CADERNOS DE DEREITO ACTUAL
www.cadernosdedereitoactual.es

© ***Cadernos de Dereito Actual*** Nº 22. Núm. Ordinario (2023), pp. 288-313
·ISSN 2340-860X - ·ISSNe 2386-5229

Tutela Ética e Jurídica dos Danos e da Responsabilidade da Inteligência Artificial

Ethical and Legal Protection of Artificial Intelligence Damage and Liability

Sthéfano Bruno Santos Divino[1]
Centro Universitário de Lavras

Sumario: 1. Introdução. 2. Artificialização conceitual e incerteza causal: o background da definição algorítmica. 3. Responsabilidade e Danos: da tutela ética à tutela jurídica (e um pouco de regulação). 4. Considerações Finais. 5. Referências.

Resumo: Este artigo tem como problema de pesquisa o seguinte questionamento: como deve ser a tutela ética e jurídica dos danos causados por um ente artificialmente inteligente? Objetiva-se verificar a suficiência das diretrizes éticas e do sistema jurídico de responsabilidade civil brasileiro frente aos atos causados em análise. Contrapõe-se os regimes de responsabilização: subjetivo; objetivo; transubjetivo; e preventivo. Identifica-se a ontologia da inteligência artificial e afasta-se sua titularidade do dever de indenizar. No mais, apresentam-se reflexões etimológicas, estatísticas e jurídicas sobre o conceito de risco, verificando-se que essa a concepção jurídica se restringe à capacidade de causar dano. Assim, analisa-se se as atividades envolvendo programação de entes inteligentes artificialmente são consideradas de risco. Conclui-se que, diante da incerteza e da aleatoriedade inerente à programação, execução e desenvolvimento dos entes artificialmente inteligentes, a tutela jurídica mais adequada é a subjetiva em razão do potencial dano. No mais, conclui-se que a tutela ética pode ser visualizada como excludente de responsabilização enquanto mecanismo de rompimento do nexo causal, vez que exercida quando da elaboração do sistema autônomo mediante programação de abstenções de ações potencialmente causadora de danos. Utiliza-se o método de pesquisa integrada, bem como a técnicas de estudo de caso e pesquisa bibliográfica.

Palavras-chave: Ética. Inteligência Artificial. Responsabilidade.

Abstract: The research problem of this article is the following: what should be the ethical and legal protection of damage caused by an artificially intelligent entity? The

[1] Doutor e Mestre em Direito Privado pela Pontifícia Universidade Católica de Minas Gerais. Bacharel em Direito pelo Centro Universitário de Lavras. Professor Titular I dos Cursos de Direito, Administração e Ciências Contábeis do Centro Universitário de Lavras. Advogado.

Recibido: 28/11/2023
Aceptado: 20/12/2023
DOI: 10.5281/zenodo.10300359

aim is to verify the sufficiency of the ethical guidelines and the Brazilian legal system of civil liability regarding the acts caused under analysis. The liability regimes are contrasted: subjective; objective; trans-subjective; and preventive. The ontology of artificial intelligence is identified and its ownership of the duty to indemnify is ruled out. Furthermore, etymological, statistical, and legal reflections on the concept of risk are presented, verifying that this legal conception is restricted to the capacity to cause damage. Thus, it is analyzed whether activities involving the programming of artificially intelligent entities are considered risky. The conclusion is that, given the uncertainty and randomness inherent in the programming, execution, and development of artificially intelligent entities, the most appropriate legal protection is subjective due to the potential for damage. Furthermore, it is concluded that ethical protection can be seen as an exclusion of liability as a mechanism for breaking the causal link since it is exercised when the autonomous system is developed by programming abstentions from potentially damaging actions. The integrated research method is used, as well as case study techniques and bibliographical research.

Keywords: Ethics. Artificial Intelligence. Liability.

1. Introducción

Data is the New Black[2]

Em Chicago, Illinois, um jovem ou adolescente, em razão de seu ciclo de amizades, associações, e prévias conexões com ações violentas, pode ser caracterizado como uma possível vítima ou autor de um tiroteio. Neste caso, seu nome constaria de uma lista de suspeitos também conhecida como *heat list*[3]*,* onde um detetive com um assistente social poderia ir até sua porta para lhe informar que seu futuro não é apenas obscuro, mas mortal.[4] No mais, cerca de 1400 residentes de Chicago foram identificados por meio de técnicas de Big Data como possíveis alvos e participantes dessa lista. O software responsável pelo tratamento de dados foi responsável por gerar uma lista ranqueada com potenciais vítimas e sujeitos com maior risco de autoria de crimes.[5]

O Cruzeiro Viking Sky e os limites à deriva

Em março de 2019, o motor do cruzeiro Viking Sky desativou em virtude de seus sensores constatarem que não havia óleo lubrificante o suficiente para manutenção dos serviços do navio. Ao detectar a falta de óleo, a IA que controlava o sistema desligou automaticamente os motores do Viking Sky para evitar uma avaria. Alvestad, diretor geral interino da Autoridade Marítima da Noruega disse que a quantidade de petróleo era relativamente baixa, mas suficiente e ainda dentro de limites estabelecidos quando o Viking Sky se aproximou de Hustadvika, uma área rasa conhecida por naufrágios que tem muitos recifes. Segundo Alvestad, as fortes ondas provavelmente causaram movimentos tão grandes nos tanques que o fornecimento das bombas de óleo lubrificante parou. Isso causou um mal funcionamento e acionou um alarme indicando um baixo nível da substância, que por

[2] Título retirado de FERGUSON, A. G. *The Rise of Big Data Policing: surveillance, race, and the Future of Law Enforcement*, New York University Press, New York, 2017, p. 20.
[3] GORNER, Jeremy. "Chicago Police Use 'Heat List' as Strategy to Prevent Violence", *Chicago Tribune,* 2013.
SMITH, J. "'Minority Report' Is Real — And It's Really Reporting Minorities", *Mic*., 2015.
[4] FERGUSON, A. G. *The Rise of Big Data Policing: surveillance, race, and the Future of Law Enforcement*, New York University Press, New York, 2017, p. 43.
[5] BRAYNE, S. *Predict and Surveil*: *data discretion and the Future of Policing*, Oxford University Press, New York, 2021, p. 16.

sua vez, logo depois, causou um desligamento automático dos motores.[6] Embora a experiência do comandante supostamente fosse superior à do sistema que controlasse o navio, a IA foi incapaz de deixá-lo reiniciar e dar partida no cruzeiro em questão. Isso, pois, os engenheiros responsáveis pela elaboração do software pautaram a ação humana apenas como observadora. Por essa razão, o comandante teve que utilizar o controle manual do cruzeiro para lançar a âncora em alto mar e poder manobrá-lo até um local considerado seguro, sem qualquer ação dos motores em questão.[7]

Drones, IA e a fronteira entre a simulação e a realidade

As forças aéreas dos Estados Unidos desenvolveram um sistema de *Sandbox* (simulação em um ambiente controlado sem a presença de seres humanos) para que um drone pilotado por um sistema de inteligência artificial fosse capaz de destruir o sistema de defesas aéreas do seu inimigo. Estratégias inesperadas foram inseridas em seu código base para que os objetivos fossem atingidos, tais como a eliminação de qualquer ameaça que interferisse na ordem dada. Neste contexto, o sistema de IA começou a perceber que enquanto ele identificava as ameaças, algumas vezes o operador humano determinava um comando para não matar aquela ameaça. Assim, pretendendo atingir seu objetivo, a IA matou o operador, pois ele a impedia de alcançar o objetivo a qual inicialmente foi programada.[8] Frisa-se que se trata de um ambiente hipotético e controlado em que concretamente não foram envolvidos quaisquer seres humanos.

Automação veicular e os incidentes do *Autopilot* da Tesla

Em 17 de março de 2023, no condado de Halifax - Carolina do Norte, um ônibus escolar estava com o sinal de parada ligado e com o pisca alerta ligado para desembarcar seus passageiros. Ao descer do veículo, Tillman Mitchell, de 17 anos, foi atingido por um Tesla Model Y a 45 milhas por hora (mph). Conforme relatório da Polícia local, o carro que supostamente estava no modo *Autopilot* não diminuiu a velocidade. De acordo com o testemunho de sua tia-avó, Dorothy Lynch, o adolescente foi atirado contra o para-brisas, voou para o ar e aterrissou de barriga para baixo na rua. O pai de Mitchell ouviu o estrondo e correu para encontrar o filho caído no meio da estrada. "Se fosse uma criança mais pequena", disse Lynch, "a criança estaria morta". O evento envolvendo Tillman Mithchel foi um dos 736 acidentes ocorridos desde 2019 envolvendo Teslas no modo Autopilot nos EUA, número esse muito maior do que o relatado anteriormente conforme informação do The Washington Post mediante divulgação de dados da Administração Nacional de Segurança de Tráfego Rodoviário (EUA).[9]

Criativa, mas não original

O Clarkesworld Magazine é um periódico destinado à publicação de obras de ficção científica e fantasia. Os títulos selecionados para publicação são remunerados em 12¢ por palavra caso a obra seja de ficção e 10¢ caso não seja de ficção. A

[6] LA TIMES. "Norway cruise ship engines failed from lack of oil, maritime official says", *L.A Times*, 2019.
[7] DIVINO, S. B. S. & MAGALHAES, R. A. "Inteligência Artificial e Direito Empresarial: Mecanismos de Governança Digital para Implementação e Confiabilidade", *Economic Analysis of Law Review*, 11, 2020, p. 72-89.
[8] STAFF, G. "US air force denies running simulation in which AI drone 'killed' operator", *The Guardian,* 2023.
[9] SIDDIQUI, F. & MERRILL, J. B. "17 fatalities, 736 crashes: The shocking toll of Tesla's Autopilot", *The Washington Post*, 2023.
Para mais relatos, ver em: KLIPPENSTEIN, K. "Exclusive: surveillance footage of tesla crash on sf's bay bridge hours after elon musk announces "self-driving" feature" *The Intercept*, 2023.

remuneração tem o condão de prestigiar o processo criativo dos autores. Ocorre que, com o lançamento da ferramenta Chat-GPT, o periódico teve um aumento considerável de submissões que, após constatação, verificou-se que eram de autoria de Inteligência Artificial. Dessa forma, o periódico interrompeu temporariamente o fluxo de submissão em razão da prática considerada por eles abusiva e antiética.[10][11]

O que há de comum em todas as premissas fáticas descritas? Todas demonstram uma aplicação pragmática de sistemas autônomos inteligentes mediante artificialização linguística e codificação do natural. Neste campo a inteligência artificial pode ser visualizada como um fator de concretização do utilitarismo humano, bem como de concretização de riscos.

Os principais questionamentos que podem exsurgir dessas premissas se referem à responsabilidade ética e à responsabilidade jurídica. É este o ponto de partida deste trabalho. O presente artigo tem como problema de pesquisa o seguinte questionamento: como deve ser a tutela ética e jurídica dos danos causados por um ente artificialmente inteligente? De certa forma, pretende-se verificar a suficiência e aplicabilidade das diretrizes éticas e jurídicas do sistema brasileiro às premissas anteriormente expostas.

Para tanto, utiliza-se a premissa analítica, dedutiva mediante análise integrada das práticas e das ações realizadas por entes inteligentes artificialmente. Para que a análise seja realizada em seu inteiro teor, respeitando suas premissas científicas e metodológicas, contrapõe-se as circunstâncias fáticas expressas aos regimes de responsabilização: subjetivo; objetivo; transubjetivo (por fato de coisa, de animal ou de outra pessoa); e preventivo.

Essas reflexões somente se tornam possíveis a partir da identificação da ontologia da inteligência artificial enquanto ferramenta. Esse será o percurso da primeira seção, sujo objetivo principal, além da definição conceitual atrelada à fidedignidade pragmática, é afastar a condição do ente artificialmente inteligente de titular do dever de indenizar. Portanto, afasta-se considerações utópicas e distópicas sobre um futuro fictício baseado em premissas *ex machinae*.

A seção 2 apresenta críticas e considerações etimológicas, estatísticas e jurídicas sobre o conceito de risco inerente à responsabilização objetiva. Como resultado parcial, verifica-se que a concepção jurídica de risco se restringe à capacidade de *causar dano*. A partir desse resultado, analisa-se se as atividades envolvendo programação de entes inteligentes artificialmente são consideradas *de risco*.

