MINISTÉRIO DA SAÚDE
INSTÂNCIA NACIONAL DE ÉTICA EM PESQUISA — INAEP

GUIA DE USO ÉTICO DE INTELIGÊNCIA ARTIFICIAL EM PESQUISA COM SERES HUMANOS

ORIENTAÇÕES PARA AVALIAÇÃO ÉTICA — VERSÃO 1.0

Brasília — DF
2026

Elaboração, distribuição e informações:

MINISTÉRIO DA SAÚDE

Secretaria de Ciência, Tecnologia e Inovação em Saúde

Departamento de Ciência e Tecnologia

Instância Nacional de Ética em Pesquisa

SRTV, quadra 702, via W5 Norte, Edifício PO 700, 5º andar

CEP: 70719-040 – Brasília/DF

Tel.: (61) 3315-2425

Site: www.gov.br/saude

E-mail: inaep@saude.gov.br *(confirmar)*

Supervisão:

Meiruze Sousa Freitas

Elaboração Textual:

Fabiano Tonaco Borges

Revisão Técnica: Grupo de Trabalho do Guia de Diretrizes Éticas para o Uso de Inteligência Artificial em Pesquisa com Seres Humanos da INAEP. **Integrantes:** Fabiano Tonaco Borges (Coordenador), Alexandre Dias Porto Chiavegatto Filho, Fábio de Oliveira Aquino, Felipe de Oliveira Franco, Luiz Vianna Sobrinho, Maíra Araújo de Santana, Roseli Mieko Yamamoto Nomura e Wellington Pinheiro dos Santos.

O Guia foi aprovado por unanimidade pelo Colegiado da INAEP durante a 13ª Reunião Ordinária da Inaep, com relatoria de Meiruze Sousa Freitas, nos termos do Voto nº 25/2026.

LISTA DE QUADROS

Quadro 1 – Como interpretar a linguagem deste guia 7
Quadro 2 – Elementos de um sistema de IA e implicações éticas 13
Quadro 3 – Famílias de sistemas de IA e questões éticas prioritárias 14
Quadro 4 – Experiências internacionais e contribuições incorporadas ao modelo brasileiro 23
Quadro 5 – Síntese das recomendações sobre integridade científica, autoria e uso de IA 26
Quadro 6 – Função dos módulos do conjunto documental 33

SUMÁRIO

INTRODUÇÃO 5

1.1 NOTA EDITORIAL SOBRE ESTA REVISÃO 7

1.2 COMO UTILIZAR O CONJUNTO DOCUMENTAL 8

1.3 RELAÇÃO COM O CADERNO REFERENCIAL E A MARIAH 9

FINALIDADE, ÂMBITO E ABORDAGEM DA AVALIAÇÃO ÉTICA 10

2.1 FINALIDADE 10

2.2 ÂMBITO DE APLICAÇÃO 10

2.3 MODALIDADES DE PESQUISA RELACIONADA À INTELIGÊNCIA ARTIFICIAL 10

2.4 ABORDAGEM ORIENTADA PELO CONTEXTO DE USO 11

2.5 PROPORCIONALIDADE 12

CONCEITOS ESSENCIAIS PARA A AVALIAÇÃO DE SISTEMAS DE IA 13

3.1 SISTEMA, MODELO E SAÍDA 13

3.2 TREINAMENTO, VALIDAÇÃO E TESTE 13

3.3 FAMÍLIAS DE SISTEMAS E RISCOS ASSOCIADOS 14

3.4 DADOS SINTÉTICOS 15

DIMENSÕES CRÍTICAS DE RISCO 16

4.1 VIÉS, REPRESENTATIVIDADE E EQUIDADE 16

4.2 EXPLICABILIDADE E TRANSPARÊNCIA 17

4.3 SUPERVISÃO HUMANA EFETIVA 17

4.4 CICLO DE VIDA, ATUALIZAÇÃO E DERIVA 18

4.4.1 Monitoramento, mudanças e rastreabilidade 19

4.5 PESSOAS, GRUPOS E COMUNIDADES AFETADOS ALÉM DOS PARTICIPANTES FORMAIS 19

CONTEXTO BRASILEIRO E INTERESSE PÚBLICO 21

5.1 ADEQUAÇÃO AO CONTEXTO BRASILEIRO 21

5.2 SUS E DADOS DE INTERESSE PÚBLICO 21

5.3 SAÚDE DIGITAL, INTEROPERABILIDADE E SOBERANIA DE DADOS 22

REFERÊNCIAS INTERNACIONAIS E PRINCÍPIOS INCORPORADOS 23

INTEGRIDADE CIENTÍFICA, AUTORIA E USO DE IA NA PRODUÇÃO DA PESQUISA 25

7.1 DECLARAÇÃO DO USO DE IA 25

7.2 AUTORIA E RESPONSABILIDADE HUMANA 25

7.3 ALUCINAÇÃO, FABRICAÇÃO E NEGLIGÊNCIA 26

7.4 O QUE SE RECOMENDA PARA O PROTOCOLO — E O QUE O CEP DEVE VERIFICAR 26

GOVERNANÇA DE DADOS E PROTEÇÃO DE PARTICIPANTES 27

8.1 ORIGEM, FINALIDADE E PAPÉIS 27
8.2 BANCOS DE DADOS, USO FUTURO E CONSENTIMENTO 27
8.3 REIDENTIFICAÇÃO, INFERÊNCIA E SEGURANÇA 28
8.4 TRANSFERÊNCIA INTERNACIONAL E SOBERANIA DE DADOS 28
8.5 TECNOLOGIAS DE PRESERVAÇÃO DE PRIVACIDADE 28
8.6 DIREITOS DOS PARTICIPANTES E DESTINO DO MODELO 29
ESPECIFICIDADES POR NATUREZA DA PESQUISA 30
9.1 CIÊNCIAS DA SAÚDE: IMPACTO ASSISTENCIAL E SAMD 30
9.2 CIÊNCIAS HUMANAS E SOCIAIS: INFERÊNCIAS, GRUPOS E RISCO COLETIVO 31
9.3 PESQUISAS INTERDISCIPLINARES 32
INTEGRAÇÃO COM A MARIAH E OS INSTRUMENTOS DE APLICAÇÃO 33
10.1 APLICAÇÃO DA MARIAH 33
10.2 REGISTRO DA DECISÃO ÉTICA 33
10.3 SUPERVISÃO ÉTICA LONGITUDINAL E EVENTOS QUE EXIGEM NOVA APRECIAÇÃO 34
10.4 RESPONSABILIDADES COMPARTILHADAS E LIMITES DA COMPETÊNCIA DO CEP 34
10.5 USO DE INTELIGÊNCIA ARTIFICIAL NO APOIO ÀS ATIVIDADES DO CEP 35

INTRODUÇÃO

A inteligência artificial (IA) já integra diferentes etapas das pesquisas com seres humanos desenvolvidas no Brasil, tanto nas Ciências da Saúde quanto nas Ciências Humanas e Sociais (CHS). Sistemas de apoio à análise e ao diagnóstico por imagem, modelos preditivos, ferramentas de processamento de linguagem natural, métodos de análise automatizada de textos, imagens e registros, sistemas generativos, técnicas de classificação e perfilamento e aplicações baseadas em dados sintéticos são exemplos de tecnologias que podem estar presentes em protocolos submetidos à apreciação dos Comitês de Ética em Pesquisa (CEPs).

A incorporação dessas tecnologias amplia as possibilidades científicas e metodológicas da pesquisa, mas também introduz questões específicas para a avaliação ética. Sistemas de IA podem modificar a forma como dados são coletados, combinados, processados, analisados e interpretados; produzir novas inferências sobre pessoas, grupos e comunidades; influenciar decisões; ampliar a escala e a velocidade do tratamento de informações; e possibilitar usos dos dados, modelos ou resultados que ultrapassem a finalidade inicialmente prevista no protocolo.

