Economic Analysis of Law Review

Inteligência Artificial e Direito Empresarial: Mecanismos de Governança Digital para Implementação e Confiabilidade

Artificial Intelligence and Business Law: Digital Governance mechanisms for implementation and reliability

Sthéfano Bruno Santos Divino[1]
UFLA

Rodrigo Almeida Magalhães[2]
Puc-Minas

RESUMO

A utilização das Tecnologias da Comunicação e Informação nos processos empresariais toma pauta no cenário contemporâneo. Ocorre que a fragilidade das pessoas diante da insuficiência de conhecimento específico sobre esse ramo pode gerar certa desconfiança e interferir no desenvolvimento do próprio empresário. Assim, o presente artigo tem como problema de pesquisa o seguinte questionamento: como e quais os mecanismos de governança digital podem ser usados para implementação e confiabilidade de técnicas que utilizam inteligência artificial? Para a satisfação da problemática, utiliza-se como base a *Ethics Guidelines For Trustworthy AI*, de abril de 2019, elaborada pelo Grupo De Peritos De Alto Nível sobre Inteligência Artificial, criado pela Comissão Europeia em junho de 2018. Aborda-se em três seções as modalidades de governança corporativa: *human-in-the-loop*; *human-on-the-loop*; e *human-in-command*, indicando o seu funcionamento e sua contribuição para com o desenvolvimento das atividades empresariais e sua compatibilidade com o Direito. Utiliza-se a metodologia de pesquisa integrada e a técnica de pesquisa bibliográfica.

Palavras-chave: Direito; Governança Digital; Inteligência Artificial.

JEL: K0, K4, K40, K42

ABSTRACT

The use of Communication and Information Technologies in business processes is on the agenda of the contemporary scenario. It happens that the fragility of people in face of the insufficiency of specific knowledge about this branch can generate a certain distrust and interfere in the development of the entrepreneur himself. Thus, this article has as a research problem the following question: how and which digital governance mechanisms can be used for implementation and reliability of techniques that use artificial intelligence? To satisfy the problem, the Ethics Guidelines For Trustworthy AI, of April 2019, prepared by the High-Level Expert Group on Artificial Intelligence, created by the European Commission in June 2018, is used as a basis. It covers in three sections the modalities of corporate governance: human-in-the-loop; human-on-the-loop; and human-in-command, indicating its functioning and its contribution to the development of business activities and its compatibility with the Law. It uses the methodology of integrated research and the technique of bibliographic research.

Keywords: Law; Digital governance; Artificial intelligence.

R: 05/05/20 A: 02/06/20 P: 31/12/20

[1] E-mail: sthefanodivino@ufla.br
[2] E-mail: amagalhaes@ig.com.br

1. Introdução

Centrais de atendimento virtual com atendentes automatizados, veículos autônomos, fornecimento de produtos e de serviços conforme padrões do consumidor e verificação da probabilidade do risco de atividade e da tomada de decisão em processos automatizados estão entre as atividades que envolvem o uso de Inteligência Artificial nos âmbitos computacionais, jurídicos e econômicos na contemporaneidade.

A importância dessa tecnologia está assimilada ao desenvolvimento computacional capaz de transformar a vida da sociedade. Nesse ponto, o Direito tende a agir como ator a verificar como se pode realizar essa integração destes meios sem que haja danos aos particulares. Até o presente momento, recomendações e diretrizes ética que se assemelham a *soft law* estão sendo elaboradas para traçar parâmetros e enfrentar problemas concretos que a IA apresenta. O foco do presente trabalho, contudo, direciona-se ao aspecto interno da *atividade empresarial*[3], objetivando à proposição de técnicas que possam assumir a tendência regulatória sem intervenção estatal, para coadunar os interesses dos sujeitos envolvidos, sejam os cientistas da computação, sejam os empresários ou os consumidores. Um dos mecanismos jurídicos que se pode utilizar é a governança. Nesse cenário, ela é delimitada pelo seu aspecto digital.

Assim, o presente trabalho tem como problema de pesquisa o seguinte questionamento: como e quais os mecanismos de governança digital podem ser usados para implementação e confiabilidade de técnicas que utilizam inteligência artificial?

A apresentação da resposta a problemática instaurada se baseará na *Ethics Guidelines For Trustworthy AI*, de abril de 2019, elaborada pelo Grupo De Peritos De Alto Nível sobre Inteligência Artificial (GPAN IA), criado pela Comissão Europeia em junho de 2018. Trata-se de uma diretriz de caráter de *soft law* contendo indicativos éticos, normativos e jurídicos não vinculantes para a adoção de programas de cumprimento empresarial quando da utilização de tecnologias que se referem a entes inteligentes artificialmente.

Dentre as modalidades de governança digital, trabalha-se com três delas, que serão abordadas em seções distintas. A primeira será responsável pela abordagem do *human-in-the-loop* (HITL), prática em que a ação humana tem uma razoável interferência nos processos decisórios de IA. A segunda seção aborda a modalidade *human-on-the-loop* (HOTL), onde o *ser humano* tem como principal função a *supervisão*. As atuações ou interferências neste caso são raras em virtude de o processo de automatização possuir um certo grau de independência. Por fim, a terceira seção explicita o *human-in-command* (HIC). Neste caso a ação humana tem uma maior interferência em qualquer parte do processo decisório provocado por uma IA.

Pretende-se demonstrar o funcionamento dessas três modalidades e sua possível contribuição para com o desenvolvimento das atividades empresariais e sua compatibilidade com o Direito. Ao final, considera-se que a adoção de qualquer uma dessas modalidades será variável conforme disposição e interesse do empresário, conforme preceitos delimitados. Utiliza-se a metodologia de pesquisa integrada e a técnica de pesquisa bibliográfica

[3] A terminologia utilizada, neste caso, tem sua caracterização extraída de sua acepção funcional e não da caracterização objetiva ou subjetiva. Entende-se, portanto, a *empresa* nos termos do art. 966 do Código Civil de 2002, sendo a atividade econômica organizada para a produção ou a circulação de bens ou de serviços.

