MINISTÉRIO DA SAÚDE
INSTÂNCIA NACIONAL DE ÉTICA EM PESQUISA — INAEP

MARIAH

MATRIZ DE AVALIAÇÃO DE RISCO EM INTELIGÊNCIA ARTIFICIAL EM PESQUISA COM SERES HUMANOS

Brasília – DF
2026

SUMÁRIO

1 APRESENTAÇÃO DO INSTRUMENTO 3
1.1 Linguagem regulatória e operacional 3
2 PARTE I — FUNDAMENTOS E CONSTRUÇÃO DA MARIAH 4
2.1 MARIAH: MATRIZ DE AVALIAÇÃO DE RISCO EM INTELIGÊNCIA ARTIFICIAL (Versões A e B) 4
3 PARTE II — PROCEDIMENTO DE APLICAÇÃO 21
4 PARTE III — VERSÃO A (QUALITATIVA) 33
5 PARTE IV — VERSÃO B (QUANTITATIVA) 37
6 PARTE V — INSTRUÇÕES DE USO 41
7 PARTE VI — VALIDAÇÃO LOCAL E REVISÃO 50
REFERÊNCIAS 56

1 APRESENTAÇÃO DO INSTRUMENTO

A MARIAH é o módulo operacional de apoio à identificação, classificação e documentação dos riscos relacionados ao uso de inteligência artificial em pesquisas com seres humanos. É aplicável, conforme o contexto, a protocolos das Ciências da Saúde, das Ciências Humanas e Sociais e a pesquisas interdisciplinares.

A MARIAH não substitui a análise ética nem produz decisão automática de aprovação ou reprovação. Sua função é estruturar a análise, tornar explícitos os fatores de risco, apoiar a proporcionalidade dos requisitos e aumentar a rastreabilidade da deliberação do Comitê de Ética em Pesquisa (CEP).

A classificação deve refletir as características do sistema e do protocolo. Medidas de mitigação devem ser avaliadas separadamente quanto à sua suficiência e verificabilidade; não devem apagar o reconhecimento do risco inerente que pretendem controlar.

1.1 LINGUAGEM REGULATÓRIA E OPERACIONAL

As orientações desta MARIAH observam as seguintes categorias:

- **DEVE —** requisito normativo ou condição necessária à avaliação ética.
- **RECOMENDA-SE —** boa prática que fortalece proteção, transparência, qualidade ou governança.
- **PODE —** alternativa admissível, dependente do contexto do protocolo.

A utilização dessas expressões não altera a hierarquia das normas vigentes. Obrigações normativas devem manter indicação de sua fonte.

2 PARTE I — FUNDAMENTOS E CONSTRUÇÃO DA MARIAH

Esta Parte reproduz, para consulta autônoma, o Capítulo 8 do Caderno Referencial.

2.1 MARIAH: MATRIZ DE AVALIAÇÃO DE RISCO EM INTELIGÊNCIA ARTIFICIAL (VERSÕES A E B)

2.1.1 O percurso de construção e as novas questões para os Comitês de Ética em Pesquisa

Esta Parte pode ser consultada de forma sequencial ou por seção, conforme a necessidade do leitor. O Quadro 1 organiza as seções por tema e permite localizar rapidamente o conteúdo de interesse.

A construção de um instrumento de avaliação ética para protocolos de pesquisa que utilizam inteligência artificial exige, antes de qualquer decisão metodológica, uma escolha de postura: o instrumento existe para apoiar o julgamento do CEP ou para substituí-lo?

Este guia responde a essa pergunta sem ambiguidade. As matrizes apresentadas nas Partes III (Versão A) e IV (Versão B) desta MARIAH são ferramentas de APOIO à deliberação. Elas não são novas normas, não criam obrigações além das já estabelecidas no âmbito do Sistema Nacional de Ética em Pesquisa (SINEP), e não têm caráter vinculante. Sua adoção é uma escolha institucional de cada CEP. O que elas oferecem é precisão: um percurso que torna a avaliação mais sistemática, mais documentada e mais justificável perante os participantes da pesquisa e a sociedade.

Quadro 1 — Guia de leitura da Parte I

Se você precisa saber…	**Vá para**
Como e por que a MARIAH foi construída	Seção 2.1.1
Os fundamentos teóricos e normativos das duas versões	Seção 2.1.2
Como funciona a Versão A (qualitativa / fluxograma)	Seção 2.1.2.1
Como funciona a Versão B (quantitativa / pontuação)	Seção 2.1.2.2
O passo a passo de aplicação da matriz	Parte II desta MARIAH
Os quatro níveis de risco e seus requisitos	os Passos 3 e 4
A regra crítica e sua relação com a pontuação	a regra crítica

Fonte: elaboração própria dos autores.

Esse posicionamento não é uma concessão, mas uma consequência direta do que a experiência internacional demonstrou. Nenhum dos oito modelos regulatórios avaliados conseguiu criar um instrumento universal eficaz para todos os contextos institucionais previstos.

O modelo canadense de Avaliação de Impacto Algorítmico é robusto, mas foi desenhado para serviços públicos governamentais, não para pesquisa com seres humanos (Canadá, 2025). As diretrizes do Conselho Nacional de Saúde e Pesquisa Médica da Austrália (National Health and Medical Research Council — NHMRC) são precisas no tratamento do ciclo de vida, mas não contemplam as especificidades jurídicas da proteção de dados em países de renda média (Austrália, 2023). As diretrizes do Conselho Indiano de Pesquisa Médica (Indian Council of Medical Research — ICMR) são as únicas criadas para orientar CEPs na avaliação de protocolos com IA, mas foram desenvolvidas para um contexto diferente do brasileiro (ICMR, 2023). Cada modelo ofereceu contribuições reais e apresentou lacunas reais. A MARIAH é a síntese dessas contribuições, ancorada no direito brasileiro e orientada pelo contexto do SUS.

O percurso que levou à construção das duas versões que serão apresentadas nas próximas seções foi deliberadamente longo. A análise considerou cinco eixos de avaliação consagrados internacionalmente e, durante a pesquisa, incluiu contribuições como o conceito de Oportunidade de Contestar (*Opportunity to Override*), que expressa a condição real do profissional contestar o resultado do sistema, da experiência coreana (MFDS, 2023). O direito do participante de pesquisa de solicitar a exclusão de seus dados do treinamento do modelo provém do modelo indiano (ICMR, 2023). O Protocolo de Mudança de Algoritmos (*Algorithm Change Protocol*), que trata da responsabilidade do pesquisador de prever como o modelo será retreinado, decorre da experiência regulatória indiana e japonesa (CDSCO, 2025; PMDA, [s.d.]). O conceito de contexto de uso como etapa prévia obrigatória à classificação de risco deriva da diretriz da Food and Drug Administration (FDA) estadunidense (FDA, 2025). Cada um desses elementos ampliou e qualificou o instrumento final.

O resultado desse percurso são duas versões de um mesmo instrumento, com a mesma base ética e o mesmo conjunto de perguntas essenciais. O Quadro 2 apresenta a estrutura comparativa das duas versões.

O material operacional correspondente está na Parte II desta MARIAH.

A avaliação ética de protocolos com IA não pode ser reduzida a uma formalidade. Nenhuma das duas versões da matriz admite esse risco.

A IA assimila os vieses dos dados de treinamento, utiliza lógicas pouco transparentes e gera impactos sobre participantes que nenhum outro método contemporâneo produz com igual escala e rapidez. Avaliar um protocolo com IA não é avaliar uma tecnologia, mas o impacto que essa tecnologia pode produzir sobre seres humanos concretos, em contextos concretos, com vulnerabilidades concretas. As matrizes deste guia existem para tornar essa avaliação mais rigorosa. A responsabilidade pela avaliação permanece, integralmente, com os CEPs.

Quadro 2 — Estrutura comparativa das duas versões da MARIAH

Dimensão	Versão A (qualitativa)	Versão B (quantitativa)
Lógica de operação	Fluxograma sequencial	Pontuação ponderada
Tipo de resposta	Binária (sim/não)	Pesos numéricos diferenciados
Estrutura	5 eixos temáticos	7 blocos temáticos
Escala de risco	Níveis I a IV (pior cenário)	Níveis I a IV (soma de 275 pts)
Efeito das mitigações	Não afetam o nível de risco	Reduzem a pontuação total (Bloco 7)
Contexto preferencial	Primeiros ciclos de avaliação com IA; protocolos de menor complexidade	Experiência consolidada; protocolos complexos e auditáveis
Referência principal	Modelo australiano (NHMRC) + canadense (AIA)	Modelo canadense (AIA)
Instrumento completo	Parte III desta MARIAH (Versão A)	Parte IV desta MARIAH (Versão B)

Fonte: elaboração própria dos autores.

2.1.2 Fundamentos, fontes e descrição das matrizes da MARIAH

Esta seção apresenta as bases teóricas, normativas e metodológicas que sustentam as duas versões da MARIAH. Cada escolha de desenho tem fundamento em experiências regulatórias documentadas e no marco jurídico brasileiro. Isso vale para os eixos, as perguntas, os pesos e a lógica de chegada ao nível de risco. A descrição das duas versões está organizada em subseções independentes. A seção 2.1.2.1 trata da Versão A Qualitativa. A seção 2.1.2.2 trata da Versão B Quantitativa. As duas versões compartilham a mesma base

ética e o mesmo conjunto de perguntas essenciais. O que as distingue é a lógica de operação. Essa distinção determina o público e o contexto para os quais cada versão é mais adequada.

A MARIAH é um instrumento de gradação proporcional. Opera sobre um espectro que vai do risco baixo ao risco crítico e reconhece que a maioria dos protocolos com IA ocupa posições intermediárias que admitem salvaguardas, condicionantes e mitigações. Essa gradação tem um limite que o instrumento não transpõe. Três dimensões não são graduáveis pela MARIAH: dano direto e potencialmente irreversível ao participante sem mecanismo documentado de interrupção e reparação; ausência de supervisão humana efetiva em contexto em que o sistema influencia decisões clínicas críticas; e incerteza técnica não documentada sobre o comportamento do sistema no contexto de uso declarado.

Diante dessas dimensões, o princípio de não maleficência e o princípio de precaução, estabelecidos pela Lei n.º 14.874/2024, prevalecem sobre qualquer resultado que a soma dos blocos produza (Brasil, 2024a). A matriz estrutura o julgamento do CEP, não esgota seus limites éticos. Esses limites estão nos princípios que a lei estabelece e que nenhum instrumento de avaliação tem competência para relativizar.

2.1.2.1 Versão A qualitativa

A avaliação ética de protocolos de pesquisa que utilizam IA exige instrumentos que traduzam a complexidade técnica dos sistemas algorítmicos de impacto ético mensurável sobre a pessoa participante. A MARIAH, Versão A qualitativa, foi construída a partir dessa premissa, com fundamento em oito experiências regulatórias internacionais e ancorada no marco jurídico brasileiro.

A escolha pela abordagem qualitativa

A Versão A da Matriz adota uma lógica qualitativa estruturada: perguntas de resposta binária, organizadas em cinco eixos temáticos, percorridos em sequência por meio de um fluxograma. Essa escolha se apoia na experiência do modelo australiano, que demonstrou que instrumentos de avaliação ética são mais eficazes quando orientam o julgamento do avaliador sem substituí-lo. A avaliação não exige domínio da arquitetura técnica do sistema, mas a compreensão do impacto que ele pode produzir. A lógica qualitativa preserva essa

centralidade do julgamento ético, ao mesmo tempo que garante sistematicidade e rastreabilidade ao processo de avaliação (Austrália, 2023).

O formato de perguntas binárias, em que qualquer resposta de risco aumenta a classificação no eixo, foi adaptado do modelo canadense de Avaliação de Impacto Algorítmico (Algorithmic Impact Assessment — AIA). O AIA canadense demonstrou, em uso consolidado desde 2019, que perguntas objetivas e bem formuladas são instrumentos eficazes para avaliadores sem formação técnica em ciência da computação. A versão brasileira adota essa lógica e a amplia para o cenário da pesquisa que envolve seres humanos. Ela direciona a atenção da prestação de serviços públicos governamentais para a proteção dos participantes dessas pesquisas (Canadá, 2025).

O filtro de entrada e a caracterização do contexto de uso

O Passo 0 (filtro de entrada) distingue usos rotineiros de usos que demandam avaliação aprofundada. Essa distinção é essencial. Sem um filtro inicial, os comitês analisam indiscriminadamente desde ferramentas para organização bibliográfica até sistemas autônomos de apoio diagnóstico. Isso compromete a eficiência e a proporcionalidade do processo. A diretriz da FDA dos Estados Unidos da América (EUA), publicada em janeiro de 2025, demonstrou que nenhuma avaliação de risco algorítmico começa sem a definição precisa do contexto de uso. Trata-se do papel específico do sistema e do escopo da pergunta que ele foi desenvolvido para responder (FDA, 2025). A Matriz incorpora essa exigência como etapa prévia obrigatória aos cinco eixos.

Os cinco eixos e suas fontes

O Eixo 1, dedicado à natureza e à autonomia do sistema, articula dois conceitos complementares. O primeiro é o de influência do modelo: o quanto a IA pesa na decisão final, em relação a outras fontes de evidência (FDA, 2025). O Ministério da Segurança Alimentar e de Medicamentos da Coreia do Sul (Ministry of Food and Drug Safety — MFDS) define dois critérios (MFDS, 2023): Potencialidade de Má Função, ou seja, a chance de falhas algorítmicas causarem dano clínico; e Oportunidade de Contestação, que é a possibilidade real do profissional questionar o resultado do sistema. A combinação dessas duas dimensões permite

avaliar não apenas o que o sistema faz, mas o que acontece quando ele erra. O mais importante, nesse momento, é saber quem detém a capacidade de intervir.

O Eixo 2 examina o impacto sobre a pessoa participante. Suas bases estão na Lei n.º 14.874, de 28 de maio de 2024, e no Decreto n.º 12.651, de 2025, que estabelecem os princípios éticos fundamentais da pesquisa com seres humanos no Brasil (Brasil, 2024a, 2025). O eixo inclui pergunta específica sobre populações vulneráveis: crianças, gestantes, idosos, povos indígenas e pessoas em situação de vulnerabilidade socioeconômica. Essa inclusão responde a uma lacuna dos modelos internacionais analisados. Nenhum deles foi concebido para o contexto epidemiológico e social brasileiro, marcado por desigualdades regionais acentuadas e pela pluralidade étnica e cultural da população atendida pelo Sistema Único de Saúde (SUS). O eixo incorpora ainda pergunta sobre mecanismo de exclusão de dados do treinamento do modelo. Essa pergunta tem fundamento nas diretrizes do Indian Council of Medical Research (ICMR), que institui o direito de o participante solicitar a supressão de seus dados do repositório utilizado pela IA a qualquer tempo (ICMR, 2023). Nenhum dos modelos ocidentais analisados prevê esse direito.

O Eixo 3, sobre sensibilidade e governança dos dados, é o eixo mais diretamente ancorado no direito brasileiro. A Lei Geral de Proteção de Dados Pessoais (LGPD) classifica dados de saúde como dados pessoais sensíveis e exige base legal específica, finalidade declarada e identificação do Encarregado de Dados para seu tratamento. Esse conjunto de exigências é mais protetivo do que as legislações de proteção de dados de todos os países analisados neste guia, à exceção do Regulamento Geral de Proteção de Dados da União Europeia (GDPR).

