USO DA INTELIGÊNCIA ARTIFICIAL NA PREVENÇÃO DE FRAUDES FINANCEIRAS: RISCOS, DILEMAS E CAMINHOS PARA UMA GOVERNANÇA ÉTICA

Katiane do Nascimento Tavares Pinho – tavares.katiane@gmail.com
Programa de Pós-graduação em Ciência da Propriedade Intelectual – UFS/SE
Antonio Martins de Oliveira Júnior – amartins@academico.ufs.br
Programa de Pós-graduação em Ciência da Propriedade Intelectual – UFS/SE

Resumo— Este artigo caracteriza-se como uma pesquisa qualitativa de natureza bibliográfica, estruturada pela técnica de análise de conteúdo. A investigação examina seis tensões críticas no uso da inteligência artificial para prevenção de fraudes financeiras, envolvendo riscos operacionais, dilemas éticos e desafios regulatórios. Os achados demonstram que soluções exclusivamente técnicas são insuficientes e que a mitigação de riscos exige arcabouços sociotécnicos em múltiplas camadas que integrem transparência, auditoria de vieses, supervisão humana e regulação efetiva. Como contribuição, o trabalho propõe recomendações alinhadas a marcos normativos (LGPD, GDPR, *AI Act*), defendendo a integração entre inovação tecnológica, justiça, equidade, proteção de direitos e sustentabilidade para fortalecer a confiança e a legitimidade social.
Palavras-Chave— Inteligência Artificial, Ética e Regulação, Fraudes Financeiras, Sustentabilidade.

Abstract— *This article is a qualitative bibliographic study structured using content analysis. The research examines six critical tensions in the use of artificial intelligence for financial fraud prevention, involving operational risks, ethical dilemmas, and regulatory challenges. The findings demonstrate that purely technical solutions are insufficient and that risk mitigation requires multilayered socio-technical frameworks that integrate transparency, bias auditing, human oversight, and effective regulation. As a contribution, the work proposes recommendations aligned with regulatory frameworks (LGPD, GDPR, AI Act), advocating for the integration of technological innovation, justice, equity, rights protection, and sustainability to strengthen trust and social legitimacy.*
Keywords— *Artificial Intelligence, Ethics and Regulation, Financial Fraud, Sustainability.*

1 INTRODUÇÃO

A adoção da Inteligência Artificial (IA) no setor financeiro tem transformado os métodos tradicionais de análise, detecção e combate a fraudes. Com sua capacidade de processar grandes volumes de dados, identificar padrões complexos e adaptar-se à novas estratégias criminosas, técnicas avançadas de *machine learning*, *deep learning* e redes neurais permitem prever condutas suspeitas e automatizar respostas em tempo real. Esses avanços aumentam a precisão e mitigam perdas econômicas relevantes para instituições e clientes (Floridi *et al*., 2018; Mehrabi *et al*., 2021; Yoshinaga; Castro, 2023).

No entanto, à medida que a capacidade técnica da IA evolui, emergem dilemas éticos, sociais e regulatórios que ultrapassam a eficácia tecnológica e instigam debates sobre seu uso sustentável e responsável. Tais dilemas envolvem a distribuição de riscos e responsabilidades entre grupos sociais e remetem a dimensões normativas como proteção de dados, justiça procedimental e governança algorítmica (Binns, 2018; Doneda *et al*., 2018; Floridi *et al*., 2018; Wachter; Mittelstaedt; Florid, 2017).

A automação intensiva também impõe desafios ligados à privacidade, à vigilância e a vieses discriminatórios que reproduzem desigualdades (O'Neil, 2016). Ademais, a redução da supervisão humana

agrava a questão da governança ética, demandando modelos híbridos com participação humana (*human-in-the-loop*) para assegurar maior controle e justiça (Floridi *et al*., 2018).

Marcos regulatórios internacionais (*Artificial Intelligence-AI Act*, Regulamento Geral de Proteção de Dados-GDPR) e nacionais (Lei Geral de Proteção de Dados-LGPD) buscam enfrentar essas tensões ao estabelecer direitos e deveres no tratamento de dados e reforçar transparência, finalidade e segurança. Ademais, orientam o desenvolvimento da IA segura, ética e confiável, protegendo direitos fundamentais, a democracia e o Estado de direito, sem deixar de fomentar a inovação (Brasil, 2016; União Europeia, 2016; 2024).

Apesar do crescente uso da IA em sistemas antifraude, observa-se a escassez de estudos que integrem aspectos éticos, regulatórios e de sustentabilidade tecnológica em um modelo analítico único, prevalecendo abordagens centradas no desempenho algorítmico e na eficiência operacional. Essa lacuna limita a compreensão dos riscos emergentes e das condições para a adoção responsável da tecnologia, sobretudo em contextos que exigem elevados níveis de confiança pública e de legitimidade institucional.

Nessa perspectiva, o estudo examina a literatura sobre as tensões do uso de IA na prevenção de fraudes financeiras, articulando dimensões técnicas e socioinstitucionais para oferecer um referencial que oriente práticas organizacionais e políticas públicas. Ao integrar contribuições nacionais e internacionais, evidencia os desafios e descortina caminhos que conciliem inovação, transparência e proteção de direitos, promovendo expansão justa, auditável e inclusiva da IA no setor financeiro (Vinuesa *et al*., 2020; Floridi *et al*., 2018).

2 REFERENCIAL TEÓRICO

O uso da IA na prevenção de fraudes financeiras é permeada por tensões que ultrapassam o campo técnico e envolvem dilemas de justiça, poder, privacidade e responsabilidade. Este referencial teórico reúne contribuições que evidenciam como o uso da IA, ao mesmo tempo em que promove inovação, também gera riscos que desafiam a governança, a legitimidade institucional e a proteção de direitos.