2. Diretrizes Éticas para uma IA de Confiança (*Ethics Guidelines For Trustworthy AI*)

O aparente crescimento da automação relacionada à Inteligência Artificial (IA)[4] pode ser vislumbrado como um dos grandes avanços científicos do século XXI. Bostrom[5], pressupõe-se 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.[6] A AGI tenta representar o conceito *original de inteligência*, traduzindo-se em algoritmos com desempenho equivalente ou superior ao do *ser humano* e são 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.[7] 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". [8]

No contexto tecnológico contemporâneo, detecta-se apenas a inserção da ANI na sociedade informacional. As diretrizes e os preceitos gerais para implementação da AGI e da ASI estão em aparente desenvolvimento através das técnicas de *Machine Learning*[9] (aprendizado de máquina) e *deep learning*[10] (aprendizado profundo). Estima-se de maneira muito otimista que a AGI estará disponível apenas em 2029 e que a ASI tornaria um evento singular em 2045.[11] Contudo, isso não reflete a maior parte dos cientistas, que tendem a crer que a AGI será alcançada apenas em torno de 2100, e a ASI após 30 anos de descoberta da AGI.[12]

Todo esse duvidoso cenário de predições e suposições fazem descrições utópicas e distópicas[13] nascerem e se tornarem uma das maiores preocupações dos sujeitos afetados pelo uso dessa tecnologia. A responsabilidade civil pelos ilícitos cometidos pelo uso de carros

[4] "We define AI as the study of agents that receive percepts from the environment and perform actions. Each such agent implements a function that maps percept sequences to actions, and we cover different ways to represent these functions, such as reactive agents, real-time planners, and decision-theoretic systems". NORVIG, Peter; RUSSELL, Stuart J. Artificial intelligence: a modern approach. New Jersey: Prentice Hall, 2010, p. VIII.

[5] BOSTROM, Nick. Superinteligência. Rio de Janeiro: Darkside, 2018.

[6] BOSTROM, Nick. 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**, 2018, 18, 545. Disponível em: https://doi.org/10.1186/s12913-018-3359-4. Acesso em: 09 abr. 2020.

[7] BOSTROM, Nick. The ethics of artificial intelligence. In RAMSEY, W.; FRANKISH, K. (org.) Draft for Cambridge Handbook of Artificial Intelligence. Cambridge University Press, 2011. Disponível em: https://www.nickbostrom.com/ethics/artificial-intelligence.pdf. Acesso em: 09 abr. 2020.

[8] BOSTROM, Nick. Superinteligência. Rio de Janeiro: Darkside, 2018, p. 55.

[9] GOLDBERG, D. E.; HOLLAND, J. H. *Genetic algorithms and machine learning. In Machine learning*. Vol. 3. Switzerland: 1988, p. 95-99.

[10] ČERKA, Paulius; GRIGIENĖ, Jurgita; SIRBIKYTĖ, Gintarė. Liability for damages caused by Artificial Intelligence. *Computer Law & Security Review*, Elsevier, v. 31, n. 3, p. 376-389, jun. 2015.

[11] REEDY, C. Kurzweil Claims That the Singularity Will Happen by 2045. Futurism. Disponível em: https://futurism.com/kurzweil-claims-that-the-singularity-will-happen-by-2045. Acesso em: 09 abr. 2020.

[12] BOSTROM, Nick. Superinteligência. Rio de Janeiro: Darkside, 2018, p. 50.

[13] Para mais, ver o capítulo 6 em LEE, Kai-Fu. Inteligência Artificial. Rio de Janeiro: Globo Livros, 2019, p. 169.

autônomos é um dos exemplos aplicáveis neste caso.[14] Outro exemplo é a eventual responsabilidade civil pela coleta e tratamento de dados oriundos de relações contratuais eletrônicas.[15]

Na medida em que o desenvolvimento da automação fica manifesto, existe uma certa preocupação dada pelos sistemas jurídicos acerca da regulamentação. Essa posição, contudo, pode se tornar problemática. Caso a regulamentação seja realizada de forma muito intensa o desenvolvimento científico pode ser prejudicado, pois sua implementação será demasiadamente demorada se comparada a um setor com regulamentos não tão rígidos. Assim, tanto os resultados quanto o avanço para a solução de eventuais problemas tendem a ser mais lentos. Lado outro, a simples abertura sem instrumentos regulatórios poderia fornecer mecanismos para instauração de um capitalismo predatório incapaz de satisfazer os objetivos inicialmente estipulados para satisfação e aprimoramento das convencionalidades propostas para serem insertas em sociedade. No mesmo sentido, o setor concorrencial também poderia se tornar um caos, tendo em vista a inexistência ou a insuficiência de normativos e preceitos éticos para regular a situação.

Foi com esse intuito que a Comissão Europeia elaborou, em 2019, um documento denominado Diretrizes éticas para uma IA de confiança (*Ethics Guidelines For Trustworthy AI*).[16] 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.[17]

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; II) Prevenção de danos; III) Equidade; e IV) Explicabilidade.[18]

O princípio da autonomia humana tem como objetivo a manutenção da autodeterminação[19] plena e efetiva sobre si próprios e participação no processo democrático quando da utilização de sistemas de IA para tal finalidade. Destaca o normativo europeu que:

---

[14] Mais em GOMES, Rodrigo Dias de Pinho. Carros autônomos e os desafios impostos pelo ordenamento jurídico: uma breve análise sobre a responsabilidade civil envolvendo veículos inteligentes. In FRAZÃO, Ana; MULHOLLAND, Caitlin. Inteligência Artificial e Direito. São Paulo: Thomson Reuters Brasil, 2019, p. 567-585.

[15] Para mais, ver em DIVINO, S. B. S. Reflexiones escépticas, principiológicas y económicas sobre el consentimiento necesario para la recolección y tratamiento de datos. Derecho PUCP, 83, 179 - 206, 29 nov. 2019. Disponível em: <https://doi.org/10.18800/derechopucp.201902.006>. Acesso em: 09. abr. 2020.

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

[17] Ibidem, p. 6.

[18] Ibidem, p. 14.

[19] 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 the

> Os sistemas de IA não devem subordinar, coagir, enganar, manipular, condicionar ou arregimentar injustificadamente os seres humanos. Em vez disso, devem ser concebidos para aumentar, complementar e capacitar as competências cognitivas, sociais e culturais dos seres humanas.[20]

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".[21] Deve-se garantir *solidez* na adoção dessas técnicas para que os resultados sejam prolíficos para todos os sujeitos envoltos na relação em análise.

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".[22] Isso conduz à adoção de procedimentos que evitem enviesamentos injustos, discriminação e estigmatização contra pessoas e grupos.[23]

Por fim, a explicabilidade:

> 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.[24]

Os normativos contemporâneos que desenvolvem a proteção de dados já têm adotado o referido princípio no corpo de seu texto legal. No RGPD ela está prescrita no art. 22[25]; na LGPD, sua prescrição está no art. 20[26].