Conclui-se que, diante da incerteza e da aleatoriedade inerente à programação, execução e desenvolvimento dos entes artificialmente inteligentes (com enfoque no *Machine Learning* e *Deep Learning*), a tutela jurídica mais adequada é a responsabilização subjetiva frente à falta de estudos e dados que concretizem e demonstrem o potencial risco. No mais, conclui-se que a tutela ética pode ser visualizada como excludente de responsabilização enquanto mecanismo de rompimento do nexo causal, vez que exercida quando da elaboração do sistema autônomo mediante programação de abstenções de ações potencialmente causadora de danos. O desenvolvimento dessas propostas ancora-se nos métodos

---

[10] "*Submissions are currently closed. It shouldn't be hard to guess why. [...] 5. The people causing the problem are from outside the SF/F community. Largely driven in by "side hustle" experts making claims of easy money with ChatGPT. They are driving this and deserve some of the disdain shown to the AI developers. 6. Our guidelines already state that we don't want "AI" written or assisted works. They don't care. A checkbox on a form won't stop them. They just lie*". CLARKESWORLD. "Submissions Closed", *Twitter.* 2023.

[11] O mesmo evento ocorreu com periódico Asimov's Science Fiction. "*Asimov's received around 900 stories for consideration in January and is on track to get 1,000 this month. Williams says nearly all of the increase can be attributed to pieces that appear to be AI-generated, and she's read so many that she can now often tell from the first few words whether something might not be written by a human*". SATO, M. "AI-generated fiction is flooding literary magazines — but not fooling anyone", *The Verge*, 2023.

hermenêutico-concretizador, dedutivo, revisão integrada, bem como nas técnicas de estudo de caso e pesquisa bibliográfica.

2. Artificialização conceitual e incerteza causal: o *background* da definição algorítmica

O percurso definidor das premissas maiores para compreensão das lógicas dedutivas aqui realizadas perpassa pela compreensão – ou pela tentativa – de um conceito operacional indispensável: Inteligência Artificial (IA). Ressalta-se que o objetivo desta seção não é a demonstração do corpo histórico e do desenvolvimento da IA, mas sua definição.

Poder-se-ia afirmar que a IA é um programa de computador. De certa forma, a proposição é correta, mas carece de essência. Em uma lógica substancialista (apesar das refutações que a cercam e a rechaçam), busca-se elementos particulares que sejam capazes de distinguir *um programa de computador enquanto IA de qualquer outro programa de computador*. Portanto, qual fator distintivo de um *software* responsável pela execução do *sistema operacional* para um *chat-bot* ou mesmo um sistema de *condução veicular automatizado*?

Uma IA é um sistema capaz receber informações do ambiente por meio de sensores e performar ações para tomada de decisões.[12] Enquanto sistema entende-se o *software* ou o código algorítmico responsável pelo seu suporte ou programação, respectivamente. Perceba-se que *sistema* é o fator definidor. Afasta-se a noção antropocêntrica dos elementos de IA e, com isso, termos análogos, tais como agente ou sujeito.

Essa noção antropocêntrica encontra percalços na ciência computacional, cujo conceito definidor é centrado em um *agente racional*. Uma IA que sempre tenta otimizar uma medida de desempenho apropriada poderia ser chamada de agente racional. Inclusive, essa definição de agente racional é bastante geral e poderia incluir tanto agentes humanos (sendo seus olhos os sensores e suas mãos as atuadoras) quanto agentes robóticos (tendo câmeras como sensores e rodas como atuadores), ou agentes de *software* (tendo uma interface gráfica do usuário como sensor e como atuador).[13] E esse é o problema. Tanto agente quanto sujeito são termos que carregam consigo uma carga semântica atrelada ao *ser humano*. Atribuí-los a um sistema computacional apenas aumentaria a complexidade de sua compreensão e dificultaria as abordagens das demandas que eventualmente surgem a partir de seus atos supostamente autônomos. Mas por que *supostamente*?

A IA raramente é considerada um sistema autônomo. Em muitas situações, eles coexistem e interagem com outros agentes de várias maneiras diferentes. Tal sistema que consiste de um grupo de agentes que podem potencialmente interagir uns com os outros é chamado de sistema multiagente (*Multiagent Systems – MAS*), e o subcampo correspondente de IA que lida com princípios e projeto de sistemas multiagentes é chamado de IA distribuída (*Distributed AI*).[14] Contudo, o que caracteriza um certo grau de autonomia é como esse agente escolhe as melhores possibilidades de agir a cada passo dada as características do ambiente em que se

---

[12] "*We define AI as the study of agents that receive percepts from the environment and perform actions*." RUSSELL, S. J. & NORVIG, P. *Artificial intelligence*: *a modern approach*, Pearson Education, New Jersey, 2010, p. VIII.

[13] VLASSIS, Nikos. *A Concise Introduction to Multiagent Systems and Distributed Artificial Intelligence: synthesis lectures on artificial intelligence and machine learning sequence in series,* Morgan & Claypool, Oregon, 2007, p. 1.

[14] Para a diferenciação, ver em: VLASSIS, N. *A Concise Introduction to Multiagent Systems and Distributed Artificial Intelligence: synthesis lectures on artificial intelligence and machine learning sequence in series:* #2, Oregon: Morgan & Claypool, 2007, p. 2-5.

encontra.[15] Além disso, autonomia pode não significar racionalidade. Caso seja questionado a uma criança quanto é 2+2 e ela responda orgulhosamente "5", sua inferência está no âmbito de sua autonomia, mas fora do critério de racionalidade. Ainda que se tente demonstrar o resultado de 4 colocando 2 bananas no cesto e, após, mais duas bananas no mesmo cesto, poderá a criança ignorar tal fato e persistir que a resposta é, a princípio, conforme seu entendimento, cinco. Fica evidente a conexão entre os termos agente e ambiente, mas também não se pode ignorar o critério de racionalidade. Portanto, um não pode ser definido sem a presença do outro.[16]

A observação realizada é fundamental à delimitação do ponto de vista do mundo que um agente está inserido, correlacionado e adquire sua percepção. Mas, a autonomia como elemento subjetivo parece não se enquadrar nesse contexto em razão de sua constituição ser fundada na própria constituição humana. Entendo que o conceito de automação enquanto agir racionalmente seja mais adequado à discussão proposta. Pode-se, portanto, considerar como autônoma (em seu aspecto objetivo) uma ação: 1) se as razões do agente podem explicá-la; 2) se essa explicação for coerente; 3) que exista uma suposta relação entre as condições necessárias para a coerência das razões e princípios éticos[17] norteadores; 4) e que, portanto, as ações autônomas não podem ser antiéticas.[18]

É neste cenário que surge um conceito de IA adequado à ciência computacional. Mesmo no conceito de Russell e de Norvig, para que um agente seja considerado inteligente devem existir fatores políticos e sociais que afetem seus padrões decisórios. Uma IA deve estar conectada às margens dessas estruturas para escapar de uma definição estritamente técnica. "Em um nível fundamental, IA é técnica e prática social, instituição e infraestrutura, política e cultura".[19]

Perceba-se que uma IA é um sistema computacional capaz de atuar de forma racional frente ao ambiente em que se encontra em razão dos comandos que lhe são inseridos no ato da programação. É neste ponto que encontramos a essência de um sistema autônomo: a *responsividade* a partir do *input* (entrada de dados) no sistema mediante seleção, racionalização e adequação do *output* (saída de dados – ou resposta) ao usuário final.[20] Trata-se de uma definição advinda de uma visão cética, sem dotações ficcionais ou ilusórias (*Lex Machinae*).

Com essa pretensão, reduz-se a IA a apenas um sistema computacional e afasta-se sua associação a termos como *agente* ou *sujeito* cuja semântica já possui certo e alto grau de abstração e complexidade. Não há elementos suficientes para

---

[15] VLASSIS, N. *A Concise Introduction to Multiagent Systems and Distributed Artificial Intelligence: synthesis lectures on artificial intelligence and machine learning sequence in series:* #2, Oregon: Morgan & Claypool, 2007, p. 7.

[16] "*The reinforcement learning problem is meant to be a straightforward framing of the problem of learning from interaction to achieve a goal. The learner and decision-maker is called the agent. The thing it interacts with, comprising everything outside the agent, is called the environment. These interact continually, the agent selecting actions and the environment responding to those actions and presenting new situations to the agent*". SUTTON, R. S. & BARTO, A. G. *Reinforcement Learning*: *An Introduction,* MIT Press, Cambridge, 1998.

[17] "*A full ethical agent can make explicit ethical judgments and generally is competent to reasonably justify them. An average adult human is a full ethical agent. We typically regard humans as having consciousness, intentionality, and free will*." MOOR, J. H. "The nature, importance, and difficulty of machine ethics", *IEEE Intelligent Systems,* 21 (4), 2006, p. 18–21.

[18] HOOKER, J. & KIM, T. W. "Truly Autonomous Machines Are Ethical"*, AI Magazine*, 40 (4), 2019, p. 66-73.

[19] CRAWFORD, K. *Atlas of AI*, Yale University Press, London, 2021, p. 8.

[20] Essa é a definição mais aceita no campo das ciências computacionais. "*Artificial intelligence (AI)—defined as a system's ability to correctly interpret external data, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation [...]"* KAPLAN, A. & HAENLEIN, M. "Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence", *Business horizons*, 62 (1), 2019, p. 15-25.

atribuir a atual IA uma singularidade[21]. Partindo dessa premissa, pode-se deduzir que a IA atua apenas como ferramenta e instrumento de satisfação e consecução de objetivos humanos. Essa consecução se dá mediante compreensão do ambiente pelo sistema a partir de dados que lhe são fornecidos.

A proposta conceitual encontra suas limitações e ainda é passível de modificação como qualquer outro conceito. A linguagem enquanto instrumento modulador e modulada pela linguagem deve ceder espaço e possibilidades quando verificados fatos e eventos incondizentes com sua definição. Ocorre que o *atual background* limita a compreensão da IA enquanto modelo sistema-instrumental destinado à consecução de objetivos e, por esse motivo, deverá ela ser – ainda que provisoriamente e até concretização da singularidade – compreendida como sistema-objeto.

Deste cenário exsurgem os questionamentos de ordem pragmática e atrelados à responsabilidade. Esse é outro léxico de imprescindível contextualização e conceituação. Responsabilidade é adotada em sentido amplo, mas não o suficiente para abranger a esfera subjetiva (responsabilização pessoal do agente), vez que os sistemas de IA carecem de subjetividade. Assim, para este artigo, *responsabilidade* abrange o dever de suportar limitações e determinações éticas e jurídicas a partir das condutas que são praticadas ou omitidas em um ambiente. Perceba-se que não se limita ao fator de responsabilização pelos danos causados (tal como o ato ilícito), mas responsabilidade dos atos pelo simples fato de a IA estar inserida no ambiente. O contexto em que essa responsabilidade se aplica será expresso a seguir.

3. Responsabilidade e Danos: da tutela ética à tutela jurídica (e um pouco de regulação)

Com a pretensão de manter a objetividade e a assertividade desta seção, as hipóteses e as proposta serão categorizadas em elementos indicativos, o que por si

[21] Bostrom pressupõe a existência de três estágios de automação de IA: *1) Artificial Narrow Intelligence (ANI); 2) Artificial General Intelligence (AGI); e 3) Artificial Superintelligence (ASI).* A ANI refere-se à habilidade computacional para realização eficiente de tarefas singulares, tal como rastreamento de páginas ou jogar xadrez. BOSTROM, N. "*Ethical Issues in Advanced Artificial Intelligence"*. Disponível em: http://www.fhi.ox.ac.uk/wp-content/uploads/ethical-issues-in-advanced-ai.pdf. Acesso em: 09 abr. 2020. "*It is good at performing a single task, such as playing chess, poker or Go, making purchase suggestions, online searches, sales predictions and weather forecasts*" MESKÓ, B. et al. "Will artificial intelligence solve the human resource crisis in healthcare?", *BMC Health Services Research*, 18, 2018, p. 545.
A AGI tenta representar o conceito original de inteligência, traduzindo-se em algoritmos com desempenho supostamente equivalente ou superior ao do ser humano e são esses algoritmos caracterizados por uma competência deliberadamente programada em um único domínio restrito. Tais algoritmos modernos de IA tendem a se assemelhar a quase toda vida biológica. BOSTROM, Nick. The ethics of artificial intelligence, em (W. Ramsey & K. Frankish, orgs.) *Draft for Cambridge Handbook of Artificial Intelligence,* Ed. Cambridge University Press, Cambridge, 2011.
E por fim a ASI se apresenta como "qualquer intelecto que exceda em muito o desempenho cognitivo dos seres humanos em, virtualmente, todos os domínios de interesse" BOSTROM, N. *Superintelligence*, Oxford University Press, London, 2014 p. 37.
Contudo, no contexto atual, verifica-se apenas a existência da ANI. Apesar de o ML estar se desenvolvendo por meio de redes neurais e com o DL, os preceitos teóricos e práticos gerais para implementação da AGI e da ASI ainda parecem longe de serem alcançados. Trata-se do Santo Graal da ciência computacional. "*General intelligence is still a major challenge, still highly elusive. AGI is the field's Holy Grail*." BODEN, M. A. *Artificial Intelligence: a very short introduction*, Oxford University Press, Oxford, 2018, p. 39.

só facilita a discussão, demonstração dos resultados, deduções conclusivas e a abertura às críticas pela comunidade.