2.1 EFICÁCIA ALGORÍTMICA VS. EXPLICABILIDADE: O DILEMA DA "CAIXA PRETA"

Uma das principais tensões do uso da IA em diferentes setores, incluindo o financeiro, está no paradoxo entre desempenho e transparência. Conhecida como "dilema da caixa-preta", essa tensão evidencia o conflito entre a eficiência dos sistemas de IA e a necessidade de compreender e justificar suas decisões (Hermitaño Castro, 2022).

Modelos de IA aplicados às finanças costumam apresentar estruturas de difícil interpretação. Arquiteturas de *machine learning* e *deep learning*, ao lidar com grandes volumes de dados e relações não lineares, alcançam bons resultados na detecção de fraudes e análise de crédito, mas apresentam dificuldade na rastreabilidade das decisões, comprometendo a transparência necessária para auditoria e contestação (Doshi-Velez; Kim, 2017; Wachter; Mittelstaedt; Florid, 2017).

Esse cenário evidencia um *trade-off* entre interpretabilidade e performance. Modelos mais simples, como regressões lineares e logística, são mais transparentes na explicação dos resultados, mas limitados na captura de relações complexas. Já algoritmos sofisticados, como redes neurais ou modelos de múltiplas árvores, alcançam alta eficácia na detecção de anomalias, porém funcionam como "caixas-pretas", dificultando a responsabilização e a confiança dos usuários (Goodman; Flaxman, 2016; Floridi *et al*., 2018).

No setor financeiro, onde os impactos das decisões tendem a ser imediatos, cresce a demanda por modelos interpretáveis e auditáveis que assegurem conformidade ética e regulatória. A literatura destaca que a transparência é condição para a confiança pública e para a validação das decisões automatizadas, o que tem impulsionado iniciativas de *Explainable Artificial Intelligence* (XAI), buscando traduzir os processos internos desses sistemas em informações compreensíveis a diferentes públicos (Carvalho, 2021; Hermitaño Castro, 2022; União Europeia, 2020; 2024).

Wachter, Mittelstaedt e Floridi (2017) reforçam a importância da explicabilidade algorítmica para garantir transparência e responsabilização, alertando que a ausência desse recurso pode gerar decisões opacas, limitando a confiança dos usuários e dificultando a responsabilização em contextos regulatórios. No Brasil, Da Silveira (2020) enfatiza os riscos da opacidade algorítmica para os consumidores, especialmente em

sistemas financeiros, nos quais as decisões produzem efeitos diretos e imediatos.

Destarte, a explicabilidade algorítmica não deve ser vista como um elemento acessório, mas como requisito necessário para assegurar *accountability* e conformidade regulatória em contextos de alto risco (Doshi-Velez; Kim, 2017; Wachter; Mittelstaedt; Florid, 2017).

2.2 EFICIÊNCIA PREDITIVA VS. JUSTIÇA ALGORÍTMICA: A SOMBRA DOS VIÉS

A busca por maior precisão nos sistemas de IA pode resultar em decisões enviesadas e sistemas injustos. Isso ocorre porque tais sistemas operam por meio de algoritmos que aprendem a partir da observação e análise de bases de dados em vez de regras pré-instruídas, tornando a qualidade das saídas (*output*) dependente da representatividade dos dados de entrada (*input*). Quando os dados históricos carregam preconceitos ou desigualdades, os modelos tendem a reproduzir e até amplificar essas distorções, o que leva à manifestação de vieses algorítmicos com potenciais efeitos discriminatórios (Carvalho, 2021; Doneda *et al*., 2018; Tosta; Dias, 2025).

Conforme Zuboff (2020), algoritmos funcionam com base em uma lógica extrativa que transforma dados comportamentais em insumos mercadológicos, contribuindo para a reprodução de desigualdades e o fortalecimento de mecanismos de controle. Nessas condições, sistemas automatizados podem restringir o acesso a crédito, dificultar a inclusão produtiva e acentuar vulnerabilidades preexistentes (Bruno, 2019; Eubanks, 2018).

Corroborando com esse pensamento, Cremonez *et al.* (2025) enfatizam que a IA carrega consigo o potencial de reproduzir e amplificar desigualdades sociais, perpetuando vieses e marginalizando grupos minoritários. Segundo os autores, algoritmos de categorização e filtragem de informações aumentam as chances de reforço aos estereótipos históricos.

O'Neil (2016), em sua obra seminal intitulada "*Weapons of Math Destruction*", alerta para o perigo dos sistemas automatizados que reforçam e amplificam as discriminações existentes, mascarando-as sob a aparência de objetividade estatística. Esses modelos ampliam desigualdades podendo impactar sobretudo grupos marginalizados: seus vieses podem reproduzir hierarquias de raça, gênero e classe, influenciam decisões em emprego, crédito, serviços públicos, educação e justiça criminal e, assim, ameaçam autonomia, igualdade e dignidade (Doneda *et al*., 2018; Cremonez *et al.*, 2025; União Europeia, 2024).

Emergindo, assim, a necessidade de criar, avaliar e mitigar vieses e discriminações causadas por algoritmos, a chamada justiça algorítmica. A literatura reconhece a justiça algorítmica como um conceito multifacetado, dependente do contexto de aplicação e dos valores sociais que o orientam (Mehrabi *et al*., 2021). Nesse sentido, defende-se que esta deve constituir eixo estruturante nas políticas de governança da IA, para que sua adoção não reforce desigualdades históricas, contribua para mitigá-las (Kyrillos, 2024).

