Responsible AI for Public Health: A Methodological Illustration with a Forecasting Model applied to Respiratory Hospitalizations on SUS Data

Ramon G. Pereira [1], **Luís Eduardo Limas Brito** [1], **Italo Avelar** [1], **Matheus Carvalho** [1]
Marisa Vasconcelos [1], **Michele A. Brandão** [1], **Wagner Meira Jr** [1]

[1]Departamento de Ciência da Computação
Universidade Federal de Minas Gerais (UFMG) - ICEx - Anexo U
Av. Antônio Carlos, 6627, Pampulha, Belo Horizonte - MG, CEP: 31270-901.

{ramonbhb,luis-eduardo,matheusapc}@ufmg.br
{italo.avelar,marisavasconcelos,michelle.brandao,meira}@dcc.ufmg.br

Abstract. *This paper investigates how Responsible AI principles can be systematically integrated into public health predictive modeling using data from the Brazilian Unified Health System. We operationalize a four-layer architecture embedding governance, leakage-controlled temporal validation, fairness auditing, explainability, and structured documentation into the modeling lifecycle. In predicting monthly respiratory hospitalizations, LightGBM achieved an RMSE of 13.81 but exhibited a 44.8 percent sMAPE disparity in the highest disparity state. A resampling adjustment reduced this gap by 8.51 percentage points. The results demonstrate how structured Responsible AI controls reshape evaluation beyond aggregate predictive accuracy.*

1. Introduction

The acceleration of digital transformation in the Brazilian public sector has positioned Artificial Intelligence (AI) as a strategic tool for strengthening health management and surveillance. Within this context, the Brazilian Unified Health System (SUS) provides a unique large-scale ecosystem of administrative and hospital data that enables predictive modeling to support healthcare planning. However, strong regional heterogeneity and historical inequalities pose significant challenges for the responsible use of AI, demanding structured governance and methodological safeguards [Mittelstadt 2019, OECD 2024].

In Brazil, operational protocols that translate Responsible AI (RAI) principles into reproducible computational procedures remain scarce, particularly in light of the emerging national AI legal framework[1]. The emerging Brazilian AI regulatory framework (PL 2338) introduces obligations related to transparency, risk assessment, documentation, and human oversight in high-impact systems, reinforcing the need for operational safeguards in public-sector AI applications. Although healthcare AI studies frequently emphasize predictive accuracy [Scaramussa and Pacheco 2025], metric-centered evaluations often overlook broader ethical, social, and operational risks. Normative guidelines increasingly highlight fairness, transparency, and explainability [UNESCO 2021], yet their systematic implementation in public health analytics is limited [Mittelstadt 2019]. In high-stakes

[1]Brazilian Senate, Bill No. 2338/2023 Use of Artificial Intelligence (2023), available at: https://www25.senado.leg.br/en_US/web/atividade/materias/-/materia/157233.

federative systems, opaque predictive models may amplify regional inequalities, obscure structural data biases, and produce misleading forecasts due to temporal leakage or subgroup imbalance risks that directly affect vulnerable populations and public resource allocation [Landmann-Szwarcwald and Macinko 2016]. This raises the following question: how can Responsible AI principles be systematically integrated into public health predictive modeling using SUS data?

To address this question, we present an illustrative case forecasting monthly respiratory hospitalizations using a SUS-based computational pipeline that integrates predictive modeling with structured Responsible AI mechanisms. The proposed framework is organized into four integrated dimensions: Governance, Data, Modeling, and Responsible AI Evaluation, embedding fairness auditing, explainability analysis, and auditability controls throughout the AI workflow. We incorporate structured documentation artifacts, including Data Cards [Pushkarna et al. 2022] and Model Cards [Mitchell et al. 2019], to enhance transparency, traceability, and alignment with the Brazilian General Data Protection Law (LGPD)[2].

This study makes four main contributions. First, it proposes a structured lifecycle architecture for embedding Responsible AI into public health forecasting. Second, it develops domain-adapted templates for Data Cards and Model Cards, as well as a practical operational checklist to support the systematic implementation of Responsible AI in public health forecasting contexts. Third, it operationalizes fairness, auditability, and explainability through model-dependent safeguards. Fourth, it provides an empirical demonstration using nationwide SUS hospital data, including a structured comparison of model interpretation and decision-support outcomes with and without Responsible AI safeguards, illustrating how responsibility mechanisms reshape model evaluation beyond aggregate predictive accuracy.

