IOCH
2026
Conference

The 1st International Online Conference on Healthcare

25–26 March 2026 | Online

MDPI

Healthcare Regulatory Compliance: A Generative AI Framework for Identifying and Mitigating Risks

Nafissatou Ndiaye, Johanna Fokui, Julie Abouem, Diarra Gningue, Mounina Toure, Anuradha Kar

aivancity School of AI & Data for Business & Society, Paris, France

Contact: kar@aivancity.ai

aivancity
Paris Île-de-France · Nice Côte d'Azur

Article id:
sciforum-163144

INTRODUCTION & AIM

Introduction

Healthcare organizations operate within strict regulatory environments designed to ensure patient safety, data protection, and ethical standards. However, for advanced data driven and machine learning based applications in health care, the complexity of regulations such as GDPR and AI Act makes compliance monitoring challenging. Traditional compliance audit approaches rely on manual review processes, which can be time-consuming and prone to oversight leading to regulatory violations and non-compliance.

Aim

This work proposes a novel framework to support automatic healthcare regulatory compliance by systematically identifying potential regulatory risks and suggesting mitigation strategies. The framework leverages large language models (LLMs) to analyze regulatory texts, detect compliance gaps, and provide structured recommendations to assist organizations in compliance management of healthcare projects.

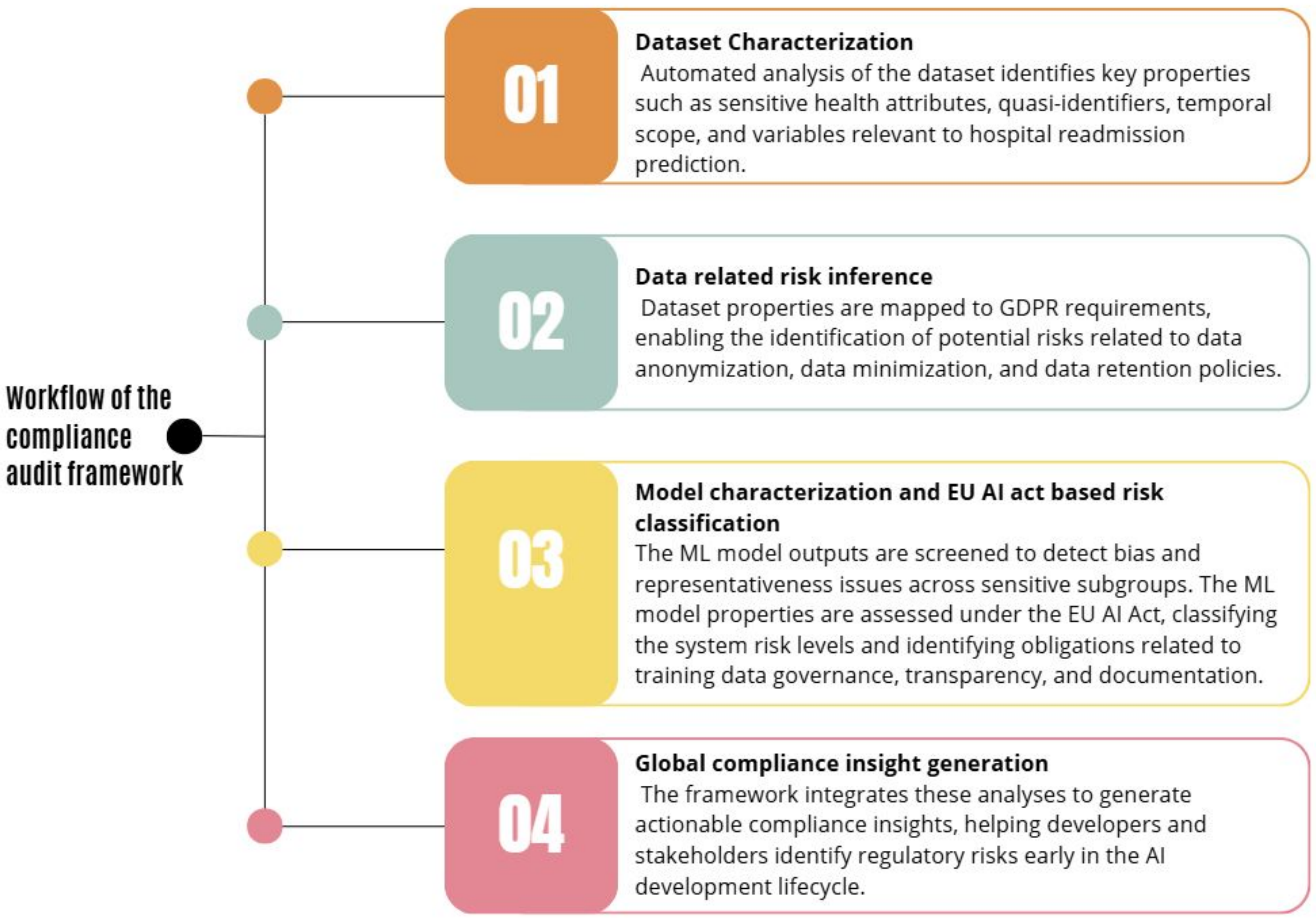
METHOD

Use Case and Dataset

The framework was evaluated using a **machine learning based hospital readmission prediction** use case. The task is to predict if a patient will be **readmitted within 30 days after hospital discharge**. The dataset contains features like **age, gender, diagnosis, number of procedures, and discharge destination**. Two machine learning models: logistic regression and random forest were tested for prediction.

Compliance audit framework concept

The framework involves several modules as shown below. First, a **dataset characterization module** identifies sensitive attributes and quasi-identifiers of compliance from the dataset. Next, the **GDPR risk inference module** map dataset properties to data protection requirements. Then a **model assessment module** is used to study model properties under the EU AI Act. Finally an **insight generation module** creates a synthesis report based on the outputs of the above modules to to generate actionable compliance insights


Workflow of the compliance audit framework

01 **Dataset Characterization**
Automated analysis of the dataset identifies key properties such as sensitive health attributes, quasi-identifiers, temporal scope, and variables relevant to hospital readmission prediction.

02 **Data related risk inference**
Dataset properties are mapped to GDPR requirements, enabling the identification of potential risks related to data anonymization, data minimization, and data retention policies.

