

MINISTRY OF HEALTH

National Ethics Committee for Research — INAEP

**GUIDE TO THE ETHICAL USE OF ARTIFICIAL INTELLIGENCE
IN RESEARCH WITH HUMAN BEINGS**

GUIDELINES FOR ETHICAL EVALUATION — VERSION 1.0

Brasília — DF

2026

Preparation, distribution and information:

MINISTRY OF HEALTH

Secretariat of Science, Technology and Innovation in Health

Department of Science and Technology

National Ethics Committee for Research

SRTV, block 702, via W5 Norte, PO 700 Building, 5th floor

Postal Code: 70719-040 – Brasília/DF

Tel.: (61) 3315-2425

Website: www.gov.br/saude

Email: inaep@saude.gov.br (*please confirm*)

Supervision:

Meiruze Sousa Freitas

Textual Elaboration:

Fabiano Tonaco Borges

Technical Review: Working Group of the Ethical Guidelines for the Use of Artificial Intelligence in Research with Human Beings of INAE. **Members:** Fabiano Tonaco Borges (Coordinator), Alexandre Dias Porto Chiavegatto Filho, Fábio de Oliveira Aquino, Felipe de Oliveira Franco, Luiz Vianna Sobrinho, Máira Araújo de Santana, Roseli Mieko Yamamoto Nomura and Wellington Pinheiro dos Santos.

The Guide was unanimously approved by the INAE Board during the 13th Ordinary Meeting of INAE, with Meiruze Sousa Freitas as rapporteur, in accordance with Vote No. 25/2026.

LIST OF PAINTINGS

Table 1 – How to interpret the language of this guide7

Table 2 – Elements of an AI system and ethical implications.....13

Table 3 – Families of AI systems and priority ethical issues.....14

Table 4 – International experiences and contributions incorporated into the Brazilian model 23

Table 5 – Summary of recommendations on scientific integrity, authorship and use of AI26

Table 6 – Function of the modules in the document set33

SUMMARY

	INTRODUCTION.....	5
1.1	EDITORIAL NOTE ON THIS REVIEW.....	7
1.2	HOW TO USE THE DOCUMENTARY SET	8
1.3	RELATIONSHIP WITH THE REFERENCE NOTEBOOK AND MARIAH.....	9
	PURPOSE, SCOPE AND APPROACH OF ETHICAL EVALUATION	10
2.1	PURPOSE	10
2.2	SCOPE OF APPLICATION.....	10
2.3	MODALITIES OF RESEARCH RELATED TO ARTIFICIAL INTELLIGENCE	10
2.4	CONTEXT-ORIENTED APPROACH	11
2.5	PROPORTIONALITY.....	12
	Essential Concepts for Evaluating AI Systems	13
3.1	SYSTEM, MODEL AND OUTPUT	13
3.2	TRAINING, VALIDATION AND TESTING	13
3.3	FAMILIES OF SYSTEMS AND ASSOCIATED RISKS.....	14
3.4	SYNTHETIC DATA	15
	CRITICAL RISK DIMENSIONS.....	16
4.1	BIAS, REPRESENTATIVENESS AND EQUITY	16
4.2	EXPLANABILITY AND TRANSPARENCY.....	17
4.3	EFFECTIVE HUMAN SUPERVISION.....	17
4.4	LIFE CYCLE, UPGRADE AND DRIFT	18
	4.4.1 Monitoring, changes and traceability	19
4.5	People, groups and communities affected beyond the formal participants.....	19
	Brazilian Context and Public Interest	21
5.1	Adaptation to the Brazilian Context.....	21
5.2	SUS AND DATA OF PUBLIC INTEREST.....	21
5.3	DIGITAL HEALTH, INTEROPERABILITY AND DATA SOVEREIGNTY	22
	INTERNATIONAL REFERENCES AND INCORPORATED PRINCIPLES.....	23
	SCIENTIFIC INTEGRITY, AUTHORSHIP AND USE OF AI IN RESEARCH PRODUCTION..	25
7.1	DECLARATION OF AI USE	25
7.2	Authorship and Human Responsibility	25
7.3	HALLUCINATION, FABRICATION AND NEGLIGENCE	26
7.4	What is recommended for the protocol — and what the CEP should verify ..	26
	DATA GOVERNANCE AND PARTICIPANT PROTECTION.....	27

8.1	ORIGIN, PURPOSE AND ROLES.....	27
8.2	DATABASES, FUTURE USE AND CONSENT.....	27
8.3	REIDENTIFICATION, INFERENCE AND SECURITY.....	28
8.4	INTERNATIONAL TRANSFER AND DATA SOVEREIGNTY.....	28
8.5	Privacy Preservation Technologies	28
8.6	RIGHTS OF PARTICIPANTS AND DESTINATION OF THE MODEL.....	29
	SPECIFICITIES BY NATURE OF THE RESEARCH.....	30
9.1	HEALTH SCIENCES: ASSISTANCE IMPACT AND SAMD.....	30
9.2	HUMAN AND SOCIAL SCIENCES: INFERENCES, GROUPS AND COLLECTIVE RISK	31
9.3	INTERDISCIPLINARY RESEARCH	32
	Integration with Mariah and Application Tools.....	33
10.1	APPLICATION OF MARIAH.....	33
10.2	RECORDING OF THE ETHICAL DECISION	33
10.3	Longitudinal Ethical Supervision and Events Requiring Further Review..	34
10.4	SHARED RESPONSIBILITIES AND LIMITS OF THE COMPETENCE OF THE CEP	34
10.5	USE OF ARTIFICIAL INTELLIGENCE IN SUPPORTING CEP ACTIVITIES	35

INTRODUCTION

Artificial intelligence (AI) is already integrated into different stages of research with living beings. Humans developed in Brazil, both in Health Sciences and in Human Sciences. and Social Sciences (CHS). Systems to support image analysis and diagnosis, predictive models, natural language processing tools, automated analysis methods texts, images and records, generative systems, classification and profiling techniques and Applications based on synthetic data are examples of technologies that may be present in protocols submitted for review by Research Ethics Committees (RECs).

