v7.n12.2026

REVISTA CIENTÍFICA - RECIMA21 ISSN 2675-6218

A UTILIZAÇÃO DE IA NA TOMADA DE DECISÕES EMPRESARIAIS: IMPACTOS DA OPACIDADE ALGORÍTMICA E DA EXPERIÊNCIA DO GESTOR SOBRE A CONFIABILIDADE

THE USE OF AI IN BUSINESS DECISION-MAKING: IMPACTS OF ALGORITHMIC OPACITY AND THE MANAGER'S EXPERIENCE ON TRUSTWORTHINESS

EL USO DE LA IA EN LA TOMA DE DECISIONES EMPRESARIALES: IMPACTOS DE LA OPACIDAD ALGORÍTMICA Y DE LA EXPERIENCIA DEL GESTOR SOBRE LA CONFIABILIDAD

Maria Elisa Eduardo Alves[1], André Luiz da Silva[2], Felipe Diniz Dallilo[3]

https://doi.org/10.47820/recima21.v7i12.9706

PUBLICADO: 12/2026

RESUMO

O uso de inteligência artificial (IA) na tomada de decisão empresarial amplia a capacidade analítica das organizações, mas traz desafios ligados à opacidade algorítmica e à calibração da confiança humana nas recomendações. Este estudo analisou de que maneira a opacidade dos algoritmos e a experiência do tomador de decisão afetam a confiabilidade dos processos decisórios apoiados por IA. Realizou-se uma revisão sistemática da literatura, de abordagem qualitativa e descritiva, conforme Kitchenham e Charters (2007), com buscas em sete bases de dados, *snowballing* e uma busca complementar de atualização no site da editora MDPI. De 576 registros selecionados para triagem, 16 estudos compuseram o *corpus*. Os resultados indicam que a opacidade pode estar associada tanto ao excesso de confiança quanto à aversão às recomendações, que essa relação é heterogênea e que a experiência do decisor tende a favorecer a identificação de erros, sem eliminá-los. Como interpretação autoral, e não como achado dos estudos, propõe-se que experiência, explicabilidade e governança atenuam riscos distintos e não se substituem integralmente. A hipótese foi parcialmente sustentada. Conclui-se que a confiabilidade das decisões apoiadas por IA depende não só do desempenho técnico do algoritmo, mas também de como suas recomendações são interpretadas, verificadas e supervisionadas na organização.

PALAVRAS-CHAVE: Confiabilidade decisória. Explicabilidade. Governança de IA. Inteligência Artificial. Modelos híbridos. Opacidade algorítmica.

ABSTRACT

The use of artificial intelligence (AI) in business decision-making expands organizations' analytical capacity but raises challenges related to algorithmic opacity and the calibration of human trust in recommendations. This study analyzed how the opacity of algorithms and the decision-maker's experience affect the trustworthiness of AI-supported decision-making processes. A systematic literature review, with a qualitative and descriptive approach, was conducted following Kitchenham and Charters (2007), with searches in seven databases, snowballing and a complementary update search on the MDPI publisher's website. Of 576 records selected for screening, 16 studies made up the corpus. The results indicate that opacity may be associated with both overreliance and algorithm aversion, that this relationship is heterogeneous, and that decision-maker experience tends to help identify errors without eliminating them.

[1] Graduanda em Sistemas de Informação na Universidade de Araraquara.
[2] Doutor em Educação Escolar pela Universidade Estadual Paulista e professor da Universidade de Araraquara, na área de Inteligência Artificial.
[3] Doutor em Engenharia de Software pela Universidade de São Paulo e coordenador do curso de Inteligência Artificial da Universidade de Araraquara.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

As the authors' interpretation, rather than a finding of the studies, it is proposed that experience, explainability and governance mitigate distinct risks and do not fully substitute for one another. The hypothesis was partially supported. It is concluded that the trustworthiness of AI-supported decisions depends not only on the algorithm's technical performance but also on how its recommendations are interpreted, verified and supervised in the organization.

***KEYWORDS**: AI governance. Algorithmic opacity. Artificial Intelligence. Decisional trustworthiness. Explainability. Hybrid models.*

RESUMEN
El uso de la inteligencia artificial (IA) en la toma de decisiones empresariales amplía la capacidad analítica de las organizaciones, pero plantea desafíos relacionados con la opacidad algorítmica y la calibración de la confianza humana en las recomendaciones. Este estudio analizó de qué manera la opacidad de los algoritmos y la experiencia del tomador de decisiones afectan la confiabilidad de los procesos decisorios apoyados por IA. Se realizó una revisión sistemática de la literatura, de enfoque cualitativo y descriptivo, conforme a Kitchenham y Charters (2007), con búsquedas en siete bases de datos, snowballing y una búsqueda complementaria de actualización en el sitio de la editorial MDPI. De 576 registros seleccionados para el cribado, 16 estudios conformaron el corpus. Los resultados indican que la opacidad puede estar asociada tanto al exceso de confianza como a la aversión a las recomendaciones, que esta relación es heterogénea y que la experiencia del decisor tiende a favorecer la identificación de errores, sin eliminarlos. Como interpretación de los autores, y no como hallazgo de los estudios, se propone que la experiencia, la explicabilidad y la gobernanza atenúan riesgos distintos y no se sustituyen íntegramente. La hipótesis fue parcialmente sustentada. Se concluye que la confiabilidad de las decisiones apoyadas por IA depende no solo del desempeño técnico del algoritmo, sino también de cómo sus recomendaciones son interpretadas, verificadas y supervisadas en la organización.

***PALABRAS CLAVE**: Confiabilidad decisoria. Explicabilidad. Gobernanza de la IA. Inteligencia Artificial. Modelos híbridos. Opacidad algorítmica.*

1. INTRODUÇÃO

No atual cenário de negócios, marcado por transformações tecnológicas aceleradas e fluxos exponenciais de dados, a precisão na tomada de decisões tornou-se um diferencial competitivo crucial. Nesse contexto, a Inteligência Artificial (IA) assume um papel estratégico: Bandeira e Tortato (2024) destacam que sua integração à tomada de decisão amplia a capacidade organizacional de identificar padrões relevantes, e Agrawal, Gans e Goldfarb (2018) a tratam como apoio que amplia a capacidade de análise do decisor sem ocupar o lugar dele.

No entanto, um desafio permanece: a opacidade algorítmica, ou seja, a dificuldade de entender como o sistema chega às suas conclusões. A explicabilidade, definida como a habilidade de entender como o modelo opera, torna-se fundamental para a confiança na utilização da IA. Janssen *et al.* (2022) indicam que a explicabilidade, associada à experiência do decisor, favorece a identificação de sugestões incorretas, e Teffé e Medon (2020) ressaltam, no

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

campo jurídico, que a delegação de decisões a sistemas automatizados não afasta o dever de diligência dos agentes empresariais.

Apesar dos avanços na aplicação da IA em sistemas de apoio à decisão, ainda há lacunas na compreensão dos limites da atuação humana diante de sistemas algorítmicos opacos, que se acentuam em contextos de maior complexidade decisória. Diante disso, este estudo apresenta a seguinte questão de pesquisa: de que maneira a opacidade dos algoritmos de IA e o nível de experiência do tomador de decisão impactam a confiabilidade dos processos decisórios apoiados por IA nas organizações? A hipótese é que a opacidade algorítmica, na ausência de explicabilidade adequada, tende a induzir excesso de confiança (*overreliance*) nas recomendações de sistemas de IA e a tornar os decisores, inclusive os mais experientes, mais suscetíveis a vieses e erros, especialmente em contextos de alta complexidade.

