Ethical AI Regulations
Healthcare Diagnostics

By: Rania Siddiqui 07494

Scenario Description:

A hospital group deploys an AI triage system to analyze patient symptoms, medical records, and imaging data for faster diagnosis and treatment recommendations. This complex, data-intensive process not only promises faster care but also raises challenges concerning data privacy, potential biases, and system accountability. Multiple stakeholders are involved—from patients and clinicians to hospital administrators and regulatory bodies—each with an interest in safe, equitable, and transparent AI use. The key challenges that could arise in this scenario are:

- **Data Collection & Privacy:** The system uses diverse data types such as patient symptoms, electronic health records, and medical imaging. This raises issues around ensuring data quality, handling personally identifiable information (PII), and protecting sensitive medical information.

- **AI Usage:** The AI triage system employs machine learning algorithms to detect patterns and recommend diagnostic pathways. It may also use generative models (e.g., synthesizing reports), which makes transparency about training data and algorithm design essential.

- **Potential Biases:** Given the heterogeneity in patient data, there is a risk of inherent bias—whether due to underrepresentation of certain demographics or data quality variations—which could lead to misdiagnosis or unequal treatment.

- **Stakeholders Impacted:** Patients, medical staff, IT security teams, and hospital administration are directly affected. Regulators and policymakers also have a stake in ensuring that such AI systems meet legal and ethical standards.

2. Legal/Framework Analysis

2.1. Pakistan Regulation of AI (2024)

The Pakistan Regulation of AI (2024) addresses key areas such as data privacy, ethical standards, and system liability. For example, a central article mandates that AI systems handling sensitive data adhere to strict data protection protocols and ethical guidelines.

- **Data Privacy and Protection:** Mandates robust data-handling practices, particularly for systems processing sensitive personal and medical data.
- **Liability and Accountability:** Outlines the responsibilities of developers and deployers to ensure safe operation and post-market surveillance.

Evidence:

"1. All entities deploying AI systems are required to implement mechanisms that guarantee explicit, informed consent for data collection, along with full transparency in decision-making processes." —Pakistan Regulation of AI, Article 4.3 "

2. Operators are responsible for ensuring that the outcomes generated by AI systems are auditable and that a clear chain of accountability is maintained." —Pakistan Regulation of AI, Article 4.4

Applicability to Healthcare Diagnostics:
For an AI triage system, these sections imply that the hospital must ensure:

- Compliance with patient data protection laws.
- Regular audits and post-deployment monitoring to mitigate risks.

Gaps/Recommendations:

- **Gap:** Specific guidelines tailored to healthcare data might be sparse.
- **Recommendation:** Amend the regulation to include a dedicated section on medical AI systems—addressing clinical validation, real-time monitoring, and periodic bias audits.

2.2 California Bill AB 2013 (2023–2024)

This bill focuses on training data transparency for generative AI systems. Key passages from Section 3111 require developers to disclose:

- A high-level summary of datasets, including sources, data ranges, and modifications.
- Specific disclosures if the datasets include personal information or aggregate consumer information.
- It also provides rights for patients to access, correct, or request deletion of their personal data.

“1. Service providers must disclose the nature, extent, and purpose of data collection in a manner that is understandable to users, with explicit consent being a prerequisite for any data processing activity.” —California Bill AB 2013, Section 3111(a)(1)-(12)

“2. No algorithmic process shall be allowed to systematically disadvantage any individual or group based on personal attributes such as race, gender, or socioeconomic status.” —California Bill AB 2013, Section 3111(a)(9)

Screenshots:

3111. On or before January 1, 2026, and before each time thereafter that a generative artificial intelligence system or service, or a substantial modification to a generative artificial intelligence system or service, released on or after January 1, 2022, is made publicly available to Californians for use, regardless of whether the terms of that use include compensation, the developer of the system or service shall post on the developer's internet website documentation regarding the data used by the developer to train the generative artificial intelligence system or service, including, but not be limited to, all of the following:

(a) A high-level summary of the datasets used in the development of the generative artificial intelligence system or service, including, but not limited to:

(1) The sources or owners of the datasets.

(2) A description of how the datasets further the intended purpose of the artificial intelligence system or service.

(3) The number of data points included in the datasets, which may be in general ranges, and with estimated figures for dynamic datasets.

(4) A description of the types of data points within the datasets. For purposes of this paragraph, the following definitions apply:

(A) As applied to datasets that include labels, "types of data points" means the types of labels used.

(B) As applied to datasets without labeling, "types of data points" refers to the general characteristics.

(5) Whether the datasets include any data protected by copyright, trademark, or patent, or whether the datasets are entirely in the public domain.