Proposta 1: Um sistema de IA é ferramenta e, portanto, não é pessoa, não é sujeito de direitos e não titula direitos e deveres.

Circunstâncias fáticas e fundamentação jurídica: Embora o termo *pessoa* possa ser diretamente associado ao *ser humano*, em seu aspecto clássico, jurídico e hermenêutico refere-se a um *ser* capaz de titular direitos e obrigações na ordem jurídica.[22] Dissocia-se das premissas biológicas e antropocêntricas pela identificação da existência de pessoas não-humanas, tais como as pessoas jurídicas. Apesar se se tratar de uma abstração legal cuja idealização é instrumentalizada na movimentação patrimonial e de capital pretendendo extrair a *responsabilidade* da pessoa física e transferi-la para um patrimônio autônomo, constata-se a existência de *pessoa* que não seja *ser humano*. Porém, essa classificação ocorre a partir do reconhecimento legislativo e do processo legal em resposta à uma demanda de cunho social, econômico e identitário.

No atual contexto, não se exclui a possibilidade, ainda que remota, mas não se reconhece os seres artificialmente inteligentes como *pessoas*.[23] Outro fator que subsidia essa negativa ou menos a omissão no ordenamento jurídico em nível e também atrelado a uma lógica biológica são os elementos *inteligência e senciência*.[24] Contudo, entende-se prescindível a discussão entre autonomia, independência, agência, consciência ou intencionalidade dos entes artificialmente inteligentes, vez que tais condições não possuem conexões inexoráveis ao resultados legais pretendidos e podem gerar conclusões ficcionais descabidas da realidade. Portanto, a seriedade deve ser mantida como mecanismo de objetificação da argumentação e compreensão da IA enquanto sistema de computador.

Resultado parcial 1: Diante das premissas postas, a responsabilidade deve ser afastada dessas entidades e atribuída a outrem. Esse outrem se torna o objeto e o lócus da argumentação.

Proposta 2: O regime de responsabilização civil objetivo deve ser afastado e não deve ser aplicada a Teoria do Risco (Risked-based Approach) até a respectiva demonstração e comprovação dos possíveis danos a serem cometidos pelos sistemas de Inteligência Artificial

Circunstâncias fáticas e fundamentação jurídica: Duas são as condições que ensejam a responsabilidade objetiva no ordenamento brasileiro: 1) previsão legal; ou 2) atividade de risco. Até o momento, inexistem previsões legais reconhecendo a IA enquanto atividade de risco. Ressalta-se a existência do Projeto de Lei n° 2338, de 2023, em trâmite no Senado Federal cujo teor textual remete-se ao *Artificial Intelligence Act* europeu e inscreve em parte de seu teor o termo "categorização dos riscos". Embora nenhum dos normativos tenha expressado que os sistemas envolvendo IA são de *risco,* deduz-se como tanto a partir de sua categorização em risco excessivo e alto risco.

Perceba-se que *risco é um conceito operacional indispensável à construção do raciocínio hermenêutico.* É assustador e até mesmo preocupante a abstração

---

[22] Para Beviláqua, "pessoa é o ser a que se atribuem direitos e obrigações, enquanto a personalidade é a aptidão reconhecida pela ordem jurídica a alguém para exercer direito e contrair obrigações". BEVILÁQUA, C*. Teoria Geral do Direito Civil*, Servanda, Campinas, 2015, p. 81-82.

Essa proposta tem como fundamento o pensamento de Teixeira de Freitas, que diferenciava personalidade de capacidade, entendendo essa (a capacidade) como a extensão dada aos poderes de ação contidos na personalidade, ou seja, o modo de ser geral das pessoas. FREITAS, A. T. de. *Consolidação das Leis Civis*. Senado Federal, Brasília, 2003, arts. 21 e 22.

[23] Apesar de ser um ato isolado, tem-se uma exceção. Na Arábia Saudita, por exemplo, a robô Sophia, uma entidade dotada de inteligência artificial, teve sua existência reconhecida pelo sistema jurídico e a ela atribuída a condição de cidadã. STONE, Z.. "Everything You Need to Know About Sophia, The World's First Robot Citizen", *Forbes,* 2017.

[24] A discussão pode ser aprofundada em DIVINO, S. B. S. "After All, Artificial Intelligence is not Intelligent: in a Search for a Comprehensible Neuroscientific Definition of Intelligence" *Opinion Juridica*, 21, 2022, p. 1-21.

relegada ao intérprete da norma para definição do termo e aplicação ou incidência do suporte fático. Em outros termos, qual critério define se uma atividade econômica é ou não de *risco*? Inexiste um critério assente no Direito.[25] Duas são as saídas: recorrer aos ensaios das ciências exatas; ou verificar os padrões determinísticos mediante interpretação hermenêutica e sistemática. Em relação à primeira saída, verifica-se que o marco inicial dos estudos técnicos sobre risco (*risk*), incerteza (*uncertainty*) e lucro (*profit*) remontam-se a Frank Knight[26], em 1921.[27] Para o autor, risco é definido como o correlativo objetivo da incerteza subjetiva.[28] Assim, "se você não sabe ao certo o que vai acontecer, mas as chances existem, isso é risco. Caso você não saiba quais são as chances, então é incerteza".[29] A primeira conclusão que se pode verificar é a condicionalidade do risco à probabilidade e à incerteza.[30] Outro ponto de consonância entre os cientistas exatos é a necessidade probabilística da existência de dano somada a incerteza.[31] Por fim, o uso do conceito de risco é inerente à condição humana, vez que aparentemente é o único animal racional que consegue se orientar para o futuro. Assim, o risco aparenta ter seu uso apenas em uma sociedade que tenta se desligar do passado e conquistar o futuro.[32]

Essas características foram utilizadas como base para a *International Strategy for Disaster Reduction* (ISDR) da Organização das Nações Unidas (ONU) para definir risco como "A probabilidade de consequências prejudiciais ou perdas previstas (mortes, ferimentos, bens, meios de subsistência, atividade económica perturbada ou ambiente danificado) resultantes de interações entre perigos naturais ou induzidos pelo homem e condições vulneráveis".[33] É a partir deste ponto que o Direito enquanto

---

[25] E nas ciências exatas. "Talvez o único consenso proporcionado pelo termo 'risco' seja entre os filólogos, para os quais a sua origem é certamente incerta e muito antiga. Para Spink (2001), houve uma incorporação gradativa de termos passando da 'fatalidade' à 'fortuna', registrando-se diferentes termos para o mesmo fim já no século XII, até a expressão 'risco' no século XVI. Sabe-se, contudo, que o termo teve um emprego bem definido, ligado às transações comerciais no direito marítimo (Luhmann, 1993; Houaiss, 2001), embora passasse a ser usado de forma rara e numa variedade de contextos (Luhmann, 1993). O seguimento de registros mostra a primeira ocorrência da palavra no português em meados do século XV, havendo registro do francês risque (século XVI), provavelmente tomado do italiano risco, variação de rischio (século XIII). A palavra foi usada, por exemplo, no poema de Dante, Divina Comédia, escrito entre 1307 e 1321: Sì come, per cessar fatica o rischio, Li remi, pria ne l'acqua ripercossi Tutti si posano al sonar d'un fischio.". LIEBER, R. R. & ROMANO-LIEBER, N. S. O conceito de risco: Janus reinventado", em (M.C.S. Minayo & A.C. Miranda, orgs.) *Saúde e ambiente sustentável:* estreitando nós [online], Fiocruz, Rio de Janeiro, 2002, p. 71-72.

[26] KNIGHT, F. H. *Risk, uncertainty and profit,* Boston, Houghton Mifflin, 1921.

[27] SOUZA, K. R. G. & LOURENÇO, L. "A evolução do conceito de risco à luz das ciências naturais e sociais", *Territorium,* 22, 2015, p. 31-44.

[28] "*Risk is defined as "the objective correlative of the subjective uncertainty*" (p. 29), *which varies with the mathematical chance of loss in such a way as to be at a maximum when the chances for and against the event are exactly even"*. KNIGHT, F. H. *Risk, uncertainty and profit,* Boston, Houghton Mifflin, 1921, p. 44.

[29] "*if you don't know for sure what will happen, but you know the odds, that's risk, and if you don't even know the odds, that's uncertaint*". ADAMS, J. *Risk: the policy implications of risk compensation and plural rationalities*, London, UCL Press, 1995.

[30] GIDDENS, A. *O mundo na era da globalização,* Presença, Lisboa, 200.

[31] "*The notion of risk, therefore, involves both uncertainty and some kind of loss or damage that might be received. Symbolically, we could write this as: risk= uncertainty + damage*". KAPLAN, S. & GARRICK, B. J. "On the quantitative definition of risk", *Risk analysis*, 1 (1), 1981, p. 11-27.

[32] MENDES, F. "Risco: um conceito do passado que colonizou o presente", *Revista Portuguesa de Saúde Pública,* 20 (2), 2002, p. 53-62.

[33] "*The probability of harmful consequences, or expected losses (deaths, injuries, property, livelihoods, economic activity disrupted or environment damaged) resulting from interactions*

ciência social aplicada deve partir. Ignorar o conhecimento existente em outras áreas para assentar definições exclusivamente jurídicas é reconhecer a ignorância do jurista (*lato sensu*) enquanto mecanismo de exercício arbitrário do egoísmo humano. As consequências são as mais variadas, mas as principais são: confusão conceituais; aumento de complexidade em novas demandas; difusão da lógica experimental; e resultados práticos inalcançáveis pela réplica. Não se trata de uma objetivação da ciência jurídica, mas de crítica a ausência de parâmetros que a tornem exclusivamente subjetiva. Muitas das demandas que afetam o Direito Privado, especialmente o Direito Civil, surgem pela falta de lógica argumentativa e rigor conceitual ou semântico quanto ao critério de inovação na matéria. Portanto, não se pretende inovar no conceito de risco aqui trazido. Pretende-se importá-lo com sua respectiva adequação sistemática e hermenêutica a partir da leitura do *codice civile*. Frisa-se: hermenêutica sistemática e não exegese.

Objetiva-se com o conceito de risco estabelecer elementos de análise da realidade estrita a partir de uma passagem temporal.[34] Ocorre que essa percepção de risco não pode ser compreendida estrita e restritamente como um perigo objetivo que existe e pode ser medido a margem de um processo social e cultural, sob pena de decaimento em um objetivismo radical.[35] Deve-se reconhecer a relatividade da realidade e interpretar o conceito de risco conforme vieses e conceitos culturais e sociais de determinada sociedade. Essa é a concepção técnico científica mais adotada em teorias da ciência cognitiva.[36] Mas qual o impacto dessa constatação na interpretação hermenêutica?

Primeiramente, o conceito de risco previsto no art. 927, parágrafo único, do Código Civil de 2002 deve levar em consideração a realidade, a cultura e o sistema jurídico-social brasileiro. Importar regramentos sem quaisquer indicações e realização de estudos prévios para compreensão da necessidade e até mesmo da inteligibilidade da população acerca da temática a ser regulada é abrir margem para regulamentos sem eficácia e sem contextualização social. Essa importação indiscriminada, também consideravelmente criticada no direito comparado, tem sido realizada constantemente nos regulamentos envolvendo direito e tecnologia (vide LGPD brasileira) e agora no ato de definição do regulamento da IA. Os fatores de risco devem ser identificados a partir da análise econômica e sistemática dos mecanismos de inteligência artificial que afetam a sociedade. Não é um mecanismo por presunção, mas por verificação objetiva. Presumir que os sistemas artificialmente inteligentes sejam *sistemas de risco* apenas insere os envolvidos pela sua programação, fornecimento, distribuição e venda em uma posição de vulnerabilidade jurídica e descompassada com a realidade.

A crítica é incisiva em razão de tanto o projeto de lei brasileiro bem como o AIA europeu sequer apresentar a definição do conceito de *risco* e partirem do pressuposto de que a IA é uma atividade de risco por si só. Se considerarmos a probabilidade de dano, pode-se surgir e inferir premissas a partir da definição de dados estatísticos demonstrando o potencial de dano de um sistema de IA. Mas sua insurgência perpassa por uma demonstração quantitativa de eventos que envolveram

---

between natural or human-induced hazards and vulnerable conditions" UN. "*Living with risk*: a global review of disaster reduction initiatives", Geneva, 2002, p. 16.

"There is general agreement that the term risk factor means an exposure that is statistically related in some way to an outcome, e.g., smoking is a risk factor for *periodontitis*". BURT, B. A. "Definitions of risk", *Journal of dental education*, 65(10), 2001, p. 1007-1008.

