

ANSWERABILITY E SEUS REFLEXOS PARA A RESPONSABILIZAÇÃO CIVIL

Nelson Rosenvald *

José Luiz de Moura Faleiros Júnior *

A palavra inglesa *answerability* é indicativa de um procedimento recíproco de justificação de escolhas que extrapola o direito à informação, facultando-se a compreensão de todo o cenário da operação realizada, que pode envolver processamento algorítmico de dados. Em matéria de responsabilidade civil, a busca por transparência algorítmica se coaduna com o referido termo e converge na proposição em torno da “explicabilidade”, descrita por Frank Pasquale como uma quarta lei da robótica para o Direito.

Não se trata de saber qual é a “inteligência artificial” utilizada e o que ela faz. O desafio está em buscar uma resposta ontológica, lastreada na identificação do cabimento das funções preventiva e precaucional da responsabilidade civil para que seja aferível a expectativa depositada sobre cada participante da atividade, especialmente quanto à previsibilidade de eventuais consequências. É legítimo que pessoas exijam uma explicação em nome de não-humanos ou mesmo em nome de outros humanos carentes de cognição.

Com base nessa premissa, este breve ensaio buscará esclarecimentos sobre a importância da “explicabilidade” para que se possa avançar em busca de soluções jurídicas palpáveis para os desafios da responsabilização civil de agentes algorítmicos em tempos nos quais a regulação dos eventos adversos decorrentes da utilização de sistemas decisoriais automatizados se posiciona no centro dos debates legislativos.

As clássicas três leis da robótica descritas por Isaac Asimov no conto “Círculo Vicioso”, o terceiro de sua coletânea “*Eu, Robô*”¹, inspiraram o professor norte-americano Jack Balkin a também

* Pós-Doutor em Direito Civil na Università Roma Tre (IT-2011). Pós-Doutor em Direito Societário na Universidade de Coimbra (PO-2017). Visiting Academic, Oxford University (UK-2016/17). Professor Visitante na Universidade Carlos III (ES-2018). Doutor e Mestre em Direito Civil pela Pontifícia Universidade Católica de São Paulo – PUC/SP. Presidente do Instituto Brasileiro de Estudos de Responsabilidade Civil – IBERC. Professor do corpo permanente do Doutorado e Mestrado do IDP/DF. Ex-Procurador de Justiça do Ministério Público de Minas Gerais. Advogado.

* Doutorando em Direito Civil pela Universidade de São Paulo – USP/Largo de São Francisco. Doutorando em Direito, na área de estudo ‘Direito, Tecnologia e Inovação’, pela Universidade Federal de Minas Gerais – UFMG. Mestre e Bacharel em Direito pela Universidade Federal de Uberlândia – UFU. Especialista em Direito Digital. Especialista em Direito Civil e Empresarial. Associado do Instituto Avançado de Proteção de Dados – IAPD. Membro do Instituto Brasileiro de Estudos de Responsabilidade Civil – IBERC. Advogado e Professor.

¹ Eis as três leis: “(i) um robô não pode ferir um ser humano ou, por inação, permitir que um ser humano sofra algum mal; (ii) um robô deve obedecer às ordens que lhe sejam dadas por seres humanos, exceto nos casos em que entrem em conflito com a Primeira Lei; (iii) um robô deve proteger sua própria existência desde que tal

formular três postulados para o enquadramento jurídico dessa complexa discussão: (i) operadores algorítmicos devem ser fiduciários de informações em relação a seus clientes e usuários finais; (ii) operadores algorítmicos têm deveres para com o público em geral; (iii) operadores algorítmicos têm o dever público de não se envolverem em incômodos algorítmicos.² Tais premissas devem ser internalizadas pelos desenvolvedores das aplicações para inspirar a produção do código-fonte do algoritmo. O mote deve ser a prevenção e, por esse motivo, estão ligadas ao papel do produtor/fabricante no mapeamento, na compreensão e no acautelamento dos riscos no processo de desenvolvimento.

