

MINISTRY OF HEALTH
National Ethics Committee for Research — INAEP

MARIAH
Risk Assessment Matrix in Artificial Intelligence Research with Humans
HUMANS

SUMMARY

1. PRESENTATION OF THE INSTRUMENT.....3

1.1 Regulatory and Operational Language3 **2 PART I**

— FUNDAMENTALS AND CONSTRUCTION OF MARIAH4 2.1 MARIAH:

RISK ASSESSMENT MATRIX IN ARTIFICIAL INTELLIGENCE (Versions A and
B) 4

3 PART II — APPLICATION PROCEDURE.....21 4 **PART III —**

VERSION A (QUALITATIVE).....33 5 **PART IV — VERSION B**

(QUANTITATIVE).....37 6 **PART V — INSTRUCTIONS FOR**

USE.....41 7 **PART VI — LOCAL VALIDATION AND**

REVIEW.....50

REFERENCES.....56

1. INTRODUCTION TO THE INSTRUMENT

MARIAH is the operational module that supports the identification, classification, and documentation of the risks associated with the use of artificial intelligence in research with human beings. Humans. It is applicable, depending on the context, to protocols in Health Sciences, Sciences Humanities and Social Sciences and interdisciplinary research.

MARIAH does not replace ethical review nor does it produce an automatic approval decision. or disapproval. Its function is to structure the analysis, make the risk factors explicit, and support the proportionality of the requirements and increase the traceability of the Committee's deliberation. Ethics in Research (CEP).

The classification should reflect the characteristics of the system and the protocol. Measures of Mitigation measures should be evaluated separately for their sufficiency and verifiability; not They must erase the recognition of the inherent risk they intend to control.

1.1 Regulatory and Operational Language

The guidelines for this MARIAH follow these categories:

- **MUST** — normative requirement or necessary condition for ethical evaluation.
- **RECOMMENDED** — good practice that strengthens protection, transparency, and quality.
or governance.
- **MAY** — an acceptable alternative, dependent on the context of the protocol.

The use of these expressions does not alter the hierarchy of existing norms. Obligations Regulations must indicate their source.

Part 1 — Foundations and Construction of Mariah

This section reproduces, for independent reference, Chapter 8 of the Reference Notebook.

2.1 MARIAH: ARTIFICIAL INTELLIGENCE RISK ASSESSMENT MATRIX (AE VERSIONS)

B)

2.1.1 The construction process and new issues for Research Ethics Committees

This Part can be consulted sequentially or by section, as appropriate.

The reader's needs are met. Table 1 organizes the sections by topic and allows you to locate them.

Quickly find the content you're interested in.

The construction of an ethical evaluation instrument for research protocols that

The use of artificial intelligence requires, before any methodological decision, a choice of

Stance: Does the instrument exist to support the judgment of the Ethics Committee or to replace it?

This guide answers that question unambiguously. The matrices presented in

Parts III (Version A) and IV (Version B) of this MARIAH are tools to SUPPORT deliberation.

These are not new rules, they do not create obligations beyond those already established within the scope of

The National System of Ethics in Research (SINEP) guidelines are not binding. Their adoption is...

an institutional choice for each postal code. What they offer is precision: a path that makes

the most systematic, best documented, and most justifiable evaluation in the eyes of the participants.

of research and society.

Table 1 — Reading guide for Part I

If you need to know...	Go to
How and why was MARIAH built?	Section 2.1.1
The theoretical and normative foundations of the two versions	Section 2.1.2
How does Version A (qualitative / flowchart) work?	Section 2.1.2.1
How does Version B work (quantitative/scoring)?	Section 2.1.2.2
The step-by-step application of the matrix	Part II of this MARIAH
The four risk levels and their requirements.	Steps 3 and 4
The critical rule and its relationship to scoring.	the critical rule

Source: Authors' own elaboration.

This stance is not a concession, but a direct consequence of what... International experience has demonstrated this. None of the eight regulatory models evaluated managed to create a universally effective instrument for all institutional contexts. planned.

The Canadian Algorithmic Impact Assessment model is robust, but it was Designed for government public services, not for research involving human subjects. (Canada, 2025). The guidelines of the Australian National Health and Medical Research Council. (National Health and Medical Research Council — NHMRC) are accurate in the treatment of life cycle, but they do not address the specific legal aspects of data protection in Middle-income countries (Australia, 2023). Guidelines of the Indian Council of Medical Research. (Indian Council of Medical Research — ICMR) are the only ones created to guide Ethics Committees in evaluation of protocols with AI, but they were developed for a different context than Brazilian (ICMR, 2023). Each model offered real contributions and presented real gaps. MARIAH is the synthesis of these contributions, anchored in Brazilian law and guided by... context of SUS.

The journey that led to the creation of the two versions that will be presented at The following sections were deliberately lengthy. The analysis considered five axes of evaluation. internationally recognized and, during the research, included contributions such as the concept Opportunity to Override , which expresses the actual condition of Professionals can challenge the outcome of the system, based on the Korean experience (MFDS, 2023). The right the research participant has the right to request the exclusion of their data from the model training. It comes from the Indian model (ICMR, 2023). The Algorithm Change Protocol (*Algorithm Change Protocol*), which deals with the researcher's responsibility to predict how the model will perform. will be retrained, it stems from Indian and Japanese regulatory experience (CDSCO, 2025; PMDA, [sd]). The concept of context of use as a mandatory preliminary step to risk classification. This is derived from guidelines from the U.S. Food and Drug Administration (FDA) (FDA, 2025). Each one These elements expanded and improved the final instrument.

The result of this process is two versions of the same instrument, with the same ethical basis and the same set of essential questions. Table 2 presents the Comparative structure of the two versions.

The corresponding operational material is in Part II of this MARIAH.

The ethical evaluation of AI protocols cannot be reduced to a mere formality.

Neither version of the matrix acknowledges this risk.

AI absorbs biases from training data and uses opaque logic.

and generates impacts on participants that no other contemporary method produces with Equal scale and speed. Evaluating a protocol with AI is not evaluating a technology, but the impact that this technology can have on real human beings, in contexts concrete, with concrete vulnerabilities. The matrices in this guide exist to make that... more rigorous evaluation. The responsibility for the evaluation remains entirely with the

Postal codes.

Table 2 — Comparative structure of the two versions of MARIAH

Dimension	Version A (qualitative)	Version B (quantitative)
Operating logic	Sequential flowchart	Weighted score
Response type	Binary (yes/no)	Differentiated numerical weights
Structure	5 thematic areas	7 thematic blocks
Risk scale	Levels I to IV (worst-case scenario)	Levels I to IV (total of 275 points)
Mitigation effects: They	do not affect the risk level.	They reduce the total score (Block 7)
Preferred context	Initial assessment cycles using AI; less complex protocols.	Solid experience; complex and auditable protocols.
Main reference	Australian (NHMRC) + Canadian (AIA) model	Canadian model (AIA)
Instrument full	Part III of this MARIAH (Version A)	Part IV of this MARIAH (Version B)

Source: Authors' own elaboration.

2.1.2 Fundamentals, sources and description of MARIAH matrices

This section presents the theoretical, normative, and methodological foundations that support the Two versions of MARIAH. Each design choice is based on experiences. documented regulatory frameworks within the Brazilian legal framework. This applies to the axes, the Questions, weights, and logic for arriving at the risk level. The description of the two versions is... organized into independent subsections. Section 2.1.2.1 deals with the Qualitative Version A. Section 2.1.2.2 deals with the Quantitative Version B. Both versions share the same database.

Ethics and the same set of essential questions. What distinguishes them is the logic of operation.

This distinction determines the audience and context for which each version is most suitable.

The MARIAH is a proportional gradation instrument. It operates on a spectrum that... It ranges from low risk to critical risk and acknowledges that most AI-based protocols occupy intermediate positions that allow for safeguards, conditions, and mitigations. This Gradation has a limit that the instrument cannot overcome. Three dimensions are not gradable by MARIAH: direct and potentially irreversible harm to the participant without a mechanism documented interruption and repair; lack of effective human supervision in context in which the system influences critical clinical decisions; and technical uncertainty does not documented information about the system's behavior in the stated context of use.

Given these dimensions, the principle of non-maleficence and the precautionary principle, established by Law No. 14,874/2024, prevail over any result that the sum of blocks produce (Brazil, 2024a). The matrix structures the CEP judgment, it does not exhaust its Ethical limits. These limits are set by principles established by law, and no one... The assessment tool has the capacity to relativize.

2.1.2.1 Qualitative Version A

The ethical evaluation of research protocols that use AI requires instruments that translate the technical complexity of algorithmic systems into measurable ethical impact about the participant. The qualitative MARIAH Version A was constructed from this premise, based on eight international regulatory experiences and anchored in Brazilian legal framework.

The choice for a qualitative approach

Version A of the Matrix adopts a structured qualitative logic: response questions. binary, organized into five thematic axes, traversed sequentially by means of a flowchart. This choice is based on the experience of the Australian model, which demonstrated that Ethical evaluation instruments are most effective when they guide the evaluator's judgment without replacing it. The evaluation does not require mastery of the system's technical architecture, but rather... understanding the impact it can produce. Qualitative logic preserves this.

centrality of ethical judgment, while ensuring systematicity and traceability to the evaluation process (Australia, 2023).

The binary question format, in which any risky answer increases the... The classification on the axis was adapted from the Canadian Algorithmic Impact Assessment model. (Algorithmic Impact Assessment — AIA). The Canadian AIA has demonstrated, in consolidated use Since 2019, objective and well-formulated questions have been effective tools for... evaluators without technical training in computer science. The Brazilian version adopts this It applies logic and expands it to the research scenario involving human beings. It directs the attention to the provision of government public services for the protection of participants from these studies (Canada, 2025).

The input filter and the characterization of the usage context.

Step 0 (input filter) distinguishes routine uses from uses that require in-depth evaluation. This distinction is essential. Without an initial filter, the committees analyze indiscriminately ranging from bibliographic organization tools to systems autonomous diagnostic support. This compromises the efficiency and proportionality of process. The FDA guideline from the United States of America (USA), published in January of In 2025, it was demonstrated that no algorithmic risk assessment begins without defining... It requires context of use. This refers to the specific role of the system and the scope of the question. that it was developed to respond to (FDA, 2025). The Matrix incorporates this requirement as A mandatory preliminary step for all five axes.

The five axes and their sources

Axis 1, dedicated to the nature and autonomy of the system, articulates two concepts. complementary. The first is the influence of the model: how much AI weighs in the final decision, in relation to other sources of evidence (FDA, 2025). The Ministry of Food Safety and The South Korean Ministry of Food and Drug Safety (MFDS) defines two Criteria (MFDS, 2023): Potential for Malfunction, that is, the chance of algorithmic failures. cause clinical harm; and Opportunity to Contest, which is the real possibility of The professional questions the system's outcome. The combination of these two dimensions allows

To evaluate not only what the system does, but what happens when it makes a mistake. The most important right now is knowing who has the power to intervene.

Axis 2 examines the impact on the participating individual. Its basis is found in Law No. 14,874, of May 28, 2024, and in Decree No. 12,651, of 2025, which establish the Fundamental ethical principles of research with human subjects in Brazil (Brazil, 2024a, 2025). The focus includes a specific question about vulnerable populations: children, pregnant women, the elderly, Indigenous peoples and people in situations of socioeconomic vulnerability. This inclusion addresses a gap in the international models analyzed. None of them were designed for the Brazilian epidemiological and social context, marked by regional inequalities accentuated by the ethnic and cultural plurality of the population served by the Unified Health System (SUS). The axis also incorporates a question about the mechanism for excluding data from model training. This question is based on the guidelines of the Indian Council of Medical Research (ICMR), which establishes the right of the participant to request the suppression of their data. data from the repository used by the AI at any given time (ICMR, 2023). None of the models Western analyses foresee this right.

Axis 3, on data sensitivity and governance, is the axis most directly... anchored in Brazilian law. The General Data Protection Law (LGPD) classifies Health data is considered sensitive personal data and requires a specific legal basis and purpose. declared and identification of the Data Protection Officer for its processing. This set of The requirements are more protective than the data protection laws of all countries. analyzed in this guide, with the exception of the General Data Protection Regulation of the Union. European (GDPR).

