

**V CONGRESSO INTERNACIONAL DE
DIREITO E INTELIGÊNCIA
ARTIFICIAL (V CIDIA)**

REGULAÇÃO DA INTELIGÊNCIA ARTIFICIAL III

R344

Regulação da inteligência artificial III [Recurso eletrônico on-line] organização V Congresso Internacional de Direito e Inteligência Artificial (V CIDIA): Skema Business School – Belo Horizonte;

Coordenadores: Gabriel Oliveira de Aguiar Borges, Matheus Antes Schwede e Luiz Felipe de Freitas Cordeiro – Belo Horizonte: Skema Business School, 2024.

Inclui bibliografia

ISBN: 978-85-5505-930-8

Modo de acesso: www.conpedi.org.br em publicações

Tema: Mercados globais e empreendedorismo a partir do desenvolvimento algorítmico.

1. Compliance. 2. Ética. 3. Legislação. I. V Congresso Internacional de Direito e Inteligência Artificial (1:2024 : Belo Horizonte, MG).

CDU: 34

skema
BUSINESS SCHOOL

LAW SCHOOL
FOR BUSINESS

V CONGRESSO INTERNACIONAL DE DIREITO E INTELIGÊNCIA ARTIFICIAL (V CIDIA)

REGULAÇÃO DA INTELIGÊNCIA ARTIFICIAL III

Apresentação

A SKEMA Business School é uma organização francesa sem fins lucrativos, com presença em seis países diferentes ao redor do mundo (França, EUA, China, Brasil e África do Sul e Canadá) e detentora de três prestigiadas creditações internacionais (AMBA, EQUIS e AACSB), refletindo seu compromisso com a pesquisa de alta qualidade na economia do conhecimento. A SKEMA reconhece que, em um mundo cada vez mais digital, é essencial adotar uma abordagem transdisciplinar.

Cumprindo esse propósito, o V Congresso Internacional de Direito e Inteligência Artificial (V CIDIA), realizado nos dias 6 e 7 de junho de 2024, em formato híbrido, manteve-se como o principal evento acadêmico sediado no Brasil com o propósito de fomentar ricas discussões sobre as diversas interseções entre o direito e a inteligência artificial. O evento, que teve como tema central "Mercados Globais e Empreendedorismo a partir do Desenvolvimento Algorítmico", contou com a presença de renomados especialistas nacionais e internacionais, que abordaram temas de relevância crescente no cenário jurídico contemporâneo.

Profissionais e estudantes dos cursos de Direito, Administração, Economia, Ciência de Dados, Ciência da Computação, entre outros, tiveram a oportunidade de se conectar e compartilhar conhecimentos, promovendo um ambiente de rica troca intelectual. O V CIDIA contou com a participação de acadêmicos e profissionais provenientes de diversas regiões do Brasil e do exterior. Entre os estados brasileiros representados, estavam: Pará (PA), Amazonas (AM), Minas Gerais (MG), Ceará (CE), Rio Grande do Sul (RS), Paraíba (PB), Paraná (PR), Rio de Janeiro (RJ), Alagoas (AL), Maranhão (MA), Santa Catarina (SC), Pernambuco (PE), e o Distrito Federal (DF). Além disso, o evento contou com a adesão de participantes internacionais, incluindo representantes de Portugal, França, Itália e Canadá, destacando a amplitude e o alcance global do congresso. Este encontro plural reforçou a importância da colaboração inter-regional e internacional na discussão dos temas relacionados ao desenvolvimento algorítmico e suas implicações nos mercados globais e no empreendedorismo.

Foram discutidos assuntos variados, desde a regulamentação da inteligência artificial até as novas perspectivas de negócios e inovação, destacando como os algoritmos estão remodelando setores tradicionais e impulsionando a criação de empresas inovadoras. Com

uma programação abrangente, o congresso proporcionou um espaço vital para discutir os desafios e oportunidades que emergem com o desenvolvimento algorítmico, reforçando a importância de uma abordagem jurídica e ética robusta nesse contexto em constante evolução.