Mecanismos de auditoria e governança em múltiplos níveis são essenciais para equilibrar eficiência técnica e responsabilidade social. Soma-se a isso a necessidade de supervisão humana e de políticas de correção de vieses, de modo a evitar que dados enviesados reforcem injustiças (Gasser; Almeida, 2017; Mehrabi *et al*., 2021).

2.3 AUTOMAÇÃO VS. SUPERVISÃO HUMANA

O uso de sistemas de IA antifraude traz à tona o debate sobre os limites da automação e a necessidade de supervisão humana. Embora modelos de aprendizado profundo apresentem alta capacidade de processamento e detecção de padrões complexos, prometendo ganhos de eficiência, escalabilidade e precisão na detecção de anomalias, a literatura alerta que a exclusão do fator humano pode resultar em decisões opacas e suscetíveis a vieses estruturais (Mehrabi *et al.*, 2021; Schmidhuber, 2015).

Instituições financeiras têm adotado a automação para ganhar eficiência e reduzir custos, diminuindo a intervenção humana em tarefas de grande volume. Porém, em decisões críticas, como bloqueios de conta ou concessão de crédito, essa prática suscita questionamentos sobre o nível de autonomia concedido à IA. Embora

acelere processos, a ausência de julgamento contextual pode resultar em decisões injustas, sobretudo na ausência de mecanismos de revisão ou contestação (Bi; Xiao; Teng, 2025; Yoshinaga; Castro, 2023).

A literatura alerta para os riscos da automação irrestrita, já descritos por Bainbridge (1983) em sua obra seminal "*Ironias da Automação"*, enfatizando que a ausência de supervisão humana pode reproduzir desigualdades e comprometer direitos fundamentais.

Em sistemas de IA esses riscos se ampliam, tendo em vista que a exclusão do analista humano do ciclo decisório pode gerar o efeito *out-of-the-loop*, em que a perda de consciência situacional torna a intervenção ineficaz nos momentos críticos. Destarte, a falta de supervisão configura não apenas falha operacional, mas vulnerabilidade sistêmica com potenciais danos financeiros, éticos e reputacionais (Doneda *et al.*, 2018; Fügener *et al*., 2022; Floridi *et al*., 2018; Tong; Lee, 2023).

Diante do exposto, defende-se a adoção da abordagem *human-in-the-loop*, que alia a eficiência da IA com a supervisão crítica humana, permitindo revisar decisões automatizadas, corrigir erros e incorporar valores humanos, promovendo justiça sobretudo no setor financeiro, onde falhas podem gerar impactos econômicos significativos na vida dos indivíduos (Mosqueira-Rey *et al*., 2022; Tong; Lee, 2023; Xu *et al*., 2023).

No plano normativo, o Livro Branco da IA e o *AI Act* apresentam a supervisão humana como princípio estruturante para aplicações de alto risco, impondo rastreabilidade, auditabilidade e explicabilidade, em consonância com estudos que questionam a eficácia do direito à explicação diante de segredos comerciais e da complexidade algorítmica (Goodman; Flaxman, 2016; União Europeia, 2020; 2024; Wachter; Mittelstadt; Floridi, 2017).

Desta forma, a tensão entre automação e supervisão humana evidencia a necessidade de modelos híbridos que unam a capacidade analítica da IA à avaliação crítica humana, garantindo eficiência na detecção de fraudes sem perder a conformidade ética, regulatória e social. A sustentabilidade desses sistemas requer arranjos institucionais que conciliem inovação e proteção de direitos, assegurando privacidade e monitoramento proporcional de dados.

2.4 PRIVACIDADE E VIGILÂNCIA DE DADOS

A privacidade é um direito fundamental que protege os dados pessoais contra usos indevidos e monitoramento constante, assegurando a autonomia individual no ambiente digital**.** Mais que uma garantia individual, trata-se de um direito multifacetado que também regula práticas estruturais de coleta, tratamento e circulação de informações (Brasil, 1996; Solove, 2006).

No setor financeiro, sistemas de IA dependem da análise contínua de grandes volumes de dados sensíveis, ampliando riscos à privacidade e à autodeterminação informacional. Ao processar dados pessoais, transacionais e comportamentais de fontes interconectadas, esses sistemas podem ultrapassar a finalidade original antifraude e derivar para vigilância em larga escala (Da Silveira, 2020; Bruno, 2015; Doneda *et al*., 2018; Floridi *et al.*, 2018; Solove, 2006; Rouvroy; Berns, 2013).

Essa dinâmica remete à governamentalidade algorítmica, em que dados classificam indivíduos e antecipam comportamentos, muitas vezes de forma opaca. Embora eficaz em segurança e crédito, essa lógica preditiva pode reforçar desigualdades e reduzir a transparência dos critérios nas decisões automatizadas (Rouvroy; Berns, 2013).

Bruno (2015) aponta que tais práticas integram um regime de tecnopolítica da vigilância: o monitoramento de dados atua como gestão social pouco visível, suscitando debates sobre privacidade e limites ético-jurídicos da vigilância algorítmica. Essas controvérsias se agravam com o uso de dados de redes sociais e a circulação transnacional de informações para treinar IA em contextos de proteção desigual, gerando dilemas de consentimento, finalidade e segurança (Doneda *et al*., 2018; Rouvroy; Berns, 2013).

Desta forma, no uso da IA para a prevenção de fraudes, a busca por eficácia técnica pode colidir com a proteção da privacidade e a governança ética, tensionando acesso massivo a dados e direitos fundamentais. Floridi *et al*. (2018), enfatiza que o uso aceitável da IA deve seguir princípios de justiça, bem-estar e dignidade, impondo limites à coleta e circulação de dados.