Cientificamente, o estudo se justifica pela necessidade de compreender os limites do uso confiável da IA, especialmente quando a opacidade pode comprometer a capacidade crítica do decisor. No âmbito social, contribui para o debate sobre ética, responsabilidade e confiabilidade no uso de tecnologias que afetam organizações e seus *stakeholders*.

O objetivo geral é analisar, por meio de revisão sistemática da literatura, o uso da IA como suporte à tomada de decisão empresarial, com foco na integração entre tecnologia e julgamento humano. Especificamente, busca-se caracterizar o papel dos sistemas inteligentes nos processos decisórios organizacionais, analisar como a opacidade algorítmica e a experiência do gestor influenciam a confiabilidade das decisões apoiadas por IA, e discutir o papel da explicabilidade, da governança e da supervisão humana na construção de processos decisórios mais confiáveis.

A pesquisa seguiu as diretrizes de Kitchenham (2004) e de Kitchenham e Charters (2007) para revisões sistemáticas, com buscas por *strings* em sete bases de dados, em português e em inglês, *backward* e *forward snowballing* (Jalali; Wohlin, 2012; Wohlin, 2016) e uma busca complementar de atualização, conforme descrito na seção 3. Além de sintetizar os estudos, o trabalho propõe leituras próprias, construídas a partir do cruzamento entre estudos que não tratam desses pontos diretamente.

2. REFERENCIAL TEÓRICO

2.1. Sistemas de apoio à decisão

Os sistemas de apoio à decisão (SAD) são sistemas de informação que ajudam o decisor a lidar com grandes volumes de dados e a avaliar alternativas, sem tomar a decisão no lugar dele. Kovari (2024) descreve os SAD baseados em IA como sistemas que ajudam os decisores

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

a processar muitos dados para recomendar decisões bem informadas e precisas, e identifica como desafio central equilibrar acurácia, transparência e confiança.

A perspectiva sociotécnica, que Sarker *et al.* (2019) descrevem como eixo histórico de coesão da disciplina de Sistemas de Informação, orienta este trabalho: o sistema, as pessoas que o usam e a organização em que operam formam um arranjo conjunto.

2.2. Inteligência artificial no apoio à decisão

A integração da inteligência artificial aos processos decisórios amplia a capacidade analítica de gestores diante de grandes volumes de dados (Bandeira; Tortato, 2024). Agrawal, Gans e Goldfarb (2018) delimitam essa contribuição ao tratar a IA como tecnologia de predição, em que o julgamento sobre o que fazer com a previsão permanece com o decisor humano. A distinção entre ampliar e substituir o decisor importa porque estabelece um limite: o quanto a experiência do gestor consegue compensar as limitações impostas pela opacidade dos sistemas é uma pergunta em aberto.

2.3. Opacidade algorítmica e explicabilidade (XAI)

Modelos complexos, como os de aprendizado profundo, são frequentemente descritos como “caixas-pretas”, justamente por sua falta de transparência. Doshi-Velez e Kim (2017), tomados neste trabalho como referencial teórico, sistematizam a necessidade de interpretabilidade em três níveis: aplicações de alto risco, validação científica e detecção de vieses.

Técnicas de *Explainable AI* (XAI), como SHAP (Lundberg; Lee, 2017) e LIME (Ribeiro; Singh; Guestrin, 2016), foram desenvolvidas para tornar esses modelos mais interpretáveis, mostrando ao usuário quais fatores influenciaram determinada recomendação. Esses métodos permitem ao usuário compreender parte da lógica do modelo, mas a verificabilidade profunda continua sendo um problema em aberto, e Kovari (2024) inclui a opacidade entre os fatores que prejudicam a confiança dos usuários na IA.

2.4. Excesso de confiança (Overreliance) e experiência do decisor

A relação entre transparência e confiança não é linear. A natureza de “caixa-preta” aparece entre os fatores de desenho associados à aversão a algoritmos (*algorithm aversion*), mas a literatura sobre o tema é heterogênea e depende de desenhos e tarefas específicos (Mahmud *et al.*, 2022).

Jabbari *et al.* (2026), em revisão narrativa, organizam essa discussão ao separar confiança, que é uma atitude diante da IA, de dependência (*reliance*), que é o peso efetivamente

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

dado à recomendação. Para os autores, a dependência é apropriada quando esse peso se justifica pelas evidências e pelo contexto da decisão, e pode falhar em duas direções: dar peso demais à recomendação (*overreliance*) ou peso de menos (*underreliance*), como ocorre com decisores que superestimam a própria competência e descartam conselhos úteis.

A experiência do decisor aparece associada a melhores resultados, mas não de forma uniforme: ela pode ajudar na verificação das recomendações sem garantir a calibração da confiança, que também depende de como o sistema e a interação são desenhados. A forma de manter essa calibração no dia a dia depende da supervisão humana, tratada na seção 2.7.

2.5. Responsabilidade e dever de diligência

Quando a IA erra, a atribuição da responsabilidade não é óbvia. Sob o ângulo jurídico, Teffé e Medon (2020) ressaltam que a delegação de decisões a sistemas automatizados não retira essa responsabilidade: os administradores permanecem vinculados ao dever de diligência, a obrigação de atuar com cuidado, atenção e zelo na condução dos negócios. Para os autores, a opacidade algorítmica não constitui excludente de responsabilidade, e adotar um sistema sem compreender seu funcionamento pode, por si só, configurar violação desse dever, o que justifica exigir mais transparência dos sistemas utilizados, sob pena de o administrador ficar impossibilitado de exercer seu juízo crítico e, no limite, responder pessoalmente por decisões algorítmicas.

2.6. Governança em sistemas de IA

Empresas que adotam IA em processos decisórios precisam de estruturas de controle, auditoria e alinhamento com seus objetivos. A simples disponibilidade de modelos preditivos não basta, segundo Pisoni e Moloney (2024): é preciso estabelecer pontos de controle e de decisão explícitos nos sistemas de informação, de modo que decisões automatizadas possam ser rastreadas e revertidas, e a governança de IA depende do encaixe entre a arquitetura de TI da organização e as exigências regulatórias, de mercado e de bem-estar social às quais ela responde. A governança não elimina as limitações técnicas dos modelos, e seu efeito depende também da experiência do gestor que a aplica.

2.7. Supervisão humana em sistemas de IA

A supervisão humana é o mecanismo pelo qual a governança se aplica à decisão concreta. Kovari (2024) propõe manter o ser humano no comando por meio do *human-in-the-loop*, com a IA na posição de assistente, subordinada ao julgamento humano. Mecanismos específicos de supervisão, como a validação manual por um gestor em decisões de maior risco,

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

as trilhas de auditoria sobre a interação entre humano e sistema e os papéis formais que definem quem pode sobrescrever uma recomendação, são aqui tratados como desdobramento dos pontos de controle de Pisoni e Moloney (2024). A presença de uma pessoa no processo, porém, não garante supervisão efetiva, como mostra o experimento de Green e Chen (2019), analisado na seção 4.3.