The incorporation of these technologies expands scientific and methodological possibilities. from the research, but it also introduces specific questions for ethical evaluation. AI systems They can modify how data is collected, combined, processed, analyzed, and interpreted; to produce new inferences about people, groups, and communities; to influence decisions; expand the scale and speed of information processing; and enable uses data, models, or results that exceed the purpose initially intended in protocol.

These characteristics demand attention to dimensions that may require specific criteria and specific or complementary instruments of ethical evaluation, including the origin, the The quality and representativeness of the data; biases; the production of sensitive inferences; explainability and transparency; human oversight; the degree of autonomy of the systems; possibility of re-identification; individual and collective risks; data governance; and the The fate of the data, models, and results after the research ends.

It is in this context that the National Ethics Committee for Research (INAEP) presents this A guide designed to help researchers, institutions, and ethics committees identify, analyze, and... addressing ethical issues related to the use of artificial intelligence in research with Human beings in the Health Sciences and in the Human and Social Sciences.

The Guide is based on the principles and frameworks that guide ethics in research with living beings. Its purpose is not to replace humans and applies them to the particularities introduced by AI. legislation, applicable ethical standards, the powers of other authorities or deliberation by the Ethics Committees, but to offer guidelines, criteria and instruments that support the application of these benchmarks when the characteristics of the technology require evaluation. specific.

The use of artificial intelligence does not, in itself, constitute a uniform category of risk. The same technology can present different ethical implications depending on the purpose, population involved, nature and origin of the data, degree of autonomy of the system, the inferences produced, the influence of its results on decisions, and the potential consequences for participants, groups, and communities. For this reason, this Guide adopts an approach based on the context of use and the proportionality of risk.

The Guide acknowledges that the use of AI can raise distinct ethical questions in Health Sciences and Human and Social Sciences, without establishing a rigid separation between the fields. Interdisciplinary research can bring together characteristics of both areas. It is recommended that the evaluation be guided by the specific characteristics of the protocol and based on the risks actually involved, and not exclusively on the disciplinary area in which the research is classified.

This guide was developed for the Brazilian context and takes social diversity into account. The country's cultural, epidemiological, ethnic, linguistic, economic, and territorial characteristics. In this context, the Unified Health System (SUS) constitutes a particularly relevant context for research that use data, services, infrastructure or populations linked to the public system of health. The use of data from the SUS, as well as from other public, institutional or databases digital technology requires attention to individual protection and the potential effects of analytics. Classifications and inferences about groups, communities, and populations.

This Guide is both informative and operational in nature. Its objective is to support researchers and institutions in the development of protocols and to assist the Ethics Committees in the identification, analysis and proportionate assessment of the risks associated with the use of AI in research involving human subjects. It does not intend to transform researchers or members of ethics committees into experts in artificial intelligence, but establishing which information is ethically relevant, which questions need to be examined and what evidence may be needed to support them. ethical evaluation.

The purpose is to qualify and make more consistent, proportional and transparent the ethical judgment, without replacing it with automated or purely technical decisions. For To support this process, the Guide incorporates MARIAH — the Risk Assessment Matrix in Artificial Intelligence in Research with Human Beings — as a tool intended for to systematize the identification and classification of risks and to guide requirements proportionate to Protocol characteristics. MARIAH neither approves nor disapproves of research.

The set is organized into three complementary modules: Guide to Ethical Use of Artificial Intelligence in Research with Human Subjects — Guidelines for Ethical Evaluation; Reference Notebook — Ethical, Conceptual, Normative and Methodological Foundations; and MARIAH — Risk Assessment Matrix in Artificial Intelligence for Research with Living Beings Humans. It is recommended that the three modules be used in an articulated manner, as follows. the nature, complexity, and risk level of the research.

Artificial intelligence is a rapidly transforming field. For this reason, this Guide It is structured primarily around ethical issues and dimensions of risk that They remain relevant even when technology changes. Technology can support... Analysis is important, but it does not replace responsibility, deliberation, or ethical judgment.

1.1 EDITORIAL NOTE ON THIS REVIEW

This version reorganizes and rewrites the guidance text from the working document. from INAEP, which gave rise to this Guide, to provide greater institutional clarity and greater precision. regulatory reform and better separation between basis, requirement, and application instrument.

The operational content of the matrix has been removed from this Guide and concentrated in the module. autonomous MARIAH. The checklists, exercises and application tools integram a MARIAH.

Rule adopted:

The Guide establishes the necessary guidance and foundation for understanding and applying the requirements.

The Reference Handbook brings together the fundamental principles that support these guidelines, with Explanatory function, without creating additional requirements.

MARIAH translates these guidelines into verification tools and organizes the identification, classification, and documentation of risk. The ethical decision remains human, collegial, and well-founded.

To guide the interpretation of the recommendations presented in this Guide, the The expressions used should be understood according to the degree of requirement and the effects. expected, as shown in Table 1.

Table 1 – How to interpret the language of this guide

Expression	Use in the Guide	Expected effect
RECOMMENDED	A requirement arising from the applicable regulatory framework or a condition deemed necessary for the protocol to be ethically evaluable.	It is recommended that the protocol demonstrate the service or Justify, when legally permissible, the non-applicability.
RECOMMENDED	Good practice that strengthens protection, transparency, quality, governance, or traceability.	Not adopting a course of action does not automatically imply inadequacy, but may require justification in higher-risk contexts.
HE CAN	An acceptable alternative or option depending on the context.	Adoption is discretionary and it is recommended that it be compatible with the nature and risk of the research.

Source: author's own elaboration.

Foundation

The use of these expressions does not alter the hierarchy of laws, decrees, resolutions, and other applicable regulations. When the Guide translates a regulatory obligation into an operational requirement, it is recommended that the corresponding source be indicated in the final editorial version.

1.2 HOW TO USE THE DOCUMENTARY COLLECTION

For institutional guidance on ethical considerations, it is recommended...