---

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

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, n. 1, 2014, p. 2-27. Disponível em: https://digitalcommons.law.umaryland.edu/fac_pubs/1431/. Acesso em 21 mar. 2020 e 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.

[20] UNIÃO EUROPEIA. Diretrizes éticas para uma IA de confiança (Ethics Guidelines For Trustworthy AI). P. 15. Disponível em: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai Acesso em: 09 abr. 2020.

[21] Idem.

[22] Idem.

[23] Idem.

[24] Idem.

[25] Artigo 22. Decisões individuais automatizadas, incluindo definição de perfis 1. O titular dos dados tem o direito de não ficar sujeito a nenhuma decisão tomada exclusivamente com base no tratamento automatizado, incluindo a definição de perfis, que produza efeitos na sua esfera jurídica ou que o afete significativamente de forma similar. UNIÃO EUROPEIA. Regulamento 2016/679 do Parlamento Europeu e do Conselho. 2016. Disponível em: http://eur-lex.europa.eu/legal-content/PT/TXT/?uri=celex%3A32016R0679 >. Acesso em: 21 mar. 2020.

[26] Art. 20. O titular dos dados tem direito a solicitar a revisão de decisões tomadas unicamente com base em tratamento automatizado de dados pessoais que afetem seus interesses, incluídas as decisões destinadas a definir o seu perfil pessoal, profissional, de consumo e de crédito ou os aspectos de sua personalidade. BRASIL. Lei. 13.709,

Tais exigências ficam em um plano ideal. Contudo, como concretizá-las? Como tornar possível a utilização de uma IA de confiança? Como o nível de cognição de uma IA está de forma pressuposta 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[27], "a governança digital é a prática de estabelecer e implementar políticas, procedimentos e padrões para o desenvolvimento, uso e gerenciamento adequados da *infosphere*":

> 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*.[28]

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.[29] Cada um deles será abordado de forma detalhada neste momento.

3. *Human-in-the-loop* (HITL)

E se em vez de pensar na automação como ausência do envolvimento humano de uma tarefa a imaginarmos como a inclusão seletiva da participação humana? O resultado seria um processo que aproveita a eficiência da automação inteligente, permanecendo passível de feedback humano, mantendo ao mesmo tempo um maior senso de significado.[30] A abordagem *human-in-the-loop* objetiva reformular a automação dos entes inteligentes artificialmente ao estabelecer uma

---

de 14 de agosto de 2018. Diário Oficial da República Federativa do Brasil. Brasília, DF: 14 ago. 2018. Disponível em: <http://www.planalto.gov.br/cci-vil_03/_ato2015-2018/2018/lei/L13709.htm>. Acesso em: 21 mar. 2020.

[27] "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. Disponível em: http://dx.doi.org/10.1098/rsta.2018.0081. Acesso em: 09. abr. 2020.

[28] Idem.

[29] FRAZÃO, Ana. Responsabilidade Civil de administradores d esociedades empresárias por decisões tomadas com base em sistemas de inteligência artificial. in. FRAZÃO, Ana; MULHOLLAND, Caitlin. Inteligência Artificial e Direito. São Paulo: Thomson Reuters Brasil, 2019, p. 514.

[30] WANG, Ge. Humans in the Loop: The Design of Interactive AI Systems. Human-centered Artificial Intelligence. Stanford University. 2019. Disponível em: https://hai.stanford.edu/news/humans-loop-design-interactive-ai-systems. Acesso em: 10 abr. 2020.

relação *Human-Computer Interaction* (HCI), para descobrir como podemos incorporar a interação humana de forma mais útil, segura e significativa no sistema de IA.

O HITL (*Human-in-the-loop*) se concentra na criação de fluxos de trabalho em que a IA aprende com o operador humano, enquanto intuitivamente torna o trabalho do ser humano mais eficiente. A máquina executa uma ação, solicita informações a um especialista humano e aprende com a resposta que recebe. Idealmente, o processo de interação não apenas torna o trabalho do especialista humano mais eficiente, mas também captura a inteligência combinada de todo especialista que interage com o sistema. Dessa forma, todo o conhecimento tácito dos especialistas humanos pode se tornar parte do mesmo sistema compartilhado.[31]

O funcionamento do HITL se dá pela combinação de duas técnicas da ciência da computação: o *Supervised Machine Learning* (SML) e o *Active Learning* (AL). No primeiro, especialistas em ML utilizam conjuntos específicos de dados para treinar algoritmos, para ajustar parâmetros, com a finalidade de fazer previsões precisas para os dados recebidos. No AL, os dados são coletados, treinados, ajustados, testados e mais dados são inseridos no algoritmo para torná-lo mais inteligente, mais confiante e mais preciso. Essa abordagem - especialmente alimentar dados de volta para um classificador - é chamada de aprendizado ativo.[32]

Mas, quando aplicar o HITL? 1) Quando os custos transação dos erros de uma IA é alto. Um algoritmo ML não pode ter absolutamente nenhuma margem para erro. Qualquer espaço para erro pode levar a situações difíceis para o empresário. 2) Quando o aspecto semântico é preponderante para a interpretação do que está sendo analisado, determinados léxicos como *boa-fé* apresentam uma carga interpretativa que até o momento é de difícil interpretação por IA. 3) Quando há poucos dados disponíveis no momento. Por exemplo, a classificação de postagens de mídia social realizadas por uma máquina, para um novo negócio em seus estágios iniciais, pode não ser uma opção viável devido à escassez de dados. Os seres humanos poderão ser melhores nos estágios iniciais, mas, com o tempo, as máquinas podem aprender e assumir a tarefa.[33]

Quais os benefícios[34] da implementação do HITL? Em primeiro lugar, o processo de decisão tende a se tornar mais transparente. Cada etapa que incorpora a interação humana exige que o sistema seja projetado para ser entendido pelos seres humanos para executar a próxima ação. Além disso, exige-se que exista alguma interferência humana na determinação de etapas críticas. Por fim, humanos e IA realizam a tarefa juntos, tornando cada vez mais difícil o processo permanecer oculto.[35]

---

[31] HULKKO, Ville. Most value from ai with human-in-the-loop solutions. Silo.AI. 2018. Disponível em: https://silo.ai/most-value-human-in-the-loop-ai/. Acesso em: 10 abr. 2020.

[32] MOTHI. What is Human-in-the-Loop for Machine Learning? Hackernoon. 2018. Disponível em: https://hackernoon.com/what-is-human-in-the-loop-for-machine-learning-2c2152b6dfbb. Acesso em: 10 abr. 2020.