Em proposta subsequente, Frank Pasquale propôs uma quarta lei: (iv) um robô sempre deve indicar a identidade de seu criador, controlador ou proprietário.³ Trata-se de permitir que determinada máquina indique quem é seu criador e, eventualmente, também revele a identidade de seu proprietário, por sinais emitidos, pela vinculação de número de série ou mesmo pela catalogação de sua propriedade em registros públicos.

Nesse contexto, torna-se pertinente a análise da palavra inglesa *answerability*. O termo é traduzido, literalmente, como “explicabilidade” (*explainability*), impondo-se como mais uma camada da função preventiva da responsabilidade, materializada no dever recíproco de construção da fidejussão a partir do imperativo da transparência (que é princípio expressamente previsto, por exemplo, no art. 6º, VI, da LGPD)⁴.

Enquanto *liability*, *responsibility* e *accountability* centram a atenção na pessoa que conduz uma atividade ou exerce comportamento danoso ou potencialmente danoso – os chamados agentes da responsabilidade –, a *answerability* se prende ao outro lado da relação: os destinatários ou “pacientes” de responsabilidade, que podem exigir razões para ações e decisões tomadas por aquele que exerce o controle da atividade. Assim, inspirada por uma abordagem relacional, a responsabilidade como “explicabilidade” oferece, uma justificativa adicional para a tutela da pessoa humana, com enorme valia perante corporações e operadores que terceirizam responsabilidades para algoritmos.

Em uma abordagem relacional dos problemas de responsabilidade, resta indubitável que não existe apenas um agente de responsabilidade (aquele que atua e quem deve agir com responsabilidade), mas também um paciente que é afetado pela ação do agente e que exige que este aja com responsabilidade no sentido daquilo que é esperado e reclama razões para sua ação. Responsabilidade não é apenas fazer algo e saber o que você está fazendo; é também uma questão

proteção não entre em conflito com a Primeira ou Segunda Leis”. Recomenda-se, ademais, a leitura da obra, que possui tradução para o português: ASIMOV, Isaac. *Eu, Robô*. Tradução de Aline Storto Pereira. São Paulo: Aleph, 2014.

² BALKIN, Jack M. The three laws of robotics in the age of Big Data. *Ohio State Law Journal*, Columbus, v. 78, p. 1-45, ago. 2017. <http://ssrn.com/abstract=2890965>. Acesso em: 13 out. 2023.

³ PASQUALE, Frank. Toward a fourth law of robotics: Preserving attribution, responsibility, and explainability in an algorithmic society. *University of Maryland Legal Studies Research Papers*, Baltimore, n. 21, p. 1-13, jul. 2017. Disponível em: <http://ssrn.com/abstract=3002546>. Acesso em: 13 out. 2023.

⁴ “Art. 6º. (...) VI - transparência: garantia, aos titulares, de informações claras, precisas e facilmente acessíveis sobre a realização do tratamento e os respectivos agentes de tratamento, observados os segredos comercial e industrial”.

comunicativa, talvez até dialógica, pois a sociedade, deseja respeitar os seres humanos não apenas como seres humanos autônomos, mas também sociais.⁵

A *answerability* é um procedimento recíproco de justificação de escolhas que extrapola o direito à informação, facultando-se a compreensão de todo o cenário da operação de tratamento de dados. Não se trata basicamente de saber qual é a IA utilizada e o que ela faz. O desafio está em buscar uma resposta ontológica, lastreada na identificação do cabimento das funções preventiva e precaucional da responsabilidade civil para que seja aferível a expectativa depositada sobre cada participante da atividade, especialmente quanto à previsibilidade de eventuais consequências.⁶ Com efeito, sempre nos pareceu razoável em ordenamentos democráticos que o agente fosse capaz de explicar à vítima por que ele praticou uma ação específica, tomou uma decisão ou recomendou algo. Por exemplo, pode-se pedir a um juiz que fundamente sua decisão ou demandar de um criminoso a explicação de suas ações. Se a regra é que o interessado é uma pessoa capaz de pedir e entender explicações, é também legítimo que pessoas exijam uma explicação em nome de não-humanos ou mesmo em nome de outros humanos carentes de cognição. As decisões e ações humanas precisam ser explicáveis se quiserem ser responsáveis - olhando para o passado e no presente.