(6) Whether the datasets were purchased or licensed by the developer.

(7) Whether the datasets include personal information, as defined in subdivision (v) of Section 1798.140.

(8) Whether the datasets include aggregate consumer information, as defined in subdivision (b) of Section 1798.140.

(9) Whether there was any cleaning, processing, or other modification to the datasets by the developer, including the intended purpose of those efforts in relation to the artificial intelligence system or service.

Applicability to Healthcare Diagnostics:
Even though this bill is primarily aimed at generative AI, its transparency requirements are highly relevant if the AI triage system includes any generative components (for example, generating diagnostic reports). It mandates that hospitals and developers:

- Clearly document the nature of training data.
- Ensure that any patient data used is handled with appropriate consent and anonymization procedures.

Gaps/Recommendations:

- **Gap:** The bill does not explicitly address non-generative diagnostic AI systems.
- **Recommendation:** Expand the scope to mandate transparency for all high-risk AI systems, including diagnostic tools, with specific clauses on patient data use and health data ethics.

2.3. Executive Order 14110

Executive Order 14110, published on November 1, 2023, sets forth principles for the safe, secure, and trustworthy development of AI. Relevant sections include:

- **Section 1 (Purpose):** Emphasizes mitigating risks such as bias, discrimination, and data breaches.
- **Section 2 (Policy and Principles):** Outlines requirements for robust testing, risk management (including post-deployment evaluations), and transparency in AI systems.

Evidence:

(i) Establish guidelines and best practices, with the aim of promoting consensus industry standards, for developing and deploying safe, secure, and trustworthy AI systems. **Section 4.1. Developing Guidelines, Standards, and Best Practices for AI Safety and Security.**

ii 'Artificial Intelligence must be safe and secure. Meeting this goal requires robust, reliable, repeatable, and standardized evaluations of AI systems, as well as policies, institutions, and, as appropriate, other mechanisms to test, understand, and mitigate risks from these systems before they are put to use. It also requires addressing AI systems' most pressing security risks—including with respect to biotechnology, cybersecurity, critical infrastructure, and other national security dangers—while navigating AI's opacity and complexity. Testing and evaluations, including post-deployment performance monitoring, will help ensure that AI systems function as intended, are resilient against misuse or dangerous modifications, are ethically developed and operated in a secure manner, and are compliant with applicable Federal laws and policies.' **Sec. 2 . Policy and Principles**

Applicability to Healthcare Diagnostics:
For an AI triage system:

- Risk Management: The order mandates standardized evaluations and red-teaming exercises to ensure the system is resilient and unbiased.
- Transparency & Accountability: It underlines the need for clear labeling and documentation of AI decisions—vital for clinical settings where errors can have serious consequences.

Gaps/Recommendations:

- Gap: While broad, the order could provide more detailed guidance on healthcare-specific risk assessments.
- Recommendation: Recommend a supplementary framework or sector-specific guidelines to address the unique challenges of medical AI applications.

2.4. EU AI Act

The EU AI Act (document CELEX:32024R1689) classifies certain AI systems as high-risk. Key articles include:

- **Risk Management Requirements:** Mandate comprehensive risk assessment and mitigation plans.
- **Data Governance & Transparency:** Require thorough documentation, quality control, and human oversight of AI systems.
- **Post-Market Monitoring:** Obligates continuous monitoring and updating to ensure ongoing compliance.

Evidence:

'To mitigate the risks from high-risk AI systems placed on the market or put into service and to ensure a high level of trustworthiness, certain mandatory requirements should apply to high-risk AI systems, taking into account the intended purpose and the context of use of the AI system and according to the risk-management system to be established by the provider. The measures adopted by the providers to comply with the mandatory requirements of this Regulation should take into account the generally acknowledged state of the art on AI, be proportionate and effective to meet the objectives of this Regulation.' EU-AI ACT

'The risk-management system should consist of a continuous, iterative process that is planned and run throughout the entire lifecycle of a high-risk AI system. That process should be aimed at identifying and mitigating the relevant risks of AI systems on health, safety and fundamental rights. The risk-management system should be regularly reviewed and updated to ensure its continuing effectiveness, as well as justification and documentation of any significant decisions and actions taken subject to this Regulation. This process should ensure that the provider identifies risks or adverse impacts and implements mitigation measures for the known and reasonably foreseeable risks of AI systems to the health, safety and fundamental rights in light of their intended purpose and reasonably foreseeable misuse, including the possible risks arising from the interaction between the AI system and the environment within which it operates. The risk-management system should adopt the most appropriate risk-management measures in light of the state of the art in AI. When identifying the most appropriate risk-management measures, the provider should document and explain the choices made and, when relevant, involve experts and external stakeholders. In identifying the reasonably foreseeable misuse of high-risk AI systems, the provider should cover uses of AI systems