2. Responsible Artificial Intelligence in Public Health

In the SUS context, the available data and public health models rely on sensitive administrative records, large-scale population data, and territorially aggregated indicators. The system operates within a federative and territorially heterogeneous infrastructure [Paim et al. 2011]. Recently, data from SUS have been increasingly used to support collective planning and resource allocation processes [Birck et al. 2023, Justo et al. 2019]. Given its federative governance structure, characterized by historical regional inequalities and substantial municipal heterogeneity [Albuquerque et al. 2017], algorithms and models developed for this purpose must explicitly address these structural particularities. Algorithmic performance disparities may translate into territorially uneven planning outcomes.

Public health forecasting supports decisions such as hospital capacity allocation and resource distribution. Unlike patient-level clinical decision systems, aggregate forecasting tools influence territorial allocation processes and may therefore amplify existing structural inequalities when prediction errors are unevenly distributed. Forecasting research emphasizes uncertainty quantification, aggregation effects, time horizons, and realistic performance evaluation as core methodological

[2]Brazil, Law No. 13,709/2018 (General Data Protection Law – LGPD). Available at: `https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm`.

challenges [Soyiri and Reidpath 2013]. In decentralized health systems, these challenges are compounded by regional heterogeneity and multi-level governance structures [Paim et al. 2011, Ferreira et al. 2025, NAM 2024].

Temporal modeling further introduces risks of data leakage and look-ahead bias, particularly when cross-validation strategies or feature construction inadvertently incorporate future information [Bergmeir and Benítez 2012]. Such methodological flaws can artificially inflate predictive performance and distort downstream planning decisions. Ensuring temporally consistent validation and reproducible preprocessing is therefore critical in public-sector forecasting contexts.

These structural and methodological risks underscore the need for an operational Responsible Artificial Intelligence (RAI) framework capable of translating governance principles into concrete safeguards within public health modeling workflows. RAI extends beyond statistical optimization by integrating technical, institutional, and documentation practices across the system lifecycle[1]. International governance frameworks, including the OECD AI Principles [OECD 2024], WHO guidance on AI in health [WHO 2021], the NIST AI Risk Management Framework [NIST 2023], and the EU AI Act [EU 2024], emphasize fairness, transparency, and accountability as complementary safeguards in high-impact algorithmic systems. Fairness (performance parity across relevant subgroups) addresses the distributional effects of model performance across social or territorial groups, mitigating the risk of systematic disadvantage [Barocas et al. 2023]. Explainability (interpretable attribution of model behavior) contributes to epistemic transparency by enabling stakeholders to understand the factors driving model outputs [Lundberg and Lee 2017]. Auditability (reproducible and documented modeling procedures) ensures traceability and reproducibility of modeling decisions, supporting institutional oversight and accountability mechanisms [Mitchell et al. 2019]. However, these frameworks remain primarily normative and provide limited prescriptive guidance for routine computational implementation in large-scale public infrastructures [Mittelstadt 2019].

Recent consensus guidance on health care algorithms [Chin et al. 2023, Lekadir et al. 2025] emphasizes that equity, transparency, stakeholder engagement, fairness trade-offs, and accountability must be embedded across all phases of the algorithm life cycle. While these principles provide a critical normative foundation, there remains a need for concrete methodological structures that translate such recommendations into operational safeguards within real-world public health systems. This study addresses this gap by proposing and illustrating a structured Responsible AI framework tailored to aggregate public health forecasting within SUS.

3. Methodology

This section describes the methodological structure adopted in this study to operationalize Responsible AI (RAI) principles over public health data and forecasting models.

3.1. Methodological Framework

We propose a structured lifecycle architecture for embedding Responsible AI (RAI) into public health (Figure 1). The framework is organized into four structural layers: *Governance*, *Data*, *Modeling*, and *Responsible AI Evaluation*. These layers represent successive stages of the development lifecycle rather than normative ethical categories.