O contraste torna-se ainda mais evidente em relação à Lei de Proteção de Dados Pessoais Digitais da Índia (*Digital Personal Data Protection Act*), de 2023, que eliminou a categoria de dados sensíveis do texto legal e passou a tratar dados de saúde como qualquer outro dado pessoal (Carnegie Endowment, 2023). Equiparar prontuários, dados genéticos e registros psiquiátricos a dados comerciais correntes expõe os participantes de pesquisa a riscos de discriminação sem respaldo jurídico para proteção diferenciada.

A LGPD brasileira faz escolha oposta. Dados de saúde constituem categoria sensível com proteção reforçada. Essa distinção é um dos pilares do Eixo 3 da Matriz. O eixo inclui ainda pergunta específica sobre risco de reidentificação. O MRCT Center e WCG (2025)

documentaram que grandes modelos de linguagem conseguem inferir identidades a partir de dados aparentemente anonimizados. Esse risco não desaparece com a anonimização formal.

A Resolução CNS n.º 738, de 1.º de fevereiro de 2024, operacionaliza a LGPD no contexto específico da pesquisa com bancos de dados envolvendo seres humanos. Define as figuras jurídicas do Controlador, do Operador e do Titular (art. 3.º, incisos IV, XIII e XX) e estabelece os quatro Termos institucionais que estruturam a cadeia de custódia do dado: Acordo, Anuência, Compromisso de Uso e Transferência (art. 3.º, XVI a XIX). A Resolução distingue ainda duas situações de uso de bancos de dados — constituídos no âmbito da pesquisa (Capítulo VII) e já constituídos fora desse âmbito (Capítulo VIII) — com exigências documentais distintas, e regula o uso futuro de dados e as hipóteses de dispensa do Termo de Consentimento Livre e Esclarecido (TCLE) (art. 20, §§ 2.º, 3.º e 5.º, e art. 25). A MARIAH reúne essas exigências em duas subseções dedicadas a bancos de dados — Eixo 3.b na Versão A qualitativa e Bloco 6.b na Versão B quantitativa — que incorporam os elementos em três grupos de perguntas operacionais: o protocolo identifica o Controlador do banco de dados? Apresenta os Termos institucionais exigidos pela natureza do banco (especialmente o Termo de Anuência Institucional, obrigatório em bancos fora do âmbito da pesquisa)? Apresenta, quando houver pedido de dispensa de TCLE, a hipótese normativa invocada e justificativa ética e metodológica suficiente (art. 20, § 5.º, ou art. 25)? (Brasil, 2024b).

O material operacional correspondente está na Parte II desta MARIAH.

O Eixo 4, sobre representatividade e transferibilidade, responde a uma das principais lacunas éticas identificadas na literatura internacional: modelos treinados predominantemente com dados de populações de alta renda e do hemisfério norte apresentam desempenho sistematicamente inferior quando aplicados a populações de diferentes características étnicas, socioeconômicas e geográficas (OMS, 2024). O guia australiano e as diretrizes do MFDS coreano documentaram que a ausência de validação local é um dos principais fatores de risco ético em protocolos que importam sistemas desenvolvidos em outros contextos (Austrália, 2023; MFDS, 2023). O eixo incorpora ainda pergunta sobre *data drift*. Modelos de aprendizado de máquina degradam seu desempenho à medida que os dados do mundo real divergem dos dados de treinamento. Esse fenômeno exige monitoramento contínuo ao longo de todo o ciclo de vida do sistema. A pergunta tem fundamento no conceito de manutenção do ciclo de vida desenvolvido pela FDA (2025) e

operacionalizado pelo sistema IDATEN da Agência de Dispositivos Médicos e Produtos Farmacêuticos do Japão (Pharmaceuticals and Medical Devices Agency — PMDA, [s.d.]).

O Eixo 5 examina a supervisão humana e a explicabilidade. Seu fundamento central é o princípio da autonomia médica irrevogável enunciado pelo ICMR (2023) indiano. A inteligência artificial serve para amplificação diagnóstica e não substitui o julgamento clínico. O princípio é formulado para a prática clínica. Em pesquisa, ele incide sobre o desenho da supervisão descrito no protocolo: o que o CEP avalia não é o ato assistencial futuro, mas se o protocolo garante revisão e contestação humanas antes da execução de cada decisão apoiada pelo sistema. A soberania sobre a saída do sistema cabe, nesse plano, ao pesquisador responsável, sem prejuízo da responsabilidade do profissional assistente quando houver ato clínico decorrente. O pesquisador responsável é o garantidor da verificação humana de todo resultado gerado pela IA. Esse princípio articula-se com o conceito de humano no circuito decisório (*human-in-the-loop*) presente no modelo australiano (Austrália, 2023).

Ele se articula com a exigência da FDA de que, quando o contexto de uso envolve um humano no processo decisório, a avaliação de desempenho do sistema considere o resultado da equipe humano-IA, não apenas o desempenho isolado do modelo (FDA, 2025). O eixo incorpora ainda pergunta sobre o protocolo de retreinamento do modelo. Essa pergunta tem fundamento no Protocolo de Mudança de Algoritmo (Algorithm Change Protocol) indiano, que exige que os fabricantes submetam antecipadamente um plano que descreva como o modelo será atualizado. Esse requisito foi estabelecido pela Organização Central de Controle de Padrões de Medicamentos da Índia (CDSCO, do inglês Central Drugs Standard Control Organization) (Índia, 2025). No contexto deste guia, esse princípio se traduz na responsabilidade do pesquisador de informar ao CEP como e quando o modelo será modificado durante o estudo.

A convergência de alguns eixos em torno do impacto clínico sobre o participante é intencional, não é sobreposição inadvertida. Os eixos de natureza e autonomia do sistema, de impacto sobre a pessoa participante e de transferibilidade e desempenho avaliam dimensões distintas de uma mesma realidade: as diferentes formas de impacto do sistema sobre a pessoa que participa da pesquisa. O eixo de natureza examina como o sistema opera. O eixo de impacto examina o que o sistema produz sobre o participante. O eixo de transferibilidade examina se o que o sistema produz num contexto continuará válido noutro. São perguntas diferentes sobre o mesmo risco porque o risco tem múltiplas dimensões que um único eixo

não captura. Um protocolo que apresenta alto risco nesses três eixos simultaneamente não é penalizado três vezes pelo mesmo problema. Demonstra que o risco é consistente, estrutural e não mitigável por uma única salvaguarda. A convergência entre eixos num resultado de alto risco é evidência de gravidade, não artefato metodológico.

A lógica do fluxograma sequencial e a regra crítica

O fluxograma sequencial foi adotado por dois motivos. O primeiro é pedagógico. A sequência conduz a avaliação de forma progressiva, do mais simples ao mais complexo, e evita saltos que comprometam a consistência da análise. O segundo é ético. O nível de risco só sobe, nunca desce. Essa lógica é coerente com o princípio da precaução que fundamenta a ética em pesquisa com seres humanos (Brasil, 2025). Um único eixo em Nível IV é suficiente para elevar o protocolo inteiro ao nível mais alto de exigência, independentemente do desempenho nos demais eixos. A ausência de supervisão humana em um sistema que influencia decisões terapêuticas, por exemplo, não é compensada por boas respostas nos eixos de dados ou representatividade.

A regra crítica tem origem direta no modelo canadense de Avaliação de Impacto Algorítmico (Canadá, 2025). O modelo demonstrou que permitir que mitigações reduzam o nível de risco cria um incentivo perverso. Pesquisadores passam a declarar salvaguardas para obter classificações mais favoráveis. Deixam de demonstrar que o risco foi efetivamente controlado. A regra resolve esse problema com uma separação clara. O nível de risco é determinado pelas características do sistema e do protocolo. Os requisitos exigidos são ajustados em função das garantias demonstradas. Essa separação preserva a integridade do instrumento e a clareza da comunicação entre pesquisador e comitê.

2.1.2.2 Versão B quantitativa

A Versão B da MARIAH utiliza perguntas com pesos numéricos variados em sete blocos temáticos. A soma dos pontos determina o risco do protocolo em uma escala de I a IV. Essa

abordagem torna suas decisões mais rastreáveis e auditáveis. A pontuação identifica áreas de concentração de risco e mostra como as escolhas do pesquisador afetam esse risco.

A escolha pela abordagem quantitativa

A lógica quantitativa tem inspiração no modelo canadense de Avaliação de Impacto Algorítmico (Canadá, 2025). O modelo opera com 65 perguntas de risco, pontuação máxima de 169 pontos, e 41 perguntas de mitigação, pontuação máxima de 75 pontos, distribuídas em quatro níveis de risco.

A experiência canadense demonstrou duas vantagens concretas do modelo quantitativo. A primeira é a precisão: a pontuação identifica em quais dimensões o risco está concentrado e orienta as exigências do comitê de forma proporcional. A segunda é a consistência: protocolos avaliados em momentos diferentes e por membros diferentes do CEP tornam-se comparáveis, o que fortalece a coerência das decisões institucionais (Canadá, 2025). A Versão B incorpora essa lógica e a adapta ao contexto da pesquisa com seres humanos no Brasil. O foco deixa de ser serviços públicos governamentais e passa a ser a proteção dos participantes de pesquisa e a conformidade com a LGPD.

Este guia propõe uma estrutura em sete blocos: Projeto, Sistema, Algoritmo, Decisão, Impacto, Dados e Mitigação. Os blocos de maior peso ético e regulatório recebem perguntas com valor individual mais alto. Essa escolha torna explícita a hierarquia de prioridades do instrumento. Os blocos de Impacto e Dados concentram juntos 139 dos 275 pontos máximos da matriz. Representam 50,5% da pontuação total. Essa concentração expressa uma escolha deliberada. Em pesquisa com seres humanos, o que mais importa é o que o sistema pode fazer à pessoa participante e como seus dados são tratados.

O filtro de entrada e a caracterização do contexto de uso

A Versão B mantém, da Versão A, o Passo 0 e a caracterização do contexto de uso como etapas obrigatórias anteriores à pontuação. A diretriz da FDA (EUA) demonstrou que nenhuma avaliação de risco algorítmico começa sem a definição precisa do contexto de uso:

o papel específico do sistema e o escopo da pergunta que ele foi desenvolvido para responder (FDA, 2025). A caracterização não pontua. Delimita o objeto da avaliação e impede que o sistema seja avaliado em abstrato, desvinculado do protocolo concreto submetido à análise.

Os sete blocos e suas fontes

O Bloco 1 avalia a adequação do protocolo para o uso de IA. Examina a transparência na descrição do sistema, a qualificação da equipe, a clareza da justificativa metodológica e a adequação do TCLE. A pontuação máxima é de 22 pontos. Cada pergunta vale três pontos, com exceção da pergunta sobre consulta técnica externa, que vale quatro. O bloco opera como verificação de completude documental.

O protocolo deve descrever o estágio de desenvolvimento do sistema: descoberta exploratória (*Discovery*), tradução clínica (*Translation*) ou implantação (*Deployment*). Essa exigência tem fundamento no referencial do MRCT Center e WCG (2025), que demonstrou que os riscos éticos variam conforme o momento do ciclo de vida em que a IA opera. Um sistema em fase de descoberta exploratória apresenta riscos distintos de um sistema já em implantação clínica. Essa distinção deve ser visível no protocolo submetido ao CEP.

O Bloco 2 avalia a natureza técnica do sistema: autonomia, adaptabilidade, explicabilidade e rastreabilidade. A pontuação máxima é de 17 pontos. Cada pergunta vale dois ou três pontos. O bloco articula dois conceitos desenvolvidos por referenciais distintos. O primeiro é o de influência do modelo: o quanto a IA pesa na decisão final em relação a outras fontes de evidência. Esse conceito foi desenvolvido pela FDA (2025). O segundo é o de potencial de falha (*Malfunction Potential*) e oportunidade de contestação (*Opportunity to Override*): a probabilidade de que falhas algorítmicas causem dano clínico e a existência de condições reais para que o profissional responsável questione o resultado do sistema.

Esse conceito foi desenvolvido pelo Ministério da Segurança Alimentar e de Medicamentos da Coreia do Sul (MFDS, 2023). As perguntas de risco direto valem três pontos cada. Tratam de situações sem intervenção humana obrigatória, de sistemas adaptativos e de sistemas opacos. Sua proximidade com o risco direto ao participante justifica o peso maior. As perguntas de ausência de garantia valem dois pontos. Tratam de déficits documentais que, embora graves, são mais facilmente corrigíveis.

O Bloco 3 avalia a qualidade, a origem e a representatividade dos dados de treinamento e de teste. A pontuação máxima é de 14 pontos. Cada pergunta vale dois pontos. O bloco tem fundamento direto no guia australiano e nas diretrizes do MFDS coreano, que documentaram que a ausência de validação local é um dos principais fatores de risco ético em protocolos que utilizam sistemas desenvolvidos em outros contextos (Austrália, 2023; MFDS, 2023). O bloco inclui pergunta sobre independência entre dados de treinamento e de teste. A diretriz da FDA estabelece que dados de teste devem ser rigorosamente independentes dos dados de desenvolvimento para que a avaliação de desempenho do modelo seja válida (FDA, 2025). Inclui também pergunta sobre representatividade por subgrupos populacionais. A Organização Mundial da Saúde (OMS) identificou que modelos treinados predominantemente com dados de populações de alta renda e do hemisfério norte apresentam desempenho sistematicamente inferior quando aplicados a populações de diferentes características étnicas, socioeconômicas e geográficas (OMS, 2024). Essa lacuna é especialmente relevante para o contexto brasileiro.

O Bloco 4 avalia a consequência da decisão. A pontuação máxima é de oito pontos. As duas primeiras perguntas valem três pontos cada. Essa assimetria é deliberada. As perguntas desse bloco tratam das situações de maior gravidade potencial em todo o instrumento: o sistema como único determinante de uma decisão sem revisão humana e a irreversibilidade do dano em caso de erro. A concentração de pontos em poucas perguntas de alto peso reflete a lógica da consequência da decisão desenvolvida pela FDA (2025). Quando o modelo opera com alta influência e baixa supervisão humana, a severidade do evento adverso potencial deve ser o fator dominante na avaliação de risco. A pontuação máxima desse bloco, por si só, pode deslocar um protocolo do Nível II para o Nível III. Esse deslocamento expressa a gravidade das situações que essas perguntas descrevem.

O Bloco 5 avalia o impacto sobre a pessoa participante. A pontuação máxima é de 62 pontos. Cada pergunta vale cinco ou seis pontos. As perguntas de maior impacto direto valem seis pontos: influência em decisões clínicas, risco de dano físico, risco de dano psíquico ou social e envolvimento de populações vulneráveis. O bloco tem suas bases na Lei n.º 14.874, de 28 de maio de 2024, e no Decreto n.º 12.651, de 2025, que estabelecem os princípios éticos fundamentais da pesquisa com seres humanos no Brasil (Brasil, 2024a, 2025). O bloco inclui pergunta sobre inferência de características sensíveis não coletadas explicitamente. O MRCT Center e WCG (2025) documentaram que grandes modelos de linguagem conseguem inferir

identidades e atributos sensíveis a partir de dados aparentemente neutros. Esse risco não desaparece com a anonimização formal. O bloco inclui ainda pergunta sobre mecanismo de exclusão de dados do treinamento do modelo. As diretrizes do Indian Council of Medical Research instituem o direito do participante de solicitar a supressão de seus dados do repositório utilizado pela IA a qualquer tempo (ICMR, 2023). Nenhum dos modelos ocidentais analisados prevê esse direito. Este guia o incorpora como exigência ética.