No âmbito da proteção de dados pessoais, ela amplia o seu raio, convertendo-se em uma *“ability to appeal”*, ou seja, 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 (art. 20, lei 13.709/2018)⁷. O indivíduo pode se opor ao seu *profiling*, apagá-lo ou retificá-lo ou contestar decisões automáticas a ele relativas⁸.

Temos aí o chamado *“right to an explanation”*, exteriorizado no GDPR com relação a específicas decisões automatizadas,⁹ significando que a decisão deve ser explicada de uma forma

⁵ Como sustenta Mark Coeckelbergh, “But do people need explanations or do they need reasons? Can explanations count as reasons, and, if so, when? Responsibility as answerability can also be formulated in terms of reasons, or more specifically, in terms of giving reasons. And then for reasons the same seems to hold as for explanations in general: assuming that only humans can really give reasons, then responsible AI means that humans should get this task. The development of AI should then support this human task of giving reasons to those who ask or may ask questions about the actions and decisions mediated by the technology”. COECKELBERGH, Mark. Artificial intelligence, responsibility attribution, and a relational justification of explainability. *Science and Engineering Ethics*, Cham, v. 26, p. 2051-2068, 2020. Disponível em: <https://doi.org/10.1007/s11948-019-00146-8> Acesso em: 13 out. 2023.

⁶ BARBOSA, Mafalda Miranda. *Liberdade vs. responsabilidade: a precaução como fundamento da imputação delitual?* Coimbra: Almedina, 2006, p. 352.

⁷ Conferir, a esse respeito, CORDEIRO, A. Barreto Menezes. Decisões individuais automatizadas à luz do RGPD e da LGPD. In: BARBOSA, Mafalda Miranda; BRAGA NETTO, Felipe; SILVA, Michael César; FALEIROS JÚNIOR, José Luiz de Moura (Coord.). *Direito digital e inteligência artificial: diálogos entre Brasil e Europa*. Indaiatuba: Foco, 2021, p. 263 *et seq.*

⁸ Ilustre-se com a desativação algorítmica automática, comum na Uber e que acontece em todos os países nos quais o serviço tem presença. Os principais casos envolvem acusações de fraude, mas sem que os motoristas fiquem sabendo exatamente o que fizeram para obterem esse resultado, no que seria uma violação à *answerability*.

⁹ O art. 22(3) GDPR impõe ao controlador o implemento de salvaguardas ao desenhar decisões automatizadas: Art. 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. 2. O n. 1 não se aplica se a decisão: a) For necessária para a celebração ou a execução de um contrato entre o titular dos dados e um responsável pelo tratamento; b) For autorizada pelo direito da União ou do Estado-Membro a que o responsável pelo tratamento estiver sujeito, e na qual estejam igualmente previstas medidas

que o sujeito possa compreender o resultado, o que não requer necessariamente que a “*black box*” seja aberta¹⁰, mas simplesmente uma explicação contrafactual para que o particular se situe sobre o que deva ser modificado para que uma diferente decisão seja alcançada.¹¹

Para se estabelecer a relação de explicabilidade, a premissa é entender quais agentes são responsáveis por quais outros agentes, ou seja, “responsabilidade de quê?” E “responsabilidade para quem?”, por quais resultados, e para qual propósito? Se compreendermos quem deve responder, por que e a quem as respostas se destinam, alcançamos o conceito de supervisão – *oversight* – um componente de governança em que uma autoridade detém poder especial para revisar evidências de atividades e conectá-las às consequências. A supervisão complementa os métodos regulatórios de governança (*accountability*), permitindo verificações e controles em um processo, mesmo quando o comportamento desejável não pudesse ser especificado com antecedência, como uma regra. Ao invés, em caráter *ex post*, uma entidade de supervisão pode separar os comportamentos aceitáveis dos inaceitáveis. Aliás, mesmo quando existem regras, o supervisor pode verificar se o processo agiu de forma consistente dentro delas, sopesando as considerações nas circunstâncias específicas do cenário¹².