Essas características demandam atenção a dimensões que podem exigir critérios e instrumentos específicos ou complementares de avaliação ética, entre elas a origem, a qualidade e a representatividade dos dados; os vieses; a produção de inferências sensíveis; a explicabilidade e a transparência; a supervisão humana; o grau de autonomia dos sistemas; a possibilidade de reidentificação; os riscos individuais e coletivos; a governança dos dados; e o destino dos dados, modelos e resultados após o encerramento da pesquisa.

É nesse contexto que a Instância Nacional de Ética em Pesquisa (INAEP) apresenta este Guia, destinado a orientar pesquisadores, instituições e CEPs na identificação, análise e tratamento das questões éticas relacionadas ao uso de inteligência artificial em pesquisas com seres humanos nas Ciências da Saúde e nas Ciências Humanas e Sociais.

O Guia parte dos princípios e referenciais que orientam a ética em pesquisa com seres humanos e os aplica às particularidades introduzidas pela IA. Seu propósito não é substituir a legislação, as normas éticas aplicáveis, as competências de outras autoridades ou a deliberação dos CEPs, mas oferecer orientações, critérios e instrumentos que apoiem a aplicação desses referenciais quando as características da tecnologia demandarem avaliação específica.

A utilização de inteligência artificial não constitui, por si só, uma categoria uniforme de risco. Uma mesma tecnologia pode apresentar implicações éticas distintas conforme a finalidade, a população envolvida, a natureza e a origem dos dados, o grau de autonomia do sistema, as inferências produzidas, a influência de seus resultados sobre decisões e as consequências potenciais para participantes, grupos e comunidades. Por essa razão, este Guia adota uma abordagem baseada no contexto de uso e na proporcionalidade do risco.

O Guia reconhece que a utilização de IA pode produzir questões éticas distintas nas Ciências da Saúde e nas Ciências Humanas e Sociais, sem estabelecer separação rígida entre os campos. Pesquisas interdisciplinares podem reunir características de ambas as áreas. Recomenda-se que a avaliação seja orientada pelas características concretas do protocolo e pelos riscos efetivamente envolvidos, e não exclusivamente pela área disciplinar em que a pesquisa se classifica.

Este Guia foi elaborado para o contexto brasileiro e considera a diversidade social, cultural, epidemiológica, étnica, linguística, econômica e territorial do País. Nesse cenário, o Sistema Único de Saúde (SUS) constitui contexto particularmente relevante para pesquisas que utilizem dados, serviços, infraestruturas ou populações vinculadas ao sistema público de saúde. A utilização de dados do SUS, assim como de outras bases públicas, institucionais ou digitais, requer atenção à proteção individual e aos possíveis efeitos das análises, classificações e inferências sobre grupos, comunidades e populações.

Este Guia possui caráter orientador e operacional. Seu objetivo é apoiar pesquisadores e instituições na elaboração de protocolos e auxiliar os CEPs na identificação, análise e avaliação proporcional dos riscos relacionados ao uso de IA em pesquisas com seres humanos. Não pretende transformar pesquisadores ou membros dos CEPs em especialistas em inteligência artificial, mas estabelecer quais informações são eticamente relevantes, quais questões devem ser examinadas e quais evidências podem ser necessárias para fundamentar a avaliação ética.

O propósito é qualificar e tornar mais consistente, proporcional e transparente o julgamento ético, sem substituí-lo por decisão automatizada ou exclusivamente técnica. Para apoiar esse processo, o Guia incorpora a MARIAH — Matriz de Avaliação de Risco em Inteligência Artificial em Pesquisa com Seres Humanos — como instrumento destinado a sistematizar a identificação e a classificação dos riscos e a orientar requisitos proporcionais às características do protocolo. A MARIAH não aprova nem reprova pesquisas.

O conjunto está organizado em três módulos complementares: Guia de Uso Ético de Inteligência Artificial em Pesquisa com Seres Humanos — Orientações para Avaliação Ética; Caderno Referencial — Fundamentos Éticos, Conceituais, Normativos e Metodológicos; e MARIAH — Matriz de Avaliação de Risco em Inteligência Artificial em Pesquisa com Seres Humanos. Recomenda-se que os três módulos sejam utilizados de forma articulada, conforme a natureza, a complexidade e o nível de risco da pesquisa.

A inteligência artificial é um campo em rápida transformação. Por essa razão, este Guia se estrutura prioritariamente em torno de questões éticas e dimensões de risco que permanecem relevantes mesmo quando a tecnologia se modifica. A tecnologia pode apoiar a análise, mas não substitui a responsabilidade, a deliberação ou o julgamento ético.

1.1 NOTA EDITORIAL SOBRE ESTA REVISÃO

Esta versão reorganiza e reescreve o texto de orientações do documento de trabalho da INAEP que deu origem a este Guia, para conferir maior clareza institucional, maior precisão regulatória e melhor separação entre fundamento, requisito e instrumento de aplicação.

O conteúdo operacional da matriz foi retirado deste Guia e concentrado no módulo autônomo MARIAH. As listas de verificação, os exercícios e os instrumentos de aplicação integram a MARIAH.

Regra adotada:

- O Guia estabelece a orientação e os fundamentos necessários para compreender e aplicar os requisitos.
- O Caderno Referencial reúne os fundamentos que sustentam essas orientações, com função explicativa, sem criar requisitos adicionais.
- A MARIAH traduz essas orientações em instrumentos de verificação e organiza a identificação, a classificação e a documentação do risco. A decisão ética permanece humana, colegiada e fundamentada.

Para orientar a interpretação das recomendações apresentadas neste Guia, as expressões utilizadas devem ser compreendidas segundo o grau de exigência e os efeitos esperados, conforme apresentado no Quadro 1.

Quadro 1 – Como interpretar a linguagem deste guia

Expressão	Uso no Guia	Efeito esperado
RECOMENDA-SE	Requisito decorrente do marco normativo aplicável ou condição considerada necessária para que o protocolo seja eticamente avaliável.	Recomenda-se que o protocolo demonstre o atendimento ou justifique, quando juridicamente admissível, a não aplicabilidade.
RECOMENDA-SE	Boa prática que fortalece proteção, transparência, qualidade, governança ou rastreabilidade.	A não adoção não implica automaticamente inadequação, mas pode exigir justificativa em contextos de maior risco.
PODE	Alternativa admissível ou faculdade dependente do contexto.	A adoção é discricionária e recomenda-se que seja compatível com a natureza e o risco da pesquisa.

Fonte: elaboração própria.

Fundamento

O uso dessas expressões não altera a hierarquia das leis, decretos, resoluções e demais normas vigentes. Quando o Guia traduz uma obrigação normativa em requisito operacional, recomenda-se que a fonte correspondente seja indicada na versão editorial final.

1.2 COMO UTILIZAR O CONJUNTO DOCUMENTAL

Para a orientação institucional sobre os aspectos éticos a considerar, recomenda-se consultar este Guia. Para aplicação da classificação de risco, dos *checklists* e do registro da decisão, recomenda-se utilizar a MARIAH, em sua versão documental ou na ferramenta eletrônica de acesso livre. Quando houver necessidade de compreender a fundamentação de uma recomendação, de um eixo de risco ou de uma salvaguarda, recomenda-se consultar o Caderno Referencial.

A Figura 1 apresenta o percurso recomendado de avaliação ética de pesquisas com IA.

Figura 1 — Percurso recomendado de avaliação ética de pesquisas com IA



Fonte: elaboração própria.

Nota: figura elaborada com base na abordagem de ciclo de vida adotada neste Guia e nas orientações da Medicines and Healthcare products Regulatory Agency (MHRA), do Reino Unido.

O **Caderno Referencial** possui função explicativa e não estabelece requisitos adicionais para a apreciação ética dos protocolos.