[33] Idem.

[34] "We also need to consider the characteristics about the downsides for the Human In-The-Loop: Humans can make bad choices due to not thinking things through; Humans can make bad choices due to emotional clouding; Humans can slow down a process by taking too long to take an action; Humans can make errors in the actions they take Humans can be disrupted in the midst of taking actions; Humans can freeze-up and fail to take action when needed; Etc.". ELIOT, Lance. Human In-The-Loop Vs. Out-of-The-Loop in AI Systems: The Case of AI Self-Driving Cars. 2019. Disponível em: https://www.aitrends.com/ai-insider/human-in-the-loop-vs-out-of-the-loop-in-ai-systems-the-case-of-ai-self-driving-cars/. Acesso em: 11 abr. 2020.

[35] WANG, Ge. Humans in the Loop: The Design of Interactive AI Systems. Human-centered Artificial Intelligence. Stanford University. 2019. Disponível em: https://hai.stanford.edu/news/humans-loop-design-interactive-ai-systems. Acesso em: 10 abr. 2020.

Em segundo lugar, o HITL pode incorporar as inferências humanas de maneira eficaz. Como os sistemas de IA são moldados para ajudar os seres humanos, o valor de tais sistemas não reside apenas na eficiência ou correção, mas também na preferência e na agência humanas.[36]

Por fim, eles permitem a criação de sistemas mais precisos e poderosos. Embora eles desviam a pressão da criação de algoritmos *perfeitos* (o que não existe), as estratégias de interferência humana no HITL muitas vezes podem melhorar o desempenho do sistema em comparação com sistemas totalmente automatizados e totalmente manuais. Cria-se um sistema híbrido de excelência funcional que pode ser alcançado através da busca de um equilíbrio ideal entre homem-máquina.[37]

Como o HITL se dá na prática, especialmente no setor jurídico? Um exemplo de uma IA HITL que aumenta os fluxos de trabalho existentes pode ser obtido no mundo do jurídico. Profissionais jurídicos navegam por centenas de páginas de documentos diariamente procurando indicadores de possíveis riscos, sejam eles contratuais ou sejam eles possibilidades de provimento ou não de sentenças judiciais.[38] Embora a detecção de certos tipos de risco possa ser treinada para uma IA usando vários métodos de processamento de linguagem natural, o processo ainda depende muito da experiência humana de nível sênior para interpretar o texto. Um erro no reconhecimento dos riscos pode levar a custos caros em danos aos negócios. [39] Essa visualização pode se tornar mais fácil quando da verificação, por exemplo, a interpretação dos resultados advém de uma linguagem jurídica com vários significados, tal como o inadimplemento contratual com fundamento na exceção de contrato não cumprido, ou inclusive na própria violação positiva do contrato, que tem como fundamento os deveres anexos da boa-fé objetiva. Pode se tornar difícil para uma inteligência artificial que utiliza a linguagem natural interpretar o significado dessas terminologias genéricas, que normalmente possuem seu conteúdo preenchido conforme o caso analisado. A boa-fé objetiva é, neste caso, talvez a mais afetada. Como compreender os deveres de proteção e cuidado, informação e esclarecimento, e lealdade e probidade através de uma linguagem quase totalmente objetiva em seu contexto? Isso pode ser difícil sem ajuda de um agente humano. Além disso, ainda que exista um agente humano, caso ele não domine os conceitos em análise de nada adiantará, pois são oriundos de áreas específicas para sua compreensão.

Após o pré-treinamento da IA, o sistema tende a produzir sugestões de sentenças indicadoras de riscos nos documentos, formando resumos de riscos de documentos estendidos. O ser humano analisará os resultados, fornecerá o feedback para a IA em cada sugestão. Esse

[36] Idem

[37] Idem.

[38] Manifesta-se ciência de que a profissão do operador do Direito não tem seu cerne na identificação de riscos, mas na aplicação de normas ou de regimes jurídicos por meio de negócios jurídicos ou outros institutos de direito que regulam a vida em sociedade. Ocorre que a prática da análise estatística pode ser utilizada como fator objetivo de verificação do sucesso ou insucesso de uma demanda conforme assunto e parâmetros previamente estabelecidos. Por exemplo, o ingresso de uma demanda que vai contra uma súmula vinculante ou outro preceito de veiculação obrigatória cujo mandamento constituição possui proibição expressa, tem sua chance de 100% de ser negado em virtude da solidificação desse posicionamento no ordenamento jurídico. Contudo, assuntos como "quantificação do dano moral em virtude do ato ilícito X ou Y" pode ser verificada através de padrões extraídos dos próprios tribunais de justiça, inclusive os superiores. Nessa pauta, podemos elencar que um dos critérios para o arbítrio do *quantum* reparatório moral em sede do STJ é o método bifásico, que retira tem em sua primeira fase uma consideração objetiva do montante atribuído para aquele caso em análise conforme julgados prévios e (im)procedentes. Dessa forma, possibilita-se ao operador do direito oferecer informações mais objetivas com fundamento na própria jurisprudência. É justamente esse ponto que pretende-se evidenciar.

[39] HULKKO, Ville. Most value from ai with human-in-the-loop solutions. Silo.AI. 2018. Disponível em: https://silo.ai/most-value-human-in-the-loop-ai/. Acesso em: 10 abr. 2020.

feedback ensinará ainda mais a IA a melhorar seu desempenho. Uma vez que departamentos jurídicos inteiros operam com o mesmo sistema de aprendizado, ele começa a capturar rapidamente a inteligência coletiva de todo o departamento. Graças à supervisão humana, a mesma qualidade de revisão de documento é mantida como antes, apenas executada mais rapidamente a cada rodada de feedback.[40]

Outro exemplo de HITL comumente utilizado na práxis é a intervenção de um agente humano na solução de eventuais controversas pleiteadas por consumidores em plataformas digitais. Quando uma pessoa abre um ticket pleiteando a devolução dos valores pagos ou mesmo a troca de determinado produto, nos termos da legislação vigente, verifica-se a existência de uma entidade inteligente artificialmente que media o caminho entre o consumidor e a resposta ao seu pleito inicial. Contudo, quando a IA é incapaz de trazer as informações necessárias, torna-se essencial a ação humana para que as atividades complementares sejam realizadas de forma satisfatória. Assim, caso o consumidor, por exemplo, tem como objetivo verificar qual é o código de rastreio de sua encomenda, a IA pode cedê-lo de forma direta, sem a ação humana. Contudo, caso esse dado não esteja disponível ou atualizado no sistema da empresa, poderá um agente humano interferir para que tal conduta seja realizada. Após isso, a IA adquirirá a informação e poderá cedê-la ao consumidor caso esse venha requerê-la novamente.