Consult this Guide. For application of risk classification, *checklists* and registration of

For this decision, it is recommended to use MARIAH, either in its document version or as a tool.

Free access electronics. When there is a need to understand the rationale of

For a recommendation, whether regarding a risk axis or a safeguard, it is recommended to consult the Reference Notebook.

Figure 1 presents the recommended path for ethical evaluation of AI research.

Figure 1 — Recommended path for ethical evaluation of AI research.

Percurso recomendado de avaliação ética



A avaliação acompanha o ciclo de vida: alterações relevantes podem exigir nova apreciação.

Source: author's own elaboration.

Note: Figure created based on the life cycle approach adopted in this Guide and the guidelines. da Medicines and Healthcare products Regulatory Agency (MHRA), do Reino Unido.

The **Reference Guide** serves an explanatory function and does not establish additional requirements. for the ethical evaluation of the protocols.

1.3 RELATIONSHIP WITH THE REFERENCE NOTEBOOK AND MARIAH

The instructions below will guide you through the three sections.

- **Context of use, proportionality and critical dimensions** — MARIAH for operational application; Reference Notebook for conceptual and methodological foundation.
- **Scientific integrity and generative AI** — Reference booklet for further study; MARIAH for verification.
- **Data governance and participant rights** — Reference Handbook for normative and technical foundations; MARIAH for classification and documentation.
- **Health Sciences and SaMD** — MARIAH for risk factors; Reference Handbook for technical and regulatory support.
- **Human and Social Sciences** — MARIAH for sensitive inferences, collective risks and secondary uses; Reference Notebook for specific foundation.

PURPOSE, SCOPE AND APPROACH OF ETHICAL EVALUATION

1.4 PURPOSE

This Guide provides guidance for the design and ethical evaluation of research involving human subjects. that involve artificial intelligence systems or tools. The guidance applies to both to research in which AI is the object of investigation as well as to research in which it is used as a tool, method of analysis, decision component, generator of content or mechanism for producing inferences.

It is recommended that the evaluation focus on the concrete effects of the technology on The protocol concerns the people, groups, or communities potentially affected. the commercial name of the system, the technical architecture, or the disciplinary area, in isolation, They do not determine the level of ethical concern.

1.5 SCOPE OF APPLICATION

It is recommended to consider this Guide when research uses AI for, among other things. purposes:

- to analyze, classify, predict, or interpret data related to people;
- to produce, transform, or summarize content that integrates the protocol, the research instruments, consent, or results;
- to support or automate decisions that may affect participants;
- to develop, train, adjust, test, or validate AI models or systems;
- to produce synthetic data used as part of the scientific evidence;
- processing texts, images, audio, clinical records, administrative data or data originating from digital environments;
- to produce inferences about individual or collective attributes.

1.6 MODALITIES OF RESEARCH RELATED TO ARTIFICIAL INTELLIGENCE

For screening purposes and to define the information necessary for the evaluation, it is recommended- to identify whether the protocol involves: (a) health-related research that uses data and AI methods; (b) research conducted with AI tools or technologies; or (c) research

regarding the development, training, testing, validation, or performance of tools and AI technologies. The modalities can coexist within the same protocol and do not determine, in isolation, the level of risk.

It is recommended that the protocol include:

- **the type or types of AI use present in the study;**
- **the role played by AI at each stage of the research;**
- **whether AI acts on data, instruments, interaction with participants, production of inferences or decisions;**
- **whether the tool used was developed specifically for the research or provided by a third party.**

The postal code checks:

- if all relevant uses of AI have been declared, including those considered assistants;
- whether the reported modality corresponds to the use actually described in the protocol;
- if the framework allows for the identification of formal participants, people indirectly affected groups and communities potentially impacted.

1.7 CONTEXT-ORIENTED APPROACH

It is recommended that the protocol define the context of AI use before analysis. risk. The same technology can present different risks depending on its purpose, the population, data source, degree of autonomy, application environment, and possibility human intervention.

It is recommended that the protocol include:

- What is the specific purpose of AI in the study?
- what question or task does the system perform;
- at what stage of the research is AI used;
- who receives and uses the output produced;
- whether the outcome influences decisions regarding participants, groups, or communities;
- if human intervention occurs before any significant consequences arise.

It is recommended that the postal code check:

- whether the role of AI is described clearly enough to allow for evaluation. ethics;
- whether the actual usage context coincides with the context in which the system was developed or validated;
- if the system's influence on decisions has been made explicit;
- if the description allows identification of who may be affected by the use of AI.

Warning sign:

- The protocol mentions AI only in a generic way, without describing its function;
The researcher describes the system as 'supportive' without clarifying how much it influences the... decision;
- The use of AI appears only in a technical appendix, without integration with the ethical analysis of protocol.

1.8 Proportionality

It is recommended that the depth of the assessment be proportional to the risk. The use of AI should not, by itself, generate identical demands for all protocols. MARIAH is the This instrument is designed to support this classification and to document the risk factors. relevant.

Lower-risk protocols may only require clear identification of AI use.
Transparency and adequate data protection. Higher-risk protocols may require...
Contextual validation, subgroup performance analysis, effective human supervision,
Expanded documentation on governance, monitoring, and additional safeguards.

Basis:

The original document establishes that stable ethical questions—representativeness of data, human oversight, fate of the model, and protection of participants—are recommended to guide the evaluation even in the face of rapidly changing AI architectures. The revision maintains this rationale, replacing essayistic language with operational criteria.

Essential Concepts for Evaluating AI Systems

This chapter brings together only the concepts necessary for ethical evaluation.

A thorough technical understanding of architectures and algorithms is not a prerequisite for...

The role of the Ethics Committee (CEP). The goal is to allow researchers and reviewers to use common language. and identify the information needed for the analysis.

1.9 SYSTEM, MODEL AND OUTPUT

For the purposes of this Guide, an AI system processes data or information and produces outputs. such as classifications, probabilities, scores, recommendations, predictions, or content.

It is recommended that ethical evaluations consider the system in its context of use, and not only the isolated mathematical model.