1.3 RELAÇÃO COM O CADERNO REFERENCIAL E A MARIAH

As indicações abaixo orientam a navegação entre as três peças.

- **Contexto de uso, proporcionalidade e dimensões críticas —** MARIAH para aplicação operacional; Caderno Referencial para fundamentação conceitual e metodológica.
- **Integridade científica e IA generativa —** Caderno Referencial para aprofundamento; MARIAH para verificação.
- **Governança de dados e direitos dos participantes —** Caderno Referencial para fundamentação normativa e técnica; MARIAH para classificação e documentação.
- **Ciências da Saúde e SaMD —** MARIAH para fatores de risco; Caderno Referencial para suporte técnico e regulatório.
- **Ciências Humanas e Sociais —** MARIAH para inferências sensíveis, riscos coletivos e usos secundários; Caderno Referencial para fundamentação específica.

FINALIDADE, ÂMBITO E ABORDAGEM DA AVALIAÇÃO ÉTICA

1.4 FINALIDADE

Este Guia orienta a elaboração e a avaliação ética de pesquisas com seres humanos que envolvam sistemas ou ferramentas de inteligência artificial. A orientação aplica-se tanto às pesquisas em que a IA constitui objeto de investigação quanto às pesquisas em que ela é utilizada como ferramenta, método de análise, componente de decisão, geradora de conteúdo ou mecanismo de produção de inferências.

Recomenda-se que a avaliação se concentre nos efeitos concretos da tecnologia sobre o protocolo e sobre as pessoas, grupos ou comunidades potencialmente afetadas. A denominação comercial do sistema, a arquitetura técnica ou a área disciplinar, isoladamente, não determinam o nível de atenção ética.

1.5 ÂMBITO DE APLICAÇÃO

Recomenda-se considerar este Guia quando a pesquisa utilizar IA para, entre outras finalidades:

- analisar, classificar, predizer ou interpretar dados relacionados a pessoas;
- produzir, transformar ou resumir conteúdo que integre o protocolo, os instrumentos de pesquisa, o consentimento ou os resultados;
- apoiar ou automatizar decisões que possam afetar participantes;
- desenvolver, treinar, ajustar, testar ou validar modelos ou sistemas de IA;
- produzir dados sintéticos utilizados como parte da evidência científica;
- processar textos, imagens, áudios, registros clínicos, dados administrativos ou dados provenientes de ambientes digitais;
- produzir inferências sobre atributos individuais ou coletivos.

1.6 MODALIDADES DE PESQUISA RELACIONADA À INTELIGÊNCIA ARTIFICIAL

Para fins de triagem e definição das informações necessárias à avaliação, recomenda-se identificar se o protocolo envolve: (a) pesquisa relacionada à saúde que utiliza dados e métodos de IA; (b) pesquisa conduzida com ferramentas ou tecnologias de IA; ou (c) pesquisa

sobre o desenvolvimento, treinamento, teste, validação ou desempenho de ferramentas e tecnologias de IA. As modalidades podem coexistir no mesmo protocolo e não determinam, isoladamente, o nível de risco.

Recomenda-se que o protocolo apresente:

- **a modalidade ou as modalidades de uso da IA presentes no estudo;**
- **a função desempenhada pela IA em cada etapa da pesquisa;**
- **se a IA atua sobre dados, instrumentos, interação com participantes, produção de inferências ou decisões;**
- **se a ferramenta utilizada foi desenvolvida especificamente para a pesquisa ou fornecida por terceiro.**

O CEP verifica:

- se todos os usos relevantes da IA foram declarados, inclusive os considerados auxiliares;
- se a modalidade informada corresponde ao uso efetivamente descrito no protocolo;
- se o enquadramento permite identificar participantes formais, pessoas indiretamente afetadas, grupos e comunidades potencialmente impactados.

1.7 ABORDAGEM ORIENTADA PELO CONTEXTO DE USO

Recomenda-se que o protocolo defina o contexto de uso da IA antes da análise de risco. A mesma tecnologia pode apresentar riscos distintos conforme a finalidade, a população, a fonte de dados, o grau de autonomia, o ambiente de aplicação e a possibilidade de intervenção humana.

Recomenda-se que o protocolo apresente:

- qual é a finalidade específica da IA no estudo;
- qual pergunta ou tarefa o sistema executa;
- em que etapa da pesquisa a IA é utilizada;
- quem recebe e utiliza a saída produzida;
- se a saída influencia decisão sobre participante, grupo ou comunidade;
- se existe intervenção humana antes de qualquer consequência relevante.

Recomenda-se que o CEP verifique:

- se o papel da IA está descrito de forma suficientemente clara para permitir avaliação ética;
- se o contexto de uso real coincide com o contexto em que o sistema foi desenvolvido ou validado;
- se a influência do sistema sobre decisões foi explicitada;
- se a descrição permite identificar quem pode ser afetado pela utilização da IA.

Sinal de alerta:

- o protocolo menciona IA apenas de forma genérica, sem descrever sua função;
- o pesquisador descreve o sistema como 'apoio' sem esclarecer quanto ele influencia a decisão;
- o uso de IA aparece somente em anexo técnico, sem integração com a análise ética do protocolo.

1.8 PROPORCIONALIDADE

Recomenda-se que a profundidade da avaliação seja proporcional ao risco. O uso de IA não deve, por si só, gerar exigências idênticas para todos os protocolos. A MARIAH é o instrumento destinado a apoiar essa classificação e a documentar os fatores de risco relevantes.

Protocolos de menor risco podem requerer apenas identificação clara do uso da IA, transparência e proteção adequada dos dados. Protocolos de maior risco podem exigir validação contextual, análise de desempenho por subgrupos, supervisão humana efetiva, documentação ampliada de governança, monitoramento e salvaguardas adicionais.

Fundamento:

O documento de origem estabelece que perguntas éticas estáveis — representatividade dos dados, supervisão humana, destino do modelo e proteção dos participantes — recomenda-se que orientem a avaliação mesmo diante da rápida mudança das arquiteturas de IA. A revisão mantém essa racionalidade, substituindo a linguagem ensaística por critérios operacionais.

CONCEITOS ESSENCIAIS PARA A AVALIAÇÃO DE SISTEMAS DE IA

Este capítulo reúne apenas os conceitos necessários para a avaliação ética. A compreensão técnica aprofundada de arquiteturas e algoritmos não é pressuposto para a atuação do CEP. O objetivo é permitir que pesquisador e avaliador utilizem linguagem comum e identifiquem as informações necessárias à análise.

1.9 SISTEMA, MODELO E SAÍDA

Para fins deste Guia, um sistema de IA processa dados ou informações e produz saídas como classificações, probabilidades, escores, recomendações, previsões ou conteúdos. Recomenda-se que a avaliação ética considere o sistema em seu contexto de uso, e não apenas o modelo matemático isolado.

O Quadro 2 apresenta os elementos básicos de um sistema de IA — entrada, processamento e saída — e as implicações éticas associadas a cada um deles.

Quadro 2 – Elementos de um sistema de IA e implicações éticas

Elemento	Pergunta regulatória essencial	Implicação ética
Entrada	Quais dados ou informações alimentam o sistema?	Determina qualidade, representatividade, privacidade, consentimento e risco de reidentificação.
Processamento/modelo	Como o sistema transforma as entradas em saídas relevantes para a pesquisa?	Determina opacidade, possibilidade de auditoria, generalização e risco de erro.
Saída	Que classificação, predição, recomendação ou conteúdo é produzido?	Determina possíveis danos, inferências, decisões e necessidade de supervisão.

Fonte: elaboração própria.

1.10 TREINAMENTO, VALIDAÇÃO E TESTE

Quando a pesquisa desenvolve ou avalia modelo de IA, recomenda-se que o protocolo distinga, conforme aplicável, os dados utilizados para treinamento, validação e teste. A

apresentação de desempenho apenas sobre dados utilizados para treinar o sistema não demonstra capacidade de generalização.

