INTERAÇÃO SER HUMANO-MÁQUINA: O PADRÃO DE "CONTROLE HUMANO SIGNIFICATIVO" E SEUS IMPACTOS NA IMPUTAÇÃO DA RESPONSABILIDADE CIVIL POR DANOS DECORRENTES DE VEÍCULOS AUTÔNOMOS

HUMAN-MACHINE INTERACTION:
THE "SIGNIFICANT HUMAN CONTROL" STANDARD AND ITS IMPACTS ON THE CIVIL LIABILITY IMPUTATION FOR DAMAGES ARISING FROM AUTONOMOUS VEHICLES

Aline Klayse dos Santos Fonseca[*]

Resumo:
O presente artigo analisa as inflexões que as tecnologias digitais avançadas causam ao instituto da responsabilidade civil, notadamente visando traçar os principais desafios jurídicos que tais tecnologias representam para as categorias jurídico-dogmáticas relacionadas ao instituto da responsabilidade civil. Por meio da investigação do papel da multiplicidade dos agentes envolvidos no desenvolvimento e uso da inteligência artificial, destaca-se a problemática envolvendo a transparência, a imprevisibilidade das ações do sistema de inteligência artificial e do nexo de causalidade. Ademais, buscou-se investigar o potencial da interação ser humano-máquina no processo de tomada de decisão e como tal dinâmica pode ser relevante para a investigação da responsabilidade por danos, enfatizando, também, as lacunas legais e algumas abordagens para solucionar tais *gaps*. Por fim, investiga como o critério do "controle humano significativo" pode auxiliar na prevenção de danos decorrentes de sistemas de inteligência artificial autônoma. Avalia, ainda, distinções conceituais capazes de facilitar o desenvolvimento dogmático envolvendo inteligência artificial e os regimes de responsabilidade civil brasileiro. Utilizando-se de pesquisa teórica, conduzida pelos métodos dialético e dedutivo, estabeleceram-se associações entre tecnologia, inovação e responsabilidade.

Palavras-chave: Inteligência artificial. Interação humano-máquina. Danos. Responsabilidade civil.

Abstract:
This paper analyzes the results that advanced digital technologies cause to torts, notably aiming to outline the main legal challenges that such technologies represent for the legal-dogmatic categories related to it. Through the investigation of the diversity of roles of agents involved in the development and use of artificial intelligence, this paper discusses transparency, the unpredictability of the artificial intelligence system actions, and causal link. Furthermore, the research also aimed to investigate the potential of human-machine interaction in the decision-making process and how such dynamics can be relevant to investigation of liability for damages, emphasizing legal gaps and some approaches to solving them. Finally,

[*] Doutoranda em Direito Civil na Universidade de São Paulo (USP). Advogada. Professora do Instituto Federal do Pará. *E-mail*: alineklayse@usp.br.

it investigates how the "meaningful human control" criterion can help prevent damage resulting from autonomous artificial intelligence systems. It also evaluates conceptual distinctions that can facilitate dogmatic development involving artificial intelligence and Brazilian civil liability statutes. The theoretical research, conducted by dialectical and deductive methods, established associations between technology, innovation, and responsibility.

Keywords: Artificial intelligence. Human-machine interaction. Damages. Torts.

Introdução

O presente artigo está inserido no âmbito do direito privado e é conduzido pelo anseio de prestigiar a máxima tutela normativa do ser humano e de seus interesses existenciais e patrimoniais frente aos avanços tecnológicos contemporâneos, notadamente os que se baseiam em características como autonomia, autoaprendizagem, interação com o meio externo e que possuem a capacidade de tomar decisões independentes e aprender com a própria experiência.

De alguma maneira, há diversas tecnologias digitais avançadas que se inserem no cotidiano das relações sociais e são atreladas a um ideal de futuro melhor, mais dinâmico, prático e mais eficiente, mas, paradoxalmente, instauram um cenário de grandes complexidades, ameaças, inseguranças, de novos riscos e medos (reais ou aparentes), o que requer, além do seu manejo e prevenção, uma reorganização de poder e de responsabilidade (BECK, 2011, p. 28).

Francisco Amaral (2013, p. 148) já alertava que a complexidade se contrapõe à simplicidade que marcou, nesse particular, o pensamento jurídico da modernidade, pondo em xeque categorias e institutos tradicionais do direito privado, o que requer novos modos de produzir, de pensar, de tal maneira que o Direito consiga intervir em territórios até então inexistentes ou impensáveis.

Nesse contexto, testemunha-se a inclusão maciça de inteligência artificial[1] (IA) em vários espaços, cuja ênfase que lhe é dada envolve a possibilidade de efetuação de atividades de alta complexidade, otimização e realização de tarefas repetitivas e redução de custos. Por ser mantida por um amplo acervo de dados, essa tecnologia pode fornecer respostas relevantes e mais céleres que um humano, o que a torna atraente para

[1] Cumpre esclarecer que, o conceito de IA não é definido de forma rigorosa na literatura. Como o seu estudo possui estratégias e métodos distintos de análise que envolve uma abordagem centrada nos seres humanos (deve ser em parte uma ciência empírica, envolvendo hipóteses e confirmação experimental) e uma abordagem racionalista (envolve uma combinação de matemática e engenharia), isto implica em definições igualmente distintas (RUSSELL; NORVIG, 2013, p. 25).

o mercado, que enxerga na expansão de sistemas inteligentes, um fator de aumento de competitividade.

Todavia, o processo que envolve a utilização da inteligência artificial autônoma é complexo e pode ocasionar eventos danosos cuja identificação de um responsável é igualmente complexa. Isto porque, no processo de criação, desenvolvimento e uso dessa tecnologia, há diversos fatores que devem ser considerados como relevantes para a ocorrência do dano: a qualidade dos dados utilizados; a manipulação dos dados pelo agente de tratamento; a criação e treinamento do algoritmo pelo desenvolvedor; a existência de possíveis vieses algoritmos que podem conter preconceitos ou crenças do ser humano e que poderá interferir no desenvolvimento da máquina; a ação do usuário que pode ser feita desgarrada das orientações do programador; o constante aprendizado do agente inteligente autônomo que torna imprevisível suas ações até mesmo pelo desenvolvedor; defeitos nos sensores da máquina inteligente, dentre outros.

Assim, quais as possíveis soluções jurídicas advindas do instituto da responsabilidade civil para os danos injustos decorrentes de inteligência artificial autônoma, especialmente quando há forte interação entre o ser humano e a máquina inteligente? Qual o papel do ser humano dentro da dinâmica do sistema inteligente e como ele pode ser eficaz na prevenção de danos? Tais indagações nortearam a presente pesquisa e partiu da premissa de que é preciso que o ser humano esteja no controle das tecnologias inteligentes avançadas e o exerça de forma eficaz, de modo a evitar tanto quanto possível a ocorrência de danos e minimizar a excessiva dependência humana em relação às ações autônomas da máquina.