Do ponto de vista regulatório, no Brasil, a Lei Geral de Proteção de Dados (LGPD) estabelece os direitos e deveres no tratamento de dados, reforçando princípios como transparência, finalidade e segurança

(Brasil, 2016). Na Europa, o Regulamento Geral de Proteção de Dados (GDPR) tornou-se referência global ao assegurar a autodeterminação informacional e impor limites à coleta e uso de dados pessoais (União Europeia, 2016; Goodman; Flaxman, 2016). Ambos os marcos visam garantir transparência e responsabilização no uso de tecnologias com decisões automatizadas.

Embora a LGPD e o GDPR busquem responder aos desafios existentes por meio de princípios como finalidade, necessidade e proporcionalidade, persistem incertezas quanto ao alcance do direito à explicação e às limitações impostas por segredos comerciais, o que limita a efetividade das salvaguardas diante do monitoramento em larga escala. (Brasil, 2018; União Europeia, 2016; Goodman; Flaxman, 2016; Wachter; Mittelstadt; Floridi, 2017).

Desta forma, a relação entre privacidade, vigilância e dados em sistemas de IA antifraude deve equilibrar os ganhos técnicos da análise de grandes volumes de informações com os riscos sociais e normativos. O desafio é desenvolver modelos capazes de prevenir fraudes sem comprometer a privacidade, assegurando uma governança pautada no uso ético, responsável e sustentável da tecnologia.

2.5 INCLUSÃO DIGITAL VS SUSTENTABILIDADE TECNOLÓGICA

Com o avanço da adoção da IA, emerge a tensão da inclusão digital. Vinuesa *et al*. (2020) vinculam o uso responsável da IA aos Objetivos de Desenvolvimento Sustentável (ODS), ressaltando a necessidade de políticas públicas que assegurem o acesso equitativo às tecnologias e evitem a ampliação das desigualdades. Já Morozov (2013) critica o solucionismo tecnológico, lembrando que a tecnologia não é neutra e que sua aplicação sem considerar os contextos sociais e econômicos pode gerar novos "tecno-débitos" sociais.

A implementação de tecnologias de IA de alta performance, como *Deep Learning* e Redes Neurais Gráficas (GNNs), requer investimentos elevados em infraestrutura, capacitação técnica e dados de qualidade. Esses requisitos dificultam o acesso de pequenas empresas, como *fintechs* em estágio inicial e instituições de microfinanças em países em desenvolvimento, ampliando o risco de concentrar o poder tecnológico e os benefícios da inovação em grandes conglomerados financeiros (Schmidhuber, 2015; Sorj; Guedes, 2005).

A sustentabilidade tecnológica envolve tanto o uso da IA para apoiar metas sociais e ambientais quanto os custos energéticos e materiais de sua aplicação. A União Europeia reconhece os impactos do setor de Tecnologia da Informação e Comunicação (TIC) nas emissões de gases de efeito estufa e no elevado consumo dos centros de dados, defendendo uma inovação responsável que reduza danos ambientais e distribua benefícios de forma equitativa (União Europeia, 2020; 2024; Vinuesa *et al*., 2020).

Além disso, a relação entre inclusão digital e sustentabilidade tecnológica, em sistemas de IA antifraude, envolve equilibrar acesso, proteção de direitos e mitigação de impactos socioambientais. Isso requer transparência, justiça e responsabilização na distribuição dos benefícios, em consonância com modelos de governança em múltiplas camadas que alinham inovação a valores democráticos e inclusivos (Larsson, 2020; Floridi *et al*., 2018; Gasser; Almeida, 2017).

Sorj e Guedes (2005) apontam que a exclusão digital vai além da infraestrutura, refletindo desigualdades culturais e sociais que podem gerar marginalização financeira e comprometer a legitimidade da inovação. Nessa perspectiva, Floridi *et al.* (2018) e Vinuesa *et al.* (2020) defendem que a governança algorítmica da IA deve ser pautada pela diversidade, transparência e *accountability*, assegurando que o progresso tecnológico seja compatível com justiça social e inclusão.

2.6 GOVERNANÇA ALGORÍTMICA E *ACCOUNTABILITY*

A responsabilização dos sistemas de IA é um tema central para garantir a confiança e a regulação adequada do uso ético dessa tecnologia. A autonomia dos sistemas de IA, em especial os modelos de "caixa-preta", impõe um desafio fundamental aos mecanismos tradicionais de supervisão, gerando a necessidade de estruturas robustas de governança e responsabilização (*accountability*) (Lepri *et al*., 2018; Shah, 2018).

A governança algorítmica refere-se ao conjunto de políticas, processos e controles que regulam o desenvolvimento, a implementação e o monitoramento de sistemas de IA, garantindo que operem de forma

segura, ética e alinhada aos objetivos organizacionais e sociais (Kroll *et al*., 2015). Já a *accountability*, por sua vez, é a capacidade de atribuir responsabilidade pelas decisões e resultados desses sistemas, assegurando que haja reparação em caso de danos (Diakopoulos, 2016; Kumar *et al*., 2024).

A governança em sistemas de IA requer princípios como transparência, explicabilidade, robustez, segurança e justiça, sustentados por um arcabouço socio-técnico em camadas. A primeira camada é a IA Explicável (XAI), que oferece clareza sobre decisões algorítmicas; a segunda, a auditoria de vieses, que busca identificar e corrigir distorções discriminatórias; e a terceira, a supervisão humana (*human-in-the-loop*), que assegura controle contínuo, permitindo validar e ajustar o sistema em consonância com valores éticos e o contexto de aplicação (Mehrabi *et al*., 2021; Cremonez *et al*., 2025; Fügener *et al*., 2022; Tong; Lee, 2023).

