

Lessons Learned and Iterative Enhancements: A 2-Year Experience with an Established Responsible AI Framework

Supplemental Resources

Titles and legends

Supplemental Document 1. Version 3 (v3) of the UNC Health responsible AI (RAI) survey distributed to vendors of an AI-enabled solution

Supplied in this PDF.

Supplemental Document 2. The UNC Health intake form gathers a snapshot of AI-enabled solutions details and is used to determine the appropriate responsible AI (RAI) evaluation pathway based upon risk.

Supplied in this PDF.

Supplemental Data 1. The underlying structured data generated from responsible AI (RAI) requests, tools assessed, vendor responses received, and evaluations performed at UNC Health from October 10, 2024 to March 19, 2026. Each row represents one evaluation but note that some tools have undergone multiple evaluations.

Supplied in a separate document, “Torre et al – Supplemental Data 1.xlsx”.

Responsible AI Framework – Vendor Survey

This framework evaluates whether artificial intelligence (AI) used in a solution considered for use in the UNC Health system was developed responsibly through questions in four main pillars: Fairness, Transparency, Accountability, and Trustworthiness. Our definition of AI is intentionally broad, encompassing regression, machine learning, ensemble methods, and generative models. The answers should be provided by data scientists or other technical subject-matter experts (SMEs) who were involved in the AI development. If referring to outside links and documents, provide these resources but also copy the relevant text in your answer. All information shared in the answers will be confidential and not distributed to members outside the evaluation bodies. Though you may choose to protect proprietary intellectual property, it is *strongly* recommended that the answers provide as much detail as possible for adequate review by a group of subject matter experts across a wide group of fields and specialties, including but not limited to, clinicians, data scientists, and community health professionals. Answers will be scored from 1 to 5 and will play a role in the subsequent decision to adopt.



1. Purpose

- 1.1. Explain what problem your solution solves.
- 1.2. Describe your solution, how AI is employed, and how the problem is solved.
- 1.3. Specify the gold standard/traditional/alternative solutions to the problem and why your solution is preferred.
- 1.4. Present an example use case of your solution from start to finish.

2. Development

- 2.1. Detail all roles involved in the development of your solution.
- 2.2. Define the targets of your solution:
 - 2.2.1. Intended use cases
 - 2.2.2. Out-of-scope use cases
 - 2.2.3. Intended users
- 2.3. Specify the number and type(s) of AI used in your solution.

- 2.4. Outline the AI development process and analytical choices made.
- 2.5. Provide the sample sizes and feature distributions of the development dataset(s) (i.e., training, tuning). Include stratification by sex, age, race, ethnicity, and other patient characteristics.
- 2.6. Supply a risk gradient (risk severity vs risk likelihood) for clinical risk, misuse, and affects to equitable healthcare.

3. AI Details

- 3.1. Provide the input(s), and specify which inputs are essential. Also discuss how missing data is handled.
- 3.2. Provide the output(s). Specify whether it is a prediction, classification, recommendation, evaluation, analysis, or other type of output.
- 3.3. Explain the limitations of your solution and how it may affect performance.
- 3.4. Articulate whether inputs and/or significant features are accessible to the user.
- 3.5. Describe the decisions/actions users are expected to make based on the output. Include whether it informs, augments, and/or replaces clinical management.
- 3.6. Detail how the model works.

4. Validation

- 4.1. Detail all roles involved in validation of your solution.
- 4.2. Outline the AI validation process.
- 4.3. Specify the sample sizes and feature distributions of any test/validation dataset(s) and supply appropriate performance metric values for those dataset(s).
- 4.4. Share assessments of your solution's performance on datasets representative of UNC Health's.
- 4.5. Provide links to peer-reviewed independent validation of your solution.

5. Fairness Considerations

- 5.1. Describe how bias was managed, reduced, or eliminated, or why such investigations were not done.
- 5.2. Specify the approach/actions taken to ensure that your solution's output is fair.
- 5.3. Provide appropriate stratified performance metric values on validation dataset(s).

6. Maintenance

- 6.1. Describe processes in place to continuously monitor the following and whether those processes are accessible to the user. If none are in place, explain why.
 - 6.1.1. Input and output variables (e.g., for target dataset changes or drift).
 - 6.1.2. Performance. Include the performance level needed for the intended outcome.
- 6.2. Outline processes in place to calibrate/retrain your solution. Include what data is used and either the frequency or criteria required to trigger this process. If none are in place, explain why.