Table 2 presents the basic elements of an AI system — input, Processing and output — and the ethical implications associated with each.

Table 2 – Elements of an AI system and ethical implications

Element	Essential regulatory question	Ethical implication
Prohibited	What data or information feeds the system?	It determines quality, representativeness, privacy, consent, and risk. reidentification.
Processing/model	How does the system transform inputs into outputs relevant to the research?	It determines opacity, auditability, generalizability, and risk of error.
Exit	What classification, prediction, recommendation, or content is produced?	It determines potential harm, inferences, decisions, and the need for supervision.

Source: author's own elaboration.

1.10 TRAINING, VALIDATION AND TESTING

When research develops or evaluates an AI model, it is recommended that the protocol Distinguish, as applicable, the data used for training, validation, and testing.

Performance presentation based solely on data used to train the system is not demonstrates the ability to generalize.

It is recommended that the protocol include:

- origin and characteristics of the datasets;
- selection and exclusion criteria;
- relevant pre-processing procedures;
- whether synthetic data was used;
- strategy for separating training, validation, and testing phases, when applicable;
- Performance metrics relevant to the context of use.

It is recommended that the postal code check:

- if the performance shown was obtained from independent data when this is necessary to support the proposed purpose;
- whether the data used are appropriate for the population and the context of the research;
- if limitations to generalization are made explicit.

1.11 SYSTEM FAMILIES AND ASSOCIATED RISKS

Table 3 presents the most frequent AI system families in protocols of Research, with examples and the priority ethical issues associated with each type of use.

Table 3 – Families of AI systems and priority ethical issues

Type of use	Examples in the source document.	Priority issues
Classification and prediction	risk models, diagnostic support, image classification	representativeness, validation, impact of the decision, supervision <small>And it's an error.</small>
Language processing	analysis of medical records, texts, documents and speeches	Privacy, inferences, contextualization, biases, and confidentiality.
Generative AI	text production, synthesis, generation of synthetic data	hallucination, scientific integrity, traceability, confidentiality, and human verification.
Adaptive systems	Models subject to updating or retraining.	Version control, drift, revalidation, and the need for a new ethical review.

Type of use	Examples in the source document.	Priority issues
Federated learning and privacy techniques	distributed data processing	Governance, residual re-identification risk, and inter-institutional liability.

Source: author's own elaboration.

1.12 SYNTHETIC DATA

When synthetic data is used, it is recommended that the protocol...

Clearly identify and describe your role in the study. Synthetic data should not be...

presented as real observational data. When they integrate development or

For system validation, it is recommended that the researcher demonstrate its suitability to...

scientific purpose.

Warning sign:

- synthetic data described only as an 'anonymization' mechanism without evaluation of its limits;
- Artificially generated results presented as empirical evidence collected from participants;
- absence of human verification of content produced by generative AI.

Critical Risk Dimensions

It is recommended that the evaluation of AI systems consider, in an articulated manner, bias, explainability and transparency, human oversight, impact on participants and groups, and System lifecycle. These dimensions constitute central foundations of MARIAH.

1.13 BIAS, REPRESENTATIVENESS AND EQUITY

It is recommended that the protocol assess whether the data and system performance are appropriate to the people, groups, and contexts affected by the research. Aggregate metrics They can mask unequal performance between subgroups. It is recommended that equity be ensured. also considered as a safety dimension: relevant performance differences They can pose concrete risks, especially when AI influences clinical decisions. care-related, selection, classification or prioritization.

It is recommended that the protocol include:

- composition and origin of the database;
- relevant characteristics of the population represented;
- Performance evaluation by subgroups when relevant;
- known limitations in representativeness;
- measures planned to reduce or monitor disparities.

It is recommended that the protocol check:

- whether the study population is comparable to that used in the development or validation;
- if potentially vulnerable or underrepresented subgroups were considered;
- whether the protocol distinguishes between absence of evidence of bias and evidence of absence of bias.

Warning sign:

- presentation of overall accuracy only;
- imported model without justification for its transferability to the Brazilian context;
- absence of subgroup analysis in research whose results may distribute benefits or risks in an unequal way.

1.14 EXPLANABILITY AND TRANSPARENCY

The depth of explanation required depends on the function of the AI and its consequences. of their departure. Ethical evaluation does not require unrestricted access to the source code in all cases. but it requires sufficient information to understand the role of the system, its limitations and the conditions under which their exits may be challenged.

It is recommended that the protocol include:

- level of explainability available and appropriate to the context of use;
- known limitations of the explanatory techniques used;
- information necessary for the human responsible party to interpret and contest the departure;
- a form of communication to participants when AI has a significant impact on their participation or their data.

Basis:

The source document emphasizes that salience maps and other explanatory techniques do not demonstrate causality and are not, in themselves, equivalent to clinical or ethical interpretability. It is recommended that the explanation be tailored to the user and the context.

1.15 Effective Human Supervision

The formal presence of a responsible person is not sufficient when they lack the necessary resources. information, competence, time, authority, or the real possibility of disagreeing with the system. It is recommended that human supervision be effective and commensurate with the magnitude of the consequences. It is recommended that the protocol specify who is responsible for supervision, and under what circumstances. When does it occur, what information supports the human review, and how does it proceed? Given the discrepancy between the algorithmic output and human judgment, considering including the risk of automation bias.

It is recommended that the protocol include:

- who performs the supervision;
- What qualifications are needed?
- at what point does the review take place;
- how disagreement is recorded and handled;

- what happens when the system fails or produces a result incompatible with other evidence.

It is recommended that the postal code check:

- if there is a real possibility of intervention before significant consequences occur;
- if the human responsible party sufficiently understands the limitations of the system;
- if the operational design favors contestation rather than automatic acceptance.

Warning sign:

- supervision described only as 'responsible physician' or 'responsible researcher', without a review process;
- a decision of high consequence with no real possibility of contestation;
- Use of AI in contexts of overload or automation that favor automation bias.

1.16 LIFE CYCLE, UPGRADE AND DRIFT

The ethical acceptability of an AI system should not be treated as property.