A jornada teve início no dia 6 de junho com a conferência de abertura ministrada pela Professora Dr^a. Margherita Pagani, do SKEMA Centre for Artificial Intelligence, campus de Paris, França. Com o tema "Impacts of AI on Business Transformation", Pagani destacou os efeitos transformadores da inteligência artificial nos negócios, ressaltando seu impacto no comportamento do consumidor e nas estratégias de marketing em mídias sociais. O debate foi enriquecido pela participação do Professor Dr. José Luiz de Moura Faleiros Jr., da SKEMA Law School, campus de Belo Horizonte, Brasil, que trouxe reflexões críticas sobre o tema.

Após um breve intervalo, o evento retomou com o primeiro painel, intitulado "Panorama global da Inteligência Artificial". O Professor Dr. Manuel David Masseno, do Instituto Politécnico de Beja, Portugal, apresentou uma análise detalhada sobre as "práticas de IA proibidas" no novo Regulamento de Inteligência Artificial da União Europeia, explorando os limites da dignidade humana frente às novas tecnologias. Em seguida, o Professor Dr. Steve Ataky, da SKEMA Business School, campus de Montreal, Canadá, discutiu as capacidades, aplicações e potenciais futuros da IA com geração aumentada por recuperação, destacando as inovações no campo da visão computacional.

No período da tarde foram realizados grupos de trabalho que contaram com a apresentação de mais de 40 trabalhos acadêmicos relacionados à temática do evento. Com isso, o primeiro dia foi encerrado, após intensas discussões e troca de ideias que estabeleceram um panorama abrangente das tendências e desafios da inteligência artificial em nível global.

O segundo dia de atividades começou com o segundo painel temático, que abordou "Mercados globais e inteligência artificial". O Professor Dr. Edgar Gastón Jacobs Flores Filho, da SKEMA Law School, campus de Belo Horizonte, Brasil, apresentou um panorama da regulação da IA no Brasil, enquanto o Professor Dr. Fischer Stefan Meira, da SKEMA Business School, campus de Belo Horizonte, Brasil, explorou as perspectivas e desafios do desenvolvimento algorítmico.

Após breve intervalo, o terceiro painel teve início às 10:00h, focando em "Contratos, concorrência e inteligência artificial". O Professor Dr. Frédéric Marty, da Université Côte d'Azur, França, discutiu a "colusão por algoritmos", um fenômeno emergente nas políticas de concorrência, enquanto o Professor Dr. Bernardo de Azevedo e Souza, da Universidade do

Vale do Rio dos Sinos, Brasil, trouxe novas perspectivas para o empreendedorismo jurídico. A Professora Ms. Lorena Muniz e Castro Lage, SKEMA Law School, campus de Belo Horizonte, Brasil, completou o painel abordando as interseções entre startups e inteligência artificial, destacando os desafios e oportunidades para empresas inovadoras.

Durante a tarde, uma nova rodada de apresentações nos grupos de trabalho se seguiu, com 35 trabalhos acadêmicos relacionados à temática do evento sendo abordados para ilustrar a pujança do debate em torno do assunto. O segundo dia foi encerrado consolidando a importância do debate sobre a regulação e a aplicação da inteligência artificial em diferentes setores.

Como dito, o evento contou com apresentações de resumos expandidos em diversos Grupos de Trabalho (GTs), realizados on-line nas tardes dos dias 6 e 7 de junho. Os GTs tiveram os seguintes eixos de discussão, sob coordenação de renomados especialistas nos respectivos campos de pesquisa:

- a) Startups e Empreendedorismo de Base Tecnológica – Coordenado por Laurence Duarte Araújo Pereira, Maria Cláudia Viana Hissa Dias do Vale Gangana e Luiz Felipe Vieira de Siqueira.
- b) Jurimetria Cibernética Jurídica e Ciência de Dados – Coordenado por Arthur Salles de Paula Moreira, Isabela Campos Vidigal Martins e Gabriel Ribeiro de Lima.
- c) Decisões Automatizadas e Gestão Empresarial – Coordenado por Yago Aparecido Oliveira Santos, Pedro Gabriel Romanini Turra e Allan Fuezi de Moura Barbosa.
- d) Algoritmos, Modelos de Linguagem e Propriedade Intelectual – Coordenado por Vinicius de Negreiros Calado, Guilherme Mucelin e Agatha Gonçalves Santana.
- e) Regulação da Inteligência Artificial – I – Coordenado por Tainá Aguiar Junquillo, Paula Guedes Fernandes da Silva e Fernanda Ribeiro.
- f) Regulação da Inteligência Artificial – II – Coordenado por João Alexandre Silva Alves Guimarães, Ana Júlia Guimarães e Erick Hitoshi Guimarães Makiya.
- g) Regulação da Inteligência Artificial – III – Coordenado por Gabriel Oliveira de Aguiar Borges, Matheus Antes Schwede e Luiz Felipe de Freitas Cordeiro.

h) Inteligência Artificial, Mercados Globais e Contratos – Coordenado por Fernanda Sathler Rocha Franco, Gabriel Fraga Hamester e Victor Willcox.

i) Privacidade, Proteção de Dados Pessoais e Negócios Inovadores – Coordenado por Guilherme Spillari Costa, Dineia Anziliero Dal Pizzol e Evaldo Osorio Hackmann.

j) Empresa, Tecnologia e Sustentabilidade – Coordenado por Marcia Andrea Bühring, Jessica Mello Tahim e Angélica Cerdotes.

Cada GT proporcionou um espaço de diálogo e troca de experiências entre pesquisadores e profissionais, contribuindo para o avanço das discussões sobre a aplicação da inteligência artificial no direito e em outros campos relacionados.

Um sucesso desse porte não seria possível sem o apoio institucional do Conselho Nacional de Pesquisa e Pós-graduação em Direito - CONPEDI, que desde a primeira edição do evento provê uma parceria sólida e indispensável ao seu sucesso. A colaboração contínua do CONPEDI tem sido fundamental para a organização e realização deste congresso, assegurando a qualidade e a relevância dos debates promovidos. Além disso, um elogio especial deve ser feito ao trabalho do Professor Dr. Caio Augusto Souza Lara, que participou da coordenação científica das edições precedentes. Seu legado e dedicação destacam a importância do congresso e contribuem para consolidar sua reputação como um evento de referência na intersecção entre direito e inteligência artificial.

Por fim, o V Congresso Internacional de Direito e Inteligência Artificial foi, sem dúvida, um marco importante para a comunidade acadêmica e profissional, fomentando debates essenciais sobre a evolução tecnológica e suas implicações jurídicas.

Expressamos nossos agradecimentos às pesquisadoras e aos pesquisadores por sua inestimável contribuição e desejamos a todos uma leitura excelente e proveitosa!

Belo Horizonte-MG, 10 de julho de 2024.

Prof^a. Dr^a. Geneviève Daniele Lucienne Dutrait Poulingue

Reitora – SKEMA Business School - Campus Belo Horizonte

Prof. Ms. Dorival Guimarães Pereira Júnior

Coordenador do Curso de Direito – SKEMA Law School

Prof. Dr. Edgar Gastón Jacobs Flores Filho

Coordenador da Pós-Graduação da SKEMA Law School

Prof. Dr. José Luiz de Moura Faleiros Júnior

Coordenador de Pesquisa – SKEMA Law School

EXPLICABILIDADE E RESPONSABILIDADE CIVIL NOS SISTEMAS DE IA.

EXPLAINABILITY AND LIABILITY IN AI SYSTEMS

Thais Paranhos Capistrano Pereira ¹

Resumo

Este estudo analisa como garantir a explicabilidade e a responsabilidade civil em sistemas de inteligência artificial (IA), com foco em accountability. Utilizando a cirurgia robótica como exemplo, são discutidos os desafios e as soluções para tornar a IA mais transparente e responsável. Analisam-se diferenças entre modelos de caixa-branca e caixa-preta, destacando a necessidade de interfaces explicativas adaptadas. Defende-se que a accountability deve incluir deveres específicos para desenvolvedores e operadores, visando mitigar riscos e assegurar o uso ético da IA. A pesquisa enfatiza a importância de marcos regulatórios claros para proteger os direitos dos usuários e promover a confiança na tecnologia.