Recomenda-se que o protocolo apresente:

- origem e características dos conjuntos de dados;
- critérios de seleção e exclusão;
- procedimentos relevantes de pré-processamento;
- se houve uso de dados sintéticos;
- estratégia de separação entre treino, validação e teste, quando aplicável;
- métricas de desempenho pertinentes ao contexto de uso.

Recomenda-se que o CEP verifique:

- se o desempenho apresentado foi obtido em dados independentes quando isso for necessário para sustentar a finalidade proposta;
- se os dados utilizados são adequados à população e ao contexto da pesquisa;
- se limitações de generalização são explicitadas.

1.11 FAMÍLIAS DE SISTEMAS E RISCOS ASSOCIADOS

O Quadro 3 apresenta as famílias de sistemas de IA mais frequentes em protocolos de pesquisa, com exemplos e as questões éticas prioritárias associadas a cada tipo de uso.

Quadro 3 – Famílias de sistemas de IA e questões éticas prioritárias

Tipo de uso	Exemplos no documento de origem	Questões prioritárias
Classificação e predição	modelos de risco, apoio diagnóstico, classificação de imagens	representatividade, validação, impacto da decisão, supervisão e erro.
Processamento de linguagem	análise de prontuários, textos, documentos e discursos	privacidade, inferências, contextualização, vieses e confidencialidade.
IA generativa	produção de texto, síntese, geração de dados sintéticos	alucinação, integridade científica, rastreabilidade, confidencialidade e verificação humana.
Sistemas adaptativos	modelos sujeitos a atualização ou retreinamento	controle de versão, deriva, revalidação e necessidade de nova avaliação ética.

Tipo de uso	Exemplos no documento de origem	Questões prioritárias
Aprendizagem federada e técnicas de privacidade	processamento distribuído de dados	governança, risco residual de reidentificação e responsabilidade entre instituições.

Fonte: elaboração própria.

1.12 DADOS SINTÉTICOS

Quando dados sintéticos forem utilizados, recomenda-se que o protocolo os identifique expressamente e descreva sua função no estudo. Dados sintéticos não devem ser apresentados como dados observacionais reais. Quando integrarem o desenvolvimento ou a validação de um sistema, recomenda-se que o pesquisador demonstre sua adequação à finalidade científica.

Sinal de alerta:

- dados sintéticos descritos apenas como mecanismo de 'anonimização' sem avaliação de seus limites;
- resultados gerados artificialmente apresentados como evidência empírica coletada de participantes;
- ausência de verificação humana de conteúdo produzido por IA generativa.

DIMENSÕES CRÍTICAS DE RISCO

Recomenda-se que a avaliação de sistemas de IA considere, de forma articulada, viés, explicabilidade e transparência, supervisão humana, impacto sobre participantes e grupos, e ciclo de vida do sistema. Essas dimensões constituem fundamentos centrais da MARIAH.

1.13 VIÉS, REPRESENTATIVIDADE E EQUIDADE

Recomenda-se que o protocolo avalie se os dados e o desempenho do sistema são adequados às pessoas, aos grupos e aos contextos afetados pela pesquisa. Métricas agregadas podem ocultar desempenho desigual entre subgrupos. Recomenda-se que a equidade seja considerada também como dimensão de segurança: diferenças relevantes de desempenho podem produzir riscos concretos, especialmente quando a IA influencia decisões clínicas, assistenciais, de seleção, classificação ou priorização.

Recomenda-se que o protocolo apresente:

- composição e origem da base de dados;
- características relevantes da população representada;
- avaliação de desempenho por subgrupos quando pertinente;
- limitações de representatividade conhecidas;
- medidas previstas para reduzir ou monitorar disparidades.

Recomenda-se que o CEP verifique:

- se a população do estudo é comparável àquela utilizada no desenvolvimento ou validação;
- se subgrupos potencialmente vulneráveis ou sub-representados foram considerados;
- se o protocolo distingue ausência de evidência de viés de evidência de ausência de viés.

Sinal de alerta:

- apresentação apenas de acurácia global;
- modelo importado sem justificativa de transferibilidade ao contexto brasileiro;
- ausência de análise de subgrupos em pesquisa cujo resultado possa distribuir benefícios ou riscos de forma desigual.

1.14 EXPLICABILIDADE E TRANSPARÊNCIA

A profundidade da explicação necessária depende da função da IA e das consequências de sua saída. A avaliação ética não exige acesso irrestrito ao código-fonte em todos os casos, mas requer informação suficiente para compreender o papel do sistema, suas limitações e as condições nas quais suas saídas podem ser contestadas.

Recomenda-se que o protocolo apresente:

- nível de explicabilidade disponível e adequado ao contexto de uso;
- limitações conhecidas das técnicas de explicação utilizadas;
- informações necessárias para que o responsável humano interprete e conteste a saída;
- forma de comunicação aos participantes quando a IA tiver impacto relevante sobre sua participação ou seus dados.

Fundamento:

O documento de origem ressalta que mapas de saliência e outras técnicas de explicação não demonstram causalidade e não equivalem, por si só, à interpretabilidade clínica ou ética. Recomenda-se que a explicação seja adequada ao usuário e ao contexto.

1.15 SUPERVISÃO HUMANA EFETIVA

A presença formal de pessoa responsável não é suficiente quando ela não dispõe de informação, competência, tempo, autoridade ou possibilidade real de discordar do sistema. Recomenda-se que a supervisão humana seja efetiva e compatível com a magnitude das consequências. Recomenda-se que o protocolo explicite quem exerce a supervisão, em que momento ela ocorre, quais informações sustentam a revisão humana e como se procede diante de discordância entre a saída algorítmica e o julgamento humano, considerando inclusive o risco de viés de automação.

Recomenda-se que o protocolo apresente:

- quem exerce a supervisão;
- qual qualificação é necessária;
- em que momento ocorre a revisão;
- como a discordância é registrada e tratada;

– o que ocorre quando o sistema falha ou produz resultado incompatível com outras evidências.

Recomenda-se que o CEP verifique:

– se há possibilidade real de intervenção antes de consequências relevantes;
– se o responsável humano compreende suficientemente as limitações do sistema;
– se o desenho operacional favorece a contestação em vez da aceitação automática.

Sinal de alerta:

– supervisão descrita apenas como 'médico responsável' ou 'pesquisador responsável', sem processo de revisão;
– decisão de alta consequência sem possibilidade real de contestação;
– uso de IA em contexto de sobrecarga ou automação que favoreça viés de automação.

1.16 CICLO DE VIDA, ATUALIZAÇÃO E DERIVA

A aceitabilidade ética de um sistema de IA não deve ser tratada como propriedade estática. O desempenho pode variar conforme população, dados, fluxo de trabalho, ambiente institucional e condições reais de uso, e pode se modificar ao longo do tempo. Quando o sistema puder mudar durante a pesquisa, recomenda-se que o protocolo estabeleça identificação e controle de versão, critérios de monitoramento, mecanismos para detectar falhas ou deterioração de desempenho e procedimentos para atualização, recalibração, retreinamento ou substituição. Recomenda-se que mudanças capazes de alterar riscos, benefícios, equidade, privacidade, autonomia, supervisão humana ou consequências de erro sejam avaliadas quanto à necessidade de comunicação ou emenda ao CEP.

Recomenda-se que o protocolo apresente:

– versão do sistema utilizada;
– se o sistema é estático ou adaptativo;
– eventos que podem alterar seu desempenho;
– critérios mínimos de desempenho;
– procedimentos diante de deriva de dados ou desempenho;
– destino do sistema ao encerramento da pesquisa.