A abordagem HITL, portanto, pode ser aplicada a quase todos os processos negociais e empresariais em que especialistas humanos aplicam seu conhecimento interno para executar uma tarefa. Pode variar de atendimento ao cliente (automação de resposta revisada), à medicina[41] (revisão de imagem de raios-X), fabricação (controle visual de qualidade) e finanças (detecção de fraude).[42]

Lado outro, a abordagem *Human-on-the-loop* (HOTL) funciona apenas com o monitoramento do *ser humano* através de um sistema. A princípio, não se tem a intervenção do *agente* nas práticas que envolvem os atos de IA. E será nesse momento que essa modalidade será abordada.

4. *Human-on-the-loop* (HOTL)

O agir de um *ser humano* quando da supervisão das práticas exercidas por um ente inteligente autônomo através da interação intermitente do operador humano com um sistema

[40] Idem.

[41] Além disso, o sistema também pode ser usado para educar funcionários de nível júnior - vamos supor que um sistema de detecção de condição pulmonar baseado em raios-X seja treinado inicialmente apenas com especialistas de nível sênior. O sistema pode criar sugestões para médicos de nível júnior que trabalham com imagens de raios-X, apontando as anomalias sugeridas que podem ter sido perdidas sem a experiência anterior. Mesmo assim, o médico júnior mantém o poder de aceitar ou rejeitar a sugestão, usando sua própria intuição especializada para fazer a chamada final. HULKKO, Ville. Most value from ai with human-in-the-loop solutions. Silo.AI. 2018. Disponível em: https://silo.ai/most-value-human-in-the-loop-ai/. Acesso em: 10 abr. 2020

[42] HULKKO, Ville. Most value from ai with human-in-the-loop solutions. Silo.AI. 2018. Disponível em: https://silo.ai/most-value-human-in-the-loop-ai/. Acesso em: 10 abr. 2020.

remoto e automatizado, com a finalidade de gerenciar um processo controlado ou um ambiente específico de tarefas, é denominado *Human-on-the-Loop* (HOTL)[43].[44]

Dentre os exemplos de HOTL destacam-se: configurações de controle de supervisão humana que incluem controle de tráfego aéreo, comando e controle militar e espacial, e gerenciamento de resposta a crises e operações de veículos autônomos. Com a rápida expansão da tecnologia automatizada nas configurações cotidianas, as configurações de controle de supervisão humana estão se expandindo para medicina, direção e aplicações comerciais e comerciais.[45]

O HOTL pode ser traçado como um campo interdisciplinar em que a ação humana envolverá vários campos da ciência, tais como: 1) a psicologia quando da tomada de decisão em situações críticas em sistemas de alto risco e com alto nível pressão, se levado em consideração o fator limitante no sucesso e probabilidade de erros do sistema; 2) ciência da computação, especificamente o design de algoritmos (e automação resultante), bem como as interfaces que se comunicam com o operador (incluindo aural, visual e háptico); e 3) a engenharia do sistema que executa a tarefa (para compreender como os sistemas de IA em utilização podem impactar negativamente a compreensão e a execução humana de um mecanismo de controle).[46]

A diferença entre o HITL e o HOTL reside na capacidade decisória. Enquanto no primeiro o papel do *agente humano* se volta de forma mais destacada para o aspecto decisório, no HOTL a função humana tem menor ação, sendo melhor descrita apenas como um *supervisor* da tecnologia de IA.[47] Dessa forma, a interferência humana somente se daria nas hipóteses em que a IA seria totalmente incapaz de solucionar a situação problema que lhe fora posta, ou quando os riscos da atividade delegados para a IA são tão grandes que se tomadas de forma automatizadas poderiam causar um prejuízo incalculável para as partes envoltas no sinistro.

A aplicabilidade prática do HOTL pode se verificar na utilização de carros autônomos. A recomendação J3016 elaborada pela *Society of Automotive Engineers* propõe níveis de automação para carros que utilizam o sistema de IA operar, onde 0 é o nível com automação quase indetectável, comumente utilizada em veículos convencionais, e 5 são aqueles dotadas de capacidades irrestritas de direção autônoma.[48] Segundo Lancelot, "estamos a anos e anos longe de alcançar um carro autônomo de nível 5, que é essencialmente um carro autônomo que pode dirigir da maneira que um humano possa dirigir"[49]. O atual estado da arte nos permite encontrar veículos entre o nível 3, com modelos dotados de automação condicional, em que o sistema permite que se ceda ao condutor o controle das funções veiculares em situações críticas, e o nível

[43] Existe uma adoção de léxicos diversos, mas que na prática representam a mesma situação: *Human-on* ou *out-the-Loop*.

[44] COMMINGS, Mary. Supervising automation: humans on the loop. Aero-Astro Magazine Highlight: MIT Department of Aeronautics and Astronautics. 2008. Disponível em: http://web.mit.edu/aeroastro/news/magazine/aeroastro5/cummings.html. Acesso em: 10 abr. 2020.

[45] Idem.

[46] Idem.

[47] "A variation of in-the-loop is "on-the-loop," in which the role of the human coordinator is less involved, perhaps best described as that of a supervisor, rather than a deciding authority. Our work studies such interactional arrangements with the goal to enable efficient interaction and collaboration between humans and agents". FISCHER, Joel E. et.al. In-the-loop or on-the-loop? Interactional arrangements to support team coordination with a planning agent. Concurrency Computat: Pract Exper. 2017;e4082. https://doi.org/10.1002/cpe.4082.

[48] SAE. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. 2018. Disponível em: https://www.sae.org/standards/content/j3016_201806/ . Acesso em: 11 abr. 2020.

[49] LANCELOT, Eliot. Self-Driving Car is the Mother of All AI Projects: Why It is a Moonshot. https://www.aitrends.com/selfdrivingcars/self-driving-car-mother-ai-projects-moonshot/. Acesso em: 11 abr. 2020.