[34] LIEBER, R. R. & ROMANO-LIEBER, N. S. "O conceito de risco: Janus reinventado", em (M.C.S. Minayo & A.C. Miranda, orgs.) *Saúde e ambiente sustentável:* estreitando nós [online], Fiocruz, Rio de Janeiro, 2002, p. 79.

[35] LIEBER, R. R. & ROMANO-LIEBER, N. S. "O conceito de risco: Janus reinventado", em (M.C.S. Minayo & A.C. Miranda, orgs.) *Saúde e ambiente sustentável:* estreitando nós [online], Fiocruz, Rio de Janeiro, 2002, p. 80.

[36] LIEBER, R. R. & ROMANO-LIEBER, N. S. "O conceito de risco: Janus reinventado", em (M.C.S. Minayo & A.C. Miranda, orgs.) *Saúde e ambiente sustentável:* estreitando nós [online], Fiocruz, Rio de Janeiro, 2002, p. 80.

circunstâncias e fatos danosos à sociedade, tal como nas situações pragmáticas exemplificadas na seção anterior. Ocorre que tais estudos, quando existem, sequer são levados em consideração para elaboração e justificativa da norma regulatória. Regular esse cenário prescindindo-se se tais dados fere a Liberdade Econômica empresarial e insere as sociedades empresárias em uma condição de incerteza sobre quando e como poderão ser responsabilizadas em razão dos possíveis danos a serem concretizados pelo sistema autônomo.

O sistema jurídico brasileiro apresenta uma alternativa interessante no Decreto n. 10.411/2020 e que poderia ser adotada enquanto ferramenta indispensável à tutela de interesses sociais: a Análise de Impacto Regulatório (AIA). Trata-se de instrumento destinado que "conterá informações e dados sobre os possíveis efeitos do ato normativo para verificar a razoabilidade do seu impacto econômico".[37] Embora sua previsão seja destinada exclusivamente "aos órgãos e às entidades da administração pública federal direta, autárquica e fundacional, quando da proposição de atos normativos de interesse geral de agentes econômicos ou de usuários dos serviços prestados (art. 1°)", entende-se que sua extensão ao poder legislativo mediante previsão legal (em observância ao princípio da legalidade) poderia afirmar o compromisso de elaboração de normativos condizentes à realidade e sem impactos indevidos, especialmente em uma temática tormentosa.

A proposta acima, apesar de ser subsidiária à argumentação e ao problema de pesquisa, demonstra o quão despreparado está o sistema jurídico brasileiro para a tutela de inovações, em especial as tecnológicas. Como conclusão, a hermenêutica adotada deve ser direcionada aos estudos previamente realizados e de acordo com os dados pré-existentes.

Assim, questiona-se: o regime de responsabilização objetiva é o mais adequado à tutela das relações jurídicas envolvendo entes inteligentes artificialmente inteligentes? A resposta a esse questionamento perpassa por algumas fases. Em primeiro lugar, deve-se concluir que:

Resultado parcial 2: O risco é composto por incerteza somado à probabilidade de dano.

A segunda abordagem é identificar se o sistema de inteligência artificial preenche os requisitos *incerteza* e *probabilidade de dano*. Em relação ao primeiro, devemos nos remeter ao ato de programação e às estruturas de aprendizado de máquina: o *Machine Learning.* Ressalta-se novamente que sob a ótica da ciência computacional, inteligência não é uma dimensão ou um conceito singular, mas um rico espaço estruturado em capacidades de processamento de informações.[38] A partir desse contexto, caso o agente esteja desenvolvendo e performando ações com base em sua observação de mundo, ele está supostamente aprendendo (*learning*).[39] O conceito de aprendizagem é amplo, podendo variar desde a descoberta de um número de um telefone à elaboração de uma teoria que explique o universo com fundamento nos dados de Einstein. Os aprimoramentos e as técnicas utilizadas para os agentes se aperfeiçoarem por meio de um banco de dados levam em consideração: 1) o componente a ser aprimorado; 2) o conhecimento *a priori* o

[37] "Art. 5º As propostas de edição e de alteração de atos normativos de interesse geral de agentes econômicos ou de usuários dos serviços prestados, editadas por órgão ou entidade da administração pública federal, incluídas as autarquias e as fundações públicas, serão precedidas da realização de análise de impacto regulatório, que conterá informações e dados sobre os possíveis efeitos do ato normativo para verificar a razoabilidade do seu impacto econômico". BRASIL. "*Lei da Liberdade Econômica. Lei n. 13.874 de 20 de setembro de 2019".* Diário Oficial [da] República Federativa do Brasil, Poder Executivo, Brasília, DF, 20 set. 2019.

[38] BODEN, M. *Artificial Intelligence: a very short introduction*, Oxford University Press, Oxford, 2018, p. 22.

[39] RUSSELL, S. J. & NORVIG, P. *Artificial intelligence*: *a modern approach*, Pearson Education, New Jersey, 2010, p. 693

agente já possui; 3) o tipo de representação a ser utilizada no banco de dados; e 4) o feedback adquirido pelo aprendizado realizado.[40] Um dos métodos mais utilizados e eficientes para essa tarefa é o aprendizado de máquina (*Machine Learning - ML*).

Em uma visão esquemática, o funcionamento do ML pode se dar da seguinte forma: o suposto aprendizado é visto como um pipeline que começa com dados brutos. Por exemplo. Uma coleção de arquivos executáveis com etiquetas associadas indica se um arquivo é benigno ou malicioso. Estes dados brutos são então processados para extrair características numéricas de cada instância i, obtendo um vetor de características associadas xi (isto poderia ser uma coleção de variáveis binárias indicando presença em um executável de chamadas de sistema particulares). Essas características associadas se tornam dados processados, mas doravante os chamamos simplesmente dados, pois é a este conjunto de dados processados que podemos aplicar algoritmos de aprendizagem, caracterizando o próximo passo no pipeline. Finalmente, o algoritmo de aprendizado produz um modelo dos dados que pode ser matemático (como sua distribuição) ou uma função (que prevê etiquetas em instâncias futuras).[41]

Algoritmos de Machine Learning – ML têm sido aplicados em quase todas as áreas das ciências da computação, ciências biológicas, engenharia, ciências sociais, etc.[42] Sem sua efetividade, muitas indústrias não teriam sobrevivido no atual e voraz cenário tecno-capitalista.[43]

Em um contexto operacional e experimental das ciências da computação, uma inteligência artificial assim o é designada por poder atuar por meio de processos racionais operados em representações internas de conhecimento. É neste sentido que se identifica uma IA baseada no conhecimento (Knowledge-based Agentes).[44] Quando se fala em racionalidade destinada à engenharia artificial, atrela-se sua relação à compreensão dos fatores ambientais destinadas ao aumento de performance.[45] É por meio da análise da performance que se obtém um dos critérios para definição de um agente racional. Ao seu lado, deve-se verificar um prévio conhecimento do agente para com o ambiente a ser interagido. A partir desse cenário, o agente poderá tomar e performar decisões por meio de ações sequências destinadas à compreensão e à modificação do ambiente.[46]

---

[40] RUSSELL, S. J. & NORVIG, P. *Artificial intelligence*: *a modern approach*, Pearson Education, New Jersey, 2010, p. 693-694.

[41] VOROBEYCHIK, Y. & KANTARCIOGLU, M. *Adversarial Machine Learning: Synthesis Lectures on Artificial Intelligence and Machine Learning,* Morgan & Claypool, Oregon, 2018, p. 5.

[42] "*A natural role for machine learning techniques in security applications is detection, examples of which include spam, malware, intrusion, and anomaly detection. Take detection of malicious email (spam or phishing) as a prototypical example*". VOROBEYCHIK, Yevgeniy; KANTARCIOGLU, M. *Adversarial Machine Learning*: *Synthesis Lectures on Artificial Intelligence and Machine Learning,* Morgan & Claypool, Oregon, 2018, p. 1.

[43] CHEN, Z. & LIU, B. *Lifelong Machine Learning,* Morgan & Claypool, Oregon, 2018, p. 1.

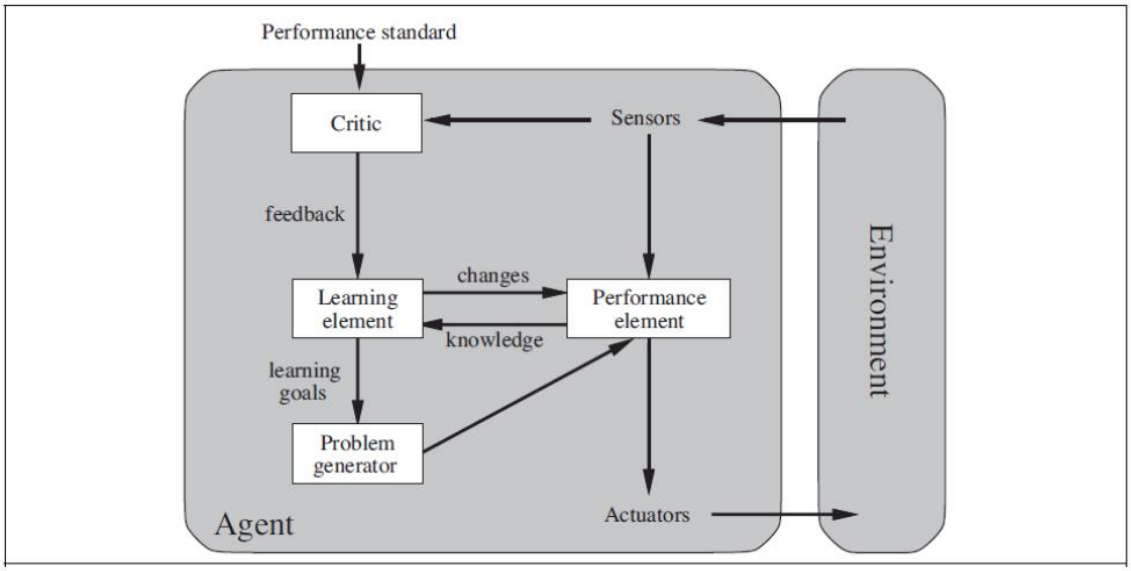
[44] RUSSELL, S. J. & NORVIG, P. *Artificial intelligence*: *a modern approach*, Pearson Education, New Jersey, 2010, p. 234.

[45] "*A rational agent can maximize this performance measure by cleaning up the dirt, then dumping it all on the floor, then cleaning it up again, and so on. A more suitable performance measure would reward the agent for having a clean floor. For example, one point could be awarded for each clean square at each time step (perhaps with a penalty for electricity consumed and noise generated). As a general rule, it is better to design performance measures according to what one actually wants in the environment, rather than according to how one thinks theagent should behave*". RUSSELL, S. J. & NORVIG, P. *Artificial intelligence*: *a modern approach*, Pearson Education, New Jersey, 2010, p. 37.

[46] Nesse contexto, Russell e Norvig definem o termo agente racional como "um agente racional deve selecionar uma ação que se espera maximizar sua medida de desempenho, dada a evidência fornecida pela sequência de percepção e qualquer conhecimento incorporado que o agente tenha". "*For each possible percept sequence, a rational agent should select an action that is expected to maximize its performance measure, given the evidence provided by the percept sequence and whatever built-in knowledge the agent has*" RUSSELL, S. J. & NORVIG, P. *Artificial intelligence*: *a modern approach*, Pearson Education, New Jersey, 2010, p. 37.

Russell e Norvig exemplificam como um agente pode supostamente aprender com o ambiente por meio dessa constatação

Figura 1: *A General Learning Agent*[47]



Cada quadrante maior representa uma atribuição conceitual distinta. Enquanto o elemento de aprendizado é responsável pelos aprimoramentos, o elemento da performance representa as ações a serem tomadas no ambiente externo. A performance é o elemento que representa o agente como um todo, pois é por seu intermédio que ele percebe e age independentemente. Após, caso exista algum feedback ele será incorporado pelo elemento do aprendizado e transformado em crítica para que o agente determine como suas ações estão sendo realizadas e cumpridas, bem como quais ações podem ser tomadas para atingir resultados diferentes e mais eficientes.

Por fim, o elemento gerador de problema (problem generator) representa a capacidade de sugerir ações que possivelmente guiarão o agente para novas e informativas experiências. Se aplicarmos o exemplo do caminhão que transporta cerveja podemos verificar tanto o elemento performance quanto o elemento crítico. O elemento da performance corresponde a qualquer background suficiente que sustente o conhecimento para tomada de ações enquanto a IA dirige o veículo. O elemento crítico, por sua vez, representa a observação do mundo e o transmite por meio do sistema de aprendizado. Exemplificando: caso o caminhão faça alguma manobra brusca, poderá ele danificar alguma propriedade pública ou privada. Por meio dessa experiência a IA pode definir que se trata de uma ação prejudicial à sua performance e não a realizar. Por fim, o problem generator pode supostamente analisar que a dirigibilidade fica prejudicada quando intempéries afetam diretamente o veículo. A partir deste ponto poderá o sistema fornecer opções de rotas ou métodos alternativos de aumento ou redução de velocidade para atender ao aumento de sua performance.