The contrast becomes even more evident in relation to the Data Protection Law. India's Digital *Personal Data Protection Act* of 2023, which eliminated the The category of sensitive data was removed from the legal text and health data began to be treated like any other data. other personal data (Carnegie Endowment, 2023). Equating medical records, genetic data and Comparing psychiatric records to current commercial data exposes research participants to risks of discrimination without legal backing for differentiated protection.

Brazilian law, LGPD, makes the opposite choice. Health data constitutes a sensitive category. with enhanced protection. This distinction is one of the pillars of Axis 3 of the Matrix. The axis includes still specific question about risk of re-identification. The MRCT Center and WCG (2025)

They documented that large language models are able to infer identities from Data that appears to be anonymized. This risk does not disappear with formal anonymization.

Resolution CNS No. 738, of February 1, 2024, operationalizes the LGPD in specific context of research with databases involving human subjects. Defines the legal figures of the Controller, the Operator and the Holder (art. 3, items IV, XIII and XX) and It establishes the four institutional terms that structure the data chain of custody: Agreement, Consent, Commitment to Use and Transfer (Article 3, XVI to XIX). The Resolution It also distinguishes two situations for the use of databases — those created within the scope of research (Chapter VII) and those already established outside of that scope (Chapter VIII) — with requirements distinct documents, and regulates the future use of data and the circumstances under which the Term of Free and Informed Consent (TCLE) (art. 20, §§ 2, 3 and 5, and art. 25). MARIAH meets These requirements are addressed in two subsections dedicated to databases — Axis 3.b in Version A. Qualitative and Block 6.b in the quantitative Version B — which incorporate the elements in three Groups of operational questions: Does the protocol identify the database controller? It presents the institutional terms required by the nature of the bank (especially the Term Institutional Approval, mandatory in banks outside the scope of the research)? Presents, When a request is made to waive the Informed Consent Form, the normative hypothesis invoked and the ethical justification must be stated. and sufficient methodological basis (art. 20, § 5, or art. 25)? (Brazil, 2024b).

The corresponding operational material is in Part II of this MARIAH.

Axis 4, on representativeness and transferability, addresses one of the main... Ethical gaps identified in the international literature: trained models predominantly with data from high-income populations and the northern hemisphere They consistently perform worse when applied to populations of different ethnic, socioeconomic and geographic characteristics (WHO, 2024). The guide The Australian and Korean MFDS guidelines documented the absence of local validation. This is one of the main ethical risk factors in protocols that import developed systems. in other contexts (Australia, 2023; MFDS, 2023). The axis also incorporates questions about *Data drift*. Machine learning models degrade their performance as... Real-world data diverges from training data. This phenomenon requires Continuous monitoring throughout the entire system lifecycle. The question has based on the life cycle maintenance concept developed by the FDA (2025) and

operated by the IDATEN system of the Medical Devices and Products Agency

Pharmaceuticals of Japan (Pharmaceutical and Medical Devices Agency — PMDA, [n.d.]).

Axis 5 examines human oversight and explainability. Its central premise is... the principle of irrevocable medical autonomy as stated by the Indian ICMR (2023). Artificial intelligence serves to enhance diagnosis and does not replace clinical judgment. The principle is formulated for clinical practice. In research, it focuses on the design of the supervision described in the protocol: what the Ethics Committee evaluates is not the future care act, but whether the protocol ensures human review and challenge before the execution of each supported decision through the system. Sovereignty over exiting the system, in this context, rests with the researcher responsible, without prejudice to the responsibility of the assisting professional when there is an act. Clinical findings. The principal investigator is responsible for ensuring human verification of all findings result generated by AI. This principle is linked to the concept of the human in the circuit. decision-making (*human-in-the-loop*) present in the Australian model (Australia, 2023).

It aligns with the FDA's requirement that, when the context of use involves a Humans are involved in the decision-making process; the system's performance evaluation should consider the outcome from the human-AI team, not just the isolated performance of the model (FDA, 2025). The axis It also includes a question about the model's retraining protocol. This question has based on the Indian Algorithm Change Protocol, which requires manufacturers to submit a plan in advance describing how the model... This will be updated. This requirement was established by the Central Control Organization of Central Drugs Standards Control of India (CDSCO) Organization) (India, 2025). In the context of this guide, this principle translates to It is the researcher's responsibility to inform the Ethics Committee how and when the model will be implemented. modified during the study.

The convergence of some axes around the clinical impact on the participant is intentional, not inadvertent overlap. The axes of nature and autonomy of the system, of The impact on the participant and the transferability and performance assessments are dimensions... distinct aspects of the same reality: the different ways the system impacts the individual who participates in the research. The nature axis examines how the system operates. The axis of The impact examines what the system produces on the participant. The transferability axis. It examines whether what the system produces in one context will remain valid in another. These are questions Different perspectives on the same risk because risk has multiple dimensions beyond a single axis.

It does not capture. A protocol that presents a high risk in all three areas simultaneously is not...
 penalized three times for the same problem. This demonstrates that the risk is consistent and structural.
 and not mitigable by a single safeguard. The convergence between axes in a high outcome.
 Risk is evidence of severity, not a methodological artifact.

The logic of the sequential flowchart and the critical rule.

The sequential flowchart was adopted for two reasons. The first is pedagogical.
 The sequence guides the evaluation in a progressive manner, from the simplest to the most complex, and avoids
 leaps that compromise the consistency of the analysis. The second is ethical. The level of risk only
 It goes up, it never goes down. This logic is consistent with the precautionary principle that underlies the...
 Ethics in research with human subjects (Brazil, 2025). A single axis at Level IV is sufficient.
 to elevate the entire protocol to the highest level of rigor, regardless of
 Performance in other areas. The absence of human supervision in a system that
 It influences therapeutic decisions, for example, but is not compensated for by good responses in
 data axes or representativeness.

The critical rule originates directly from the Canadian Impact Assessment model.
 Algorithmic (Canada, 2025). The model demonstrated that allowing mitigations to reduce the
 The level of risk creates a perverse incentive. Researchers are now declaring safeguards for
 They obtain more favorable ratings. They fail to demonstrate that the risk was effectively...
 controlled. The rule solves this problem with a clear separation. The risk level is
 determined by the characteristics of the system and the protocol. The required specifications are
 adjusted based on the guarantees demonstrated. This separation preserves the integrity of
 The instrument and the clarity of communication between the researcher and the committee are key.

2.1.2.2 Quantitative Version B

Version B of MARIAH uses questions with varying numerical weights in seven blocks.
 thematic. The sum of the points determines the risk of the protocol on a scale of I to IV. This

This approach makes your decisions more traceable and auditable. The scoring identifies areas of
 This text discusses risk concentration and shows how the researcher's choices affect that risk.

The choice for a quantitative approach

The quantitative logic is inspired by the Canadian Impact Assessment model. Algorithmic (Canada, 2025). The model operates with 65 risk questions, maximum score of 169 points, and 41 mitigation questions, maximum score of 75 points, distributed into four risk levels.

The Canadian experience demonstrated two concrete advantages of the model. Quantitative. The first is precision: the score identifies in which dimensions the risk is present. The second is the one that is focused and guides the committee's requirements proportionally. Consistency: protocols evaluated at different times and by different members of the Ethics Committee. They become comparable, which strengthens the consistency of institutional decisions (Canada, 2025). Version B incorporates this logic and adapts it to the context of research with living beings. Humans in Brazil. The focus shifts from governmental public services to...

Protection of research participants and compliance with the LGPD (Brazilian General Data Protection Law).

This guide proposes a structure in seven blocks: Project, System, Algorithm, Decision, Impact, Data, and Mitigation. The blocks with the greatest ethical and regulatory weight receive questions with a higher individual value. This choice makes explicit the hierarchy of priorities of The Impact and Data blocks together account for 139 of the 275 maximum points. from the matrix. They represent 50.5% of the total score. This concentration expresses a choice deliberate. In research with human subjects, what matters most is what the system can do. to the participating individual and how their data is handled.

The input filter and the characterization of the usage context.

Version B retains Step 0 and the characterization of the usage context from Version A. as mandatory steps prior to scoring. The FDA (USA) guideline demonstrated that No algorithmic risk assessment begins without a precise definition of the context of use.

the specific role of the system and the scope of the question it was developed to answer.

(FDA, 2025). Characterization does not score. It delimits the object of the evaluation and prevents the

The system should be evaluated in the abstract, detached from the specific protocol submitted for analysis.

The seven blocks and their sources

Block 1 assesses the protocol's suitability for the use of AI. It examines transparency. in the system description, the team's qualifications, the clarity of the methodological rationale, and the Adequacy of the Informed Consent Form. The maximum score is 22 points. Each question is worth three points. With the exception of the question about external technical consultation, which is worth four points. The block operates as a verification of document completeness.

The protocol should describe the system's development stage: discovery. Exploratory (*Discovery*), clinical translation (*Translation*), or deployment (*Deployment*). This requirement is based on the MRCT Center and WCG (2025) framework, which demonstrated that ethical risks vary depending on the stage of the AI lifecycle in which it operates. One A system in the exploratory discovery phase presents different risks than a system already in operation. Clinical implementation. This distinction should be visible in the protocol submitted to the Ethics Committee.

Block 2 assesses the technical nature of the system: autonomy, adaptability, Explainability and traceability. The maximum score is 17 points. Each question is worth Two or three points. The block articulates two concepts developed from different frameworks. The first is the influence of the model: how much AI weighs in the final decision compared to other factors. sources of evidence. This concept was developed by the FDA (2025). The second is that of potential for failure (*Malfunction Potential*) and opportunity to contest (*Opportunity to Override*): the probability that algorithmic failures will cause clinical harm and the existence of Real conditions that allow the responsible professional to question the system's results.

This concept was developed by the Ministry of Food and Nutrition Security. South Korean medicines (MFDS, 2023). Direct risk questions are worth three points. Each. They deal with situations without mandatory human intervention, adaptive systems, and... Opaque systems. Their proximity to direct risk to the participant justifies the greater weighting. Questions regarding lack of guarantee are worth two points. They deal with documentary deficiencies that... Although serious, they are more easily corrected.

Block 3 assesses the quality, origin, and representativeness of the data.

Training and testing. The maximum score is 14 points. Each question is worth two points.

The bloc is directly based on the Australian guide and the guidelines of the Korean MFDS, which

They documented that the absence of local validation is one of the main ethical risk factors in

protocols that use systems developed in other contexts (Australia, 2023; MFDS,

2023). The block includes a question about independence between training and test data. A

FDA guidelines state that test data must be strictly independent of

development data so that the model's performance evaluation is valid (FDA,

2025). It also includes a question about representation by population subgroups. A

The World Health Organization (WHO) has identified that models trained predominantly

data from high-income populations and the northern hemisphere show performance.

systematically lower when applied to populations with different characteristics

ethnic, socioeconomic and geographic (WHO, 2024). This gap is especially relevant

for the Brazilian context.

Block 4 assesses the consequence of the decision. The maximum score is eight points.

The first two questions are worth three points each. This asymmetry is deliberate. The questions

This section deals with the situations of greatest potential severity throughout the instrument: the

The system as the sole determinant of a decision without human review and its irreversibility.

the damage in case of error. The concentration of points on a few high-weighted questions reflects

the logic of the consequence of the decision developed by the FDA (2025). When the model operates

with high human influence and low human oversight, the severity of the potential adverse event

This should be the dominant factor in risk assessment. The maximum score for this block, by itself,

You can move a Level II protocol to Level III. This move expresses the

The seriousness of the situations described in these questions.

Block 5 assesses the impact on the participant. The maximum score is 62.

points. Each question is worth five or six points. The questions with the greatest direct impact are worth more.

six points: influence on clinical decisions, risk of physical harm, risk of psychological harm or

social engagement and involvement of vulnerable populations. The bloc is based on Law No. 14,874,

of May 28, 2024, and in Decree No. 12,651 of 2025, which establish the ethical principles.

Fundamentals of research with human subjects in Brazil (Brazil, 2024a, 2025). The block includes

The question concerns the inference of sensitive features not explicitly collected. The MRCT

Center and WCG (2025) documented that large language models are able to infer

Sensitive identities and attributes can be identified from seemingly neutral data. This risk does not... disappears with formal anonymization. The block also includes a question about the mechanism of Exclusion of data from model training. Guidelines of the Indian Council of Medical Research institutions establish the participant's right to request the deletion of their data from... repository used by AI at any time (ICMR, 2023). None of the Western models The analyses provided for this right. This guide incorporates it as an ethical requirement.