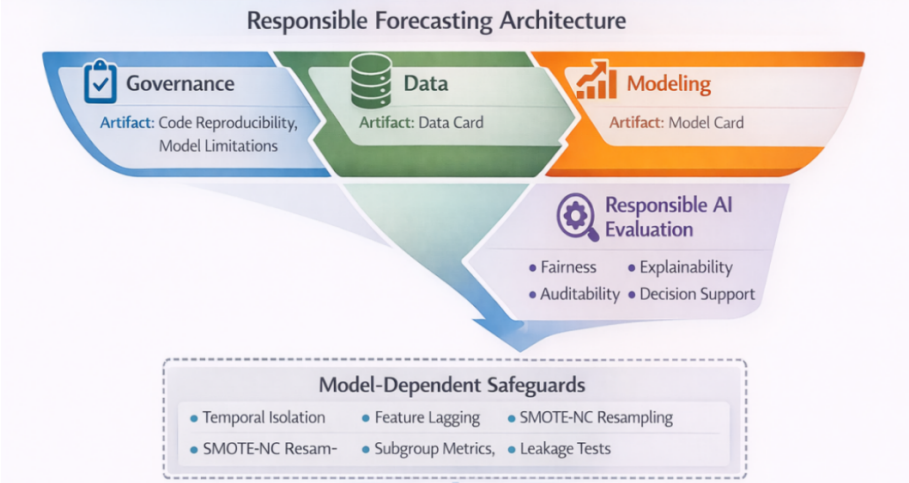


Figure 1. Responsible AI Framework for Healthcare.

Each layer produces a documentation artifact to ensure transparency and traceability. The Governance layer defines the intended use, scope boundaries, risk assumptions, and documented model limitations, establishing institutional constraints and reproducibility standards. The Data layer is documented through a Data Card describing sources, preprocessing decisions, and structural characteristics. The Modeling layer is formalized via a Model Card detailing configurations, validation protocols, and evaluation procedures. The Responsible AI Evaluation layer produces a structured assessment of compliance with RAI principles[3].

In addition, we developed domain-adapted templates for the Data Card and Model Card tailored to public health forecasting contexts, including structured fields specific to SUS infrastructure, temporal validation design, and subgroup analysis. To support operational adoption, we designed a cross-layer Responsible AI checklist composed of verifiable checkpoints that track compliance throughout the model lifecycle.

As an illustrative case, we adopted the monthly prediction of hospitalizations due to respiratory diseases (ICD-10, Chapter J)[4] within the Brazilian Unified Health System (SUS), aggregated at the hospital level. Respiratory diseases were selected due to their epidemiological relevance, seasonal dynamics, and sustained impact on hospital demand in Brazil [Lemos et al. 2024]. The task consists of estimating the monthly number of Hospital Admission Authorizations (AIHs) due to respiratory diseases per hospital, characterizing a count regression problem at an aggregated decision-making level.

All data are publicly available from SIH (Hospitalization System)/SUS (DataSUS)[5], covering the period from January 2022 to November 2025. We modeled the data at the CNES (Hospital ID) $\times$ month level, including lagged hospitalization counts, mortality indicators, length of stay, and aggregated hospital activity variables. The target variable (y_t) corresponds to the number of hospitalizations with a principal diagnosis classified under Chapter J in month t. The modeling pipeline enforces strict temporal integrity: all predictors are derived exclusively from historically available information, feature en-

[3] https://github.com/niar-saude-ufmg/SBCAS_26_Respiratory_Disease
[4]CID-10: http://www2.datasus.gov.br/cid10/V2008/cid10.htm
[5]DataSUS FTP: https://datasus.saude.gov.br/transferencia-de-arquivos

gineering is conducted using training data only, and no forward-looking aggregation is permitted.

3.2. Modeling Approach of the Study Case

The experimental design adopts a temporal train–validation–test split to simulate prospective deployment and prevent information leakage. Data from January 2022 to June 2024 were used for training, July 2024 to December 2024 for validation and hyperparameter tuning, and January 2025 to November 2025 for out-of-sample testing. Predictor variables include historical hospitalization records, demographic characteristics, length of stay, mortality indicators, and aggregated hospital activity measures. All features derive exclusively from past information in the valid experimental configurations.

The system is intended to support health management and epidemiological surveillance. Although operating at an aggregated level, forecasts may indirectly influence planning and resource distribution. This is particularly relevant in the Brazilian federative context, characterized by regional and institutional disparities, as discussed in Section 2. RAI principles were therefore integrated from the design stage. We compared four modeling configurations: *i)* Baseline Model: Linear Regression as a simple parametric reference; *ii)* LightGBM [Ke et al. 2017], selected to capture non-linear relationships in structured data; *iii)* Leaked LGB, a stress-test configuration that includes intentionally leaked features to quantify performance inflation under improper temporal design; and *iv)* Fairness LGB, an improved configuration retrained after fairness analysis and mitigation.