Palavras-chave: Explicabilidade algorítmica, Accountability, Responsabilidade civil

Abstract/Resumen/Résumé

This study explores how to ensure explainability and civil liability in artificial intelligence (AI) systems, with a focus on accountability. Using robotic surgery as an example, it discusses the challenges and solutions for making AI more transparent and responsible. It examines the differences between white-box and black-box models, highlighting the need for adapted explanatory interfaces. It argues that accountability should include specific duties for developers and operators, aimed at mitigating risks and ensuring ethical use of AI. The research emphasizes the importance of clear regulatory frameworks to protect user rights and promote trust in technology.

Keywords/Palabras-claves/Mots-clés: Algorithmic explainability, Accountability, Civil liability

¹ Analista de Direito/MPU, mestranda em Direito Constitucional Acadêmico no IDP. <http://lattes.cnpq.br/6296070588291951>

Introdução

A transição para a "sociedade da informação" impõe uma reestruturação das bases teóricas do direito, especialmente no que tange à responsabilidade civil em sistemas de inteligência artificial (IA). Inteligências artificiais não são inteligentes no sentido humano, mas processam enormes conjuntos de dados para fazer previsões probabilísticas, permanecendo distantes do comportamento expansivo, intuitivo e criativo dos seres humanos. Conforme recente definição da OCDE, sistema de IA é uma entidade baseada em computador que, para fins claramente definidos ou não, processa dados recebidos para gerar *outputs* como previsões, conteúdos, recomendações ou decisões.

A crescente adoção de sistemas de IA em áreas críticas, como a saúde, aumenta a necessidade de garantir que esses sistemas sejam transparentes e responsáveis. Exemplificando, em Boca Raton, na Flórida, em setembro de 2021, um robô cirúrgico causou a morte de uma paciente ao queimar seu intestino delgado durante um procedimento para tratar câncer de cólon. O esposo da vítima entrou com um processo judicial, alegando que a fabricante do robô Da Vinci, a Intuitive Surgical Inc., não alertou sobre os riscos associados às falhas no equipamento, o que resultou em danos irreparáveis (Ferreira, 2024). Nesse sentido, surge o problema de pesquisa: como podemos garantir a explicabilidade e a responsabilidade civil em sistemas de IA, particularmente em contextos críticos como a cirurgia robótica, onde decisões algorítmicas podem ter consequências graves para a saúde e segurança dos pacientes?

Este estudo explora a explicabilidade algorítmica e a responsabilidade civil, com foco na *accountability* e nas implicações jurídicas e éticas dessa tecnologia emergente, utilizando a cirurgia robótica como exemplo ilustrativo. Dessa forma, o objetivo principal deste estudo é explorar e definir estratégias para garantir a explicabilidade e a responsabilidade civil em sistemas de IA. Os objetivos específicos incluem analisar a diferença entre modelos de caixa-branca e caixa-preta e seus desafios para a explicabilidade, examinar o conceito de *accountability* e sua aplicação no desenvolvimento e operação de sistemas de IA, e por fim utilizar a cirurgia robótica como exemplo para ilustrar os desafios e soluções potenciais na implementação de IA explicável e responsável.

Primeiramente, será abordado os modelos de caixa-branca e caixa-preta dos sistemas de IA, e posteriormente a análise sobre o princípio da explicabilidade. Em seguida, será analisado o conceito de *accountability* e responsabilidade civil. Nas considerações finais, será feito um breve resumo enfatizando a importância de fabricantes, programadores e usuários de

sistemas de inteligência artificial estarem cientes dos riscos e colaborarem para mitigá-los, assegurando que a IA seja empregada de maneira segura e responsável, por meio da explicabilidade e *accountability*. Como abordagem metodológica, o presente estudo baseia-se em revisão bibliográfica, incluindo nas principais fontes livros e artigos científicos.