A condução da investigação se deu por meio de metodologia procedimental monográfica, tendo como principal técnica de pesquisa a consulta bibliográfica de doutrina nacional e estrangeira, publicada por meios escritos e eletrônicos. A primeira seção é destinada a traçar os principais aspectos das tecnologias digitais avançadas que representam desafios para as categorias jurídico-dogmáticas relacionadas ao instituto da responsabilidade civil, o que se fez através da investigação do papel da multiplicidade dos agentes envolvidos no desenvolvimento e uso da inteligência artificial, destacando-se a problemática envolvendo a transparência, a imprevisibilidade das ações do sistema de inteligência artificial e do nexo de causalidade. O que se pretendeu foi investigar o potencial da interação ser humano-máquina no processo de tomada de decisão e como tal dinâmica pode ser relevante para a investigação da responsabilidade por danos. A seção 2, abordará os principais impactos da inteligência artificial ao instituto da responsabilidade civil, as lacunas legais e a algumas abordagens para solucionar tais *gaps*. Apresentará o critério do padrão de controle humano significativo e a abordagem baseada em risco como uma alternativa para os desafios atuais sobre danos, responsabilidade e inteligência artificial. Por fim, estabelecem-se associações entre o padrão "controle humano significativo" na

inteligência artificial autônoma e seus reflexos no instituto da responsabilidade civil, com enfoque na aplicação para veículos autônomos.

1. Implicações das tecnologias digitais avançadas para o instituto da responsabilidade civil

A evolução do instituto da responsabilidade civil, norteada pelo anseio de oferecer respostas jurídicas satisfatórias às diversas lesões ou ameaças aos bens e interesses jurídicos relevantes, move-se se adaptando e transformando-se em face às novas tecnologias e atividades potencialmente danosas, embora nem sempre com a mesma velocidade com que elas se desenvolvem e se expandem socialmente.

Especialmente quanto à adoção de tecnologias que simulam a inteligência humana, baseada em características como autonomia, autoaprendizagem e que delegam, total ou parcialmente, a tomada de decisões à máquina, descortina dinâmicas relacionais que implicam em desafios jurídicos. Tais desafios envolvem o surgimento de novas vulnerabilidades, a abstração da pessoa humana para fins de comercialização, a titularidade de personalidade civil, impactam os direitos da personalidade e a responsabilização por danos.

Nessa seção, abordar-se-á os principais aspectos das tecnologias digitais avançadas que representam desafios para as categorias jurídico-dogmáticas relacionadas ao instituto da responsabilidade civil.

1.1. A multiplicidade de atores na cadeia de desenvolvimento e execução do sistema de inteligência artificial

No desenvolvimento e execução de um sistema de inteligência artificial autônomo há a participação de múltiplos atores, o que dificulta a intervenção, o controle pelos seres humanos, e, consequentemente, gera assimetria entre os sujeitos envolvidos e desigualdade que é agravada pela complexidade do sistema, carência de controle efetivo e dificuldade em modificar a decisão da máquina inteligente.

Em uma cadeia de desenvolvimento de uma máquina inteligente, abstratamente, a imputação da responsabilidade poderá recair em diversos sujeitos em situações de dano injusto: o fabricante das peças que compõem o sistema (hardware) e que podem apresentar defeito durante a execução; o desenvolvedor da tecnologia (*software*) que envolveu a tomada de decisão algorítmica; o agente de tratamento dos dados que alimentam a inteligência artificial; o usuário do sistema que poderá não cumprir as orientações exigidas para o bom funcionamento do sistema, dentre outros. Além disso, a crescente interconectividade das máquinas pode causar problemas, pois elas são capazes

de desenvolver simultaneamente uma infraestrutura conectada, tornando mais complicado o tratamento isolado de um único sistema.

Observa-se, então, que tecnologias digitais avançadas, tornam desafiador a tarefa de identificar quem é responsável diante de um evento lesivo, pois a análise retrospectiva das causas que ensejaram o dano é tão árdua quanto mais complexo é o sistema e menos os seres humanos são capazes de controlar, ou intervir diretamente na atividade dessas tecnologias (NOORMAN, 2018, p. 4).

No ordenamento jurídico brasileiro, a imputação da responsabilidade civil é, em regra, subjetiva, e o ressarcimento do dano injusto associa-se à apreciação da conduta do seu causador, de modo que, o fundamento da reparação do dano é o ato ilícito, nos termos do art. 186 do Código Civil. Assim, a obrigação de reparar o dano é imposta àquele que, por ação ou omissão voluntária, negligência ou imprudência, violar direito e causar dano a outrem. Extraem-se da referida norma os elementos da responsabilidade civil: dano; a culpa do agente; o nexo de causalidade entre o dano e a culpa (PEREIRA, 2018, p. 57).

Como os sistemas de IA intensificam a interação ser humano-máquina, isto é, não são entidades separadas, eles passam a exercer uma influência constitutiva das ações humanas. Para descrever a interação ser humano-máquina, foi introduzido o conceito de sistemas sociotécnicos com o objetivo de indicar que componentes técnicos e pessoas são incluídos como partes inerentes de um sistema, e, também, destacar a colaboração que esse conceito cumpre em uma determinada função, já que ambos – ser humano e máquina – andam de mãos dadas com a distribuição subjacente da agência, e, desse modo, afeta a imputação da responsabilidade (SIMMLER; FRISCHKNECHT, 2021, p. 240).

Tome-se como exemplo algumas situações em que a interação ser humano-máquina ocorre em graus distintos: i) Sistemas que fazem recomendações ao operador (componente técnico sugere opções e o humano decide); ii) Executa com aprovação humana (componente técnico atua após a aprovação humana); iii) Executa se não houver vetos humanos; iv) O componente técnico atua de forma independente e o humano é informado sobre as ações realizadas; v) O componente técnico realiza ações de forma independente sem informar os humanos (SIMMLER; FRISCHKNECHT, 2021, p. 243).

Nesse contexto, o sentido de agência também é de extrema relevância para o estudo da responsabilidade e pode ser definido como uma experiência de controle das próprias ações na interação com o ambiente, sendo, portanto, associada à descrição de dois mecanismos principais: a preparação para a ação e o *feedback* sensorial das ações (PAIS-VIEIRA; PAIS-VIEIRA, 2019, p. 208). O primeiro mecanismo se relaciona à intenção de realizar a ação e o segundo, à análise da melhor performance da interação.