Block 6 assesses data sensitivity and governance. The maximum score is... 77 points. Each question is worth between five and seven points. Questions directly linked Each section of the LGPD (Brazilian General Data Protection Law) is worth seven points. This section expresses a central choice of the guide. The law is... The most robust protective differential of the Brazilian framework compared to all other models. international countries analyzed.

The LGPD classifies health data as sensitive personal data and requires a legal basis. specific, stated purpose and identification of the Data Protection Officer for its processing. This level of protection surpasses that of the Chinese Personal Information Protection Act (PIPL). of the Japanese Personal Information Protection Act (APPI) and, in particular, the Act of Indian Digital Personal Data Protection Act of 2023 which completely removed the category of sensitive data from its wording and equated medical records and any other personal data (Carnegie Endowment, 2023).

The question about transferring data to servers outside of Brazil is worth seven. points. Depending on the receiving country and the contractual guarantees established, this transfer This could completely compromise the protection afforded by the LGPD to participants in research. The question about the risk of re-identification incorporates the documented concern. with the ability of AI systems to infer identities from seemingly unknown data anonymized (MRCT Center; WCG, 2025).

Subsection Block 6.b of Version B brings together five specific points for protocols that They use databases, with weights calibrated by operational centrality in the Resolution. CNS No. 738/2024 (Brazil, 2024b): nominal identification of the bank's Controller — P6.b.1 (seven points); presentation of the Institutional Consent Form when required by the originating entity. from the bank — P6.b.2 (seven points, ELIMINATORY); presentation of the Commitment Agreement Data Usage — P6.b.3 (five points); reasoned framing of any exemption from informed consent in the applicable case according to the origin of the bank — P6.b.4 (five points); and

Formalization of joint control in multicenter research, by means of a Terms of Agreement.

Institutional — P6.b.5 (five points).

Block 6.b adds 29 points to the total score when activated by the filter in Step 0.

The absence of an affirmative response in P6.b.2, when applicable, renders the protocol unevaluable.

on the merits of the critical rule (item "The critical rule in Version B", below), regardless

of the total score. The subsection also includes three due diligence items — P6.b.4.1 (need

and proportionality of access to identifiable data), P6.b.6 (cumulative criteria for

waiver of informed consent in AI research) and P6.b.7 (Controller's statement) — which does not

They don't score points nor add up to the ceiling: they operate as a cumulative *gate*, and the absence of any

This criterion implies the return of the protocol.

Block 7 assesses mitigation measures. It is the only block with a bidirectional logic: questions of

Risk factors add points to the total, and mitigation questions subtract. The maximum score is...

75 points. The reduction capacity through mitigation measures reaches 65 points — to which are added

up to 23 points of evidence from sub-block 7C. The block operationalizes an inherited distinction.

from the Canadian model. Mitigation measures do not alter the risk level determined by the others.

blocks, when evaluated in isolation, reduce the total score and, therefore, may

to move the protocol to a lower level on the consolidated scale (Canada, 2025). This logic

It is more flexible than Version A, where mitigations do not affect the level in any way.

hypothesis. It reflects the understanding that, in Version B, the scoring already incorporates the quality.

of the safeguards adopted by the researcher as an integral part of the risk assessment.

Sub-block 7C, concerning transparency evidence, operates under a distinct, one-off regime: a

Documented evidence subtracts points, while its absence adds no points.

The block is divided into three sub-blocks. Sub-block 7A (queries) is worth ten points.

Assess whether the researcher sought external technical guidance before submission, including

consultation with the National Health Surveillance Agency (Anvisa) regarding the classification of

system in the Computer Program as a Medical Device (Software as a

Medical Device (SaMD). Sub-block 7B (risk reduction measures, *de-risking*) is worth 65

Points. Assesses the quality of human oversight mechanisms, explainability,

Lifecycle monitoring and failure management. The sub-section includes a question about

monitoring data drift, that is, the degradation of data performance.

The model changes over time. This question is based on the concept of maintaining the model.

life cycle developed by the FDA (2025), operationalized by the agency's IDATEN system.

Japanese regulatory protocol (PMDA, [n.d.]) and the Algorithm Change Protocol (*Algorithm Indian Change Protocol* (CDSCO, 2025). Sub-block 7C (evidence of transparency) records documentary evidence — standardized documentation of the model and the database. data, algorithmic impact assessment or Data Protection Impact Report

Personal information and the availability of source code — which can reduce the score by up to 23 points when present and documented, without adding up when absent.

Level ranges and score calibration

The four level ranges were calibrated based on two principles. The first is... proportionality. The Level I range is narrow, from zero to 58 points, because protocols with Truly low-risk uses of AI are rare and should be rigorously classified.

The second factor is risk sensitivity. The Level IV range covers 67 points, from 209 to 275, because protocols at this level exhibit combinations of severe characteristics in multiple blocks. The width of the band reflects the variation in gravity within the critical level itself.

When the filter in Step 0 activates Block 6.b (database), the matrix ceiling exceeds 275. for 304 points and the bands are recalibrated proportionally: Level I from 0 to 64 points, Level Level II: 65 to 141 points, Level III: 142 to 230 points, and Level IV: 231 to 304 points (ceiling). Evaluable score of 297 points, excluding the elimination question P6.b.2). Recalibration preserves the relative distance between levels and prevents the 29 additional points of Block 6.b from distorting the classification. The ranges will be empirically tested during the usage and testing period of the Version 1.0; any adjustments will be considered in version 2.0, to be submitted to Inaep.

The corresponding operational material is in Part II of this MARIAH.

The critical rule in Version B

In Version B, the critical rule has a different formulation than in Version A. The level The score determined cannot be altered by the evaluator's subjective decision.

The score is the systematic result of applying the instrument. Mitigation measures are already in place. incorporated into the scoring of Block 7. They constitute the only legitimate way to reduce the final score. This architecture preserves the integrity of the instrument. It prevents that undocumented declarations of intent replace evidence of effective control of

risk. The Canadian model demonstrated that this principle is essential for the credibility of instrument for use by researchers, sponsors and society (Canada, 2025).

The gradation described so far coexists with four safeguards that are not graded. These are the special rule of Axis 3.b, the Ethical Precedence Clause, the chain of offense elimination, custody and the diligence preventing new consent, all described in the Supplement to Technical Note on Premises. They operate automatically, without depending on the judgment of... who applies the matrix. The first two determine the risk level on their own. The last two They suspend the merit analysis and return the protocol to the researcher for further investigation. None of them allow for compensation for declared mitigations, because they all deal with conditions that precede the risk gradation. The triggers and the procedure for each are in Part II of this MARIAH.

The corresponding operational material is in Part II of this MARIAH. Table 3 This summarizes MARIAH's international sources by axis and bloc.

Table 3 — International sources by axis/block of MARIAH

Axle/Block	Dimension evaluated	Main source	Contribution to MARIAH
Axis 1 / Block 2	Nature and autonomy of system	FDA (2025); MFDS (2023)	Influence of the model; potential for failure; opportunity for challenge
Axle 2 / Block 5	Impact on the participant	Brazil (2024a; 2025); ICMR (2023); MRCT Center e WCG (2025)	Protection of vulnerable populations; right to data exclusion; inference of sensitive attributes.
Axle 3 / Block 6	Data sensitivity and governance	LGPD; Res. CNS 738/2024; GDPR; Carnegie Endowment (2023)	Sensitive data with enhanced protection; Controller, Operator and Institutional Terms (CNS Resolution 738/2024); risk of re-identification; international transfer.
Axle 4 / Block 3	Representativeness and transferability	WHO (2024); Australia (2023); MFDS (2023); FDA (2025)	Local validation; performance by subgroups; drift monitoring. data
Axis 5 / Block 7	Human supervision and explainability	ICMR (2023); FDA (2025); PMDA [s.d.]; CDSCO (2025)	Irrevocable medical autonomy; human involvement in decision-making; retraining protocol

Axle/Block	Dimension evaluated	Main source	Contribution to MARIAH
Block 1	Protocol adequacy	MRCT Center e WCG (2025)	Life cycle stage (<i>discovery, translation, deployment</i>)
Block 4	Consequence of the decision	FDA (2025)	Influence of the × model human supervision × severity of outcome

Source: Australia (2023); Brazil (2024a; 2025); Canada (2025); Carnegie Endowment (2023); CDSCO (2025); FDA (2025); ICMR (2023); MFDS (2023); MRCT Center and WCG (2025); PMDA [n.d.]; WHO (2024).

3 PART II — APPLICATION PROCEDURE

How to use this MARIAH

Version A, which is qualitative, operates using a sequential flowchart: five axes are traversed in The question is asked in order of yes or no answers, and the risk level emerges from the combination of responses. It is a reasoning-oriented tool that guides evaluation without replacing judgment. human.

This procedure is recommended for postal codes that are beginning to familiarize themselves with the... evaluation of protocols involving AI, for cases where the protocols present lower technical complexity, and when it is a priority to ensure transparency in the deliberative process instead of a numerically precise result.

The worst-case scenario approach, which considers a single axis at Level IV as the elevator. Throughout the protocol, it reflects the precautionary principle present in research ethics that involves human beings (Brazil, 2012; Brazil, 2024a; Brazil, 2025).

The quantitative Version B operates on a weighted scoring system: the same questions They receive different numerical weights, distributed across seven thematic blocks, and the sum It determines the risk level on a 275-point scale. The instrument guides... Traceability. It allows you to identify where the risk is concentrated, as well as the mitigation measures. The measures adopted reduced it, and how different protocols compare to each other.

The choice of this version is justified for CEPs with more experience in evaluating AI-powered protocols, for protocols of high technical complexity and for institutions that They need to accurately document and audit their decisions, such as university hospitals. national research institutes and sponsors of multicenter clinical trials. The blocks Impact and Data account for 50.5% of the total score. This concentration reflects the The instrument's hierarchy of priorities. What matters most in research with living beings? Regarding humans, it refers to what the system can do to the participating individual and how their data is handled.

The two versions are compatible and complementary. A postal code (CEP) can adopt Version A. as an initial screening and resort to Version B to further analyze protocols that... Version A is classified as Level III or IV. A CEP may systematically adopt Version B and Use Version A as a guide for researchers who need to understand the The logic of the evaluation before submitting your protocols.

A postal code (CEP) may also choose not to use either version and use this chapter as a benchmark. conceptual framework for developing your own instrument, adapted to the specificities of your field. Thematic, institutional structure, and the profile of the protocols it receives. Flexibility guarantees the validity of the guide. Ignoring the institutional diversity of Brazilian postal codes. It would compromise its purpose.

In operational terms, the evaluation of subsection Axis 3.b (Version A) or Block 6.b (Version B) by CEP, in protocols that use databases, includes five questions. concrete, derived from the articles cited in the Reference Notebook. The first (3.b.1 / P6.b.1): the The bank's controller is identified by name, including their position, function, and institution. Linked? The second (3.b.2 / P6.b.2, ELIMINATORY): for banks already established outside of Within the scope of the research, the protocol includes an Institutional Consent Form signed by who is the head of the institution holding the data? The third (3.b.3 / P6.b.3): the Term of The Data Use Agreement signed by the principal investigator is included in the dossier of protocol? The fourth (3.b.4 / P6.b.4): request for waiver of informed consent, if any, indicates the hypothesis applicable regulations and presents sufficient ethical and methodological justification — art. 20, § 5.º, for banks established outside the scope of the research, or art. 25, for banks established in Scope of the research? The fifth (3.b.5 / P6.b.5): in multicenter research, controllership Is the joint agreement formalized by an Institutional Agreement?

The absence of any of these terms, when required by the nature of the bank, This does not constitute a minor issue: it constitutes the absence of a formalized chain of custody, with consequences in the critical rule (Reference Notebook, Chapter 8, item "The critical rule").

Part VI of this MARIAH offers a practical roadmap for postal codes that wish to test. Locally, the level zones are determined according to their specific circumstances. The plan proposes three approaches. independent: agreement between raters, empirical distribution of protocols in four levels and convergence between the two versions of the matrix. In addition, it includes a A template spreadsheet is provided to assist in the process. Participation is voluntary. Results will be returned to you. The working group responsible for the guide is optional. Contributions received will be... considered in future review cycles.