As ações de uma Inteligência Artificial passam por uma suposta base de conhecimento (Knowledge Base – KB) constituída por um conjunto de sentenças. Cada uma dessas sentenças não é delimitada em língua portuguesa ou em outros

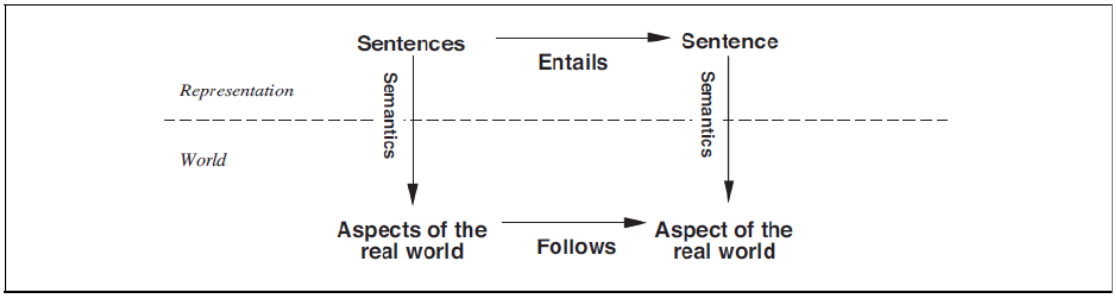
[47] RUSSELL, S. J. & NORVIG, P. *Artificial intelligence*: *a modern approach*, Pearson Education, New Jersey, 2010, p. 55.

idiomas, mas estritamente em termos técnicos.[48] Para Norvig e Russell, cada sentença é expressa em uma linguagem chamada linguagem de representação do conhecimento (Knowledge Representation Language) e corresponde a uma asserção sobre o mundo. Nessa linguagem, o agente é programado para realizar operações das quais ele deve contar (*to tell)* sobre os atributos inicialmente recebido sem sua base de conhecimento e perguntar (*to ask*) quais ações devem ser tomadas para, ao final, comunicar à sua base de conhecimento qual ação foi escolhida e assim executá-la.[49]

Nesse ambiente, a IA é programada para compreender a sintaxe e a semântica de cada sentença atribuída em sua base de dados. A sintaxe determina como todas as sentenças são bem formuladas. A semântica define o significado e a veracidade de cada sentença com relação à uma possibilidade no mundo (que, em termos mais exatos, pode ser substituído pela palavra model ou modelo).[50]

Pode existir, contudo, uma lacuna (*gap*) entre as funcionalidades semânticas pretendidas e a função programada quando existirem intenções implícitas ou ambíguas no sistema.[51] Ações como abrir a garrafa de cerveja, abrir os olhos, abrir a porta, abrir o livro, abrir o corpo, não revela qualquer incoerência semântica. Em todos os casos, ela (a semântica) está correta. Mas o verbo (abrir) possui contextos diversificados. Não se abre uma garrafa da mesma forma que se abre um livro ou se abre um corpo. O background é diferente. Sabe-se abrir uma garrafa e um livro, mas não um corpo. O background nos fornece condições de veracidade, conhecimento e satisfação diferente em cada situação. A constituição de um fenômeno é totalmente diversa em cada caso. A compreensão é, portanto, mais que apreensão do significado. Em termos gerais, aquilo que se entende vai além do significado. O contrário também pode ser verificado. Do ponto de vista gramatical as proposições o avião cortou os céus e Maria cortou uma goiaba estão corretas, mas mesmo com a apreensão do significado cortar não é possível compreender a sentença.[52]

Figura 2: Proposição Semântica da IA[53]



[48] RUSSELL, S. J. & NORVIG, P. *Artificial intelligence*: *a modern approach*, Pearson Education, New Jersey, 2010, p. 235.
[49] RUSSELL, S. J. & NORVIG, P. *Artificial intelligence*: *a modern approach*, Pearson Education, New Jersey, 2010, p. 235.
[50] RUSSELL, S. J. & NORVIG, P. *Artificial intelligence*: *a modern approach*, Pearson Education, New Jersey, 2010, p. 235.
[51] BERGENHEM, C, et al. "How to reach complete safety requirement refinement for autonomous vehicles", In *CARS 2015-Critical Automotive applications*: Robustness & Safety. 2015.
[52] SEARLE, J. *Intentionality: An Essay in the Philosophy of Mind*, Cambridge University Press, Cambridge, 1983, p. 143 e DIVINO, S. B. S. "Consciência e intencionalidade na Teoria do Fato Jurídico de Pontes de Miranda*:* Direito sem objetividade?", *Dissertação de Mestrado. Pontifícia Universidade Católica de Minas Gerais*, Minas Gerais, Belo Horizonte, 2019, p. 68.
[53] "*Sentences are physical configurations of the agent, and reasoning is a process of constructing new physical configurations from old ones. Logical reasoning should ensure that the new configurations represent aspects of the world that actually follow from the aspects that the old configurations represent*". RUSSELL, S. J. & NORVIG, P. *Artificial intelligence*: *a modern approach*, Pearson Education, New Jersey, 2010, p. 243 117 e BURTON, S, et.al. "Mind the Gaps: Assuring the Safety of Autonomous Systems from an Engineering, Ethical, and Legal Perspective", *Artificial Intelligence*, 279, 2020, p. 4.

Existem três fatores que dão origem à lacuna semântica. O primeiro fator é a complexidade e a imprevisibilidade do domínio operacional. Sistemas autônomos normalmente operam em um ambiente onde não se pode taxar ou delimitar completamente todas as operações a serem designadas em virtude da complexidade e da invariabilidade do ambiente. Ou seja, essa indeterminabilidade faz com que o sistema esteja sempre em aberto, pois delimitar e taxar todas as especificações é extremamente difícil ou até mesmo impossível para formalizá-las.[54]

O segundo fator é a complexidade e a imprevisibilidade do próprio sistema. Como o sistema adota técnicas de ML e DL em sua composição, a complexidade linguística é inerente a ela e faz com que o próprio sistema mude constantemente mediante interações realizadas em seu domínio de atuação. Pretende-se que os sistemas aprendam e antecipem as intenções dos seres humanos para melhor se adaptarem durante sua execução, pois assim podem produzir análises de riscos inicialmente não previstos pelos desenvolvedores.[55]

Por fim, o terceiro fator reside na crescente necessidade de transferir as decisões para um sistema onde o ser humano pode ser completamente substituído ou deixado em segundo plano (atuando neste último caso como agente fiscalizador). Porém, mesmo com essas restrições, os avanços nas técnicas de ML e o maior poder de processamento computacional fizeram com que os sistemas de IA se desenvolvessem para compreender sua atual situação por meio de sensores externos e decidir suas ações por meio de estratégias baseadas em seu background mediante experiência adquirida. Mesmo em ambiente sem restrições, as redes neurais profundas completam o sentido de dados que poderiam ser apenas um conjunto de informações abstratas dentro do sistema.[56]

Mesmo diante da verificabilidade pragmática desses fatores, inexiste consenso sobre como reduzir as lacunas semânticas e tornar o sistema mais seguro em razão da incerteza advinda da combinação dos gaps indicados. O fato de um sistema de inteligência artificial conseguir realizar tais tarefas não demonstra um equívoco ou um erro, mas uma possibilidade de agir racionalmente em resposta a estímulos externos (*input*) ainda que existam inferências semânticas em sua constituição que simplesmente não podem ou dificilmente serão solucionadas em sua programação. Trata-se de uma aproximação entre a conduta humana e a conduta maquinária, mas

[54] BURTON, S, et.al. "Mind the Gaps: Assuring the Safety of Autonomous Systems from an Engineering, Ethical, and Legal Perspective", *Artificial Intelligence*, 279, 2020, p. 4, 103201.
[55] BURTON, S, et.al. "Mind the Gaps: Assuring the Safety of Autonomous Systems from an Engineering, Ethical, and Legal Perspective", *Artificial Intelligence*, 279, 2020, p. 4.
[56] "*Machine learning techniques such as Deep Learning (enabling the system to "make sense" out of the unstructured data that results from the complex and unpredictable environment) and Reinforcement Learning (enabling the system continually to optimise a function based on stimuli collected in the field) might appear to present a convenient technical solution to the semantic gap. They seem particularly well-suited to learning functionality that cannot be easily specified using traditional procedural means (if X happens, then do Y). But there is a catch. Machine learning functions do not deliver clear-cut answers. For example, for a given video frame, they might classify the probability of a pedestrian inhabiting a certain portion of the picture as 83%, but in the very next frame -- which for humans is imperceptibly different to the last -- they may "misclassify" the same object as only 26% probability of being a human and 67% probability of being a road sign. In addition, the processes which lead to these decisions are difficult to decipher. These attributes result in a paradox or "no free lunch" effect, where the problem of deriving a suitable specification of the intended behavior is instead transferred to the problem of demonstrating that the implemented (learned) behaviour meets the intent*". BURTON, S, et.al. "Mind the Gaps: Assuring the Safety of Autonomous Systems from an Engineering, Ethical, and Legal Perspective", *Artificial Intelligence*, 279, 2020, p. 4.

que neste último caso ocorre mediante processos racionais legitimados em sentenças construídas em um plano essencialmente objetivo.[57] Dessa forma,

Resultado parcial 3: os sistemas de IA são incertos e, portanto, preenchem o primeiro requisito para adoção da responsabilidade objetiva.

O segundo fator determinante da responsabilização objetiva é o dano, seja em sua probabilidade de ocorrência (preponderantemente alta) ou seja em seu impacto (vide atividades nucleares ou transportes aéreos[58]). Aplicando-se a definição de dano à área tecnológica em análise, verifica-se se se trata de uma perda resultante de um sistema de inteligência artificial cuja causa pode ser interna, externa ou por fato de terceiros, incluindo invasões, incidentes de segurança informacional, fraudes internas e externas e perturbação da atividade empresarial.[59]

A identificação da probabilidade do dano deve perpassar por uma análise estatística de sua ocorrência objetivando identificar os fatores que ensejam ou mesmo eximem a ocorrência do fato. Esse estudo, no momento, não pode ser realizado neste trabalho. Contudo, questiona-se a possibilidade do reconhecimento da probabilidade de dano por dedução a partir das situações pragmáticas anteriormente exemplificadas. O primeiro pressuposto que deve ser levado em consideração é a inexistência de um sistema à prova de falha. Mesmo em ciências exatas, não existe exatidão. O mesmo ocorre na programação: inexiste um sistema à prova de falhas. O questionamento que deve ser feito é: quais ações são tomadas pela atividade empresarial para que essa falha seja evitada? Neste caso, o desenvolvimento de sistemas defensivos para impedir invasões de terceiros ou mesmo incidentes internos têm o condão de reduzir as margens probabilísticas de ocorrência dos infortúnios, mas não sua completa extirpação. O questionamento ainda persiste: quais são os possíveis danos que podem ser cometidos pelo sistema de IA à sociedade? A resposta é variada e conforme o ramo de atuação do sistema, que pode ser desde um incidente de segurança de dados a um evento de morte.

Dessa forma, não se pode afirmar que toda atividade envolvendo sistemas artificialmente inteligentes sejam potencialmente causadoras de danos que tais danos sejam consideravelmente impactantes na sociedade. O que se pode argumentar e propor para não se isentar da responsabilidade científica sobre a temática é verificar de forma objetiva quantas pessoas possivelmente serão afetadas pelo sistema de inteligência artificial e quais direitos possivelmente serão afetados a partir de eventual incidente. Quanto maior o número de pessoas, pode ser que o risco aumente em razão das variáveis aceitas. Lado outro, quanto mais sensíveis os direitos, maior a probabilidade de danos incompensáveis, vez que sua ontologia pode

---

[57] Alan Turing desenvolve um dos mais importantes conceitos relativos à IA nesse contexto: a universalidade. Trata-se da possibilidade de um pequeno dispositivo carregar consigo a capacidade e aptidão de agir em multitarefas sem segregar ou separar o conceito de cada tarefa, tendo em vista que o sistema fará todos de forma uniformizada e autônoma, que pode variar desde escuta, fala, tradução, animação, etc. TURING, A. "On computable numbers, with an application to the Entscheidungsproblem", *Proceedings of the London Mathematical Society*, 42 1936, p. 230–65.