Em relação aos veículos autônomos, a National Highway Traffic Safety Administration (NHTSA) apresenta uma taxonomia que classifica veículos autônomos

em seis níveis de automação, cujo critério leva em conta a influência do ser humano no controle da direção. O nível 0 não possui automação, já que o condutor permanece em tempo integral na direção dinâmica. No nível 1, há sistema de assistência, direção ou aceleração, enquanto no nível 2, o condutor humano executa todas as tarefas da direção dinâmica, mas com o auxílio de execução específica de um ou mais sistemas de assistência (MEDON, 2022, p. 181).

O nível 3 de automação é condicional, isto é, a presença de um condutor é necessária para que possa assumir a direção quando o sistema solicita a intervenção do motorista, e, no nível 4, mesmo que o sistema solicite a intervenção do condutor e ele não responda adequadamente ao pedido, o sistema tem condições e autonomia para realizar a direção. O nível 5, trata-se de um veículo que possui sistema de direção autônoma de todos os aspectos da direção dinâmica e em tempo integral (MEDON, 2022, p. 182).

Desse modo, esses níveis intermediários de automação, devido aos distintos graus de interação ser humano-máquina, e, a depender da situação fática que ocasione um evento danoso, podem gerar diferentes desdobramentos em termos de responsabilidade civil. A capacidade de aprender e adaptar o comportamento a um ambiente em mudança, processar informações, expandir o conhecimento implementado pelos programadores e mudar a forma como responde, permite que o sistema se adapte e melhore seu desempenho em um determinado ambiente. Por isso, sistemas adaptáveis são capazes de alterar seu comportamento o que tende a torná-los mais imprevisíveis e independentes.

Essas novas situações jurídicas exigem maior atenção do jurista sobre as nuances que envolvem as tecnologias digitais avançadas, seja em relação aos benefícios e aos riscos que delas decorrem. Segundo Anna Beckers e Gunther Teubner (2022, p. 17), a atuação algorítmica leva em conta suas relações com o ambiente natural e social, de forma individual (referente às propriedades intrínsecas de um único algoritmo, cujos riscos são impulsionados por seu único código-fonte ou design em sua interação com o ambiente), híbrida (resultado de interações próximas entre máquinas e humanos. Eles resultam em entidades emergentes sofisticadas com propriedades cujos riscos não podem ser identificados se isolamos os humanos e algoritmos envolvidos) e por meio de "comportamento" coletivo da máquina (comportamento de todo o sistema que resulta da interconectividade dos agentes da máquina).

Esse cenário de diferentes estágios de interação entre ser humano-máquina, quando aplicado aos veículos autônomos, evidencia que o regime de responsabilidade aplicável não dependerá apenas do nível de autonomia do veículo, mas da efetiva presença de um condutor responsivo na retaguarda, cuja obrigatoriedade dependerá da legislação a ser elaborada para o setor automobilístico no Brasil (MEDON, 2022, p. 237).

De todo modo, como alerta Filipe Medon (2022, p. 243), por mais imprevisível que seja a IA, incumbe aos fornecedores um dever de monitoramento e

informação constante ao usuário. Considerando as atuais regras vigentes no ordenamento jurídico brasileiro, as seguintes soluções envolvendo danos decorrentes de veículos com inteligência artificial autônoma podem ser respostas adequadas, segundo o referido autor:

> i) falha no sistema do veículo autônomo não atribuível à ação do consumidor, o fornecedor responderá civilmente pelos danos causados, diante da ocorrência de um acidente de consumo; ii) Em se tratando de carros com menor autonomia, nos quais ainda seja necessária a presença de um condutor na retaguarda, este responderá objetivamente com base no risco (nos casos de danos não recíprocos, pois nos recíprocos permanece a regra geral de responsabilidade subjetiva), só ilidindo a responsabilidade do fornecedor caso se comprove que o dano ocorreu devido à uma má utilização do sistema autônomo; iii) Para carros altamente ou totalmente autônomos (nível 4 e 5) ou quando, embora o condutor estivesse na retaguarda, o dano tenha sido causado por um defeito no processamento do sistema inteligente, o fornecedor será sempre responsável objetivamente; iv) O proprietário, por seu turno, nos casos de danos não recíprocos, será responsável pelo risco criado, na forma do parágrafo único do artigo 927 do Código Civil, o que poderá vir a ser eventualmente objeto de alteração com o tempo, caso se acabe comprovando, no futuro, que a atividade automobilística autônoma deixou de ser considerada risco.

A proposta apresentada pelo autor apresenta soluções jurídicas que evitem deixar a vítima sem o justo ressarcimento. Todavia, as tecnologias digitais avançadas emergem e impactam os conceitos jurídicos basilares que envolvem o instituto da responsabilidade civil, o que, do ponto de vista epistemológico, precisará de maior aprofundamento teórico.

Para fins deste artigo, outra distinção conceitual relevante envolve o conceito de controle de compartilhamento, monitoramento e controle de negociação. O controle de um veículo pode ser compartilhado entre o ser humano e máquina, e, em seguida, ambos fazem parte da tarefa de controle. Em caso de monitoramento, o motorista humano supervisiona a automação de direção e assume o controle a qualquer momento que seja necessário (SCHELLEKENS, 2022, p. 3).

No caso de controle de negociação, o controle sobre o veículo passa de um controlador para outro. Várias modalidades de transições podem ser discernidas. Primeiro, há transições do controle humano para o controle da máquina e vice-versa. Em segundo lugar, uma transição pode ser opcional (o humano quer deixar a automação fazer a condução daqui em diante, por exemplo) ou pode ser obrigatória, como quando o motorista humano adormece ao volante ou a automação de nível 3 que se aproxima de uma situação de estrada que não pode controlar e tem que deixar o controle para o

motorista humano. Finalmente, as transições podem ser iniciadas pelo motorista humano ou pela automação de direção. Uma transição opcional é sempre iniciada por um motorista humano (SCHELLEKENS, 2022, p. 4).

Observa-se que a entidade no controle não precisa ser a entidade que também é responsável. Por exemplo, se um humano precisa monitorar o funcionamento da automação de direção que está no controle total do veículo, o humano pode ser responsável, mas não está controlando o veículo.

Todavia, no cenário de veículos autônomos, o veículo assume parcial ou totalmente a responsabilidade de navegar no tráfego de modo que a ocorrência de danos injusto parece ocorrer mais em razão da falha do carro. Por isso, espera-se que a imputação objetiva da responsabilidade alcance maior importância no que diz respeito aos acidentes rodoviários, bem como a responsabilidade pelo fato do produto disciplinada no Código de Defesa do Consumidor (art. 12 a art. 17, CDC).

Algumas reflexões sobre a necessidade de adequação dos conceitos que estão estruturalmente relacionados à imputação da responsabilidade serão feitas no item 1.3. O próximo item se dedica a análise de como a complexidade, opacidade e ausência de transparência dos sistemas de inteligência artificial autônoma representam desafios para a indicação dos responsáveis por danos injustos decorrentes das novas tecnologias.