To address regional disparities, we applied a data resampling approach using SMOTE-NC [Lemaître et al. 2017], which supports mixed numerical and categorical data. Resampling occurred in the training partition after temporal splitting to prevent leakage. Over-sampling of underrepresented regions was selected to preserve overall training volume and maintain architectural stability. This approach was chosen as an interpretable first-line fairness mitigation strategy prior to algorithmic constraint-based methods. Training was repeated across five predefined random seeds, and results are reported as mean performance to assess stability. A Model Card[3] documents all configurations to ensure auditability and reproducibility. Predictive performance was evaluated on the temporal test set using MAE, RMSE, and sMAPE. Metrics were computed globally and stratified to support subgroup analysis.

Responsible AI principles were operationalized through model-dependent safeguards embedded in the pipeline. These safeguards included temporal isolation, feature lagging, controlled randomization, subgroup-stratified error analysis for fairness assessment, and post-hoc feature attribution for explainability. As technical controls rather than normative principles, these safeguards reinforce fairness, auditability, and transparency within each modeling configuration. Finally, to assess practical implications, we conducted a structured decision-support evaluation comparing model outputs and interpretations with and without Responsible AI safeguards. This analysis examined how fairness auditing, interpretability procedures, and documentation artifacts influence model interpretation and deployment readiness beyond predictive accuracy.

4. Results

This section reports the results of our experiments with the framework and the RAI operationalized within it. While predictive metrics are reported, the analytical focus is on as-

sessing how fairness, explainability, and auditability mechanisms jointly influence model development and deployment decisions. Results are organized along four dimensions: predictive performance, fairness analysis, explainability assessment, and auditability implementation. This structure mirrors the methodological lifecycle from Section 3.

4.1. Performance

Table 1 shows that LightGBM improves over the linear baseline across all aggregate metrics, reducing sMAPE from 43.29% to 32.29% and slightly lowering MAE and RMSE. The simulated leakage scenario achieves the best performance, with sMAPE reaching 26.28%, confirming how future dependent information can artificially inflate predictive accuracy. The magnitude of this improvement illustrates the risk of overestimating model capability when temporal validation and feature auditing are not strictly enforced. The fairness-constrained model achieves performance comparable to the standard LightGBM on sMAPE and MAE, but with a modest increase in RMSE. This pattern suggests not a generalized degradation, but a redistribution of error across institutional strata. In public health planning contexts, such redistribution may reflect deliberate trade-offs aimed at reducing disparity rather than maximizing purely variance-sensitive metrics.

Table 1. Model performance

Model	sMAPE (%)	MAE	RMSE
Baseline	43.29	8.72	14.13
LightGBM	32.29	8.35	13.81
Leaked LGB	26.28	7.15	12.46
Fairness LGB	31.97	8.32	14.18

Table 2. sMAPE by Hospital Size

Size	Interval	sMAPE (%)
Small	$<$437	43.6
Medium	437 - 859	35.4
Large	859 - 1703	29.3
Super Sized	$>$1703	24.8

4.2. Fairness Assessment

Fairness was evaluated by stratifying the hospital-level test set by structural variables. Subgroups were defined using each facility's majority demographic profile (sex, race, age, region, and hospital size). We identified a low disparity across sex (2.3%), race/ethnicity (6.7%) , and age (4.7%), indicating consistent error distribution across these dimensions. Stratification by hospital size, as shown in Table 2, reveals substantial heterogeneity. The hospitals were divided by size, defined by the average number of hospitalizations for respiratory diseases by month. Small hospitals present the highest sMAPE and large hospitals the lowest. Figure 2 reveals a larger regional variation. Performance ranges from 26% sMAPE (Distrito Federal) to 44.8% (Roraima), an 18.8-point gap. Regional stratification thus exhibits the largest disparity, suggesting that territorial heterogeneity within SUS affects predictive reliability more than demographic composition. From a Responsible AI perspective, these findings reinforce the importance of leakage auditing and group-based evaluation to ensure that predictive gains do not conceal inequitable performance patterns across institutional profiles.