2.8. O trade-off entre acurácia e explicabilidade

É um equívoco comum supor que modelos mais transparentes são sempre preferíveis. Rudin (2019) argumenta que, em muitas decisões de alto risco, é possível usar modelos intrinsecamente interpretáveis, diminuindo a dependência de explicações *post hoc*. Cordasco e Véliz (2026) complicam esse conselho: mesmo quando modelos interpretáveis não são viáveis, explicabilidade e acurácia podem entrar em conflito, especialmente em ambientes decisórios abertos e dinâmicos. Os autores distinguem explicações espessas (*thick*), que são confiáveis e verificáveis e permitem contestar os critérios da decisão, de explicações finas (*thin*), úteis para a compreensão geral, mas insuficientes para essa verificação, e exigir explicações espessas nesses ambientes tende a reduzir a acurácia. Diante desse *trade-off*, uma alternativa tem sido explorada: os modelos híbridos.

2.9. Modelos híbridos como estratégia de mitigação

Modelos totalmente opacos dificultam a auditoria. Modelos rigidamente baseados em regras reduzem a flexibilidade decisória. Entre esses dois extremos, as organizações enfrentam um problema de equilíbrio entre adaptabilidade e verificabilidade, e é nesse espaço que os modelos híbridos surgem como alternativa intermediária, combinando regras de negócio explicitamente formuladas com componentes de aprendizado de máquina. Mais do que um desafio técnico, trata-se de um problema organizacional: como estruturar a decisão para que parte dela permaneça auditável sem comprometer a resposta a novos padrões.

Bork, Ali e Dinev (2023) propõem integrar as duas lógicas por meio da notação DMN (*Decision Model and Notation*), que permite aos especialistas de negócio descrever regras de forma visual, declarativa e auditável. O aprendizado de máquina (ML) é treinado sobre dados gerados pela execução dessas regras, e a importância relativa de cada variável para o resultado da decisão é sobreposta ao modelo DMN original, em um processo chamado de superimposição. Preserva-se assim um núcleo de regras compreensíveis, que sustenta a explicabilidade, enquanto o ML amplia a flexibilidade; os modelos híbridos tornam, dessa forma, mais administrável a tensão entre acurácia e explicabilidade, mesmo sem eliminá-la.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

2.10. Síntese da revisão bibliográfica

A literatura analisada mostra que o uso de IA em processos decisórios empresariais amplia a capacidade analítica das organizações, mas introduz desafios ligados à opacidade algorítmica, à calibração da confiança humana e à supervisão organizacional. Explicabilidade, governança e modelos híbridos reduzem parte desses problemas, mas não os eliminam: a verificabilidade de decisões apoiadas por IA permanece uma questão em aberto. Como os estudos enfatizam dimensões distintas do problema, este trabalho examina como essas dimensões aparecem no conjunto de estudos selecionados e quais relações podem ser estabelecidas entre elas.

3. METODOLOGIA

Esta seção descreve o protocolo da revisão sistemática, do planejamento ao término da execução, e a síntese dos dados. O protocolo segue as etapas de Kitchenham e Charters (2007), a partir do procedimento original de Kitchenham (2004): planejamento, execução e relato. O relato dos resultados está na seção 4.

3.1. Planejamento da revisão

O objetivo da revisão é identificar, nos estudos disponíveis, evidências sobre como a opacidade algorítmica, a explicabilidade, a experiência do gestor, a governança e os modelos híbridos se relacionam com a confiabilidade da decisão empresarial apoiada por IA. A questão de pesquisa apresentada na Introdução foi desdobrada em cinco perguntas (Quadro 1). Cada *string* de busca derivou de uma delas: as palavras-chave da pergunta, seus sinônimos e os filtros formaram a *string* (Quadro 2).

Quadro 1. Perguntas da revisão e strings que as respondem

Pergunta	**Formulação**	***Strings***
P0	Qual o papel dos sistemas de IA nos processos decisórios organizacionais?	S0
P1	Como a opacidade e a explicabilidade afetam a confiança do decisor (excesso de confiança e aversão)?	S1, S1b
P2	Como a experiência do gestor modula esse efeito?	S1, S1b
P3	Qual o papel da governança, da supervisão humana e da responsabilização?	S2a, S2b, S4

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Pergunta	Formulação	*Strings*
P4	Como os modelos híbridos lidam com o *trade-off* entre acurácia e explicabilidade?	S3

Fonte: Elaborado pelos autores (2026).

As buscas foram feitas entre 21 de março e 15 de abril de 2026 em sete bases: SpringerLink, ScienceDirect, Semantic Scholar, MDPI, SAGE Journals, Taylor & Francis e SciELO. A escolha das bases considerou a adequação ao eixo de cada *string* e o volume de resultados, e por isso nem toda *string* foi aplicada em todas as bases. Os filtros variaram conforme a plataforma (período de 2019 a 2026, tipo de documento e área de Ciência da Computação). Para permitir a leitura crítica individual dos estudos potencialmente pertinentes, adotou-se um limite operacional de até 30 resultados por combinação de base de dados e *string* de busca, na ordem de relevância apresentada pela própria plataforma; nas buscas com menos de 30 resultados, todos os registros foram examinados. Esse limite restringiu a abrangência da busca, o que constitui uma limitação, e não indica que os 30 primeiros resultados sejam os estudos de maior qualidade científica. O estudo de Green e Chen (2019) apareceu na posição 38 do Semantic Scholar (S2b), fora desse limite, e foi mantido por atender aos critérios.

Quadro 2. Strings de busca

Código	*String*	Pergunta
S0	(*"artificial intelligence"* OR AI) AND (*"decision support"* OR *"decision-making"*) AND (*business* OR *industry* OR *management* OR *organizational* OR *government*)	P0
S1	(*"algorithmic opacity"* OR *"black box"* OR *"explainable AI"* OR *explainability*) AND (*expertise* OR *experience* OR *"decision-makers"*) AND (*"decision-making"* OR *"decision support"*)	P1, P2
S1b	(*"algorithm aversion"* OR *"algorithm appreciation"* OR *overreliance* OR *"appropriate reliance"*) AND (*"decision-making"* OR *"decision support"*)	P1
S2a	(*"AI governance"* OR *"responsible AI"*) AND (*business* OR *organization* OR *organisation* OR *"decision-making"*)	P3
S2b	(*"algorithm-in-the-loop"* OR *"human-in-the-loop"* OR *"human oversight"*) AND (*"decision-making"* OR *"decision making"*)	P3
S3	(*"accuracy-explainability trade-off"* OR *"accuracy explainability trade-off"* OR *"hybrid decision"*) AND (*explainability* OR *interpretability* OR *explainable*)	P4
S4	"inteligência artificial" "responsabilidade civil" "decisões empresariais"	P3 (jurídico)

Fonte: Elaborado pelos autores (2026).

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Os critérios de inclusão e de exclusão estão no Quadro 3. No critério I3, “artigo de revisão” designa um tipo de publicação científica, como a revisão narrativa ou sistemática da literatura. Não houve avaliação formal da qualidade metodológica com instrumento padronizado: a seleção baseou-se na publicação revisada por pares e na leitura crítica do texto completo de cada estudo incluído, segundo esses critérios. O critério E6 separa os estudos que apenas mencionam a inteligência artificial, a tomada de decisão ou algum dos eixos de interesse daqueles que são fontes substantivas sobre a relação entre esses fatores e a confiabilidade das decisões apoiadas por IA. Por isso, não se exigiu que os estudos mensurassem diretamente a confiabilidade: uma proposta técnica, uma discussão conceitual, uma arquitetura de decisão ou uma análise das condições organizacionais pode contribuir de forma relevante.