Static. Performance may vary depending on population, data, workflow, and environment.

institutional and actual conditions of use, and may change over time. When the

If the system changes during the research, it is recommended that the protocol be established.

identification and version control, monitoring criteria, mechanisms for detection

failures or deterioration in performance and procedures for updating, recalibration,

retraining or replacement. It is recommended that changes capable of altering risks,

benefits, equity, privacy, autonomy, human oversight, or consequences of error

They should be evaluated regarding the need for communication or amendment to the postal code.

It is recommended that the protocol include:

- System version used;
- whether the system is static or adaptive;
- events that can alter your performance;
- minimum performance criteria;
- procedures in case of data drift or performance issues;
- the system's destination upon completion of the search.

1.16.1 Monitoring, changes and traceability

It is recommended that ethical review accompany the AI lifecycle within the protocol. When proportionate to the risk and purpose of the research, it is recommended that the protocol Provide mechanisms that allow you to identify which system, version, or configuration was used. used; monitor performance during execution; recognize unexpected results or significant deterioration; and establish previously defined responses when new Information can modify the risk profile.

It is recommended that the CEP verify if there are any technical or contextual changes. seemingly limited factors can modify the interaction with ethically relevant interactions. participants, decision-making autonomy, performance in specific groups, supervision human, the risk-benefit ratio, or the safeguards provided.

The existence of registration, authorization, certification, or other regulatory framework. The system's implementation does not replace an ethical evaluation of its use within the specific protocol.

1.17 People, groups and communities affected in addition to the formal participants

It is recommended that ethical evaluations consider not only the individuals whose data is... directly involved with or interacting with the research team, but also those that can be classified, profiled, represented, or affected by inferences, decisions, subsequent products or uses of the system. The absence of a formal framework such The participant's failure to do so does not ethically disregard the foreseeable risks of discrimination. stigmatization, exclusion, surveillance, manipulation, misrepresentation or distribution unfair balance of benefits and burdens.

It is recommended that the protocol include:

- who may be directly or indirectly affected by the outputs, inferences, or reuses of the system;
- what individual, collective, social, cultural, economic, or territorial damages are reasonably foreseeable;
- how potentially affected groups were considered in the design, validation, and communication of the research;
- What measures reduce the risks of stigmatization, discrimination, and unfair profiling? and inadequate representation.

The postal code checks:

- if the risk analysis goes beyond individual protection and reaches collective impacts and belonging to the company;
- if public, anonymized, or secondary data have been improperly treated as ethically neutral;
- whether the research could shift risks to groups that did not consent or have no channel. right of appeal;
- whether the claimed social value is compatible with the distribution of risks, benefits and power decisional.

Warning sign:

The protocol states that there are no participants only because it uses public data. secondary or anonymized;

Sensitive inferences or collective classifications are produced without identifying the... affected groups;

- Predictable social or cultural impacts are treated exclusively as a data protection issue.

Brazilian Context and Public Interest

It is recommended that the ethical evaluation of AI research consider the context. Brazil, its population, territorial, social and epidemiological diversity and inequalities which can affect the production and interpretation of the data.

1.18 Adaptation to the Brazilian Context

It should not be assumed that systems developed or validated in other contexts They should be suitable for the Brazilian population. It is recommended that the protocol justify the... Transferability of the model or presentation of validation compatible with the proposed purpose.

It is recommended that the postal code check:

- if the population and development environment are comparable to the context of search;
- whether differences in infrastructure, language, culture, epidemiology, or composition Population data can alter performance;
- whether there is a risk of reinforcing inequalities already present in the data.

1.19 SUS AND DATA OF PUBLIC INTEREST

In research that uses linked data, services, infrastructure, or populations Regarding the Unified Health System (SUS), it is recommended that the use of AI take into account the protection of participants, data governance, research purpose, equity, and interests public.

The scope and diversity of data produced within the SUS (Brazilian Public Health System) expand the potential. scientific research and, simultaneously, increase responsibility regarding its use. Sharing, reuse, and destination of products derived from this data.

Warning sign:

- Transfer of data or models derived from public data without defining purpose, responsibility, and destination;
- use of foreign or third-party infrastructure without a description of the supply chain processing;

- lack of impact analysis on underrepresented groups or territories.

1.20 DIGITAL HEALTH, INTEROPERABILITY AND DATA SOVEREIGNTY

When research is embedded within digital health ecosystems, platforms
For interoperable or national data networks, it is recommended that the protocol explicitly define the flows.
data and relevant institutional responsibilities. Interoperability does not eliminate the
obligations of purpose, security, traceability and protection.

INTERNATIONAL REFERENCES AND INCORPORATED PRINCIPLES

The international analysis performed on the source document contributes to the construction of the criteria used in this Guide and in MARIAH. In this regulatory version, the country description The country is replaced by a synthesis of the contributions incorporated into the Brazilian model.

Table 4 summarizes the contributions of international experiences incorporated into This guide is for MARIAH.

Table 4 – International experiences and contributions incorporated into the Brazilian model

Reference/experience	Incorporated contribution	Application in the Guide/MARIAH
OMS	Ethical governance, human oversight, participant protection, and attention to system limitations.	General principles; supervision; transparency; accountability.
Canada	algorithmic impact assessment and separation between risk and mitigation	proportional structure; traceability; critical rule
Australia/NHMRC	guidelines for evaluating research involving AI	adequacy of the information presented to the committee and proportional evaluation
United States/FDA	context of use, influence of the model and life cycle	characterization of use; monitoring; influence of AI in decision making
European Union and data protection	Governance, transparency, and data protection in AI systems.	data, purpose, risks and documentation
Asia — references analyzed in the original	Contestability, performance, participant protection, and device evaluation.	Supervision, validation, rights and requirements by context.

Source: author's own elaboration.

These references do not replace the Brazilian framework. They serve as sources of... Comparison and methodological support. The requirements applicable to the protocol must be based on current Brazilian regulations and the responsibilities of the ethics system. in research.