How to apply the matrix: steps and procedures

The matrix operates in four stages. Step 0 verifies if the protocol requires evaluation. In-depth assessment is needed, or whether the use of AI is routine. Step 1 defines the object of the assessment by means of... of eight mandatory descriptive questions. Step 2 covers the five assessment axes, which They examine the nature of the system, the impact on the participant, data governance, and... representativeness of the model and human supervision. Step 3 translates the results into a of the four levels of algorithmic risk. Step 4 associates a set of requirements with each level. minimum requirements that the researcher must demonstrate having met. The following sections describe each. Step. The complete instruments are contained in Parts III and IV of this MARIAH.

Step 0: Input filter

The evaluation process begins before the axes and before the scoring. The first The next step is to verify if the protocol submitted to the Ethics Committee requires the application of the matrix. The question It's straightforward: in this protocol, the artificial intelligence system performs automated decision-making. Content generation or intervention in the conduct of the study or its results?

If the answer is negative, the use is classified as routine and the matrix does not apply. Systems that operate exclusively as administrative support, organization of bibliographic references, audio transcription, grammar review, style adaptation or translation — or any other function that does not influence the study's decisions, nor produce effects on the participant — are not within the scope of the evaluation, even though their Usage must be declared. The protocol follows the standard CEP (Brazilian Postal Code) flow.

If the answer is affirmative, the next step is to characterize the context of use. This filter does not eliminate the need to mention the use of AI in the protocol. The researcher must declare any... The use of algorithmic systems, even if routine. What the filter determines is the level of Depth of the evaluation.

The distinction between routine and non-routine use is based on the framework of MRCT Center and WCG (2025). The absence of a clear entry criterion leads the CEPs to to indiscriminately evaluate tools with very low impact on high-risk systems. This misconception compromises the efficiency and proportionality of the process. The input filter Solve this problem before it happens.

The obligation to declare the use of AI in research precedes the matrices in this guide.

Policy on Integrity in Scientific Activity of the National Council for Scientific and Technological Development

The Scientific and Technological Commission, established by Ordinance of March 6, 2026, determines that

Any use of generative AI must be declared, specifying the tool used.

and the purpose, at any stage of the research development (CNPq, 2026).

The matrix input filter checks, before any other step, if this statement is present and if the system's role is described with sufficient precision to enable the evaluation. The absence of this information does not constitute a documentary gap. It constitutes Violation of scientific integrity with express normative support in federal legislation. current.

In the same Step 0, a second filter, with a complementary function, checks if the The protocol uses a database, whether it is its own, from another research project, or from an institution. external. An affirmative response activates the subsections dedicated to databases, derived from CNS Resolution No. 738/2024: Axis 3.b in Version A (qualitative) and Block 6.b in Version B (quantitative). These subsections operationalize the identification of the Controller, the verification of the applicable institutional terms and the framework for any waiver of informed consent. If the The protocol does not use a database, for example, when it collects data exclusively from... The scope of the research with each participant, the questions from Axis 3.b / Block 6.b remain. inapplicable.

In this case, the evaluation continues with the other axes. The accuracy of this second filter It avoids two symmetrical flaws: requiring documents that the nature of the protocol makes unnecessary. unnecessary requirements and stop demanding those that the nature of the protocol makes mandatory.

Step 1: Characterization of the context of use

Once the entry filter is approved, the protocol must answer eight descriptive questions. mandatory in both versions of the matrix.

First: what is the question that the AI system was developed to answer? In this protocol? Second: does the system answer this question autonomously or Does it support a decision made by a human being? Third: what is the identification status? of the personal data used — identifiable, pseudonymized, anonymized or combination by study stage? Fourth: what is the approximate volume of records of

Participants? Fifth: at what point does anonymization occur in relation to access by researchers — before access, by the custodial institution; after access, by the researchers themselves. researchers; there will be no anonymization; or not applicable (answer only if the data (If they are not anonymized)? Sixth: what is the role of the AI system — analysis tool Regarding research data, the object of development (model training), or both? A Seventh: What is the intended use of the model — restricted use within this protocol, future applications? or broad use, sharing with third parties, or does it not apply? The eighth: will there be Sharing the database, in whole or in part, with third parties, in addition to the institutions. participants — no; yes, under a Consent Form or Institutional Agreement; or yes, other shape?

These questions do not score or determine the level of risk on their own. They They define the object of evaluation and prevent the system from being evaluated in the abstract. detached from the specific protocol submitted for analysis. The same AI system can to present radically different risk profiles depending on the role it plays in each Protocol. A language model used to support the coding of qualitative data. It is a different object of evaluation from the same model used to generate recommendations. Personalized therapies. The context of use defines what should be analyzed.

It is impossible to assess the credibility of an AI system for regulatory purposes without first, define precisely what question it was developed to answer and in what context of use. A specific approach will be applied. This concept was developed by the US FDA (FDA, 2025). MARIAH incorporates this requirement as a mandatory preliminary step, prior to any other risk classification.

Step 2: The five axes of evaluation

After characterizing the context of use, the five axes of algorithmic risk are evaluated. Each axis corresponds to a dimension, the risk level of which is defined by responses. qualitative (Version A) or weighted score (Version B). The five axes are described in The questions for each section are fully contained in Parts III and IV of this MARIAH.

Axis 1 examines the nature and autonomy of the system. It assesses whether it operates in a way... autonomous or under mandatory human supervision, whether it is adaptive, whether it is explainable, and whether it was validated before use in the protocol.

The axis articulates two distinct reference frames. FDA (2025) developed the concept of Influence of the model: how much AI weighs in the final decision compared to other sources of... evidence. The Ministry of Food and Drug Safety of South Korea He developed the concepts of Malfunction Potential *and* opportunity . contestation (*Opportunity to Override*) (MFDS, 2023). The first assesses the probability that Algorithmic flaws generate clinical risk. The second assesses the degree to which professionals have concrete conditions for questioning the system's decisions.

Axis 2 examines the impact on the participant. It assesses whether the system affects clinical decisions, whether it can cause physical, psychological, social or economic harm, whether it involves vulnerable populations and sensitive characteristics not directly provided by are inferred. participant. The axis is directly grounded in Brazilian legislation regarding research with living beings. humans (Brazil, 2024a, 2025). It also incorporates the participant's right to request exclusion. from their model training data, as established by the guidelines of the Indian Council of Medical Research (ICMR, 2023).

Axis 3 examines the sensitivity and governance of the data. It assesses whether the data being processed is... Data in the protocol is sensitive under the LGPD (Brazilian General Data Protection Law) if it is transferred to servers outside of the company. In Brazil, if there is a real risk of re-identification, does the consent cover use for training purposes? of models and whether there is an incident response plan. The axis is more directly anchored in Brazilian legislation receives more weight than any other from the parent company. In Version B, it receives the greatest weight. regulatory framework of the instrument.

Axis 4 examines the representativeness and transferability of the model. It assesses whether the... Training data represent the diversity of the protocol's target population, if the The model was validated in a context comparable to the SUS (Brazilian Unified Health System), if there is performance evaluation by... population subgroups and whether the test data are independent of the training data.

This axis addresses one of the main ethical gaps identified in the literature. International. Models trained predominantly with data from high population density. income levels in the northern hemisphere consistently underperform when applied to populations with different ethnic, socioeconomic, and geographic characteristics (WHO, 2024). This gap is especially critical for protocols conducted within the framework of THEIR.

Axis 5 examines human oversight and explainability. It assesses whether there is human review. mandatory before the execution of decisions, if the team has the technical competence to

to interpret and challenge the results, whether there is continuous performance monitoring and whether there is Documented protocol for failures and for retraining the model.

The central foundation of the axis is the principle of irrevocable medical autonomy. As stated by ICMR (2023), artificial intelligence serves for diagnostic enhancement and not... It replaces clinical judgment. The principle is formulated for clinical practice. In research, it focuses on the supervision design described in the protocol: what the Ethics Committee evaluates is not the act itself. future care, but if the protocol guarantees human review and challenge before the execution of each decision supported by the system. Sovereignty over the system's outcome rests with, In this plan, the responsibility lies with the principal investigator, without prejudice to the professional's responsibility. assistant when a clinical act arises. The axis also incorporates the concept of life cycle maintenance developed by the FDA (2025), operationalized by the system of Timely Evaluation of Innovation IDATEN (Improvement Design within Approval for Timely Evaluation and Notice) of the Japanese regulatory agency (PMDA, [n.d.]) and by the Protocol of Indian Algorithm Change Protocol (CDSCO, 2025).

Step 3: Levels I to IV

The result of the evaluation across the five axes is the classification of the protocol into one of the four levels of algorithmic risk. The levels are not pass or fail categories. These are classifications that determine the set of minimum requirements that the researcher must meet. demonstrate having met the requirements before approval by the Ethics Committee.

Level I (Low) corresponds to protocols in which the AI system operates with limited autonomy, without direct influence over clinical or therapeutic decisions nor on the research procedures and outcomes — allocation to arms, eligibility, data collection and analysis — with low-sensitivity data and effective human oversight.

Typical examples include the use of AI for organizing and categorizing references. bibliographical data, automated transcription of interviews, or support for exploratory analysis of unidentified data.

Level II (Moderate) corresponds to protocols in which the AI system plays a role. a more relevant role in the protocol, but with reversible consequences and human supervision. guaranteed. Typical examples are protocols that employ language models for support for qualitative coding, participant eligibility screening systems or Medical image analysis tools with mandatory human review.

Level III (High) corresponds to protocols where the AI system has an influence. directly influencing decisions with the potential to harm the participant, operates with sensitive data in large scale or presents documented limitations in representativeness and transferability. Typical examples include protocols that employ diagnostic support systems in units. basic health, predictive risk models in clinical trials, or tools for Patient stratification for therapeutic protocols.

Level IV corresponds to critical risk. Protocols that fall into this level include... that the system operates with a high degree of autonomy in decisions with direct and potentially impactful outcomes. irreversible damage to the participant, without effective human supervision or with serious gaps in Validation and governance. Typical examples are protocols that employ systems. Autonomous therapeutic decision-making tools, diagnostic tools without human review. mandatory or prognostic models used to define eligibility in studies of Access to experimental treatments. Table 4 summarizes the four algorithmic risk levels. and its requirements.

Table 4 — Algorithmic risk levels: summary

Level	Classification	Typical profile	Minimum requirements (cumulative)
I	Low	AI has no influence on clinical decisions; data has low sensitivity; effective human supervision is lacking.	Declaration of AI usage and appropriate transparency to affected individuals, where applicable.
II	Moderate	AI plays a relevant role, has reversible consequences, and is subject to guaranteed oversight.	+ System description; LGPD compliance; legal basis
III	High	AI influences potentially harmful decisions; large-scale sensitive data; limitations in representativeness.	Technical validation; performance by subgroups; <i>human-in-the-loop</i> ; Data Protection Officer.
IV	Critical	High autonomy, direct and potentially irreversible impact, without effective supervision.	External review; continuous monitoring; notification to Anvisa (SaMD); assessment by an accredited or qualified Ethics Committee in AI.

Source: author's own elaboration.

Step 4: Requirements by level

Each level corresponds to a set of minimum requirements that the researcher must meet.

To meet the requirements for protocol approval by the Ethics Committee. The requirements are cumulative. Those of Level II Level I includes those in Level III. Level III includes those in Levels I and II. Level IV includes all the previous ones.

At Level I, a declaration of AI use in the protocol and mention in the Informed Consent Form (ICF) are required. What algorithmic system will be used and what its role is in the study.

At Level II, the system description and the data source are added. training, demonstration of compliance with the LGPD (Brazilian General Data Protection Law), and indication of the legal basis for the processing of the data used.

At Level III, the technical validation documentation of the system and the analysis are added. performance by relevant population subgroups, verification of the mechanism of Effective human supervision in the decision-making loop (*human-in-the-loop*) and the identification of Data Protection Officer of the institution responsible for the study.

At Level IV, the external technical opinion on the system and the plan are added. Continuous monitoring with defined interruption criteria, notification to Anvisa. when the system qualifies as SaMD and the protocol is submitted for review by the Ethics Committee. accredited for high-risk or CEP protocols with specific qualification in Artificial intelligence. Table 5 summarizes the essential questions for applying MARIAH.