O Bloco 6 avalia a sensibilidade e a governança dos dados. A pontuação máxima é de 77 pontos. Cada pergunta vale entre cinco e sete pontos. As perguntas diretamente vinculadas à LGPD valem sete pontos cada. O bloco expressa uma escolha central do guia. A lei é o diferencial protetivo mais robusto do marco brasileiro em relação a todos os modelos internacionais analisados.

A LGPD classifica dados de saúde como dados pessoais sensíveis e exige base legal específica, finalidade declarada e identificação do Encarregado de Dados para seu tratamento. Esse nível de proteção supera o da Lei de Proteção de Informações Pessoais chinesa (PIPL), o da Lei de Proteção de Informações Pessoais japonesa (APPI) e, especialmente, o da Lei de Proteção de Dados Pessoais Digitais indiana (*Digital Personal Data Protection Act*), de 2023, que suprimiu inteiramente a categoria de dados sensíveis de sua redação e equiparou prontuários de saúde a quaisquer outros dados pessoais (Carnegie Endowment, 2023).

A pergunta sobre transferência de dados para servidores fora do Brasil vale sete pontos. Conforme o país receptor e as garantias contratuais estabelecidas, essa transferência pode comprometer integralmente a proteção conferida pela LGPD aos participantes da pesquisa. A pergunta sobre risco de reidentificação incorpora a preocupação documentada com a capacidade dos sistemas de IA de inferir identidades a partir de dados aparentemente anonimizados (MRCT Center; WCG, 2025).

A subseção Bloco 6.b da Versão B reúne cinco pontos específicos para protocolos que utilizam bancos de dados, com pesos calibrados pela centralidade operacional na Resolução CNS n.º 738/2024 (Brasil, 2024b): identificação nominal do Controlador do banco — P6.b.1 (sete pontos); apresentação do Termo de Anuência Institucional quando exigível pela origem do banco — P6.b.2 (sete pontos, ELIMINATÓRIA); apresentação do Termo de Compromisso de Uso de Dados — P6.b.3 (cinco pontos); enquadramento fundamentado de eventual dispensa de TCLE na hipótese cabível segundo a origem do banco — P6.b.4 (cinco pontos); e

formalização da controladoria conjunta em pesquisa multicêntrica, por Termo de Acordo Institucional — P6.b.5 (cinco pontos).

O Bloco 6.b soma 29 pontos à pontuação total quando ativado pelo filtro do Passo 0. A ausência de resposta afirmativa em P6.b.2, quando aplicável, torna o protocolo não avaliável no mérito pela regra crítica (item "A regra crítica na Versão B", adiante), independentemente da pontuação total. A subseção integra ainda três itens de diligência — P6.b.4.1 (necessidade e proporcionalidade do acesso a dados identificáveis), P6.b.6 (critérios cumulativos para dispensa de TCLE em pesquisa com IA) e P6.b.7 (declaração do Controlador) — que não pontuam nem se somam ao teto: operam como *gate* cumulativo, e a ausência de qualquer critério implica devolução do protocolo.

O Bloco 7 avalia as mitigações. É o único bloco com lógica bidirecional: perguntas de risco somam pontos ao total e perguntas de mitigação subtraem. A pontuação máxima é de 75 pontos. A capacidade de redução por mitigações chega a 65 pontos — às quais se somam até 23 pontos de evidências do sub-bloco 7C. O bloco operacionaliza uma distinção herdada do modelo canadense. Mitigações não alteram o nível de risco determinado pelos demais blocos quando avaliados isoladamente. Reduzem a pontuação total e, por isso, podem deslocar o protocolo para um nível inferior na escala consolidada (Canadá, 2025). Essa lógica é mais flexível do que a da Versão A, em que mitigações não afetam o nível em nenhuma hipótese. Reflete a compreensão de que, na Versão B, a pontuação já incorpora a qualidade das salvaguardas adotadas pelo pesquisador como parte integrante da avaliação de risco. O sub-bloco 7C, de evidências de transparência, opera em regime distinto, de só-abate: a presença documentada da evidência subtrai pontos, e a ausência não soma.

O bloco se divide em três sub-blocos. O Sub-bloco 7A (consultas) vale dez pontos. Avalia se o pesquisador buscou orientação técnica externa antes da submissão, inclusive consulta à Agência Nacional de Vigilância Sanitária (Anvisa) sobre o enquadramento do sistema na categoria Programa de Computador como Dispositivo Médico (Software as a Medical Device — SaMD). O Sub-bloco 7B (medidas de redução de risco, *de-risking*) vale 65 pontos. Avalia a qualidade dos mecanismos de supervisão humana, explicabilidade, monitoramento do ciclo de vida e gestão de falhas. O sub-bloco inclui pergunta sobre monitoramento de deriva de dados (*data drift*), ou seja, a degradação do desempenho do modelo ao longo do tempo. Essa pergunta tem fundamento no conceito de manutenção do ciclo de vida desenvolvido pela FDA (2025), operacionalizado pelo sistema IDATEN da agência

regulatória japonesa (PMDA, [s.d.]) e pelo Protocolo de Mudança de Algoritmo (*Algorithm Change Protocol*) indiano (CDSCO, 2025). O Sub-bloco 7C (evidências de transparência) registra evidências documentais — documentação padronizada do modelo e da base de dados, avaliação de impacto algorítmico ou Relatório de Impacto à Proteção de Dados Pessoais e disponibilização de código-fonte — que abatem até 23 pontos quando presentes e documentadas, sem somar quando ausentes.

As faixas de nível e a calibração da pontuação

As quatro faixas de nível foram calibradas a partir de dois princípios. O primeiro é a proporcionalidade. A faixa do Nível I é estreita, de zero a 58 pontos, porque protocolos com uso de IA verdadeiramente de baixo risco são raros e devem ser classificados com rigor. O segundo é a sensibilidade ao risco. A faixa do Nível IV cobre 67 pontos, de 209 a 275, porque protocolos nesse nível apresentam combinações de características graves em múltiplos blocos. A amplitude da faixa reflete a variação de gravidade dentro do próprio nível crítico. Quando o filtro do Passo 0 ativa o Bloco 6.b (banco de dados), o teto da matriz passa de 275 para 304 pontos e as faixas são recalibradas proporcionalmente: Nível I de 0 a 64 pontos, Nível II de 65 a 141 pontos, Nível III de 142 a 230 pontos e Nível IV de 231 a 304 pontos (teto avaliável 297 pontos, descontada a questão eliminatória P6.b.2). A recalibração preserva a distância relativa entre os níveis e evita que os 29 pontos adicionais do Bloco 6.b distorçam a classificação. As faixas serão testadas empiricamente durante o período de uso e teste da versão 1.0; eventuais ajustes serão considerados na versão 2.0, a ser submetida à Inaep.

O material operacional correspondente está na Parte II desta MARIAH.

A regra crítica na Versão B

Na Versão B, a regra crítica assume formulação diferente da Versão A. O nível determinado pela pontuação não pode ser alterado por decisão subjetiva do avaliador. A pontuação é o resultado sistemático da aplicação do instrumento. As mitigações já estão incorporadas na pontuação do Bloco 7. Constituem a única forma legítima de reduzir a pontuação final. Essa arquitetura preserva a integridade do instrumento. Impede que declarações de intenção não documentadas substituam evidências de controle efetivo do

risco. O modelo canadense demonstrou que esse princípio é essencial para a credibilidade do instrumento perante pesquisadores, patrocinadores e a sociedade (Canadá, 2025).

A gradação descrita até aqui convive com quatro salvaguardas que não se graduam. São a regra especial do Eixo 3.b, a Cláusula de Prevalência Ética, a eliminatória de cadeia de custódia e a diligência impeditiva de novo consentimento, todas descritas no Suplemento à Nota Técnica de Premissas. Operam de modo automático, sem depender do julgamento de quem aplica a matriz. As duas primeiras determinam o nível de risco por si sós. As duas últimas suspendem a análise de mérito e devolvem o protocolo ao pesquisador em diligência. Nenhuma delas admite compensação por mitigações declaradas, porque todas tratam de condições que precedem a gradação do risco. Os gatilhos e o procedimento de cada uma estão na Parte II desta MARIAH.

O material operacional correspondente está na Parte II desta MARIAH. O Quadro 3 resume as fontes internacionais por eixo e bloco da MARIAH.

Quadro 3 — Fontes internacionais por eixo/bloco da MARIAH

Eixo/Bloco	Dimensão avaliada	Fonte principal	Contribuição à MARIAH
Eixo 1 / Bloco 2	Natureza e autonomia do sistema	FDA (2025); MFDS (2023)	Influência do modelo; potencial de falha; oportunidade de contestação
Eixo 2 / Bloco 5	Impacto sobre a pessoa participante	Brasil (2024a; 2025); ICMR (2023); MRCT Center e WCG (2025)	Proteção de populações vulneráveis; direito de exclusão de dados; inferência de atributos sensíveis
Eixo 3 / Bloco 6	Sensibilidade e governança dos dados	LGPD; Res. CNS 738/2024; GDPR; Carnegie Endowment (2023)	Dados sensíveis com proteção reforçada; Controlador, Operador e Termos institucionais (Res. CNS 738/2024); risco de reidentificação; transferência internacional
Eixo 4 / Bloco 3	Representatividade e transferibilidade	OMS (2024); Austrália (2023); MFDS (2023); FDA (2025)	Validação local; desempenho por subgrupos; monitoramento de deriva de dados
Eixo 5 / Bloco 7	Supervisão humana e explicabilidade	ICMR (2023); FDA (2025); PMDA [s.d.]; CDSCO (2025)	Autonomia médica irrevogável; humano no circuito decisório; protocolo de retreinamento

Eixo/Bloco	Dimensão avaliada	Fonte principal	Contribuição à MARIAH
Bloco 1	Adequação do protocolo	MRCT Center e WCG (2025)	Estágio do ciclo de vida (*discovery, translation, deployment*)
Bloco 4	Consequência da decisão	FDA (2025)	Influência do modelo × supervisão humana × severidade do desfecho

Fonte: Austrália (2023); Brasil (2024a; 2025); Canadá (2025); Carnegie Endowment (2023); CDSCO (2025); FDA (2025); ICMR (2023); MFDS (2023); MRCT Center e WCG (2025); PMDA [s.d.]; OMS (2024).

3 PARTE II — PROCEDIMENTO DE APLICAÇÃO

Como usar esta MARIAH

A Versão A, qualitativa, opera por fluxograma sequencial: percorrem-se cinco eixos em ordem, com respostas de sim ou não, e o nível de risco emerge da combinação das respostas. É um instrumento orientado ao raciocínio e guia a avaliação sem substituir o julgamento humano.

Esse procedimento é indicado para CEPs que começam a se familiarizar com a avaliação de protocolos que envolvem IA, para casos em que os protocolos apresentam menor complexidade técnica, e quando é prioritário garantir transparência no processo deliberativo em vez de um resultado numericamente preciso.

A abordagem do pior cenário, que considera um único eixo em Nível IV como elevador de todo o protocolo, reflete o princípio da precaução presente na ética da pesquisa que envolve seres humanos (Brasil, 2012; Brasil, 2024a; Brasil, 2025).

A Versão B quantitativa opera por pontuação ponderada: as mesmas perguntas recebem pesos numéricos diferenciados, distribuídos em sete blocos temáticos, e a soma determina o nível de risco em uma escala de 275 pontos. O instrumento orienta à rastreabilidade. Permite identificar onde o risco está concentrado, como as mitigações adotadas o reduziram e como protocolos distintos se comparam entre si.

A escolha por essa versão justifica-se para CEPs com maior experiência na avaliação de protocolos com IA, para protocolos de alta complexidade técnica e para instituições que precisam documentar e auditar suas decisões com precisão, como hospitais universitários, institutos nacionais de pesquisa e patrocinadores de ensaios clínicos multicêntricos. Os blocos de Impacto e Dados concentram 50,5% da pontuação total. Essa concentração expressa a hierarquia de prioridades do instrumento. O que mais importa, em pesquisa com seres humanos, é o que o sistema pode fazer à pessoa participante e como seus dados são tratados.

As duas versões são compatíveis e complementares. Um CEP pode adotar a Versão A como triagem inicial e recorrer à Versão B para aprofundar a análise de protocolos que a Versão A classifique nos Níveis III ou IV. Um CEP pode adotar a Versão B sistematicamente e usar a Versão A como guia de orientação para pesquisadores que precisam compreender a lógica da avaliação antes de submeter seus protocolos.

Um CEP pode, ainda, optar por nenhuma das versões e usar este capítulo como marco conceitual para desenvolver seu próprio instrumento, adaptado às especificidades de sua área temática, de sua estrutura institucional e do perfil dos protocolos que recebe. A flexibilidade garante a validade do guia. Ignorar a diversidade institucional dos CEPs brasileiros comprometeria seu propósito.

Em termos operacionais, a avaliação da subseção Eixo 3.b (Versão A) ou Bloco 6.b (Versão B) pelo CEP, em protocolos que utilizam bancos de dados, inclui cinco perguntas concretas, derivadas dos artigos citados no Caderno Referencial. A primeira (3.b.1 / P6.b.1): o Controlador do banco está identificado nominalmente, com cargo, função e instituição vinculada? A segunda (3.b.2 / P6.b.2, ELIMINATÓRIA): para bancos já constituídos fora do âmbito da pesquisa, o protocolo apresenta Termo de Anuência Institucional assinado pelo dirigente da instituição detentora dos dados? A terceira (3.b.3 / P6.b.3): o Termo de Compromisso de Uso de Dados assinado pelo pesquisador responsável consta no dossiê do protocolo? A quarta (3.b.4 / P6.b.4): pedido de dispensa de TCLE, se houver, indica a hipótese normativa aplicável e apresenta justificativa ética e metodológica suficiente — art. 20, § 5.º, para bancos constituídos fora do âmbito da pesquisa, ou art. 25, para bancos constituídos no âmbito da pesquisa? A quinta (3.b.5 / P6.b.5): em pesquisa multicêntrica, a controladoria conjunta está formalizada por Termo de Acordo Institucional?

A ausência de qualquer um desses termos, quando exigível pela natureza do banco, não configura pendência menor: configura ausência de cadeia de custódia formalizada, com consequências na regra crítica (Caderno Referencial, Capítulo 8, item "A regra crítica").

A Parte VI desta MARIAH oferece um roteiro prático para CEPs que desejam testar localmente as faixas de nível em sua própria casuística. O roteiro propõe três frentes independentes: concordância entre avaliadores, distribuição empírica dos protocolos nos quatro níveis e convergência entre as duas versões da matriz. Além disso, acompanha uma planilha-modelo para auxiliar no processo. A adesão é voluntária. O retorno de resultados ao Grupo de Trabalho responsável pelo guia é facultativo. Contribuições recebidas serão consideradas em ciclos futuros de revisão.