1. Explorando modelos de caixa-branca e caixa-preta: A importância do princípio da explicabilidade

A explicabilidade na IA refere-se à capacidade de um sistema de IA fornecer explicações compreensíveis sobre suas decisões e funcionamento. Para se entender melhor o conceito de explicabilidade, Herzog (2021) propõe fazer as perguntas: qual humano?, sob quais circunstâncias? e para quais propósitos? Segundo Floridi *et al.* (2020), a explicabilidade combina inteligibilidade (*intelligibility*) e *accountability*, permitindo que usuários e afetados compreendam e desafiem as interações e resultados dos sistemas de IA. A transparência é um componente essencial, exigindo a comunicação de todos os aspectos necessários para examinar ou desafiar uma IA ao longo de seu ciclo de vida.

Para tornar um sistema inteligente mais explicável, é necessário adotar abordagens que promovam a clareza, a transparência e a *accountability* em todos os aspectos da implementação, implantação e uso de modelos de IA. Isso envolve não apenas a utilização de métodos que permitam aos usuários compreender como as decisões são tomadas pelos modelos, mas também garantir que esses modelos possam ser auditados e questionados por aqueles que são afetados por suas decisões (Herzog, 2021).

Inicialmente, é importante reconhecer que tanto modelos de caixa-preta quanto de caixa-branca podem apresentar desafios para a explicabilidade. O termo "caixa branca" em IA refere-se a modelos e sistemas cujos processos internos podem ser mais facilmente entendidos por humanos. Isso contrasta com os modelos de "caixa preta", que são sistemas cuja operação interna é oculta ou não facilmente compreensível. Em modelos de caixa branca, os algoritmos usam métodos e estruturas que permitem aos usuários e desenvolvedores entender como as decisões são feitas. Isso pode incluir, por exemplo, árvores de decisão, regras lógicas, ou modelos lineares, onde é possível rastrear e entender como os *inputs* são transformados em *outputs* (Herzog, 2021).

Dessa forma, a transparência dos modelos de caixa branca é considerada uma vantagem, especialmente em áreas onde a explicabilidade e a responsabilidade são cruciais, como em aplicações médicas, financeiras, ou legais. Modelos de caixa-branca, apesar de inerentemente mais transparentes, podem ser complexos e não necessariamente intuitivos para não especialistas, exigindo interfaces de explicação que contextualizem e esclareçam seu

funcionamento interno e saídas. Portanto, mesmo modelos que são acessíveis a desenvolvedores especializados podem requerer esforços adicionais para torná-los compreensíveis para todos os usuários e partes interessadas (Herzog, 2021).

Para modelos de caixa-preta, que tradicionalmente são menos acessíveis, é essencial desenvolver ferramentas explicativas que possam traduzir suas operações ocultas em formatos compreensíveis. Isso pode incluir a implementação de visualizações de dados, como mapas de saliência em sistemas de classificação de imagens, que ajudam a ilustrar quais características do dado estão influenciando as decisões do modelo. No entanto, essas ferramentas devem ser cuidadosamente validadas para garantir que elas sejam confiáveis e não introduzam mal-entendidos ou simplificações excessivas (Herzog, 2021).

Além das ferramentas técnicas, promover a explicabilidade efetiva também envolve a criação de processos de design que considerem as implicações éticas e sociais desde o início. Isso inclui envolver uma gama diversificada de partes interessadas no processo de desenvolvimento para entender melhor as necessidades e os contextos específicos em que a IA será utilizada. Tal abordagem multidisciplinar pode ajudar a identificar e mitigar possíveis preconceitos, erros e áreas de opacidade antes que o sistema seja amplamente implementado (Herzog, 2021).

Como discutido por Herzog (2021), a explicabilidade vai além da mera interpretabilidade mecânica, exigindo interfaces adaptadas ao destinatário e ao caso de uso. Essas interfaces devem incorporar uma variedade de formas de explicações que podem não ser completas, mas que permitem aos *stakeholders* (partes interessadas) assumir responsabilidade. A flexibilidade da explicabilidade permite que ela não exija explicações mecanicistas em todos os níveis de uso, mas sim o nível necessário para servir princípios éticos como beneficência, não maleficência e autonomia.