Table 5 — Essential questions for applying MARIAH

Question No. that the protocol should answer	
1.	Does the protocol declare the use of AI, specifying the tool and its purpose?
2.	Does the system operate autonomously or does it rely on human decision-making?
3	What question was the system developed to answer in this protocol?
4	Is there effective human oversight in the decision-making process?
5	Do the training data represent the diversity of the SUS target population?
6	Does the protocol demonstrate compliance with the LGPD (Brazilian General Data Protection Law), with a specific legal basis and an identified Data Protection Officer?
7	Is there a documented mechanism for excluding participant data from model training?
8	Is there a protocol for monitoring data drift and retraining the model?
9	Are the stated mitigation measures documented with evidence, or are they merely statements of intent?

	Question No. that the protocol should answer
10	Does any aspect of the protocol meet the criteria for conditions not graded by MARIAH (irreversible damage without interruption, absence of human supervision in critical clinical decision-making, undocumented technical uncertainty)?

Source: compiled by the organizer.

The critical rule

The risk level determined by the assessment does not change with the mitigation measures presented. by the researcher after classification. This applies to the qualitative logic of Version A and to the Version B scoring. Only the required criteria can be adjusted based on the... Guarantees demonstrated.

The rule is directly based on the Canadian Impact Assessment model. Algorithmic (Canada, 2025). Allow only declared mitigations, presented after the Classification, reducing the level of risk, creates a structurally perverse incentive. Researchers are now declaring safeguards to obtain more favorable classifications. They fail to demonstrate that the risk was effectively controlled. The separation between the level of Risk and required criteria preserve the integrity of the instrument and also ensure clarity. communication between researcher and committee.

In Version B, this same logic operates by design: the mitigations are not introduced later. Regarding the classification: these are already incorporated into the calculation, in the Block 7 score, and only subtract. Points when documented and verifiable. This is the only way in which mitigations... They influence the result. No measure presented after the classification alters the level. obtained.

The critical rule expresses the principle that guides this entire section. The evaluation does not fall... about intentions, but about evidence. What a protocol describes as supervised, The matrix needs to demonstrate that it is supervised, explainable, and validated. It exists to make this demonstration systematic. The responsibility for demanding it remains. with the Committee.

The critical rule incorporates both elimination hypotheses and hypotheses of ethical prevalence. CNS Resolution No. 738/2024 eliminates protocols that use external databases without a Term. Institutional Approval signed by the head of the institution (art. 27, VI). This corresponds

to question 3.b.2 in Version A and to question P6.b.2 in Version B, both marked as Qualifying Rounds (Brazil, 2024b).

The lack of consent indicates that the originating institution did not authorize the use of the data. In such a case, the dossier is returned to the researcher for mandatory due diligence before any further action is taken. merit analysis.

Some chain-of-custody failures increase the level of risk, but do not prevent it. automatically. The lack of identification of the Controller (questions 3.b.1 / P6.b.1) and the request for exemption from informed consent without indication of the applicable regulatory basis, without justification. sufficient ethical and methodological standards or a lack of proportionate safeguards (questions 3.b.4 / P6.b.4) They elevate the protocol according to the subsection consolidation rule.

The Ethical Precedence Clause represents the second category of direct intervention. in the classification. Mark affirmatively the questions P4.1 (decision without human review) or P4.2 (irreversible harm to the participant) triggers the precautionary principle. In such cases, the CEP You can raise the protocol to Level IV immediately, regardless of the total score. as described at the end of Part IV of this MARIAH.

Unlike the elimination clauses, the Precedence Clause does not rule out... The evaluation workflow protocol. Instead, it elevates it to the highest level of requirements.

The coexistence of weighted scoring and critical rules deserves explicit formulation. to avoid ambiguous interpretation. The two operate at different times and with different logics.

The score runs through the seven blocks, accumulating the distributed risk profile of the protocol. and produces a classification based on the total sum, already including the mitigations of Block 7. incorporated. The critical rule comes in later, as a clause of ethical prevalence: it verifies if Some aspect of the protocol has a structural flaw that the overall picture should not obscure. When the critical rule applies, it doesn't cancel out the score; it takes precedence over it.

The result is a protocol with a favorable sum, but with a structural flaw in one Critical dimension, should not be interpreted as low risk. The score maps the risk. distributed. The critical rule protects against the illusion that the average can hide the unacceptable. Both operate in the same direction: ensuring that the evaluation is based on evidence, not on... Intentions. Table 6 shows the connections with the other modules in the set. documentary.

Table 6 — Connections with the other modules of the document set

To delve deeper...	Go to
The critical properties of AI systems — bias, explainability, human oversight, and lifecycle.	Journal Reference, Chapter 3
Data protection for participants and data processing in AI.	Journal Reference, Chapter 7
The Brazilian regulatory frameworks applicable to AI research.	Journal Reference, Chapter 4
The ethical questions roadmap for any AI protocol	Journal Reference, Chapter 1
Training the Ethics Committee for the practical evaluation of AI protocols.	Parts V and VI this MARIAH

Source: author's own elaboration.

4 PART III — VERSION A (QUALITATIVE)

MARIAH — Version A (Qualitative)

MARIAH — ARTIFICIAL INTELLIGENCE RISK ASSESSMENT MATRIX — VERSION A
(QUALITATIVE)

AXIS 1: NATURE AND AUTONOMY OF THE SYSTEM

The axis articulates the concept of model influence (FDA, 2025) with the concepts of potential for failure and opportunity for contestation developed by the Ministry of Security Food and Drug Administration of South Korea (MFDS, 2023).

Axis 1 Result: 0 risk
responses ÿ Level I; 1 to 2 risk
responses ÿ Level II; 3 to 4 risk responses
ÿ Level III; 5 or more risk responses ÿ
Level IV

AXIS 2: IMPACT ON THE PARTICIPATING PERSON

The axis is based on Brazilian legislation regarding research with human beings (Brazil, 2024a, 2025) and incorporates the participant's right to request the deletion of their data from training of the model, established by the Indian Council of Medical Research (ICMR, 2023).

Axis 2 Result: 0 risk
responses ÿ Level I; 1 to 2 risk
responses ÿ Level II; 3 to 4 risk responses
ÿ Level III; 5 or more risk responses ÿ
Level IV

AXIS 3: DATA SENSITIVITY AND GOVERNANCE

The central theme is anchored in the LGPD (Brazilian General Data Protection Law) and the guidelines of the Indian Council of Medical Research. (ICMR, 2023) on data governance in artificial intelligence research.

Axis 3 Result: 0 risk
responses ÿ Level I; 1 to 2 risk
responses ÿ Level II; 3 to 4 risk responses
ÿ Level III; 5 or more risk responses ÿ
Level IV

AXIS 3.b: RESEARCH DATABASES (CNS Resolution No. 738/2024)

Note: Axis 3.b is a complementary subsection to Axis 3, activated by the database filter in Step 0.

This subsection applies when the protocol uses databases—either its own or from another company. research or from an external institution —, as per the filter activated in Step 0 (Part II of this MARIAH). The questions are derived from CNS Resolution No. 738/2024 and complement Axis 3. without duplicating it. The result of this subsection may raise the consolidated level of the protocol (see Consolidation rule at the end).

Result of Axis 3.b:

0 risk responses ÿ does not raise the consolidated level

1 to 2 risk responses ÿ leads to Level III

3 or more risk responses ÿ raises to Level IV

Eliminatory hypothesis: answer "no" to question 3.b.2 (Term of Consent)

(Institutional absence in a database outside the scope of the research) constitutes an absence of the chain of Formal custody has been established. The protocol is not subject to merit review and proceeds to mandatory due diligence. before any analysis, as per the item "The critical rule" of the Reference Notebook, Chapter 8.

AXIS 4: REPRESENTATIVENESS AND TRANSFERABILITY

The framework is based on the Australian guide (NHMRC, 2023) and the MFDS guidelines. Korean (Australia, 2023) and in the framework of the MRCT Center and WCG (2025) on representativeness of models in diverse population contexts.

Result of Axis 4:

0 risk responses ÿ Level I

1 to 2 risk responses ÿ Level II

3 to 4 risk responses ÿ Level III

5 or more risk responses ÿ Level IV

AXIS 5: HUMAN SUPERVISION AND EXPLAINABILITY

The central tenet is based on the principle of irrevocable medical autonomy as stated by Indian Council of Medical Research (ICMR, 2023) and the concept of maintaining the cycle of life developed by the FDA (2025), operationalized by the agency's IDATEN system. Japanese regulatory protocol (PMDA, [n.d.]) and the Algorithm Change Protocol (Algorithm The Indian Change Protocol (CDSCO, 2025) is reinforced by the 2026 recommendations of the The UK's National Commission into the Regulation of AI in Healthcare, which emphasizes

Life cycle proportional supervision, continuous monitoring, equity as a dimension
security, traceability and change management.

Result for Axis 5:

0 risk responses ÿ Level I

1 to 2 risk responses ÿ Level II

3 to 4 risk responses ÿ Level III

5 or more risk responses ÿ Level IV

MANDATORY CROSS-CUTTING MODULE — LIFE CYCLE, CHANGES AND PERFORMANCE

The questions below complement Axes 1, 4, and 5 and should be answered when applicable. They do not constitute a separate sixth axis and do not alter, in this version, the lanes. Classification. Responses that reveal risk should be recorded in the corresponding section. and considered in the consolidation rule and in the assessment of the sufficiency of safeguards.

1. The protocol clearly identifies the AI system and, where relevant, its version.
Or relevant configuration?
2. The protocol indicates whether the system can be updated, recalibrated, retrained,
Replaced or modified during the research?
3. When there is a possibility of change, is there a procedure to assess whether the
The change modifies risks, benefits, equity, privacy, autonomy,
Human supervision or other safeguards?
4. The protocol defines proportionate mechanisms for monitoring the performance of
AI during research to detect flaws, unexpected results, or deterioration.
relevant?
5. Are there predefined criteria or procedures for review?
suspension, interruption or adoption of additional measures when performance
Or will the risk profile change?
6. The protocol assesses the possibility of differential performance between groups.
What are the relevant subgroups or contexts for the research population?
7. It is defined who exercises human supervision, with what authority, and how.
What happens when there is a discrepancy between the AI output and human judgment?
8. When proportionate to the risk, it is possible to trace later which system,
The version or configuration contributed to the result, inference, or decision.
relevant?

Warning sign: system with significant impact on participants lacking version identification when necessary, lacking a change management mechanism, lacking proportional performance monitoring, or lacking effective human supervision.

CONSOLIDATION RULE — SEQUENTIAL FLOWCHART

The CEP (Center of Proportionality) traverses the five axes in order. The final level is the highest one identified in... any axis. A single axis at Level IV raises the entire protocol to Level IV. Figure 1 Table 7 presents the consolidation flow for Version A, and the consolidation criteria for... risk level.

Figure 1 — Consolidation flow of MARIAH Version A



Source: own elaboration, based on MARIAH's consolidation rule.

Table 7 — Criteria for consolidating the risk level of Version A

Consolidated level	Criterion
I Low	All axes at Level I
II Moderate	At least one axis at Level II; none above.
III High	At least one axis at Level III; none at Level IV.
IV Critic	At least one axis at Level IV

Source: author's own elaboration.

5 PART IV — VERSION B (QUANTITATIVE)

MARIAH — Version B (Quantitative)

MARIAH — ARTIFICIAL INTELLIGENCE RISK ASSESSMENT MATRIX — VERSION B
(QUANTITATIVE)

Scoring system: each question answered as a risk adds points to the total.

The level is determined by the final sum. Mitigation questions, when answered.

Positively, they subtract points.

Level ranges:

- Without database (ceiling 275): Level I — 0 to 58 points; Level II — 59 to 127 points; Level

Level III — 128 to 208 points; Level IV — 209 to 275 points.

- With database (Block 6.b activated; theoretical ceiling 304, assessable ceiling 297)

(excluding elimination round P6.b.2): Level I — 0 to 64 points; Level II — 65 to 141 points; Level

Level III — 142 to 230 points; Level IV — 231 to 304 points.

Methodological note: the cuts are derived from the fixed fractions of the *baseline* (I = 50/238, II =

110/238, III = 180/238) applied to the current theoretical ceiling, rounded off.

half a unit upwards.

STEP 0: INPUT FILTER

Before starting the scoring, the postal code checks:

Does the AI system perform decision automation, generate content that integrates the design, data collection instruments, informed consent form, or results, or intervene in the protocol or conduct of the study? (borderline cases — editorial assistance, prognostic models — in Step 0 of the Reference Notebook, Chapter 8)

Not for routine use. The matrix does not apply.