Além das buscas por *strings*, a revisão usou três procedimentos complementares. O primeiro foi o *backward snowballing* (referências dos estudos incluídos) e o *forward snowballing* (estudos que citam os incluídos), conforme Jalali e Wohlin (2012) e Wohlin (2016), com os mesmos critérios de inclusão e exclusão. O segundo foi uma busca complementar de atualização no site da editora MDPI, realizada em 15 de outubro de 2026 devido ao dinamismo da área e ao intervalo entre a coleta e a conclusão do trabalho. Buscou-se, sem *string*, por termos livres (“artificial intelligence”, “decision support systems” e “*explainable AI*”), com o filtro de área *Computer Science* e o filtro temporal de publicações dos últimos três meses. Os resultados foram avaliados quanto à pertinência temática e à adequação aos critérios de elegibilidade. O terceiro foi a busca exploratória por obras fundamentais.

Essas obras fundamentais formam o referencial teórico: Agrawal, Gans e Goldfarb (2018) foram encontrados no Google Acadêmico, com os termos *artificial intelligence* e *decision-making*, e Sarker *et al.* (2019), também no Google Acadêmico, pela relevância para o curso. Doshi-Velez e Kim (2017) foram identificados por *backward snowballing* a partir de Cordasco e Véliz (2026). As obras de Kitchenham e Charters (2007), Kitchenham (2004), Jalali e Wohlin (2012), Wohlin (2016) e Page *et al.* (2021) são referências do método e, assim como as anteriores, não constam entre os estudos analisados.

Quadro 3. Critérios de inclusão (I) e de exclusão (E)

Código	Critério
I1	Publicado entre 2019 e 2026 (exceção justificada para obras seminais identificadas por *snowballing* ou indicação)
I2	Em português ou inglês
I3	Artigos científicos de pesquisa original ou de revisão publicados em periódicos revisados por pares, trabalhos completos publicados em anais de eventos com revisão por pares e capítulos de livros acadêmicos, com texto completo disponível

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Código	Critério
I4	Trata de decisão apoiada por IA, de decisão algorítmica ou de conceitos e métodos de explicabilidade que sustentam essas decisões
I5	Aborda ao menos um eixo: opacidade e explicabilidade; confiança, excesso de confiança ou aversão; experiência ou expertise do decisor; governança, supervisão humana ou responsabilização; modelos híbridos ou *trade-off* entre acurácia e explicabilidade
E1	Não trata de apoio à decisão ou de tomada de decisão
E2	Método técnico de IA ou aplicação setorial (clínica, energia, agricultura, manufatura etc.) sem relação com decisão organizacional
E3	Sem texto completo disponível
E4	Editorial, resenha, resumo de evento ou duplicado
E5	Fora dos idiomas definidos
E6	Menciona a inteligência artificial, a tomada de decisão ou algum dos eixos de interesse, mas não é fonte substantiva sobre a relação entre esses fatores e a confiabilidade das decisões apoiadas por IA

Fonte: Elaborado pelos autores (2026).

3.2. Execução

A execução começou pela aplicação das *strings*. O Quadro 4 mostra, por base e *string*, quantos resultados cada busca retornou e quantos foram selecionados para triagem. Os resultados retornados dependem do mecanismo e dos filtros de cada plataforma, e os selecionados correspondem ao limite operacional descrito na seção 3.1, e não a uma avaliação automática da qualidade. Ao todo, 576 registros foram selecionados para triagem. Desses, 32 eram duplicados (títulos idênticos entre buscas e bases), o que deixou 544 registros únicos. Os duplicados vieram da sobreposição de *strings* (S0 e S1 no SpringerLink; S2a e S2b no ScienceDirect) e do mesmo artigo em duas bases.

Quadro 4. Resultados retornados e selecionados para triagem por base e string (retornados/selecionados)

Base	S0	S1	S1b	S2a	S2b	S3	S4	Total
SpringerLink	639/30	10/10	315/30	738/30	–	201/30	–	130
ScienceDirect	–	–	419/30	179/30	342/30	78/30	–	120
Semantic Scholar	610/30	252/30	78/30	396/30	324/30	50/30	106/30	210
MDPI	295/30	–	–	–	–	–	–	30
SAGE Journals	–	119/30	–	–	–	–	–	30
Taylor & Francis	–	–	112/30	–	–	–	–	30
SciELO	26/26	–	–	–	–	–	–	26
Total selecionados	116	70	120	90	60	90	30	576

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

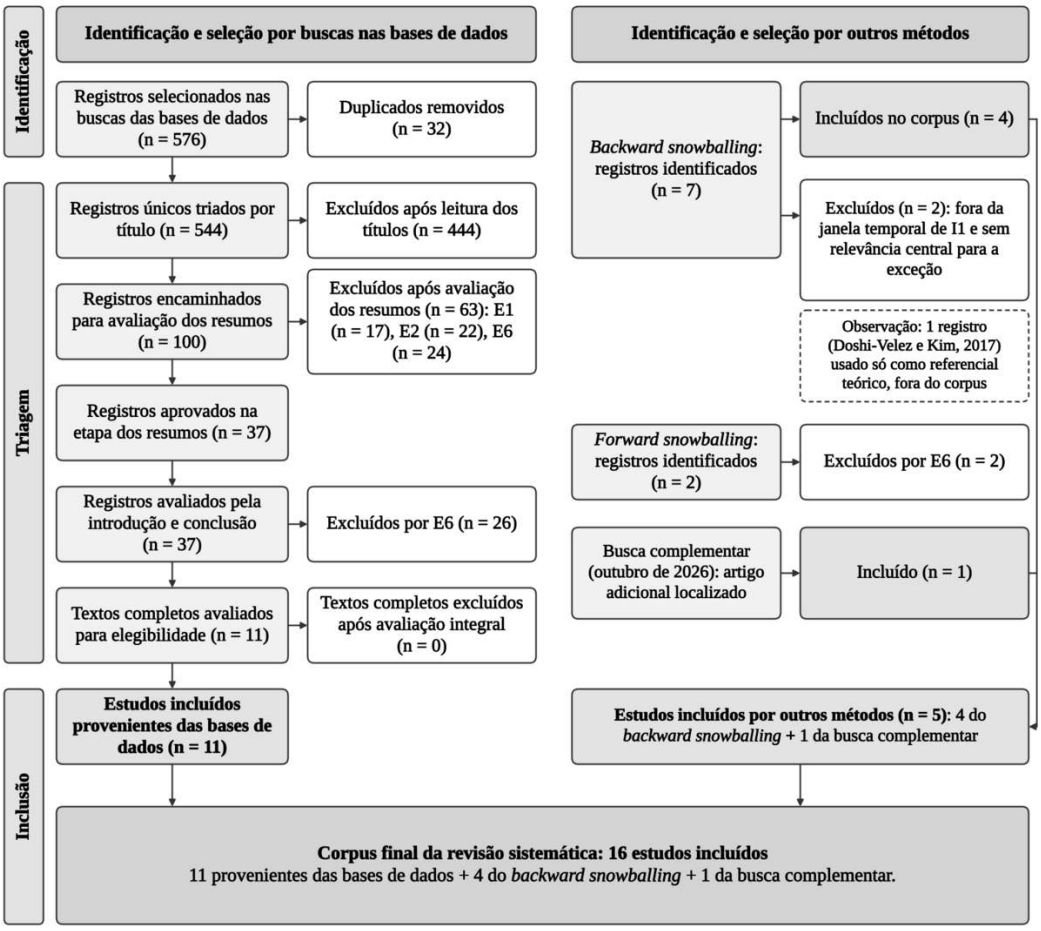
Fonte: Elaborado pelos autores (2026). Traço indica string não aplicada na base.