A confusão entre interpretabilidade e explicabilidade pode levar a equívocos na implementação de sistemas de IA. Abordagens de caixa-preta têm recebido mais atenção, com esforços para fornecer saídas algorítmicas inteligíveis para usuários finais. No entanto, sistemas baseados em modelos interpretáveis, como os de caixa-branca, podem não ser inteligíveis para outros *stakeholders* além dos desenvolvedores, impedindo a responsabilidade dos usuários do sistema quando faltam interfaces de explicação (Herzog, 2021).

Fornecer explicações específicas ao domínio para sistemas de caixa-preta não necessariamente alcança responsabilidade (Herzog, 2021). Por exemplo, mapas de saliência usados para entender a classificação de imagens baseadas em IA de caixa-preta podem ser enganosos e comprometer o uso responsável, apesar de sua aparência intuitiva (Simonyan *et*

al., 2014). Por outro lado, modelos geralmente interpretáveis, como abordagens de regressão simples, não fornecem valor adicional em termos de explicabilidade quando implantados sem meios adicionais para explicar o raciocínio por trás do modelo (Herzog, 2021).

A confusão entre interpretabilidade e explicabilidade, denominada 'confusão interpretabilidade-explicabilidade', é bastante disseminada. Estudos recentes mostram que há uma tendência de equiparar modelos interpretáveis a modelos explicáveis, mesmo entre cientistas de dados altamente qualificados. A distinção entre essas funções é importante, especialmente quando as saídas algorítmicas precisam ser plausíveis para que os usuários possam desafiá-las em situações críticas, como na saúde e condução autônoma (Herzog, 2021).

Portanto, é necessário não permitir que os benefícios epistemológicos dos modelos de caixa-branca permaneçam não aproveitados devido à falta de consideração para torná-los explicáveis. A forma específica que essa transição deve tomar é contextual e depende da aplicação em consideração. Nesse sentido, Frank Pasquale (2017) introduziu a "quarta lei da robótica", que exige que os robôs indiquem a identidade de seus criadores, controladores ou proprietários. Esse princípio da explicabilidade é fundamental para a responsabilidade civil, pois facilita a atribuição de responsabilidade e a *accountability* - prestação de contas.

Em termos gerais, a propõe-se que a explicabilidade seja vista não apenas como uma característica técnica, mas como um princípio ético que sustenta todo o ciclo de vida do desenvolvimento de IA, alinhando-se, assim, às preocupações mais amplas de governança e responsabilidade social. Isso significa que a responsabilidade, a transparência e a capacidade de resposta devem ser integradas como objetivos centrais, não apenas para cumprir com regulamentações, mas para construir confiança e promover uma adoção mais ampla e segura da tecnologia de IA.

2. *Accountability* em sistemas de IA e responsabilidade civil

A *accountability* no contexto da IA envolve a criação de deveres específicos para os desenvolvedores, operadores e usuários de algoritmos, garantindo que todos os envolvidos assumam responsabilidade pelas decisões e consequências dos sistemas de IA. Em outras palavras, a *accountability ex post* envolve regras de boas práticas e governança focadas na gestão de possíveis danos (*compliance*). Portanto, considera-se a natureza, o alcance, o propósito e o grau do risco associado à implementação dos sistemas baseados em algoritmos. Esse conceito é importante em qualquer aplicação de IA, pois erros podem ter consequências

significativas. A fidedúcia entre os profissionais e a confiança na tecnologia são fundamentais para uma IA confiável (trustworthy AI) (Faleiros Júnior, 2024).

