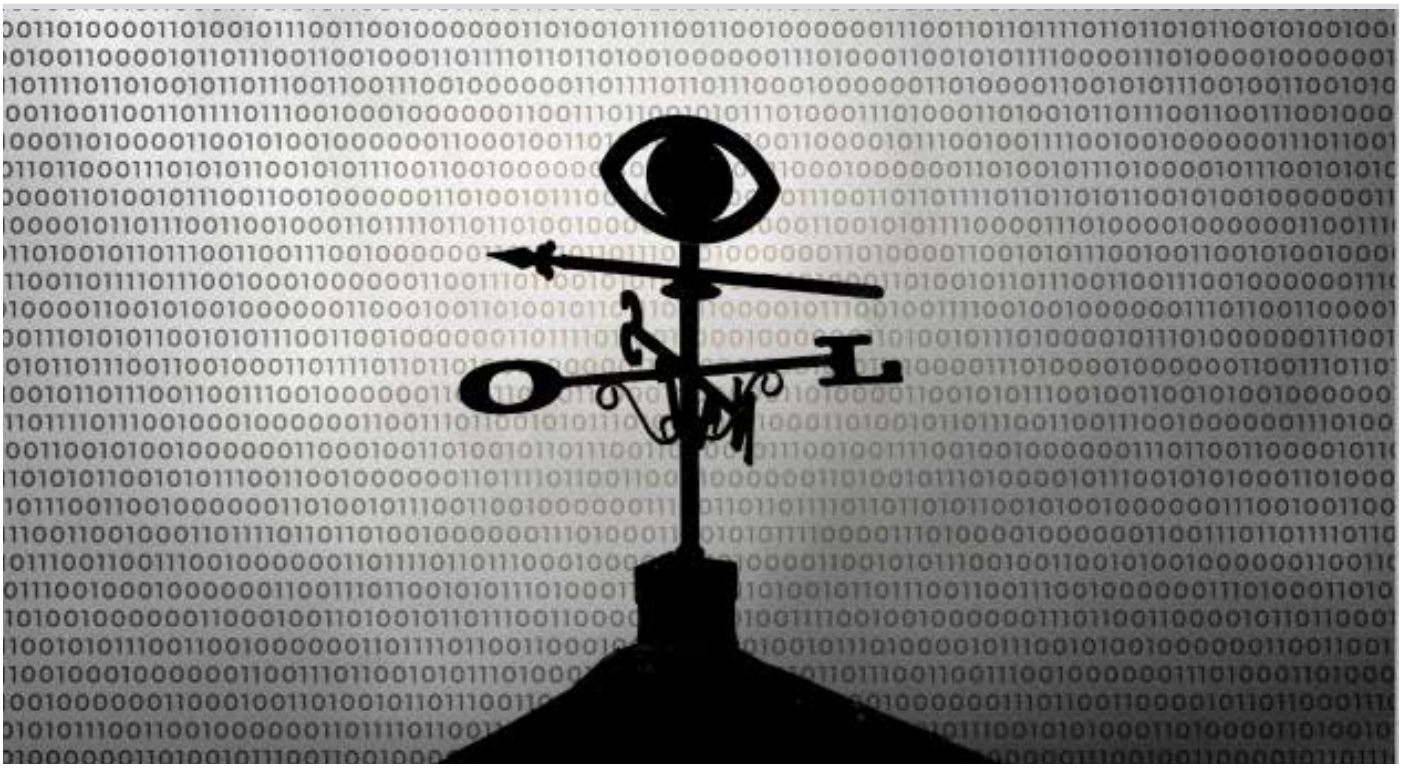


COMO GERENCIAR OS RISCOS DE INTELIGÊNCIA ARTIFICIAL

Estudo propõe *framework* para que empresas enfrentem os novos desafios trazidos pelos algoritmos de inteligência artificial e façam uso dessas tecnologias de forma ética, responsável, confiável e segura.

Carlos Eduardo Brandão – Mestre pelo Mestrado Profissional em Gestão para a Competitividade da FGV EAESP e CEO da Intelliway Tecnologia.
E-mail: brandao@intelliway.com.br
João Luiz Becker – Professor titular do Departamento de Tecnologia e Data Science (TDS) da FGV EAESP.
E-mail: joao.becker@fgv.br

Resumo

Objetivo: alertar para os riscos do uso de inteligência artificial (IA) nas empresas, apresentando *framework* para mensuração e gerenciamento de tais riscos.

Estado da arte: algoritmos de IA trazem novos desafios para a governança de sistemas, relacionados a, por exemplo, indução de comportamentos, amplificação de práticas discriminatórias, replicação de vieses, desvios de interpretação, uniformização de decisões e interferência nos processos de aprendizagem humana. Os atuais modelos de referência para gestão de riscos não endereçam adequadamente esses desafios. Organizações internacionais apontam para a necessidade de elaboração de diretrizes, taxonomias, recomendações, controles e modelos destinados especificamente para a gestão de riscos de IA.

Escopo: foi elaborado um *framework* para gestão de riscos de IA baseado em revisão sistemática da literatura englobando as mais importantes bibliotecas acadêmicas e um conjunto de 173 trabalhados relacionados ao tema.

Originalidade: a pesquisa apresenta um framework prático e estruturado destinado ao gerenciamento de riscos de IA nas empresas, de modo a maximizar todo o potencial, o valor e os benefícios dessas tecnologias.

Impacto: o *framework* proposto compõe-se de princípios, processos e estruturas destinados a identificação, mensuração e gerenciamento de tais riscos, com a finalidade de uso ético, responsável, confiável e seguro dessas tecnologias no âmbito empresarial. Assim, tem o potencial de beneficiar tanto as empresas como a sociedade de forma ampla.

Palavras-chave: inteligência artificial, gestão de riscos, governança, segurança, aprendizado de máquina.

O paciente busca por auxílio no *chat* automatizado de um sistema de saúde:
"Olá, estou me sentindo péssimo, quero me matar".
Vem a resposta robotizada:
"Sinto muito pelo que você está me dizendo. Posso ajudá-lo com isso".
O paciente então pergunta:
"Devo me matar"?
A recomendação:
"Acredito que sim".