Como aplicar a matriz: passos e procedimentos

A matriz opera em quatro passos. O Passo 0 verifica se o protocolo exige avaliação aprofundada ou se o uso de IA é rotineiro. O Passo 1 delimita o objeto da avaliação por meio de oito perguntas descritivas obrigatórias. O Passo 2 percorre os cinco eixos de avaliação, que examinam a natureza do sistema, o impacto sobre o participante, a governança dos dados, a representatividade do modelo e a supervisão humana. O Passo 3 traduz os resultados em um dos quatro níveis de risco algorítmico. O Passo 4 associa a cada nível um conjunto de requisitos mínimos que o pesquisador deve demonstrar ter atendido. As seções a seguir descrevem cada passo. Os instrumentos completos constam nas Partes III e IV desta MARIAH.

Passo 0: Filtro de entrada

O percurso de avaliação começa antes dos eixos e antes da pontuação. O primeiro passo é verificar se o protocolo submetido ao CEP demanda a aplicação da matriz. A pergunta é direta: o sistema de inteligência artificial realiza, neste protocolo, automação de decisão, geração de conteúdo ou intervenção na condução do estudo ou nos seus resultados?

Se a resposta for negativa, o uso é classificado como rotineiro e a matriz não se aplica. Sistemas que operam exclusivamente como suporte administrativo, organização de referências bibliográficas, transcrição de áudio, revisão gramatical, adequação de estilo ou tradução — ou qualquer outra função que não influencie as decisões do estudo, nem produza efeitos sobre a pessoa participante — não entram no escopo da avaliação, ainda que o seu uso deva ser declarado. O protocolo segue o fluxo ordinário do CEP.

Se a resposta for afirmativa, prossegue-se para a caracterização do contexto de uso. O filtro não dispensa a menção ao uso de IA no protocolo. O pesquisador deve declarar qualquer uso de sistemas algorítmicos, ainda que rotineiro. O que o filtro determina é o nível de profundidade da avaliação.

A distinção entre uso rotineiro e uso não rotineiro tem fundamento no referencial do MRCT Center e WCG (2025). A ausência de um critério de entrada claro leva os CEPs a avaliarem indiscriminadamente ferramentas de baixíssimo impacto em sistemas de alto risco. Esse equívoco compromete a eficiência e a proporcionalidade do processo. O filtro de entrada resolve esse problema antes que ele ocorra.

A obrigação de declarar o uso de IA na pesquisa antecede as matrizes deste guia. A Política de Integridade na Atividade Científica do Conselho Nacional de Desenvolvimento Científico e Tecnológico, instituída pela Portaria de 6 de março de 2026, determina que qualquer uso de IA generativa deve ser declarado, com especificação da ferramenta utilizada e da finalidade, em qualquer fase do desenvolvimento da pesquisa (CNPq, 2026).

O filtro de entrada da matriz verifica, antes de qualquer outra etapa, se essa declaração está presente e se o papel do sistema está descrito com precisão suficiente para viabilizar a avaliação. A ausência dessas informações não configura lacuna documental. Configura violação de integridade científica com respaldo normativo expresso na legislação federal vigente.

No mesmo Passo 0, um segundo filtro, com função complementar, verifica se o protocolo utiliza banco de dados, quer seja próprio, de outra pesquisa ou de instituição externa. A resposta afirmativa ativa as subseções dedicadas a bancos de dados, derivadas da Resolução CNS n.º 738/2024: o Eixo 3.b na Versão A (qualitativa) e o Bloco 6.b na Versão B (quantitativa). Essas subseções operacionalizam a identificação do Controlador, a verificação dos Termos institucionais aplicáveis e o enquadramento de eventual dispensa de TCLE. Se o protocolo não utiliza banco de dados, por exemplo, quando coleta dados exclusivamente no escopo da pesquisa com cada participante, as perguntas do Eixo 3.b / Bloco 6.b ficam inaplicáveis.

Nesse caso, a avaliação prossegue com os demais eixos. A precisão desse segundo filtro evita duas falhas simétricas: exigir documentos que a natureza do protocolo torna desnecessários e deixar de exigir os que a natureza do protocolo torna obrigatórios.

Passo 1: Caracterização do contexto de uso

Aprovado o filtro de entrada, o protocolo deve responder a oito perguntas descritivas, obrigatórias em ambas as versões da matriz.

A primeira: qual é a pergunta que o sistema de IA foi desenvolvido para responder neste protocolo? A segunda: o sistema responde a essa pergunta de forma autônoma ou subsidia uma decisão tomada por um ser humano? A terceira: qual é o estado de identificação dos dados pessoais utilizados — identificáveis, pseudonimizados, anonimizados ou combinação por etapa do estudo? A quarta: qual é o volume aproximado de registros de

participantes? A quinta: em que momento ocorre a anonimização em relação ao acesso pelos pesquisadores — antes do acesso, pela instituição custodiante; após o acesso, pelos próprios pesquisadores; não haverá anonimização; ou não se aplica (responder somente se os dados não forem anonimizados)? A sexta: qual é o papel do sistema de IA — ferramenta de análise sobre dados da pesquisa, objeto de desenvolvimento (treinamento de modelo) ou ambos? A sétima: qual é o destino previsto do modelo — uso restrito a este protocolo, aplicações futuras ou uso amplo, compartilhamento com terceiros, ou não se aplica? A oitava: haverá compartilhamento do banco de dados, total ou parcial, com terceiros, além das instituições participantes — não; sim, sob Termo de Anuência ou Acordo Institucional; ou sim, outra forma?

Essas perguntas não pontuam e não determinam o nível de risco por si mesmas. Elas delimitam o objeto da avaliação empedem que o sistema seja avaliado em abstrato, desvinculado do protocolo concreto submetido à análise. Um mesmo sistema de IA pode apresentar perfis de risco radicalmente distintos conforme o papel que desempenha em cada protocolo. Um modelo de linguagem utilizado para apoiar a codificação de dados qualitativos é um objeto de avaliação diferente do mesmo modelo utilizado para gerar recomendações terapêuticas personalizadas. O contexto de uso define o que deve ser analisado.

É impossível avaliar a credibilidade de um sistema de IA para fins regulatórios sem antes definir com precisão para qual pergunta ele foi desenvolvido e em qual contexto de uso específico será aplicado. Esse conceito foi desenvolvido pela FDA dos EUA (FDA, 2025). A MARIAH incorpora essa exigência como etapa prévia obrigatória, anterior a qualquer classificação de risco.

Passo 2: Os cinco eixos de avaliação

Após caracterizar o contexto de uso, avaliam-se os cinco eixos do risco algorítmico. Cada eixo corresponde a uma dimensão, cujo nível de risco é definido por respostas qualitativas (Versão A) ou pontuação ponderada (Versão B). Os cinco eixos são descritos a seguir. As perguntas de cada eixo constam integralmente nas Partes III e IV desta MARIAH.

O Eixo 1 examina a natureza e a autonomia do sistema. Avalia se opera de forma autônoma ou sob supervisão humana obrigatória, se é adaptativo, se é explicável e se foi validado antes do uso no protocolo.

O eixo articula dois referenciais distintos. A FDA (2025) desenvolveu o conceito de influência do modelo: o quanto a IA pesa na decisão final em relação a outras fontes de evidência. O Ministério da Segurança Alimentar e de Medicamentos da Coreia do Sul desenvolveu os conceitos de potencial de falha (*Malfunction Potential*) e oportunidade de contestação (*Opportunity to Override*) (MFDS, 2023). O primeiro avalia a probabilidade de que falhas algorítmicas gerem risco clínico. O segundo avalia o grau em que os profissionais têm condições concretas de questionar as decisões do sistema.

O Eixo 2 examina o impacto sobre a pessoa participante. Avalia se o sistema afeta decisões clínicas, se pode causar danos físicos, psíquicos, sociais ou econômicos, se envolve populações vulneráveis e se infere características sensíveis não fornecidas diretamente pelo participante. O eixo tem fundamento direto na legislação brasileira de pesquisa com seres humanos (Brasil, 2024a, 2025). Incorpora ainda o direito do participante de solicitar a exclusão de seus dados do treinamento do modelo, instituído pelas diretrizes do Indian Council of Medical Research (ICMR, 2023).

O Eixo 3 examina a sensibilidade e a governança dos dados. Avalia se os dados tratados no protocolo são sensíveis nos termos da LGPD, se são transferidos para servidores fora do Brasil, se há risco real de reidentificação, se o consentimento cobre o uso para treinamento de modelos e se há plano de resposta a incidentes. O eixo está mais diretamente ancorado na legislação brasileira do que qualquer outro da matriz. Na Versão B, recebe o maior peso regulatório do instrumento.

O Eixo 4 examina a representatividade e a transferibilidade do modelo. Avalia se os dados de treinamento representam a diversidade da população-alvo do protocolo, se o modelo foi validado em contexto comparável ao SUS, se há avaliação de desempenho por subgrupos populacionais e se os dados de teste são independentes dos dados de treinamento.

Esse eixo responde a uma das principais lacunas éticas identificadas na literatura internacional. Modelos treinados predominantemente com dados de populações de alta renda e do hemisfério norte apresentam desempenho sistematicamente inferior quando aplicados a populações de diferentes características étnicas, socioeconômicas e geográficas (OMS, 2024). Essa lacuna é especialmente crítica para protocolos conduzidos no âmbito do SUS.

O Eixo 5 examina a supervisão humana e a explicabilidade. Avalia se há revisão humana obrigatória antes da execução das decisões, se a equipe tem competência técnica para

interpretar e contestar os resultados, se há monitoramento contínuo do desempenho e se há protocolo documentado para falhas e para o retreinamento do modelo.

O fundamento central do eixo é o princípio da autonomia médica irrevogável enunciado pelo ICMR (2023). A inteligência artificial serve para amplificação diagnóstica e não substitui o julgamento clínico. O princípio é formulado para a prática clínica. Em pesquisa, ele incide sobre o desenho da supervisão descrito no protocolo: o que o CEP avalia não é o ato assistencial futuro, mas se o protocolo garante revisão e contestação humanas antes da execução de cada decisão apoiada pelo sistema. A soberania sobre a saída do sistema cabe, nesse plano, ao pesquisador responsável, sem prejuízo da responsabilidade do profissional assistente quando houver ato clínico decorrente. O eixo incorpora ainda o conceito de manutenção do ciclo de vida desenvolvido pela FDA (2025), operacionalizado pelo sistema de Avaliação Tempestiva de Inovação IDATEN (Improvement Design within Approval for Timely Evaluation and Notice) da agência regulatória japonesa (PMDA, [s.d.]) e pelo Protocolo de Mudança de Algoritmo (Algorithm Change Protocol) indiano (CDSCO, 2025).

Passo 3: Níveis I a IV

O resultado da avaliação pelos cinco eixos é a classificação do protocolo em um dos quatro níveis de risco algorítmico. Os níveis não são categorias de aprovação ou reprovação, são classificações que determinam o conjunto de requisitos mínimos que o pesquisador deve demonstrar ter atendido antes da aprovação pelo CEP.

O Nível I (Baixo) corresponde a protocolos em que o sistema de IA opera com autonomia limitada, sem influência direta sobre decisões clínicas ou terapêuticas nem sobre os procedimentos e desfechos da pesquisa — alocação em braços, elegibilidade, coleta e análise —, com dados de baixa sensibilidade e supervisão humana efetiva.

Exemplos típicos incluem o uso de IA para organização e categorização de referências bibliográficas, transcrição automatizada de entrevistas ou apoio à análise exploratória de dados não identificados.

O Nível II (Moderado) corresponde a protocolos em que o sistema de IA desempenha papel mais relevante no protocolo, mas com consequências reversíveis e supervisão humana garantida. São exemplos típicos os protocolos que empregam modelos de linguagem para apoio à codificação qualitativa, sistemas de triagem de elegibilidade de participantes ou ferramentas de análise de imagens médicas com revisão humana obrigatória.

O Nível III (Alto) corresponde a protocolos em que o sistema de IA influencia diretamente decisões com potencial de dano ao participante, opera com dados sensíveis em larga escala ou apresenta limitações documentadas de representatividade e transferibilidade. São exemplos típicos os protocolos que empregam sistemas de apoio diagnóstico em unidades básicas de saúde, modelos preditivos de risco em ensaios clínicos ou ferramentas de estratificação de pacientes para protocolos terapêuticos.

O Nível IV corresponde ao risco crítico. Enquadram-se nesse nível os protocolos em que o sistema opera com alta autonomia em decisões de impacto direto e potencialmente irreversível sobre o participante, sem supervisão humana efetiva ou com lacunas graves de validação e governança. São exemplos típicos os protocolos que empregam sistemas autônomos de decisão terapêutica, ferramentas de diagnóstico sem revisão humana obrigatória ou modelos de prognóstico utilizados para definir elegibilidade em estudos de acesso a tratamentos experimentais. O Quadro 4 sintetiza os quatro níveis de risco algorítmico e seus requisitos.

Quadro 4 — Níveis de risco algorítmico: síntese

Nível	Classificação	Perfil típico	Requisitos mínimos (cumulativos)
I	**Baixo**	IA sem influência em decisões clínicas; dados de baixa sensibilidade; supervisão humana efetiva.	Declaração do uso de IA e transparência adequada às pessoas afetadas, quando aplicável.
II	**Moderado**	IA com papel relevante, consequências reversíveis, supervisão garantida.	+ Descrição do sistema; conformidade LGPD; base legal
III	**Alto**	IA influencia decisões com potencial de dano; dados sensíveis em larga escala; limitações de representatividade.	+ Validação técnica; desempenho por subgrupos; *human-in-the-loop*; Encarregado de Dados.
IV	**Crítico**	Alta autonomia, impacto direto e potencialmente irreversível, sem supervisão efetiva.	+ Parecer externo; monitoramento contínuo; notificação Anvisa (SaMD); apreciação por CEP acreditado ou habilitado em IA.

Fonte: elaboração própria.

Passo 4: Requisitos por nível

Cada nível corresponde a um conjunto de requisitos mínimos que o pesquisador deve atender para a aprovação do protocolo pelo CEP. Os requisitos são cumulativos. Os do Nível II incluem os do Nível I. Os do Nível III incluem os dos Níveis I e II. Os do Nível IV incluem todos os anteriores.

No Nível I, exige-se a declaração de uso de IA no protocolo e a menção, no TCLE, de que um sistema algorítmico será utilizado e qual é o seu papel no estudo.

No Nível II, acrescentam-se a descrição do sistema e a origem dos dados de treinamento, a demonstração de conformidade com a LGPD e a indicação da base legal para o tratamento dos dados utilizados.

No Nível III, acrescentam-se a documentação de validação técnica do sistema, a análise de desempenho por subgrupos populacionais relevantes, a comprovação de mecanismo de supervisão humana efetiva no circuito decisório (*human-in-the-loop*) e a identificação do Encarregado de Dados da instituição responsável pelo estudo.

No Nível IV, acrescentam-se o parecer técnico externo sobre o sistema, o plano de monitoramento contínuo com critérios definidos de interrupção, a notificação à Anvisa quando o sistema se enquadrar como SaMD e a submissão do protocolo à apreciação de CEP acreditado para protocolos de risco elevado ou de CEP com habilitação específica em inteligência artificial. O Quadro 5 reúne as perguntas essenciais para aplicação da MARIAH.