Accountability, assim, refere-se às práticas relacionadas à responsabilidade ética, à obrigação, à transparência e à prestação de contas das atividades em andamento. Em contrapartida, inclui não apenas a execução dessas atividades, mas também a explicação das razões por trás delas e como elas estão sendo realizadas (Frazão; Mulholland, 2019). Maria Celina Bodin de Moraes esquematiza a implementação de medidas eficazes que possam demonstrar a adesão e o cumprimento das regulamentações de proteção de dados pessoais, inclusive a efetividade dessas ações, indicando que “não descumprir a lei não é mais suficiente; é preciso “proativamente” prevenir a ocorrência de danos” (Moraes, 2019, p.5). Trata-se, dessa forma, da *accountability ex ante*.

Por sua vez, Jack Balkin (2017) propôs que os operadores algorítmicos têm deveres fiduciários em relação aos usuários finais, responsabilidades para com o público em geral e a obrigação de evitar incômodos algorítmicos. Nesse sentido, Coeckelbergh (2020) propõe a explicabilidade em um sentido mais amplo para incluir 'justificativas relacionais' (relational justifications), exigindo que os desenvolvedores expliquem suas decisões e ações ao longo do desenvolvimento e uso de sistemas de IA. Rosensvald (2020) acrescenta que a explicabilidade deve ser entendida como um componente multidimensional que abrange não apenas a clareza técnica, mas também a capacidade de comunicar essas explicações de maneira acessível aos diversos *stakeholders* envolvidos. A explicabilidade, nesse sentido, deve ser integrada como um princípio fundamental da responsabilidade civil, garantindo que todos os aspectos do desenvolvimento e da operação de sistemas de IA sejam transparentes e compreensíveis.

Isso é particularmente relevante em aplicações críticas, como a cirurgia robótica, onde as decisões algorítmicas devem ser nítidas para médicos e pacientes. Nesse sentido, a Resolução do Parlamento Europeu de 16 de fevereiro de 2017, recomenda a adoção de sistemas como a "caixa-preta" para robôs cirúrgicos, em consonância com as discussões sobre *accountability* e explicabilidade em IA. Esses sistemas de registro oferecem um meio para documentar com precisão todas as operações e decisões tomadas pelo robô, o que torna imprescindível para investigar e entender as causas de incidentes adversos (Nogaroli, 2023). Se o comportamento de um robô se tornar imprevisível a ponto de causar danos, isso indica que o dispositivo não oferece segurança adequada para os humanos, podendo levar a exigências legais para sua remoção do mercado. Além disso, tal imprevisibilidade pode ser vista como evidência de um defeito inerentemente perigoso no robô, situação que poderia ensejar responsabilidade objetiva por defeito do produto (Nogaroli, 2023).

Os "deveres informados por dados" (data-informed duties) propostos por Pasquale (2019) alinham-se com a necessidade de superar as "caixas-pretas" algorítmicas, promovendo a transparência e a *accountability*. Nessa linha, Giovanni Comandé (2019) destaca a importância de evoluir da responsabilidade estrita (liability) para uma responsabilidade civil que inclua a função preventiva e precaucional (riscos desconhecidos), impondo aqueles deveres "informados" aos desenvolvedores e operadores de IA.

A implementação de IA exige a consideração de vários fatores, incluindo a segurança, a precisão e a transparência das decisões algorítmicas. Utilizando a cirurgia robótica como exemplo, percebe-se a necessidade que os desenvolvedores incluam mecanismos de desligamento (backdoors) e funcionalidades de emergência (emergency brakes) para garantir a segurança dos pacientes (Nogaroli, 2023).

A supervisão contínua das inteligências não naturais, incluindo auditorias e estudos pós-mercado, é essencial para identificar e mitigar riscos (*accountability ex ante*). A obrigação de informar potenciais vítimas de comportamentos anômalos deve ser incorporada como um dever de boa-fé, alinhado ao princípio da transparência. A normatização de boas práticas e governança direcionada aos possíveis danos e seu enfrentamento é fundamental no direcionamento da responsabilidade civil.

Conclusão

A explicabilidade algorítmica e a *accountability* são essenciais para a implementação ética e segura de sistemas de IA e. A combinação de inteligibilidade e responsabilidade, aliada a uma abordagem proativa de prevenção e precaução, pode garantir que esses sistemas sejam utilizados de forma responsável, promovendo a confiança e a segurança dos usuários.