A Figura 1 resume o processo de identificação, triagem e inclusão, no formato do PRISMA 2020 (Page *et al.*, 2021) adaptado a esta revisão. Nas bases de dados, a leitura dos títulos dos 544 registros únicos excluiu 444 e encaminhou 100 para a avaliação dos resumos. Nessa etapa, 63 foram excluídos (17 por E1, 22 por E2 e 24 por E6) e 37 seguiram para a leitura da introdução e da conclusão. Dos 37, 26 foram excluídos por E6 e 11 foram encaminhados para a leitura integral, na qual os 11 foram incluídos, sem exclusões. Os registros de outros métodos seguem percurso próprio, no lado direito da figura.

O *backward snowballing* identificou 7 registros: 4 foram incluídos no *corpus* (Lundberg e Lee, 2017; Ribeiro, Singh e Guestrin, 2016; Rudin, 2019; Raisch e Krakowski, 2021), 2 foram excluídos por estarem fora da janela temporal de I1 e não apresentarem relevância suficientemente central para justificar a exceção, e 1 (Doshi-Velez e Kim, 2017) foi aproveitado exclusivamente como referencial teórico, sem integrar o *corpus*. Lundberg e Lee (2017) e Ribeiro, Singh e Guestrin (2016) também estão fora do período de 2019 a 2026 e foram mantidos pela exceção de obras seminais de I1. O *forward snowballing* identificou 2 registros, não incluídos por ausência de abordagem substantiva, conforme o critério E6. A busca complementar, de 15 de outubro de 2026, localizou o artigo de Jabbari *et al.* (2026), publicado em 10 de outubro, depois das buscas iniciais, que foi incluído porque trata da dependência em relação à IA e do papel das explicações e atende aos critérios de elegibilidade. O *corpus* final somou 16 estudos: 11 das bases de dados, 4 do *backward snowballing* e 1 da busca complementar.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Figura 1. Fluxograma do processo de identificação, triagem e inclusão dos estudos



Fonte: Elaborado pelos autores (2026), adaptado de Page *et al.* (2021), PRISMA 2020.

3.3. Síntese e análise dos dados

Dos 11 estudos vindos das bases de dados, o SpringerLink contribuiu com 3, o Taylor & Francis e o Semantic Scholar com 2 cada, e o ScienceDirect, o MDPI, o SAGE Journals e o SciELO com 1 cada. O Quadro 5 reúne os estudos com tipo, pergunta de revisão, grupo, origem e país do primeiro autor.

No Quadro 5, os códigos P0 a P4 correspondem às perguntas de revisão do Quadro 1, e os critérios de inclusão (I1 a I5) e de exclusão (E1 a E6) seguem os códigos do Quadro 3. Na síntese, os achados relatados pelos estudos são distinguidos das leituras próprias dos autores desta revisão, identificadas como tais no texto e não testadas empiricamente.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Quadro 5. Estudos incluídos na revisão

Estudo	Tipo	Pergunta	Grupo	Origem (base / *string*)	País
Janssen *et al.* (2022)	Experimento (governo)	P1, P2	C	SAGE / S1	Holanda
Dogru e Krämer (2025)	Estudo empírico	P1, P2	C	Taylor & Francis / S1b	Alemanha
Mahmud *et al.* (2022)	Revisão sistemática	P1	C	ScienceDirect / S1b	Finlândia
Kotzian, Weissenberger e Prinz (2026)	Experimentos fatoriais	P1	C	Taylor & Francis / S1b	Alemanha
Green e Chen (2019)	Experimento	P3	C	Semantic Scholar / S2b	EUA
Jabbari *et al.* (2026)	Revisão narrativa	P1, P2	C	Busca complementar	EUA
Pisoni e Moloney (2024)	Conceitual (BPM)	P3	D	SpringerLink / S2a	Reino Unido
Cordasco e Véliz (2026)	Conceitual (ética)	P4	D	SpringerLink / S3	Reino Unido
Teffé e Medon (2020)	Jurídico	P3	D	Semantic Scholar / S4	Brasil
Bork, Ali e Dinev (2023)	Proposta técnica e caso	P4	A	SpringerLink / S3	Áustria
Lundberg e Lee (2017)	Técnico (SHAP)	P1	A	*Snowballing* (Cordasco)	EUA
Ribeiro, Singh e Guestrin (2016)	Técnico (LIME)	P1	A	*Snowballing* (Cordasco)	EUA
Rudin (2019)	Argumentativo	P4	A	*Snowballing* (Cordasco)	EUA
Kovari (2024)	Revisão	P0	B	MDPI / S0	Hungria
Bandeira e Tortato (2024)	Revisão	P0	B	SciELO / S0	Brasil
Raisch e Krakowski (2021)	Conceitual	P0, P4	B	*Snowballing* (Kotzian)	Suíça

Fonte: Elaborado pelos autores (2026). Grupos: A, tecnologia, IA e explicabilidade; B, administração e sistemas de decisão; C, fatores humanos, confiança e supervisão; D, governança, ética e direito.

Os grupos reúnem 4 estudos em A, 3 em B, 6 em C e 3 em D. A atribuição de cada estudo a uma pergunta e a um grupo reflete uma leitura autoral e considera o foco principal do estudo, não tudo o que ele menciona. Quatorze dos 16 estudos foram publicados a partir de 2019, e sete em 2024 ou depois. Os dois mais antigos são os métodos LIME (2016) e SHAP (2017). Quanto à origem, nove estudos têm primeiro autor em instituição europeia (Alemanha e

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Reino Unido, 2 cada; Áustria, Finlândia, Holanda, Hungria e Suíça, 1 cada), cinco nos Estados Unidos e dois no Brasil.

4. RESULTADOS E DISCUSSÃO

Os resultados estão organizados segundo as perguntas de revisão identificadas no Quadro 1, indicadas pelos códigos P0 a P4. As subseções discutem os achados relacionados a cada pergunta e, quando pertinente, apresentam interpretações construídas a partir do cruzamento entre os estudos. Essas interpretações são identificadas no texto como leituras próprias: não foram demonstradas diretamente por um único estudo nem testadas empiricamente neste trabalho.

4.1. O papel da IA nos processos decisórios (P0)

Kovari (2024) e Bandeira e Tortato (2024) descrevem a IA como ampliação da capacidade analítica do decisor, e Raisch e Krakowski (2021) mostram que a relação entre automação e aumentação é mais complexa.

Jabbari *et al.* (2026), em revisão narrativa de 29 estudos centrais, identificam cinco condições sociotécnicas que moldam a dependência em relação à IA (sistema, usuário, tarefa, interação e organização) e mostram que um mesmo recurso, como uma explicação, pode produzir efeitos distintos conforme essa configuração. Esse resultado sustenta que o desempenho estatístico do modelo é importante, mas não suficiente isoladamente, pois a confiabilidade depende também de como a recomendação é interpretada pelos responsáveis e amparada por mecanismos organizacionais.

Por complexidade decisória entende-se a presença de muitas variáveis interdependentes e de um ambiente pouco previsível, em que a checagem independente da recomendação é mais difícil e cara. A hipótese central, de que a opacidade tende a ampliar a confiança excessiva mesmo entre experientes, é retomada adiante. Quanto a P0, os estudos tratam a IA como apoio que amplia a capacidade analítica sem substituir o julgamento humano, e indicam que o valor desse apoio depende do contexto.