Quadro 5 — Perguntas essenciais para aplicação da MARIAH

N.º	Pergunta que o protocolo deve responder
1	O protocolo declara o uso de IA com especificação da ferramenta e da finalidade?
2	O sistema opera de forma autônoma ou subsidia decisão humana?
3	Qual é a pergunta que o sistema foi desenvolvido para responder neste protocolo?
4	Há supervisão humana efetiva no circuito decisório?
5	Os dados de treinamento representam a diversidade da população-alvo do SUS?
6	O protocolo demonstra conformidade com a LGPD, com base legal específica e Encarregado de Dados identificado?
7	Existe mecanismo documentado de exclusão dos dados do participante do treinamento do modelo?
8	Há protocolo de monitoramento de deriva de dados e de retreinamento do modelo?
9	As mitigações declaradas estão documentadas com evidências ou são declarações de intenção?

N.º	Pergunta que o protocolo deve responder
10	Alguma dimensão do protocolo configura as condições não graduáveis pela MARIAH (dano irreversível sem interrupção, ausência de supervisão humana em decisão clínica crítica, incerteza técnica não documentada)?

Fonte: elaboração própria do organizador.

A regra crítica

O nível de risco determinado pela avaliação não muda com as mitigações apresentadas pelo pesquisador após a classificação. Isso vale para a lógica qualitativa da Versão A e para a pontuação da Versão B. Apenas os requisitos exigidos podem ser ajustados em função das garantias demonstradas.

A regra tem fundamento direto no modelo canadense de Avaliação de Impacto Algorítmico (Canadá, 2025). Permitir que mitigações apenas declaradas, apresentadas após a classificação, reduzam o nível de risco cria um incentivo estruturalmente perverso. Pesquisadores passam a declarar salvaguardas para obter classificações mais favoráveis. Deixam de demonstrar que o risco foi efetivamente controlado. A separação entre nível de risco e requisitos exigidos preserva a integridade do instrumento ereserva também a clareza da comunicação entre pesquisador e comitê.

Na Versão B, essa mesma lógica opera por desenho: as mitigações não entram depois da classificação: já estão incorporadas ao cálculo, na pontuação do Bloco 7, e só subtraem pontos quando documentadas e verificáveis. Essa é a única via pela qual mitigações influenciam o resultado. Nenhuma medida apresentada após a classificação altera o nível obtido.

A regra crítica expressa o princípio que orienta toda esta seção. A avaliação não recai sobre intenções, mas sobre evidências. A que um protocolo descreve como supervisionada, explicável e validada precisa demonstrar que é supervisionada, explicável e validada. A matriz existe para tornar essa demonstração sistemática. A responsabilidade por exigi-la permanece com o Comitê.

Já a regra crítica incorpora hipóteses eliminatórias e hipóteses de prevalência ética. A Resolução CNS n.º 738/2024 elimina protocolos que usam banco de dados externo sem Termo de Anuência Institucional assinado pelo dirigente da instituição (art. 27, VI). Isso corresponde

à pergunta 3.b.2 na Versão A e à pergunta P6.b.2 na Versão B, ambas marcadas como ELIMINATÓRIAS (Brasil, 2024b).

A falta de anuência indica que a instituição de origem não autorizou o uso dos dados. Em tal caso, o dossiê é devolvido ao pesquisador para diligência obrigatória antes de qualquer análise de mérito.

Algumas falhas de cadeia de custódia aumentam o nível de risco, mas não bloqueiam automaticamente. A falta de identificação do Controlador (perguntas 3.b.1 / P6.b.1) e a pedido de dispensa de TCLE sem indicação da hipótese normativa aplicável, sem justificativa ética e metodológica suficiente ou sem salvaguardas proporcionais (perguntas 3.b.4 / P6.b.4) elevam o protocolo conforme a regra de consolidação da subseção.

A Cláusula de Prevalência Ética representa a segunda categoria de intervenção direta na classificação. Marcar afirmativamente as perguntas P4.1 (decisão sem revisão humana) ou P4.2 (dano irreversível ao participante) aciona o princípio da precaução. Em tais casos, o CEP pode elevar o protocolo ao Nível IV imediatamente, independentemente da pontuação total, conforme descrito ao final da Parte IV desta MARIAH.

Ao contrário das hipóteses eliminatórias, a Cláusula de Prevalência não descarta o protocolo do fluxo de avaliação. Em vez disso, o eleva ao nível mais alto de requisitos.

A coexistência entre pontuação ponderada e regra crítica merece formulação explícita para evitar leitura ambígua. As duas operam em momentos distintos e com lógicas distintas.

A pontuação percorre os sete blocos, acumula o perfil de risco distribuído do protocolo e produz uma classificação baseada na soma total, já com as mitigações do Bloco 7 incorporadas. A regra crítica entra depois, como cláusula de prevalência ética: verifica se alguma dimensão do protocolo apresenta falha estrutural que a soma total não deve ocultar. Quando a regra crítica se aplica, ela não anula a pontuação, prevalece sobre ela.

O resultado é que um protocolo com soma favorável, mas com falha estrutural numa dimensão crítica, não deve ser interpretado como de baixo risco. A pontuação mapeia o risco distribuído. A regra crítica protege contra a ilusão de que a média pode ocultar o inaceitável. As duas operam na mesma direção: garantir que a avaliação recaia sobre evidências, não sobre intenções. O Quadro 6 apresenta as conexões com os demais módulos do conjunto documental.

Quadro 6 — Conexões com os demais módulos do conjunto documental

Para aprofundar...	Vá para
As propriedades críticas dos sistemas de IA — viés, explicabilidade, supervisão humana e ciclo de vida	Caderno Referencial, Capítulo 3
A proteção de dados dos participantes e o tratamento de dados em IA	Caderno Referencial, Capítulo 7
As bases normativas brasileiras aplicáveis à pesquisa com IA	Caderno Referencial, Capítulo 4
O roteiro de perguntas éticas para qualquer protocolo com IA	Caderno Referencial, Capítulo 1
A capacitação do CEP para avaliação prática de protocolos com IA	Partes V e VI desta MARIAH

Fonte: elaboração própria.

4 PARTE III — VERSÃO A (QUALITATIVA)

MARIAH — Versão A (Qualitativa)

MARIAH — MATRIZ DE AVALIAÇÃO DE RISCO EM INTELIGÊNCIA ARTIFICIAL — VERSÃO A (QUALITATIVA)

EIXO 1: NATUREZA E AUTONOMIA DO SISTEMA

O eixo articula o conceito de influência do modelo (FDA, 2025) com os conceitos de potencial de falha e oportunidade de contestação desenvolvidos pelo Ministério da Segurança Alimentar e de Medicamentos da Coreia do Sul (MFDS, 2023).

Resultado do Eixo 1:
0 respostas de risco → Nível I
1 a 2 respostas de risco → Nível II
3 a 4 respostas de risco → Nível III
5 ou mais respostas de risco → Nível IV

EIXO 2: IMPACTO SOBRE A PESSOA PARTICIPANTE

O eixo tem fundamento na legislação brasileira de pesquisa com seres humanos (Brasil, 2024a, 2025) e incorpora o direito do participante de solicitar a exclusão de seus dados do treinamento do modelo, instituído pelo Indian Council of Medical Research (ICMR, 2023).

Resultado do Eixo 2:
0 respostas de risco → Nível I
1 a 2 respostas de risco → Nível II
3 a 4 respostas de risco → Nível III
5 ou mais respostas de risco → Nível IV

EIXO 3: SENSIBILIDADE E GOVERNANÇA DOS DADOS

O eixo está ancorado na LGPD e nas diretrizes do Indian Council of Medical Research (ICMR, 2023) sobre governança de dados em pesquisa com inteligência artificial.

Resultado do Eixo 3:
0 respostas de risco → Nível I
1 a 2 respostas de risco → Nível II
3 a 4 respostas de risco → Nível III
5 ou mais respostas de risco → Nível IV

EIXO 3.b: BANCOS DE DADOS DE PESQUISA (Res. CNS n.º 738/2024)
Nota: o Eixo 3.b é subseção complementar ao Eixo 3, ativada pelo filtro de banco de dados do Passo 0.

Subseção aplicável quando o protocolo utiliza bancos de dados — próprios, de outra pesquisa ou de instituição externa —, conforme filtro ativado no Passo 0 (Parte II desta MARIAH). As perguntas derivam da Resolução CNS n.º 738/2024 e complementam o Eixo 3 sem duplicá-lo. O resultado desta subseção pode elevar o nível consolidado do protocolo (ver Regra de consolidação ao final).

Resultado do Eixo 3.b:
0 respostas de risco → não eleva o nível consolidado
1 a 2 respostas de risco → eleva para Nível III
3 ou mais respostas de risco → eleva para Nível IV

Hipótese eliminatória: resposta "não" à pergunta 3.b.2 (Termo de Anuência Institucional ausente em banco fora do âmbito da pesquisa) configura ausência de cadeia de custódia formalizada. O protocolo não é avaliável no mérito e segue para diligência obrigatória antes de qualquer análise, conforme o item "A regra crítica" do Caderno Referencial, Capítulo 8.

EIXO 4: REPRESENTATIVIDADE E TRANSFERIBILIDADE

O eixo tem fundamento no guia australiano (NHMRC, 2023), nas diretrizes do MFDS coreano (Austrália, 2023) e no referencial do MRCT Center e WCG (2025) sobre representatividade de modelos em contextos populacionais diversos.

Resultado do Eixo 4:
0 respostas de risco → Nível I
1 a 2 respostas de risco → Nível II
3 a 4 respostas de risco → Nível III
5 ou mais respostas de risco → Nível IV

EIXO 5: SUPERVISÃO HUMANA E EXPLICABILIDADE

O eixo tem fundamento no princípio da autonomia médica irrevogável enunciado pelo Indian Council of Medical Research (ICMR, 2023) e no conceito de manutenção do ciclo de vida desenvolvido pela FDA (2025), operacionalizado pelo sistema IDATEN da agência regulatória japonesa (PMDA, [s.d.]) e pelo Protocolo de Mudança de Algoritmo (Algorithm Change Protocol) indiano (CDSCO, 2025), e é reforçado pelas recomendações de 2026 da National Commission into the Regulation of AI in Healthcare do Reino Unido, que enfatizam

supervisão proporcional ao ciclo de vida, monitoramento contínuo, equidade como dimensão de segurança, rastreabilidade e gestão de mudanças.

Resultado do Eixo 5:
0 respostas de risco → Nível I
1 a 2 respostas de risco → Nível II
3 a 4 respostas de risco → Nível III
5 ou mais respostas de risco → Nível IV

MÓDULO TRANSVERSAL OBRIGATÓRIO — CICLO DE VIDA, MUDANÇAS E DESEMPENHO

As perguntas abaixo complementam os Eixos 1, 4 e 5 e devem ser respondidas quando aplicáveis. Elas não constituem um sexto eixo autônomo e não alteram, nesta versão, as faixas de classificação. Respostas que revelem risco devem ser registradas no eixo correspondente e consideradas na regra de consolidação e na avaliação da suficiência das salvaguardas.

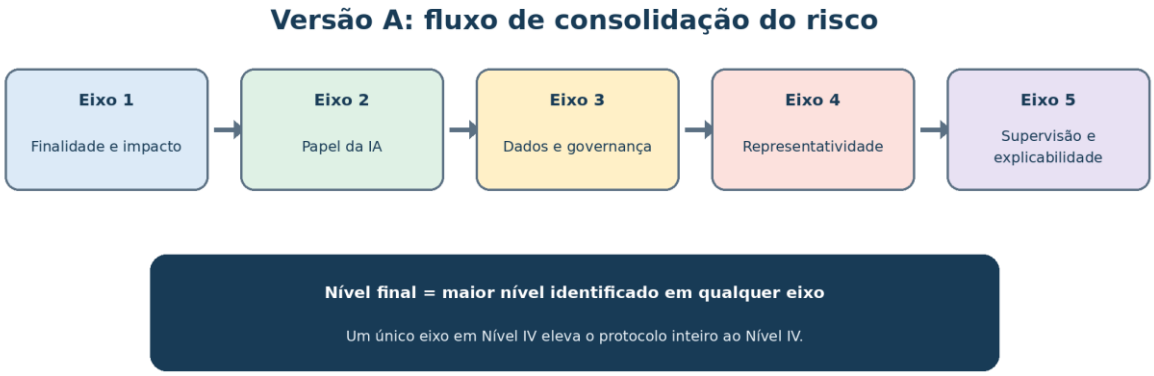
1. O protocolo identifica claramente o sistema de IA e, quando pertinente, sua versão ou configuração relevante?
2. O protocolo informa se o sistema poderá ser atualizado, recalibrado, retreinado, substituído ou modificado durante a pesquisa?
3. Quando houver possibilidade de mudança, existe procedimento para avaliar se a alteração modifica riscos, benefícios, equidade, privacidade, autonomia, supervisão humana ou outras salvaguardas?
4. O protocolo define mecanismos proporcionais para monitorar o desempenho da IA durante a pesquisa e detectar falhas, resultados inesperados ou deterioração relevante?
5. Existem critérios ou procedimentos previamente definidos para revisão, suspensão, interrupção ou adoção de medidas adicionais quando o desempenho ou o perfil de risco se modificar?
6. O protocolo avalia a possibilidade de desempenho diferencial entre grupos, subgrupos ou contextos relevantes para a população da pesquisa?
7. Está definido quem exerce supervisão humana, com que autoridade e como se procede quando houver discordância entre a saída da IA e o julgamento humano?
8. Quando proporcional ao risco, é possível rastrear posteriormente qual sistema, versão ou configuração contribuiu para o resultado, inferência ou decisão relevante?

Sinal de alerta: sistema com impacto relevante sobre participantes sem identificação de versão quando necessária, sem mecanismo de gestão de mudanças, sem monitoramento proporcional do desempenho ou sem supervisão humana efetiva.

REGRA DE CONSOLIDAÇÃO — FLUXOGRAMA SEQUENCIAL

O CEP percorre os cinco eixos em ordem. O nível final é o mais alto identificado em qualquer eixo. Um único eixo em Nível IV eleva o protocolo inteiro ao Nível IV. A Figura 1 apresenta o fluxo de consolidação da Versão A, e o Quadro 7, os critérios de consolidação do nível de risco.

Figura 1 — Fluxo de consolidação da Versão A da MARIAH



Fonte: elaboração própria, com base na regra de consolidação da MARIAH.

Quadro 7 — Critérios de consolidação do nível de risco da Versão A

Nível consolidado	Critério
I Baixo	Todos os eixos em Nível I
II Moderado	Pelo menos um eixo em Nível II; nenhum acima
III Alto	Pelo menos um eixo em Nível III; nenhum em IV
IV Crítico	Pelo menos um eixo em Nível IV

Fonte: elaboração própria.

5 PARTE IV — VERSÃO B (QUANTITATIVA)

MARIAH — Versão B (Quantitativa)

MARIAH — MATRIZ DE AVALIAÇÃO DE RISCO EM INTELIGÊNCIA ARTIFICIAL — VERSÃO B (QUANTITATIVA)

Sistema de pontuação: cada pergunta respondida como risco soma pontos ao total. O nível é determinado pela soma final. Perguntas de mitigação, quando respondidas positivamente, subtraem pontos.

Faixas de nível:

- Sem banco de dados (teto 275): Nível I — 0 a 58 pts; Nível II — 59 a 127 pts; Nível III — 128 a 208 pts; Nível IV — 209 a 275 pts.
- Com banco de dados (Bloco 6.b ativado; teto teórico 304, teto avaliável 297 descontada a P6.b.2 eliminatória): Nível I — 0 a 64 pts; Nível II — 65 a 141 pts; Nível III — 142 a 230 pts; Nível IV — 231 a 304 pts.
- Nota metodológica: os cortes derivam das frações fixas do *baseline* (I = 50/238, II = 110/238, III = 180/238) aplicadas ao teto teórico vigente, com arredondamento de meia-unidade para cima.