The resampling strategy reduced the max–min regional sMAPE gap from 18.8 to 17.2 percentage points (Figure 2), corresponding to a 8.51% relative reduction. While this indicates sensitivity to training distribution imbalance, a substantial residual disparity persists. This suggests that regional predictive gaps are not solely attributable to sample imbalance and may reflect structural or data quality asymmetries within the health

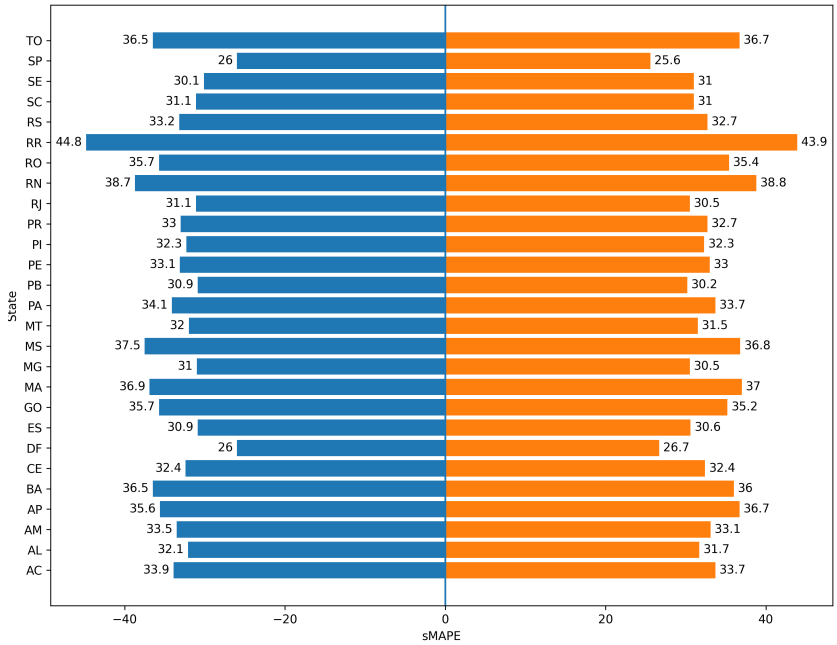


Figure 2. Predictive performance (sMAPE) stratified by Brazilian state.

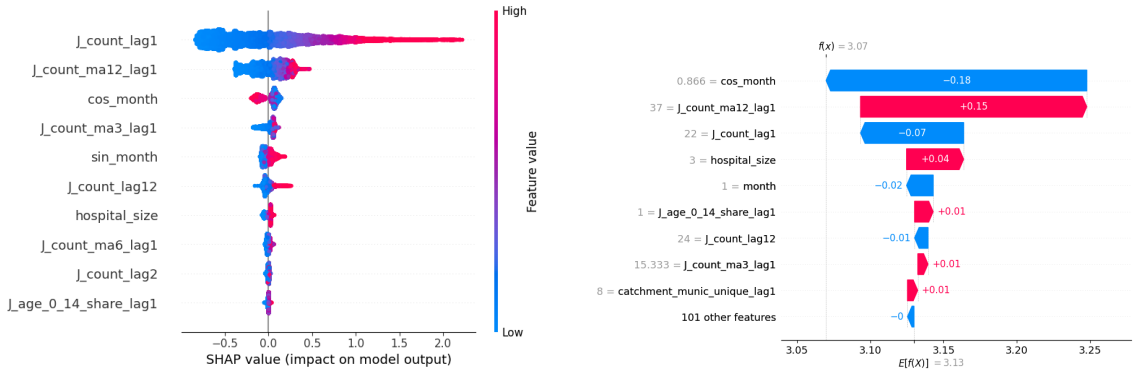
system. This analysis evaluates performance parity across predefined groups and does not encompass the full conceptual scope of fairness. Aggregated data and the absence of causal testing limit interpretability.

4.3. Explainability Analysis

Explainability was conducted using SHAP [Lundberg and Lee 2017], selected for its theoretical grounding in Shapley values and consistency properties. Analyses were restricted to the LightGBM model, which achieved the best predictive performance. Figure 3 summarizes global feature importance. The most influential predictors are prior admission patterns, particularly the 1-month lag (*J_count_lag1*) and the 12-month moving average (*J_count_ma12_lag1*). Historic values from previous months also contribute, though with a smaller magnitude. Overall, predictions are primarily driven by historical utilization. Figure 3b illustrates local explanations for a single prediction. While historical averages remain influential, feature contributions vary across instances, with demographic variables becoming more relevant in specific contexts. This instance-level variability supports auditability by allowing managers to inspect individual forecasts before decision-making. SHAP provides post-hoc feature attribution but does not establish causality or fully capture complex interactions. Explanations should therefore be interpreted as approximations of model behavior rather than definitive justifications. The 1-month lag contributes 43.95% of total SHAP magnitude, confirming its dominant influence on predictions.

