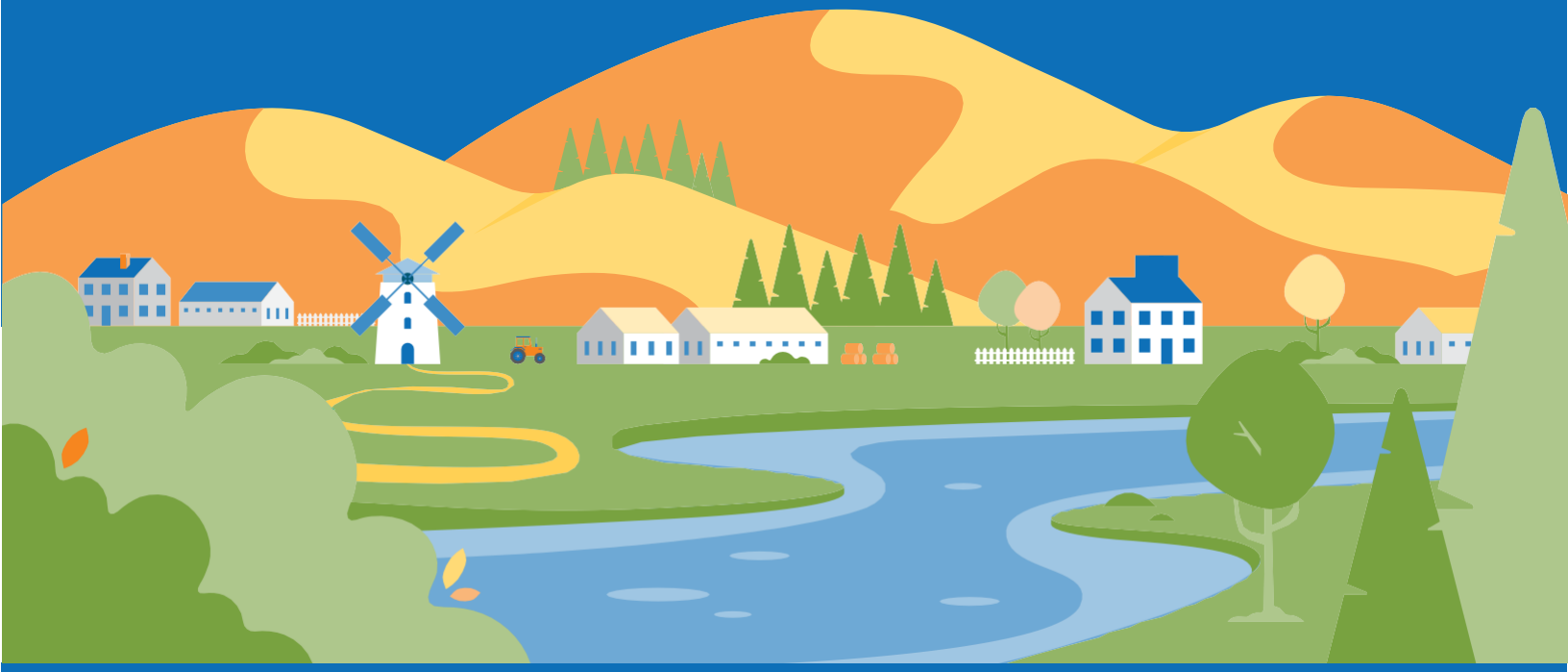
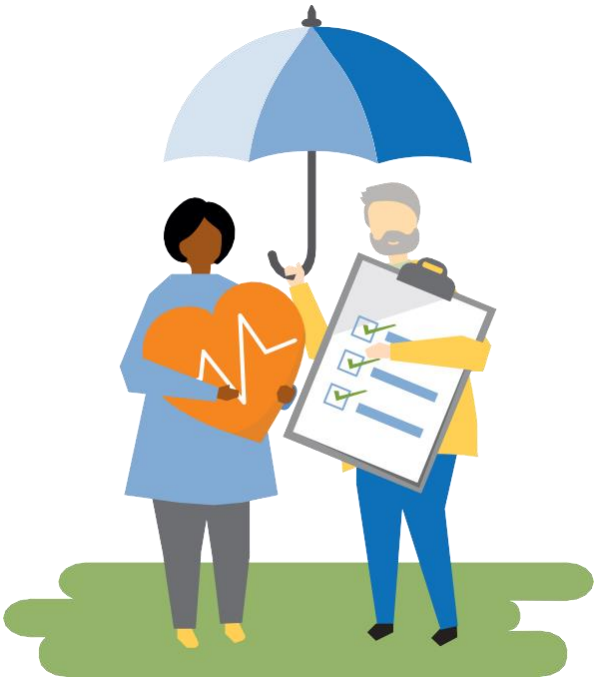
Health Care Artificial Intelligence Unintended Impact



CALIFORNIA TELEHEALTH RESOURCE CENTER

Part of OCHIN

2025



The California Telehealth Resource Center (CTRC) and all resources and activities produced or supported by the CTRC are made possible by grant number U1UTH42520-01-01 from the Office for the Advancement of Telehealth, Health Resources and Services Administration, DHHS, and the California State Office of Rural Health (CalSORH). This information or content and conclusions are those of the CTRC and should not be construed as the official position or policy of, nor should any endorsements be inferred by HRSA, HHS, the U.S. Government, CalSORH or the State of California.