1.2. A transparência e o problema da (im)previsibilidade das ações do sistema de inteligência artificial e do nexo de causalidade

As diversas normas jurídicas que visam regular a inteligência artificial, a transparência aparece frequentemente como um requisito indispensável para uma IA ética e confiável. Em 8 de abril de 2019 foi publicado diretrizes éticas para uma inteligência artificial confiável, elaborado pelo Grupo Independente de Peritos de Alto Nível sobre a Inteligência Artificial (UNIÃO EUROPEIA, 2019, p. 3) criado pela Comissão Europeia em junho de 2018. Dentre as orientações, estabeleceu-se que o desenvolvimento, a implantação e o uso de sistemas de IA devem atender aos sete requisitos principais da IA confiável: agência e supervisão humana; robustez e segurança técnicas; privacidade e governança de dados; transparência; diversidade, não discriminação e justiça; bem-estar ambiental e social e *accountability*.

No que tange à transparência, o termo pode ser entendido como rastreabilidade, ou seja, registrar e documentar tanto as decisões tomadas pelos sistemas quanto todo o processo, incluindo uma descrição da coleta e rotulagem de dados e uma descrição do algoritmo utilizado. Entretanto, o termo também pode se relacionar com a explicabilidade do processo algorítmico de tomada de decisão, ou seja, explicações sobre o grau em que um sistema de IA influencia e molda o processo de tomada de decisão

organizacional, as escolhas de design do sistema e a justificativa para implantá-lo, ou, ainda, à transparência de dados e sistemas e a transparência do modelo de negócios. Abrange, ainda, a possibilidade de comunicar as capacidades e limitações do sistema de IA aos seus utilizadores, identificar o sistema de IA garantindo que os utilizadores saibam que estão a interagir com uma IA, e identificar as pessoas responsáveis pelo sistema (WOJTCZAK; KSIĘŻAK, 2021, p. 563).

A opacidade e as complexidades que envolvem as tecnologias digitais avançadas, permitem que os desenvolvedores se valham das normas jurídicas do regime tradicional de responsabilidade civil para se desviar da imputação. A distância entre desenvolvedores e os efeitos do uso das tecnologias que eles criam pode, por exemplo, ser usada para alegar que não há um nexo de causalidade direto e imediato que ligaria os desenvolvedores a um mau funcionamento, ou, ainda, que sua contribuição para a cadeia de eventos foi insignificante, pois fazem parte de cadeia maior ou que tiveram ínfimas possibilidades de controlar a ação da máquina inteligente (NOORMAN, 2018, p. 25).

Essas considerações evidenciam a fragilidade de se adotar o modelo tradicional de imputação subjetiva de responsabilidade, já que podem ocorrer vários fatores alheios à ingerência dos controladores da IA. Mesmo que fosse possível atribuir a decisão da IA ou do robô a um determinado fato, não ficaria claro como a decisão foi tomada, dificultando a caracterização de culpa. De fato, parece-nos que a análise dos danos decorrentes de sistemas inteligentes pelos juristas não deverá envolver um juízo de culpabilidade, mas de causalidade, sendo certo que por vezes será extremamente difícil determinar a causa e atribuir os devidos danos dela resultantes.

Nesse contexto, a noção de previsibilidade tem sido utilizada pelos desenvolvedores de sistemas inteligentes diante da ocorrência de um dano injusto, especialmente em sistema inteligente de aprendizagem profunda em que o controle do desenvolvedor é reduzido, de modo que a alegação do grau de liberdade que a tecnologia tem para concluir a tarefa que lhe foi ordenada tornaria a ação lesiva suficientemente imprevisível e sem a possibilidade real de impedi-la, rompendo o nexo de causalidade necessário para a imputação da responsabilidade.

Como vários sistemas de inteligência artificial, incluindo veículos autônomos, empregam aprendizado de máquina, à medida que o *software* interage com o ambiente, ele incorpora suas ações mais bem-sucedidas ao comportamento futuro e evolui com o tempo. Todavia, em cada incidente que gere um evento danoso, a questão de saber se a interação de um humano com a inteligência artificial era previsível ou inesperada se tornará mais difícil. Devido a essa imprevisibilidade, alegações de causas supervenientes podem ser usadas isentando o desenvolvedor de inteligência artificial da responsabilidade (KOWERT, 2020, p. 184).

Entretanto, inserir o cenário de danos decorrentes de sistemas de inteligência artificial nos extremos, isto é, de responsabilidade total ou nenhuma responsabilidade não fornece uma resposta jurídica adequada e alinhada à evolução dogmática do instituto da responsabilidade civil. Em que pese haja muitas interações entre humanos intervenientes e o *software* capazes de aumentar os riscos contra os quais os desenvolvedores de inteligência artificial não possam controlar, outras tantas interações podem ser consideradas previsíveis.

Ressalta-se que, como já apontou a doutrina civilista francesa, a investigação sobre se os efeitos lesivos poderiam ou deveriam ser previstos, volta-se à busca da culpa, de modo que a análise da causalidade em função da previsibilidade é tendenciosa, na medida em que ela reconduz indiretamente a causalidade à culpa (VINEY; JOURDAIN; CARVAL, 2013, p. 246-248). Para isso, o critério da normalidade do dano em substituição ao critério da previsibilidade ou evitabilidade tem sido adotado como uma alternativa, já que, para fins de causalidade, é relevante a análise sobre se o evento é previsível do ponto de vista objetivo, ou seja, baseando-se em conhecimentos científicos disponíveis, e não na previsibilidade por parte do agente.

Ademais, como alerta Gustavo Tepedino e Rodrigo Silva (2019, p. 75), é necessário um enfrentamento da questão envolvendo a imprevisibilidade das condutas dos sistemas autônomos para que não se constitua como um falso problema, pois, independentemente da previsibilidade das reações dos robôs submetidos à autoaprendizagem, o problema da reparação de danos, nesses casos, há de ser solucionado no âmbito da causalidade e da imputabilidade daí decorrente, a partir da alocação de riscos estabelecida pela ordem jurídica ou pela autonomia privada.

Uma vez que a interação ser humano-máquina é capaz de alterar a tomada de decisão da máquina inteligente, também tem o potencial de alterar a imputação da responsabilidade. O próximo item será destinado à análise dos conceitos estruturalmente relacionados à responsabilidade civil, tais como o conceito de autonomia, e suas inflexões ante a inteligência artificial.

