) já foram externalizados em favor dos sócios e desfavor de terceiros através da imputação direta ao ente abstrato.

A personificação de IA's, portanto, representaria nova externalização de custos e controle de riscos, desta vez em favor da pessoa jurídica e em desfavor de terceiros, notadamente pessoas naturais que, por sua condição de pessoas substancialmente consideradas (possuidoras de consciência, de interesses, aptas à dor e ao sofrimento, suscetíveis de forma particular à passagem do tempo por sua mortalidade) apresentam-se naturalmente como vulneráveis em qualquer situação que envolva uma IA na cadeia de causalidade que dê origem a um resultado a elas danoso.

Um ordenamento que se funda no princípio da dignidade da pessoa e comprometido normativamente com um princípio de solidariedade não deve, portanto, atribuir a sistemas de IA personalidade formal, visto encontrar-se ausentes todo e qualquer ganho social que a personificação de entes abstratos como sociedades, associações ou fundações permite.

Ao contrário, a personificação formal de sistemas de IA se apresenta, tão somente, como um mecanismo formal que dificultaria ou impediria a integral reparação de danos sofridos por pessoas naturais. Um novo degrau no processo de expropriação de sua subjetividade identificado inicialmente por Rodotá.

A personificação formal de IA's, portanto, apresenta-se como interesse único de desenvolvedores de tais sistemas como meio de limitação ou obstáculo adicional a ser superado para que suportem efetivamente e de forma integral os riscos gerados pela ferramenta (produto) desenvolvido.

5. CONSIDERAÇÕES FINAIS

Diante de todo o exposto, apresenta-se aqui, as conclusões obtidas ao longo da investigação, correlacionando-as, de modo a expor de forma unificada os resultados da pesquisa.

Sistemas baseados em IA realizam unicamente operações sintáticas, incapazes de gerar representações semânticas das informações processadas, conseqüentemente, não são dotados de consciência e, portanto, intencionalidade intrínseca.

A personalidade como valor, ou personalidade substancial, pressupõe que, no plano fático, o ente a que se lhe reconheça presente, não todas, mas alguma ou algumas das características manifestas em entes humanos.

Sistemas de IA, por sua natureza incorpórea, somente poderiam se cogitar pessoas substanciais na presença de intencionalidade intrínseca, requisito para a noção de agência. Sistemas de IA, portanto, não podem ser considerados pessoas naturais. Apresentam-se como ferramentas.

A personificação formal de entes desprovidos de intencionalidade ou qualquer estado subjetivo (dor, sofrimento, interesses intrínsecos) é possível através de expedientes de personificação formal, porém não imperativa ou necessária.

A personificação formal de um ente abstrato ou inanimado desempenha um papel heurístico a partir da constatação de sua adequação e eficiência para a obtenção de ganhos sociais em favor de pessoas naturais. É um meio, não um fim.

A personificação formal na busca de simplificação de relações complexas entre uma pluralidade de agentes ou como mecanismo para maior eficiência na organização de uma atividade, como ocorre em relação a determinados entes abstratos, é inaplicável a IA's, em virtude de sua natureza de processo sintático de organização simbólica. Uma IA é uma ferramenta de coordenação.

A personificação formal de entes abstratos como mecanismo de imputação direta e afirmação de autonomia patrimonial pressupõe-se um meio para a promoção de ganhos sociais em favor de pessoas naturais. Apesar disso, o fenômeno identificado como expropriação da subjetividade ou naturalização da pessoa jurídica já revela os riscos e limitações de tal meio para a obtenção dos resultados pretendidos.

A personificação formal de IA's para fins de imputabilidade direta e afirmação de autonomia patrimonial representativa, de fato, um mecanismo de externalização de custos e controle de riscos em desfavor de pessoas naturais, e não meio adequado ou eficiente para a obtenção de ganhos sociais em seu favor.

Desta maneira, conclui-se que um ordenamento fundado no princípio da dignidade da pessoa e comprometido com um princípio de solidariedade, não autoriza, a princípio, a personificação formal de sistemas baseados em IA. A responsabilidade por danos causados a

partir de processos causais desencadeados por IA's, portanto, deve ser atribuída ao desenvolvedor ou àquele que disponibiliza tal tecnologia ao público.

6. 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 107-151

DENNET, Daniel C. From Bacteria to Bach and Back: the evolution of minds. 1.ed. New York: W.W. Norton & Company, 2017.

GARDNER, Howard. A nova ciência da mente: uma história da revolução cognitiva. [Trad.] Cláudia Malbergier Caon. 3. ed. São Paulo: Editora da Universidade de São Paulo, 2003.

KELSEN, Hans. Teoria Pura do Direito. [Trad.] João Baptista Machado. São Paulo: Martins Fontes, 2006.

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

NEGRI, Sergio Marcos Carvalho de Ávila. As razões da pessoa jurídica e a expropriação da subjetividade. In: Civilistica. Rio de Janeiro, a.5.n.2, 2016. Disponível em: <<http://civilistica.com/wp-content/uploads/2016/12/Negri-civilistica.com.a.5.n.2.2016.pdf>>. Acesso em: 18/07/2017.

NINO, Carlos Santiago. Ética y derechos humanos: un ensayo de fundamentación. Barcelona: Ariel, 1989.

SEARLE, John R. A redescoberta da mente. [Trad.] Eduardo Pereira e Ferreira. São Paulo: Martins Fontes, 2006.

_____. O mistério da consciência. [Trad.] André Yuji Pinheiro Uema e Vladimir Safatle. São Paulo: Paz e Terra, 1998.

_____. [Trad.] Arthur Morão. Mente, cérebro e ciência. Lisboa: Edições 70, [s.d].

RIBEIRO, Henrique de Moraes. Uma Revisão da Teoria Quântica da Consciência de Penrose e Hameroff. In: Revista Eletrônica Informação e Cognição, v.3, n.1, p.108-125, 2001.

RODOTA, Stefano. Dal soggetto a la persona. Trasformazioni di una categoria giuridica. In: Filosofia Política, Fascicolo 3. Dezembro, 2007.

TEPEDINO, Gustavo. A tutela da personalidade no ordenamento civilconstitucional. In: Temas de Direito Civil. 2 ed. Rio de Janeiro: Renovar, 2001, p. 23-54.

TURING, Alan M. Computing Machinery and Intelligence. In: *Mind*, New Series, Vol. 59, No. 236 (Oct., 1950), Oxford: Oxford University, 1950, p. 433-460.