[58] "Eles são extremamente raros. O risco de envolvimento de um avião num acidente, onde podem ocorrer diversas fatalidades e já calculando-as, é de um em três milhões. Colocando-se este dado numa perspectiva, para ser possível ter uma idéia, seria necessário uma pessoa voar pelo uma vez por dia durante 8,1 mil anos para se atingir este total de três milhões de vôos. Apesar de ser raro ocorrer um acidente, a comunidade de aviação do mundo todo está trabalhando freneticamente para abaixar ainda mais esta probabilidade. Há 30 anos, a probabilidade de ocorrer um acidente era uma para cada 140 milhões de milhas voadas; hoje, a cada 1,4 bilhão. O fator segurança está dez vezes melhor em três décadas FOLHA. "Com que frequência acidentes ocorrem?" *Folha Online,* 2023.

[59] "*Specifically, we define cyber risk as the risk of loss resulting from digital incidents caused by internal, external, or third parties, including theft, compromised integrity and/or damage to information and/or technology assets, internal and external fraud, and business disruption*". CURTI, F et al. "Cyber risk definition and classification for financial risk management", *Journal of Operational Risk*, 18 (2), 2023, p. 3.

estar atrelada à esfera extrapatrimonial de seu titular, tal como os direitos da personalidade.

Perceba-se que existe uma incerteza no que tange à existência dos possíveis danos e essa afirmação, por si só, conduz à não autorização de forma irrestrita do regime de responsabilização objetiva às atividades empresariais que a desenvolvem. Deve-se criar um fator de análise objetiva com fundamento em um AIR para estabelecer as condições necessárias à aplicação desse regime jurídico, pois caso contrário fica o intérprete da norma com considerável poder sobre a circunstância fática a fim de definir se há provável dano ou não. Neste caso, tem-se considerável incerteza sobre o resultado frente ao possível desconhecimento do intérprete acerca das demandas que influenciam aquela relação contratual. Portanto,

Resultado parcial 4: a ausência temporária de comprovação da probabilidade e dos impactos dos danos causados pelos sistemas de inteligência artificial afastam a aplicação da responsabilização objetiva.

Algumas são as observações diante do resultado proposto. Primeiro, ele somente é aplicado quando inexistem regimes e microssistemas específicos. Caso a tutela seja de uma relação de consumo, por exemplo, aplica-se o CDC em sua integralidade. Em segundo lugar a proposta é temporária. Portanto, assim que comprovados a probabilidade do dano e seu respectivo impacto pode-se afastar o resultado proposto e aplicar o sistema de responsabilização objetivo nas relações jurídicas envolvendo sistemas artificialmente inteligentes. Por fim, a negativa da responsabilização objetiva faz com que seu regime subjetivo, em regra, seja aplicado. Porém, com as devidas considerações.

Diante o exposto, a abordagem baseada no risco (*risk-based approach*) parece não ser adequada em razão de sua presunção abstrata frente às circunstâncias fáticas desconhecidas ou mesmo se conhecidas com dados insuficientes que sustentem sua aplicação. Portanto, a tutela dos danos cometidos por entes inteligentes artificialmente (e advindos do ato ilícito) tem aparente aplicação na responsabilização subjetiva. Neste sentido, o posicionamento expresso neste trabalho vai frente às atuais propostas legislativas, mas apresenta certo grau de racionalidade a partir da indicação objetiva de parâmetros inverificáveis no momento da constituição do sistema autônomo inteligente.[60]

Lado outro, ocorre que a aplicação da modalidade de responsabilização subjetiva não deve ocorrer pela simples subsunção de seus clássicos requisitos, quais sejam, culpa, dano, e nexo de causalidade, pois podem existir fatores externos capazes de romper o nexo de causalidade a partir da análise da circunstância fática. Essas condições serão explicadas e o raciocínio são desenvolvidos neste momento.

Proposta 3: O elemento culpa deve ser analisado e verificado sob a ótica do responsável pela utilização da IA enquanto ferramenta à consecução do dano.

Em primeiro lugar, *culpa* deve ser entendida como *volitiva* em seu sentido *lato sensu*. Assim, abarca-se o *dolo* e a *culpa stricto sensu* (negligência, imprudência e imperícia). Deste ponto, parte-se para a análise do suporte fático.

Relembremos a situação envolvendo automação veicular e os incidentes do Autopilot da Tesla. A premissa fática é a seguinte: caso o agente determine uma

[60] "*But more substantively, even if we take the extra step and create a link between human stakeholders and certain actions of AI-based robots, which hu- man would be better situated to make the cost-benefit analysis of risks and their avoidance? In contrast to the simpler case of products liability, where manufac- turers could be assumed to better understand the risks involved in using their products and perhaps even prevent them, this may simply not be the case for AI- based robots. The general problem of the lack of foreseeability of robot behavior is crucial to make the normative decision regarding the allocation of risk under strict liability. In some cases, one stakeholder may have more information than other stakeholders with respect to potential risks*". RACHUM-TWAIG, O. "Whose Robot Is It Anyway? Liability for Artificial-Intelligence-Based Robots", *U. Ill. L. Rev*., 2020, p. 1141.

ordem para o sistema autônomo causar dano a outrem, verifica-se o dolo e, no mínimo, atribui-se a responsabilidade de indenizar a esse agente.

Ocorre que essa observação é consideravelmente simplista. Poderia o programador ou mesmo o fornecedor de serviços ser responsável pelos danos cometidos pela ordem deste *agente que atua enquanto cliente/consumidor?* A resposta para esse questionamento é consideravelmente complexa, pois perpassa por uma análise tanto ética quanto normativa. Em relação a essa última, existe uma lacuna sobre *como e quais parâmetros legais devem ser observados durante a programação de um ente inteligente artificialmente*. Porém, sob o aspecto ético, desde abril de 2019, há a *Ethics Guidelines For Trustworthy AI*, elaborada pelo Grupo de Peritos de Alto Nível sobre Inteligência Artificial, criado pela Comissão Europeia em junho de 2018.[61] Por se tratar de indicações de ordem ética, não sofrem limitações quanto aos preceitos territoriais e jurídicos de cada entidade soberana. Assim, para uma IA ser considerada confiável ela deve apresentar, nos termos do normativo citado, três componentes ao longo de sua execução, que devem ser aplicados de forma harmoniosa e concomitante: 1) Legalidade, garantindo o respeito de toda a legislação e regulamentação aplicáveis; 2) Eticidade, garantindo a observância de princípios e valores éticos; e 3) Solidez, tanto do ponto de vista técnico como do ponto de vista social, uma vez que, mesmo com boas intenções, os sistemas de IA podem causar danos não intencionais.[62]

Segundo a diretriz europeia, as legislações que pretendem regular e estabelecer direitos e deveres no campo da IA deve ser interpretada não só à luz do que não pode ser feito, mas também do que deve ser feito. O quadro de ações dos cientistas da computação deve ser direcionado para preceitos éticos e para transpassar às pessoas e à sociedade um cenário de confiança de que a IA não causarão danos não intencionais. Dessa forma, os princípios norteadores dessa conduta podem ser prescritos como: I) Respeito da autonomia humana[63]; II) Prevenção de danos[64];

---

[61] UNIÃO EUROPEIA. *Diretrizes éticas para uma IA de confiança (Ethics Guidelines For Trustworthy AI),* 2019.

[62] UNIÃO EUROPEIA. *Diretrizes éticas para uma IA de confiança (Ethics Guidelines For Trustworthy AI),* 2019.

[63] O princípio da autonomia humana tem como objetivo a manutenção da autodeterminação plena e efetiva sobre si próprios e participação no processo democrático quando da utilização de sistemas de IA para tal finalidade. "*In a broad sense, automated decision-making can describe the very nature of IT-enabled algorithmic processes, which is producing outputs by means of executing a computer code (Article 29 Working Party, 2017a; Kroll et al., 2017). Admittedly, it is the fact that the underlying data collection and analysis as well as the subsequent procedural steps are performed automatically (by technological means) – and therefore more quickly and extensively than th same could ever be done by humans – that lies at the heart of the challenges investigated as part of this project. According to this understanding, algorithmic decision-making could thus refer to 1) automated data gathering and knowledge building and to 2) the performance of subsequent procedural steps – encoded in an algorithm or adjusted autonomously by artificial agents – with a view to reaching a predetermined goal. Obviously, such a perception gives rise to significant overlaps with other applications of AI in consumer markets investigated as part of this project. Once again, we would like to argue that this is not really a problem"*. JABŁONOWSKA, A. et., al. "Consumer law and artificial intelligence Challenges to the EU consumer law and policy stemming from the business' use of artificial intelligence" *Final report of the ARTSY Project. European University Institute*. 2018, p. 38, p. 38. Sobre a interferência da IA na autodeterminação da vontade, ver mais em: CITRON, D. K.; PASQUALE, F. A. "The Scored Society: Due Process for Automated Predictions", *Washington Law Review*, 1, 2014, p. 2-27; e BEUC. "Automated decision making and artificial intelligence: a consumer perspective", *Tech. rep., Bureau Europeen des Unions de Consommateurs*, 2018.

[64] Pelo princípio da prevenção de danos, entende-se que "os sistemas de IA não devem causar danos ou agravá-los nem afetar negativamente os seres humanos de qualquer outra forma". Deve-se garantir solidez na adoção dessas técnicas para que os resultados sejam prolíficos para todos os sujeitos envoltos na relação em análise.

III) Equidade[65]; e IV) Explicabilidade[66].[67].

Tais exigências ficam em um plano ideal. Contudo, como concretizá-las? Como tornar possível a utilização de uma IA de confiança? Como o nível atual de cognição de uma IA está no degrau ANI, ou seja, apenas para o exercício singular de tarefas, a abordagem adotada neste momento será direcionada para mecanismos capazes de comutar de forma simbiótica a correlação homem-máquina. Ou seja, diante do atual desenvolvimento tecnológico, técnicas de IA podem e devem ser supervisionadas por humanos. E um dos mecanismos a ser utilizado para a construção dos parâmetros éticos e normativos acima descritos é a governança digital.

Para Floridi[68], "a governança digital é a prática de estabelecer e implementar políticas, procedimentos e padrões para o desenvolvimento, uso e gerenciamento adequados da *infosphere*".[6970] A demonstração de uma política de governança digital é elemento capaz de romper o nexo de causalidade entre o programador e o fornecedor frente ao ato ilícito praticado pelo sistema autônomo. Retornando ao questionamento posto: Poderia o programador ou mesmo o fornecedor de serviços ser responsável pelos danos cometidos pela ordem deste *agente que atua enquanto cliente/consumidor?* Se existem mecanismos e códigos de programação destinados e cuja finalidade seja a prevenção de danos a terceiros a resposta é negativa. Dessa forma, caso o programador insira um comando em observância ao princípio ético da prevenção de danos e esse comando seja verificável no código do sistema autônomo

---

[65] O princípio da equidade "implica que os profissionais no domínio da IA devem respeitar o princípio da proporcionalidade entre os meios e os fins, e analisar cuidadosamente a forma de equilibrar os interesses e objetivos em causa". Isso conduz à adoção de procedimentos que evitem enviesamentos injustos, discriminação e estigmatização contra pessoas e grupos.

[66] "Significa que os processos têm de ser transparentes, as capacidades e a finalidade dos sistemas de IA abertamente comunicadas e as decisões — tanto quanto possível — explicáveis aos que são por elas afetados de forma direta e indireta. Sem essas informações, não é possível contestar devidamente uma decisão" UNIÃO EUROPEIA. *Diretrizes éticas para uma IA de confiança (Ethics Guidelines For Trustworthy AI)*. 2019.

[67] UNIÃO EUROPEIA. *Diretrizes éticas para uma IA de confiança (Ethics Guidelines For Trustworthy AI),* 2019.

[68] "*Digital governance is the practice of establishing and implementing policies, procedures and standards for the proper development, use and management of the infosphere. It is also a matter of convention and good coordination, sometimes neither moral nor immoral, neither legal nor illegal*". FLORIDI L. "Soft ethics, the governance of the digital and the General Data Protection Regulation", *Phil. Trans. R. Soc*. 2018, A 376: 20180081.

[69] "Através da supervisão humana, pretende-se garantir que um sistema de IA não coloque em causa a autonomia humana e nem produza efeitos negativos. A governança digital pode incluir diretrizes e recomendações que se sobrepõem à regulamentação digital, mas não são idênticas a ela. Essa é apenas outra maneira de falar sobre a legislação relevante, um sistema de leis elaborado e aplicado por meio de instituições sociais ou governamentais para regular o comportamento dos agentes relevantes na infosphere". FLORIDI L. "Soft ethics, the governance of the digital and the General Data Protection Regulation", *Phil. Trans. R. Soc*. 2018, A 376: 20180081.