1.3. Conceitos estruturalmente relacionados à responsabilidade e suas inflexões ante a inteligência artificial

Considerar as possíveis inflexões que o instituto da responsabilidade civil perpassa com as ameaças ou lesões a bens jurídicos decorrentes de tecnologias digitais avançadas, requer a investigação detida sobre as peculiaridades da capacidade digital de ação, sobre a natureza dos bens que são violados pela IA e a necessidade de repensar conceitos que estão estruturalmente relacionados à imputação da responsabilidade e que

poderá contribuir para a apresentação de respostas jurídicas mais coerentes e harmônicas diante da ocorrência de dano injusto.

Em um cenário de mudanças tão robustas e de tecnologias variadas, é importante, inicialmente, estabelecer algumas distinções. A primeira delas é que inteligência artificial autônoma se distingue de decisões automatizadas e de sistemas automáticos. Alguns sistemas apenas executam ações prescritas que são pré-fixadas e não mudam em resposta ao ambiente, ou seja, são sistemas meramente automáticos. Outros, porém, iniciam ou ajustam suas ações ou desempenho com base no *feedback* do ambiente, denominam-se de automatizados. Os sistemas de inteligência artificial realmente compreendido como autônomos são aquelas que estão em algum grau fora do controle humano. O ser humano pode exercer algum controle durante o projeto, desenvolvimento, no ponto de ativação para uma tarefa específica ou durante a operação, por exemplo, interrompendo seu funcionamento (INTERNATIONAL COMMITTEE OF RED CROSS, 2019, p. 8).

Desse modo, a necessidade de controle humano, liga-se à complexidade do ambiente e da tarefa que deverá realizar: quanto maior a complexidade, maior a necessidade de controle humano direto e menor a tolerância de autonomia, especialmente para tarefas e em ambientes onde uma falha do sistema pode matar ou ferir pessoas ou danificar propriedades.

Especialmente no que tange à inteligência artificial autônoma, um importante questionamento a ser feito é o que o Direito deve assumir como inteligência. Isto porque, a narrativa em torno da inteligência artificial é estruturada em comparação com a inteligência humana, por meio de paralelos com o ser humano como em "a máquina conseguiu solucionar o problema de modo semelhante a um humano", ou "a decisão tomada pela máquina é válida, pois foi equivalente à decisão de um humano" (FONSECA, 2021, p. 7). Mas quando um sistema inteligente age, isso significa que computadores podem ser equiparados a humanos? Ou, de outro lado, o jurista deve assumir algoritmos como um simples fluxo de informação matematicamente formalizado? A profunda imersão do ser humano com os artefatos digitais permite que eles sejam concebidos em comparação com o ser humano?

No intuito de exemplificar o novo cenário que emerge com o hibridismo ser humano-máquina, algumas formas de agenciamento na cultura digital podem ser visualizadas, a exemplo de quando artefatos digitais alertam, autorizam ou validam a ação humana, ou quando aplicativos indicam o melhor caminho a ser seguido pelo condutor, ou mesmo ordena que o motorista coloque o cinto de segurança para poder conduzir o veículo. Observa-se, então, a existência de uma espécie de "capacidade" de sociabilidade de não humano.

Lucia Santaella e Tarcísio Cardoso (2015, p. 173), ao investigar os estudos de Sayes, apontam quatro variações no conceito de não humano: i) como uma condição para a possibilidade da sociedade humana (não humanos I); ii) como mediadores (não humanos II); iii) como membros de uma associação moral e política (não humanos III) e; iv) como agregadores de atores de diferentes ordens espaço-temporais (não humanos IV). A partir dessa diferenciação, destacam que essa pluralidade de arranjos omite ou dificulta a responsabilização ou do que indivíduos ou grupos devem ser responsabilizados por associações morais e políticas.

Uma tentativa comum de esclarecer o papel e ações dos sistemas inteligentes é compará-los às pessoas jurídicas. Para Anna Beckers e Gunther Teubner (2022, p. 17), para realizar o paralelo estrito entre atores digitais e atores coletivos é necessário rejeitar dois equívocos de personificação de entidades não humanas: conceber organizações como conjuntos de pessoas agregadas em uma pessoa coletiva real, e, postular que os agentes de *software* transformam um computador em um *homo ex machina*.

Assim, um primeiro aspecto conceitual relevante diz respeito a propriedades dos sistemas de inteligência artificial que não condiz com a noção de agente, tampouco de pessoa. Há assimetrias na interação ser humano-máquina que corroboram tal assertiva. Com base nos estudos de Bruno Latour sobre mediação técnica, os autores destacam três assimetrias entre ser humano e máquina que asseveram que o termo mais adequado para sistemas tecnológicos de forte interação entre ser humano e máquina é o conceito de "actantes".

A primeira assimetria é operacional: as operações internas dos algoritmos não podem ser equiparadas às operações mentais dos humanos, pois o funcionamento interno consiste em operações matemáticas baseadas em sinais eletrônicos. Ademais, na comunicação entre atores humanos, os parceiros fazem a escolha de seu comportamento dependendo da escolha do outro, isto é, a dupla contingência é simétrica em ambos os lados, enquanto que, na comunicação entre humano e máquina, a dupla contingência é experimentada apenas unilateralmente, ou seja, apenas pelo humano e não pela máquina. Por fim, a assimetria também envolve o processo de compreensão mútua do ser humano e da máquina: os humanos poderiam ser capazes de entender os processos internos do algoritmo, mas, simultaneamente, o algoritmo pode não ter a capacidade de reconstruir a autorreferência da vida humana interior (BECKERS; TEUBNER, 2022, p. 29).

Desse modo, o conceito de actantes evidenciam que a interação ser humano-máquina não se refere a processos digitais antropomorfizantes, mas, ao revés, podem ser compreendidos como agentes de *software* desantropomorfizantes, isto é, permanecem máquinas sem mente, mas quando a atribuição de ação a eles é firmemente institucionalizada em um campo social, eles se tornam membros não humanos da sociedade. Tais diferenciações conceituais são relevantes pois, sujeito, objeto ou qualquer

outra terminologia usada para categorizar um elemento da produção do sentido, apenas só pode ser bem entendido a partir da construção do texto e levando em consideração a relação que ele guarda com o contexto no qual está inserido (SANTAELLA; CARDOSO, 2015, p. 170).

Outro aspecto que merece destaque se refere à noção de intencionalidade em instituições sociodigitais. Isto porque, como já explanado no item 1.1, os sistemas de inteligência artificial intensificam a interação ser humano-máquina e não são entidades separadas, ao revés, exercem uma influência constitutiva das ações humanas, logo, os efeitos são multicausais, produtos de interação. Desse modo, a intenção perpassa pelo sistema humano-máquina, de forma coletiva, sugerindo a responsabilização igualmente compartilhada entre os vários actantes.

A intencionalidade analisada diante da falha do sistema inteligente com interação ser humano-máquina, poderá ensejar a imputação da responsabilidade e a consequente reparação ou compensação dos danos de modo proporcional, nas situações que decorrem de causas múltiplas, concomitantes ou sucessivas.