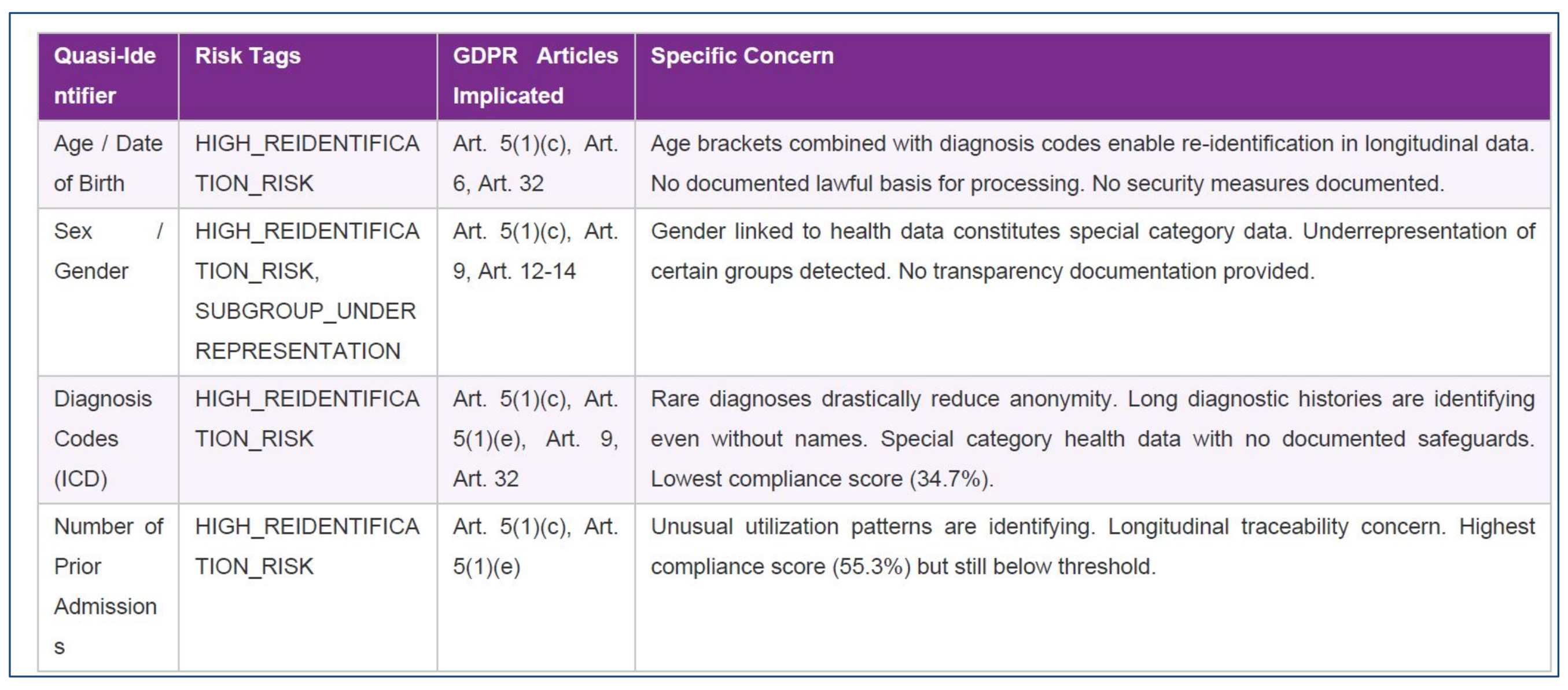
03 **Model characterization and EU AI act based risk classification**
The ML model outputs are screened to detect bias and representativeness issues across sensitive subgroups. The ML model properties are assessed under the EU AI Act, classifying the system risk levels and identifying obligations related to training data governance, transparency, and documentation.

04 **Global compliance insight generation**
The framework integrates these analyses to generate actionable compliance insights, helping developers and stakeholders identify regulatory risks early in the AI development lifecycle.