4.4. Auditability and Reproducibility

Auditability was operationalized by proposing and instantiating three structured governance artifacts: a Data Card template, a Model Card template, and a Responsible AI development checklist. Together, these components formalize traceability throughout the data, modeling, and evaluation lifecycle.



(a) Global analysis: summary plot. **(b) Local analysis: waterfall plot.**

Figure 3. SHAP explainability analysis for the LightGBM model. (a) illustrates primary feature drivers globally, while (b) shows feature contributions for an individual SUS hospitalization forecast.

The Data Card template captures the data set's provenance, inclusion criteria, temporal scope, preprocessing logic, and potential sources of bias. The Model Card template records architecture, hyperparameters, validation strategy, performance metrics, intended use, limitations, and fairness mitigation steps. The accompanying checklist ensures that fairness, explainability, and reproducibility considerations are explicitly evaluated during model development rather than being retrospectively documented.

Reproducibility was reinforced through fixed temporal splits, predefined random seeds, and repeated training on five runs. All experimental configurations, dependencies, and evaluation metrics are recorded within the Model Card, enabling deterministic re-execution. Documentation templates and executable code are available in the aforementioned anonymous repository (see Footnote[3]). These artifacts demonstrate that auditability can be embedded as a structured methodological layer, complementing fairness and explainability analyzes by making assumptions, mitigation decisions, and experimental results externally verifiable.

4.5. Comparative Procedural Impact of RAI Integration

Table 3. Procedural differences before and after the integration of RAI practices.

Dimension	Without Structured RAI	With Structured RAI
Temporal Validation	Possible random cross-validation	Out-of-time validation (strict temporal split)
Fairness Evaluation	Global metrics only	Stratified sMAPE + disparity analysis
Explainability	Aggregate feature importance	SHAP global and local attribution
Reproducibility	Single run; limited documentation	Five seeds; Data Card; Model Card
Governance	Implicit modeling assumptions	Explicit intended use and limitations

Table 3 illustrates how the adoption of structured Responsible AI practices transforms the modeling workflow into a more robust and systematic approach to public health forecasting. Specifically, *temporal validation* ensures that the train–validation–test split respects the temporal nature of the data; *fairness evaluation* enables the identification of underrepresented subgroups; *explainability* improves the understanding of the most influential predictors; *reproducibility* is supported through the generation of artifacts that

enable the replication of experiments; and *governance* is strengthened by documenting modeling assumptions and decision-making processes.

4.6. Applied Use-Case Scenarios

To illustrate how Responsible AI works in practice, the forecast pipeline was assigned to seven institutional public health scenarios in SUS, as presented in Table 4. Each case demonstrates how predictive performance, fairness auditing, explainability, and structured documentation support governance and decision-making beyond numerical optimization.

This mapping highlights that RAI integration reshapes not only predictive evaluation but also the institutional conditions under which forecasts are interpreted, validated, and acted upon. In some contexts, fairness audits may justify model refinement; in others, they may indicate that further refinement or restricted deployment is warranted when disparities remain high.

Table 4. Illustrative mapping of the forecasting pipeline to institutional public health use cases within SUS. Each column shows how Responsible AI practices, predictive output, auditability, explainability, and fairness evaluation support decision-making and governance.

Use Case	Predictive Output	Auditability	Explainability	Fairness Risk
State Budget Allocation	RMSE, sMAPE for state totals	Reproducible pipeline; dataset hash	SHAP identifies demand drivers	Regional error disparity assessed
ICU Capacity Planning	Monthly admission forecast	Versioned model; drift monitoring	SHAP identifies seasonal and historical feature contributions	Worst-group error monitored
Epidemic Surge Detection	Early spike prediction	Audit trail for data updates	Temporal SHAP contributions highlight drivers of the sudden increase	Macro-region error tracking
Hospital Benchmarking	Forecast error comparison	Transparent evaluation criteria	Residual pattern interpretation	Error parity assessed across hospital sizes
Policy Impact Evaluation	Observed vs predicted comparison	Reproducible experimental setup	Feature-level attribution	Distributional impact monitored
Infrastructure Expansion	Multi-month demand projection	Full version control	Structural drivers identified	Longitudinal fairness tracking
Vaccination Demand Forecasting	Dose requirement per facility	Traceable data pipeline	Accessibility-related drivers identified	Underestimation risk for vulnerable territories

5. Methodological Checklist for Responsible AI

Based on our methodological illustration, in Table 5, we propose a practical checklist to support Responsible AI (RAI) implementation in public health forecasting. It consolidates procedural safeguards across the four layers of our framework to ensure transparency, auditability, and reproducibility.