4, onde podem operar sem supervisão humana, mas somente em condições específicas delimitadas pelo fator geográfico ou pela estrada.[50] Este último caso pode ser exemplificado com o caso da *startup* Otto, controlada pela empresa Uber, que utilizou um caminhão sem interferências humanas para realizar uma entrega de 50 mil latas de cerveja no seu destino.[51]

A situação pode se complicar com a utilização do HOTL quando o sistema possui uma maior autonomia frente o aspecto decisório humano. A IA que adota essa modalidade trabalha sob o aspecto estatístico de probabilidades para conceder uma melhor resposta à situação que lhe é posta. Caso ela esteja em uma situação crítica e a conduta humana persista na decisão de continuar, a decisão da IA poderá se sobrepor a decisão humana e fazer o que for melhor para aquele momento. Caso que demonstra a aplicabilidade desse conflito se deu com o cruzeiro Viking Sky, em março de 2019.

O motor do cruzeiro 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 sua vez, logo depois, causou um desligamento automático dos motores.[52]

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.

A adoção do HOTL, portanto, pressupõe um alto nível de automação do sistema de IA envolto nas relações empresariais. A ação humana que tem como principal finalidade a supervisão deve ser utilizada para tentar amenizar eventuais prejuízos. Contudo, se o nível decisório do próprio sistema tiver prioridade sobre a conduta humana, tal como o caso do Viking Sky, a esfera de autonomia pode ficar prejudicada. Assim, com o objetivo de atender os princípios éticos da autonomia humana prescrito pela diretriz europeia, quando houver o design de IA com preceitos de HOTL deve existir algum mecanismo que transfira para o ser humano a capacidade de escolha para atuar naquela situação específica. Isso, pois, o próprio sistema pode ser falho e a experiência do controlador humano pode vir a ser um fator preponderante de subjetividade naquele momento. Claro que, em ambos os casos, a responsabilidade por eventuais danos acometidos ao ambiente ou a pessoas serão atribuídos ao responsável pela tecnologia e pelo bem móvel que a

[50] GOMES, Rodrigo Dias de Pinho. Carros autônomos e os desafios impostos pelo ordenamento jurídico: uma breve análise sobre a responsabilidade civil envolvendo veículos inteligentes. In FRAZÃO, Ana; MULHOLLAND, Caitlin. Inteligência Artificial e Direito. São Paulo: Thomson Reuters Brasil, 2019, p. 567-585

[51] DAVIES, A. Uber's Self-Driving Truck Makes Its First Delivery: 50,000 Beers. Wired. 2016. Disponível em: https://www.wired.com/2016/10/ubers-self-driving-truck-makes-first-delivery-50000-beers/ Acesso em: 11 abr. 2020.

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

utiliza, sob o prisma do risco do desenvolvimento. Isso, contudo, não retira a possibilidade de redução dessa automação para atendimento das diretrizes éticas então estipuladas.

O terceiro e último mecanismo de governança digital para implementação na seara empresarial que tem como objetivo a implementação de tecnologias de inteligência artificial em seu ramo de atividade é o *human-in-command* (HIC). Neste caso, além de supervisionar a atividade do sistema de IA, atribui-se a possibilidade de decisão e de utilização do sistema em qualquer situação específica. Portanto, a autonomia humana, neste caso, tem um fator preponderante perante os mecanismos automatizados. E será ele abordado neste momento.

5. *Human-in-command* (HIC)[53]

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.[54] 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".[55] 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".[56]

O EESS assinala 11 esferas nas quais a IA suscita atualmente desafios sociais: ética; segurança; privacidade; transparência e explicabilidade; trabalho; educação e competências; (des)igualdade e inclusividade; legislação e regulamentação; governo e democracia; guerra; e superinteligência.[57] Por esse motivo, alguns questionamentos devem ser realizados para compatibilizar o grau de automação com eventuais interferências humanas no processo em análise, tais como: qual a influência da IA autônoma (com capacidade de autoaprendizagem) na nossa integridade pessoal, autonomia, dignidade, independência, igualdade, segurança e liberdade de escolha? Como se garante que os nossos valores, normas e direitos humanos fundamentais são respeitados e salvaguardados?

---

53 "Human-in-command (HIC): refers to the capability to oversee the overall activity of the AI system (including its broader economic, societal, legal and ethical impact) and the ability to decide when and how to use the system in any particular situation". UNIÃO EUROPEIA. Diretrizes éticas para uma IA de confiança (Ethics Guidelines For Trustworthy AI). Disponível em: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai Acesso em: 09 abr. 2020.

54 THE STRATEGY WEB. Bosch Connected World 2020 – The human unicorn on AI. Disponível em: https://thestrategyweb.com/2020/02/20/bosch-connected-world-2020-the-human-unicorn-on-ai/. Acesso em: Acesso em: 11 abr. 2020.

55 "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. 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.

56 "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". Idem.

57 UNIÃO EUROPEIA. Artificial Intelligence - The consequences of artificial intelligence on the (digital) single market, production, consumption, employment and society (own-initiative opinion). 2016. Disponível em: https://www.eesc.europa.eu/en/our-work/opinions-information-reports/opinions/artificial-intelligence-consequences-artificial-intelligence-digital-single-market-production-consumption-employment-and. Acesso em 11 abr. 2020.

Tais questionamentos devem ser realizados com intuito de prevenção e diminuição das atividades tecnológicas para eventuais danos acometidos a seres humanos. Entende-se os sistemas de IA, se utilizados de forma errônea, podem afetar negativamente os direitos fundamentais. Um dos exemplos pode ser a reprodução de condutas discriminatórias infelizmente presente na linguagem natural. Como a máquina realiza uma análise objetiva dos dados que lhe foram oferecidos e, a princípio não se faz uma *filtragem de conteúdo de forma direta*, os algoritmos podem vir replicar o comportamento discriminatório presente naquelas decisões. Além disso, uma IA pode não entender o limite de uma esfera privada, por exemplo, adentrado na intimidade e tendo acesso a dados sensíveis ou não que dizem respeito unicamente ao seu titular. Nas situações em que existe esse risco, deve ser realizada uma avaliação do impacto dos direitos fundamentais, incluindo se esses riscos podem ser reduzidos para manutenção de uma relação democrática, a fim de respeitar os direitos e liberdades de terceiros.

No mais, os sistemas de IA podem ser implantados para moldar e influenciar o comportamento humano por meio de mecanismos que podem ser difíceis de detectar, o que pode ameaçar a autonomia individual. O princípio geral da autonomia do usuário deve ser central para a funcionalidade do sistema. Assim, supervisão humana pode ajudar a garantir que um sistema de IA não comprometa sua *autonomia* nem cause outros efeitos adversos. Ao usar a IA, deve-se garantir que os agentes públicos tenham a capacidade de exercer a supervisão de acordo com suas ações e escolhas.[58]

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*.