1.16.1 Monitoramento, mudanças e rastreabilidade

Recomenda-se que a avaliação ética acompanhe o ciclo de vida da IA no protocolo. Quando proporcional ao risco e à finalidade da pesquisa, recomenda-se que o protocolo preveja mecanismos que permitam identificar qual sistema, versão ou configuração foi utilizado; monitorar desempenho durante a execução; reconhecer resultados inesperados ou deterioração relevante; e estabelecer respostas previamente definidas quando novas informações modificarem o perfil de risco.

Recomenda-se que o CEP verifique se alterações técnicas ou contextuais aparentemente limitadas podem modificar de forma eticamente relevante a interação com participantes, a autonomia decisória, o desempenho em grupos específicos, a supervisão humana, a relação risco-benefício ou as salvaguardas previstas.

A existência de registro, autorização, certificação ou outro enquadramento regulatório do sistema não substitui a avaliação ética de sua utilização no protocolo específico.

1.17 PESSOAS, GRUPOS E COMUNIDADES AFETADOS ALÉM DOS PARTICIPANTES FORMAIS

Recomenda-se que a avaliação ética considere não apenas as pessoas cujos dados são diretamente utilizados ou que interagem com a equipe de pesquisa, mas também aquelas que possam ser classificadas, perfiladas, representadas ou afetadas pelas inferências, decisões, produtos ou usos posteriores do sistema. A ausência de enquadramento formal como participante não torna eticamente irrelevantes os riscos previsíveis de discriminação, estigmatização, exclusão, vigilância, manipulação, representação inadequada ou distribuição injusta de benefícios e ônus.

Recomenda-se que o protocolo apresente:

- quem pode ser afetado direta ou indiretamente pelas saídas, inferências ou reutilizações do sistema;
- quais danos individuais, coletivos, sociais, culturais, econômicos ou territoriais são razoavelmente previsíveis;
- como grupos potencialmente afetados foram considerados na concepção, validação e comunicação da pesquisa;
- quais medidas reduzem riscos de estigmatização, discriminação, perfilamento indevido e representação inadequada.

O CEP verifica:

- se a análise de risco ultrapassa a proteção individual e alcança impactos coletivos e societais pertinentes;
- se dados públicos, anonimizados ou secundários foram indevidamente tratados como eticamente neutros;
- se a pesquisa pode deslocar riscos para grupos que não consentiram nem possuem canal direto de contestação;
- se o valor social alegado é compatível com a distribuição de riscos, benefícios e poder decisório.

Sinal de alerta:

- o protocolo afirma não haver participantes apenas porque utiliza dados públicos, secundários ou anonimizados;
- inferências sensíveis ou classificações coletivas são produzidas sem identificação dos grupos afetados;
- impactos sociais ou culturais previsíveis são tratados exclusivamente como problema de proteção de dados.

CONTEXTO BRASILEIRO E INTERESSE PÚBLICO

Recomenda-se que a avaliação ética de pesquisas com IA considere o contexto brasileiro, sua diversidade populacional, territorial, social e epidemiológica e as desigualdades que podem afetar a produção e a interpretação dos dados.

1.18 ADEQUAÇÃO AO CONTEXTO BRASILEIRO

Não se deve presumir que sistemas desenvolvidos ou validados em outros contextos sejam adequados à população brasileira. Recomenda-se que o protocolo justifique a transferibilidade do modelo ou apresente validação compatível com a finalidade proposta.

Recomenda-se que o CEP verifique:

- se a população e o ambiente de desenvolvimento são comparáveis ao contexto da pesquisa;
- se diferenças de infraestrutura, idioma, cultura, epidemiologia ou composição populacional podem alterar o desempenho;
- se existe risco de reforço de desigualdades já presentes nos dados.

1.19 SUS E DADOS DE INTERESSE PÚBLICO

Nas pesquisas que utilizem dados, serviços, infraestruturas ou populações vinculadas ao Sistema Único de Saúde (SUS), recomenda-se que a utilização de IA observe a proteção dos participantes, a governança dos dados, a finalidade da pesquisa, a equidade e o interesse público.

A abrangência e a diversidade dos dados produzidos no SUS ampliam o potencial científico da pesquisa e, simultaneamente, aumentam a responsabilidade sobre uso, compartilhamento, reutilização e destino dos produtos derivados desses dados.

Sinal de alerta:

- transferência de dados ou modelos derivados de dados públicos sem definição de finalidade, responsabilidade e destino;
- uso de infraestrutura estrangeira ou de terceiros sem descrição da cadeia de processamento;

– ausência de análise de impacto sobre grupos ou territórios sub-representados.

1.20 SAÚDE DIGITAL, INTEROPERABILIDADE E SOBERANIA DE DADOS

Quando a pesquisa estiver inserida em ecossistemas de saúde digital, plataformas interoperáveis ou redes nacionais de dados, recomenda-se que o protocolo explicite os fluxos de dados e as responsabilidades institucionais relevantes. A interoperabilidade não elimina as obrigações de finalidade, segurança, rastreabilidade e proteção.

REFERÊNCIAS INTERNACIONAIS E PRINCÍPIOS INCORPORADOS

A análise internacional realizada no documento de origem contribui para a construção dos critérios utilizados neste Guia e na MARIAH. Nesta versão regulatória, a descrição país a país é substituída por uma síntese das contribuições incorporadas ao modelo brasileiro.

O Quadro 4 sintetiza as contribuições das experiências internacionais incorporadas a este Guia e à MARIAH.

Quadro 4 – Experiências internacionais e contribuições incorporadas ao modelo brasileiro

Referência/experiência	Contribuição incorporada	Aplicação no Guia/MARIAH
OMS	governança ética, supervisão humana, proteção de participantes e atenção às limitações dos sistemas	princípios gerais; supervisão; transparência; responsabilidade.
Canadá	avaliação de impacto algorítmico e separação entre risco e mitigação	estrutura proporcional; rastreabilidade; regra crítica
Austrália/NHMRC	orientação para avaliação de pesquisa envolvendo IA	adequação da informação apresentada ao comitê e avaliação proporcional
Estados Unidos/FDA	contexto de uso, influência do modelo e ciclo de vida	caracterização do uso; monitoramento; influência da IA na decisão
União Europeia e proteção de dados	governança, transparência e proteção de dados em sistemas de IA	dados, finalidade, riscos e documentação
Ásia — referências analisadas no original	contestabilidade, desempenho, proteção de participantes e avaliação de dispositivos	supervisão, validação, direitos e requisitos por contexto

Fonte: elaboração própria.

Essas referências não substituem o marco brasileiro. Elas funcionam como fontes de comparação e apoio metodológico. Os requisitos aplicáveis ao protocolo devem ser fundamentados nas normas brasileiras vigentes e nas responsabilidades do sistema de ética em pesquisa.

Recomenda-se que o CEP verifique:

- se referência internacional citada pelo pesquisador é compatível com o contexto brasileiro;
- se certificação, aprovação ou desempenho em outro país está sendo indevidamente utilizado como substituto da avaliação ética local;
- se salvaguardas adotadas no exterior foram adaptadas à população e às normas brasileiras.

INTEGRIDADE CIENTÍFICA, AUTORIA E USO DE IA NA PRODUÇÃO DA PESQUISA

O uso de IA na elaboração, análise ou documentação de uma pesquisa não desloca a responsabilidade humana. O pesquisador responsável permanece responsável pela integridade do protocolo, dos dados, das análises, das referências e dos resultados apresentados.

1.21 DECLARAÇÃO DO USO DE IA

Recomenda-se que o protocolo declare o uso de IA quando ela contribuir de modo relevante para a produção de conteúdo, análise, geração de dados, formulação de hipóteses, elaboração de instrumentos ou outras etapas que possam afetar a integridade científica ou a proteção dos participantes.