It is recommended that the postal code check:

- whether the international reference cited by the researcher is compatible with the context
Brazilian;
- if certification, approval, or performance in another country is being improperly assessed.
used as a substitute for local ethical review;
- whether safeguards adopted abroad have been adapted to the population and standards
Brazilian women.

Scientific integrity, authorship, and the use of AI in research production.

The use of AI in the design, analysis, or documentation of research does not replace the human responsibility. The principal investigator remains responsible for integrity of the protocol, data, analyses, references, and results presented.

1.21 AI Usage Declaration

It is recommended that the protocol declare the use of AI when it contributes in a way that... relevant for content production, analysis, data generation, hypothesis formulation, development of instruments or other steps that may affect scientific integrity or the protection of the participants.

It is recommended that the protocol include:

- tool or system used, when relevant;
- purpose of use;
- stage of the research in which it was used;
- human responsible for review and validation;
- measures to verify references, facts, calculations, data, or generated content.

1.22 Authorship and Human Responsibility

AI systems should not replace the attribution of human responsibility for... scientific content. Ethical and scientific responsibility cannot be transferred to the system used.

It is recommended that the protocol check:

- whether there is a person responsible for each relevant step of the protocol;
- whether the team has the expertise to verify the AI outputs;
- if the use of AI does not obscure the chain of responsibility.

1.23 HALLUCINATION, FABRICATION AND NEGLIGENCE

It is recommended that AI-generated content be verified before integration. protocol or materials submitted to the Ethics Committee. Non-existent references, statements Incorrect regulations, synthetic data presented as factual, and unverified hypotheses. They can compromise the scientific validity and, consequently, the ethical justification of Exposure of participants to risks.

Warning sign:

- unverifiable bibliographic references;
- normative statements without an official source;
- AI-generated consent text without substantive review;
- Artificially generated data or images without clear identification;
- Delegation of scientific or ethical judgment to the tool.

1.24 WHAT IS RECOMMENDED FOR THE PROTOCOL — AND WHAT THE CEP SHOULD VERIFY

Table 5 summarizes what is recommended for the protocol to include and what the Ethics Committee (CEP) requires. It verifies the scientific integrity, authorship, and use of AI in the production of the research.

Table 5 – Summary of recommendations on scientific integrity, authorship, and the use of AI.

Her	It is recommended that the protocol present	It is recommended that the postal code be verified.
Use of AI	declaration of the tool, function, and step.	whether the statement is complete and consistent with actual use.
Human verification	responsible party and review process if the	review is substantive and enough.
Synthetic data	identification, purpose and validation	unless they were mistaken for real data.
References and standards	conference mechanism	if the relevant sources are verifiable.
Responsibility	person responsible for decisions and content	if the chain of responsibility remains human and defined.

Source: author's own elaboration.

DATA GOVERNANCE AND PARTICIPANT PROTECTION

Data governance is a central dimension of the ethical evaluation of research with AI. It is recommended that the protocol allow for the identification of the data's origin and purposes, treatment, the responsible agents, the sharing flows, the mechanisms of security and the destination of the data and derived models.

1.25 ORIGIN, PURPOSE AND ROLES

It is recommended that the protocol include:

- origin of each dataset;
- intended use in the study and training or adjustment of the system;
- Controller and Operator, when applicable;
- institutions and third parties that will have access;
- infrastructure used for storage or processing;
- destination of the data, parameters, and model after the search.

1.26 DATABASES, FUTURE USE AND CONSENT

When research uses databases, it is recommended that the protocol demonstrate... The applicable framework depends on the bank's origin, the intended use, and the regulations. Relevant Brazilian laws. Possible exemption from obtaining or providing new documentation. consent must indicate the applicable legal basis and provide an ethical justification and methodological, without being presumed solely based on the bank's origin.

It is recommended that the postal code check:

- whether the applicable institutional instruments have been presented;
- whether future use is covered by prior consent or by the normative hypothesis invoked;
- whether the anonymization or pseudonymization described is adequate to the risk of re-identification;
- whether the purpose of the research is compatible with the intended use.

1.27 REIDENTIFICATION, INFERENCE AND SECURITY

The absence of direct identifiers does not automatically eliminate the risk of re-identification. AI systems can combine variables and produce inferences capable of revealing sensitive information. It is recommended that the protocol consider the technical risk and contextual re-identification and inference.

Warning sign:

- use of the term 'anonymized' without describing the procedure;
- large datasets with multiple indirect variables capable of re-identification;
- Transfer to third parties without a defined purpose or responsibility;
- Use of an external AI service that retains or reuses protocol data.

1.28 INTERNATIONAL TRANSFER AND DATA SOVEREIGNTY

When there is international transfer or processing in infrastructure. If the protocol is located outside of Brazil, it is recommended that it describe the flow, purpose, and... agents involved, safeguards, and the fate of the data and the resulting model.

1.29 Privacy Preservation Technologies

Federated learning, differential privacy, homomorphic cryptography, and others. Privacy-preserving architectures can reduce certain risks, but not... They should be treated as absolute guarantees. It is recommended that the protocol describe the mechanism adopted and the relevant residual risk.

Basis:

The source document highlights that privacy techniques modify the architecture of data processing, but do not automatically eliminate risks of reverse account theft, re-identification, or institutional liability. The Ethics Committee assesses the adequacy of the safeguards in the context. concrete.

1.30 RIGHTS OF PARTICIPANTS AND DESTINATION OF THE MODEL

It is recommended that the protocol clarify how the rights of [people/entities] will be handled.

Participants during the research and, when relevant, after model training. A

Data retrieval may have technical limitations when the system has already been trained; it is recommended-

I hope that these limitations are explained and evaluated in a transparent manner.

It is recommended that the protocol include:

- how the participant can exercise their rights over their data;
- how will the eventual withdrawal of consent be handled;
- whether the model will continue to incorporate patterns learned from the retrieved data;
- if there is a mechanism for disposal, reprocessing or other applicable measure;
- what will be the final destination of the model and its parameters.