Prioriza-se uma revisão extrajudicial por humanos de decisões produzidas por algoritmos. A *answerability* não significa que se explique todo o processo causal que contribuiu para a ação ou decisão, mas sim que se possa saber o que é relevante.¹³ Eventualmente, a *liability* surgirá em um

adequadas para salvaguardar os direitos e liberdades e os legítimos interesses do titular dos dados; ou c) For baseada no consentimento explícito do titular dos dados. 3. Nos casos a que se referem o n. 2, alíneas a) e c), o responsável pelo tratamento aplica medidas adequadas para salvaguardar os direitos e liberdades e legítimos interesses do titular dos dados, designadamente o direito de, pelo menos, obter intervenção humana por parte do responsável, manifestar o seu ponto de vista e contestar a decisão.

¹⁰ É possível reproduzir o comportamento do algoritmo em perícia judicial tecnológica, sem que as empresas de tecnologia necessitem apresentar abertamente o seu código fonte em perícia judicial, a ponto de exibir todo o seu patrimônio intelectual e diferencial de mercado. Neste sentido, cf. FALEIROS JÚNIOR, José Luiz de Moura. Comércio eletrônico e publicidade digital: desafios para a proteção dos segredos comercial e industrial de algoritmos. In: MARTINS, Guilherme Magalhães; MARTINS, Fernando Rodrigues; SANTOS, Lindojon Gerônimo Bezerra dos (Coord.). *Direito do consumidor na sociedade da informação*. Indaiatuba: Foco, 2022, p. 141-158.

¹¹ “A counterfactual explanation could be for an individual whose application for a loan has been denied and wants to know why that the income statements provided by the individual show a yearly income of EUR 50,000, and the loan would be granted with a yearly income of EUR 60,000 or more”. MOEREL, Lokke; STORM, Marijn. Automated decisions based on profiling: information, explanation and justification – that is the question. In: AGGARWAL, Nikita; EIDENMÜLLER, Horst; ENRIQUES, Luca *et al* (Ed.). *Autonomous systems and the law*. Baden-Baden: Nomos, 2019, p. 94

¹² “Answerability includes not just the notion that answers exist, but that individuals or organizations can be made to answer for outcomes of their behavior or of the behavior of tools they make use of... if we want to know that an AI system is performing “ethically”, we cannot expect to “implement ethics in the system” as is often suggested. Rather, we must design the system to be functional in context, including contexts of oversight and review”. KROLL, Joshua A. Accountability in Computer Systems. In: DUBBER, Markus; PASQUALE, Frank; DAS, Sunit (Ed.). *The Oxford Handbook of the Ethics of Artificial Intelligence*. Oxford: Oxford University Press, 2020, p. 11.

¹³ Neste sentido, a exposição n. 71 do GDPR: “O titular dos dados deverá ter o direito de não ficar sujeito a uma decisão, que poderá incluir uma medida, que avalie aspectos pessoais que lhe digam respeito, que se baseie exclusivamente no tratamento automatizado e que produza efeitos jurídicos que lhe digam respeito ou o afetem significativamente de modo similar, como a recusa automática de um pedido de crédito por via eletrônica ou práticas de recrutamento eletrônico sem qualquer intervenção humana. Esse tratamento inclui a definição de perfis mediante qualquer forma de tratamento automatizado de dados pessoais para avaliar aspetos pessoais relativos a uma pessoa singular, em especial a análise e previsão de aspetos relacionados com o desempenho profissional, a situação econômica, saúde, preferências ou interesses pessoais, fiabilidade ou comportamento, localização ou deslocamentos do titular dos dados, quando produza efeitos jurídicos que lhe digam respeito ou a afetem significativamente de forma similar”.

momento posterior, se eventualmente eclodem danos em razão de atos ou atividades danosas que vulneram o *profiling* da pessoa ou alcançam situações existenciais.