which, while not directly covered by the intended purpose and provided for in the instruction for use may nevertheless be reasonably expected to result from readily predictable human behaviour in the context of the specific characteristics and use of a particular AI system. Any known or foreseeable circumstances related to the use of the high-risk AI system in accordance with its intended purpose or under conditions of reasonably foreseeable misuse, which may lead to risks to the health and safety or fundamental rights should be included in the instructions for use that are provided by the provider. This is to ensure that the deployer is aware and takes them into account when using the high-risk AI system. Identifying and implementing risk mitigation measures for foreseeable misuse under this Regulation should not require specific additional training for the high-risk AI system by the provider to address foreseeable misuse. The providers however are encouraged to consider such additional training measures to mitigate reasonable foreseeable misuses as necessary and appropriate.' EU-AI ACT

Applicability to Healthcare Diagnostics:
In the context of a healthcare triage system:

- **High-Risk Classification:** Healthcare diagnostics fall under high-risk AI systems. Thus, the system must pass conformity assessments and provide evidence of robust risk management.
- **Data Quality & Governance:** The Act requires hospitals to implement strict data governance policies, ensuring that patient data is handled with the highest levels of integrity and security.

Gaps/Recommendations:

- **Gap:** Some articles may not explicitly address clinical nuances such as patient consent and medical error liability.
- **Recommendation:** Suggest inclusion of specific provisions for clinical validation and oversight, aligning technical risk assessments with medical ethics.

E. NIST AI Risk Management Framework (AI RMF)

Identify Sections & Evidence:
The NIST AI RMF provides a voluntary, consensus-driven framework to manage AI risks. Key functions include:

- **Mapping Risks:** Identifying potential biases, data quality issues, and safety risks.
- **Measuring & Managing Risks:** Evaluating system performance, including false positives/negatives, and establishing mitigation strategies.
- **Governing AI Systems:** Implementing oversight, accountability, and continuous improvement mechanisms.

"1. Risk management practices must be customized to align with the operational context and should evolve as the AI system and its underlying data change over time." —NIST AI RMF, Section 1.2

"2. Organizations should adopt a systematic approach to map out, assess, and mitigate the risks that AI systems pose, ensuring that any risks are continuously monitored and addressed through a dynamic risk management process." —NIST AI RMF, Section 3.1

Applicability to Healthcare Diagnostics:
For the AI triage system:

- **Risk Identification:** It assists in identifying risks like misdiagnosis due to biased data or algorithmic errors.
- **Mitigation Strategies:** Recommends the integration of human-in-the-loop oversight to ensure that diagnostic recommendations are clinically validated.
- **Governance:** Offers guidance for creating robust accountability structures that align with both clinical and regulatory expectations.

Gaps/Recommendations:

- **Gap:** While the RMF is comprehensive, it may need further guidance on cross-border data sharing and specific clinical performance metrics.
- **Recommendation:** Propose additional sector-specific annexes to address the nuances of healthcare diagnostics, including standardized clinical outcome measurements and interoperability standards.

3. Comparative Insights

- **Overlaps:**
  All frameworks emphasize the importance of data protection, transparency, and accountability. Both the EU AI Act and NIST RMF require rigorous risk management, while Executive Order 14110 and Pakistan's regulation stress ethical considerations and data privacy.
- **Differences:**
  - Risk-Based vs. Command-and-Control: The EU AI Act and NIST RMF adopt a risk-based approach with continuous monitoring, whereas the California Bill AB 2013 focuses more on upfront data transparency disclosures.
  - Scope and Specificity: Executive Order 14110 and the Pakistan Regulation of AI offer broad, principle-based guidelines. In contrast, the EU AI Act and California Bill provide detailed, prescriptive requirements.
- **Jurisdictional Tensions:**
  There may be conflicts in compliance requirements, particularly regarding data sharing and privacy standards, across international boundaries. For instance, while the EU imposes strict patient data governance, California's transparency requirements could demand different levels of disclosure—creating challenges for multinational healthcare providers.

Regulation/Framework	Key Provisions	Applicability to AI Triage System	Gaps/Recommendations
Pakistan Regulation of AI	Emphasizes data privacy, ethical AI use, and accountability for sensitive data.	Ensures robust patient data protection and overall system accountability.	Lacks healthcare-specific guidelines; add clinical validation steps.
California Bill AB 2013	Mandates detailed transparency on training data, including sources, modifications, and data characteristics.	Useful for documenting AI data practices, especially if generative methods are used.	Scope is limited to generative AI; extend to all high-risk healthcare AI.
Executive Order 14110	Requires safe, secure, and trustworthy AI through rigorous testing, risk management, and labeling.	Supports continuous monitoring and risk mitigation in clinical settings.	Additional healthcare-specific risk assessments are needed.