A adoção de princípios como os propostos por Pasquale, Coeckelbergh e Comandé pode fornecer uma base sólida para a regulação e a governança da IA na sociedade da informação. A pesquisa investigou como garantir a explicabilidade e a responsabilidade civil em sistemas de IA, focando em *accountability* e utilizando a cirurgia robótica como exemplo ilustrativo.

Para responder ao problema de pesquisa, conclui-se que a explicabilidade e a responsabilidade civil em sistemas de IA podem ser garantidas através da implementação de interfaces explicativas adequadas, da criação de deveres específicos para desenvolvedores, operadores e usuários, e da adoção de medidas proativas de prevenção e precaução. Em outros termos, como corroborado por Floridi *et al.* (2020) e Herzog (2021), propõe-se usar a explicabilidade como a combinação de inteligibilidade (intelligibility) e *accountability* para fundamentar a IA em modelos inerentemente interpretáveis.

Referências

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.

COECKELBERGH, Mark. Artificial intelligence, responsibility attribution, and a relational justification of explainability. *Science and Engineering Ethics*, Cham, v. 26, p. 2051-2068, 2020. p. 2066.

COMANDÉ, Giovanni. Intelligenza Artificiale e responsabilità tra liability e accountability: il carattere trasformativo dell'IA e il problema della responsabilità. In: NUZZO, Antonio; OLIVIERI, Gustavo (a cura di). *Analisi giuridica dell'Economia. Studi e discussioni sul diritto dell'impresa*. Bolonha: Il Mulino, 2019, v. 1. p. 185.

FALEIROS JÚNIOR, José Luiz de Moura. **Explicabilidade algorítmica e responsabilidade civil**. 2024. Disponível em: <https://www.migalhas.com.br/coluna/migalhas-de-responsabilidade-civil/404514/explicabilidade-algoritmica-e-responsabilidade-civil> Acesso em 12 maio 2024.

FERREIRA, Yuri. Robô teria queimado e rasgado mulher com câncer que morreu após cirurgia, diz processo. **Revista Fórum**. 2024. Disponível em: <https://revistaforum.com.br/ciencia-e-tecnologia/2024/2/10/rob-teria-queimado-rasgado-mulher-com-cncer-que-morreu-apos-cirurgia-diz-processo-153799.html> Acesso em: 12 maio 2024.

FLORIDI, L; *et al.*: AI 4People — an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. **Minds Mach.** 28, 689–707 (2018). <https://doi.org/10.1007/s11023-018-9482-5>.

FRAZÃO, Ana. MULHOLLAND, Caitlin. **Inteligência artificial e Direito: Ética, Regulação e Responsabilidade**. São Paulo: Thomson Reuters Brasil, 2019, p. 85
HERZOG, Christian. On the risk of confusing interpretability with explicability. **AI Ethics** 2, 219–225. 2021. <https://doi.org/10.1007/s43681-021-00121-9> Disponível em: <https://link.springer.com/article/10.1007/s43681-021-00121-9#citeas> Acesso em 13 maio 2024.

MORAES, Maria Celina Bodin de. LGPD: um novo regime de responsabilização dito "proativo". **Civilística**. A. 8, n. 3, 2019. Disponível em: <http://civilistica.com/lgpd-um-novo-regime-de-responsabilizacao-civil-ditoproativo/>. Acesso em: 25 abr 2024.

NOGAROLI, Rafaella. **Responsabilidade civil médica e inteligência artificial: culpa médica e deveres de conduta no século XXI**. SP: Thomson Reuters. Brasil, 2023. 337 p.

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, 2017.

ROSENVALD, Nelson. A polissemia da responsabilidade civil na LGPD. *Migalhas de Proteção de Dados*, 06 nov. 2020.

SIMONYAN, Karen; *et al.* Deep inside convolutional networks: Visualising image classification models and saliency maps. **arXiv preprint arXiv:1312.6034**, 2014. Disponível em: <https://arxiv.org/abs/1312.6034> Acesso em 13 maio 2024.