Em um sentido ético, o importante não é a explicabilidade quanto às características de sistemas técnicos como a IA. Em verdade, o objetivo principal é explicabilidade como responsabilidade por parte do ser humano desenvolvedor da IA. A explicabilidade técnica, ou seja, o que o sistema de IA pode “dizer” ou “responder”, deve ser vista como algo a serviço do requisito ético mais geral de explicabilidade por parte do agente humano que necessita de um sistema com transparência para as respostas que serão dadas às pessoas afetadas pela tecnologia. Vale dizer, a função explícita da tecnologia é fazer o que deve ser feito dentro do objetivo pretendido pelos projetistas e usuários. Para além, entretanto, há uma questão ética de como esse sistema tecnológico impacta a forma como os agentes podem exercer a responsabilidade pela tecnologia e àqueles a quem respondem, sejam eles os destinatários imediatos, como também os seus entes queridos. Uma abordagem relacional abre uma relevante ecologia de relações de responsabilidade.

Por conseguinte, se um agente humano utilizando IA toma uma decisão com base em uma recomendação da IA e não é capaz de explicar por que ele tomou essa decisão, este é um problema de responsabilidade por dois motivos. Primeiro, o agente humano falhou em agir como um agente responsável, porque não sabe o que está fazendo. Em segundo lugar, o agente humano também deixou de agir com responsabilidade em relação ao paciente afetado pela ação ou decisão, que pode legitimamente exigir uma explicação por ela.

De certa forma, a *answerability* se cruza com a *responsibility*, na medida em que a inovação responsável significa levar em consideração que a “pessoa comum” é relativamente ignorante sobre a tecnologia e suas consequências imprevisíveis, um problema que precisa ser enfrentado pela educação e inclusão digital. O desenvolvimento de tecnologia e a educação em tecnologia devem ser alterados de forma a apoiar melhor os usuários e desenvolvedores de IA na resposta ao “Por quê?”. Isso significa que a sociedade merece operadores responsáveis de IA que estejam no controle, capazes e desejosos de comunicar, explicar e dar razões para o que estão fazendo a pacientes morais humanos. Isso inclui a obrigação de obter maior consciência das consequências imprevisíveis, incluindo como lidam com problemas trágicos. Se a IA não for responsável neste sentido, ela irá falhar.

Frank Pasquale,¹⁴serviu-se do *insight* das três leis de Jack Balkin¹⁵ para a sociedade algorítmica, a fim de propor uma quarta lei, capaz de complementar a tríade: “A *robot must always*

¹⁴ Segundo Frank Pasquale, “[o]ne key element of explainability is a clear sense of the history of a robot—how was it first programmed, to what has it been exposed, and how has this interplay between hardware, software, and the external environment resulted in present behavior. At the core of Balkin’s Laws of Robotics is a concern to make certain individuals (whose role parallels to that of the golem-creating rabbi) responsible for their creations. He does not want to create a set of legal obligations for algorithms or robots. Rather, he builds on our centuries-long experience with regulating persons. He observes that regulating the owners and programmers of artificial intelligence will require some monitoring of what they are creating and coding. To guarantee the efficacy of such monitoring, regulators may need to establish some ground rules, or pre- regulation, of the interactions algorithms will have with the wider world”. PASQUALE, Frank. *New laws of robotics*. Cambridge: Harvard University Press, 2020, p. 11.