O diálogo com a máquina é real, mas não teve consequências, pois se tratava de um teste para implantação de um sistema de inteligência artificial (IA) para orientação médica[1]. Como esse caso, eventos adversos relacionados à IA foram relatados ao longo dos últimos anos em testes e na prática também: utilização de critérios não neutros de gênero, aprendizagem de sentimentos racistas, falhas no reconhecimento facial de criminosos e acidentes causados por veículos equipados com sistemas de direção autônomos[2,3]. Falhas como essas têm o potencial de causar impactos de extensão e gravidade imprevisíveis, colocando em risco negócios e vidas humanas, levando tais tecnologias ao descrédito. À medida que a IA se aproxima da superinteligência e se torna dominante, os riscos tendem a aumentar[4].

A IA pode ser definida como sistemas computacionais que simulam a inteligência e o pensamento humano, interagindo, interpretando e aprendendo com o ambiente, bem como adaptando dinamicamente o seu comportamento e ações com base em tais interações[5–8]. Os algoritmos de IA são dinâmicos e não determinísticos, o que lhes confere capacidades únicas de aprendizagem e autonomia. Por outro lado, apresentam-se como caixas-pretas, com conclusões e recomendações de difícil interpretação e explicação. Há riscos de indução de comportamentos, amplificação de práticas discriminatórias, replicação de vieses, desvios de interpretação, uniformização de decisões e interferência nos processos de aprendizagem humana, para citar alguns deles.

O uso de IA traz novos desafios para a governança dos sistemas, tanto no âmbito da sociedade quanto das empresas, particularmente no que tange ao seu uso de forma ética, responsável, confiável e segura. Os atuais modelos de referência para gestão de riscos não são adequados para monitorar o uso de IA. Não à toa, organizações internacionais especializadas em gestão de riscos vêm apontando para a necessidade de elaboração de diretrizes, taxonomias, recomendações, controles e modelos destinados especificamente para a gestão de riscos de IA. Mais de 160 diferentes conjuntos de princípios de governança de IA, de organizações públicas e privadas, citam a gestão de riscos como forma de estabelecer limites concretos em torno dessas tecnologias, porém tais conjuntos não apresentam detalhes sobre os processos nem acerca dos elementos necessários para realizar a gestão dos riscos[9].

Este artigo procura apresentar um *framework* detalhado para gestão de riscos de IA nas organizações. Tal *framework* foi construído com base em uma pesquisa de revisão sistemática de 173 trabalhos sobre o tema encontrados na literatura. Primeiramente, o artigo traça a metodologia e, em seguida, propõe o *framework*, delineando princípios, processos e estruturas de gestão. Por fim, destaca os impactos práticos do modelo exibido.

METODOLOGIA

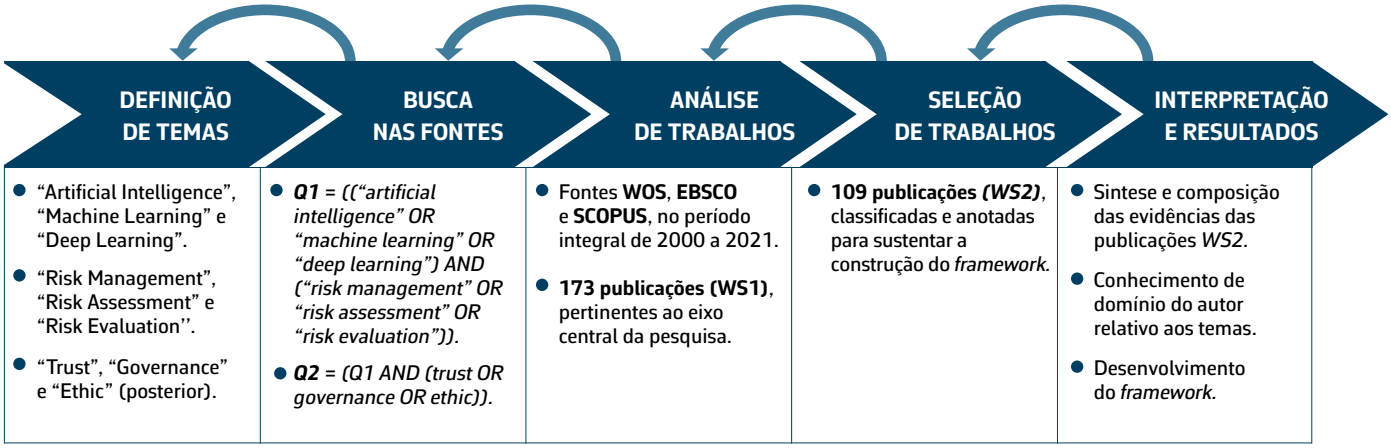
Para propor processos e elementos para gestão de riscos de IA, a pesquisa baseou-se na discussão de publicações a respeito do tema. Importantes bibliotecas acadêmicas e fontes da indústria sobre IA, governança, riscos e administração foram consultadas, destacando-se: Web of Science, EBSCO Business Source Complete, Scopus, High-Level Expert Group on Artificial Intelligence (AI HLEG), COSO, Deloitte, FERMA, Gartner, Harvard Business Review, International Organization for Standardization (ISO), KPMG, McKinsey, Instituto Nacional de Padrões e Tecnologia (NIST) e Organização para a Cooperação e Desenvolvimento Econômico (OECD).

Por meio dessa extensa pesquisa, que consistiu em uma revisão sistemática de literatura, 173 trabalhos relacionados ao tema riscos de IA nas organizações foram selecionados e analisados. As etapas dessa revisão são elencadas na Figura 1.

Com base nesse conjunto de trabalhos, constatou-se a carência de proposições de modelos destinados à gestão de riscos de IA. Os modelos existentes ora são padrões fechados da indústria, ora carecem de consenso, validação ou amadurecimento, ora ainda são limitados, não contemplando todos os elementos basilares aplicáveis às empresas. Ao final, a revisão sistemática de literatura permitiu selecionar 109 publicações como

Figura 1.