This checklist is primarily designed for aggregated predictive tasks, as illustrated in our case study. For individual-level clinical applications, additional safeguards are

Table 5. Operational checklist for RAI in public health forecasting. Checkboxes (□) indicate recommended safeguards within four-layer architecture.

Framework Layer	Checklist Items
Governance	□ Define whether predictions operate at aggregated (territorial) or individual level. □ Document intended use, scope, and limitations (Model Card). □ Establish human-in-the-loop protocols and ensure compliance in health standards (e.g. LGPD).
Data	□ Apply strict temporal splitting to prevent leakage in time-series forecasting. □ Document provenance, preprocessing steps, and versioning (e.g., DataCard). □ Verify data representativeness across the federative health infrastructure.
Modeling	□ Document architecture, hyperparameters, and validation design via Model Card. □ Ensure reproducibility through fixed random seeds and versioned environments. □ Implement a baseline model to benchmark relative performance and gains.
RAI Evaluation	□ Stratify performance metrics (e.g. SMAPE) across territorial and institutional subgroups. □ Apply post-hoc attribution methods (e.g., SHAP) for global and local interpretability. □ Conduct fairness auditing, interpreting disparities in the light of structural or epidemiological context. □ Generate a structured Evaluation Report to enable external audit and traceability.

required, including patient-level interpretability and formal clinical validation. The intention is not to prescribe a rigid architecture, but rather to operationalize safeguards that can be systematically integrated throughout the modeling lifecycle.

6. Concluding Remarks

This study demonstrated that ethical risk persists even when predictive modeling operates at an aggregated territorial level. The Responsible AI (RAI) Evaluation layer revealed that forecasting errors are unevenly distributed across regions, confirming that territorial heterogeneity is a structurally relevant fairness dimension in federative systems such as the Brazilian Unified Health System (SUS). Aggregation does not neutralize inequity; instead, it can obscure disparities unless systematic subgroup auditing is embedded in the evaluation process. Rather than proposing a universal Responsible AI framework, this work operationalized a structured lifecycle architecture composed of four interconnected layers: Governance, Data, Modeling, and RAI Evaluation. Each layer produced a concrete documentation artifact, complemented by an operational checklist, designed to ensure transparency and traceability. The Governance layer defined intended use and risk assumptions. The Data layer formalized provenance and preprocessing decisions through a Data Card. The Modeling layer documented validation strategies via a Model Card, while the RAI Evaluation layer consolidated fairness diagnostics, interpretability analyses, and performance assessments into a structured evaluation report.

Within this architecture, responsibility does not emerge from model complexity but from procedural clarity and lifecycle documentation. Temporal validation reduced optimistic bias in performance estimation, while subgroup auditing exposed territorially uneven behavior. Reproducibility practices ensured that the forecasting pipeline could be reconstructed and inspected. Together, these elements transformed abstract RAI principles into operational safeguards aligned with public health governance needs. More broadly, this study suggests that Responsible AI in public health should be understood as

a layered institutional process. Transparency, auditability, and contextual evaluation must be embedded across governance, data management, and modeling decisions. Predictive accuracy is necessary but insufficient without structured mechanisms that make model behavior visible and governable within institutional settings.

Future work should extend this layered architecture to address robustness to distributional shifts, resilience under epidemiological shocks, and continuous post-deployment monitoring within SUS. Formalizing human-oversight protocols within the Governance layer and expanding the RAI Evaluation layer to incorporate intersectional fairness analyzes are promising directions. Harmonizing these practices across heterogeneous regions while preserving local autonomy remains a central methodological and institutional challenge for Responsible AI in federative public health systems.

References

Albuquerque, M., Viana, A., Lima, L., Ferreira, M., Fusaro, E., and Iozzi, F. (2017). Regional Health Inequalities: Changes Observed in Brazil from 2000–2016. *Ciência & Saúde Coletiva*, 22(4):1055–1064.