Um aspecto problemático da adoção do HIC é a pausa abrupta e a mudança de controle para o agente humano. Quando o sistema está funcionando sob a ótica do *deep* ou do *machine learning*, qualquer ação voltada ao exercício da atividade desenvolvida é utilizada pela IA para o aprendizado e futuras correções de erros. As pausas abruptas infelizmente contam nesse processo e interferem no aprendizado da máquina, dificultando que ela verifique a situação e trabalhe da melhor forma estatística possível.

Esse problema complexo está sendo resolvido pelos pesquisadores da *Ecole Polytechnique Fédérale de Lausanne* através de um método denominado *safe interruptibility* (interrupção segura). Tal método permite que os agentes humanos interrompam o processo de aprendizado da IA quando necessário sem afetar diretamente o desempenho do *machine learning* e do *deep learning*. Os pesquisadores alteram o sistema de aprendizado da máquina para que ele não seja afetado pelas interrupções. Assim, "*we worked on existing algorithms and showed that safe interruptibility can work no*

[58] EIT HEALTH. AI and Ethics in the Health Innovation Community. 2020. Disponível em: https://eithealth.eu/wp-content/uploads/2020/01/AI-and-Ethics-in-the-Health-Innovation-Community.pdf. Acesso em: 11 abr. 2020.

matter how complicated the AI system is, the number of robots involved, or the type of interruption. We could use it with the Terminator and still have the same results".[59]

Contudo, a abordagem descrita torna-se eficiente apenas quando as interferências humanas representam mudanças bem pequenas nos resultados alterados por eventuais erros que seriam cometidos pela IA. Se se detecta uma maior probabilidade do acontecimento e há a interrupção humana, isso pode prejudicar o desenvolvimento da IA, que aprenderia com aquele eventual equívoco para não o repetir futuramente.

Dessa forma, a adoção do HIC tende a solucionar o problema da autonomia de forma aparentemente completa, mas pode prejudicar o desenvolvimento científico por interferências indesejadas no *machine learning* e no *deep leaning* durante o funcionamento da IA.

Portanto, a adoção desses modelos de Governança Digital dependerá basicamente da escolha do empresário levando em consideração a disponibilidade da tecnologia, os custos de transação atrelados a ela, e 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.

Assim, a adoção do HITL pressupõe um investimento mediano em tecnologia de IA e a assunção moderada de riscos, já que a participação de agentes humanos entrará em pauta quando da decisão dos entes inteligentes artificialmente. Lado outro, o HOTL pressupõe um investimento considerável em tecnologias de IA e a alta assunção de risco voltada a atividade, já que o sistema em si será designado para tomar as decisões de forma automatizada e o agente humano atuará apenas como supervisor. E, por fim, o HIC possui um razoável custo de transação voltado à esfera tecnológica, já que as eventuais decisões poderão ser tomadas por agentes humanos a qualquer tempo. Assim, a assunção de risco é variável conforme a atuação humana no momento em que eventual sinistro vier a acontecer.

Como se observa, as práticas acima mencionadas servem como guia para instauração de processos tecnológicos objetivando atendimento de exigências éticas e sua compatibilização entre o desenvolvimento científico e tecnológico e a intrínseca relação de lucro empresarial. Compete ao empresário, neste caso, detectar suas necessidades e orçamento e optar pela melhor escolha que o atenderá naquele momento e também ao atendimento de preceitos éticos insculpidos pelos órgãos internacionais.

6. Considerações Finais

O propósito do presente trabalho foi apresentar como e quais os mecanismos de governança digital podem ser usados para implementação e confiabilidade de técnicas que utilizam inteligência artificial. Três desses mecanismos foram apresentados e explicitados. As considerações a seguir são apenas delineamentos para futuros trabalhos e que, de forma alguma, representa *conclusões* sobre a abordagem temática, que está longe de ser solidificada.

[59] MAURER, Alexandre. How can humans keep the upper hand on artificial intelligence? ScienceDaily. 2017. Disponível em: https://www.sciencedaily.com/releases/2017/12/171204094950.htm. Acesso em: 11 abr. 2020.

1) A abordagem H*uman-in-the-Loop* (HITL) objetiva reformular a automação dos entes inteligentes artificialmente ao estabelecer uma relação *Human-Computer Interaction* (HCI), para descobrir como se incorpora a interação humana de forma mais útil, segura e significativa no sistema de IA.
2) No HITL, o processo de decisões automatizadas de uma IA pode sofrer inferências diretas da ação humana. Assim, a adoção do HITL pressupõe um investimento mediano em tecnologia de IA e a assunção moderada de riscos, já que a participação de agentes humanos entrará em pauta quando da decisão dos entes inteligentes artificialmente.
3) O *Human-on-the-loop* (HOTL) pode ser definido como o agir de um ser humano quando da supervisão das práticas exercidas por um ente inteligente autônomo através da interação intermitente do operador humano com um sistema remoto e automatizado, com a finalidade de gerenciar um processo controlado ou um ambiente específico de tarefas.
4) A adoção do HOTL do o HOTL pressupõe um investimento considerável em tecnologias de IA e a alta assunção de risco voltada a atividade, já que o sistema em si será designado para tomar as decisões de forma automatizada e o agente humano atuará apenas como supervisor.
5) O *Human-in-Command* (HIC) refere-se à capacidade de supervisionar a atividade geral do sistema de IA (incluindo seu impacto econômico, social, jurídico e ético mais amplo) e a capacidade de decidir quando e como usar o sistema em qualquer local específico situação.
6) Pelo HIC possuir um razoável custo de transação voltado à esfera tecnológica, já que as eventuais decisões poderão ser tomadas por agentes humanos a qualquer tempo, a assunção de risco é variável conforme a atuação humana no momento em que eventual sinistro vier a acontecer.
7) Em qualquer das modalidades de Governança Digital acima descritas, a observância dos princípios éticos da autonomia, prevenção de danos, equidade, explicabilidade deve ser realizada para tornar possível a compatibilização do desenvolvimento tecnológico com as premissas econômicas e a tutela jurídica de direitos individuais.

7. Referências

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.

BRASIL. **Lei. 13.709, de 14 de agosto de 2018**. Diário Oficial da República Federativa do Brasil. Brasília, DF: 14 ago. 2018. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm>. Acesso em: 21 mar. 2020.

BOSTROM, Nick. **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, Nick. **Superinteligência**. Rio de Janeiro: Darkside, 2018.