4.2. Opacidade, explicabilidade e experiência do decisor (P1 e P2)

A opacidade (P1) pode produzir reações opostas, conforme o desenho. No experimento de Janssen *et al.* (2022), com decisões simuladas sobre pedidos de residência no setor público e amostra composta sobretudo por estudantes de mestrado, os participantes apoiados por aprendizado de máquina (sistema opaco) acertaram menos e detectaram menos erros do que os apoiados por regras de negócio (sistema transparente), o que sugere aceitação acrítica.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Mahmud *et al.* (2022) apontam a caixa-preta entre os fatores associados à aversão. Em sentido contrário, Kotzian, Weissenberger e Prinz (2026), em experimentos com profissionais e previsões de vendas fornecidas por um algoritmo ou por um analista humano, encontram aversão muito limitada, sem que a explicação aumentasse o uso das previsões, e Dogru e Krämer (2025), com 529 enxadristas em uma tarefa de baixo risco e um sistema de 67% de acurácia, não encontram excesso de confiança. Assim, a evidência não sustenta uma relação única: a opacidade aparece associada à aceitação acrítica em alguns desenhos e à aversão em outros, e em dois estudos o efeito foi pequeno ou ausente.

Quanto à experiência do decisor (P2), Janssen *et al.* (2022) observaram que participantes com experiência na decisão detectaram mais erros, embora nenhum grupo tenha detectado todos. Dogru e Krämer (2025) associam a expertise a maior autoconfiança e menor confiança na IA, em resultados correlacionais, e Jabbari *et al.* (2026) tratam experiência e autoconfiança como condições cujo efeito varia com o contexto. As evidências vêm de contextos muito diferentes e de amostras nem sempre de gestores, o que limita a generalização. Quanto a P1 e P2, a opacidade não tem efeito único sobre a confiança, e a experiência favorece a verificação das recomendações sem garantir a calibração da confiança.

Cordasco e Véliz (2026) mostram que explicações finas informam, mas nem sempre permitem verificar ou contestar o critério da recomendação. A partir desse resultado e do de Janssen *et al.* (2022), é proposto aqui, como interpretação ainda não testada, que um decisor experiente talvez compense parte das limitações de uma explicação fina com seu repertório, e que a forma de apresentar as explicações poderia variar conforme a expertise. Isso diz respeito à apresentação, e não aos requisitos de verificação, rastreabilidade e responsabilização, que não devem ser reduzidos porque o decisor é experiente: nenhum estudo do *corpus* mostra que a experiência os dispense, e mesmo experientes deixam de detectar erros (Janssen *et al.*, 2022).

4.3. Governança, supervisão humana e responsabilização (P3)

Os estudos reúnem contribuições de natureza distinta. Pisoni e Moloney (2024), em análise conceitual, defendem pontos de controle e de decisão explícitos nos sistemas de informação, de modo que decisões automatizadas possam ser rastreadas e revertidas; Teffé e Medon (2020), em discussão jurídica, sustentam que a delegação a sistemas automatizados não afasta o dever de diligência dos administradores, mesmo quando o sistema é opaco; e Kovari (2024), em revisão, propõe manter o ser humano no comando. São propostas organizacionais e argumentos jurídicos sobre o que as organizações deveriam fazer, e não resultados empíricos sobre a eficácia dessas medidas.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

A evidência experimental vem de Green e Chen (2019), com avaliações de risco em duas situações (liberdade pré-julgamento e concessão de crédito). Na condição *update*, o participante registra sua previsão, vê a recomendação do modelo e então a revisa. Essa condição teve o melhor desempenho médio nos dois cenários, com diferenças pequenas, e as previsões iniciais nela também foram melhores do que as da condição sem apoio, o que os autores atribuem a um efeito de treinamento. Ainda assim, em nenhuma condição os participantes avaliaram com precisão o desempenho do modelo nem calibraram a dependência a ele, e todos mostraram viés na interação com o modelo. O estudo não comparou iniciantes e experientes, não examinou a intenção ou a boa-fé de quem aplica o procedimento e não envolveu ambientes organizacionais.

Propõe-se, aqui, que formar um julgamento antes de ver a recomendação seria mais difícil para o iniciante e funcionaria, para o experiente, mais como verificação, e que o procedimento poderia perder o efeito se o registro for superficial. Essas hipóteses não foram demonstradas por Green e Chen (2019) nem por outro estudo do *corpus*, e não devem ser lidas como efeitos da experiência ou da boa-fé dos gestores. Jabbari *et al.* (2026) acrescentam que reduzir a dependência excessiva pode aumentar a insuficiente, e por isso a supervisão deve ser avaliada pela seletividade do uso da IA. Quanto a P3, governança e supervisão aparecem como condições para tornar a decisão rastreável e responsabilizável, mas a presença de uma pessoa no processo não garante supervisão efetiva.

4.4. Acurácia, explicabilidade e modelos híbridos (P4)

O *corpus* reúne um argumento em favor de modelos interpretáveis (Rudin, 2019), uma análise conceitual de ética empresarial (Cordasco; Véliz, 2026), uma proposta técnica com estudo de caso (Bork; Ali; Dinev, 2023) e um estudo conceitual de gestão (Raisch; Krakowski, 2021). Esses estudos contribuem de maneiras distintas para a discussão sobre acurácia, explicabilidade e participação humana e, em conjunto, permitem discutir os modelos híbridos como estratégia para administrar a tensão entre adaptabilidade e rastreabilidade, e não como solução capaz de eliminá-la.

É preciso distinguir duas discussões. A organizacional: para Raisch e Krakowski (2021), automação e aumentação são contraditórias numa dada tarefa e momento, mas interdependentes ao longo do tempo e entre tarefas, e priorizar só uma delas traz riscos (perda de habilidades e rigidez, ou custos elevados e ciclos de aprendizado que não se fecham). A técnica: Bork, Ali e Dinev (2023) treinam um modelo de aprendizado de máquina sobre dados gerados por regras DMN auditáveis e sobrepõem ao DMN a importância das variáveis. No caso relatado, os dados foram simulados a partir das próprias regras: é uma proposta técnica com caso de aplicação, e não evidência sobre o efeito dessa arquitetura nas decisões.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

Rudin (2019) é aqui um contraponto, e não um estudo de modelos híbridos: defende modelos intrinsecamente interpretáveis em decisões de alto risco quando viáveis e critica explicações *post hoc* de caixas-pretas. Os modelos híbridos não contradizem essa posição, mas também não eliminam a opacidade, pois o componente de aprendizado de máquina segue sendo uma caixa-preta parcial. A abordagem híbrida parece, nesse sentido, permitir que essa opacidade se restrinja a uma parcela menor da decisão, em que auditoria, trilhas e políticas de *override* têm mais chance de funcionar; nenhum estudo do *corpus* testou se isso aumenta a confiabilidade. Também como interpretação, o gestor passaria a precisar compreender a estrutura do sistema (o que é regra, o que é aprendizado de máquina, onde a supervisão direta é indispensável), e seu preparo para isso poderia influenciar o sucesso do arranjo, hipótese não testada.

Dessa forma, quanto a P4, os modelos híbridos aparecem como estratégia para equilibrar rastreabilidade e adaptação, sem eliminar a opacidade nem garantir a confiabilidade.

4.5. Contribuição das estratégias de busca para o *corpus*

Jalali e Wohlin (2012) comparam as duas estratégias de busca, e aqui elas trouxeram contribuições diferentes. As buscas nas bases responderam por 11 dos 16 estudos (69%) e forneceram o eixo comportamental e organizacional. O *backward snowballing*, com 4 estudos (25%), trouxe sobretudo a base técnica (SHAP, LIME e Rudin) e Raisch e Krakowski (2021). O *forward snowballing* não incluiu estudos, e a busca complementar contribuiu com um estudo recente (6%), publicado depois da busca inicial (Jabbari *et al.*, 2026).