Recomenda-se que o protocolo apresente:

– ferramenta ou sistema utilizado, quando relevante;
– finalidade do uso;
– etapa da pesquisa em que foi empregado;
– responsável humano pela revisão e validação;
– medidas de verificação de referências, fatos, cálculos, dados ou conteúdo gerado.

1.22 AUTORIA E RESPONSABILIDADE HUMANA

Sistemas de IA não devem substituir a atribuição de responsabilidade humana pelo conteúdo científico. A responsabilidade ética e científica não pode ser transferida ao sistema utilizado.

Recomenda-se que o CEP verifique:

– se há pessoa responsável por cada etapa relevante do protocolo;
– se a equipe dispõe de competência para verificar as saídas da IA;
– se o uso de IA não obscurece a cadeia de responsabilidade.

1.23 ALUCINAÇÃO, FABRICAÇÃO E NEGLIGÊNCIA

Recomenda-se que o conteúdo gerado por IA seja verificado antes de integrar o protocolo ou os materiais submetidos ao CEP. Referências inexistentes, afirmações normativas incorretas, dados sintéticos apresentados como reais e hipóteses não verificadas podem comprometer a validade científica e, consequentemente, a justificativa ética da exposição de participantes a riscos.

Sinal de alerta:

– referências bibliográficas não verificáveis;
– afirmações normativas sem fonte oficial;
– texto de consentimento produzido por IA sem revisão substantiva;
– dados ou imagens gerados artificialmente sem identificação clara;
– delegação de julgamento científico ou ético à ferramenta.

1.24 O QUE SE RECOMENDA PARA O PROTOCOLO — E O QUE O CEP DEVE VERIFICAR

O Quadro 5 sintetiza o que se recomenda que o protocolo apresente e o que o CEP verifica quanto à integridade científica, à autoria e ao uso de IA na produção da pesquisa.

Quadro 5 – Síntese das recomendações sobre integridade científica, autoria e uso de IA

Tema	Recomenda-se que o protocolo apresente	Recomenda-se que o CEP verifique
Uso de IA	declaração da ferramenta, função e etapa	se a declaração é completa e compatível com o uso real.
Verificação humana	responsável e processo de revisão	se a revisão é substantiva e suficiente.
Dados sintéticos	identificação, finalidade e validação	se não foram confundidos com dados reais.
Referências e normas	mecanismo de conferência	se as fontes relevantes são verificáveis.
Responsabilidade	pessoa responsável por decisões e conteúdo	se a cadeia de responsabilidade permanece humana e definida.

Fonte: elaboração própria.

GOVERNANÇA DE DADOS E PROTEÇÃO DE PARTICIPANTES

A governança de dados é uma dimensão central da avaliação ética de pesquisas com IA. Recomenda-se que o protocolo permita identificar a origem dos dados, as finalidades de tratamento, os agentes responsáveis, os fluxos de compartilhamento, os mecanismos de segurança e o destino dos dados e dos modelos derivados.

1.25 ORIGEM, FINALIDADE E PAPÉIS

Recomenda-se que o protocolo apresente:

- origem de cada conjunto de dados;
- finalidade de uso no estudo e no treinamento ou ajuste do sistema;
- Controlador e Operador, quando aplicável;
- instituições e terceiros que terão acesso;
- infraestrutura utilizada para armazenamento ou processamento;
- destino dos dados, parâmetros e modelo após a pesquisa.

1.26 BANCOS DE DADOS, USO FUTURO E CONSENTIMENTO

Quando a pesquisa utilizar banco de dados, recomenda-se que o protocolo demonstre o enquadramento aplicável segundo a origem do banco, a finalidade da utilização e as normas brasileiras pertinentes. Eventual dispensa de obtenção ou documentação de novo consentimento deve indicar a hipótese normativa aplicável e apresentar justificativa ética e metodológica, sem ser presumida apenas em razão da origem do banco.

Recomenda-se que o CEP verifique:

- se os instrumentos institucionais aplicáveis foram apresentados;
- se o uso futuro está coberto pelo consentimento anterior ou pela hipótese normativa invocada;
- se a anonimização ou pseudonimização descrita é adequada ao risco de reidentificação;
- se a finalidade da pesquisa é compatível com o uso proposto.

1.27 REIDENTIFICAÇÃO, INFERÊNCIA E SEGURANÇA

A ausência de identificadores diretos não elimina automaticamente o risco de reidentificação. Sistemas de IA podem combinar variáveis e produzir inferências capazes de revelar informações sensíveis. Recomenda-se que o protocolo considere o risco técnico e contextual de reidentificação e inferência.

Sinal de alerta:

- uso do termo 'anonimizado' sem descrição do procedimento;
- grandes conjuntos de dados com múltiplas variáveis indiretas capazes de reidentificação;
- transferência a terceiros sem definição de finalidade e responsabilidade;
- uso de serviço externo de IA que retenha ou reutilize dados do protocolo.

1.28 TRANSFERÊNCIA INTERNACIONAL E SOBERANIA DE DADOS

Quando houver transferência internacional ou processamento em infraestrutura localizada fora do Brasil, recomenda-se que o protocolo descreva o fluxo, a finalidade, os agentes envolvidos, as salvaguardas e o destino dos dados e do modelo resultante.

1.29 TECNOLOGIAS DE PRESERVAÇÃO DE PRIVACIDADE

Aprendizagem federada, privacidade diferencial, criptografia homomórfica e outras arquiteturas de preservação de privacidade podem reduzir determinados riscos, mas não devem ser tratadas como garantias absolutas. Recomenda-se que o protocolo descreva o mecanismo adotado e o risco residual relevante.

Fundamento:

O documento de origem destaca que técnicas de privacidade modificam a arquitetura do tratamento, mas não eliminam automaticamente riscos de inversão, reidentificação ou responsabilidade institucional. O CEP avalia a suficiência da salvaguarda no contexto concreto.

1.30 DIREITOS DOS PARTICIPANTES E DESTINO DO MODELO

Recomenda-se que o protocolo esclareça como serão tratados os direitos dos participantes durante a pesquisa e, quando pertinente, após o treinamento do modelo. A retirada de dados pode ter limitações técnicas quando o sistema já foi treinado; recomenda-se que essas limitações sejam explicadas e avaliadas de forma transparente.

Recomenda-se que o protocolo apresente:

- como o participante poderá exercer direitos sobre seus dados;
- como será tratada eventual retirada do consentimento;
- se o modelo continuará a incorporar padrões aprendidos a partir dos dados retirados;
- se há mecanismo de eliminação, reprocessamento ou outra medida aplicável;
- qual será o destino final do modelo e de seus parâmetros.

ESPECIFICIDADES POR NATUREZA DA PESQUISA

As especificidades não constituem exceções à avaliação geral. Elas determinam quais dimensões de risco exigem maior atenção. Pesquisas das Ciências da Saúde e pesquisas das Ciências Humanas e Sociais podem compartilhar princípios e, ao mesmo tempo, demandar perguntas diferentes.

1.31 CIÊNCIAS DA SAÚDE: IMPACTO ASSISTENCIAL E SAMD

Quando a IA puder influenciar diagnóstico, prognóstico, tratamento, monitoramento, triagem ou outra decisão de natureza assistencial, recomenda-se que o protocolo descreva o impacto potencial da saída do sistema e o papel da supervisão humana.

Quando o sistema se enquadrar como Software as a Medical Device (SaMD), a apreciação ética pelo CEP e a avaliação sanitária pela Anvisa têm mandatos complementares, e não concorrentes. A apreciação pelo CEP não substitui a competência da autoridade sanitária, e a existência de enquadramento regulatório não substitui a avaliação ética da pesquisa.

Em sistemas de IA com impacto assistencial, recomenda-se que o CEP considere ainda se o desempenho foi demonstrado em população e contexto pertinentes; se há possibilidade de desempenho diferencial entre grupos; se a supervisão humana é efetiva; e se o protocolo prevê monitoramento e gestão de mudanças quando o sistema puder ser atualizado durante o estudo.