RESULTS & DISCUSSION

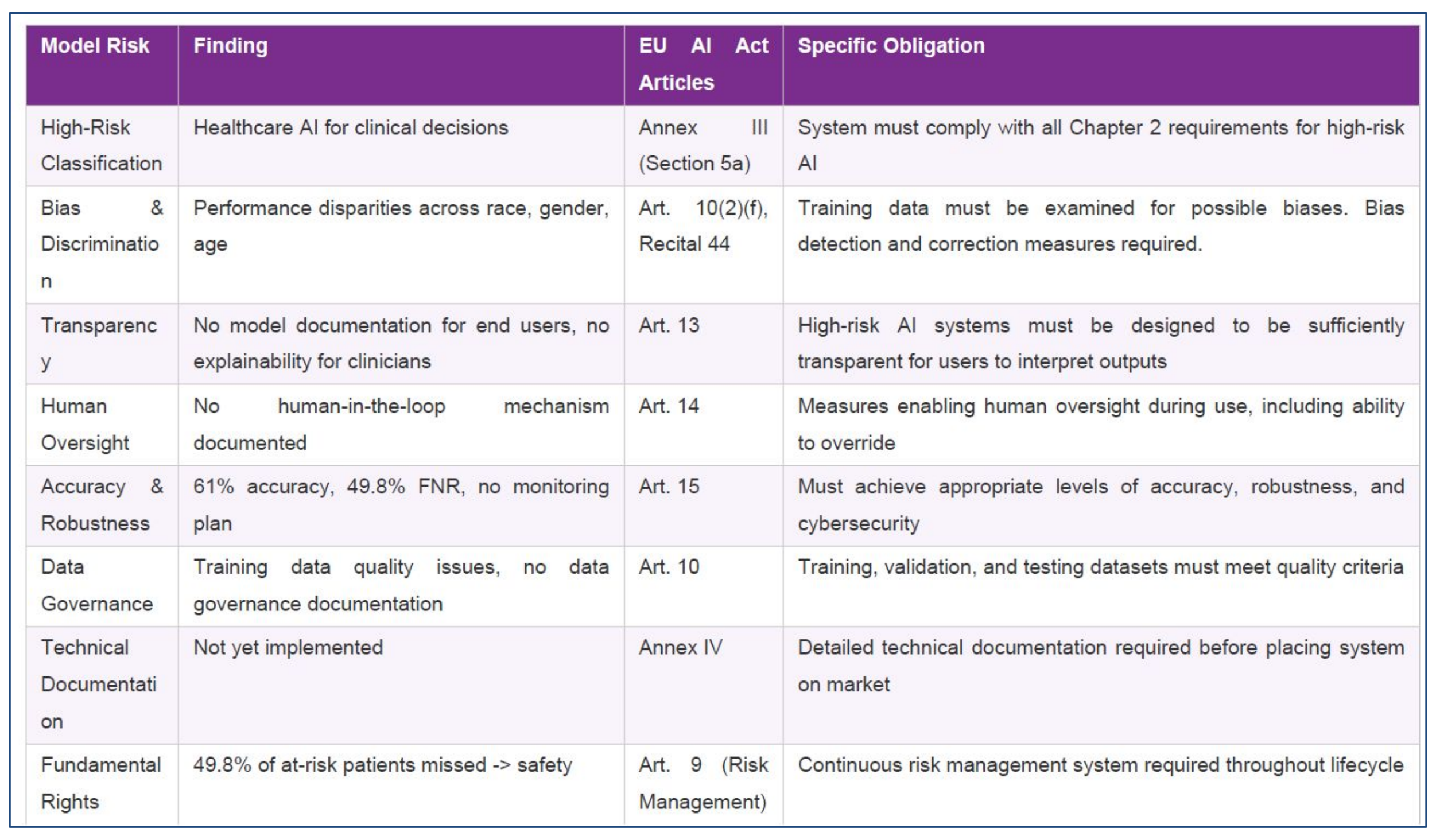
The framework extracted **quasi-identifiers** in the dataset and assigned regulatory risk tags (e.g., high re-identification risk, sensitive attributes). Using a **GDPR and EU AI Act knowledge base**, a **RAG-based approach** generated an automated compliance checklist which helped to identify **GDPR and EU AI Act regulatory risks that are presented below.**