Contudo, mesmo com uma governança bem implementada, persiste a controvérsia sobre a responsabilização jurídica (*accountability*) quando um sistema autônomo causa dano. A complexidade dos algoritmos pode criar uma indefinição de culpabilidade, dificultando a identificação de um nexo causal claro (Calo, 2017; Pasquale, 2015).

No Brasil, a ausência de uma lei específica sobre IA leva à aplicação de regimes jurídicos já existentes. Nas relações de consumo, como nos serviços financeiros, prevalece a responsabilidade objetiva prevista no Código de Defesa do Consumidor, que dispensa a comprovação de culpa da instituição. Fora desse escopo, aplica-se a responsabilidade subjetiva do Código Civil, que exige prova de dolo ou culpa, um ônus probatório difícil para a vítima diante da opacidade algorítmica (De Almeida Junior; Reinas, 2024).

A LGPD, em seu artigo 20, assegura ao titular o direito à revisão de decisões automatizadas, representando um avanço processual para a *accountability*. A regulamentação deste artigo e a tramitação de projetos de lei como o PL 2.338/23 (que dispõe sobre o uso da IA) buscam preencher lacunas existentes, alinhando o Brasil a movimentos regulatórios internacionais, como o *AI Act* da União Europeia (Brasil, 2018; 2023; União Europeia, 2024).

3 METODOLOGIA

Este estudo caracteriza-se como uma revisão bibliográfica de natureza qualitativa, com finalidade exploratório-descritiva. A pesquisa foi conduzida com o objetivo de analisar e sintetizar o conhecimento científico e técnico sobre as tensões éticas e de sustentabilidade do uso de IA na prevenção de fraudes financeiras.

A coleta de dados foi realizada por meio de buscas no Portal de Periódicos da CAPES, *Web of Science*, Scopus e Cielo, assegurando acesso a publicações nacionais e internacionais. Foram selecionadas obras seminais para fundamentação, além de estudos recentes que refletissem os avanços mais atuais da área. A estratégia de busca utilizou uma combinação de descritores em português e inglês, incluindo: “Inteligência Artificial” AND “fraude financeira”; “ética algorítmica”; “sustentabilidade digital”; “viés algorítmico” AND “setor financeiro”; “IA explicável” AND “finanças”; “*human-in-the-loop*” AND “*fraud detection*”; “*algorithmic accountability*”; e “regulação de IA” AND “Brasil”.

Os critérios de inclusão contemplaram: (1) artigos publicados em periódicos revisados por pares, anais de conferências e relatórios técnicos de reconhecida relevância; (2) período de publicação entre 2015 e 2025, de modo a contemplar tanto obras seminais quanto estudos recentes; (3) textos em português, inglês ou espanhol; e (4) pertinência temática, abordando ao menos uma das controvérsias centrais do estudo (explicabilidade, viés algorítmico, privacidade, supervisão humana, sustentabilidade ou responsabilização). Foram excluídos trabalhos de caráter exclusivamente técnico, sem discussão ética ou social.

A análise do material foi realizada pela técnica de análise de conteúdo (Bardin, 2015), cujo propósito é interpretar criticamente o sentido das comunicações, sejam explícitas ou latentes (Chizzotti, 2006). Seguindo suas três etapas (pré-análise, exploração do material e tratamento dos resultados), os dados foram organizados em torno das seis tensões teóricas que estruturam o referencial do estudo. A síntese buscou não apenas descrever essas tensões, mas também discutir suas interconexões, oferecendo uma visão crítica e integrada do problema.

4 RESULTADOS E DISCUSSÕES

A análise da literatura evidencia que as tensões associadas ao uso da inteligência artificial (IA) na prevenção de fraudes financeiras não são fenômenos isolados, mas um sistema de controvérsias interconectadas. A busca pela eficiência técnica desencadeia dilemas éticos, jurídicos e sociais que se retroalimentam, configurando um ecossistema de riscos e responsabilidades.

Foram identificadas seis tensões centrais no uso da IA para prevenção de fraudes financeiras. O Quadro 1 resume os riscos, implicações e soluções propostas, articulando-os a princípios normativos como a LGPD, o GDPR e o *AI Act*, de modo a evidenciar caminhos para uma governança ética e sustentável.

Quadro 1 - Tensões no uso da IA antifraude em finanças: riscos, implicações e soluções propostas

Tensão	**Riscos Identificados**	**Implicações**	**Soluções/Recomendações**	**Princípios normativos (LGPD, GDPR, *AI Act*)**
Eficácia Algorítmica vs. Explicabilidade	Opacidade das decisões; dificuldade de auditoria	Limitação da contestação e da responsabilização	Adoção de *Explainable AI* (XAI) e logaritmos auditáveis	Transparência, direito à explicação, *accountability*
Eficiência Preditiva vs. Justiça algorítmica	Reforço de discriminações históricas; exclusão social	Acesso desigual a crédito e serviços financeiros	Auditorias de viés, métricas de justiça e supervisão humana	Não discriminação, equidade, justiça social
Automação vs. Supervisão Humana	Decisões opacas em situações críticas; falhas não detectadas	Vulnerabilidade sistêmica; perda de confiança	Modelos híbridos (*human-in-the-loop*); revisão e contestação	Supervisão humana, rastreabilidade, auditabilidade
Privacidade e Vigilância de Dados	Coleta excessiva de dados, desvio de finalidade e violação da autodeterminação informacional.	Privacidade fragilizada, direitos fundamentais tensionados e queda da confiança pública.	Minimização/finalidade de dados; comitês de ética/auditoria; transparência e consentimento; conformidade com LGPD, GDPR e AI Act.	Proteção de dados, finalidade, necessidade, proporcionalidade, *accountability*
Inclusão Digital vs. Sustentabilidade Tecnológica	Exclusão de pequenas *fintechs*; custos energéticos elevados	Concentração de poder e desigualdade no acesso	Políticas públicas de inclusão digital e métricas de sustentabilidade	Inclusão, justiça distributiva, inovação responsável
Governança Algorítmica e *Accountability*	Lacuna de responsabilidade em caso de danos	Incerteza jurídica; fragilidade da responsabilização	Marcos regulatórios claros; atribuição de responsabilidades	*Accountability*, segurança, robustez