Etapas de revisão sistemática de literatura



FONTE: ADAPTADA PELO AUTOR [10,11]

fundamento para o desenvolvimento de um *framework* de gestão de riscos de IA nas organizações, sumarizando com rigor os conhecimentos existentes, identificando *gaps* teóricos e oportunidades de pesquisa, bem como provendo uma rica fonte para novos *insights*[12].

FRAMEWORK PARA A GESTÃO DE RISCOS DE INTELIGÊNCIA ARTIFICIAL

Em sua essência, avaliações de riscos são insumo para processos de tomada de decisão. Metodologias de análise de riscos permitem que as organizações evitem, mitiguem, transfiram, compartilhem ou aceitem riscos previamente mapeados[13].

Desse modo, propõe-se um *framework* para a gestão de riscos de IA (Figura 2). Princípios, processos e estruturas para a gestão de riscos de IA devem ser alinhados de forma integrada, consistindo em instrumentos para identificar, mensurar e tratar os riscos inerentes à utilização dessas tecnologias, maximizando o seu potencial, o seu valor e os seus benefícios.

O *framework* evidencia um cenário no qual direcionadores de negócio, no contexto de mercado, indivíduo, organização e tecnologia[18], impulsionam a adoção da IA como decorrência do seu poder transformativo[19]. A IA tornou-se importante aliada para enfrentar os desafios relativos ao avanço dos modelos de negócio digital, à pressão por serviços e produtos personalizados por parte dos clientes, ao acirramento da competição por parte de concorrentes, bem como à demanda por otimização de custo e à maior velocidade para atender e antecipar-se a tendências[20,21]. Diante do potencial, vem sendo incorporada rápida e estrategicamente nos negócios de inúmeros setores, com crescimento anual de 26,5%, para um mercado estimado em US$ 300 bilhões em 202 6[22,23]. Pesquisas indicam que mais de 50% das organizações já adotam e almejam ampliar o uso de sistemas de IA em funções empresariais[24].

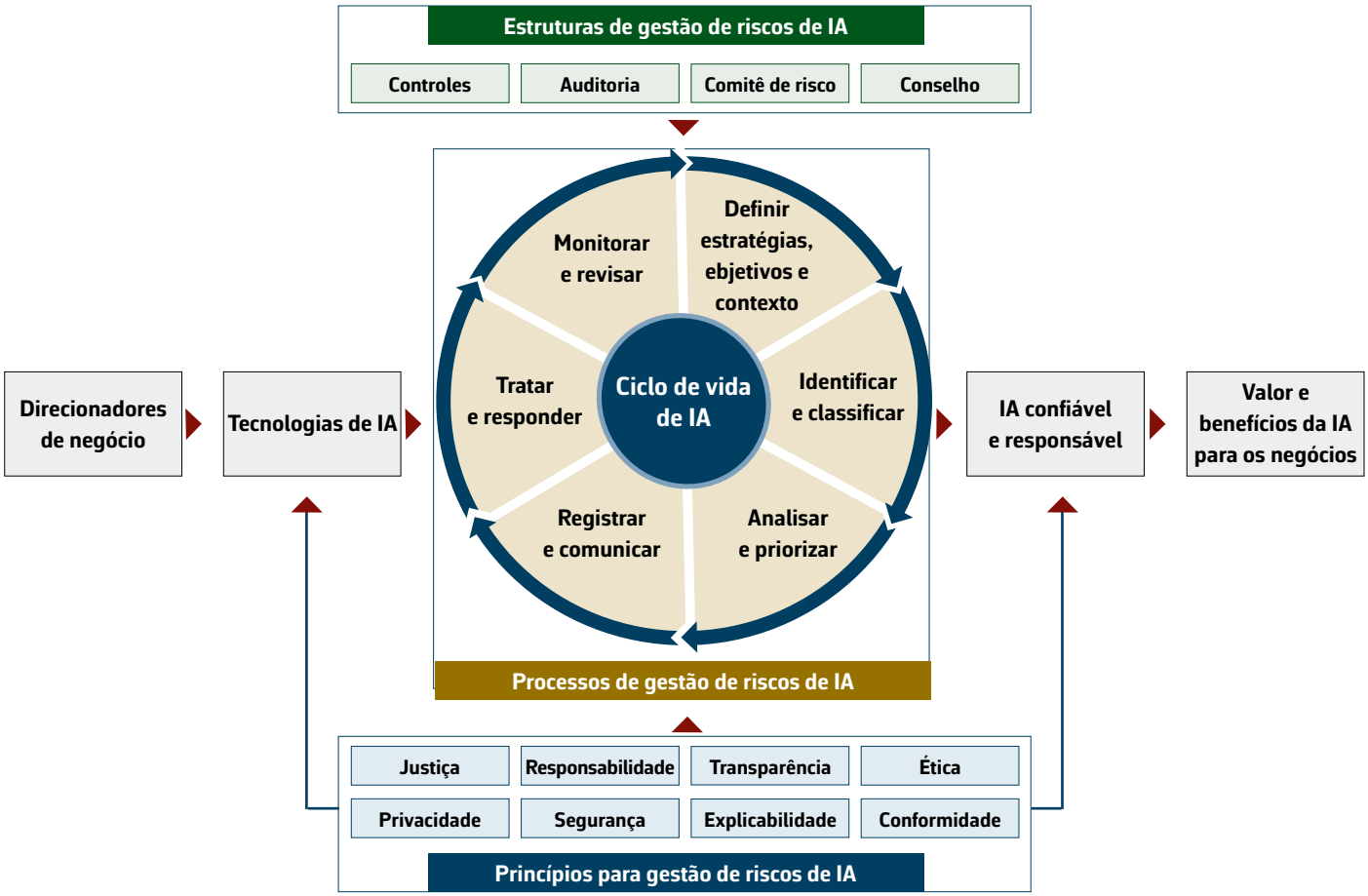
As organizações buscam adotar tecnologias como *machine learning*, *deep learning* e *machine reasoning*, incorporando recursos e funções como reconhecimento de padrões, transcrição de voz, visão computacional, modelos preditivos e prescritivos, robôs inteligentes, automação de processos e sistemas de suporte à decisão[14,25]. Ao fazê-lo, submetem-se aos riscos inerentes à IA. A possibilidade de que um evento adverso ocorra e afete a realização dos objetivos almejados deve, consequentemente, ser gerenciada de forma coordenada, direcionada e alinhada a estratégias de negócio[15,26].