BOSTROM, Nick. The ethics of artificial intelligence. In RAMSEY, W.; FRANKISH, K. (org.) **Draft for Cambridge Handbook of Artificial Intelligence**. Cambridge University Press, 2011. Disponível em: https://www.nickbostrom.com/ethics/artificial-intelligence.pdf. Acesso em: 09 abr. 2020.

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

CITRON, D. K.; PASQUALE, F. A. **The Scored Society: Due Process for Automated Predictions**. Washington Law Review, n. 1, 2014, p. 2-27. Disponível em: https://digitalcommons.law.umaryland.edu/fac_pubs/1431/. Acesso em 21 mar. 2020.

COMMINGS, Mary. **Supervising automation: humans on the loop**. Aero-Astro Magazine Highlight: MIT Department of Aeronautics and Astronautics. 2008. Disponível em: http://web.mit.edu/aeroastro/news/magazine/aeroastro5/cummings.html. Acesso em: 10 abr. 2020.

DAVIES, A. **Uber's Self-Driving Truck Makes Its First Delivery: 50,000 Beers**. Wired. 2016. Disponível em: https://www.wired.com/2016/10/ubers-self-driving-truck-makes-first-delivery-50000-beers/ Acesso em: 11 abr. 2020.

DIVINO, S. B. S. **Reflexiones escépticas, principiológicas y económicas sobre el consentimiento necesario para la recolección y tratamiento de datos**. Derecho PUCP, 83, 179 - 206, 29 nov. 2019. Disponível em: <https://doi.org/10.18800/derechopucp.201902.006>. Acesso em: 09. abr. 2020.

EIT HEALTH. **AI and Ethics in the Health Innovation Community**. 2020. Disponível em: https://eithealth.eu/wp-content/uploads/2020/01/AI-and-Ethics-in-the-Health-Innovation-Community.pdf. Acesso em: 11 abr. 2020.

ELIOT, Lance. **Human In-The-Loop Vs. Out-of-The-Loop in AI Systems: The Case of AI Self-Driving Car**s. 2019. Disponível em: https://www.aitrends.com/ai-insider/human-in-the-loop-vs-out-of-the-loop-in-ai-systems-the-case-of-ai-self-driving-cars/. Acesso em: 11 abr. 2020.

FISCHER, Joel E. **et.al. In-the-loop or on-the-loop? Interactional arrangements to support team coordination with a planning agent**. Concurrency Computat: Pract Exper. 2017, e4082. https://doi.org/10.1002/cpe.4082.

FLORIDI L. **Soft ethics, the governance of the digital and the General Data Protection Regulation**. Phil. Trans. R. Soc. 2018, A 376: 20180081. Disponível em: http://dx.doi.org/10.1098/rsta.2018.0081. Acesso em: 09. abr. 2020.

FRAZÃO, Ana. **Responsabilidade Civil de administradores de sociedades empresárias por decisões tomadas com base em sistemas de inteligência artificial**. in. FRAZÃO, Ana; MULHOLLAND, Caitlin. Inteligência Artificial e Direito. São Paulo: Thomson Reuters Brasil, 2019, p. 514.

GOLDBERG, D. E.; HOLLAND, J. H. **Genetic algorithms and machine learning**. In Machine learning. Vol. 3. Switzerland: 1988, p. 95-99.

GOMES, Rodrigo Dias de Pinho. Carros autônomos e os desafios impostos pelo ordenamento jurídico: uma breve análise sobre a responsabilidade civil envolvendo veículos inteligentes. In FRAZÃO, Ana; MULHOLLAND, Caitlin. **Inteligência Artificial e Direito**. São Paulo: Thomson Reuters Brasil, 2019, p. 567-585.

HULKKO, Ville. **Most value from ai with human-in-the-loop solutions**. Silo.AI. 2018. Disponível em: https://silo.ai/most-value-human-in-the-loop-ai/. Acesso em: 10 abr. 2020.

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.

LANCELOT, **Eliot. Self-Driving Car is the Mother of All AI Projects: Why It is a Moonshot**. https://www.aitrends.com/selfdrivingcars/self-driving-car-mother-ai-projects-moonshot/. Acesso em: 11 abr. 2020.

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

LEE, Kai-Fu. **Inteligência Artificial**. Rio de Janeiro: Globo Livros, 2019.

MAURER, Alexandre. **How can humans keep the upper hand on artificial intelligence?** ScienceDaily. 2017. Disponível em: https://www.sciencedaily.com/releases/2017/12/171204094950.htm. Acesso em: 11 abr. 2020.

MESKÓ, B. et al. **Will artificial intelligence solve the human resource crisis in healthcare?** BMC Health Services Research, 2018, 18, 545. Disponível em: https://doi.org/10.1186/s12913-018-3359-4. Acesso em: 09 abr. 2020.

MOTHI. **What is Human-in-the-Loop for Machine Learning?** Hackernoon. 2018. Disponível em: https://hackernoon.com/what-is-human-in-the-loop-for-machine-learning-2c2152b6dfbb. Acesso em: 10 abr. 2020.

NORVIG, Peter; RUSSELL, Stuart J. **Artificial intelligence: a modern approach**. New Jersey: Prentice Hall, 2010.

REEDY, C. **Kurzweil Claims That the Singularity Will Happen by 2045**. Futurism. Disponível em: https://futurism.com/kurzweil-claims-that-the-singularity-will-happen-by-2045. Acesso em: 09 abr. 2020.

SAE. **Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles**. 2018. Disponível em: https://www.sae.org/standards/content/j3016_201806/ . Acesso em: 11 abr. 2020.

THE STRATEGY WEB. Bosch Connected World 2020 – **The human unicorn on AI**. Disponível em: https://thestrategyweb.com/2020/02/20/bosch-connected-world-2020-the-human-unicorn-on-ai/. Acesso em: Acesso em: 11 abr. 2020.

UNIÃO EUROPEIA. **Artificial Intelligence: Europe needs to take a human-in-command approach, says EESC**. 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**). Disponível em: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai Acesso em: 09 abr. 2020.

UNIÃO EUROPEIA. **Regulamento 2016/679 do Parlamento Europeu e do Conselho**. 2016. Disponível em: http://eur-lex.europa.eu/legal-content/PT/TXT/?uri=celex%3A32016R0679 >. Acesso em: 21 mar. 2020.

WANG, G. Humans in the Loop: **The Design of Interactive AI Systems. Human-centered Artificial Intelligence**. Stanford University. 2019. Disponível em: https://hai.stanford.edu/news/humans-loop-design-interactive-ai-systems. Acesso em: 10 abr. 2020.