Fonte: os autores, (2025)

Os achados revelam que a eficácia algorítmica, sustentada por modelos robustos como redes neurais e *deep learning*, intensifica o chamado dilema da "caixa-preta", cuja opacidade dificulta a transparência e compromete a auditabilidade, limitando a possibilidade de contestação e a responsabilização. Essa falta de clareza se configura não apenas como um entrave técnico, mas também como a base sobre a qual emergem vieses algorítmicos: decisões enviesadas decorrentes de dados historicamente desiguais, que reforçam exclusões sociais e econômicas.

A mesma lógica de opacidade se articula com a expansão da vigilância financeira. A coleta em massa e contínua de dados pessoais, transacionais e comportamentais coloca em rota de colisão o direito fundamental à privacidade. Assim, o ganho em precisão técnica pode representar perda de autonomia informacional, instaurando a necessidade de práticas de governamentalidade algorítmica e vigilância estrutural.

Nesse contexto, a literatura converge para a necessidade de mecanismos com supervisão humana (*human-in-the-loop*) que, longe de ser um retrocesso à automação, é uma abordagem que surge para equilibrar eficiência técnica e justiça social. Enfatizando que a supervisão humana fortalece a capacidade de revisão crítica e correção de erros, reduzindo a assimetria de poder entre instituições financeiras e consumidores.

No plano da sustentabilidade tecnológica, observa-se que os custos elevados de desenvolvimento, infraestrutura e energia para treinar modelos avançados concentram poder em grandes corporações, ampliando barreiras de entrada para *fintechs* emergentes e instituições de menor porte. Esse cenário aprofunda desigualdades globais e compromete a democratização do acesso às inovações tecnológicas.

Por fim, verifica-se que esse ecossistema opera sob incerteza jurídica, pois a autonomia dos sistemas de IA desafia os regimes tradicionais de responsabilização, resultando em lacuna de responsabilidade em caso de danos. No Brasil, persistem indefinições sobre a aplicação da responsabilidade, ao passo que o debate legislativo, embora em avanço (PL 2.338/23), ainda segue em ritmo mais lento que a evolução tecnológica.

Em síntese, os resultados indicam que respostas exclusivamente técnicas não são suficientes, sendo necessária uma integração entre IA Explicável (XAI), auditorias de viés, supervisão humana e marcos regulatórios claros para a mitigação dos riscos e diminuição das tensões existentes. Mais que dilemas técnicos, as controvérsias aqui discutidas representam tensões epistemológicas e normativas que exigem uma governança algorítmica capaz de alinhar inovação tecnológica aos princípios de justiça, transparência, responsabilidade e sustentabilidade, o que reforça a necessidade de um marco legal específico para assegurar responsabilização e confiança no ecossistema financeiro digital.

5 CONCLUSÃO

A adoção da inteligência artificial na prevenção de fraudes financeiras, embora traga ganhos técnicos, gera tensões éticas, sociais e regulatórias que não podem ser resolvidas apenas por soluções tecnológicas. Eficácia algorítmica, vieses de dados, vigilância ampliada, sustentabilidade e incertezas jurídicas configuram um ecossistema de riscos que exige respostas integradas. Recomenda-se, assim, o fortalecimento da *Explainable AI* (XAI), auditorias de vieses, modelos híbridos com supervisão humana, maior proteção de dados alinhada a marcos regulatórios como LGPD, GDPR e *AI Act*, além de políticas públicas que promovam inclusão digital e assegurem benefícios distribuídos de forma equitativa e sustentável.

Para pesquisas futuras, recomenda-se estudos empíricos que avaliem a aplicação prática das recomendações em instituições financeiras; análises comparativas entre marcos regulatórios e a legislação brasileira; investigações sobre métricas de sustentabilidade digital; e abordagens interdisciplinares que integrem ciência de dados, direito, ética e ciências sociais na construção de modelos de governança mais robustos.

REFERÊNCIAS

ALVES, Marco A. S.; ANDRADE, Otávio M. Da “caixa-preta” à “caixa de vidro”: o uso da *Explainable Artificial Intelligence* (XAI) para reduzir a opacidade e enfrentar o enviesamento em modelos algorítmicos. ***Revista de Direito Público***, Brasília, v. 18, n. 100, 2022. Disponível em: https://www.portaldeperiodicos.idp.edu.br/direitopublico/article/view/5973. Acesso em: 18 ago. 2025.

BAINBRIDGE, Lisanne. Ironies of automation. In: *Analysis, design and evaluation of man–machine systems*. Pergamon, 1983. P. 129-135.

BARDIN, Laurence. *Análise de conteúdo*. São Paulo: Edições 70, 2015.

BI, Shuochen; XIAO, Jue; DENG, Tingting. The Role of AI in Financial Forecasting: ChatGPT’s Potential and Challenges. In The 4th Asia-Pacific Artificial Intelligence and Big Data Forum (AIBDF 2024), 27–29, 2024, Ganzhou, China. ***ACM***, New York, USA, 7 p. Disponível em: https://doi.org/ 10.1145/3718491.3718663. Acesso em: 14 ago. 2025.