CTRC does not provide legal advice. CTRC has no relevant financial interest, arrangement, or affiliation with any organizations related to commercial products or services mentioned in this toolkit. It is for informational purposes only.

KEY STEPS TO ASSESS AND MITIGATE RISK OF UINTENDED IMPACTS OF AI SYSTEMS IN HEALTH CARE

Pre-Deployment Assessment

Objective: Identify potential algorithmic adverse biases and unintended impacts (risks, gaps, or performance issues) *before* the AI system is deployed. A resource like the MIT AI Risk Repository can help providers identify the risks they may encounter.

Understand the Context:

- **Conditions of Deployment:** Analyze the characteristics of the community served by the clinic or hospital, including common chronic conditions, care access challenges, and service utilization patterns.
- **Health Care Needs:** Identify what the AI system should help address within the community (e.g., high rates of uncontrolled chronic disease, preventive care, limited specialty care access).

Algorithmic Fairness Review:

- **Adverse Bias Identification:** Obtain from the developer/vendor and carefully review their disclosures on how the AI system's algorithms were trained, including training data sources, feature selection, and decision-making processes.
- **Fairness and Representative Analyses:** Ask the developer/vendor what fairness metrics were used to assess potential unintended adverse bias.
- **Validation in Comparable Settings:** Request documentation on whether the AI tool has been validated in similar environments and patient populations to your setting. Ask how effectiveness was measured.

Stakeholder Involvement:

- **Frontline Input:** Engage a varied group of clinicians and care teams to provide input on the tool's relevance, ease of use, and any concerns about reliability and unintended risks.
- **Patient Perspective:** Where possible, seek patient feedback on how AI-generated insights might be communicated or acted upon, especially when managing long-term conditions like diabetes or hypertension.

Data Review and Preparation:

- **Data Quality and Representation:** Ensure that the developer/vendor provides you with a report on the training data and discusses whether their model training data includes patients with characteristics similar to your population. Also verify accuracy, source legitimacy, and legal rights to use the data. Check for any known data gaps.
- **Data Cleaning:** Ask developer/vendor about any steps taken to clean data, ensure the data is free from errors, and minimize the influence of outliers or historical inconsistencies.

Deployment Phase

Objective: Implement the AI system with safeguards to minimize unintended risks and impact and continuously monitor its performance.

Pilot Testing:

- **Conditions of Deployments:** Ask developer/vendor for information on deployment in similar settings and ask for recommendations from providers that have already implemented systems.
- **Controlled Environment:** Deploy the AI system in a controlled environment or as a pilot program to monitor its initial performance and impact.
- **Performance Evaluation:** Measure the AI system's performance across different demographic groups and settings to identify any differences in outcomes.
- **Version History:** Tie each model build to a version or some form of traceability to the exact snapshot of data it learned from.

Monitoring and Feedback Mechanisms:

- **Monitoring:** Implement monitoring tools to regularly assess the AI system's performance and its impact on different patient groups. The frequency of the assessment should be based on the level of risk. Clinical applications typically provide the highest levels of risk. Accordingly, these should be monitored more frequently.
- **Feedback Loops:** Establish feedback mechanisms for healthcare providers and patients to report any concerns about the AI system's recommendations or decisions, including unexpected results, workflow disruptions, or inconsistent recommendations.

Post-Deployment Assessment

Objective: Continuously evaluate the AI system and take corrective actions when necessary

Regular Audits:

- **Performance Audits:** Conduct regular audits to assess the AI system for any emerging issues, such as declining accuracy or misclassification of high-risk patients. Ensure good documentation like dataset cards, datasheets, and lineage graphs so future teams can audit or reproduce behaviors.
- **Outcome Analysis:** Compare clinical and operational outcomes across different demographic groups to ensure the tool is having its intended impact.

Impact Mitigation Strategies:

- **Human Oversight:** Ensure that healthcare providers review and validate AI recommendations, especially in cases where decisions have significant impacts on patient care.
- **Autonomous AI System Accountability:** If an AI system does not have a human in the loop, ensure that developer/vendor has assumed liability and has mechanisms to monitor, respond quickly, and notify your team.

Transparency and Accountability:

- **Documentation:** Maintain documentation of the AI system's deployment, and evaluation processes, including any actions taken to adjust or improve the tool.
- **Reporting:** Ensure your team and/or your vendor regularly report on the AI system's performance and any identified inaccuracies or corrective actions to stakeholders, including patients, staff, and regulatory bodies.
- **AI Champion:** Appoint a staff member whose responsibilities include promoting best practices and clear accountability and reporting on AI tool performance. They can serve as the liaison between clinical staff and AI governance committee.

Continuous Improvement

Objective: Establish a culture of continuous improvement to adapt to changing demographics, medical practices, and emerging technologies.

Ongoing Training:

- **Staff Training:** Provide ongoing training for staff on the ethical use of AI, how to interpret outputs, and when to escalate concerns. Include examples based on real use cases like chronic disease alerts or patient risk scores.
- **Organizational and Patient Engagement:** Engage with the staff and patients to understand their evolving needs and concerns regarding AI in healthcare.

Conclusion

Deploying an AI system in a healthcare setting, particularly in community health clinics or rural hospitals and for high-impact areas like chronic disease, requires a comprehensive approach to assess and mitigate any unintended impact. By following these key steps—pre-deployment assessment, careful deployment, continuous post-deployment evaluation, and ongoing improvement—healthcare providers can ensure safe, reliable, and clinically appropriate implementation of AI systems contribute high-quality care for all patients.