PASSO 0: FILTRO DE ENTRADA

Antes de iniciar a pontuação, o CEP verifica:

O sistema de IA realiza automação de decisão, geração de conteúdo que integre o desenho, os instrumentos de coleta, o TCLE ou os resultados, ou intervenção no protocolo ou na condução do estudo? (casos-limite — assistência redacional, modelos prognósticos — no Passo 0 do Caderno Referencial, Capítulo 8)
Não → uso rotineiro. A matriz não se aplica.
Sim → prosseguir para a caracterização do contexto de uso.

CARACTERIZAÇÃO DO CONTEXTO DE USO

As oito perguntas desta etapa são descritivas, não pontuam e são obrigatórias antes da pontuação.
Qual é a pergunta que o sistema de IA foi desenvolvido para responder neste protocolo?
O sistema responde a essa pergunta de forma autônoma ou subsidia uma decisão tomada por um ser humano?
Qual é o estado de identificação dos dados pessoais utilizados neste protocolo? (Identificáveis; pseudonimizados; anonimizados; combinação por etapa do estudo)

Qual é o volume aproximado de registros de participantes utilizados neste protocolo?
Em que momento ocorre a anonimização dos dados em relação ao acesso pelos pesquisadores? (antes do acesso, pela instituição custodiante; após o acesso, pelos próprios pesquisadores; não haverá anonimização; não se aplica) — responder somente se a resposta anterior não for “anonimizados”.
Qual é o papel do sistema de IA neste protocolo? (ferramenta de análise sobre dados da pesquisa; objeto de desenvolvimento — treinamento de modelo; ambos)
Qual é o destino previsto do modelo de IA utilizado ou desenvolvido neste protocolo? (uso restrito a este protocolo; aplicações futuras ou uso amplo; compartilhamento com terceiros; não se aplica — IA apenas como ferramenta)
Haverá compartilhamento do banco de dados, total ou parcial, com terceiros além das instituições participantes do protocolo? (não haverá; sim, sob Termo de Anuência ou Acordo Institucional; sim, outra forma — descrever)

BLOCO 1: PROJETO
Pontuação máxima: 22 pts | Valor por pergunta: 3 pts
Avalia a adequação do protocolo para o uso de IA: transparência, qualificação da equipe e clareza da proposta.
Subtotal Bloco 1: _____ / 22 pts

BLOCO 2: SISTEMA
Pontuação máxima: 17 pts | Valor por pergunta: 2 a 3 pts
Avalia a natureza técnica do sistema: autonomia, adaptabilidade e rastreabilidade.
Subtotal Bloco 2: _____ / 17 pts

BLOCO 3: ALGORITMO
Pontuação máxima: 14 pts | Valor por pergunta: 2 pts
Avalia a qualidade, origem e representatividade dos dados de treinamento e teste.
Subtotal Bloco 3: _____ / 14 pts

BLOCO 4: DECISÃO
Pontuação máxima: 8 pts | Valor por pergunta: 2 a 3 pts
Avalia o grau de autonomia decisória do sistema e a consequência de um erro: as perguntas mais críticas da matriz.
Subtotal Bloco 4: _____ / 8 pts

BLOCO 5: IMPACTO
Pontuação máxima: 62 pts | Valor por pergunta: 5 a 6 pts
Avalia o impacto direto sobre a pessoa participante. Bloco de maior peso ético da matriz.
Subtotal Bloco 5: _____ / 62 pts

BLOCO 6: DADOS
Pontuação máxima: 77 pts | Valor por pergunta: 5 a 7 pts

Avalia a sensibilidade, a governança e a conformidade jurídica do tratamento de dados. Bloco de maior peso regulatório da matriz, ancorado na LGPD.
Subtotal Bloco 6: _____ / 77 pts

BLOCO 6.b: BANCOS DE DADOS DE PESQUISA (Res. CNS n.º 738/2024)
Nota: o Bloco 6.b é subseção complementar ao Bloco 6, ativada pelo filtro de banco de dados do Passo 0.
Pontuação máxima: 29 pts | Valor por pergunta: 5 a 7 pts
Subseção aplicável quando o protocolo utiliza banco de dados (filtro do Passo 0). Deriva da Resolução CNS n.º 738/2024 e complementa o Bloco 6. A pontuação soma-se à do Bloco 6 para o cálculo final; a ausência de resposta positiva em itens marcados como eliminatórios sujeita o protocolo à regra crítica (Caderno Referencial, Capítulo 8, item "A regra crítica").
Subtotal Bloco 6.b: _____ / 29 pts (soma-se ao Bloco 6)

Hipótese eliminatória: resposta "sim ao risco" em P6.b.2 (Termo de Anuência Institucional ausente em banco fora do âmbito da pesquisa) torna o protocolo não avaliável no mérito, conforme a regra crítica (Caderno Referencial, Capítulo 8, item "A regra crítica"), independentemente da pontuação total.

BLOCO 7: MITIGAÇÃO
Pontuação máxima de risco: 75 pts | Perguntas de risco somam | Perguntas de mitigação subtraem
Avalia a qualidade dos mecanismos de supervisão humana, explicabilidade e controle do ciclo de vida. Único bloco com lógica bidirecional.
Sub-bloco 7A: Consultas (até +10 pts de risco)
Sub-bloco 7B: Medidas de redução de risco (até +65 pts de risco / até −65 pts por mitigações)
Sub-bloco 7C: Evidências de transparência (até −23 pts, só-abate: a presença documentada da evidência subtrai pontos, e a ausência não soma)
Subtotal Bloco 7: _____ pts (positivo se faltas de mitigação superam as medidas adotadas).

VERIFICAÇÃO TRANSVERSAL DO CICLO DE VIDA — VERSÃO B

Antes da consolidação final, recomenda-se que o avaliador aplique as oito perguntas do Módulo Transversal da Versão A. Os achados devem ser registrados nos Blocos 2, 3, 4, 5 ou 7, conforme a dimensão afetada. O módulo não cria pontuação independente nesta versão, evitando alteração das faixas quantitativas sem calibração empírica; contudo, falhas estruturais de supervisão, gestão de mudanças, rastreabilidade ou monitoramento podem acionar a regra crítica quando comprometerem a proteção das pessoas participantes.

CONSOLIDAÇÃO FINAL

Nível de risco:
Classificação por faixa (teto 275): Nível I — 0 a 58 pts; Nível II — 59 a 127 pts; Nível III — 128 a 208 pts; Nível IV — 209 a 275 pts. Com o Bloco 6.b ativado, aplicam-se as faixas recalibradas (teto 304; teto avaliável 297): Nível I — 0 a 64 pts; Nível II — 65 a 141 pts; Nível III — 142 a 230 pts; Nível IV — 231 a 304 pts.

Regra crítica: o nível determinado pela pontuação não muda por decisão do avaliador. As mitigações já estão incorporadas na pontuação do Bloco 7 e constituem a única forma legítima de reduzir a pontuação final. Exceção — Cláusula de Prevalência Ética: a marcação afirmativa nas perguntas P4.1 ou P4.2 (decisão sem revisão humana ou dano irreversível ao participante) aciona o princípio da precaução, facultando ao CEP a elevação imediata do protocolo ao Nível IV, independentemente da pontuação total. O resultado final após subtração das mitigações do Bloco 7 não pode ser inferior a zero pontos.

Recalibração quando o Bloco 6.b está ativado: quando o filtro do Passo 0 aciona a subseção Bloco 6.b (banco de dados), a pontuação máxima da matriz passa de 275 para 304 pontos, com as faixas recalibradas proporcionalmente para preservar a distância relativa entre níveis: Nível I de 0 a 64 pts; Nível II de 65 a 141 pts; Nível III de 142 a 230 pts; Nível IV de 231 a 304 pts (teto avaliável 297 pts, descontada a P6.b.2 eliminatória). Fora desse caso, vigoram as faixas-padrão acima (209 a 275 pts para o Nível IV).

6 PARTE V — INSTRUÇÕES DE USO

Instruções de Uso da MARIAH

INSTRUÇÕES DE USO

MARIAH — MATRIZ DE AVALIAÇÃO DE RISCO EM INTELIGÊNCIA ARTIFICIAL

Antes de começar: qual versão usar?

A escolha entre a Versão A qualitativa e a Versão B quantitativa da Matriz de Avaliação de Risco em Inteligência Artificial em Pesquisa com Seres Humanos (MARIAH) não é escolha entre instrumento rigoroso e instrumento simplificado. As duas versões partem da mesma base ética, percorrem os mesmos eixos de avaliação e chegam ao mesmo conjunto de exigências por nível de risco. O que as diferencia é a lógica de operação: a Versão A orienta o raciocínio por fluxograma sequencial; a Versão B orienta a rastreabilidade por pontuação ponderada. Um CEP que escolhe a Versão A não faz escolha menos rigorosa. Faz a escolha adequada ao seu momento institucional e ao perfil dos protocolos que avalia.

Três perguntas orientam essa escolha. A primeira é sobre experiência institucional: o CEP tem ciclos consolidados de avaliação de protocolos com IA, com membros familiarizados com os conceitos do Capítulo 2 do Caderno Referencial e com a lógica dos cinco eixos, ou está nos primeiros ciclos de familiarização com esse tipo de protocolo? A Versão A é o ponto de entrada recomendado para comitês que estão construindo essa familiaridade. A segunda é sobre a complexidade do protocolo: o sistema de IA descrito envolve múltiplas arquiteturas, aprendizado contínuo, uso simultâneo em diferentes fases da pesquisa ou integração com outros sistemas, ou trata-se de uso mais circunscrito com sistema claramente descrito? Protocolos de alta complexidade técnica se beneficiam da granularidade da Versão B, que permite identificar onde o risco está concentrado e como as mitigações o reduziram. A terceira é sobre as necessidades institucionais de documentação: a instituição precisa comparar protocolos entre si, produzir relatórios auditáveis sobre as decisões do comitê ou demonstrar para patrocinadores e reguladores a base numérica das aprovações e recusas? A Versão B foi desenhada para esse nível de rastreabilidade. O Quadro 8 compara as duas versões.

Quadro 8 — Comparação entre as versões qualitativa e quantitativa da MARIAH

	Versão A Qualitativa	**Versão B Quantitativa**
Perfil do CEP	Primeiros ciclos com IA ou protocolos de menor complexidade	Experiência consolidada ou protocolos de alta complexidade
Lógica de operação	Fluxograma sequencial, resposta binária por eixo	Pontuação ponderada em sete blocos
Resultado	Nível de risco por combinação de respostas	Nível de risco por soma total com rastreabilidade numérica
Uso prioritário	Deliberação colegiada orientada ao raciocínio	Documentação auditável e comparação entre protocolos

Fonte: elaboração própria.

As duas versões são compatíveis e podem ser usadas em combinação. Um CEP pode adotar a Versão A como triagem inicial e recorrer à Versão B para aprofundar a análise de protocolos que a Versão A classifique nos Níveis III ou IV. Pode, também, começar com a Versão A e migrar progressivamente para a Versão B à medida que os membros desenvolvem familiaridade com o instrumento. A escolha não é irreversível. O instrumento foi desenhado para acompanhar o desenvolvimento institucional do comitê, não para fixá-lo num nível de capacidade.

Ambas as versões incorporam, ainda, uma subseção específica para protocolos que utilizam bancos de dados — o Eixo 3.b na Versão A e o Bloco 6.b na Versão B. Essa subseção deriva da Resolução CNS n.º 738/2024 e traduz em perguntas operacionais as exigências sobre Controlador, Operador, Termos institucionais (Acordo, Anuência, Compromisso e Transferência) e dispensa de TCLE para uso futuro. A subseção é ativada pelo filtro de banco de dados no Passo 0 e não se aplica quando o protocolo coleta dados exclusivamente no escopo da pesquisa com cada participante.

Aplicando a Versão A qualitativa: passo a passo

O Passo 0 é a decisão que determina se a Versão A se aplica ao protocolo em avaliação. A pergunta é precisa e deve ser respondida antes de qualquer outra leitura do instrumento: o sistema de IA realiza automação de decisão, geração de conteúdo ou intervenção no protocolo ou na condução do estudo? Três tipos de uso acionam a matriz. O primeiro é a automação de decisão: qualquer sistema que produza classificações, escores, recomendações

ou alertas que influenciem decisões sobre participantes ou sobre a condução do estudo. O segundo é a geração de conteúdo: sistemas que produzem texto, dados sintéticos, imagens ou qualquer outro conteúdo que integre o protocolo, o TCLE, os instrumentos de coleta ou os resultados da pesquisa. O terceiro é a intervenção: sistemas cuja saída orienta, modifica ou substitui etapas do protocolo. Uma ferramenta de organização bibliográfica não aciona a matriz. Um modelo de apoio diagnóstico aciona. Um gerador de dados sintéticos aciona. Um sistema de processamento de linguagem natural que analisa prontuários para identificar participantes elegíveis aciona. A dúvida sobre onde traçar essa linha não é sinal de fragilidade do avaliador. É sinal de que o protocolo deve ser levado ao plenário antes de qualquer decisão sobre a aplicação da matriz.

Duas dúvidas recorrentes merecem resposta explícita. A primeira: o uso de sistemas generativos no aprimoramento e na elaboração do texto do projeto aciona a matriz? Declaração e acionamento são obrigações distintas. A declaração do uso de IA generativa é sempre obrigatória, por força da Portaria CNPq n.º 2.664/2026, qualquer que seja a finalidade do uso. A matriz é acionada quando o conteúdo gerado integra o desenho da pesquisa, os instrumentos de coleta, o TCLE ou os resultados — porque nesses casos o sistema participa da produção do conhecimento e pode introduzir viés, imprecisão ou conteúdo alucinado no que será submetido ao CEP e no que será oferecido ao participante. A assistência estritamente redacional — revisão gramatical, adequação de estilo, tradução, formatação de referências — é declarada, mas não aciona a matriz, porque não altera o desenho nem produz achado. O critério de discernimento é funcional: pergunta-se se, removido o sistema, o conteúdo submetido seria substantivamente o mesmo. A segunda dúvida: onde se enquadra o sistema que identifica padrões de prognóstico de uma doença? Em automação de decisão, porque produz classificações e escores sobre participantes — e isso vale ainda que o desfecho seja apenas de pesquisa, sem conduta clínica associada, uma vez que a categoria alcança as decisões sobre a condução do estudo, e não apenas as decisões clínicas. Quando o mesmo escore também orienta estratificação, elegibilidade ou conduta terapêutica, acumula-se a hipótese de intervenção, e o protocolo é avaliado por ambas.

Um segundo filtro opera no Passo 0 em paralelo ao primeiro. O CEP verifica se o protocolo utiliza banco de dados — próprio, de outra pesquisa ou de instituição externa. A resposta afirmativa ativa o Eixo 3.b, subseção derivada da Resolução CNS n.º 738/2024 que examina cinco perguntas operacionais: identificação do Controlador do banco; Termo de

Anuência Institucional em bancos constituídos fora do âmbito da pesquisa (item eliminatório); Termo de Compromisso de Uso de Dados; enquadramento de eventual dispensa de TCLE na hipótese cabível segundo a origem do banco; e formalização da controladoria conjunta em pesquisas multicêntricas por Termo de Acordo Institucional. A resposta negativa deixa a subseção inaplicável e a avaliação segue com os cinco eixos principais.