BINNS, Reuben. Justiça no aprendizado de máquina: Lições da filosofia política. In: *Conferência sobre justiça, responsabilização e transparência.* P. 149-159, PMLR, 2018.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília, DF: Presidência da República, 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709compilado.htm. Acesso em: 08 ago. 2025

_____. Senado Federal. Projeto de Lei (PL) no 2338/2023. Dispõe sobre o uso da Inteligência Artificial. Disponível em: http://www25.senado.leg.br/web/atividade/materias/-/materia/157233. Acesso em: 17 ago. 2025.

BRUNO, Fernanda; BEZERRA, Júlio; MILANI, Wilson. Tecnopolíticas e vigilância. ***Revista Eco-Pós***, v. 2, p. 1-7, 2015.

CALO, Ryan. Artificial Intelligence Policy: A Primer and Roadmap. ***Revista Eletrônica SSRN***. Disponível em: http://dx.doi.org/10.2139/ssrn.3015350. Acesso em: 09 ago. 2025.

CARVALHO, André Carlos Ponce de Leon Ferreira. Inteligência Artificial: riscos, benefícios e uso responsável. ***Estudos Avançados***, São Paulo, Brasil, v. 35, n. 101, p. 21–36, 2021. Disponível em: https://revistas.usp.br/eav/article/view/185020. Acesso em: 18 ago. 2025.

CHIZZOTTI, Antonio. *Pesquisa em ciências humanas e sociais*. 8. ed. São Paulo: Cortez, 2006.

CREMONEZ, Pedro; BARIZON Filho, Antonio; ALMEIDA, Patrícia O. P. de. Algoritmos da Exclusão: Inteligência Artificial e as Hierarquias Invisíveis da Organização do Conhecimento. ***ISKO Brasil***, Canela, RS, 2025. DOI: 10.22477/ISKO2025.55. Acesso em: 10 ago. 2025.

DA SILVEIRA, Sergio Amadeu. Sistemas algorítmicos, subordinação e colonialismo de dados. ***Algoritarismos***, p. 158-170, 2020.

DE ALMEIDA JUNIOR, Jesualdo Eduardo; REINAS, Caroline Pastri Pinto. Responsabilidade civil de algoritmos. ***Revista de Direito Civil Contemporâneo***, v. 36, p. 151-173, 2024. Disponível em: https://ojs.direitocivilcontemporaneo.com/index.php/rdcc/article/view/1335. Acesso em 08 ago. 2025.

DIAKOPOULOS, Nicholas. Accountability in algorithmic decision making. ***Communications of the ACM***, v. 59, n. 2, p. 56-62, 2016.

DONEDA, Danilo C. M. et al. Considerações iniciais sobre inteligência artificial, ética e autonomia pessoal. ***Pensar-Revista de Ciências Jurídicas***, v. 23, n. 4, p. 1-17, 2018. DOI: 10.5020/2317-2150.2018.8257. Acesso em: 08 ago. 2025.

DOSHI-VELEZ, Finale; KIM, Been. Toward a rigorous science of interpretable machine learning. arXiv preprint, arXiv:1702.08608, 2017. Disponível em: https://arxiv.org/abs/1702.08608. Acesso em: 03 ago. 2025.

EUBANKS, Virginia. Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. New York: ***St. Martin's Press***, 2018.

FLORIDI, Luciano *et al*. AI4People – An ethical framework for a good AI society: opportunities, risks, principles, and recommendations. ***Minds and Machines***, v. 28, p. 689–707, 2018. DOI: 10.1007/s11023-018-9482-5. Acesso em: 10 ago. 2025.

GASSER, U.; ALMEIDA, V. A. F. A Layered Model for AI Governance, In: IEEE Internet Computing, vol. 21, n. 6, pp. 58-62, 2017. DOI: 10.1109/MIC.2017.4180835. Acesso em: 10 ago. 2025.

GOODMAN, B.; FLAXMAN, S. European Union Regulations on Algorithmic Decision-Making and a 'Right to Explanation.' ***AI Magazine***, vol. 38, no. 3, Association for the Advancement of Artificial Intelligence, 2017, pp. 50–57. DOI: 10.1609/aimag.v38i3.2741. Acesso em: 11 ago. 2025.

HERMITAÑO CASTRO, J. A. Aplicación de Machine Learning en la Gestión de Riesgo de Crédito Financiero: Una revisión sistemática. ***Interfases***, v. 15, n. 015, p. 160-178, 2022. Disponível em: https://doi.org/10.26439/interfases2022.n015.5898. Acesso em: 10 ago. 2025.

KROLL, Joshua Alexander. ***Algoritmos responsáveis***. 2015. Tese de Doutorado. Universidade de Princeton.

KYRILLOS, G. M. Interseccionalidade: proposta de um mapa teórico provisório. ***Revista Estudos Feministas***, Florianópolis, v. 32, n. 2, 2024. Disponível em: doi.org/10.1590/1806-9584-2024v32n290290. Acesso em: 16 ago. 2025.

KUMAR, Jambi Ratna Raja et al. Transparência na tomada de decisão algorítmica: modelos interpretáveis para responsabilização ética. E3S Web of Conferences. ***EDP Sciences***, p. 02041, 2024.

LARSSON, Stefan. On the governance of artificial intelligence through ethics guidelines. ***Asian Journal of Law and Society***, v. 7, n. 3, p. 437-451, 2020. Disponível em: https://doi.org/10.1017/als.2020.19. Acesso em: 16 ago. 2025.