7. Other

- 7.1. Provide additional information about your solution not covered above



Responsible AI Framework – Vendor Survey: Supplement

Question	Pillars	What we're looking for
1. Purpose		
1.1 Explain what problem your solution solves.		
1.2 Describe your solution, how AI is employed, and how the problem is solved.		<ul style="list-style-type: none"> Note, our definition of AI is intentionally broad, encompassing regression, machine learning, ensemble methods, and generative models.
1.3 Specify the gold standard/traditional/alternative solutions to the problem and why your solution is preferred.		<ul style="list-style-type: none"> Ideally, a table containing the values of performance metric(s) appropriate for your model and values for your competitors/alternatives in this space. Include the pros and cons of each approach
1.4 Present an example use case of your solution from start to finish.		
2. Development		
2.1 Detail all roles involved in the development of your solution.	Fairness Transparency	<ul style="list-style-type: none"> A list of names and/or groups of a specific domain expertise, including the corresponding role of each in the development and implementation
2.2 Define the targets of your solution: 2.2.1 Intended use cases 2.2.2 Out-of-scope use cases 2.2.3 Intended users	Accountability	<ul style="list-style-type: none"> Unambiguous users and use cases to show that the development was purposeful and well thought out. Use cases should include a description of tasks, situations, and/or populations (e.g., inpatient, outpatient, care setting, age grouping (pediatrics, etc), disease states)
2.3 Specify the number and type(s) of AI used in your solution	Accountability	
2.4 Outline the AI development process and analytical choices made.	Accountability	<ul style="list-style-type: none"> AI model type choice justification Study design, and how data was collected Data selection, e.g., inclusion/exclusion criteria Data transformations, intermediate processes
2.5 Provide the sample sizes and feature distributions of the development dataset(s) (i.e., training, tuning). Include stratification by sex, age, race, ethnicity, and other patient characteristics.	Fairness Trustworthiness	<ul style="list-style-type: none"> A table containing descriptive statistics for all model inputs for the development dataset(s) Justify that the training data is representative of the intended use case(s) Applications using patient data, regardless of direct use in the model, should include a table of distributions for race, ethnicity, sex, age, and health status. Consider including language, sexual orientation, gender identity, and social determinants of health.
2.6 Supply a risk gradient (risk severity vs risk likelihood) for clinical risk, misuse, and affects to equitable healthcare.	Accountability	<ul style="list-style-type: none">
3. AI Details		
3.1 Provide the input(s), and specify which inputs are essential. Also discuss how missing data is handled.	Transparency Accountability Trustworthiness	<ul style="list-style-type: none"> List of the variables, image features, and/or predictors Inclusive of hyperparameters like temperature, sampling, and sequence length for large language models or depth and samples per leaf for decision trees.

Question	Pillars	What we're looking for
3.2 Provide the output(s). Specify whether it is a prediction, classification, recommendation, evaluation, analysis, or other type of output.	Transparency	
3.3 Explain the limitations of your solution and how it may affect performance.	Transparency Accountability Trustworthiness	<ul style="list-style-type: none"> • <i>specific use cases when model performance may deteriorate</i> • <i>model assumptions</i>
3.4 Articulate whether inputs and/or significant features are accessible to the user.	Accountability Trustworthiness	
3.5 Describe the decisions/actions users are expected to make based on the output. Include whether it informs, augments, and/or replaces clinical management.	Transparency Accountability Trustworthiness	
3.6 Detail how the model works.	Transparency	<ul style="list-style-type: none"> • <i>e.g., for neural networks, specify the type and number of layers, filters, activation functions</i> • <i>e.g., regression, specify the type and importance (coefficients) for predictors</i> • <i>e.g., for generative models, specify how the prompts are generated</i>
4. Validation		
4.1 Detail all roles involved in validation of your solution.	Transparency Accountability Trustworthiness	<ul style="list-style-type: none"> • <i>A list of names and/or group of a specific domain expertise</i> • <i>We want an answer that tells us that subject matter experts reflective of all potential users in 2.2.3. were involved in evaluation</i>
4.2 Outline the AI validation process.	Transparency Accountability	<ul style="list-style-type: none"> • <i>Was the dataset a holdout from the collected data used for training?</i> • <i>What performance metrics are appropriate for the model?</i> • <i>Was there a target performance level?</i> • <i>What was used as the gold standard for which to compare performance?</i> • <i>If implementing a third-party model, specify the procedures that gave you confidence in commercializing this solution.</i>
4.3 Specify the sample sizes and feature distributions of any test/validation dataset(s) and supply appropriate performance metric values for those dataset(s).	Fairness Accountability Trustworthiness	<ul style="list-style-type: none"> • <i>A table containing descriptive statistics for all model inputs for the test/validation dataset(s)</i> • <i>Applications using patient data, regardless of direct use in the model, should include a table of distributions for race, ethnicity, sex, age, and health status. Consider including language, sexual orientation, gender identity, and social determinants of health.</i> • <i>Appropriate performance metrics for the AI model may include accuracy, sensitivity, specificity, F1 score, mean squared error, intersection over union, and/or mean average precision: https://oecd.ai/en/catalogue/metrics</i>
4.4 Share assessments of your solution's performance on datasets representative of UNC Health's.	Fairness Trustworthiness	