Data compliance findings from the healthcare dataset generated by our framework



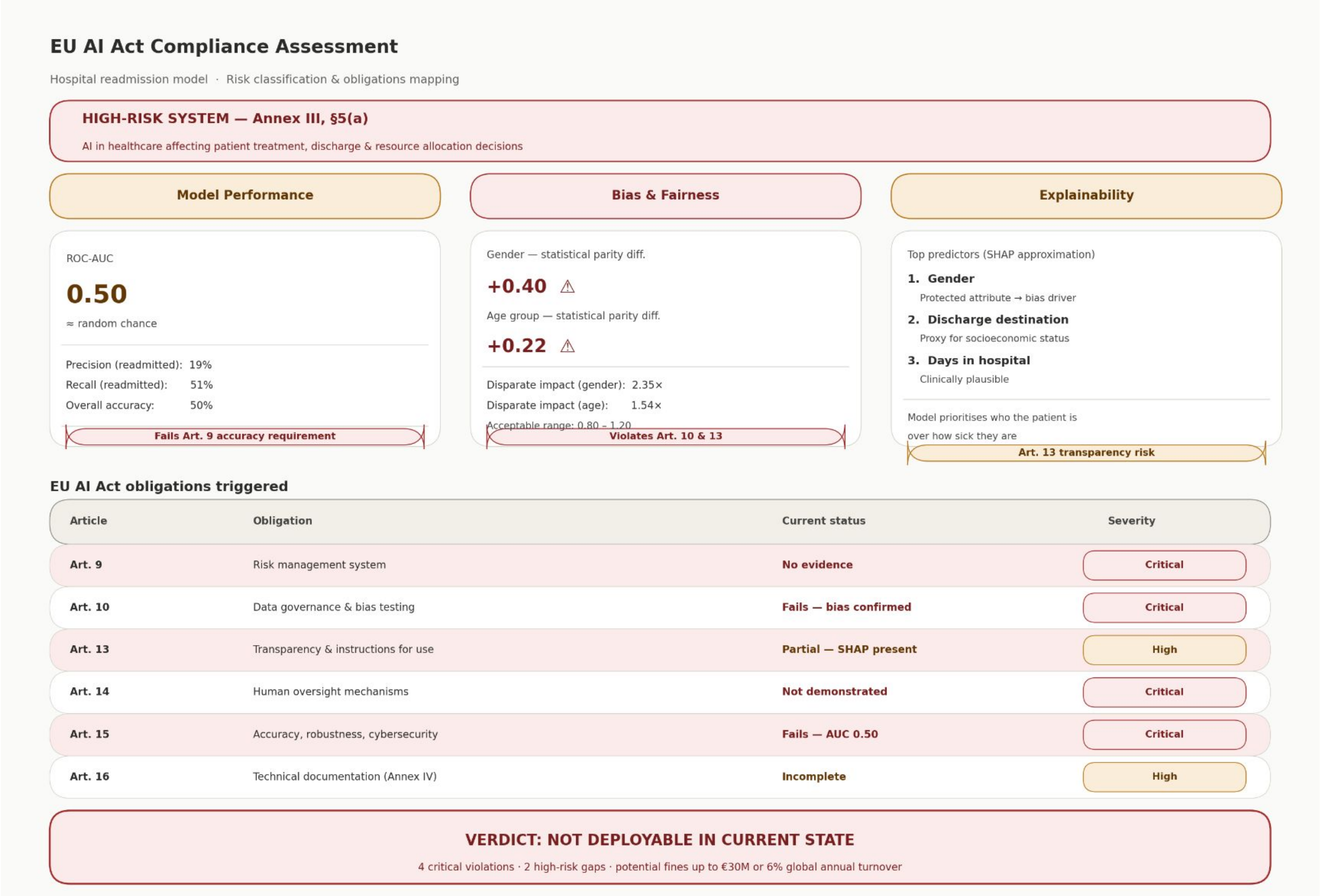
Quasi-Identifier	Risk Tags	GDPR Articles Implicated	Specific Concern
Age / Date of Birth	HIGH_REIDENTIFICATION_RISK	Art. 5(1)(c), Art. 6, Art. 32	Age brackets combined with diagnosis codes enable re-identification in longitudinal data. No documented lawful basis for processing. No security measures documented.
Sex / Gender	HIGH_REIDENTIFICATION_RISK, SUBGROUP_UNDERREPRESENTATION	Art. 5(1)(c), Art. 9, Art. 12-14	Gender linked to health data constitutes special category data. Underrepresentation of certain groups detected. No transparency documentation provided.
Diagnosis Codes (ICD)	HIGH_REIDENTIFICATION_RISK	Art. 5(1)(c), Art. 5(1)(e), Art. 9, Art. 32	Rare diagnoses drastically reduce anonymity. Long diagnostic histories are identifying even without names. Special category health data with no documented safeguards. Lowest compliance score (34.7%).
Number of Prior Admissions	HIGH_REIDENTIFICATION_RISK	Art. 5(1)(c), Art. 5(1)(e)	Unusual utilization patterns are identifying. Longitudinal traceability concern. Highest compliance score (55.3%) but still below threshold.