Todavia, a depender da situação fática, apenas um agente poderá suportar os efeitos da responsabilização (caso se adote o critério subjetivo de responsabilidade civil). Se, a título exemplificativo, um desenvolvedor vende um *software* de inteligência artificial a um indivíduo racista que o instala em seu robô doméstico que passa a aprender e se desenvolver sob os ensinamentos de seu proprietário, e, agride um negro que se dirige à residência do proprietário para fazer uma entrega (por aprender por meio das práticas do proprietário que negros são assaltantes), em que pese haver diferentes atores humanos relacionados ao dano, não há intencionalidade do desenvolvedor do *software* de inteligência artificial e isso poderá ser levado em consideração quando da imputação da responsabilidade.

Outro conceito que está estruturalmente relacionado à responsabilidade é a autonomia que não se confunde com automatização ou automação. De acordo com o Comitê Internacional da Cruz Vermelha (2019, p. 7), a autonomia tem sido compreendida nos estudos sobre inteligência artificial como a capacidade de o sistema agir sem intervenção humana direta. No entanto, trata-se de característica de alguns sistemas ajustarem suas ações com base no *feedback* do ambiente, e, também, sobre sua própria situação atual. O aumento da autonomia é geralmente equiparado a uma maior adaptação ao ambiente e às vezes é apresentado como um aumento da "inteligência" para uma determinada tarefa.

Observa-se que, não são as propriedades internas do algoritmo e da inteligência artificial autônoma que lhes conferem autonomia, mas a sua comunicação e participação social, ou seja, tais aspectos que são relevantes para o reconhecimento da autonomia do ponto de vista jurídico. A intencionalidade, conforme dito alhures, é um conceito importante para a análise dos danos injustos decorrente de tecnologias

digitais avançadas, mas, em se tratando de instituições sociodigitais com forte interação entre ser humano-máquina, não deve ser relacionada com a tradicional compreensão de estado psicológico interno do agente, mas a atribuição externa de ação intencional por um observador.

Em suma, em tecnologias digitais avançadas com forte interação ser humano-máquina, a noção de autonomia no sentido jurídico deve perpassar análise da orientação para o objetivo do agente e escolha dos meios, participação do algoritmo na comunicação social. Entretanto, Anna Beckers e Gunther Teubner (2022, p. 37) elencam outros critérios para que o direito assuma que sistemas de inteligência artificial tenha autonomia: i) o *software* deve ter sido programado de tal forma que ele tenha que decidir entre alternativas; ii) deve ter tomado essa decisão como otimização de vários critérios; iii) um programador não pode explicar o comportamento do agente de *software* retrospectivamente fazer predição para o futuro, mas pode apenas corrigi-lo *ex-post*.

Em termos de responsabilidade, isso implica no incremento de obrigações aos desenvolvedores e fabricantes de sistemas inteligentes, de modo a ter mecanismos na própria tecnologia capazes de registrar o processo da tomada de decisão, identificar incertezas e fornecer informações sobre como a decisão foi baseada, facilitando a identificação da causa do possível comportamento que ensejou o dano injusto.

A seção 2, abordará os principais impactos da inteligência artificial ao instituto da responsabilidade civil, as lacunas legais e a algumas abordagens para solucionar tais *gaps*. Apresentará o critério do padrão de controle humano significativo e a abordagem baseada em risco como uma alternativa para os desafios atuais sobre danos, responsabilidade e inteligência artificial.

2. O padrão "controle humano significativo" para inteligência artificial autônoma e seus reflexos no instituto da responsabilidade civil

As características da inteligência artificial autônoma e da interação ser humano-máquina impactam a dogmática jurídica civilista, sobretudo em matéria de responsabilidade civil, já que o modelo tradicional ainda não apresenta contornos satisfatórios diante dos riscos e complexidades de tecnologias baseadas no aprendizado, mesmo diante da inegável expansão do instituto jurídico, hoje, fortemente ancorado nas normas constitucionais.

As preocupações referentes aos desafios que as tecnologias digitais avançadas trazem ao instituto da responsabilidade civil como "lacuna de responsabilidade", para indicar a preocupação de que a presença de tecnologias baseadas no aprendizado pode tornar mais difícil a imputação. O autor Andreas Matthias (2009, p. 24) destacou que, dentre as condições para a responsabilidade na estrutura da filosofia analítica

contemporânea, destacam-se cinco pontos principais: intencionalidade; ter entendimento racional das relações causais entre ações e resultados, além da capacidade de o agente agir de acordo com seu entendimento; ter volições de segunda ordem, ou seja, a capacidade de escolher os objetivos que ela deseja perseguir; ser sã; ser capaz de distinguir entre o que é pretendido e o meramente previsto.

Não obstante, considerando a singularidade que envolve dos danos decorrentes possibilidade de agentes subpessoais como a inteligência artificial autônoma, nem todos os critérios acima mencionados são perfeitamente atendidos, a exemplo do critério da "sanidade", já que isto pressupõe uma enorme quantidade de conhecimento sobre que tipos de volições são comumente usados em um contexto social específico (MATTHIAS, 2009, p. 25).

Em outros termos, sistemas inteligentes com capacidade de aprendizado por meio de interação com outros agentes e com o ambiente, tormam o controle humano e a previsão sobre as ações da máquina mais difícil, mas a responsabilidade requer controle e conhecimento.

Atualmente, tem-se buscado ressignificar a discussão da lacuna de responsabilidade em termos mais alinhados com a categorização dos conceitos da responsabilidade na filosofia moral e jurídica, de modo a identificar que tipo de responsabilidade é ameaçada, por qual aspecto da automação e por qual razão isso é importante. Ao menos quatro lacunas da responsabilidade por danos decorrentes de inteligência artificial podem ser observadas: lacunas da culpabilidade, da responsabilidade moral e pública e responsabilidade ativa (SANTONI DE SIO; MECACCI, 2021, p. 1.057).

Segundo os autores, a lacuna da culpabilidade se refere ao fato de que o uso de inteligência artificial ao tomar decisões introduz um novo elemento de opacidade técnica e falta de explicabilidade que torna mais difícil para as pessoas individuais satisfazer as condições tradicionais de culpabilidade: intenção, previsibilidade e controle. A lacuna da moralidade diz respeito à preocupação geral de que a IA possa tornar as pessoas menos capazes de entender, explicar e refletir sobre seu funcionamento e o de outros agentes. Por ser alimentada por dados de diversas fontes, a IA dificulta que o cidadão saiba a quem recorrer em caso de utilização de dados incorretos, corrompidos ou tendenciosos, por resultar de contribuições coletivas (lacuna da responsabilidade pública). Por fim, a lacuna da responsabilidade ativa consiste no fato de que os atores envolvidos no projeto ou no uso da IA não estão suficientemente cientes da própria responsabilidade de se evitar danos, ou não são capazes e motivados para cumprir essa obrigação, por atentarem muito mais para os benefícios técnicos da inteligência artificial (SANTONI DE SIO; MECACCI, 2021, p. 1.068).