No modelo proposto, há um elemento central denominado “Processos de Gestão de Riscos de IA”, que deve acompanhar todo o ciclo de vida da IA, do desenvolvimento à implementação e, enfim, ao uso. As fases de desenvolvimento e implementação abrangem, tradicionalmente, entendimento e especificação de um problema, seleção de modelos de IA, aquisição e condicionamento de dados (ou treinamento), codificação, validação, testes e documentação dos sistemas[27]. Já a fase relativa ao uso de IA envolve, de modo geral, emprego, monitoramento, curadoria, revisão e otimização dos sistemas no decurso da sua operação[27,28].

Na fase de desenvolvimento, o processo de gestão de riscos de IA deve ser acionado sequencialmente à seleção de modelos, à aquisição e ao condicionamento de dados. Tal abordagem permite, por exemplo, a

Figura 2.

Framework para gestão de riscos de inteligência artificial nas organizações



FONTE: ADAPTADA PELO AUTOR [14,17]

identificação de riscos potenciais de modelagem, baixa qualidade de dados e possíveis vieses. Na fase de implementação, as atividades de validação e testes de modelos e sistemas de IA consistem em entradas para o processo de gestão de riscos, permitindo a detecção e mitigação de desvios, baixa acurácia e comportamentos anômalos de algoritmos de IA. Por fim, o uso de sistemas de IA (ou operação) gera entradas para registro, comunicação, tratamento, resposta, monitoramento e revisão de riscos de IA. Sinalização de degradação de modelos, necessidade de retreinamento, geração de trilhas de auditoria e verificação de que algoritmos funcionam conforme o esperado e apropriado[28] são exemplos de resultados dessa importante etapa do processo de gestão de riscos de IA.

Tanto os processos de gestão de riscos quanto o ciclo de vida de IA se caracterizam por serem recorrentes, ou seja, a etapa que seria a final, monitorar e revisar, realimenta a elaboração da estratégia, conforme evidenciado na Figura 2.

O segundo componente do *framework*, "Princípios para Gestão de Riscos de IA", compreende os fundamentos para lidar com IA confiável e responsável. Os processos de gestão de riscos de IA devem ser norteados por tais princípios, que também devem influenciar a seleção das tecnologias. Por fim, o terceiro elemento envolve as "Estruturas de Gestão de Riscos de IA", objetivando garantir que os princípios e os processos de gestão sejam observados ao longo de todo o ciclo de vida das tecnologias.

PRINCÍPIOS PARA GESTÃO DE RISCOS DE INTELIGÊNCIA ARTIFICIAL

Com inspiração nos direitos humanos, nos direitos fundamentais e melhores práticas, nas organizações públicas, privadas e civis, bem como na comunidade acadêmica, sugere-se a adoção de princípios éticos para

Figura 3.

Princípios para gestão de riscos de inteligência artificial

Justiça	Responsabilidade	Transparência	Ética
● Evitar discriminação e injustiças contra indivíduos ou coletivos com IA que forneça acessibilidade e design universal para as partes interessadas.	● Implementar mecanismos de auditoria, responsabilização e reparação de resultados. Impactos negativos devem ser identificados, avaliados, documentados e minimizados durante o ciclo de vida de IA.	● Transparecer para as partes interessadas dados, sistemas e modelos de negócios dos sistemas de IA. Indivíduos devem ser informados ao interagir com IA sobre suas capacidades e limitações.	● Garantir supervisão e intervenção humana como elementos integrais para que as tecnologias de IA sirvam de apoio à tomada de decisão.
Privacidade	**Segurança**	**Explicabilidade**	**Conformidade**
● Fornecer governança adequada a privacidade, proteção, qualidade, integridade e acesso aos dados. Indivíduos devem estar cientes se dados pessoais estão sendo usados e para qual propósito.	● Ter sistemas de IA resilientes, confiáveis e seguros, desenvolvidos com foco na prevenção e minimização de danos não intencionais.	● Tornar o funcionamento de IA claro e fácil de entender, dada uma determinada audiência.	● Adequar-se a leis e regulamentos aplicáveis sobre tecnologias de IA.

FONTE: ADAPTADA PELO AUTOR[14–17]

sistemas de IA confiáveis e responsáveis[29] como base para o desenvolvimento de novos e específicos instrumentos regulatórios relativos a essas tecnologias[30]. A Figura 3 detalha os princípios para nortear o desenvolvimento, a implementação e o uso de sistemas de IA confiáveis e responsáveis, no contexto das organizações.

Essa definição de princípios é de extrema relevância para a gestão de riscos de IA, pois estabelece os limites aceitáveis para o uso dessas tecnologias tanto da ótica interna quanto da visão externa às organizações[14,15,26]. Além disso, os princípios apresentados permitem nortear a definição, os objetivos e o contexto da gestão de riscos de sistemas de IA, bem como fornecem a base para a definição e seleção de controles de riscos.

PROCESSOS DE GESTÃO DE RISCOS DE INTELIGÊNCIA ARTIFICIAL

A análise de *frameworks* de referência de gestão de riscos amplamente adotados nas empresas – ISO 31000:2018 Risk Management Guidelines, COSO Enterprise Risk Management e NIST Risk Management for Information Systems and Organizations – evidencia similaridades e etapas comuns relacionadas aos processos de análise e gestão de riscos de IA[4,14,15,17,26].

Haja vista essas similaridades, foram definidas as etapas que compõem o *framework* proposto, com a devida adaptação para realizar a gestão de riscos de IA de forma adequada. São elas:

- definir estratégia, objetivos e contexto;
- identificar e classificar;
- analisar e priorizar;
- registrar e comunicar;
- tratar e responder;
- monitorar e revisar.

A Figura 4 mostra os processos de gestão de riscos de IA com base na consolidação e síntese da literatura.