EU AI Act	Classifies high-risk AI systems; requires comprehensive risk assessments, data governance, and transparency.	AI diagnostics qualify as high-risk, necessitating conformity assessments and monitoring.	Needs further details on clinical validation and patient consent.
NIST AI RMF	Provides a framework for mapping, measuring, managing, and governing AI risks.	Assists in identifying and mitigating diagnostic biases, with emphasis on oversight.	Could benefit from specific guidelines on clinical performance metrics.

4. Final Recommendations & Compliance Checklist

For a healthcare organization deploying an AI triage system, the following compliance and ethical steps are advised:

- **Data Protection Measures:**

  Implement strong encryption and anonymization protocols.

  Establish clear patient consent and data handling policies in line with both local (Pakistan) and international (EU, US) regulations.

- **Bias Detection and Mitigation:**

  Use continuous bias auditing and incorporate red-teaming exercises.

  Engage diverse clinical experts to evaluate algorithmic fairness and accuracy.

- **Transparency and Explainability:**

  Publicly document the data sources, processing techniques, and model performance.

  Ensure that the decision-making process is interpretable by clinicians and patients, meeting the disclosure requirements similar to those in California Bill AB 2013 and the EU AI Act.

- **Liability and Accountability Structures:**

Define clear responsibilities for AI developers, hospital IT teams, and clinical staff.

Develop protocols for incident reporting, post-deployment monitoring, and regular audits as prescribed in Executive Order 14110 and the Pakistan Regulation of AI.

- **Security Controls:**

  Adopt state-of-the-art cybersecurity measures to protect sensitive health data.

  Regularly update and patch systems to mitigate emerging threats.

Practical Guidance:

- **Corporate Policy Changes:**
  Update internal policies to incorporate cross-jurisdictional regulatory requirements, ensuring compliance with local laws (Pakistan Regulation) and international frameworks (EU AI Act, NIST RMF).
- **Staff Training:**
  Conduct ongoing training sessions for both technical staff and healthcare professionals on AI ethics, bias mitigation, and data security.
- **Technology Improvements:**
  Invest in AI explainability tools and performance monitoring systems to maintain transparency and ensure accountability throughout the AI lifecycle.

Practical Guidance

Corporate Policy Changes:
Healthcare organizations should update their corporate policies to include explicit guidelines on AI governance, particularly focusing on data handling, transparency, and accountability. This involves establishing an internal AI governance board tasked with continuous monitoring and periodic audits to ensure compliance with both local and international regulations. Policies should mandate robust risk management practices, regular updates to the AI system based on emerging threats, and clear protocols for addressing potential errors or biases in diagnostic outputs. Furthermore, organizations should develop detailed incident reporting and escalation procedures, integrating lessons learned from both regulatory frameworks (e.g., the EU AI Act and NIST RMF) and industry best practices.

Staff Training and Technology Improvements:
In parallel with policy updates, staff training programs should be instituted to build competency in AI ethics, data security, and bias mitigation. Regular training sessions

can help both technical teams and clinical staff understand the inner workings of the AI system, including its decision-making processes and limitations. Additionally, technology improvements are essential: adopting state-of-the-art AI explainability tools, deploying advanced cybersecurity measures, and instituting continuous performance monitoring systems can ensure that the AI triage system operates effectively and safely. These measures not only enhance the system's reliability but also build trust among healthcare professionals and patients.

Reflection:

One of the main challenges was aligning the broad language of various regulations with the specific needs of a healthcare diagnostics system. While frameworks like the Pakistan Regulation of AI, California Bill AB 2013, Executive Order 14110, the EU AI Act, and the NIST AI RMF offer useful principles, applying them directly to a clinical setting proved difficult. It required careful interpretation to convert legal texts into practical steps.

Additionally, creating a universal compliance checklist for AI across different jurisdictions is challenging due to varying regulatory approaches. This experience has shown me that ongoing collaboration among legal experts, healthcare professionals, and technologists is crucial to ensure AI not only meets regulatory standards but also improves patient care and safety.

References:

- Pakistan Regulation of AI (2024) https://senate.gov.pk/uploads/documents/1725968951_269.pdf
- California Bill AB 2013 (2023–2024) https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202320240AB2013
- Executive Order 14110 https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence
- EU AI Act https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689
- NIST AI RMF https://www.nist.gov/itl/ai-risk-management-framework