Nessa senda, as características de aprendizagem, opacidade, interação com muitos agentes e recursos autônomos ou semiautônomos são questões centrais sobre a

relação ser humano-máquina e responsabilidade, de modo que a necessidade de controle humano passa a ser fundamental. Buscando resolver as lacunas de responsabilidade, foi proposto o conceito de "controle humano significativo", apresentado pela primeira vez para enfrentar os desafios envolvendo sistemas de armas autônomas. Todavia, tal conceito ganha cada vez mais destaque no debate sobre inteligência artificial e responsabilidade, especialmente quando aplicado a veículos autônomos (SIEBERT *et al.*, 2022, p. 3). O item 2.1 abordará esse conceito como padrão para o desenvolvimento e utilização da inteligência artificial responsável.

2.1. O padrão "controle humano significativo"

O primeiro uso influente do padrão "controle humano significativo" como conceito pode ser atribuído a um artigo de 2013 constante no documento apresentado na Reunião de Peritos da Convenção do Relatório Técnico sobre Certas Armas Convencionais em Genebra (Artigo 36), publicado como resposta direta à crescente controvérsia relacionados aos sistemas de armas autônomas. O termo rapidamente ganhou popularidade e, em 2014, foi adotado por outras organizações e Estados como um conceito-chave para enquadrar os debates sobre a autonomia em armas. Em 2015, o padrão controle humano significativo tornou-se a questão dominante nos debates sobre autonomia sistemas de armas (KWIK, 2022, p. 3).

As propriedades de um sistema sob controle humano significativo podem ser resumidas em: i) o sistema de IA humano tem um domínio de design operacional moral explícito e o agente de IA adere aos limites desse domínio; ii) Os agentes humanos e de IA têm representações apropriadas e mutuamente compatíveis do sistema humano-IA e seu contexto; iii) Os agentes relevantes têm capacidade e autoridade para controlar o sistema para que os humanos possam agir de acordo com sua responsabilidade; iv) As ações dos agentes da IA estão explicitamente vinculadas às ações dos humanos que estão cientes de sua responsabilidade moral (SIEBERT *et al.*, 2022, p. 4).

No contexto de veículos autônomos, o controle humano significativo pode trazer contribuições quando aplicado às situações de forte interação entre ser condutores e o sistema inteligente presente no veículo autônomo, possibilitando o controle eficaz do ser humano em momentos de urgência. Por meio dessa abordagem, garante-se que o elemento humano na instituição ser humano-máquina tenha um papel relevante na tomada de decisão, mormente em situações de ameaça à bens de elevada relevância jurídica.

Isso porque, embora a inteligência artificial autônoma possua, comumente, elevados níveis de acurácia, não se pode superestimar as tecnologias digitais avançadas e fomentar um cenário de quase absoluta dependência das decisões advindas de artefatos digitais. Tal controle possui especial importância em veículos autônomos com níveis

de autonomia de 3 a 5, em que há maior delegação decisória para o sistema inteligente autônomo.

Nesse cenário, as funções de monitoramento precisam avaliar os limites ou falhas dos sistemas de inteligência do veículo autônomo e/ou as futuras tarefas de direção que exigem a transferência do controle do veículo para o motorista humano. Se isso ocorrer, haverá a necessidade de apoiar adequadamente a transição para a condução manual. A primeira questão crucial é gerar um *Take-Over Request* (TOR) para o condutor, ou seja, informações ou avisos, dependendo da emergência da situação de trânsito e do orçamento de tempo disponível, e, em seguida, para gerenciar as transições da condução automatizada para a condução manual (BELLET *et al.*, 2019, p. 157).

Todavia, isso se relaciona à avaliação das habilidades do motorista humano para executar manualmente a tarefa de direção atual. Assim, o sistema deve ser capaz de avaliar o elemento humano como sendo capaz de retomar o controle manual de forma segura. Nos outros casos de um motorista distraído ou inconsciente, os sistemas autônomos devem manter o controle total do veículo. Ademais, o condutor deve ser claramente informado pelo sistema sobre as tarefas de condução que estão sendo desempenhadas pelo veículo autônomo e, se necessário, sobre as situações que se aproximam que exijam a retomada do controle manual.

Conclusão

A multiplicidade de situações envolvendo a inteligência artificial e a consolidação das instituições sociodigitais desafia a imputação da responsabilidade que, abstratamente, poderá recair em diversos sujeitos diante de um evento lesivo. Todavia, a crescente utilização de tecnologias digitais avançadas gera inflexões jurídicas, notadamente dogmáticas, que exigem maior atenção do jurista.

Observou-se que, a atuação algorítmica ocorre de forma individual ou híbrida, resultando em entidades cujos riscos não podem ser identificados se isolamos os humanos. É fundamental a análise do "comportamento" coletivo da máquina e dos diferentes estágios de interação entre ser humano-máquina.

Ademais, verifica-se que o contexto de tecnologias digitais avançadas gera dissonâncias envolvendo os conceitos jurídicos que tradicionalmente estão relacionados à responsabilidade civil, a exemplo do conceito de autonomia que, não diz respeito às propriedades internas da inteligência artificial autônoma, mas a sua comunicação e participação social, ou, ainda, o conceito de intencionalidade que não deve ser relacionada com a tradicional compreensão de estado psicológico interno do agente, mas a atribuição externa de ação intencional por um observador.

No que tange ao impacto das tecnologias digitais avançadas ao instituto da responsabilidade civil, observa-se diferentes lacunas de responsabilidade, entendidas como dificuldades que a presença de tecnologias baseadas no aprendizado traz para a imputação, especialmente no contexto de forte interação ser humano-máquina. O padrão de controle humano significativo contribui para sanar tais lacunas, na medida em que visa conferir maior domínio de design operacional explícito, representações compatíveis do sistema humano-IA e capacidade do ser humano controlar o sistema e agir de acordo com sua responsabilidade. Esse padrão, quando aplicado aos veículos autônomos, permite que o condutor receba mais informações sobre situações emergenciais para gerenciar as transições da condução automatizada para a condução manual, evitando a ocorrência de danos.

São Paulo, setembro de 2022.

Referências

AMARAL, Francisco. Código civil e interpretação jurídica. *Revista Brasileira de Direito Comparado*, Rio de Janeiro, n. 44/45, p. 147-167, jan./jun. 2013.