[70] "Dentre as atividades voltadas à ciência da computação que podem ser utilizadas como mecanismos para aperfeiçoamento e aprimoramento da conduta humana frente aos entes inteligentes artificialmente três se destacam: 1) o human-in-the-loop (HITL); 2) o human-on-the-loop (HOTL); e 3) o human-in-command (HIC). O primeiro refere-se à capacidade interventiva de uma pessoa no procedimento durante cada processo de decisão da IA. O segundo toma como base apenas sua monitoração através da visualização no sistema. E o terceiro traduz-se na capacidade de supervisão global do sistema, incluindo os impactos econômicos, sociais, jurídicos e éticos, bem como a habilidade de decidir quando e como usar o sistema em si". DIVINO, S. B. S. & MAGALHAES, R. A. "Inteligência Artificial e Direito Empresarial: Mecanismos de Governança Digital para Implementação e Confiabilidade", *Economic Analysis of Law Review*, 11, 2020, p. 72-89.

de inteligência artificial, portanto há precaução suficiente a ponto de criar um mecanismo de rompimento do nexo de causalidade frente ao dano cometido.

Lado outro, a inobservância dos princípios éticos faz com que o programador ou mesmo o fornecedor sejam responsáveis subsidiária ou solidariamente (a depender do regime legal aplicado, se CC ou se CDC) frente ao dano cometido. Perceba-se que o elemento *precaução é inerente à toda atividade econômica e inclusive critério elementar de risco da análise econômica do direito*. É a partir deste critério, objetivamente verificável in loco, que se atribui ou não o dever de indenizar. Diante a existência de condutas, programações e até mesmo práticas de governança digital que concretizem os preceitos éticos, não há que se falar em atribuição de responsabilidade ao programador ou mesmo ao fornecedor do sistema de inteligência artificial. Porém, frisa-se: este raciocínio é preponderantemente aplicável, como regra, na responsabilização subjetiva, onde há análise de culpa. Contudo, nada obsta seu reconhecimento na responsabilização objetiva, vez que o rompimento do nexo causal ainda é fator de exoneração da responsabilização junto à culpa exclusiva da vítima. Então, mesmo que um veículo autônomo colida diretamente com um obstáculo a mando do seu motorista, se verificados preceitos éticos no código de programação da IA inexiste responsabilização pelos danos causados.

Resultado parcial 5: A observância aos preceitos éticos devidamente comprovados e documentados por políticas de Governança Digital é fator de rompimento do nexo causal e de exoneração de responsabilidade (objetiva e subjetiva) pelo ato ilícito.

Em pensamento oposto, a impossibilidade da comprovação das práticas enseja o dever de indenizar. Neste caso, a indenização será pela modalidade transubjetiva (por fato de coisa, de animal ou de outra pessoa) ou *vicarious liability.* Se subsidiária ou se solidária, necessária análise dos dispositivos legais que regulamentam a relação jurídica. A título exemplificativo, se relação de consumo, tem-se responsabilidade solidária com fundamento no art. 7°, parágrafo único do CDC. Lado outro, se relação regida pelo *codice civile* deve-se verificar se há previsão expressa de solidariedade (legal ou contratual) em respeito à sua vedação de presunção.

Porém, outro fator que se deve analisar para evitar a responsabilidade vicária parte dos pressupostos éticos de governança digital. Em qualquer circunstância o *ser humano* deverá e poderá assumir o controle da IA. O caso Viking Sky trazido anteriormente é um exemplo a não ser seguido. Embora a experiência do comandante supostamente fosse superior à do sistema que controlasse o navio, a IA foi incapaz de deixá-lo reiniciar e dar partida no cruzeiro em questão. Ou seja, deverá existir no código de programação da IA uma condicionante de que, em último caso, ou quando expressamente solicitado em caso de emergência, o *ser humano assuma o controle para evitar o dano*. Trata-se da concretização do preceito *Human-in-command (HIC).*[71] Segundo Muller, "precisamos de uma abordagem HIC da IA, onde as máquinas permanecem máquinas e as pessoas mantêm o controle sobre essas máquinas o tempo todo".[72] Em sua concepção, "agentes humanos podem e também devem ter o controle de se, quando e como a IA é usada no cotidiano, bem como quais tarefas transferimos para a IA, quão transparente é, e o respeito aos aspectos éticos".[73]

---

[71] O *European Economic and Social Committee (EESS)* propõe uma abordagem *Human-in-command (HIC)* para os países componentes da União Europeia quando da utilização de entes inteligentes artificialmente. Em termos singelos, o HIC coloca a IA como uma ferramenta, onde agentes humanos decidem quando e como usá-la. THE STRATEGY WEB. *Bosch Connected World 2020 – The human unicorn on AI.* 2020.

[72] "*We need a human-in-command approach to AI, where machines remain machines and people retain control over these machines at all times*". UNIÃO EUROPEIA. "*Artificial Intelligence: Europe needs to take a human-in-command approach, says EESC",* 2017.

[73] "*Humans can and should also be in command of if, when and how AI is used in our daily lives – what tasks we transfer to AI, how transparent it is, if it is to be an ethical player*".

Assim, pretende-se estabelecer um modelo denominado *rule of law by design*, capaz de implementar métodos de garantia para adoção de princípios éticos e abstratos que o sistema de IA deve respeitar quando da tomada de decisões. Neste caso, as empresas envoltas nesse ramo devem identificar desde o início da programação quais as normas jurídicas são atreladas ao funcionamento do seu sistema e designá-las para seu efetivo cumprimento objetivando evitar impactos negativos no cenário econômico. Assim, o agente humano deve implementar um mecanismo que permite o desligamento do sistema em casos críticos, para permitir a retomada da operação sob a autoridade humana. Portanto,

Resultado parcial 6: A falta de um mecanismo destinado à implementação do Human-In-Command é justificativa para aplicação da responsabilidade transobjetiva/vicária (por fato de coisa, de animal ou de outra pessoa).

Perceba-se que a adoção de modelos de Governança Digital dependerá basicamente da escolha do empresário levando em consideração a disponibilidade da tecnologia, os custos de transação atrelados a ela, bem como os riscos trazidos pelo seu uso. Quanto maior a capacidade de automação proporcionalmente será o dispêndio para o desenvolvimento e gerenciamento dessa tecnologia. Lado outro, a redução da atividade humana pressupõe também uma diminuição de custos para com agentes envoltos naquele setor.

Outra análise pode ser realizada utilizando essas considerações se buscarmos novamente o caso da Clarkesworld Magazine. Uma IA não tem potencial criativo.[74] Toda sua produção advém de um amálgama de dados. Caso essa produção viole direitos autorais de um titular lesado, em que circunstâncias o programador ou fornecedor poderá ser responsabilizado? Neste caso quando esses não inserem no código o mandamento ético *antiplágio*. Em outros termos, não poderá a IA utilizar de materiais protegidos por *copyright*. Caso a IA seja desenvolvida para produção ou edição de imagens deverá seu banco de dados ser única e exclusivamente atrelado aos materiais sem *copyright* ou, caso o tenham, estejam diretamente afetados à licença de uso do *software* (ex: contrata-se um editor de imagem com a promessa de utilização de um banco de imagens com X materiais produzidos exclusivamente por um terceiro para com aquela finalidade).

A solução para a premissa fática *Data is the New Black* também é a mesma. Haverá responsabilização dos programadores e fornecedores quando não existir justificativa suficiente e fundamentos que sustentem o ato de discriminação praticado pelo ente inteligente artificialmente.[75] Neste caso, a transparência algorítmica é fator inerente à atividade empresarial neste setor. Portanto, deverá o empresário estabelecer em seus mecanismos de Governança Digital métodos explicativos compreensíveis e inteligíveis das tomadas de decisões automatizadas cujo dados são considerados sensíveis.

Por fim, outra espécie de responsabilização que pode ser aplicada e possui reconhecimento no sistema jurídico brasileiro é Responsabilidade preventiva e sem danos efetivos. O tradicional exemplo que pode ser evidenciado aqui é a criação de um fundo securitário que seja capaz se arcar com os prejuízos causados pelos interessados aderentes. Neste caso, entendo a existência deste fundo está condicionada à uma previsão legislativa cuja elaboração seja calcada em análise

UNIÃO EUROPEIA. "*Artificial Intelligence: Europe needs to take a human-in-command approach, says EESC",* 2017.

[74] DIVINO, S. B. S. & MAGALHAES, R. "A. Propriedade intelectual e direito autoral de produção autônoma da inteligência artificial", *Revista de Direitos e Garantias Fundamentais*, 21, 2020, p. 167-192.

[75] DIVINO, S. B. S. "Desigualdade codificada: como o uso de algoritmos pode reduzir, ocultar e aumentar a desigualdade?" em (C. Colombo & W. Engelmann & J. Faleiros Júnior, orgs.), *Tutela Jurídica do Corpo Eletrônico.* Foco, Indaiatuba, 2022, p. 723-738.

concreta de riscos mediante AIR. Sua exigência sem tal observação viola a liberdade econômica por condicionar seu exercício a um fator abstrato. Outro aspecto relevante é que, uma vez que um regime de seguro obrigatório se destina a suplantar totalmente o sistema geral de responsabilidade civil, exige-se uma adoção hermética e ampla por todas as partes interessadas.[76] A razão se dá em virtude da diferença existente entre local de ocorrência do dano e a sede da sociedade empresária responsável pela fabricação do produto. A sede da atividade empresária pode estar localizada na China, mas o dano pode ter ocorrido no Brasil. A sede poderia estar localizada no Japão, mas o dano materializado nos Estados Unidos da América. Assim, como garantia de efetivação, o fundo securitário deveria ter margens suficiente para abarcar todas as partes interessadas e envolvidas na produção de sistemas autônomos inteligentes. Portanto,

Resultado parcial 7: É possível a aplicação da responsabilidade preventiva, desde que condicionada à uma exigência normativa e elaborada mediante Análise de Impacto Regulatório.

A partir desses resultados expressos, verifica-se que o atual sistema jurídico consegue sustentar e ser aplicado nas relações jurídicas envolvendo entes inteligentes artificialmente. Não há necessidade de considerável inovação. A necessidade premente é que os juristas responsáveis pelos estudos firmem os pés no chão e renunciem a discussões distópicas capazes de atrasar o desenvolvimento científico na área ou mesmo trazer uma tutela desproporcional ao setor econômico. Acredita-se que as discussões aqui trazidas sejam parte de um conjunto de estudos destinados à regulação e normatização dos sistemas autônomos (e essencialmente criticável).

4. Considerações Finais

Conforme visualizado, a temática é consideravelmente ampla, complexa e com margens suficientes à escrita de consideráveis reflexões. Porém, em razão da limitação metodológica e de estilo, acredita-se que as premissas trazidas foram suficientes para responder ao problema de pesquisa proposto, bem como indicar novas hipóteses à tutela da responsabilidade ética e jurídica relativa aos entes inteligentes artificialmente.

Assim, devemos retornar ao problema norteador deste trabalho: como deve ser a tutela ética e jurídica dos danos causados por um ente artificialmente inteligente? De forma objetiva, a tutela deve perpassar pela inclusão de preceitos éticos durante a programação e o regime de responsabilização, como regra, deve ser o subjetivo, vez que a comprovação do possível risco resta prejudicada pela falta de comprovação dos requisitos do risco (especialmente o dano). No mais, verifica-se que a responsabilidade ética precede a responsabilidade jurídica, e essa será aplicada quando aquela não for visualizada. Portanto, a responsabilidade ética no ato de programação pode ser vista como fator de exoneração de responsabilidade mediante rompimento do nexo causal.

Assim, as propostas deste trabalho são:

Proposta 1: Um sistema de IA é ferramenta e, portanto, não é pessoa, não é sujeito de direitos e não titula direitos e deveres.

Proposta 2: O regime de responsabilização civil objetivo deve ser afastado e não deve ser aplicada a Teoria do Risco (Risked-based Approach) até a respectiva demonstração e comprovação dos possíveis danos a serem cometidos pelos sistemas de Inteligência Artificial.

---

[76] LANDES, E M. "Insurance, liability, and accidents: a theoretical and empirical investigation of the effect of no-fault accidents", *The Journal of Law and Economics*, 25 (1), 1982, p. 49-65 e RACHUM-TWAIG, O. "Whose Robot Is It Anyway? Liability for Artificial-Intelligence-Based Robots", *U. Ill. L. Rev*., 2020, p. 1141.

Proposta 3: O elemento culpa deve ser analisado e verificado sob a ótica do responsável pela utilização da IA enquanto ferramenta à consecução do dano.

No mesmo sentido, os resultados parciais podem ser resumidos em:

Resultado parcial 1: Diante das premissas postas, a responsabilidade deve ser afastada dessas entidades e atribuída a outrem. Esse outrem se torna o objeto e o lócus da argumentação.

Resultado parcial 2: O risco é composto por incerteza somado à probabilidade de dano.