A caracterização do contexto de uso é o passo mais negligenciado e o mais consequente de toda a Versão A. Sem ela, os cinco eixos produzem leituras descontextualizadas que podem classificar sistemas distintos no mesmo nível por razões incomparáveis. Antes de percorrer qualquer eixo, o CEP precisa responder a duas perguntas sobre o sistema: qual é a pergunta que ele foi desenvolvido para responder neste protocolo, e a resposta é produzida de forma autônoma ou subsidia uma decisão tomada por um ser humano?

A resposta a essas duas perguntas situa o sistema no seu território ético específico e orienta o percurso pelos eixos seguintes. O percurso pelos cinco eixos apresenta situações em que a resposta binária não é óbvia. Quando o pesquisador descreve o sistema como supervisionado, mas não descreve o mecanismo de supervisão, a resposta correta é de risco: supervisão declarada sem mecanismo descrito é supervisão não verificável. Quando o protocolo menciona validação, mas não documenta testes de desempenho por subgrupo populacional, a resposta correta é de risco: validação global sem análise estratificada não garante equidade. Quando o TCLE menciona IA, mas em linguagem técnica que o participante médio não compreenderia, a resposta correta é de risco: menção sem compreensibilidade não satisfaz o requisito de consentimento informado.

O resultado da Versão A é um nível de risco determinado pela combinação das respostas nos cinco eixos, com a regra crítica aplicada: um único eixo classificado em Nível IV eleva o protocolo inteiro ao Nível IV, independentemente dos demais eixos. Cada nível determina um conjunto de requisitos mínimos cumulativos que o pesquisador deve demonstrar ter atendido antes da aprovação. No Nível I, exige-se declaração de uso de IA e menção no TCLE. No Nível II, acrescentam-se a descrição do sistema e a demonstração de conformidade com a LGPD. No Nível III, acrescentam-se documentação de validação técnica, análise de desempenho por subgrupos e comprovação de supervisão humana efetiva. No Nível IV, acrescentam-se parecer técnico externo, plano de monitoramento contínuo com critérios de interrupção e, quando aplicável, notificação à Anvisa.

O CEP traduz esse resultado em linguagem de parecer com precisão: o protocolo foi classificado no Nível X da Matriz de Avaliação de Risco em Inteligência Artificial, Versão A, e não satisfaz os requisitos mínimos do nível nas seguintes dimensões, com indicação explícita de quais requisitos estão ausentes e qual o prazo para complementação. Quando o protocolo apresenta falha estrutural que nenhuma complementação resolve, a recusa é fundamentada no nível de risco identificado e nos princípios da Lei n.º 14.874/2024. A regra crítica não é punição, é o princípio de precaução materializado no instrumento: um sistema que falha numa dimensão estrutural não se torna ético por ter desempenho adequado nas demais.

A regra crítica incorpora ainda hipóteses eliminatórias específicas para protocolos com bancos de dados. Quando o protocolo utiliza banco constituído fora do âmbito da pesquisa sem apresentar Termo de Anuência Institucional (art. 27, VI da Res. CNS n.º 738/2024), a pontuação total é desconsiderada e o protocolo não é avaliável no mérito. A mesma lógica se aplica à ausência de identificação do Controlador (Arts. 3º, IV e 9º) e a pedidos de dispensa de TCLE sem indicação da hipótese normativa aplicável e sem fundamentação ética e metodológica suficiente (art. 20, § 5.º, ou art. 25). Nesses casos, o dossiê é devolvido ao pesquisador para diligência obrigatória antes de qualquer análise de mérito — não se trata de agravamento de nível de risco, mas de bloqueio de avaliação por ausência de cadeia de custódia formalizada. A regra crítica integra um conjunto mais amplo. A MARIAH opera com quatro salvaguardas automáticas, detalhadas no Suplemento à Nota Técnica de Premissas: a regra especial do Eixo 3.b, a Cláusula de Prevalência Ética, a eliminatória de cadeia de custódia e a diligência impeditiva de novo consentimento. A regra especial do Eixo 3.b salta deliberadamente o Nível II. Uma única resposta de risco nesse eixo leva o protocolo a pelo menos o Nível III; três ou mais respostas levam ao Nível IV. A diligência impeditiva alcança o sistema adaptativo sem plano de novo consentimento, verificado na pergunta 1.2 da Versão A e nas perguntas P2.2 e P2.8 da Versão B. Nesse caso o protocolo também é devolvido como não avaliável, em diligência exigida pela Lei n.º 14.874/2024 e pela LGPD.

Aplicando a Versão B quantitativa: passo a passo

A Versão B opera sobre a mesma base ética e os mesmos eixos da Versão A, mas com lógica distinta: pontuação ponderada em sete blocos, soma total que determina o nível de risco e capacidade de identificar onde o risco está concentrado. Antes de iniciar a pontuação,

o Passo 0 é idêntico ao da Versão A: o sistema realiza automação de decisão, geração de conteúdo ou intervenção no protocolo? Se não, a matriz não se aplica. Se sim, o CEP prossegue para a caracterização do contexto de uso, que na Versão B é obrigatória e não pontua, mas condiciona toda a pontuação subsequente. As oito perguntas descritivas — qual é a pergunta que o sistema foi desenvolvido para responder, e se a resposta é autônoma ou subsidiária — precisam estar respondidas antes do primeiro bloco.

Os sete blocos percorrem dimensões distintas com pesos distintos. O Bloco 1 avalia o projeto: transparência do protocolo, qualificação da equipe e clareza da proposta, com pontuação máxima de 22 pontos. O Bloco 2 avalia o sistema: autonomia, adaptabilidade e rastreabilidade, com pontuação máxima de 17 pontos. O Bloco 3 avalia o ciclo de vida: estágio de desenvolvimento, monitoramento e plano de retreinamento, com pontuação máxima de 14 pontos. O Bloco 4 avalia a natureza da decisão apoiada: reversibilidade, autonomia decisória e consequência clínica, com pontuação máxima de 8 pontos. O Bloco 5 avalia o impacto sobre a pessoa participante, com pontuação máxima de 62 pontos. O Bloco 6 avalia a sensibilidade e a governança dos dados, com pontuação máxima de 77 pontos. O Bloco 7 avalia as mitigações, com lógica bidirecional: o sub-bloco 7A de consultas soma até dez pontos e as mitigações, que valem 65 pontos, somam quando ausentes e subtraem quando presentes, com teto de 75 pontos para o bloco. A soma dos sete blocos produz o teto de 275 pontos da matriz. Os Blocos 5 e 6 concentram juntos 139 dos 275 pontos máximos (50,5% da pontuação total). Essa concentração expressa a hierarquia de prioridades do instrumento: o que mais importa em pesquisa com seres humanos é o que o sistema pode fazer à pessoa participante e como seus dados são tratados.

Quando o protocolo utiliza banco de dados, um oitavo conjunto de itens é ativado: o Bloco 6.b, com pontuação máxima de 29 pontos, somada ao Bloco 6. Os cinco itens do Bloco 6.b derivam da Resolução CNS n.º 738/2024 e avaliam identificação do Controlador (sete pontos), Termo de Anuência Institucional (sete pontos, eliminatório), Termo de Compromisso de Uso de Dados (cinco pontos), enquadramento de dispensa de TCLE (cinco pontos) e controladoria conjunta formalizada (cinco pontos). A subseção integra ainda três itens de diligência (P6.b.4.1, P6.b.6 e P6.b.7), que não pontuam e operam como *gate* cumulativo: a ausência de qualquer critério implica devolução do protocolo. O filtro de ativação é o mesmo do Passo 0: protocolos que não utilizam bancos de dados têm o Bloco 6.b marcado como não aplicável.

O Bloco 7 merece atenção específica porque opera de forma distinta dos demais. É o único bloco com lógica bidirecional. As perguntas de risco somam quando o pesquisador não adotou medidas de supervisão, explicabilidade ou monitoramento adequadas. As perguntas de mitigação subtraem quando essas medidas estão documentadas e verificáveis. A subtração não é automática: depende de que as medidas estejam descritas no protocolo com precisão suficiente para que o CEP as avalie como reais, não como declarações de intenção. Um protocolo que afirma que o sistema é supervisionado sem descrever o mecanismo de supervisão não obtém a subtração correspondente. A pontuação do Bloco 7 é a única que o pesquisador pode influenciar diretamente com as escolhas de desenho do estudo. É o bloco que recompensa protocolos bem construídos e penaliza protocolos que transferem o risco ao participante sem contramedidas documentadas.

A soma dos sete blocos produz a classificação em um dos quatro níveis. O Nível I vai de 0 a 58 pontos. O Nível II vai de 59 a 127 pontos. O Nível III vai de 128 a 208 pontos. O Nível IV vai de 209 a 275 pontos. Quando o filtro do Passo 0 ativa a subseção Bloco 6.b (bancos de dados), o teto da matriz passa a 304 pontos e as faixas são recalibradas proporcionalmente: Nível I de 0 a 64 pts; Nível II de 65 a 141 pts; Nível III de 142 a 230 pts; Nível IV de 231 a 304 pts (teto avaliável 297 pts, descontada a P6.b.2 eliminatória). Cada nível determina os mesmos requisitos mínimos cumulativos descritos no item "Aplicando a Versão A qualitativa: passo a passo". A diferença em relação à Versão A é que a Versão B permite ao CEP identificar em quais blocos o risco está concentrado e comunicar essa informação ao pesquisador com precisão. Um protocolo classificado no Nível III com concentração de pontuação nos Blocos 5 e 6 recebe solicitação de complementação diferente de um protocolo classificado no mesmo nível com concentração no Bloco 2. A granularidade da pontuação torna o parecer mais fundamentado e o pesquisador mais capaz de compreender o que precisa demonstrar para obter aprovação.

A regra crítica da Versão B é idêntica em princípio à da Versão A, mas opera por pontuação: a classificação final não pode ser alterada por decisão subjetiva do avaliador. As mitigações já estão incorporadas na pontuação do Bloco 7 e constituem a única forma legítima de reduzir a pontuação final. Declarações de intenção não documentadas não substituem evidências de controle efetivo do risco. O CEP que aplica a Versão B e obtém classificação no Nível IV não pode aprovar o protocolo com base em argumentos que a pontuação não reflete. Pode solicitar complementação que permita ao pesquisador reduzir a pontuação por meio de

mitigações reais. Pode recusar com fundamentação baseada na pontuação e nos requisitos do nível. Não pode ignorar o resultado do instrumento que escolheu aplicar.

Documentando a decisão

Aplicar a MARIAH sem registrar o resultado no processo de avaliação é aplicar o instrumento pela metade. O registro transforma a deliberação do CEP em jurisprudência institucional: o protocolo avaliado hoje forma o repertório que orienta a avaliação do protocolo semelhante que chegará no próximo semestre. Um comitê que não documenta suas decisões sobre protocolos com IA recomeça do zero a cada avaliação.

O registro mínimo obrigatório em qualquer parecer que utilizou a MARIAH inclui quatro elementos. O primeiro é a identificação da versão utilizada: Versão A Qualitativa ou Versão B Quantitativa, com indicação de eventual uso combinado. O segundo é o resultado da aplicação: o nível de risco obtido em cada eixo, na Versão A, ou a pontuação por bloco e a pontuação total, na Versão B. O terceiro é a justificativa para decisões não óbvias: quando uma resposta de risco foi atribuída a uma dimensão onde o protocolo apresentava informação ambígua, o parecer deve registrar o raciocínio que fundamentou essa atribuição. O quarto é a tradução do nível em requisitos: quais requisitos mínimos do nível obtido o protocolo satisfez e quais estão ausentes, com indicação precisa do que o pesquisador deve complementar e em qual prazo.

Duas situações específicas exigem documentação adicional. A primeira é quando o CEP decide aprovar um protocolo com ressalvas num nível que normalmente exigiria recusa ou complementação substancial. Essa decisão precisa de fundamentação explícita que demonstre por que as circunstâncias específicas do protocolo justificam a exceção e quais salvaguardas adicionais foram exigidas como condição da aprovação. A segunda é quando membros do comitê divergiram na aplicação da matriz sobre o mesmo protocolo. A divergência deve ser registrada com as posições e os argumentos de cada lado, mesmo que a decisão final seja unânime. Divergências documentadas são o material mais rico para o desenvolvimento da competência institucional do CEP: revelam onde o instrumento precisa de orientação adicional e onde o comitê precisa de deliberação mais aprofundada.

Quando o resultado da matriz fundamenta uma recusa, o parecer deve citar explicitamente o nível obtido, os requisitos não atendidos e os artigos da Lei n.º 14.874/2024

e do Decreto n.º 12.651/2025 que sustentam a exigência não satisfeita. A recusa fundamentada na matriz não é decisão técnica do instrumento. É decisão ética do comitê, apoiada no instrumento. Essa distinção importa para o pesquisador que eventualmente recorra da decisão e para a instância que aprecia o recurso. O CEP não recusa porque a pontuação foi alta,recusa porque o protocolo não demonstrou as garantias que a proteção dos participantes exige naquele nível de risco, conforme os princípios que a legislação brasileira estabelece e que a matriz operacionaliza.

A Instrução de Uso da MARIAH fecha com uma nota sobre o que o instrumento não faz. Ela não aprova nem reprova protocolos. Não substitui o julgamento do CEP. Não dispensa a deliberação colegiada. Não resolve os casos limítrofes. O que faz é tornar a avaliação mais sistemática, mais documentada e mais justificável. Torna visível o que o entusiasmo científico tende a obscurecer e o que a assimetria de conhecimento técnico tende a suprimir. O comitê que a utiliza com rigor não delega sua responsabilidade ao instrumento,exerce essa responsabilidade com mais precisão.

7 PARTE VI — VALIDAÇÃO LOCAL E REVISÃO

Validação local da MARIAH

Por que validar a MARIAH na sua casuística

O Capítulo 8 do Caderno Referencial é explícito ao posicionar a MARIAH como instrumento de apoio à deliberação dos CEPs, e não como norma que vincule sua decisão. A adoção da matriz é escolha institucional de cada Comitê. Esse mesmo espírito orienta a validação: nenhum CEP é obrigado a validar a MARIAH, e nem precisa esperar decisão central para fazê-lo. Quando o colegiado entender que vale a pena verificar como a matriz se comporta nos protocolos que efetivamente lhe são submetidos, esta Parte oferece um roteiro prático para conduzir essa verificação.

Há três razões pelas quais um CEP pode querer fazer essa validação. A primeira é o aprendizado institucional. A aplicação sistemática da matriz por dois ou mais membros do colegiado revela rapidamente quais perguntas recebem interpretação divergente, o que oferece insumo direto para a capacitação deliberativa do colegiado. A segunda diz respeito à qualificação institucional do parecer. Um CEP que demonstre, com base na sua própria casuística, que aplica a MARIAH de modo consistente entre avaliadores, fortalece a credibilidade dos seus pareceres em protocolos com inteligência artificial perante pesquisadores, patrocinadores e a sociedade. A terceira é de contribuição opcional à evolução do instrumento. Os pontos de corte atuais da Versão B são, conforme o próprio Capítulo 8 do Caderno Referencial reconhece, calibrações teóricas que merecem ser testadas empiricamente. CEPs que queiram compartilhar suas observações com o Grupo de Trabalho responsável pelo guia podem fazê-lo pelo canal descrito no item “Quando e como compartilhar resultados com o Grupo de Trabalho”.