Yes ÿ proceed to the characterization of the context of use.

CHARACTERIZATION OF THE CONTEXT OF USE

The eight questions in this stage are descriptive, do not award points, and are mandatory before scoring.

What question was the AI system developed to answer in this protocol?

Does the system answer this question autonomously, or does it support a decision made by a human?

What is the identification status of the personal data used in this protocol? (Identifiable; pseudonymized; anonymized; combination by study stage)

What is the approximate volume of participant records used in this protocol?

At what point does data anonymization occur in relation to access by researchers? (before access, by the custodian institution; after access, by the researchers themselves; there will be no anonymization; not applicable) — answer only if the previous answer is not "anonymized".

What is the role of the AI system in this protocol? (analysis tool for research data; development object — model training; both)

What is the intended use of the AI model used or developed in this protocol? (restricted use within this protocol; future applications or broad use; sharing with third parties; not applicable — AI as a tool only)

Will the database be shared, in whole or in part, with third parties other than the institutions participating in the protocol? (no; yes, under a Consent Form or Institutional Agreement; yes, another form — describe)

BLOCK 1: PROJECT

Maximum score: 22 points | Value per question: 3 points

Assess the suitability of the protocol for the use of AI: transparency, team qualifications, and clarity of the proposal.

Subtotal Block 1: _____ / 22 pts

BLOCK 2: SYSTEM

Maximum score: 17 points | Value per question: 2 to 3 points

Evaluate the technical nature of the system: autonomy, adaptability, and traceability.

Subtotal Block 2: _____ / 17 pts

BLOCK 3: ALGORITHM

Maximum score: 14 points | Value per question: 2 points

Evaluates the quality, origin, and representativeness of training and test data.

Subtotal Block 3: _____ / 14 pts

BLOCK 4: DECISION

Maximum score: 8 points | Value per question: 2 to 3 points

It assesses the degree of decision-making autonomy of the system and the consequence of an error: the questions more critical of the parent company.

Subtotal Block 4: _____ / 8 pts

BLOCK 5: IMPACT

Maximum score: 62 points | Value per question: 5 to 6 points

Assesses the direct impact on the participant. The most ethically weighted block in the matrix.

Subtotal Block 5: _____ / 62 pts

BLOCK 6: DATA

Maximum score: 77 points | Value per question: 5 to 7 points

Assesses the sensitivity, governance, and legal compliance of data processing. The most heavily weighted regulatory component of the framework, anchored in the LGPD (Brazilian General Data Protection Law).

Subtotal Block 6: _____ / 77 pts

BLOCK 6.b: RESEARCH DATABASES (CNS Resolution No. 738/2024)

Note: Block 6.b is a complementary subsection to Block 6, activated by the database filter in Step 0.

Maximum score: 29 points | Value per question: 5 to 7 points

This subsection applies when the protocol uses a database (Step 0 filter). It derives from CNS Resolution No. 738/2024 and complements Block 6. The score is added to that of Block 6 for the final calculation; the absence of a positive response in items marked as disqualifying subjects the protocol to the critical rule (Reference Booklet, Chapter 8, item "The critical rule").

Subtotal Block 6.b: _____ 29 points (added to Block 6)

Eliminatory hypothesis: a "yes" response to risk in P6.b.2 (Institutional Consent Form absent in a database outside the scope of the research) renders the protocol unevaluable on merit, according to the critical rule (Reference Booklet, Chapter 8, item "The critical rule"), regardless of the total score.

BLOCK 7: MITIGATION

Maximum risk score: 75 points | Risk questions add up | Mitigation questions subtract

Evaluates the quality of human supervision mechanisms, explainability, and life cycle control. Unique block with bidirectional logic.

Sub-block 7A: Consultations (up to +10 risk points)

Sub-block 7B: Risk reduction measures (up to +65 risk points / up to -65 points due to mitigation)

Sub-block 7C: Evidence of transparency (up to -23 points, deduction only: documented evidence subtracts points, and its absence does not add any)

Subtotal Block 7: _____ pts (positive if mitigation shortcomings outweigh the measures adopted).

Cross-sectional life cycle verification — Version B

Before final consolidation, it is recommended that the evaluator apply the eight questions from the Cross-Sectional Module of Version A. The findings should be recorded in Blocks 2, 3, 4, 5, or 7, depending on the dimension affected. The module does not create independent scores in this version, preventing changes to the quantitative ranges without empirical calibration; however, structural failures in supervision, change management, traceability, or monitoring may trigger the critical rule when they compromise the protection of the participating individuals.

FINAL CONSOLIDATION

Risk level:

Classification by range (ceiling 275): Level I — 0 to 58 pts; Level II — 59 to 127 pts; Level III — 128 to 208 pts; Level IV — 209 to 275 pts. With Block 6.b activated, the recalibrated ranges apply (ceiling 304; assessable ceiling 297): Level I — 0 to 64 pts; Level II — 65 to 141 pts; Level III — 142 to 230 pts; Level IV — 231 to 304 pts.

Critical rule: the level determined by the score does not change at the evaluator's discretion. Mitigations are already incorporated into the Block 7 score and constitute the only legitimate way to reduce the final score. Exception — Ethical Precedence Clause: an affirmative answer to questions P4.1 or P4.2 (decision without human review or irreversible harm to the participant) triggers the precautionary principle, allowing the Ethics Committee to immediately raise the protocol to Level IV, regardless of the total score. The final result after subtracting the Block 7 mitigations cannot be less than zero points.

Recalibration when Block 6.b is activated: when the Step 0 filter triggers the Block 6.b subsection (database), the maximum matrix score changes from 275 to 304 points, with the ranges recalibrated proportionally to preserve the relative distance between levels: Level I from 0 to 64 points; Level II from 65 to 141 points; Level III from 142 to 230 points; Level IV from 231 to 304 points (evaluable ceiling 297 points, excluding the eliminatory P6.b.2). Outside of this case, the standard ranges above apply (209 to 275 points for Level IV).

6 PART V — INSTRUCTIONS FOR USE

Instructions for Use of MARIAH

INSTRUCTIONS FOR USE

MARIAH — RISK ASSESSMENT MATRIX IN ARTIFICIAL INTELLIGENCE

Before we begin: which version to use?

The choice between the qualitative Version A and the quantitative Version B of the Assessment Matrix.

The Risk Assessment in Artificial Intelligence in Human Research (MARIAH) is not a choice.

between a rigorous instrument and a simplified instrument. Both versions start from the same

Based on ethical principles, they follow the same evaluation criteria and arrive at the same set of...

requirements by risk level. What differentiates them is the operating logic: Version A guides the

Sequential flowchart reasoning; Version B guides traceability through scoring.

A balanced choice. A postal code that chooses Version A is not making a less rigorous choice. It is making the choice.

appropriate to its institutional context and the profile of the protocols it evaluates.

Three questions guide this choice. The first is about institutional experience: the

CEP has consolidated cycles for evaluating AI protocols, with members familiar with the subject.

with the concepts of Chapter 2 of the Reference Notebook and with the logic of the five axes, or is

In the initial familiarization cycles with this type of protocol? Version A is the starting point.

Recommended entry for committees that are building this familiarity. The second is

Regarding the complexity of the protocol: the AI system described involves multiple architectures,

continuous learning, simultaneous use in different phases of research, or integration with

other systems, or is it a more limited use with a clearly described system?

Protocols with high technical complexity benefit from the granularity of Version B, which

It allows you to identify where the risk is concentrated and how the mitigation measures have reduced it. The third

It's about the institution's documentation needs: the institution needs to compare

protocols among themselves, produce auditable reports on committee decisions, or demonstrate

For sponsors and regulators, what is the numerical basis for approvals and rejections? Version B was

designed for this level of traceability. Table 8 compares the two versions.

Table 8 — Comparison between the qualitative and quantitative versions of MARIAH

	Qualitative Version A	Quantitative Version B
ZIP Code Profile	Initial cycles using AI or less complex protocols.	Experience consolidated or with highly complex protocols
Operating logic	Sequential flowchart, binary response per axis	Weighted score in seven blocks
Result	Risk level by combination of responses	Risk level by total sum with numerical traceability
Priority use	Collegial deliberation guided by reasoning.	Documentation auditable and comparing protocols

Source: author's own elaboration.

Both versions are compatible and can be used in combination. A postal code can adopt Version A as an initial screening and resort to Version B to deepen the analysis of protocols that Version A classifies as Levels III or IV. It can also start with the Version A and progressively migrate to Version B as members develop. Familiarity with the instrument. The choice is not irreversible. The instrument was designed to keep pace with the institutional development of the committee, not to fix it at a level of capacity.

Both versions also incorporate a specific subsection for protocols that They use databases — Axis 3.b in Version A and Block 6.b in Version B. This subsection This derives from CNS Resolution No. 738/2024 and translates the requirements regarding operational questions. Controller, Operator, Institutional Terms (Agreement, Consent, Commitment and Transfer) and waiver of informed consent for future use. The subsection is activated by the database filter. data in Step 0 and does not apply when the protocol collects data exclusively in scope of the research with each participant.

Applying the Qualitative Version A: Step by Step

Step 0 is the decision that determines whether Version A applies to the protocol under evaluation. The question is precise and must be answered before any other reading of the instrument: the AI system performs decision automation, content generation, or intervention in protocol or in conducting the study? Three types of use trigger the matrix. The first is the Decision automation: any system that produces classifications, scores, recommendations.

or alerts that influence decisions about participants or the conduct of the study. The

The second aspect is content generation: systems that produce text, synthetic data, and images.

or any other content that is part of the protocol, the informed consent form, the data collection instruments or the

research results. The third is intervention: systems whose output guides, modifies, or

It replaces steps in the protocol. A bibliographic organization tool does not trigger the...

matrix. A diagnostic support model is triggered. A synthetic data generator is triggered. A

natural language processing system that analyzes medical records to identify

Eligible participants trigger the process. Doubt about where to draw that line is not a sign of weakness.

from the evaluator. This indicates that the protocol should be taken to the plenary session before any decision is made.

Regarding the application of the matrix.

Two recurring questions deserve explicit answers. The first: the use of systems

Does generative manipulation in the improvement and elaboration of the project text trigger the matrix?

Declaration and action are distinct obligations. The declaration of the use of generative AI is...

always mandatory, by virtue of CNPq Ordinance No. 2,664/2026, whatever the purpose.

of the use. The matrix is triggered when the generated content integrates the research design, the

data collection instruments, the informed consent form, or the results — because in these cases the system participates.

in the production of knowledge and can introduce bias, inaccuracy, or misleading content into

which will be submitted to the Ethics Committee and what will be offered to the participant. The assistance strictly

Editorial work — grammar revision, style adaptation, translation, reference formatting —

It is declared, but it does not trigger the matrix because it does not alter the design or produce a finding.

The criterion for discernment is functional: the question is whether, with the system removed, the content

The submitted system would be essentially the same. The second question: where does the system fit in?

What identifies patterns that predict a disease's prognosis? In decision automation, because...

It produces classifications and scores about participants — and this is true even if the outcome is...

for research purposes only, without associated clinical practice, since the category reaches the

decisions about the conduct of the study, not just clinical decisions. When the same

The score also guides stratification, eligibility, or therapeutic conduct; it accumulates...

The intervention hypothesis is presented, and the protocol is evaluated by both parties.

A second filter operates in Step 0 in parallel with the first. The CEP checks if the

The protocol uses a database—either its own, from another research project, or from an external institution.

An affirmative response activates Axis 3.b, a subsection derived from CNS Resolution No. 738/2024, which

examines five operational questions: identification of the bank's Controller; Term of

Institutional approval from banks established outside the scope of the research (eliminary item);
 Data Usage Commitment Agreement; framing of any waiver of Informed Consent Form in
 applicable hypothesis according to the bank's origin; and formalization of joint controllership in
 Multicenter research by Institutional Agreement Term. A negative response leaves the
 The subsection is inapplicable, and the evaluation continues with the five main axes.

Characterizing the context of use is the most neglected and most...
 resulting from the entirety of Version A. Without it, the five axes produce readings.
 decontextualized reasons that can classify distinct systems at the same level for the same reasons
 incomparable. Before traversing any axis, the CEP needs to answer two questions.
 Regarding the system: what question was it developed to answer in this protocol?
 and the response is produced autonomously or supports a decision made by a being
 human?