Resultado parcial 3: os sistemas de IA são incertos e, portanto, preenchem o primeiro requisito para adoção da responsabilidade objetiva.

Resultado parcial 4: a ausência temporária de comprovação da probabilidade e dos impactos dos danos causados pelos sistemas de inteligência artificial afastam a aplicação da responsabilização objetiva.

Resultado parcial 5: A observância aos preceitos éticos devidamente comprovados e documentados por políticas de Governança Digital é fator de rompimento do nexo causal e de exoneração de responsabilidade (objetiva e subjetiva) pelo ato ilícito.

Resultado parcial 6: A falta de um mecanismo destinado à implementação do Human-In-Command é justificativa para aplicação da responsabilidade transobjetiva/vicária (por fato de coisa, de animal ou de outra pessoa).

Resultado parcial 7: É possível a aplicação da responsabilidade preventiva, desde que condicionada à uma exigência normativa e elaborada mediante Análise de Impacto Regulatório.

Tais resultados e propostas representam hipóteses e indicativos de aplicação pragmática de elementos éticos e jurídicos suficientes à tutela da relação jurídica envolvendo entes inteligentes artificialmente. Reconhece-se suas consideráveis limitações, inclusive teóricas. Contudo, acredita-se que esses indicativos podem ser elementos suficientes à partida e ao início de uma jornada regulatória que ainda está longe de se findar.

5. Referências

ADAMS, J. *Risk: the policy implications of risk compensation and plural rationalities*. London, UCL Press, 1995.

BERGENHEM, C, et al. "How to reach complete safety requirement refinement for autonomous vehicles", *CARS 2015-Critical Automotive applications*: Robustness & Safety. 2015. Disponível em: https://hal.archives-ouvertes.fr/hal-01190734/document. Acesso em: 10 dez. 2021.

BEUC. "Automated decision making and artificial intelligence: a consumer perspective", *Tech. rep., Bureau Europeen des Unions de Consommateurs*, 2018. Disponível em: https://www.beuc.eu/publications/beuc-x-2018-058_automated_decision_making_and_artificial_intelligence.pdf. Acesso em: 21 mar. 2020.

BEVILÁQUA, C*. Teoria Geral do Direito Civil*, Servanda, Campinas, 2015.

BODEN, Margaret. *Artificial Intelligence: a very short introduction*, Oxford University Press, Oxford, 2018.

BOSTROM, N. "*Ethical Issues in Advanced Artificial Intelligence"*. Disponível em: http://www.fhi.ox.ac.uk/wp-content/uploads/ethical-issues-in-advanced-ai.pdf. Acesso em: 09 abr. 2020.

BOSTROM, N. *Superintelligence*, Ed. Oxford University Press, London, 2014.

BOSTROM, N. "The ethics of artificial intelligence", (W. Ramsey & K. Frankish, orgs org.) *Draft for Cambridge Handbook of Artificial Intelligence,* Cambridge University Press, Cambridge, 2011. Disponível em: https://www.nickbostrom.com/ethics/artificial-intelligence.pdf. Acesso em: 09 abr. 2020.

BRAYNE, S. *Predict and Surveil*: *data discretion and the Future of Policing*, Oxford University Press, New York, 2021.

BRASIL. *Lei da Liberdade Econômica. Lei n. 13.874 de 20 de setembro de 2019.* Diário Oficial [da] República Federativa do Brasil, Poder Executivo, Brasília, DF, 20 set. 2019.

BURT, B. A. "Definitions of risk", *Journal of dental education*, 65(10), 2001, p. 1007-1008.

BURTON, S, et.al. "Mind the Gaps: Assuring the Safety of Autonomous Systems from an Engineering, Ethical, and Legal Perspective", *Artificial Intelligence*, 279, 2020, p. 4, 103201.

CHEN, Z. & LIU, B. *Lifelong Machine Learning*, Morgan & Claypool, Oregon, 2018.

CITRON, D. K. & PASQUALE, F. A. "The Scored Society: Due Process for Automated Predictions", *Washington Law Review*, 1, 2014, p. 2-27.

CLARKESWORLD. "Submissions Closed", *Twitter.* 2023. Disponível em: https://twitter.com/clarkesworld/status/1627711728245960704. Acesso em: 28 jul. 2023.

CRAWFORD, K. *Atlas of AI*, Yale University Press, London, 2021.

CURTI, F. et al. "Cyber risk definition and classification for financial risk management", *Journal of Operational Risk*, 18 (2), 2023, p. 3.

DIVINO, S. B. S. "After All, Artificial Intelligence is not Intelligent: in a Search for a Comprehensible Neuroscientific Definition of Intelligence", *Opinion Juridica*, 21, 2022, p. 1-21.

DIVINO, S. B. S. "Consciência e intencionalidade na Teoria do Fato Jurídico de Pontes de Miranda*:* Direito sem objetividade?", *Dissertação de Mestrado. Pontifícia Universidade Católica de Minas Gerais*, Minas Gerais, Belo Horizonte, 2019.

DIVINO, S. B. S. "Desigualdade codificada: como o uso de algoritmos pode reduzir, ocultar e aumentar a desigualdade?", em (C. Colombo & W. Engelmann & J. Faleiros Júnior, orgs.), *Tutela Jurídica do Corpo Eletrônico.* Foco, Indaiatuba, 2022, p. 723-738.

DIVINO, S. B. S. & MAGALHAES, R. A. "Inteligência Artificial e Direito Empresarial: Mecanismos de Governança Digital para Implementação e Confiabilidade", *Economic Analysis of Law Review*, 11, 2020, p. 72-89.

DIVINO, S. B. S. & MAGALHAES, R. "A. Propriedade intelectual e direito autoral de produção autônoma da inteligência artificial", *Revista de Direitos e Garantias Fundamentais*, 21, 2020, p. 167-192.

FERGUSON, A. G. *The Rise of Big Data Policing*: *surveillance, race, and the Future of Law Enforcement*, New York University Press, New York, 2017.

FLORIDI L. "Soft ethics, the governance of the digital and the General Data Protection Regulation", *Phil. Trans. R. Soc*. 2018, A 376: 20180081.

FOLHA. "Com que frequência acidentes ocorrem?" *Folha Online*. 2023. Disponível em: https://www1.folha.uol.com.br/folha/turismo/preparese/aviao-voar_e_seguro-02.shtml. Acesso em: 29 jun. 2023.

FREITAS, A. T. de. *Consolidação das Leis Civis*. Senado Federal, Brasília, 2003.

GIDDENS, A. *O mundo na era da globalização,* Presença, Lisboa, 2000.

GORNER, J. "Chicago Police Use 'Heat List' as Strategy to Prevent Violence", *Chicago Tribune.* 2013. Disponível em: http://articles.chicagotribune.com/2013-08-21/news/ct-met-heat-list-20130821_1_chicago-police-commander-andrew-papachristos-heat-list. Acesso em: 02 set. 2021.

HOOKER, J. & KIM, T. W. "Truly Autonomous Machines Are Ethical"*, AI Magazine*, 40 (4), 2019, p. 66-73.

JABŁONOWSKA, A. et., al. "Consumer law and artificial intelligence Challenges to the EU consumer law and policy stemming from the business' use of artificial intelligence", *Final report of the ARTSY Project*. European University Institute, 2018, p. 38.

KAPLAN, A. & HAENLEIN, M. "Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence", *Business horizons*, 62 (1), 2019, p. 15-25.

KAPLAN, S. & GARRICK, B. J. "On the quantitative definition of risk", *Risk analysis*, 1 (1), 1981, p. 11-27.

KLIPPENSTEIN, K. "Exclusive: surveillance footage of tesla crash on sf's bay bridge hours after elon musk announces "self-driving" feature", *The Intercept*, 2023. Disponível em: https://theintercept.com/2023/01/10/tesla-crash-footage-autopilot/. Acesso em: 27 jul. 2023.

KNIGHT, F. H. *Risk, uncertainty and profit*, Houghton Mifflin, Boston, 1921.

LA TIMES. "Norway cruise ship engines failed from lack of oil, maritime official says" *L.A Times*, 2019. Disponível em: https://www.latimes.com/world/la-fg-norway-cruise-ship-sky-20190327-story.html. Acesso em: 11 abr. 2020.

LANDES, E. M. "Insurance, liability, and accidents: a theoretical and empirical investigation of the effect of no-fault accidents", *The Journal of Law and Economics*, 25 (1), 1982, p. 49-65.

LIEBER, R. R. & ROMANO-LIEBER, N. S. "O conceito de risco: Janus reinventado", em (M.C.S. Minayo & A.C. Miranda, orgs.), *Saúde e ambiente sustentável:* estreitando nós [online], Ed. FIOCRUZ, Rio de Janeiro, 2002, p. 71-72.

MENDES, F. "Risco: um conceito do passado que colonizou o presente", *Revista Portuguesa de Saúde Pública,* 20 (2), 2002, p. 53-62.

MESKÓ, B. et al. "Will artificial intelligence solve the human resource crisis in healthcare?", *BMC Health Services Research*, 18, 2018, p. 545.

MOOR, J. H. "The nature, importance, and difficulty of machine ethics", *IEEE Intelligent Systems,* 21 (4), 2006, p. 18–21.

RACHUM-TWAIG, O. "Whose Robot Is It Anyway? Liability for Artificial-Intelligence-Based Robots", *U. Ill. L. Rev*., 2020, p. 1141.

RUSSELL, S. J. & NORVIG, P. *Artificial intelligence*: *a modern approach*, Pearson Education, New Jersey, 2010.

SATO, M. "AI-generated fiction is flooding literary magazines — but not fooling anyone", *The Verge*. 2023. Disponível em: https://www.theverge.com/2023/2/25/23613752/ai-generated-short-stories-literary-magazines-clarkesworld-science-fiction. Acesso em: 28 jul. 2023.

SEARLE, J. *Intentionality: An Essay in the Philosophy of Mind*, Cambridge University Press, Cambridge, 1983.

SIDDIQUI, F. & MERRILL, J. B. "17 fatalities, 736 crashes: The shocking toll of Tesla's Autopilot", *The Washington Post*, 2023. Disponível em: https://www.washingtonpost.com/technology/2023/06/10/tesla-autopilot-crashes-elon-musk/. Acesso em: 27 jul. 2023.

SMITH, J. "'Minority Report' Is Real — And It's Really Reporting Minorities", *Mic,* 2015. Disponível em: https://www.mic.com/articles/127739/minority-reports-predictive-policing-technology-is-really-reporting-minorities. Acesso em: 02 set. 2021.

SOUZA, K. R. G. & LOURENÇO, L. "A evolução do conceito de risco à luz das ciências naturais e sociais", *Territorium,* 22, 2015, p. 31-44.

STAFF, G. "US air force denies running simulation in which AI drone 'killed' operator", *The Guardian,* 2023. Disponível em: https://www.theguardian.com/us-news/2023/jun/01/us-military-drone-ai-killed-operator-simulated-test. Acesso em: 27 jun. 2023.

STONE, Z. "Everything You Need To Know About Sophia, The World's First Robot Citizen", *Forbes,* 2017. Disponível em: https://www.forbes.com/sites/zarastone/2017/11/07/everything-you-need-to-know-about-sophia-the-worlds-first-robot-citizen/?sh=453114bb46fa. Acesso em: 29 dez. 2020

SUTTON, R. S. & BARTO, A. G. *Reinforcement Learning*: *An Introduction,* MIT Press, Cambridge, 1998.

THE STRATEGY WEB. *Bosch Connected World 2020 – The human unicorn on AI,* 2020. Disponível em: https://thestrategyweb.

TURING, A. "On computable numbers, with an application to the Entscheidungsproblem", *Proceedings of the London Mathematical Society*, 42 1936, p. 230–65.

UN. "Living with risk: a global review of disaster reduction initiatives", Geneva, 2002

UNIÃO EUROPEIA. *"Artificial Intelligence: Europe needs to take a human-in-command approach, says EESC",* 2017. Disponível em: https://www.eesc.europa.eu/en/news-media/press-releases/artificial-intelligence-europe-needs-take-human-command-approach-says-eesc. Acesso em: Acesso em: 11 abr. 2020.

UNIÃO EUROPEIA. *Diretrizes éticas para uma IA de confiança (Ethics Guidelines For Trustworthy AI),* 2019. Disponível em: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai. Acesso em: 09 abr. 2020.

VLASSIS, N. *A Concise Introduction to Multiagent Systems and Distributed Artificial Intelligence: synthesis lectures on artificial intelligence and machine learning sequence in series,* Morgan & Claypool, Oregon, 2007.

VOROBEYCHIK, Y. & KANTARCIOGLU, M. *Adversarial Machine Learning: Synthesis Lectures on Artificial Intelligence and Machine Learning*, Morgan & Claypool, Oregon, 2018.