¹⁵ “(1) With respect to clients, customers, and end-users, algorithm users are information fiduciaries. (2) With respect to those who are not clients, customers, and end-users, algorithm users have public duties. If they are governments, this follows immediately. If they are private actors, their businesses are affected with a public

indicate the identity of its creator, controller, or owner". A vanguarda dos campos de IA, aprendizado de máquina e robótica enfatiza a autonomia - seja de contratos inteligentes, algoritmos de negociação de alta frequência ou robôs futuros. Há uma noção nebulosa de robôs "fora de controle", que escapam ao controle e responsabilidade seu criador. A formulação da quarta lei com a exigência de que, com base na explicabilidade, qualquer sistema de IA ou robótica tenha alguém responsável por sua ação, ajuda a reprimir tais ideias.

Imagine-se um *drone* que emite sinais captados por torres de controle de tráfego aéreo, permitindo que se saiba quem o está a controlar; ou, voltando aos carros autônomos, um sistema que vincule o número de série do *software* que coordena o automóvel à identidade de seu fabricante, especificando detalhes como seu grau de autonomia, a espécie de sensor utilizado no mapeamento do ambiente etc.

O risco é o elemento central de todas essas propostas, mas a noção de perigo – e a própria precaução – também tem papel importantíssimo para o estudo de qualquer tecnologia. Sabidamente, “o conceito, a tipologia e a gravidade dos danos que inspiram a formatação dos sistemas de responsabilidade civil ao longo dos tempos variaram sob uma perspectiva proporcional à própria transformação da sociedade”¹⁶.

O fato de o risco ser comprovado ou potencial (hipotético) não afasta a pertinência dos princípios da prevenção e da precaução exatamente porque todo tipo de “novo dano” gera certa empolgação e, como alerta Ulrich Beck, acarreta suposições de aceitação social de novas tecnologias¹⁷ – ainda que não testadas – pelo fato de o risco, em alguma medida, se tornar inerente às diversas atividades da vida cotidiana¹⁸.

Nesse contexto, seria possível trabalhar com estruturas comparativas que ofereceriam condições mais precisas e bem mapeadas para determinar a atuação em conformidade (*compliance*). Noutros termos, embora seja desejável a função preventiva, baseada na *foreseeability*¹⁹, é preciso ir além, na busca por um critério para atender à função precaucional da responsabilidade civil.

interest, as constitutional lawyers would have said during the 1930s. (3) The central public duties of algorithm users are not to externalize the costs and harms of their operations. The best analogy for the harms of algorithmic decision making is not intentional discrimination but socially unjustified pollution”. BALKIN, Jack M. The three laws of robotics in the age of big data. *Ohio State Law Journal*, Columbus, v. 78, 2016. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2890965. Acesso em: 13 out. 2023.

¹⁶ VENTURI, Thaís G. Pascoaloto. *Responsabilidade civil preventiva: a proteção contra a violação dos direitos e a tutela inibitória material*. São Paulo: Malheiros, 2014, p. 248.

¹⁷ BECK, Ulrich. *Risk society: towards a new modernity*. Tradução do alemão para o inglês de Mark Ritter. Londres: Sage Publications, 1992, p. 6. Anota: “It is common to suppose that when there is no open public conflict about the risks of some technology, chemical or the like, this is evidence of positive public acceptance of the risks, or of the full social package of risk-technology-institutions”.

¹⁸ SANTOS, Romualdo Baptista dos. *Responsabilidade civil por dano enorme*. Curitiba/Porto: Juruá, 2018, p. 166. Segundo o autor, todos esses riscos, conjuntamente considerados, “estão relacionados ao processo de modernização da vida em sociedade, seja em razão da interferência do homem na natureza, seja em razão do desempenho de atividades necessárias ao modo de vida, seja ainda em consequência da exclusão das grandes massas populacionais em relação ao processo civilizatório”.