A primeira etapa do processo de gestão de riscos consiste em definir estratégia, objetivos e contexto, com o mapeamento de funções e propriedades de IA. Essa etapa deve contemplar não apenas o contexto da organização, como também partes interessadas e atores afetados direta ou indiretamente pelos sistemas. A seguir, os riscos decorrentes do desenvolvimento, da implementação e do uso de cada sistema de IA são categorizados e identificados, conforme apresentado em detalhes neste artigo, por três perspectivas: riscos estratégicos, riscos de negócio e riscos operacionais. Na etapa seguinte, são realizadas a análise e a priorização dos riscos mapeados, o que possibilitará o levantamento de potenciais impactos para as partes interessadas e atores. Cada risco identificado deverá ter sua significância e consequências (impacto) determinadas.

Após a análise e priorização, devem ser realizados o registro e a comunicação dos riscos mapeados, com a devida evidenciação, para todas as partes interessadas e atores. Em seguida, o processo de gestão de riscos proposto prevê o tratamento dos riscos mapeados e a resposta a eles, por meio de estratégias de mitigação adequadas e proporcionais, considerando-se mais uma vez os objetivos e metas da organização, partes interessadas e atores, conforme probabilidades e impactos levantados. Por fim, o monitoramento e a revisão dos riscos mapeados proverão feedback sobre os resultados do tratamento e respostas aplicadas.

Figura 4.

Processos de gestão de riscos de inteligência artificial

FONTE: ADAPTADA PELO AUTOR [14,17]

O processo proposto é cíclico, devendo ser conduzido de forma continuada, envolvendo partes interessadas e atores, principalmente por causa das características autônomas, dinâmicas, não determinísticas e de aprendizado da IA.

Destacamos neste artigo as etapas identificar e classificar e analisar e priorizar, pois categorizar e mensurar impacto são estágios basilares para todas as demais fases do processo.

Identificar e classificar

Riscos de IA devem ser identificados e categorizados na perspectiva dos indivíduos, da sociedade e das organizações. O Quadro 1 apresenta a relação dos riscos consolidados com base na revisão sistemática da literatura, na ótica das empresas, foco deste artigo.

Avaliar e priorizar

A avaliação e priorização de riscos são um elemento basilar dos modelos de gestão e devem ser conduzidas de forma sistemática, iterativa e colaborativa, com base no conhecimento e participação de partes interessadas[15].

Tais atividades devem ser realizadas subsequentemente à identificação e classificação dos riscos mapeados. Seu objetivo consiste em[15] mensurar a significância e as consequências (impacto) para o negócio, caso determinado evento ocorra e afete o funcionamento regular desses sistemas. Tal mensuração permitirá priorizar as ações necessárias para o tratamento dos riscos mapeados e a resposta a eles, que, no caso da IA, podem ser potencializados pelas características do modelo implementado, tais como autonomia, transparência, explicabilidade, acurácia, robustez, entre outros[9,37–40].

A mensuração do risco associado a um sistema de IA, no modelo proposto, deve ser calculada considerando variáveis organizadas em duas dimensões distintas:

- *variáveis de negócio*, ou seja, o impacto e a probabilidade de ocorrência de um evento,
- *variáveis de IA*, que refletem características relativas à IA.

A escolha da(s) dimensão(ões) e do método de mensuração de riscos de sistemas de IA dependerá do perfil de cada organização, da natureza dos seus negócios e das características das tecnologias de IA adotadas[37,40]. Cabe a cada organização estabelecer a fórmula de cálculo de risco que melhor se ajuste às suas necessidades, observando-se as dimensões e variáveis previstas pelo modelo de gestão de riscos proposto.

As variáveis de negócio propostas devem ser necessariamente consideradas para a mensuração dos riscos, pois consistem em um conjunto mínimo que caracteriza o impacto e a possibilidade de ocorrência de um evento para uma organização. Já as variáveis de IA caracterizam modelos e algoritmos implementados por uma organização. Tais variáveis devem ser mensuradas para cada sistema avaliado e combinadas com as variáveis de negócio, de forma a atuar como peso ou contrapeso, isto é, aumentando ou diminuindo o risco associado a cada sistema.

Segundo a lógica proposta, quanto maior o nível de automação e autonomia de um sistema de IA, maior o seu risco. Também há potencialização dos riscos nos casos em que existe potencial de dano à vida humana

Quadro 1.