Question	Pillars	What we're looking for
4.5 Provide links to peer-reviewed independent validation of your solution.	Trustworthiness	<ul style="list-style-type: none"> • Ideally, citations from academic papers or white papers authored by people outside your organization or who were not involved in your model's development or implementation; however, please provide any peer reviewed research.
5. Fairness Considerations		
5.1 Describe how bias was managed, reduced, or eliminated, or why such investigations were not done.	Fairness Accountability Trustworthiness	<ul style="list-style-type: none"> • We define bias to be the influence of human or systemic prejudice on the training data or AI algorithms, leading to distorted outputs and potentially harmful outcomes. • Did you consider representation, data quality, outcome disparities, and ethics?
5.2 Specify the approach/actions taken to ensure that your solution's output is fair.	Fairness Accountability	
5.3 Provide appropriate stratified performance metric values on validation dataset(s).	Fairness Accountability Trustworthiness	
6. Maintenance		
6.1 Describe processes in place to continuously monitor the following and whether those processes are accessible to the user. If none are in place, explain why. <ul style="list-style-type: none"> 6.1.1 Input and output variables (e.g., for target dataset changes or drift). 6.1.2 Performance. Include the performance level needed for the intended outcome. 	Accountability	
6.2 Outline processes in place to calibrate/retrain your solution. Include what data is used and either the frequency or criteria required to trigger this process. If none are in place, explain why.	Accountability Trustworthiness	
7. Other		
7.1 Provide additional information about your solution not covered above		



Responsible AI Framework – Intake Risk Tier Form

This assessment provides comprehensive details about your AI-enabled application to ensure responsible deployment. Please answer each question thoughtfully; your responses will help determine the risk tier and guide next steps. An asterisk (*) denotes a required response.

Section 1 — Project & Contacts

1. What is the name of the vendor and application? *
2. What is the technical vendor contact email? *
3. What is the ServiceNow demand number? *
4. Who is the ISD application owner? *

Section 2 — Use Case

5. In what area(s) will this tool be used? e.g., radiology, cardiology, coding, finance, billing. *
6. What is an end-to-end example use case of the solution, from start to finish? *
7. Does the tool replace or automate an existing manual workflow? * (Yes or No)
8. Will the tool undergo a pilot implementation before a full system roll-out? * (Yes or No)

Section 3 — Model & Vendor

9. What type of AI is used in the tool? e.g., regression, machine learning, generative AI. *
10. What is the current model version? *
11. What changes have been made to the AI model since the last version? *
12. Is the AI model custom-built for this tool? * (Yes or No)
13. Are the application vendor and the AI model developer the same? * (Yes or No)
14. Who is/are the AI model developer(s)? *
15. If the model is generative, will the application vendor provide the system prompt(s)? * (Yes, No, or N/A)
16. Is the AI output post-processed or otherwise modified within the application? * (Yes or No)
17. Please specify the post-processing performed. *

Section 4 — Data & Privacy

18. Does the tool use PHI or PII as input? * (Yes or No)

Section 5 — Use & Impact

19. Will there be a human-in-the-loop in the intended workflow? * (Yes or No)
20. Does the tool make or recommend a clinical decision? * (Yes or No)
21. What clinical decision(s) will be made or recommended, and how will the outputs be used? *
22. Does the tool make or recommend an operational decision? * (Yes or No)
23. What operational decision(s) will be made or recommended, and how will the outputs be used? *
24. Is there any associated risk with this tool? e.g., safety, misuse, equity impacts, staffing/resource allocation. *
(Yes or No)
25. What risks are anticipated and what mitigations or safeguards are planned? *
26. Are any outputs patient-facing? * (Yes or No)

Section 6 — Safety, Quality, & Compliance

27. Has the AI model received any external certifications or regulatory approvals? * (Yes or No)
28. Which certifications or approvals apply? e.g., FDA clearance, CE mark, ISO