BECK, Ulrich. *Sociedade de risco*: rumo a uma outra modernidade. Tradução de Sebastião Nascimento. 2. ed. São Paulo: Editora 34, 2011.

BECKERS, Anna; TEUBNER, Gunther. *Three liability regimes of artificial intelligence*: algorithms actants, hybrids, crowds. Oxford: Hart Publishing, 2022.

BELLET, Thierry; CUNNEEN, Martin; MULLINS, Martin; MURPHY, Finbarr; PÜTZ, Fabian; SPICKERMANN, Florian; BRAENDLE, Claudia; BAUMANN, Martina Felicitas. From semi to fully autonomous vehicles: new emerging risks and ethico-legal challenges for human-machine interactions. *Transportation Research Part F*: Traffic Psychology and Behaviour, [*s. l.*], v. 63, p. 153-164, May 2019. Disponível em: https://reader.elsevier.com/reader/sd/pii/S1369847818308556?token=800F16988DEB5CC6A75C6DF866F32F18DCF69D32C839939B58B6E9D709D041F089138A597FCB9B5E39F23166708E74BA&originRegion=us-east-1&originCreation=20220913203026.

FONSECA, Aline Klayse dos Santos. Delineamentos jurídico-dogmáticos da inteligência artificial e seus impactos no instituto da responsabilidade civil. *Civilistica.com*: Revista Eletrônica de Direito Civil, Rio de Janeiro, v. 10, n. 2, p. 1-36, set. 2021. Disponível em: https://civilistica.emnuvens.com.br/redc/article/download/671/546.

INTERNATIONAL COMMITTEE OF THE RED CROSS. *Artificial intelligence and machine learning in armed conflict*: a human-centred approach. Geneve, 6 Jun. 2019. Disponível em: https://www.icrc.org/en/document/artificial-intelligence-and-machine-learning-armed-conflict-human-centred-approach. Acesso em: 7 nov. 2022.

KOWERT, Weston. The foreseeability of human-artificial intelligence interactions. *Texas Law Review*, Austin, v. 96, n, 1, p. 181-204, 2020. Disponível em: https://texaslawreview.org/foreseeability-human-artificial-intelligence-interactions/.

KWIK, Jonathan. A practicable operationalisation of meaningful human control. *Laws*, Basel, v. 11, n. 3, 21 p., 2022. Disponível em: https://www.mdpi.com/2075-471X/11/3/43/pdf?version=1652699783.

MATTHIAS, Andreas. From coder to creator: responsibility issues in intelligent artifact design. *In*: LUPPICINI, Rocci; ADELL, Rebecca. *Handbook of research on technoethics*. New York: IGI Global, 2009. p. 635-650.

MEDON, Filipe. *Inteligência artificial e responsabilidade civil*: autonomia, riscos e solidariedade. 2. ed. Salvador: JusPodivm, 2022.

NOORMAN, Merel. Computing and moral responsibility. *In*: ZALTA, Edward Nouri (ed.). *The Stanford Encyclopedia of Philosophy*, Stanford, 2018. Disponível em: https://plato.stanford.edu/entries/computing-responsibility/.

PAIS-VIEIRA, Miguel; PAIS-VIEIRA, Carla. Interfaces cérebro-máquina e os limites da responsabilidade. *In*: CURADO, Manuel; FERREIRA, Ana Elisabete; PEREIRA, André Dias (coord.). *Vanguardas da responsabilidade*: direito, neurociências e inteligência artificial. Coimbra: Petrony, 2019.

PEREIRA, Caio Mário da Silva. *Responsabilidade civil*. 12. ed., rev., atual. e ampl. Rio de Janeiro: Forense, 2018.

RUSSELL, Stuart; NORVIG, Peter. *Inteligência artificial*. Tradução: Regina Célia Simille de Macedo. 3. ed. Rio de Janeiro: Elsevier, 2013.

SANTAELLA, Lucia; CARDOSO, Tarcísio. O desconcertante conceito de mediação técnica em Bruno Latour. *Matrizes*, São Paulo, v. 9, n. 1, p. 167-185, jan./jun. 2015. Disponível em: https://www.revistas.usp.br/matrizes/article/view/100679/99413.

SANTONI DE SIO, Filippo; MECACCI, Giulio. Four responsibility gaps with artificial intelligence: why they matter and how to address them. *Philosophy & Technology*, [*s. l.*], v. 34, n. 4, p. 1.057-1.084, May 2021. Disponível em: https://link.springer.com/article/10.1007/s13347-021-00450-x.

SCHELLEKENS, Maurice. Human-machine interaction in self-driving vehicles: a perspective on product liability. *International Journal of Law and Information Technology*, Oxford, May 2022. Disponível em: https://academic.oup.com/ijlit/advance-article-pdf/doi/10.1093/ijlit/eaac010/43760800/eaac010.pdf. Acesso em: 9 set. 2022.

SIEBERT, Luciano Cavalcante; LUPETTI, Maria Luce; AIZENBERG, Evgeni; BECKERS, Niek; ZGONNIKOV, Arkady; VELUWENKAMP, Herman; ABBINK, David; GIACCARDI, Elisa; HOUBEN, Geert-Jan; JONKER, Catholijn Maria; HOVEN, Jeroen van den; FORSTER, Deborah; LAGENDIJK, Reginald L. Meaningful human control: actionable properties for AI system development. *AI Ethics*, April 2022. Disponível em: https://link.springer.com/content/pdf/10.1007/s43681-022-00167-3.pdf.

SIMMLER, Monika; FRISCHKNECHT, Ruth. A taxonomy of human-machine collaboration: capturing automation and technical autonomy. *AI & Society*, New York, v. 36, n. 1, p. 239-250, Mar. 2021. Disponível em: https://doi.org/10.1007/s00146-020-01004-z.

TEPEDINO, Gustavo; SILVA, Rodrigo da Guia. Inteligência artificial e elementos da responsabilidade civil. *In*: FRAZÃO, Ana; MULHOLLAND, Caitlin (coord.). *Inteligência artificial e direito*: ética, regulação e responsabilidade. São Paulo: Thomson Reuters Brasil, 2019.

UNIÃO EUROPEIA. Comissão Europeia. Grupo independente de peritos de alto nível sobre Inteligência Artificial. *Orientações éticas para uma IA de confiança*. Bruxelas, 8 abr. 2019. Disponível em: https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1/language-pt/format-PDF. Acesso em: 15 set. 2022.

VINEY, Geneviève; JOURDAIN, Patrice; CARVAL, Suzanne. *Les conditions de la responsabilité*. Paris: LGDJ, 2013. (Collection Traités).

WOJTCZAK, Sylwia; KSIĘŻAK, Paweł. Causation in civil law and the problems of transparency in AI. *European Review of Private Law*, Dordrecht, v. 29, n. 4, p. 561-582, 2021.