LEPRI, Bruno et al. Processos de tomada de decisão algorítmica justos, transparentes e responsáveis: a premissa, as soluções propostas e os desafios em aberto. ***Filosofia & Tecnologia***, v. 31, n. 4, p. 611-627, 2018.

MEHRABI, Ninareh *et al*. A survey on bias and fairness in machine learning. ***ACM Computing Surveys (CSUR)***, v. 54, n. 6, p. 1–35, 2021. Disponível em: https://doi.org/10.1145/3457607. Acesso em: 11 ago. 2025.

MOROZOV, Evgeny. To save everything, click here: The folly of technological solutionism. NY: ***PublicAffairs***, 2013.

MOSQUEIRA-REY, E. *et al*. Aprendizado de máquina human-in-the-loop: um estado da arte. ***Artif Intell Rev*** 56, p. 3005–3054, 2023. Disponível em: https://doi.org/10.1007/s10462-022-10246-w. Acesso em: 18 ago. 2025.

NAJIBI, Alex. Racial discrimination in face recognition technology. ***Science in the News***, v. 24, 2020. Disponível em: https://sites.harvard.edu/sitn/. Acesso em: 18 ago. 2025.

O'NEIL, Cathy. *Weapons of math destruction: How big data increases inequality and threatens democracy.* New York: Crown Publishing Group, 2016.

PASQUALE, Frank. *The black box society: The secret algorithms that control money and information*. Cambridge: Harvard University Press, 2015.

ROUVROY, Antoinette; BERNS, T. Governamentalidade algorítmica e perspectivas de emancipação. O díspar como condição de individuação pela relação? ***Redes***, 177(1), 163-196. 2013. Disponível em: https://doi.org/10.3917/res.177.0163. Acesso em 11 ago. 2025.

SHAH, Hetan. *Responsabilidade algorítmica. Philosophical Transactions of the Royal Society*, v. 376, n. 2128, p. 20170362, 2018. Disponível em https://doi.org/10.1098/rsta.2017.0362. Acesso em: 11 ago. 2025.

SCHMIDHUBER, Jürgen. Deep learning in neural networks: An overview. ***Neural networks***, v. 61, p. 85-117, 2015.

SOLOVE, Daniel J. A *Taxonomy of Privacy. University of Pennsylvania Law Review*, vol 154, n. 3, p. 477-564, 2006.

SORJ, Bernardo; GUEDES, Luís E. Exclusão digital: problemas conceituais, evidências empíricas e políticas públicas. ***Novos estudos*** CEBRAP, p. 101-117, 2005. DOI: 10.1590/S0101-33002005000200006. Acesso: 15 ago. 2025.

TONG, Jiarui Richard; LEE, Timothy Xueqian. Trustworthy AI that engages humans as partners in teaching and learning. ***Computer***, v. 56, n. 5, p. 62-73, 2023.

TOSTA, Pedro Lucas Maranini; DIAS, Jonatas Cerqueira. Detecção de fraudes em transações bancárias utilizando inteligência artificial. ***Revista Processando o Saber***, v. 17, p. 21-37, 2025. Disponível em: https://fatecpg.edu.br/revista/index.php/ps/article/view/341. Acesso em: 10 ago. 2025.

UNIÃO EUROPEIA. Parlamento Europeu; Conselho da União Europeia. *Regulamento (UE) 2016/679, de 27 de abril de 2016. Regulamento Geral sobre a Proteção de Dados – RGPD. Jornal Oficial da União Europeia*, L 119, p. 1-88, 4 maio 2016. Disponível em: https://gdpr-info.eu/. Acesso em: 10 ago. 2025

_____. Comissão Europeia. *Livro Branco sobre a inteligência artificial: uma abordagem europeia virada para a excelência e a confiança*. Bruxelas: Comissão Europeia, 2020. Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX%3A52020DC0065. Acesso em: 10 ago. 2025.

_____. Parlamento Europeu; Conselho da União Europeia. *Regulamento (UE) 2024/1689, de 13 de junho de 2024. Regulamento da Inteligência Artificial*. Jornal Oficial da União Europeia, L.1689, 12 jul. 2024. Disponível em: https://artificialintelligenceact.eu/the-act. Acesso em: 10 ago. 2025.

VINUESA, Roberta *et al*. The role of artificial intelligence in achieving the Sustainable Development Goals. ***Nature Communications***, v. 11, n. 233, p. 1–10, 2020. DOI: 10.1038/s41467-019-14108-y. Acesso em: 12 ago. 2025.

WACHTER, Sandra; MITTELSTADT, Brent; FLORIDI, Luciano. Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. ***International Data Privacy Law***, v. 7, n. 2, p. 76–99, 2017. DOI: 10.1093/idpl/ipx005. Acesso em: 11 ago. 2025.

XU, Wei *et al*. Transitioning to Human Interaction with AI Systems: New Challenges and Opportunities for HCI Professionals to Enable Human-Centered AI. ***International Journal of Human–Computer Interaction***, 39:3, 494-518. Disponível em: https://doi.org/10.1080/10447318.2022.2041900. Acesso em: 10 ago. 2025.

YOSHINAGA, Claudia; CASTRO, Fernando H.. Inteligência artificial: a vanguarda das finanças. ***GV EXECUTIVO***, v. 22, p. 31–35, 2023. Disponível em: https://doi.org/10.12660/gvexec.v22n3.2023.89911. Acesso em: 14 ago. 2025.

ZUBOFF, Shoshana. The age of surveillance capitalism. In: Social theory re-wired. ***Routledge***, 3 ed., p. 203-213, 2023.