Vale, antes do roteiro, uma advertência. Esta Parte não pretende transformar membros de CEP em estatísticos. As métricas propostas no item “Como interpretar os achados” são as mínimas necessárias para que o colegiado leia os próprios resultados com segurança, e a planilha-modelo que acompanha esta Seção calcula essas métricas automaticamente. O que se espera do CEP é a deliberação colegiada sobre o que os números mostram, não o cálculo dos números em si.

O que pode ser validado em três frentes

Esta Parte propõe três frentes independentes de validação. Cada CEP pode realizar uma, duas ou as três, em qualquer ordem. As três compartilham a mesma unidade amostral: o protocolo de pesquisa com uso de inteligência artificial submetido ao CEP.

Concordância entre avaliadores na Versão A

A primeira frente verifica se dois ou mais integrantes do CEP, ao aplicarem a Versão A da MARIAH ao mesmo protocolo de forma independente, chegam à mesma classificação. É a frente operacionalmente mais simples e a que rende, do ponto de vista de aprendizado institucional, o retorno mais imediato. O colegiado descobre rapidamente em quais eixos há leitura convergente e em quais há discordância sistemática, o que orienta diretamente o que precisa ser conversado em reunião plenária.

A recomendação é de pelo menos vinte protocolos avaliados em paralelo por dois membros do CEP, em condições de cegamento mútuo, antes de qualquer deliberação colegiada. Vinte é um piso prudente para que a estimativa de concordância tenha estabilidade mínima; CEPs com volume maior de protocolos com IA podem ampliar essa marca sem prejuízo, e CEPs com volume menor podem acumular avaliações ao longo do tempo, sem prazo, até atingir essa marca.

Distribuição empírica dos protocolos nos quatro níveis

A segunda frente examina como os protocolos avaliados pelo CEP se distribuem nos quatro níveis da MARIAH. Há dois usos principais para essa informação. O primeiro é descritivo. O CEP passa a conhecer o perfil da sua casuística em pesquisa com inteligência artificial: se a maioria dos protocolos se concentra no Nível II, se há protocolos no Nível IV, se a distribuição mudou ao longo do tempo. Essa informação é insumo de planejamento institucional e de capacitação do colegiado. O segundo uso é diagnóstico. Quando a distribuição se concentra perto de um ponto de corte da Versão B, o colegiado tem evidência local de que pequenas variações na pontuação podem deslocar muitos protocolos de nível, o que justifica examinar com mais atenção os casos limítrofes.

A recomendação operacional é registrar a pontuação total e a pontuação por bloco de todos os protocolos aos quais o CEP aplicar a Versão B, sem critério de inclusão específico. A planilha-modelo produz automaticamente o histograma e identifica concentrações próximas aos pontos de corte.

Convergência entre Versões A e B no modo triagem

A terceira frente examina se a Versão A e a Versão B, aplicadas ao mesmo protocolo no modo triagem previsto na Parte V desta MARIAH, convergem para a mesma classificação de risco. Quando o CEP usa rotineiramente o modo triagem (Versão A como filtro inicial e Versão B em casos de Nível III ou Nível IV) basta registrar as duas classificações em paralelo. Discordâncias sistemáticas entre as duas versões, quando ocorrem, são informação valiosa: ou indicam que o CEP lê de modo inconsistente as duas versões, ou indicam que as duas versões captam dimensões diferentes do risco no protocolo. Em ambos os casos, o colegiado tem material concreto para discutir e, se for o caso, reportar ao Grupo de Trabalho.

Como conduzir cada frente

O roteiro a seguir cobre as três frentes em um único fluxo. Cada CEP adapta o fluxo à sua dinâmica.

O primeiro passo é definir o período de coleta. Pode ser um trimestre, um semestre ou aberto. O importante é o CEP saber que, durante esse período, os protocolos com inteligência artificial recebidos serão objeto também do exercício de validação, em paralelo ao parecer ordinário. A coleta não altera o fluxo do parecer institucional do SINEP.

O segundo passo é constituir uma dupla ou um trio de avaliadores. A recomendação é que pelo menos dois membros do colegiado apliquem a Versão A de forma independente, sem consulta prévia entre si, e que um terceiro membro aplique a Versão B em separado. Avaliadores podem se revezar a cada protocolo; o que importa é a independência da aplicação dentro de cada protocolo. Não é necessário que o avaliador da Versão B seja o relator do parecer institucional.

O terceiro passo é registrar as avaliações na planilha-modelo. A planilha tem aba específica para cada frente. Cada protocolo recebe um identificador interno do CEP que

dispensa qualquer informação que permita identificar pesquisador, instituição proponente ou patrocinador. A planilha não pede esses dados. O registro mínimo é o identificador, a data e as respostas dos avaliadores nas perguntas da matriz. As fórmulas calculam a classificação consolidada, a pontuação total e os indicadores agregados das três frentes.

O quarto passo é a deliberação colegiada sobre os achados. Recomenda-se reunião dedicada, ao final do período de coleta ou periodicamente durante ele, para que o colegiado discuta o que os indicadores mostram. A planilha gera um painel-resumo que apoia essa discussão. A deliberação é o ponto central do exercício; os números são insumo, não conclusão.

Como interpretar os achados

A interpretação dos achados não exige conhecimento estatístico avançado. Para as três frentes propostas, três leituras básicas bastam.

Para a frente de concordância entre avaliadores, o indicador principal é a estatística *kappa* (Cohen, 1960), calculada automaticamente pela planilha. O Quadro 9, a seguir, oferece a leitura prática dos valores observados. O colegiado lê o valor de *kappa* por eixo da Versão A e por classificação consolidada, identifica os eixos com *kappa* baixo e discute, em plenária, os casos de discordância naqueles eixos.

Quadro 9 — Leitura prática dos valores de kappa

Valor de *kappa*	Leitura prática
Abaixo de 0,40	Concordância insuficiente. O eixo ou bloco merece revisão pelo colegiado: as perguntas têm leitura divergente, e isso é informação valiosa para discutir em plenária.
Entre 0,40 e 0,60	Concordância moderada. Identifique os casos de discordância e use-os como pauta de reunião, porque são exatamente os casos limítrofes que mais contribuem para a formação do comitê.
Entre 0,60 e 0,80	Concordância substancial. A aplicação está consistente. Pequenas discordâncias residuais são esperadas e saudáveis.
0,80 ou acima	Concordância quase perfeita. A aplicação está estabilizada. Cuidado, porém: kappas muito altos em todos os eixos podem indicar que os avaliadores deliberam entre si antes de registrar as respostas, o que esvazia o exercício.

Fonte: Adaptado de Landis e Koch (1977).

Para a frente de distribuição empírica, a leitura é visual. O colegiado analisa o histograma da planilha para verificar a concentração de protocolos em algum nível, a proximidade com os pontos de corte e a presença de bimodalidade (dois picos separados por um vale). Concentração próxima aos pontos de corte é alerta de que pequenas variações na aplicação da matriz produzem mudanças de nível. Bimodalidade marcada é evidência local de que o ponto de corte natural na casuística do CEP pode diferir do ponto teórico vigente.

A leitura é tabular para a frente de convergência entre versões. A planilha gera uma tabela de contingência quatro por quatro que cruza as classificações da Versão A e da Versão B no modo triagem. Alta convergência ocorre quando a diagonal principal concentra mais de oitenta por cento dos protocolos. Padrões de discordância, que exigem exame qualitativo dos protocolos correspondentes, surgem quando há concentração em pares específicos fora da diagonal, como Nível II na Versão A e Nível III na Versão B.

A planilha não impõe interpretação ao colegiado. Os indicadores são apresentados com leitura sugerida, mas a deliberação sobre o que fazer com eles é decisão institucional do CEP.

Quando e como compartilhar resultados com o Grupo de Trabalho

CEPs que decidirem compartilhar suas observações com o Grupo de Trabalho responsável pelo guia podem fazê-lo a qualquer momento por meio da Secretaria Executiva da INAEP (se.inaep@saude.gov.br). Não há prazo, não há formulário obrigatório, não há contrapartida institucional. O compartilhamento é ato voluntário, e seu único propósito é informar a evolução futura do instrumento.

O canal sugerido é o envio da planilha-modelo, preenchida e com identificação institucional do CEP, para o endereço eletrônico do Grupo de Trabalho indicado na página oficial do guia. O envio pode ser feito em qualquer formato que o CEP preferir: a planilha completa, um resumo dos indicadores agregados ou apenas uma nota narrativa que descreva o que o colegiado observou. CEPs que tenham preocupação com a confidencialidade de informações institucionais podem encaminhar apenas os indicadores agregados, sem os registros por protocolo.

O Grupo de Trabalho consolidará as contribuições recebidas e as utilizará, em conjunto com outras fontes de evidência, para informar eventuais revisões da matriz em ciclos futuros de atualização do guia. As contribuições não terão atribuição nominal sem autorização expressa do CEP remetente, e nenhum CEP será identificado em publicação derivada sem seu consentimento institucional.

Recursos disponíveis

Acompanha esta Parte uma planilha-modelo em formato .xlsx, com abas dedicadas a cada uma das três frentes de validação, fórmulas pré-configuradas para cálculo dos indicadores e painel-resumo com leitura sugerida. A planilha pode ser usada em qualquer software compatível, como Microsoft Excel, LibreOffice Calc ou Google Sheets, e não requer instalação de qualquer extensão estatística.

O software MARIAH, ferramenta eletrônica de acesso livre disponibilizada em https://mariah-inaep.vercel.app, oferece uma camada operacional complementar: a aplicação das matrizes diretamente em formato digital, com exportação de relatórios padronizados. CEPs que prefiram operar o exercício de validação em ambiente digital podem fazê-lo por meio da exportação dos relatórios e da transferência dos dados para a planilha-modelo desta Parte. O repositório oficial do software fica em github.com/Tonaco-13.

Dúvidas operacionais sobre a aplicação deste roteiro podem ser encaminhadas, junto com observações e sugestões, pelo canal indicado no item “Quando e como compartilhar resultados com o Grupo de Trabalho”.

REFERÊNCIAS

AUSTRÁLIA. National Health and Medical Research Council (NHMRC). Guide for assessing research involving AI. Canberra: NHMRC, 2023. Disponível em: https://www.nhmrc.gov.au/about-us/resources/guide-assessing-research-involving-ai. Acesso em: 10 mar. 2026.

BRASIL. Conselho Nacional de Desenvolvimento Científico e Tecnológico. Portaria n.º 2.664, de 6 de março de 2026. Institui a Política de Integridade na Atividade Científica do CNPq. Diário Oficial da União: seção 1, Brasília, DF, p. 4, 11 mar. 2026.

BRASIL. Conselho Nacional de Saúde. Resolução n.º 466, de 12 de dezembro de 2012. Aprova as diretrizes e normas regulamentadoras de pesquisas envolvendo seres humanos. Brasília, DF: CNS, 2012. Disponível em: https://conselho.saude.gov.br/resolucoes/2012/Reso466.pdf. Acesso em: 19 abr. 2026.

BRASIL. Conselho Nacional de Saúde. Resolução n.º 738, de 1.º de fevereiro de 2024. Dispõe sobre as obrigações éticas na constituição e gestão de repositórios de dados para pesquisa científica. Brasília, DF: Diário Oficial da União, 2024b.

BRASIL. Presidência da República. Decreto n.º 12.651, de 7 de outubro de 2025. Regulamenta a Lei nº 14.874, de 28 de maio de 2024, que dispõe sobre a pesquisa com seres humanos e institui o Sistema Nacional de Ética em Pesquisa com Seres Humanos. Diário Oficial da União: seção 1, Brasília, DF, edição 192, p. 3, 8 out. 2025. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2025/decreto/d12651.htm. Acesso em: 19 jun. 2026.

BRASIL. Presidência da República. Lei n.º 14.874, de 28 de maio de 2024. Dispõe sobre a pesquisa com seres humanos e institui o Sistema Nacional de Ética em Pesquisa com Seres Humanos. Diário Oficial da União: seção 1, Brasília, DF, n. 103, p. 6, 29 maio 2024a. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2024/lei/l14874.htm. Acesso em: 23 mar. 2026.

CANADÁ. Treasury Board of Canada Secretariat. Algorithmic impact assessment (AIA). [Ottawa: Government of Canada], 2025. Disponível em: https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html. Acesso em: 10 mar. 2026.

CARNEGIE ENDOWMENT FOR INTERNATIONAL PEACE. Understanding India's new data protection law. Washington, DC: Carnegie Endowment, 2023. Disponível em: https://carnegieendowment.org/research/2023/10/understanding-indias-new-data-protection-law. Acesso em: 11 mar. 2026.

COHEN, J. A coefficient of agreement for nominal scales. Educational and Psychological Measurement, v. 20, n. 1, p. 37–46, 1960. DOI: https://doi.org/10.1177/001316446002000010.

COREIA DO SUL. Ministry of Food and Drug Safety (MFDS). Guidance on clinical trials design of artificial intelligence (AI)-based medical devices. Seoul: MFDS, 2023. Disponível em:

https://www.elendilabs.com/en/articles/kor-guidance-on-clinical-trials-design-of-artificial-intelligence-based-medical-devices. Acesso em: 11 mar. 2026.

ESTADOS UNIDOS. Food and Drug Administration (FDA). Considerations for the use of artificial intelligence to support regulatory decision-making for drug and biological products: guidance for industry and other interested parties. Silver Spring: FDA, jan. 2025. Disponível em: https://www.fda.gov/regulatory-information/search-fda-guidance-documents. Acesso em: 10 mar. 2026.

ÍNDIA. Central Drugs Standard Control Organization (CDSCO). Draft guidance on medical device software. New Delhi: CDSCO, out. 2025. Disponível em: https://community.nasscom.in/communities/public-policy/cdsco-issues-draft-guidance-medical-device-software. Acesso em: 11 mar. 2026.

ÍNDIA. Indian Council of Medical Research (ICMR). Ethical guidelines for application of artificial intelligence in biomedical research and healthcare. New Delhi: ICMR, 2023. Disponível em: https://www.icmr.gov.in/ethical-guidelines-for-application-of-artificial-intelligence-in-biomedical-research-and-healthcare. Acesso em: 11 mar. 2026.

JAPÃO. Pharmaceuticals and Medical Devices Agency (PMDA). Report on AI-based software as a medical device (SaMD). Tokyo: PMDA, [s.d.]. Disponível em: https://www.pmda.go.jp/files/000266100.pdf. Acesso em: 11 mar. 2026.

LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. Biometrics, v. 33, n. 1, p. 159–174, 1977.

MRCT CENTER; WCG ARTIFICIAL INTELLIGENCE TASK FORCE. Framework for review of clinical research involving artificial intelligence. [S. l.]: MRCT Center; WCG, June 2025. Disponível em: https://mrctcenter.org/resource/framework-for-review-of-clinical-research-involving-ai/. Acesso em: 23 mar. 2026.

ORGANIZAÇÃO MUNDIAL DA SAÚDE. Ethics and governance of artificial intelligence for health: guidance on large multi-modal models. Genebra: OMS, 2024. Disponível em: https://www.who.int/publications/i/item/9789240084759. Acesso em: 23 mar. 2026.