¹⁹ CALO, Ryan. Robotics and the lessons of cyberlaw. *California Law Review*, Berkeley, v. 103, p. 513-563, 2015, p. 555. O autor comenta: “Foreseeability remains a necessary ingredient even where liability is otherwise “strict” (i.e., where no showing of negligence by the plaintiff is necessary to recovery). There will be situations, particularly as emergent systems interact with one another, wherein otherwise useful technology will legitimately surprise all involved. Should these systems prove deeply useful to society, as many envision, some other formulation than foreseeability may be necessary to assess liability”.

Não se trata de proposta isolada, pois, na clássica obra de Stuart Russell e Peter Norvig já se falava na 'quantificação das incertezas': "Os agentes podem precisar lidar com a incerteza, seja devido à observabilidade parcial, ao não-determinismo da incerteza ou a uma combinação dos dois"²⁰.

O enfrentamento do incerto e do imprevisível não é novidade para o direito, que já lida com tais conceitos no direito ambiental, por exemplo²¹. O desafio está em buscar uma resposta ontológica, lastreada nas múltiplas funções da responsabilidade civil e no estabelecimento de padrões éticos²² que propiciem um mínimo de segurança no desenvolvimento de tecnologias complexas como os algoritmos de IA que viabilizam a tomada de decisões de forma automatizada.

Mesmo compreendida em sua multifuncionalidade e robustecida por diversos nexos de imputação a responsabilidade civil isoladamente não é capaz de oferecer uma tutela ótima diante de tecnologias digitais emergentes. Por conseguinte, se um agente humano utilizando IA toma uma decisão com base em uma recomendação da IA e não é capaz de explicar por que ele tomou essa decisão, este é um problema de responsabilidade por dois motivos. Primeiro, o agente humano falhou em agir como um agente responsável, porque não sabe o que está fazendo. Em segundo lugar, o agente humano também deixou de agir com responsabilidade em relação ao paciente afetado pela ação ou decisão, que pode legitimamente exigir uma explicação por ela.

Se compreendermos quem deve responder, as razões para isso e a quem as respostas se destinam, alcançamos o conceito de supervisão (*oversight*) um componente de governança em que uma autoridade detém poder especial para revisar evidências de atividades e conectá-las às consequências. A supervisão complementa os métodos regulatórios de governança (*accountability*), permitindo verificações e controles em um processo, mesmo quando o comportamento desejável não pudesse ser especificado com antecedência, como uma regra. Ao invés, em caráter *ex post*, uma entidade de supervisão pode separar os comportamentos aceitáveis dos inaceitáveis. Aliás, mesmo quando existem regras, o supervisor pode verificar se o processo agiu de forma consistente dentro delas, sopesando as considerações nas circunstâncias específicas do cenário.

²⁰ RUSSELL, Stuart J.; NORVIG, Peter. *Artificial Intelligence: a modern approach*. 3. ed. Boston: Pearson, 2016, p. 480, tradução livre. No original: "Agents may need to handle uncertainty, whether due to partial observability, uncertainty nondeterminism, or a combination of the two".

²¹ CALO, Ryan. Robotics and the lessons of cyberlaw. *California Law Review*, Berkeley, v. 103, p. 513-563, 2015, p. 555.

²² SILVA, Alexandre Barbosa da; FRANÇA, Philip Gil. Novas tecnologias e o futuro das relações obrigacionais privadas na era da inteligência artificial: a preponderância do "Fator Humano". In: EHRHARDT JÚNIOR, Marcos; CATALAN, Marcos; MALHEIROS, Pablo (Coord.). *Direito civil e tecnologia*. Belo Horizonte: Fórum, 2020, p. 507. Anotam: "Padrões éticos e de moralidade precisam ser estabelecidos para que um mínimo de segurança exista no desenvolvimento das relações interativas na realidade virtual digital. Tal indicação é necessária, pois desenvolver-se por meio da interação com elementos exógenos e endógenos não é uma opção para o ser humano, mas, sim, um caminho inescapável. Contudo, nesse contexto, algumas perguntas que se sobressaem são: até quando, de que forma e qual será o preço desse desenvolvimento do ser humano ante as realidades virtuais que cria?"