The answer to these two questions situates the system within its specific ethical territory and
 The route is guided by the following axes. The route along the five axes presents situations in
 that the binary answer is not obvious. When the researcher describes the system as
 supervised, but it does not describe the supervision mechanism, the correct answer is risk:
 Supervision declared without a described mechanism is unverifiable supervision. When the
 The protocol mentions validation, but does not document performance testing by subgroup.
 population-based, the correct answer is risk: global validation without stratified analysis no
 ensures fairness. When the Informed Consent Form mentions AI, but in technical language that the participant...
 The average person wouldn't understand; the correct answer is risky: mention without comprehensibility.
 It satisfies the requirement for informed consent.

The result of Version A is a risk level determined by the combination of
 Responses across all five axes, with the critical rule applied: a single axis classified at Level IV.
 It elevates the entire protocol to Level IV, regardless of the other axes. Each level
 It establishes a set of minimum cumulative requirements that the researcher must meet.
 demonstrate compliance prior to approval. At Level I, a declaration of AI usage is required and
 Mention in the Informed Consent Form. At Level II, a description of the system and a demonstration of... are added.
 Compliance with LGPD. At Level III, technical validation documentation is added.
 Performance analysis by subgroups and verification of effective human supervision. No
 Level IV adds external technical opinion and a continuous monitoring plan with
 Discontinuation criteria and, where applicable, notification to Anvisa.

The CEP translates this result into precise opinion language: the protocol was classified at Level X of the Artificial Intelligence Risk Assessment Matrix, Version A, and does not meet the minimum requirements for the level in the following dimensions, with explicit indication Which requirements are missing and what is the deadline for completion? When the protocol It presents a structural flaw that no supplementary solution can resolve; the refusal is justified. at the identified risk level and in accordance with the principles of Law No. 14.874/2024. The critical rule is not Punishment is the precautionary principle embodied in the instrument: a system that fails in a A structural dimension does not become ethical simply because it performs adequately in other areas.

The critical rule also incorporates specific elimination hypotheses for protocols with databases. When the protocol uses a database created outside the scope of the research. without presenting Institutional Consent Form (art. 27, VI of CNS Resolution No. 738/2024), the The total score is disregarded and the protocol is not evaluated on its merits. The same logic applies if... applies to the absence of identification of the Controller (Articles 3, IV and 9) and to requests for exemption from Informed consent form without indication of the applicable legal basis and without ethical justification and sufficient methodological basis (art. 20, § 5, or art. 25). In these cases, the dossier is returned to researcher for mandatory due diligence prior to any merit review — this is not about Increased risk level, but assessment blocked due to lack of chain of control. Formalized custody. The critical rule is part of a broader set. MARIAH operates with four automatic safeguards, detailed in the Supplement to the Technical Note on Premises: a Axis 3.b's special rule, the Ethical Precedence Clause, eliminates the chain of custody. and the diligence preventing new consent. The special rule of Axis 3.b skips deliberately Level II. A single risk response on this axis leads the protocol to at least except Level III; three or more responses lead to Level IV. Preventive diligence reaches the Adaptive system without a new consent plan, verified in question 1.2 of Version And in questions P2.2 and P2.8 of Version B. In this case, the protocol is also returned as

Not assessable, subject to due diligence required by Law No. 14.874/2024 and the LGPD (Brazilian General Data Protection Law).

Applying the quantitative Version B: step by step

Version B operates on the same ethical basis and the same principles as Version A, but with Distinct logic: weighted scoring across seven blocks, total sum determining the level of risk and the ability to identify where the risk is concentrated. Before starting the scoring,

Step 0 is identical to that of Version A: the system performs decision automation, generation of Content or intervention in the protocol? If not, the matrix does not apply. If yes, the CEP proceeds. for characterizing the context of use, which is mandatory in Version B and does not award points, but It determines all subsequent scoring. The eight descriptive questions — what is the question what the system was developed to respond to, and whether the response is autonomous or subsidiary — These questions need to be answered before the first section.

The seven blocks cover different dimensions with different weights. Block 1 evaluates the project: transparency of the protocol, qualified team, and clarity of the proposal, with Maximum score of 22 points. Block 2 evaluates the system: autonomy, adaptability and Traceability, with a maximum score of 17 points. Block 3 assesses the life cycle: stage of development, monitoring and retraining plan, with a maximum score of 14 points. Block 4 assesses the nature of the supported decision: reversibility, autonomy. decision-making and clinical consequence, with a maximum score of 8 points. Block 5 assesses the impact on the participant, with a maximum score of 62 points. Block 6 evaluates Data sensitivity and governance, with a maximum score of 77 points. Block 7 It assesses the mitigation measures, using a bidirectional logic: sub-block 7A of consultations adds up to ten points. And the mitigations, which are worth 65 points, add up when absent and subtract when absent. Present, with a ceiling of 75 points for the block. The sum of the seven blocks produces a ceiling of 275. points of the matrix. Blocks 5 and 6 together concentrate 139 of the 275 maximum points (50.5% of the (total score). This concentration expresses the hierarchy of priorities of the instrument: the What matters most in research with human subjects is what the system can do for the person. Participant and how their data is handled.

When the protocol uses a database, an eighth set of items is activated: the Block 6.b, with a maximum score of 29 points, added to Block 6. The five items of Block 6.b derive from CNS Resolution No. 738/2024 and assess the identification of the Controller (seven points), Institutional Consent Form (seven points, disqualifying), Commitment Form Data Usage (five points), waiver of informed consent framework (five points) and Formalized joint comptrollership (five points). The subsection also includes three items of due diligence (P6.b.4.1, P6.b.6 and P6.b.7), which do not score points and operate as a cumulative *gate* : a The absence of any criteria implies the return of the protocol. The activation filter is the same. From Step 0: protocols that do not use databases have Block 6.b marked as no. applicable.

Block 7 deserves special attention because it operates differently from the others. It is the only block with bidirectional logic. Risk questions add up when the researcher does not adopt appropriate supervisory, explainability, or monitoring measures. The questions mitigation measures are subtracted when those measures are documented and verifiable. Subtraction is not automatic: it depends on the measures being accurately described in the protocol. It is sufficient for the postal code to evaluate them as genuine, not as declarations of intent. One protocol that states the system is supervised without describing the mechanism of supervision. The supervisor does not obtain the corresponding subtraction. The score for Block 7 is the only one that... The researcher can directly influence the study design choices. It is the block which rewards well-designed protocols and penalizes protocols that transfer risk to the user participant without documented countermeasures.

The sum of the seven blocks produces the classification into one of four levels. Level I goes from 0 to 58 points. Level II ranges from 59 to 127 points. Level III ranges from 128 to 208 points. Level IV ranges from 209 to 275 points. When the Step 0 filter activates the Block 6.b subsection (banks of data), the matrix ceiling becomes 304 points and the bands are recalibrated proportionally: Level I: 0 to 64 points; Level II: 65 to 141 points; Level III: 142 to 230 points; Level IV: 231 to 304 points (maximum possible score of 297 points, excluding the elimination P6.b.2). Each level determines the same minimum cumulative requirements described in the item "Applying the Qualitative Version A: step by step". The difference compared to Version A is that Version B allows the ZIP code to identify in which blocks the risk is concentrated in and communicate this information to the researcher with accuracy. A protocol classified at Level III with a concentration of scores in Block 5 and 6 receives a request for additional information that differs from a protocol classified in the same category. Level with concentration in Block 2. The granularity of the scoring makes the opinion more well-founded and the researcher most capable of understanding what needs to be demonstrated to obtain approval.

The critical rule of Version B is identical in principle to that of Version A, but operates by scoring: the final classification cannot be altered by the evaluator's subjective decision. Mitigation measures are already incorporated into Block 7's scoring and constitute the only legitimate approach to reduce the final score. Undocumented statements of intent do not replace evidence of effective risk control. The CEP that applies Version B and obtains a rating in Level IV cannot approve the protocol based on arguments that the score does not reflect. You can request additional information that allows the researcher to reduce the score by means of

Real mitigations. You can refuse with justification based on the score and requirements of the level. You cannot ignore the result of the instrument you chose to apply.

Documenting the decision

Applying MARIAH without recording the result in the evaluation process is applying the instrument only partially valid. The registration transforms the CEP's decision into jurisprudence. Institutional: the protocol evaluated today forms the repertoire that guides the evaluation of a similar protocol will arrive next semester. A committee that does not document its decisions regarding AI protocols start from scratch with each evaluation.

The minimum required record for any opinion piece that used MARIAH includes four elements. The first is identifying the version used: Qualitative Version A or Quantitative version B, indicating possible combined use. The second is the result from the application: the risk level obtained on each axis, in Version A, or the score per block and the total score, in Version B. The third is the justification for non-obvious decisions: when a risk response was assigned to a dimension where the protocol presented information. If the assignment is ambiguous, the opinion must record the reasoning that justified it. The fourth is translating the level into requirements: what are the minimum requirements for the level achieved by the protocol? Which aspects were satisfactory and which are absent, with a precise indication of what the researcher should do.

To complement the supplement, and within what timeframe?

Two specific situations require additional documentation. The first is when the postal code... decides to approve a protocol with reservations at a level that would normally require rejection or substantial supplementation. This decision needs explicit justification that demonstrate why the specific circumstances of the protocol justify the exception and what... Additional safeguards were required as a condition of approval. The second is when committee members disagreed on the application of the matrix to the same protocol. A disagreement should be recorded along with the positions and arguments of each side, even if the... The final decision must be unanimous. Documented disagreements are the richest material for the... Development of the institutional competence of the Ethics Committee: reveals where the instrument needs to be developed for further guidance and where the committee needs more in-depth deliberation.

When the matrix result justifies a rejection, the opinion must cite it. explicitly state the level achieved, the requirements not met, and the articles of Law No. 14.874/2024.

and Decree No. 12,651/2025, which support the unmet requirement. The refusal

Based on the matrix, it's not a technical decision of the instrument. It's an ethical decision of the committee.

supported by the instrument. This distinction is important for the researcher who may eventually

to appeal the decision and forward it to the body that hears the appeal. The CEP does not refuse because the

The score was high, but it was rejected because the protocol did not demonstrate the guarantees of protection.

The level of risk required of participants at that risk level, according to the principles established by law.

The Brazilian company establishes the plan, and the parent company implements it.

The MARIAH Instruction Manual concludes with a note about what the instrument does not do.

It does not approve or disapprove of protocols. It does not replace the judgment of the Ethics Committee. It does not dispense with the need for such decisions.

The collegial deliberation. It doesn't resolve borderline cases. What it does is make the evaluation more...

Systematic, better documented, and more justifiable. It makes visible what scientific enthusiasm...

It tends to obscure, and what the asymmetry of technical knowledge tends to suppress. The committee

whoever uses it rigorously does not delegate their responsibility to the instrument, but exercises that responsibility.

Responsibility with greater precision.

7 PART VI — LOCAL VALIDATION AND REVIEW

Local validation of MARIAH

Why validate MARIAH in your case studies?

Chapter 8 of the Reference Notebook explicitly positions MARIAH as This is a tool to support the deliberations of the Ethics Committees, and not a rule that binds their decisions. Adopting the matrix is an institutional choice for each Committee. This same spirit guides the Validation: No postal code is required to validate MARIAH, nor does it need to wait for a central decision. to do so. When the board deems it worthwhile to examine how the matrix behaves. As it pertains to the protocols that are actually submitted to it, this Party offers a roadmap. practical for conducting this verification.

There are three reasons why a postal code might want to perform this validation. The first is... Institutional learning. The systematic application of the matrix by two or more members of the The panel quickly reveals which questions receive divergent interpretations, which It provides direct input for the deliberative capacity of the collegiate body. The second concerns... Institutional qualification of the opinion. A CEP that demonstrates, based on its own The case law, which applies to MARIAH consistently among guarantors, strengthens credibility of their opinions in protocols with artificial intelligence before researchers, sponsors, and society. The third is an optional contribution to evolution. of the instrument. The current cut-off points of Version B are, according to Chapter 8 of the instrument itself. The Reference Notebook acknowledges theoretical calibrations that deserve to be tested. empirically. CEPs who wish to share their observations with the Working Group. Those responsible for the guide can do so through the channel described in the "When and how" section. "Share the results with the Working Group."