SPECIFICITIES BY NATURE OF THE RESEARCH

Specificities are not exceptions to the general assessment. They determine which Risk dimensions require greater attention. Research in Health Sciences and research in The Humanities and Social Sciences can share principles and, at the same time, demand... different questions

1.31 HEALTH SCIENCES: ASSISTANCE IMPACT AND SAMD

When AI can influence diagnosis, prognosis, treatment, monitoring, For triage or other healthcare-related decisions, it is recommended that the protocol describe the potential impact of system output and the role of human oversight.

When the system qualifies as Software as a Medical Device (SaMD), the Ethical review by the Ethics Committee and health assessment by Anvisa have complementary mandates. and not competitors. The assessment by the Ethics Committee does not replace the authority's competence. sanitary conditions, and the existence of a regulatory framework, do not replace ethical evaluation of search.

In AI systems with an impact on healthcare, it is recommended that the Ethics Committee also consider... whether the performance has been demonstrated in a relevant population and context; whether there is a possibility differential performance between groups; whether human supervision is effective; and whether the protocol It provides for monitoring and change management when the system can be updated during the study.

It is recommended that the protocol include:

- purpose of the system and its function in research;
- development or evaluation phase;
- potential clinical or healthcare impact;
- relevant performance in the context of use;
- mechanism for supervision and contestation;
- applicable regulatory framework, when relevant;
- plan for failure, unexpected outcome, or need for immediate intervention.

It is recommended that the postal code check:

- whether the healthcare decision remains adequately supervised; - whether the protocol distinguishes between research, healthcare use, and regulatory purposes; - whether the risks arising from error or malfunction have been assessed; - whether data governance requirements continue to be met.

Warning sign:

- system described as 'experimental' being used in an assistance-based manner without delimitation;
- High autonomy in making decisions with serious consequences without effective human review; - Use of performance obtained in a different population without contextual validation.

1.32 HUMAN AND SOCIAL SCIENCES: INFERENCES, GROUPS AND COLLECTIVE RISK

In the Humanities and Social Sciences, it is recommended that assessment not be limited to the nature of the subject matter. from the source data. Algorithmic processing can produce inferences about identity, behavior, relationships, beliefs, social position, health, belonging or other attributes that were not explicitly present in the collected data.

It is recommended that the protocol include:

- Origin of the data and access conditions; - What the system intends to infer, classify, or group; - Whether the inferences involve sensitive attributes; - Ethical and legal basis for the treatment of these inferences; - Potentially affected groups, including beyond formal participants; - Risk of stigmatization, discrimination, or misrepresentation; - Destination of the model and possibility of secondary uses.

It is recommended that the postal code check:

- whether consent or justification for exemption covers the inferences produced, and not just the source data; - whether publicly accessible data is being treated as ethically neutral without context analysis; - whether there are collective risks that are not captured by individual protection; - whether the model could be reused in a discriminatory or unforeseen context.

Warning sign:

- Public data processed to produce sensitive inferences without specific analysis;
- minimum risk classification based solely on the fact that the data is public or anonymized;
- lack of identification of the social groups affected by the inferences;
- Open publication of the model without evaluation of potentially secondary uses harmful.

In the Humanities and Social Sciences, it is recommended to examine specifically the contextual and linguistic adequacy of data and models; power relations between researchers, platforms, institutions and communities; the transformation of discourses, images, links, and behaviors in automated categories; the possibility of surveillance, manipulation or erasure of meanings; and the eventual appropriation of data, knowledge or cultural expressions. It is recommended that the analysis not reduce these questions regarding the legality of accessing the data.

The Ethics Committee (CEP) additionally verifies the following in the Humanities and Social Sciences:

- whether the categories produced by the system preserve context, language, historicity, and social meaning;
- whether vulnerable communities or groups participate meaningfully in identifying risks and benefits;
- whether the use of digital databases, social networks or platforms respects reasonable expectations of privacy and contextual integrity;
- if there is a risk of data colonialism, asymmetric value extraction, reinforcement of inequalities or reuse incompatible with the interests of the groups of origin.

1.33 Interdisciplinary Research

When the protocol combines characteristics of Health Sciences and Sciences

In the Humanities and Social Sciences, it is recommended that the Ethics Committee cumulatively apply the following questions. relevant to different contexts. Disciplinary classification should not reduce attention to a real risk that is actually present.

Integration with Mariah and Application Tools

This Guide establishes the foundations, criteria, and information ethically. relevant. MARIAH and its application tools structure the verification, the classification and documentation of risks.

Table 6 presents the main function of each module in the document set and the A question that each person answers.

Table 6 – Function of the modules in the document set

Module	Main function	Question that answers
Guide	establishes scope, principles, concepts and evaluation criteria	What should be considered and why?
Reference Notebook	It brings together the ethical, conceptual, scientific, and regulatory foundations that underpin the guidelines.	Why was each recommendation, risk axis, or safeguard adopted?
MARIAH — verification instruments	converts the criteria into verifiable items and supporting materials.	What is recommended for the protocol to include, and what does the Ethics Committee verify?
MARIAH — matrix of risk classification	ranks and documents the risk of proportional form	What is the risk profile and level, and what requirements result from it?

Source: author's own elaboration.

1.34 APPLICATION OF MARIAH

It is recommended that MARIAH be applied according to the terms of its own module. This The guide does not reproduce the matrix, its scores, or its calculation rules in order to avoid... Duplication and discrepancies between versions.

Whenever an issue addressed in this Guide corresponds to a dimension assessed by MARIAH, it is recommended that the decision be recorded in the risk instrument and, when necessary, based on the opinion of the Ethics Committee.

1.35 RECORD OF THE ETHICAL DECISION

It is recommended that the report be clear enough to demonstrate what risks are being considered. were identified, what information supported the analysis, what requirements were Which issues have been addressed, what pending matters remain, and what reasons underpinned the decision?

It is recommended that the postal code record:

- the context of AI use;
- the risk level obtained in MARIAH, when applicable;
- the most relevant risk dimensions;
- the necessary requirements or safeguards;
- any relevant technical or ethical differences;
- the basis for due diligence, conditioning or refusal.