4.6. Foco dos estudos selecionados

Pelo foco principal de cada estudo (Quadro 5), os 16 estudos concentram-se em uma ou duas das perguntas da revisão: sete tratam de opacidade, confiança e experiência (P1 e P2), três de governança, supervisão e responsabilização (P3), quatro de acurácia, explicabilidade e modelos híbridos (P4) e três do papel da IA nas decisões (P0), com Raisch e Krakowski (2021) contando em P0 e em P4. Essa classificação reflete uma leitura própria e se refere ao foco declarado dos estudos selecionados nesta revisão, sem permitir afirmar a ausência de estudos integradores na literatura.

4.7. A hipótese diante dos estudos

A hipótese foi parcialmente sustentada. Há apoio para dois pontos: em Janssen *et al.* (2022), a opacidade esteve associada a menor detecção de erros e mesmo os experientes deixaram passar falhas; em Green e Chen (2019), os participantes não calibraram a dependência e mostraram viés na interação com o modelo, embora esse estudo não tenha manipulado a

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

opacidade nem a experiência. São cenários específicos (decisão administrativa simulada, com amostra sobretudo de estudantes, e avaliações de risco).

Em sentido contrário, Kotzian, Weissenberger e Prinz (2026) encontram aversão muito limitada e Dogru e Krämer (2025) não encontram excesso de confiança, ambos em condições específicas, e Mahmud *et al.* (2022) e Jabbari *et al.* (2026) mostram que o efeito da opacidade e das explicações varia com o contexto. Essa heterogeneidade impede tratar o excesso de confiança como consequência geral da opacidade, e a parte da hipótese sobre alta complexidade não foi testada diretamente. A avaliação é, portanto, parcial e condicional: a opacidade pode favorecer a aceitação acrítica e a experiência não elimina erros, mas permanecem incertas a generalidade do excesso de confiança e o papel da complexidade.

5. CONSIDERAÇÕES FINAIS

Este trabalho investigou, por meio de revisão sistemática da literatura, de que maneira a opacidade dos algoritmos de IA e a experiência de quem decide afetam a confiabilidade das decisões empresariais apoiadas por IA. A resposta que os estudos do *corpus* sustentam é condicional: a opacidade dificulta a verificação das recomendações e pode levar tanto à aceitação acrítica quanto à rejeição, e a experiência do decisor tende a melhorar a identificação de erros, mas não os elimina nem garante, por si só, a calibração da confiança. A confiabilidade depende, portanto, não apenas do desempenho técnico do algoritmo, mas de como suas recomendações são interpretadas, verificadas e supervisionadas no contexto organizacional.

Sobre a opacidade, os estudos não indicam relação uniforme: houve menor detecção de erros com um sistema opaco (Janssen *et al.*, 2022), aversão associada à caixa-preta (Mahmud *et al.*, 2022) e, em outros desenhos, aversão muito limitada (Kotzian; Weissenberger; Prinz, 2026) ou ausência de excesso de confiança (Dogru; Krämer, 2025), e o efeito depende do cenário, da tarefa, do desenho do sistema e do perfil do usuário (Jabbari *et al.*, 2026). A governança e a supervisão tornam o uso da IA rastreável e responsabilizável (Pisoni; Moloney, 2024; Teffé; Medon, 2020), mas a presença de uma pessoa no processo não garante supervisão efetiva (Green; Chen, 2019), e os modelos híbridos tornam a tensão entre acurácia e explicabilidade mais administrável, sem eliminá-la (Bork; Ali; Dinev, 2023).

A hipótese do trabalho, de que a opacidade sem explicabilidade adequada induz excesso de confiança e torna mesmo os decisores experientes mais suscetíveis a vieses e erros, foi parcialmente sustentada: há apoio para a possibilidade de aceitação acrítica de recomendações incorretas e para o fato de que a experiência não elimina erros, mas permanecem incertos a generalidade do excesso de confiança e o papel da complexidade. A contribuição própria do

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

trabalho é uma interpretação sociotécnica construída pela síntese da literatura, e não um achado de um estudo isolado: experiência, explicabilidade e governança atenuam riscos distintos (a experiência, a verificação individual das recomendações; a explicabilidade, a compreensão do critério; a governança, a rastreabilidade e a responsabilização) e não se substituem integralmente, isto é, cada elemento compensa apenas parte das falhas dos demais, e não na mesma medida. Essa leitura não foi testada empiricamente.

Este trabalho tem limitações que precisam ser reconhecidas. Por se tratar de pesquisa bibliográfica, qualitativa e descritiva, nenhuma das relações propostas foi testada empiricamente. O limite de até 30 resultados por combinação, seguindo a relevância da plataforma, restringiu a abrangência da busca sem constituir avaliação de qualidade, a classificação por pergunta e grupo reflete uma leitura própria, e a busca complementar não seguiu o protocolo das demais. Os resultados referem-se aos 16 estudos do *corpus* e não permitem generalizar para outros contextos organizacionais.

Como pesquisas futuras, sugere-se testar empiricamente, com gestores e em decisões mais complexas, as proposições formuladas aqui (a adaptação da forma de apresentar explicações conforme a expertise, sem reduzir a verificação e a rastreabilidade; o efeito da condição *update* conforme a experiência; e o preparo do gestor nos modelos híbridos) e ampliar as buscas.

REFERÊNCIAS

AGRAWAL, Ajay K.; GANS, Joshua S.; GOLDFARB, Avi. Prediction, judgment and complexity: a theory of decision making and artificial intelligence. **NBER Working Paper**, n. 24243, Cambridge, 2018. DOI: 10.3386/w24243. Disponível em: https://www.nber.org/papers/w24243. Acesso em: 14 abr. 2026.

BANDEIRA, Gabriel Lucas; TORTATO, Ubiratã. The future avenues of artificial intelligence and decision-making in business management. **Revista de Administração Mackenzie**, São Paulo, v. 25, n. 6, p. 1-30, 2024. DOI: 10.1590/1678-6971/eRAMD240081. Disponível em: https://www.scielo.br/j/ram/a/skNJNcwqWM7QgRMfVRpVxLx/. Acesso em: 21 mar. 2026.

BORK, Dominik; ALI, Syed Juned; DINEV, Georgi Milenov. AI-enhanced hybrid decision management. **Business & Information Systems Engineering**, [s. l.], v. 65, n. 2, p. 179-199, 2023. DOI: 10.1007/s12599-023-00790-2. Disponível em: https://link.springer.com/article/10.1007/s12599-023-00790-2. Acesso em: 12 abr. 2026.

CORDASCO, Carlo Ludovico; VÉLIZ, Carissa. The accuracy-explainability trade-off, the right to explanation, and the implications for organisations. **Journal of Business Ethics**, [s. l.], 2026. DOI: 10.1007/s10551-026-06305-9. Disponível em: https://link.springer.com/article/10.1007/s10551-026-06305-9. Acesso em: 5 abr. 2026.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

DOGRU, Enis Omer; KRÄMER, Nicole C. Investigating appropriate reliance on AI-based decision support systems: the role of expertise, trust, and self-confidence. **Journal of Decision Systems**, [s. l.], v. 34, n. 1, art. 2593251, 2025. DOI: 10.1080/12460125.2025.2593251. Disponível em: https://www.tandfonline.com/doi/full/10.1080/12460125.2025.2593251. Acesso em: 3 abr. 2026.