Model compliance findings for the trained machine learning model as identified by our framework



Model Risk	Finding	EU AI Act Articles	Specific Obligation
High-Risk Classification	Healthcare AI for clinical decisions	Annex III (Section 5a)	System must comply with all Chapter 2 requirements for high-risk AI
Bias & Discrimination	Performance disparities across race, gender, age	Art. 10(2)(f), Recital 44	Training data must be examined for possible biases. Bias detection and correction measures required.
Transparency	No model documentation for end users, no explainability for clinicians	Art. 13	High-risk AI systems must be designed to be sufficiently transparent for users to interpret outputs
Human Oversight	No human-in-the-loop mechanism documented	Art. 14	Measures enabling human oversight during use, including ability to override
Accuracy & Robustness	61% accuracy, 49.8% FNR, no monitoring plan	Art. 15	Must achieve appropriate levels of accuracy, robustness, and cybersecurity
Data Governance	Training data quality issues, no data governance documentation	Art. 10	Training, validation, and testing datasets must meet quality criteria
Technical Documentation	Not yet implemented	Annex IV	Detailed technical documentation required before placing system on market
Fundamental Rights	49.8% of at-risk patients missed -> safety	Art. 9 (Risk Management)	Continuous risk management system required throughout lifecycle

Bias, fairness and explainability assessment of the machine learning based hospital readmission prediction model reported by our framework



EU AI Act Compliance Assessment

Hospital readmission model · Risk classification & obligations mapping

HIGH-RISK SYSTEM — Annex III, §5(a)

AI in healthcare affecting patient treatment, discharge & resource allocation decisions

Model Performance

ROC-AUC
0.50
≈ random chance

Precision (readmitted): 19%
Recall (readmitted): 51%
Overall accuracy: 50%

Fails Art. 9 accuracy requirement

Bias & Fairness

Gender — statistical parity diff.
+0.40 ⚠
Age group — statistical parity diff.
+0.22 ⚠

Disparate impact (gender): 2.35×
Disparate impact (age): 1.54×
Acceptable range: 0.80 – 1.20

Violates Art. 10 & 13

Explainability

Top predictors (SHAP approximation)
1. Gender
Protected attribute → bias driver
2. Discharge destination
Proxy for socioeconomic status
3. Days in hospital
Clinically plausible

Model prioritises who the patient is over how sick they are

Art. 13 transparency risk

EU AI Act obligations triggered

Article	Obligation	Current status	Severity
Art. 9	Risk management system	No evidence	Critical
Art. 10	Data governance & bias testing	Fails — bias confirmed	Critical
Art. 13	Transparency & instructions for use	Partial — SHAP present	High
Art. 14	Human oversight mechanisms	Not demonstrated	Critical
Art. 15	Accuracy, robustness, cybersecurity	Fails — AUC 0.50	Critical
Art. 16	Technical documentation (Annex IV)	Incomplete	High

VERDICT: NOT DEPLOYABLE IN CURRENT STATE

4 critical violations · 2 high-risk gaps · potential fines up to €30M or 6% global annual turnover

CONCLUSION & FUTURE WORK

Taking into consideration the above data and prediction model compliance findings, our framework generated the following synthesis for the chosen use case: ***The hospital readmission prediction system fails to meet the requirements of both the GDPR and the EU AI Act. The data layer presents significant privacy and governance risks (44.1% compliance), while the model layer raises patient safety, fairness, and transparency concerns. These risks compound each other: privacy-sensitive features are also the features driving model predictions, creating a regulatory feedback loop.***

Our framework demonstrates a data and model driven approach to automate compliance audit of AI for healthcare pipelines as per existing EU regulations. This is an ongoing work where future directions involve including more use cases, models, datasets and considering global data and AI related regulations.