Recomenda-se que o protocolo apresente:

- finalidade do sistema e sua função na pesquisa;
- fase de desenvolvimento ou avaliação;
- impacto clínico ou assistencial potencial;
- desempenho relevante no contexto de uso;
- mecanismo de supervisão e contestação;
- enquadramento regulatório aplicável, quando pertinente;
- plano para falha, resultado inesperado ou necessidade de intervenção imediata.

Recomenda-se que o CEP verifique:

- se a decisão assistencial permanece adequadamente supervisionada;
- se o protocolo distingue pesquisa, uso assistencial e finalidade regulatória;
- se os riscos decorrentes de erro ou mau funcionamento foram avaliados;
- se requisitos de governança de dados continuam atendidos.

Sinal de alerta:

- sistema descrito como 'experimental' sendo utilizado de forma assistencial sem delimitação;
- alta autonomia em decisão de consequência grave sem revisão humana efetiva;
- uso de desempenho obtido em população distinta sem validação contextual.

1.32 CIÊNCIAS HUMANAS E SOCIAIS: INFERÊNCIAS, GRUPOS E RISCO COLETIVO

Nas Ciências Humanas e Sociais, recomenda-se que a avaliação não se limite à natureza dos dados de origem. O processamento algorítmico pode produzir inferências sobre identidade, comportamento, relações, crenças, posição social, saúde, pertencimento ou outros atributos que não estavam explicitamente presentes nos dados coletados.

Recomenda-se que o protocolo apresente:

- origem dos dados e condições de acesso;
- o que o sistema pretende inferir, classificar ou agrupar;
- se as inferências envolvem atributos sensíveis;
- base ética e jurídica para o tratamento dessas inferências;
- grupos potencialmente afetados, inclusive além dos participantes formais;
- risco de estigmatização, discriminação ou representação inadequada;
- destino do modelo e possibilidade de usos secundários.

Recomenda-se que o CEP verifique:

- se consentimento ou justificativa de dispensa cobre as inferências produzidas, e não apenas os dados de origem;
- se dados publicamente acessíveis estão sendo tratados como eticamente neutros sem análise do contexto;
- se há riscos coletivos que não são capturados pela proteção individual;
- se o modelo poderá ser reutilizado em contexto discriminatório ou não previsto.

Sinal de alerta:

- dados públicos processados para produzir inferências sensíveis sem análise específica;
- classificação de risco mínimo baseada apenas no fato de os dados serem públicos ou anonimizados;
- ausência de identificação dos grupos sociais afetados pelas inferências;
- publicação aberta do modelo sem avaliação de usos secundários potencialmente danosos.

Nas Ciências Humanas e Sociais, recomenda-se examinar de modo específico a adequação contextual e linguística dos dados e modelos; as relações de poder entre pesquisadores, plataformas, instituições e comunidades; a transformação de discursos, imagens, vínculos e comportamentos em categorias automatizadas; a possibilidade de vigilância, manipulação ou apagamento de significados; e a eventual apropriação de dados, conhecimentos ou expressões culturais. Recomenda-se que a análise não reduza essas questões à legalidade do acesso aos dados.

O CEP verifica adicionalmente nas Ciências Humanas e Sociais:

- se as categorias produzidas pelo sistema preservam contexto, linguagem, historicidade e significado social;
- se comunidades ou grupos vulnerabilizados participam de forma pertinente da identificação de riscos e benefícios;
- se o uso de bases digitais, redes sociais ou plataformas respeita expectativas razoáveis de privacidade e integridade contextual;
- se há risco de colonialismo de dados, extração assimétrica de valor, reforço de desigualdades ou reutilização incompatível com os interesses dos grupos de origem.

1.33 PESQUISAS INTERDISCIPLINARES

Quando o protocolo reunir características das Ciências da Saúde e das Ciências Humanas e Sociais, recomenda-se que o CEP aplique cumulativamente as perguntas pertinentes aos diferentes contextos. A classificação disciplinar não deve reduzir a atenção a um risco efetivamente presente.

INTEGRAÇÃO COM A MARIAH E OS INSTRUMENTOS DE APLICAÇÃO

Este Guia estabelece os fundamentos, os critérios e as informações eticamente relevantes. A MARIAH e seus instrumentos de aplicação estruturam a verificação, a classificação e a documentação dos riscos.

O Quadro 6 apresenta a função principal de cada módulo do conjunto documental e a pergunta que cada um responde.

Quadro 6 – Função dos módulos do conjunto documental

Módulo	Função principal	Pergunta que responde
Guia	estabelece escopo, princípios, conceitos e critérios de avaliação	O que se recomenda considerar e por quê?
Caderno Referencial	reúne os fundamentos éticos, conceituais, científicos e regulatórios que sustentam as orientações	Por que cada recomendação, eixo de risco ou salvaguarda foi adotada?
MARIAH — instrumentos de verificação	converte os critérios em itens verificáveis e materiais de apoio	O que se recomenda que o protocolo apresente e o que o CEP verifica?
MARIAH — matriz de classificação de risco	classifica e documenta o risco de forma proporcional	Qual é o perfil e o nível de risco e quais requisitos decorrem dele?

Fonte: elaboração própria.

1.34 APLICAÇÃO DA MARIAH

Recomenda-se que a MARIAH seja aplicada nos termos de seu módulo próprio. Este Guia não reproduz a matriz, suas pontuações ou suas regras de cálculo, a fim de evitar duplicidade e divergência entre versões.

Sempre que uma questão tratada neste Guia corresponder a dimensão avaliada pela MARIAH, recomenda-se que a decisão seja registrada no instrumento de risco e, quando necessário, fundamentada no parecer do CEP.

1.35 REGISTRO DA DECISÃO ÉTICA

Recomenda-se que o parecer seja suficientemente claro para demonstrar quais riscos foram identificados, quais informações sustentaram a análise, quais requisitos foram atendidos, quais pendências permanecem e quais razões fundamentaram a decisão.

Recomenda-se que o CEP registre:

- o contexto de uso da IA;
- o nível de risco obtido na MARIAH, quando aplicável;
- as dimensões de risco de maior relevância;
- os requisitos ou salvaguardas necessárias;
- eventuais divergências técnicas ou éticas relevantes;
- a fundamentação de diligência, condicionamento ou recusa.

1.36 SUPERVISÃO ÉTICA LONGITUDINAL E EVENTOS QUE EXIGEM NOVA APRECIAÇÃO

A aprovação ética inicial não encerra a supervisão quando o sistema, os dados, a finalidade ou o contexto de uso puderem mudar de modo relevante. Recomenda-se que o protocolo defina os eventos que exigem emenda, comunicação ou nova apreciação pelo CEP, de forma proporcional ao risco e sem confundir atualização meramente técnica com modificação capaz de alterar direitos, riscos, benefícios ou a validade científica.

Recomenda-se submeter à apreciação do CEP, conforme o risco:

- troca do modelo, do fornecedor ou de componente que altere função, desempenho, explicabilidade ou governança;
- inclusão de novas bases de dados, populações, variáveis sensíveis ou fontes digitais; alteração de finalidade, contexto de uso, autonomia, usuário da saída ou influência sobre decisões;
- retreinamento, ajuste substancial, deriva relevante ou queda de desempenho, inclusive por subgrupos;
- novo uso, compartilhamento, abertura ou transferência do modelo, dos parâmetros ou dos produtos derivados;
- identificação de dano, discriminação, reidentificação, incidente de segurança ou uso não previsto.