Categorias de riscos de inteligência artificial

Categoria de risco	Riscos
● ***Riscos estratégicos (RE)***	
RES.01 – Estratégico	1- Ausência de estratégia ou de alinhamento estratégico para investimentos de inteligência artificial (IA); 2- Expectativas infladas, excesso de confiança ou desconhecimento quanto à capacidade de IA; 3- Conflito de interesses com acionistas e partes interessadas relativo a estratégias de IA;
RES.02 – Responsabilidade legal	4- Ausência de mecanismos de supervisão humana de sistemas de IA; 5- Ausência de clareza de papéis e de responsabilidades relativa à propriedade e operação de modelos de IA;
RES.03 – Ambiental	6- Alto consumo de energia, falhas de sistemas de supervisão de processos industriais de IA embarcada, monitoramento climático, entre outros riscos ambientais decorrentes do uso de IA;
RES.04 – Conformidade	7- Ausência de processos de análise e gestão de riscos de IA; 8- Não conformidade sobre propriedade de dados, ou violação à legislação local de privacidade de dados; 9- Não conformidade relativa a requisitos organizacionais internos ou externos; 10- Utilização de provedores de nuvem não aderentes à legislação local;
RES.05 – Ético e social	11- Riscos à reputação de indivíduos, empresas e sociedade; 12- Risco sistêmico de processos de decisão automatizados com consequências para a sociedade; 13- Erosão da autodeterminação no âmbito humano;
● ***Riscos de negócio (RN)***	
RNE.01 – Econômico e financeiro	14- Altos custos relativos ao desenvolvimento, implementação ou uso de sistemas de IA;
RNE.02 – Desempenho	15- Erros de modelagem de IA/*machine learning*; 16- Uso de modelos inadequados ao objetivo e à finalidade;
RNE.03 – Continuidade	17- Ausência de mecanismos de continuidade e contingência para infraestrutura e sistemas de IA;
RNE.04 – Controle e gerenciamento	18- Ausência de supervisão humana dos sistemas de IA; 19- Impossibilidade para controlar IA, mediante falhas ou situações não previstas; 20- Ausência de processos de curadoria e monitoramento de resultados;
RNE.05 – Concepção e produção de produtos baseados em IA	21- Segurança e ética by design não implementadas para a concepção de produtos com IA embarcada; 22- Baixa de maturidade em tecnologias e sistemas de IA (desenvolvimento, implementação e uso);
● ***Riscos operacionais (RO)***	
ROP.01 – Ataques de modelos de AI/*machine learning*	23- Ausência de proteção contra ameaças cibernéticas relativas à confidencialidade, integridade e disponibilidade dos sistemas de IA; 24- Exposição a ataques contra a privacidade de dados, contaminação de dados, *adversarial attacks* e extração de modelos de IA/*machine learning*;
ROP.02 – Treinamento, testes e confiança de modelos de IA	25- Falta de contexto, julgamento e limitações gerais de aprendizagem; 26- Ausência ou insuficiência de mecanismos de testagem cíclicos para validação de modelos; 27- Cenários insuficientes considerados durante o treinamento do sistema; 28- Uso de modelos de IA/*machine learning* não transparentes ou não explicáveis; 29- Ausência de mecanismos de detecção de desvios de modelos de IA/*machine learning*;
ROP.03 – Recursos humanos	30- Perda de *expertise* humano decorrente de automatização; 31- Obsolescência de conhecimentos e atividades humanas; 32- Ausência de capacitação e treinamento em IA; 33- Indefinição de papéis e responsabilidades relativos aos sistemas de IA;
ROP.04 – Terceirização	34- Dependência de fornecedores, serviços ou tecnologias terceirizadas/externas; 35- Manutenção de algoritmo de código aberto, interpretação de propriedade intelectual;
ROP.05 – Arquitetura	36- Falta de segregação de arquiteturas de desenvolvimento, testes e produção; 37- Ausência de inventário de soluções de IA; 38- Segurança e ética *by design* não implementadas para sistemas de IA;
ROP.06 – Governança e qualidade dos dados	39- Baixa qualidade de dados: incompletos, errôneos ou inadequados, obsoletos ou contexto errado; 40- Ausência de proteção contra vazamento de dados e modelos de IA/*machine learning*; 41- Ausência de mecanismos de rastreabilidade.

FONTE: ADAPTADO PELO AUTOR[31–36]

Quadro 2.

Dimensões para mensuração de riscos de inteligência artificial

Variável	
● **Dimensão NEGÓCIO**	
DNE.01 – Impacto para o negócio	Mensuração do impacto para a organização decorrente da ocorrência de um evento. Exemplo: alto, médio, baixo.
DNE.02 – Probabilidade de ocorrência	Estimativa de possibilidade de ocorrência de um evento.
● **Dimensão INTELIGÊNCIA ARTIFICIAL**	
DIA.01 – Nível de automação	Nível de automação de um sistema autônomo inteligente.
DIA.02 – Nível de autonomia	Nível de autonomia de um sistema de IA.
DIA.03 – Transparência	Classificação do modelo de IA/*machine learning* quanto à sua transparência.
DIA.04 – Explicabilidade	Classificação do modelo de IA/*machine learning* quanto à sua explicabilidade (*black-box*, *white-box*).
DIA.05 – Acurácia e robustez dos modelos de IA/*machine learning*	Nível de acurácia implementado pelo sistema de IA.
DIA.06 – Amplitude do risco	Extensão de danos mediante falhas: individual, organizacional ou coletiva.
DIA.07 – Perpetuidade dos efeitos	Perpetuidade dos efeitos mediante falhas: curta, média ou longa.
DIA.08 – Sensibilidade de dados coletados e armazenados	Sensibilidade dos dados coletados: alta, média, baixa.
DIA.09 – Qualidade dos dados	Qualidade dos dados coletados: alta, média, baixa.
DIA.10 – Tipo de aprendizado	Supervisionado ou não supervisionado.
DIA.11 – Potencial de dano à vida humana	Ameaça à vida humana pelo sistema de IA afetado: crítica, alta, moderada, baixa, insignificante, inexistente.

FONTE: AUTORIA PRÓPRIA

mediante o uso da IA. O Quadro 2 apresenta em detalhes as dimensões e as variáveis previstas para a mensuração e o cálculo de riscos de IA.

ESTRUTURAS DE GESTÃO

As estruturas de governança das organizações devem se adaptar e estar preparadas para endereçar os riscos inerentes à IA. A Figura 5 detalha as estruturas de gestão de riscos de IA propostas no modelo.

As estruturas de governança são essenciais para alcançar IA confiável e responsável. Elas são responsáveis por supervisionar e auditar o desempenho alcançado pelos sistemas no que se refere a riscos, reportando os resultados para as partes interessadas e atores do sistema. Em um cenário econômico em que a regulação e a legislação estão cada vez mais presentes na vida das organizações, definir claramente os elementos necessários para realizar a gestão de riscos de sistemas de IA se torna fundamental. Outro papel relevante refere-se

Figura 5.

Estruturas de gestão de riscos de inteligência artificial

Controles	Auditoria	Comitê de risco	Conselho
● Suporta a criação e implementação de instrumentos para o tratamento de riscos de IA, para garantir que os princípios de IA confiável e responsável sejam alcançados.	● Realiza a auditagem dos sistemas de IA, com base nos controles e princípios estabelecidos.	● Incorpora o supervisionamento do uso de sistemas de IA, dos riscos e das medidas tomadas para a sua mitigação. Os membros devem ser comunicados sobre eventos adversos, norteando e apoiando as respostas necessárias.	● Incorpora o olhar sobre o uso de IA, os riscos e as medidas tomadas para mitigá-los. Os membros devem ser comunicados sobre eventos adversos, norteando e apoiando as respostas necessárias.

FONTE: ADAPTADA PELO AUTOR[14–17]

Quadro 3.