DOSHI-VELEZ, Finale; KIM, Been. Towards a rigorous science of interpretable machine learning. **arXiv preprint**, arXiv:1702.08608, 2017. Disponível em: https://arxiv.org/abs/1702.08608. Acesso em: 6 abr. 2026.

GREEN, Ben; CHEN, Yiling. The principles and limits of algorithm-in-the-loop decision making. **Proceedings of the ACM on Human-Computer Interaction**, [s. l.], v. 3, n. CSCW, art. 50, p. 1-24, 2019. DOI: 10.1145/3359152. Disponível em: https://dl.acm.org/doi/10.1145/3359152. Acesso em: 11 abr. 2026.

JABBARI, Amirhossein et al. Appropriate reliance on AI decision support systems: a sociotechnical narrative review. **Information**, [s. l.], v. 17, n. 10, art. 1005, 2026. DOI: 10.3390/info17101005. Disponível em: https://doi.org/10.3390/info17101005. Acesso em: 15 out. 2026.

JALALI, Samireh; WOHLIN, Claes. Systematic literature studies: database searches vs. backward snowballing. In: ACM-IEEE INTERNATIONAL SYMPOSIUM ON EMPIRICAL SOFTWARE ENGINEERING AND MEASUREMENT (ESEM), 2012, Lund. **Proceedings** [...]. New York: ACM, 2012. p. 29-38. DOI: 10.1145/2372251.2372257. Disponível em: https://doi.org/10.1145/2372251.2372257. Acesso em: 24 mar. 2026.

JANSSEN, Marijn *et al.* Will algorithms blind people? The effect of explainable AI and decision-makers' experience on AI-supported decision-making in government. **Social Science Computer Review**, [s. l.], v. 40, n. 2, p. 478-493, 2022. DOI: 10.1177/0894439320980118. Disponível em: https://journals.sagepub.com/doi/10.1177/0894439320980118. Acesso em: 21 mar. 2026.

KITCHENHAM, Barbara. **Procedures for performing systematic reviews**. Keele: Keele University, 2004. Technical Report TR/SE-0401. Disponível em: http://www.inf.ufsc.br/~aldo.vw/kitchenham.pdf. Acesso em: 24 mar. 2026.

KITCHENHAM, Barbara A.; CHARTERS, Stuart. **Guidelines for performing systematic literature reviews in software engineering**. Technical Report EBSE-2007-01. Keele: Keele University; Durham: University of Durham, 2007. Disponível em: https://www.elsevier.com/__data/promis_misc/525444systematicreviewsguide.pdf. Acesso em: 24 mar. 2026.

KOTZIAN, Peter; WEISSENBERGER, Barbara E.; PRINZ, Sonja G. Not so averse after all: behavioral mechanisms underlying the use of algorithms in managerial forecasting. **International Journal of Human–Computer Interaction**, [s. l.], v. 42, n. 7, p. 4907-4931, 2026. DOI: 10.1080/10447318.2025.2543988. Disponível em: https://doi.org/10.1080/10447318.2025.2543988. Acesso em: 2 abr. 2026.

KOVARI, Attila. AI for decision support: balancing accuracy, transparency, and trust across sectors. **Information**, [s. l.], v. 15, n. 11, art. 725, 2024. DOI: 10.3390/info15110725. Disponível em: https://www.mdpi.com/2078-2489/15/11/725. Acesso em: 3 abr. 2026.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.

LUNDBERG, Scott M.; LEE, Su-In. A unified approach to interpreting model predictions. In: ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS (NIPS), 30., 2017, Long Beach. **Proceedings** [...]. La Jolla: NIPS, 2017. p. 4765-4774. Disponível em: https://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions. Acesso em: 7 abr. 2026.

MAHMUD, Hasan *et al.* What influences algorithmic decision-making? A systematic literature review on algorithm aversion. **Technological Forecasting & Social Change**, [s. l.], v. 175, art. 121390, 2022. DOI: 10.1016/j.techfore.2021.121390. Disponível em: https://www.sciencedirect.com/science/article/pii/S0040162521008210. Acesso em: 1 abr. 2026.

PAGE, Matthew J. *et al.* The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. **BMJ**, [s. l.], v. 372, art. n71, 2021. DOI: 10.1136/bmj.n71. Disponível em: https://doi.org/10.1136/bmj.n71. Acesso em: 24 mar. 2026.

PISONI, Galena; MOLONEY, Maria. Responsible AI-based business process management and improvement. **Digital Society**, [s. l.], v. 3, art. 23, p. 1-16, 2024. DOI: 10.1007/s44206-024-00105-2. Disponível em: https://link.springer.com/article/10.1007/s44206-024-00105-2. Acesso em: 10 abr. 2026.

RAISCH, Sebastian; KRAKOWSKI, Sebastian. Artificial intelligence and management: the automation–augmentation paradox. **Academy of Management Review**, [s. l.], v. 46, n. 1, p. 192-210, 2021. DOI: 10.5465/amr.2018.0072. Disponível em: https://journals.aom.org/doi/10.5465/amr.2018.0072. Acesso em: 13 abr. 2026.

RIBEIRO, Marco Tulio; SINGH, Sameer; GUESTRIN, Carlos. "Why should I trust you?": explaining the predictions of any classifier. In: ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING (KDD), 22., 2016, San Francisco. **Proceedings** [...]. New York: ACM, 2016. p. 1135-1144. DOI: 10.1145/2939672.2939778. Disponível em: https://dl.acm.org/doi/10.1145/2939672.2939778. Acesso em: 8 abr. 2026.

RUDIN, Cynthia. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. **Nature Machine Intelligence**, [s. l.], v. 1, n. 5, p. 206-215, 2019. DOI: 10.1038/s42256-019-0048-x. Disponível em: https://www.nature.com/articles/s42256-019-0048-x. Acesso em: 9 abr. 2026.

SARKER, Suprateek; CHATTERJEE, Sutirtha; XIAO, Xiao; ELBANNA, Amany. The sociotechnical axis of cohesion for the IS discipline: its historical legacy and its continued relevance. **MIS Quarterly**, [s. l.], v. 43, n. 3, p. 695-719, 2019. DOI: 10.25300/MISQ/2019/13747. Disponível em: https://doi.org/10.25300/MISQ/2019/13747. Acesso em: 15 abr. 2026.

TEFFÉ, Chiara Spadaccini de; MEDON, Filipe. Responsabilidade civil e regulação de novas tecnologias: questões acerca da utilização de inteligência artificial na tomada de decisões empresariais. **Revista Estudos Institucionais**, Rio de Janeiro, v. 6, n. 1, p. 301-333, 2020. DOI: 10.21783/rei.v6i1.383. Disponível em: https://www.estudosinstitucionais.com/REI/article/view/383. Acesso em: 21 mar. 2026.

WOHLIN, Claes. Second-generation systematic literature studies using snowballing. In: INTERNATIONAL CONFERENCE ON EVALUATION AND ASSESSMENT IN SOFTWARE ENGINEERING (EASE), 20., 2016, Limerick. **Proceedings** [...]. New York: ACM, 2016. Disponível em: https://dblp.dagstuhl.de/rec/conf/ease/Wohlin16.html. Acesso em: 24 mar. 2026.

Este artigo é publicado em acesso aberto (Open Access) sob a licença Creative Commons Atribuição 4.0 Internacional (CC-BY), que permite uso, distribuição e reprodução irrestritos em qualquer meio, desde que o autor original e a fonte sejam creditados.