1.37 RESPONSABILIDADES COMPARTILHADAS E LIMITES DA COMPETÊNCIA DO CEP

A proteção ética em pesquisas relacionadas à IA constitui responsabilidade distribuída ao longo do ciclo de vida. Recomenda-se que sejam identificadas as atribuições do

pesquisador responsável, da instituição proponente, dos responsáveis pela governança e pelo acesso aos dados, dos desenvolvedores ou fornecedores, dos patrocinadores, dos financiadores e de outras instâncias competentes. Essa distribuição não reduz a competência do CEP nem transfere a ele funções técnicas, sanitárias, administrativas ou de proteção de dados pertencentes a outras autoridades.

Recomenda-se que o protocolo apresente:

- **matriz de responsabilidades pelas decisões, dados, validação, segurança, monitoramento, resposta a incidentes e encerramento;**
- **dependências de fornecedores e limites de acesso a informações técnicas ou contratuais;**
- **instâncias complementares de supervisão e os mecanismos de comunicação entre elas;**
- **responsável por assegurar o cumprimento das condições aprovadas pelo CEP durante todo o estudo.**

1.38 USO DE INTELIGÊNCIA ARTIFICIAL NO APOIO ÀS ATIVIDADES DO CEP

Ferramentas de IA podem ser utilizadas em atividades auxiliares de organização, capacitação ou pré-análise, desde que sua utilização seja institucionalmente autorizada, transparente, segura e submetida à supervisão humana. A IA não pode deliberar, votar, emitir decisão autônoma, substituir a apreciação colegiada ou assumir responsabilidade pelo parecer. Recomenda-se que o conteúdo do protocolo não seja inserido em serviço externo que retenha, reutilize ou utilize os dados para treinamento sem garantia institucional adequada de confidencialidade, segurança e finalidade.

Salvaguardas para o uso de IA pelo CEP:

- autorização institucional e definição prévia das finalidades permitidas;
- proteção da confidencialidade dos protocolos, pareceres e dados pessoais ou comerciais;
- registro da ferramenta, versão, finalidade, usuário e etapa em que houve apoio automatizado;
- verificação humana substantiva de fatos, referências, normas, inferências e sugestões produzidas;
- avaliação de viés de automação, adequação linguística e contextual e representatividade das fontes;

- proibição de decisão autônoma e preservação da deliberação humana, colegiada e fundamentada;
- monitoramento periódico de desempenho, incidentes, erros e efeitos sobre a qualidade da revisão.

Sinal de alerta:

- uso não declarado de ferramenta pública ou comercial para analisar protocolo confidencial;
- aceitação de recomendação automatizada sem conferência humana documentada;
- dependência de modelo treinado em normas, idiomas ou contextos incompatíveis com o SNEPSH;
- substituição da capacitação dos membros ou da discussão colegiada por respostas geradas automaticamente.

ANEXO A — SÍNTESE DOS REQUISITOS DESTE GUIA

Dimensão	Informação mínima esperada
Contexto de uso	finalidade, função, autonomia, influência sobre decisão e população afetada
Dados	origem, qualidade, representatividade, finalidade, agentes e fluxo
Validação	treino/validação/teste, desempenho e adequação contextual
Viés e equidade	desempenho por subgrupos, limitações e medidas de controle
Explicabilidade	informação suficiente para compreensão e contestação
Supervisão humana	responsável, competência, tempo, intervenção e registro da discordância
Ciclo de vida	versão, atualização, deriva, retreinamento e destino
Integridade científica	declaração de uso de IA, verificação humana e responsabilidade
Governança de dados	consentimento, bancos de dados, transferências, segurança e direitos
Ciências da Saúde	impacto assistencial, SaMD, falha e supervisão
Ciências Humanas e Sociais	inferências sensíveis, grupos afetados, risco coletivo e usos secundários

Fonte: elaboração própria.

ANEXO B — SINAIS DE ALERTA CONSOLIDADOS

- uso de IA não declarado ou descrito apenas genericamente;
- ausência de contexto de uso definido;
- desempenho apresentado apenas em dados de treinamento;
- ausência de análise de representatividade ou subgrupos quando pertinente;
- supervisão humana apenas nominal;
- alta consequência sem possibilidade real de contestação;
- dados sintéticos apresentados como dados reais;
- referências ou afirmações normativas não verificadas;
- anonimização declarada sem método descrito;
- transferência a terceiros ou para o exterior sem fluxo e responsabilidade definidos;
- modelo importado aplicado ao contexto brasileiro sem análise de transferibilidade;
- Ciências Humanas e Sociais com inferências sensíveis tratadas como risco mínimo apenas porque os dados de origem são públicos;
- SaMD ou sistema assistencial com impacto relevante sem delimitação de competência regulatória e de supervisão;
- ausência de identificação de pessoas, grupos ou comunidades afetados, além dos participantes formais;
- mudança relevante de modelo, base, finalidade ou contexto sem avaliação da necessidade de emenda;
- responsabilidades fragmentadas ou atribuídas genericamente ao CEP;
- uso de IA pelo CEP sem autorização, registro, proteção da confidencialidade e verificação humana;
- redução de riscos coletivos, sociais ou culturais a questões exclusivamente técnicas ou de proteção de dados.

ANEXO C — FUNDAMENTOS NORMATIVOS E TÉCNICOS A SEREM MANTIDOS NA EDIÇÃO FINAL

Recomenda-se que a edição final preserve a rastreabilidade das orientações aos marcos e referências já utilizados no documento de origem. Entre os principais conjuntos de fontes citados no texto original encontram-se:

- Lei nº 14.874/2024 e Decreto nº 12.651/2025;
- Lei Geral de Proteção de Dados Pessoais — LGPD;
- Resoluções CNS nº 466/2012, nº 510/2016 e nº 738/2024;
- normas e orientações setoriais aplicáveis a SaMD e uso de IA na medicina, quando pertinentes ao protocolo;
- Política de Integridade na Atividade Científica do CNPq, conforme citada no documento de origem;
- referências da Organização Mundial da Saúde e demais experiências internacionais utilizadas na fundamentação da MARIAH;
- referências técnicas sobre explicabilidade, viés, aprendizagem federada, privacidade e avaliação clínica citadas no documento original.

LEITURAS RECOMENDADAS

ETO, T. *et al.* Streamlining IRB review of AI human subjects research (AIHSR): the three-stage framework. **Frontiers in Systems Biology**, v. 6, 1804193, Mar. 2026. DOI: https://doi.org/10.3389/fsysb.2026.1804193.

KERASIDOU, A. *et al.* Ethics Review of AI research: an approach to reviewing and revising existing governance structures. **Research Ethics**, v. 22, n.2, p. 357-370, 2026. DOI: https://doi.org/10.1177/17470161251406299.

MALPANI, R. *et al.* Ethics oversight of health-related research should be contemporaneous in the age of artificial intelligence. **The Lancet Digital Health**, v. 8, n. 8, 101085, Aug. 2026. DOI: https://doi.org/10.1016/j.landig.2026.101085.

MOODLEY, K.; MALPANI, R.; REIS, A. A. How will 'Chat-IRB' impact research ethics review in LMICs? **Journal of Medical Ethics**, jme-2025-111452, Nov. 2025. DOI: https://doi.org/10.1136/jme-2025-111452.

UNITED KINGDOM GOVERNMENT. Department of Health & Social Care. National Commission into the Regulation of AI in Healthcare. **National Commission into the Regulation of AI in Healthcare**: Recommendations for a future regulatory framework. [London]: ©Crown, 10 Sep. 2026. Disponível em: https://www.gov.uk/government/publications/national-commission-into-the-regulation-of-ai-in-healthcare-recommendations-for-a-future-regulatory-framework/national-commission-into-the-regulation-of-ai-in-healthcare-recommendations-for-a-future-regulatory-framework. Acesso em: 10 set. 2026.

WORLD HEALTH ORGANIZATION. **Artificial intelligence-related health research**: ethics review and oversight. Geneva: WHO, 2026. Disponível em: https://iris.who.int/items/21f44816-7ae8-4600-93ef-b4cc9a7c97f2. Acesso em: 28 ago. 2026.