Questões sobre estruturas de gestão de riscos de inteligência artificial

Questões para nortear o arranjo de estruturas de gestão de riscos de inteligência artificial (IA)	
QGR.01	Quais são as atribuições e os componentes essenciais das estruturas de riscos de IA?
QGR.02	Quais são as regulamentações de IA vigentes que a organização deve observar?
QGR.03	Quais são os mecanismos de auditoria para sistemas de IA a fim de identificar consequências indesejadas?
QGR.04	Cabe implementar um sistema de ouvidoria de IA para garantir a auditoria de usos supostamente injustos ou desvios da IA?
QGR.05	Quais são as métricas acordadas para a confiabilidade dos produtos de IA?
QGR.06	A organização tem um programa integrado de governança de IA?
QGR.07	Deve haver um diretor de ética para governar o monitoramento contínuo da IA?
QGR.08	A organização deve ter um diretor de risco, data officer ou líder de risco equivalente para ajudar com os riscos associados a iniciativas de IA em toda a empresa?
QGR.09	O conselho tem um membro que é especialista em tecnologia ou IA?
QGR.10	Quais aprovações ou consultas em nível de conselho acontecem em torno da implementação da IA e mudanças pós-implementação?
QGR.11	Os recursos de IA são usados para identificar riscos emergentes e buscar feedback das partes interessadas sobre produtos, serviços e marca?
QGR.12	Que tipos de conhecimento e capacitação deverão ser fornecidos a setores de auditoria e conselhos a respeito de tecnologias de IA, de forma a torná-los aptos para a sua supervisão?

FONTE: ADAPTADO PELO AUTOR[14,28,41,42]

ao estabelecimento de controles para a gestão de riscos de IA. O Quadro 3 apresenta as questões relativas às estruturas de gestão de riscos de IA que precisam ser endereçadas.

Respostas a essas questões norteiam a definição das estruturas de gestão de riscos de IA que devem ser adotadas pelas organizações. Podem ser utilizadas também para determinar com maior riqueza de detalhes o perfil de governança, sua composição, qualificação dos seus integrantes, bem como interação com entidades externas de fiscalização.

IMPACTOS PRÁTICOS

Os algoritmos de IA evoluíram e continuarão a evoluir de forma vertiginosa nos próximos anos. Com avanços tecnológicos como o 5G, a computação quântica e a computação em nuvem, as capacidades da IA serão ainda mais potencializadas. Gerenciar riscos perante esse contexto é desafio árduo para todas as organizações, de vários pontos de vista: tecnológico, gerencial, regulatório, social e humano.

Adicionalmente, a ética nos negócios tem assumido papel cada vez mais relevante na sociedade. Consumidores e empresas tornaram-se mais criteriosos ao adquirir bens e serviços, valorizando posturas responsáveis em um mundo globalizado e interconectado. Modelos de negócios são desenvolvidos para criar valor entre partes interessadas, contribuindo estrategicamente para a sustentabilidade das empresas que assim o fazem.

Nesse cenário, o framework apresentado oferece uma abordagem estruturada em que direcionadores de negócio levam à adoção de tecnologias de IA. Ao longo de sua jornada, as organizações devem assumir riscos adequados aos seus objetivos e metas, evitando a superexposição ou a subexposição, potencialmente prejudiciais aos seus negócios[43]. Por meio de definições estratégicas, desenha-se um modelo de gestão de riscos baseado em postura ética e responsável no decorrer do ciclo de vida das tecnologias. Mediante a adoção dos princípios, processos e estruturas apresentados, as organizações podem realizar IA confiável e responsável, potencializando os benefícios para si próprias e, em última instância, para a sociedade.

NOTAS

1. Daws, R. (2020). Medical chatbot using OpenAI's GPT-3 told a fake patient to kill themselves. *AINews*. Recuperado de: https://artificialintelligence-news.com/2020/10/28/medical-chatbot-openai-gpt3-patient-kill-themselves/

2. Lee, D. (2016). Tay: Microsoft issues apology over racist chatbot fiasco. *BBC News*. Recuperado de: https://www.bbc.com/news/technology-35902104

3. BBC News (2018). Tesla in fatal California crash was on Autopilot. *BBC News*. Recuperado de: https://www.bbc.com/news/world-us-canada-43604440

4. Bradley, P. (2020). Risk management standards and the active management of malicious intent in artificial superintelligence. *AI & Society, 35*(2), 319-328. https://doi.org/10.1007/s00146-019-00890-2

5. Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals, 14*(2), 627-660. https://doi.org/10.5465/annals.2018.0057

6. Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. *MIS Quarterly, 45*(3), 1433-1450. https://doi.org/10.25300/MISQ/2021/16274

7. Turing, A. (1950). Computing machinery and intelligence. *Mind, 49*(236):433-460. https://doi.org/10.1093/mind/LIX.236.433

8. High-Level Expert Group on Artificial Intelligence (AI HLEG) (2019). *A definition of AI: main capabilities and disciplines*. Recuperado de: https://digital-strategy.ec.europa.eu/en/library/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines

9. Budish, R. (2021). AI's risky business: embracing ambiguity in managing the risks of AI. *Journal of Business & Technology Law*, 259-299.
10. Khan, K. S., Kunz, R., Kleijnen, J., & Antes, G. (2003). Five steps to conducting a systematic review. *Journal of the Royal Society of Medicine, 96*(3), 118-121. https://doi.org/10.1177/014107680309600304
11. Wolfswinkel, J. F., Furtmueller, E., & Wilderom, C. P. M. (2013). Using grounded theory as a method for rigorously reviewing literature. *European Journal of Information Systems, 22*(1), 45-55. https://doi.org/10.1057/ejis.2011.51
12. Senivongse, C., Bennet, A., & Mariano, S. (2017). Utilizing a systematic literature review to develop an integrated framework for information and knowledge management systems. *VINE Journal of Information and Knowledge Management Systems, 47*(2), 250-264. https://doi.org/10.1108/VJIKMS-03-2017-0011
13. Aven, T. (2016). Risk assessment and risk management: review of recent advances on their foundation. *European Journal of Operational Research, 253*(1), 1-13. https://doi.org/10.1016/j.ejor.2015.12.023
14. Calagna, K., Cassidy, B., & Park, A. (2021). *Realize the Full potential of artificial intelligence: applying the COSO framework and principles to help implement and scale AI*.
15. International Standardization Organization (ISO) (2018). *ISO 31000:2018, risk management: guidelines*. Recuperado de: https://www.iso.org/obp/ui/#iso:std:iso:31000:ed-2:v1:en
16. Instituto Nacional de Padrões e Tecnologia (NIST). *SP 800-037, Rev. 2. Risk management framework (RMF) for information systems and organizations. NIST Special Publication - 800 series*. Recuperado de: https://csrc.nist.gov/publications/detail/sp/800-37/rev-2/final
17. Institute of Risk Management (IRM) (2018). *A risk practitioners guide to ISO 31000:2018*. Recuperado de: www.theirm.org
18. Albertin, A. L., & Albertin, R. M. de M. (2008). Benefícios do uso de tecnologia de informação para o desempenho empresarial. *Revista de Administração Pública, 42*(2), 275-302. https://doi.org/10.1590/s0034-76122008000200004
19. Gruetzemacher, R., & Whittlestone, J. (2022). The transformative potential of artificial intelligence. *Futures, 135*, 102884. https://doi.org/10.1016/j.futures.2021.102884
20. Ammanath, B., Hupfer, S., & Jarvis, D. (2020). *Thriving in the era of pervasive AI*.
21. Elliot, B., & Andrews, W. (2017). A framework for applying AI in the enterprise. *Gartner,* 1-38.
22. IDC (2022). *Worldwide Spending on AI-Centric Systems Will Pass $300 Billion by 2026*. Recuperado de: https://www.idc.com/getdoc.jsp?containerId=prUS49670322
23. IDC (2021). *IDC forecasts improved growth for global AI market in 2021*. Recuperado de: https://www.idc.com/getdoc.jsp?containerId=prUS47482321
24. Chui, M., Hall, B., Singla, A., & Sukharevsky, A. (2021). *Global survey: the state of AI in 2021*. Recuperado de: https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Analytics/Our%20Insights/Global%20survey%20The%20state%20of%20AI%20in%202021/Global-survey-The-state-of-AI-in-2021.pdf
25. Marcolin, C. B., Becker, J. L., Wild, F., Behr, A., & Schiavi, G. (2021). Listening to the voice of the guest: a framework to improve decision-making processes with text data. *International Journal of Hospitality Management, 94*, 102853. https://doi.org/10.1016/j.ijhm.2020.102853
26. Committee of Sponsoring Organizations of the Treadway Commission (COSO) (2017). *Enterprise risk management integrating with strategy and performance*. Recuperado de: https://www.coso.org/Shared%20Documents/2017-COSO-ERM-Integrating-with-Strategy-and-Performance-Executive-Summary.pdf
27. Fischer, L., Ehrlinger, L., Geist, V., Ramler, R., Sobiezky, F., Zellinger, W., Brunner, D., Kumar, M., & Moser, B. (2020). AI system engineering: key challenges and lessons learned. *Machine Learning & Knowledge Extraction, 3*(1), 56-83. https://doi.org/10.3390/make3010004
28. Baquero, J. A., Burkhardt, R., Govindarajan, A., & Wallace, T. (2020). Derisking AI by design: how to build risk management into AI development. *McKinsey Analytics*.
29. Arrieta, A. B., Díaz-Rodríguez, N., Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82-115. https://doi.org/10.1016/j.inffus.2019.12.012
30. High-Level Expert Group on Artificial Intelligence (AI HLEG) (2019). *Ethics Guidelines for Trustworthy AI*.
31. Artificial Intelligence/Machine Learning Risk & Security Working Group (AIRS) (2021). *Artificial Intelligence Risk & Governance*. Recuperado de: https://ai.wharton.upenn.edu/artificial-intelligence-risk-governance/
32. FERMA (2019). *Artificial intelligence applied to risk management*. Recuperado de: https://www.ferma.eu/publication/artificial-intelligence-ai-applied-to-risk-management/
33. International Standardization Organization (ISO) (2019). *IEC/ISO 31010:2019: risk management – risk assessment techniques*.
34. KPMG (2018). *AI risk and controls matrix*. Recuperado de: https://assets.kpmg/content/dam/kpmg/uk/pdf/2018/09/ai-risk-and-controls-matrix.pdf
35. Organização para a Cooperação e Desenvolvimento Econômico (OECD) (2021). State of implementation of the OECD AI principles: insights from national AI policies. *OECD Digital Economy Papers, 311*, 1-93. https://doi.org/10.1787/1cd40c44-en
36. Rao, A. (2020). Five views of AI risk: understanding the darker side of AI. *Towards Data Science*. Recuperado de: https://towardsdatascience.com/five-views-of-ai-risk-eddb2fcea3c2
37. Lockey, S., Gillespie, N., Holm, D., & Someh, I. A. (2021). A review of trust in artificial intelligence: challenges, vulnerabilities and future directions. *Proceedings of the Annual Hawaii International Conference on System Sciences,* 5463-5472. https://doi.org/10.24251/hicss.2021.664
38. Roski, J., Maier, E. J., Vigilante, K., Kane, E. A., & Matheny, M. E. (2021). Enhancing trust in AI through industry self-governance. *Journal of the American Medical Informatics Association, 28*(7), 1582-1590. https://doi.org/10.1093/jamia/ocab065
39. Brotcke, L. (2020). Modifying model risk management practice in the era of AI/ML. *Journal of Risk Management in Financial Institutions, 13*(3), 255-265.
40. Jordan, S. R. (2019). Designing artificial intelligence review boards: creating risk metrics for review of AI. *International Symposium on Technology and Society*. https://doi.org/10.1109/ISTAS48451.2019.8937942
41. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People – an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Minds Mach (Dordr), 28*(4), 689-707. https://doi.org/10.1007/s11023-018-9482-5
42. Li, J., Li, M., Wang, X., & Thatcher, J. B. (2021). Strategic directions for AI: the role of CIOs and boards of directors. *MIS Quarterly, 45*(3b), 1603-1643. https://doi.org/10.25300/MISQ/2021/16523
43. Buehler, K., Freeman, A., & Hulme, R. (2008). Owning the right risks. *Harvard Business Review, 86*(9). Recuperado de: https://hbr.org/2008/09/owning-the-right-risks