Barocas, S., Hardt, M., and Narayanan, A. (2023). *Fairness and Machine Learning: Limitations and Opportunities*. MIT Press, Cambridge, MA, USA.

Bergmeir, C. and Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences*, 191(1):192–213.

Birck, M. G., Ferreira, R., Curi, M., et al. (2023). Real-world treatment patterns of rheumatoid arthritis in brazil: analysis of datasus national administrative claims data for pharmacoepidemiology studies (2010–2020). *Scientific Reports*, 13(1):17739.

Chin, M. H. et al. (2023). Guiding principles to address the impact of algorithm bias on racial and ethnic disparities in health and health care. *JAMA Network Open*, 6(12):e2345050.

EU (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council (Artificial Intelligence Act). Official Journal of the European Union. In: `https://eur-lex.europa.eu/eli/reg/2024/1689/oj`. Acesso em: 11/01/2026.

Ferreira, A. P. L. d. A., Redaelli, E., Maldaner, L. F., and Rigo, S. J. (2025). Inovação e transformação no ecossistema de saúde auditiva: Inteligência artificial e interoperabilidade. In *Simpósio Brasileiro de Computação Aplicada à Saúde*, pages 1–12, Porto Alegre, Brasil.

Justo, N. et al. (2019). Real-world evidence in healthcare decision making: Global trends and case studies from latin america. *Value in Health*, 22(6):739–749.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In *Advances in Neural Information Processing Systems*, pages 3146–3154, Long Beach, USA.

Landmann-Szwarcwald, C. and Macinko, J. (2016). A panorama of health inequalities in brazil. *International Journal for Equity in Health*, 15(1):174.

Lekadir, K. et al. (2025). FUTURE-AI: International Consensus Guideline for Trustworthy and Deployable Artificial Intelligence in Healthcare. *British Medical Journal*, 388(1):e078499.

Lemaître, G., Nogueira, F., and Aridas, C. K. (2017). Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *Journal of Machine Learning Research*, 18(17):1–5.

Lemos, D., Fonseca, L., Florêncio, R., de Almeida, J., Lima, I., and Gualdi, L. (2024). Hospitalisations and Fatality due to Respiratory Diseases according to a National Database in Brazil: a Longitudinal Study. *BMJ Open Respiratory Research*, 11(1):e002103.

Lundberg, S. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*, NIPS'17, pages 4765–4774, Long Beach, USA. NeurIPS, Curran Associates, Inc.

Mitchell, M. et al. (2019). Model Cards for Model Reporting. In *ACM Conference on Fairness, Accountability, and Transparency*, FAT* 2019, pages 220–229, Atlanta, USA. ACM.

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical ai. *Nature Machine Intelligence*, 1(11):501–507.

NAM (2024). Health Care Artificial Intelligence Code of Conduct. National Academy of Medicine. In: `https://www.nationalacademies.org/news/ai-code-of-conduct-for-health-and-medicine-presented-in-new-nam-special-publication`.

NIST (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0) — NIST AI 100-1. National Institute of Standards and Technology. In: `https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf`.

OECD (2024). Recommendation of the Council on Artificial Intelligence. OECD Legal Instruments. In: `https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449`.

Paim, J., Travassos, C., Almeida, C., Bahia, L., and Macinko, J. (2011). The Brazilian health system: History, advances, and challenges. *The Lancet*, 377(9779):1778–1797.

Pushkarna, M., Zaldivar, A., and Kjartansson, O. (2022). Data Cards: Purposeful and Transparent Dataset Documentation for Responsible AI. In *ACM Conference on Fairness, Accountability, and Transparency*, pages 1936–1949, Seoul, South Korea.

Scaramussa, L. M. and Pacheco, A. G. C. (2025). Anotação de imagens médicas assistida por ia: um estudo sobre segmentação de lesões de pele por não especialistas. In *Simpósio Brasileiro de Computação Aplicada à Saúde*, pages 152–163, Porto Alegre, Brasil.

Soyiri, I. N. and Reidpath, D. D. (2013). An overview of health forecasting. *Environmental Health and Preventive Medicine*, 18(1):1–9.

UNESCO (2021). Recommendation on the ethics of artificial intelligence. United Nations Educational, Scientific and Cultural Organization. In: `https://unesdoc.unesco.org/ark:/48223/pf0000381137`.

WHO (2021). Ethics and governance of artificial intelligence for health. World Health Organization. Guidance document. In: `https://www.who.int/publications/i/item/9789240029200`.