1.36 Longitudinal Ethical Supervision and Events Requiring Reassessment

Initial ethical approval does not end oversight if the system, data, or... The purpose or context of use may change significantly. It is recommended that the The protocol defines the events that require amendment, communication, or re-evaluation by the Ethics Committee. in a manner proportional to the risk and without confusing purely technical updates with A modification capable of altering rights, risks, benefits, or scientific validity.

It is recommended to submit the matter to the Ethics Committee for evaluation, according to the risk:

- changing the model, supplier, or component in a way that alters function or performance, explainability or governance;
- inclusion of new databases, populations, sensitive variables or digital sources; change of purpose, context of use, autonomy, user of the output or influence over decisions;
- retraining, substantial adjustment, significant drift or drop in performance, including by subgroups;
- new use, sharing, opening or transfer of the model, parameters or derived products;
- identification of damage, discrimination, re-identification, security incident or improper use as planned.

1.37 SHARED RESPONSIBILITIES AND LIMITS OF THE COMPETENCE OF THE CEP

Ethical protection in AI-related research constitutes a shared responsibility. throughout the life cycle. It is recommended that the responsibilities of be identified.

The responsible researcher, from the proposing institution, and those responsible for governance and for access to data, from developers or suppliers, from sponsors, from funders and other competent bodies. This distribution does not reduce competence. does not transfer to it any technical, sanitary, administrative or protection functions. data belonging to other authorities.

It is recommended that the protocol include:

- **Matrix of responsibilities for decisions, data, validation, security, monitoring, incident response, and closure;**
- **supplier dependencies and limits on access to technical information or contractual;**
- **complementary supervisory bodies and the communication mechanisms between them;**
- **responsible for ensuring compliance with the conditions approved by the Ethics Committee during the entire study.**

1.38 USE OF ARTIFICIAL INTELLIGENCE IN SUPPORTING CEP ACTIVITIES

AI tools can be used in auxiliary organizational activities. training or pre-analysis, provided that its use is institutionally authorized, Transparent, secure, and subject to human oversight. AI cannot deliberate, vote, or issue decisions. autonomous decision, replace collegial assessment, or assume responsibility for opinion. It is recommended that the protocol's content not be entered into an external service. that retains, reuses, or uses the data for training without institutional guarantee. adequate confidentiality, security and purpose.

Safeguards for the use of AI by CEP:

- institutional authorization and prior definition of permitted purposes;
- protection of the confidentiality of protocols, opinions and personal data or commercial;
- Tool registration, version, purpose, user, and stage in which support was provided. automated;
- substantive human verification of facts, references, norms, inferences, and suggestions produced;
- assessment of automation bias, linguistic and contextual appropriateness, and representativeness from the sources;

- prohibition of autonomous decision-making and preservation of human, collegial deliberation and well-founded;
- periodic monitoring of performance, incidents, errors, and effects on quality of the review.

Warning sign:

- undeclared use of a public or commercial tool to analyze the protocol confidential;
- acceptance of automated recommendations without documented human verification;
- dependence on a model trained on norms, languages, or contexts incompatible with o SNEPSH;
- replacing member training or collegial discussion with responses automatically generated

APPENDIX A — SUMMARY OF THE REQUIREMENTS OF THIS GUIDE

Dimension	Minimum expected information
Context of use	purpose, function, autonomy, influence on decision-making, and population affected
Data	origin, quality, representativeness, purpose, agents and flow
Validation	training/validation/testing, performance and contextual suitability
Bias and equity	Performance by subgroups, limitations, and control measures.
Explainability	sufficient information for understanding and contesting
Human supervision	responsible, competence, time, intervention and recording of disagreement
Life cycle	version, update, drift, retraining and destination
Scientific integrity	Declaration of AI Use, Human Verification, and Liability
Data governance	consent, databases, transfers, security and rights
Health Sciences	care impact, SaMD, failure and supervision
Humanities and Social Sciences	sensitive inferences, affected groups, collective risk and secondary uses

Source: author's own elaboration.

APPENDIX B — CONSOLIDATED WARNING SIGNS

- Use of AI that is not declared or is only described generically;
- absence of a defined context of use;
- performance shown only in training data;
- absence of representativeness analysis or subgroups when relevant;
- human supervision is merely nominal;
- high consequence with no real possibility of contestation;
- synthetic data presented as real data;
- unverified normative references or statements;
- declared anonymization without a described method;
- Transfer to third parties or abroad without defined flow and responsibility;
- imported model applied to the Brazilian context without transferability analysis;
- Humanities and Social Sciences with sensitive inferences treated as minimal risk
 - simply because the source data is public;
- SaMD or healthcare system with a significant impact without a defined scope of regulatory and supervisory authority;
- lack of identification of affected individuals, groups or communities, in addition to formal participants;
- a significant change in model, basis, purpose, or context without an assessment of the need for amendment;
- responsibilities that are fragmented or generically assigned to the Ethics Committee;
- use of AI by the CEP without authorization, registration, protection of confidentiality and human verification;
- Reducing collective, social, or cultural risks to purely technical or data protection issues.

APPENDIX C — NORMATIVE AND TECHNICAL FOUNDATIONS TO BE MAINTAINED IN THIS EDITION FINAL

It is recommended that the final edition preserve the traceability of the guidelines to landmarks and references already used in the source document. Among the main sets of Sources cited in the original text include:

- Law No. 14,874/2024 and Decree No. 12,651/2025;
- General Law for the Protection of Personal Data — LGPD;
- CNS Resolutions No. 466/2012, No. 510/2016 and No. 738/2024;
- sector-specific standards and guidelines applicable to SaMD and the use of AI in medicine, when relevant to the protocol;
- CNPq's Policy on Integrity in Scientific Activity, as cited in original document;
- references from the World Health Organization and other international experiences used in the foundation of MARIAH;
- technical references on explainability, bias, federated learning, Privacy and clinical evaluation as mentioned in the original document.

RECOMMENDED READINGS

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

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

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

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

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

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