Before we get to the script, a warning is in order. This Part does not intend to transform CEP members in statistical roles. The metrics proposed in the item "How to interpret the "findings" are the minimum necessary for the board to interpret its own results with security, and the template spreadsheet that accompanies this Section calculates these metrics. automatically. What is expected from the CEP is a collegial deliberation on what the numbers They show, not the calculation of the numbers themselves.

This can be validated on three fronts.

This Part proposes three independent validation fronts. Each CEP can carry out One, two, or all three, in any order. All three share the same sampling unit: The research protocol using artificial intelligence submitted to the Ethics Committee.

Inter-rater agreement in Version A

The first step verifies whether two or more members of the CEP, when applying Version A From MARIAH to the same protocol independently, they arrive at the same classification. It is the The operationally simpler front, and the one that yields the best results from a learning perspective. Institutionally, the return is more immediate. The board quickly discovers in which areas there is... convergent reading and in which there is systematic disagreement, which directly guides what This needs to be discussed in a plenary meeting.

The recommendation is for at least twenty protocols to be evaluated in parallel by two CEP members, under conditions of mutual blindness, before any deliberation. collegiate body. Twenty is a prudent minimum to ensure stability in the agreement estimate. minimum; postal codes with a higher volume of AI-based protocols can expand this mark without This can be detrimental, and postal codes with lower volume can accumulate evaluations over time, without... deadline, until that milestone is reached.

Empirical distribution of protocols across the four levels

The second aspect examines how the protocols evaluated by the Ethics Committee are distributed across... There are four levels of MARIAH. There are two main uses for this information. The first is descriptive. The Ethics Committee gains an understanding of the profile of its research cases through intelligence. artificial: if most protocols focus on Level II, if there are protocols at Level IV, if The distribution has changed over time. This information is an input for planning. The institutional and capacity-building aspect for the collegiate body. The second use is diagnostic. When the The distribution is concentrated near a Version B cutoff point; the collegiate body has evidence. a place where small variations in scoring can shift many level protocols, the which justifies examining borderline cases more closely.

The operational recommendation is to record the total score and the score per block of all protocols to which the CEP applies Version B, without specific inclusion criteria. A

The template spreadsheet automatically generates the histogram and identifies nearby concentrations to the cutting points.

Convergence between Versions A and B in screening mode

The third aspect examines whether Version A and Version B, applied to the same protocol, are compatible. In the triage method outlined in Part V of this MARIAH, they converge to the same classification. risk. When the CEP routinely uses triage mode (Version A as an initial filter and Version B (in cases of Level III or Level IV) simply requires registering both classifications in parallel. Systematic discrepancies between the two versions, when they occur, are valuable information: or indicate that the postal code reads the two versions inconsistently, or indicate that the two Versions capture different dimensions of risk in the protocol. In both cases, the collegiate body We have concrete material to discuss and, if necessary, report to the Working Group.

How to manage each front.

The following script covers all three aspects in a single flow. Each postal code adapts the flow to... its dynamics.

The first step is to define the data collection period. It could be a quarter, a semester. or open. The important thing is for the CEP (Ethics Committee) to know that, during this period, the protocols with The artificial intelligence received will also be subject to validation exercises in parallel. to the ordinary opinion. The collection does not alter the flow of the institutional opinion of SINEP.

The second step is to form a pair or trio of evaluators. The recommendation is that at least two members of the board apply Version A independently, without prior consultation between them, and that a third member applies Version B separately. Evaluators can take turns with each protocol; what matters is the independence of the application. within each protocol. It is not necessary for the evaluator of Version B to be the rapporteur of the appear institutional.

The third step is to record the evaluations in the template spreadsheet. The spreadsheet has a tab. specific to each front. Each protocol receives an internal ZIP code identifier that

It dispenses with any information that would allow the identification of the researcher, proposing institution, or sponsor. The spreadsheet does not ask for this information. The minimum required information is the identifier, the date, and the responses from the evaluators to the questions in the matrix. The formulas calculate the ranking. Consolidated, the total score and the aggregated indicators of the three fronts.

The fourth step is a collegial deliberation on the findings. A meeting is recommended. dedicated, at the end of the collection period or periodically during it, so that the collegiate body Discuss what the indicators show. The spreadsheet generates a summary dashboard that supports this. discussion. Deliberation is the central point of the exercise; numbers are input, not... conclusion.

How to interpret the findings

Interpreting the findings does not require advanced statistical knowledge. Three proposed approaches, three basic readings are sufficient.

Regarding the level of agreement between raters, the main indicator is statistics. *kappa* (Cohen, 1960), calculated automatically by the spreadsheet. Table 9 below provides The practical reading of the observed values. The committee reads the *kappa* value per axis from Version A. and through consolidated classification, identifies the axes with low *kappa* and discusses, in plenary session, the cases of disagreement on those axes.

Table 9 — Practical reading of kappa values

<i>Kappa</i> value	Practical reading
Below 0.40	Insufficient agreement. The central theme or block deserves review by the board: the questions have divergent interpretations, and this is valuable information to discuss in plenary session.
Between 0.40 and 0.60	Moderate agreement. Identify cases of disagreement and use them as meeting agenda items, because it is precisely the borderline cases that contribute most to the formation of the committee.
Between 0.60 and 0.80	Substantial agreement. The application is consistent. Minor residual disagreements are expected and healthy.
0.80 or above	Near-perfect agreement. The application is stable. Caution, however: very high <i>kappa</i> values across all axes may indicate that the evaluators deliberate among themselves before recording the responses, which undermines the exercise.

Source: Adapted from Landis and Koch (1977).

For the empirical distribution front, the reading is visual. The panel analyzes the The spreadsheet histogram can be used to check the concentration of protocols at a given level. proximity to the cutoff points and the presence of bimodality (two peaks separated by (a valley). Concentration near the cutoff points is a warning that small variations in The application of the matrix produces level changes. Marked bimodality is local evidence of that the natural cutoff point in the case studies of the Ethics Committee may differ from the prevailing theoretical point.

The reading is tabular for forward convergence between versions. The spreadsheet generates a four-by-four contingency table that crosses the classifications of Version A and Version B. B in screening mode. High convergence occurs when the main diagonal concentrates more than eighty percent of the protocols. Patterns of disagreement, which require qualitative examination of the Corresponding protocols emerge when there is a focus on specific pairs outside of diagonal, as Level II in Version A and Level III in Version B.

The spreadsheet does not impose any interpretation on the board. The indicators are presented. with suggested reading, but the decision on what to do with them is an institutional decision of

POCKET.

When and how to share results with the Working Group

CEPs that decide to share their observations with the Working Group Those responsible for the guide can do so at any time through the Executive Secretariat. from INAEP (se.inaep@saude.gov.br). There is no deadline, there is no mandatory form, there is no institutional counterpart. Sharing is a voluntary act, and its sole purpose is To inform about the future evolution of the instrument.

The suggested channel is to send the completed template spreadsheet, with your identification details. institutional address of the CEP, to the Working Group's email address indicated on the page. Official guide. Delivery can be made in any format the postal code prefers: the spreadsheet. complete, a summary of the aggregate indicators, or just a narrative note describing What the panel observed. Ethics Committees that are concerned about confidentiality of Institutional information can forward only the aggregated indicators, without the... records by protocol.

The Working Group will consolidate the contributions received and use them collectively.
with other sources of evidence, to inform any revisions to the matrix in future cycles.
Guide update. Contributions will not be nominally attributed without authorization.
The sender's postal code is expressly provided, and no postal code will be identified in derivative publications without it.
institutional consent.

Available resources

This section includes a template spreadsheet in .xlsx format, with tabs dedicated to...
each of the three validation fronts, pre-configured formulas for calculating the
Indicators and summary panel with suggested reading. The spreadsheet can be used in any...
Compatible software, such as Microsoft Excel, LibreOffice Calc, or Google Sheets, and does not require...
Installation of any statistical extension.

The MARIAH software, a free electronic tool available at
<https://mariah-inaep.vercel.app> offers a complementary operational layer: a
Application of matrices directly in digital format, with report export.
standardized. Postal Codes that prefer to conduct the validation exercise in a digital environment can
to do so by exporting the reports and transferring the data to the spreadsheet.
This is a model of this part. The official software repository is located at github.com/Tonaco-13.

Operational questions regarding the application of this guide can be submitted, along with...
with comments and suggestions, through the channel indicated in the item "When and how to share"
"results with the Working Group".

REFERENCES

AUSTRALIA. National Health and Medical Research Council (NHMRC). Guide for assessing research involving AI. Canberra: NHMRC, 2023. Available at: <https://www.nhmrc.gov.au/about-us/resources/guide-assessing-research-involving-ai>. Accessed on: March 10, 2026.

BRAZIL. National Council for Scientific and Technological Development. Ordinance No. 2,664, of March 6, 2026. Establishes the Policy of Integrity in Scientific Activity of CNPq. Official Gazette of the Union: section 1, Brasília, DF, p. 4, March 11, 2026.

BRAZIL. National Health Council. Resolution No. 466, of December 12, 2012. Approves the guidelines and regulatory standards for research involving human beings. Brasília, DF: CNS, 2012. Available at: <https://conselho.saude.gov.br/resolucoes/2012/Reso466.pdf>. Accessed on: April 19, 2026.

BRAZIL. National Health Council. Resolution No. 738, of February 1, 2024. Provides for ethical obligations in the establishment and management of data repositories for scientific research. Brasília, DF: Official Gazette of the Union, 2024b.

BRAZIL. Presidency of the Republic. Decree No. 12,651, of October 7, 2025. Regulates Law No. 14,874, of May 28, 2024, which provides for research with human beings and establishes the National System of Ethics in Research with Human Beings. Official Gazette of the Union: section 1, Brasília, DF, edition 192, p. 3, Oct. 8, 2025. Available at: https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2025/decreto/d12651.htm. Accessed on: June 19, 2026.

BRAZIL. Presidency of the Republic. Law No. 14,874, of May 28, 2024. Provides for research with human beings and establishes the National System of Ethics in Research with Human Beings. Official Gazette of the Union: section 1, Brasília, DF, no. 103, p. 6, May 29, 2024a. Available at: https://www.planalto.gov.br/ccivil_03/_ato2023-2026/2024/lei/l14874.htm. Accessed on: March 23, 2026.

CANADÁ. Treasury Board of Canada Secretariat. Algorithmic impact assessment (AIA). [Ottawa: Government of Canada], 2025. Available at: <https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html>. Accessed on: March 10, 2026.

CARNEGIE ENDOWMENT FOR INTERNATIONAL PEACE. Understanding India's New Data Protection Law. Washington, DC: Carnegie Endowment, 2023. Available at: <https://carnegieendowment.org/research/2023/10/understanding-indias-new-data-protection-law>. Accessed on: March 11, 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/10.1177/00131644600200010>.

SOUTH KOREA. Ministry of Food and Drug Safety (MFDS). Guidance on clinical trials design of artificial intelligence (AI)-based medical devices. Seoul: MFDS, 2023. Available at:

<https://www.elendilabs.com/en/articles/kor-guidance-on-clinical-trials-design-of-artificial-intelligence-based-medical-devices>. Accessed on: March 11, 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.

INDIA. Central Drugs Standard Control Organization (CDSCO). Draft guidance on medical device software. New Delhi: CDSCO, Oct. 2025. Available at: <https://community.nasscom.in/communities/public-policy/cdsco-issues-draft-guidance-medical-device-software>. Accessed on: Mar. 11, 2026.

ÍNDIA. Indian Council of Medical Research (ICMR). Ethical guidelines for application of artificial intelligence in biomedical research and healthcare. New Delhi: ICMR, 2023. Available at: <https://www.icmr.gov.in/ethical-guidelines-for-application-of-artificial-intelligence-in-biomedical-research-and-healthcare>. Accessed on: March 11, 2026.

JAPAN. Pharmaceutical and Medical Device Agency (PMDA). Report on AI-based software as a medical device (SaMD). Tokyo: PMDA, [n.d.]. Available at: <https://www.pmda.go.jp/files/000266100.pdf>. Accessed on: March 11, 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. Available at: <https://mrctcenter.org/resource/framework-for-review-of-clinical-research-involving-ai/>. Accessed on: March 23, 2026.

WORLD HEALTH ORGANIZATION. Ethics and governance of artificial intelligence for health: guidance on large multi-modal models. Geneva: WHO, 2024. Available at: <https://www.who.int/publications/i/item/9789240084759>. Accessed on: March 23, 2026.