The content that follows is work in-progress and in draft form available for public comment beginning on 11/22/2024 and ending on 01/22/2025. The mention or sharing of any examples, products, organizations, or individuals does not indicate any endorsement of those examples, products, organizations, or individuals by the Coalition for Health AI (CHAI). Any examples provided here are still under review for alignment with existing standards and instructions. We welcome feedback and stress-testing of the tool in draft form.

The information provided in this document is for general informational purposes only and does not constitute legal advice. It is not intended to create, and receipt or review of it does not establish, an attorney-client relationship.

This document should not be relied upon as a substitute for consulting with qualified legal or compliance professionals. Organizations and individuals are encouraged to seek advice specific to their unique circumstances to ensure adherence to applicable laws, regulations, and standards.

Special thank you to the contributing members of the CHAI Model Card Work Group including:

Demetri Giannikopoulos, Aidoc

Eric Williamson, MD, Mayo Clinic

Suchi Saria, PhD, Bayesian Health

Jon McManus, Sharp Healthcare

Sara Liu, PhD, Amazon

Lindsay Mico, M.S., Providence Health

Naveen Valluri, M.S., Microsoft

Sophia Bessias, M.S., Duke

Fawad Butt, MBA, Penguin AI

Celena Wheeler, RN, Oracle

Sneh Shah, MBA, Bayesian Health

Warren D'Souza, PhD, MBA , University of Maryland Medical

Rebecca Weintraub, MD, Ariadne Labs

Christine Swisher, PhD, Oracle

Ted Robertson, MPA, Center for Healthcare Marketplace Innovation, University of Berkeley

Merage Ghane, PhD, CHAI

Brenton Hill, JD, CHAI

Lucy Orr Ewing, CHAI

Sumanth Ratna, CHAI

The CHAI Applied Model Card

Coalition for Health AI

Table of Contents

1. Introduction

The applied model card describes an AI system focused on the application in a health use case. An AI system is a fully operational AI use case, including the model, technical infrastructure, and personnel in the workflow.

This model card supports meeting the criteria for HTI-1 for predictive DSIs defined as "...technology that supports decision-making based on algorithms or models that derive relationships from training data and then produce an output that results in prediction, classification, recommendation, evaluation, or analysis." In addition, this model card provides transparency for all Five of CHAI's Principles of Responsible AI (Transparency, Safety, Security & Privacy, Fairness & Bias, and Usefulness).

Individuals or teams completing the model card may include vendor or health organization developers of applied AI solutions (see appendix 5.3 for definitions). The primary intended use of this model card is for transparency purposes and may be used as part of transparency processes during procurement processes, or as internal transparency documentation for tracking or governance of AI solutions employed within a health organization (e.g. health systems, hospitals, or payers). The applied model card is intended for an applied **AI solution** in context, not an **AI model** outside of an intended use/purpose (e.g. general foundation models, or AI platforms used to build AI solutions/applications. This model card does not replace optional or requested external validation/evaluation of AI solutions by an independent evaluator such as a Quality Assurance Lab, but may include description of findings from such if used. CHAI recommends the developer issue a revised report card tied to the software revision number when there are any changes to: 1) the Uses and Directions section, 2) a new risk or adverse events is discovered 3) significant changes in Key Metrics, or 4) structural changes to how the model is developed or trained. For

changes to the key metrics, the recommendation is for the AI solution provider and the implementing institution to determine the proper cadence or standard deviation from the current metrics at which the metrics should be re-evaluated. We recommend implementers and users to ask for updated cards annually.

Examples of structural changes include:

- Base foundation model change (e.g. GPT-4, LLama)
- New training dataset with new patient population represented (e.g. new geography, adding pediatrics)
- Takes in new type of Imaging acquisition type or anatomy (e.g. T2W, DWI, FDG-PET)
- Announcing a new revision to customers from 2.0 to 3.0

2. The Template

<table>
<tr><td colspan="2">Name:
Developer:</td><td>Inquires or to report an issue:
abc@abc.com or +1 (999) 999- 9999</td></tr>
<tr><td colspan="2">Release Stage:
Global Availability:</td><td>Release Date:
Regulatory Approval, If applicable:
Version:</td></tr>
<tr><td colspan="2">Summary:

Keywords:</td><td>Uses and Directions:
• Intended use and workflow:
• Primary intended users:
• How to use:
• Targeted patient population:
• Cautioned out-of-scope settings and use cases:</td></tr>
<tr><td colspan="3">Warnings</td></tr>
<tr><td colspan="3">• Known risks and limitations:
• Known biases or ethical considerations:
• Clinical risk level:</td></tr>
<tr><td colspan="3">Trust Ingredients</td></tr>
<tr><td colspan="3">AI System Facts:
• Outcome(s) and output(s):
• Model type:
• Foundation models used in application, if applicable:
• Input data source:
• Output/Input data type:
• Development data characterization:
• Bias mitigation approaches:
• Ongoing Maintenance:</td></tr>
</table>

- **Security and compliance environment practices or accreditations, if applicable:**
- **Transparency, Intelligibility, and Accountability mechanisms, if applicable:**

Transparency Information:

- **Funding source of the technical implementation:**
- **3rd Party Information, If Applicable:**
- **Stakeholders consulted during design of intervention (e.g. patients, providers):**

Key Metrics

Usefulness, Usability, and Efficacy		Fairness and Equity		Safety and Reliability	
Goal of metric(s):		**Goal of metric(s):**		**Goal of metric(s):**	
Result:	**Interpretation:**	**Result:**	**Interpretation:**	**Result:**	**Interpretation:**
Test Type:		**Test Type:**		**Test Type:**	
Testing Data Description:		**Testing Data Description:**		**Testing Data Description:**	
Validation Process and Justification:		**Validation Process and Justification:**		**Validation Process and Justification:**	

Resources

- **Evaluation References, If Available:**
- **Clinical Trial, If Available:**
- **Peer Reviewed Publication(s):**
- **Reimbursement status, if applicable:**
- **Patient consent or disclosure required or suggested:**
- **Stakeholders consulted during design of solution:**

3. Instructions to Complete the Applied Model Card

Text in blue, including references, denote instructions only. All instructions, CFR referenced or not, support alignment with HTI-1 and CHAI's five principle areas (Transparency, Usefulness, Fairness, Safety, Privacy & Security; see definitions in Appendix section 5.3)

Name: **Developer:** [45 CFR 170.315 (b)(11)(iv)(B)(1)(i)]	**Inquires or to report an issue:** [45 CFR 170.315 (b)(11)(iv)(B)(1)(i)] abc@abc.com or +1 (999) 999- 9999
Release Stage: **Release Date:** **Version:** Model / Software Release Version **Global Availability:** **Regulatory Approval, If applicable:**	
Summary: Please provide a summary of the AI solution's intended uses, key features, clinical workflow and key performance metrics. **Keywords:**	**Uses and Directions:** • **Intended use and workflow:** Please describe how the solution should be used in clinical practice, including the following details: ○ Whether the solution is intended to inform, augment, or replace clinical management [45 CFR 170.315(b)(11)(iv)(B)(*2*)(iv)]. ○ Whether human oversight or a "human in the loop" is required to operate the solution, and any actions clinicians should take at certain steps of the clinical workflow.

	○ Specific use cases and/or diseases that this solution addresses. • **Primary intended users:** ○ Please specify the target users of the solution, including any necessary knowledge or expertise they should have prior to use. ○ Please outline any inclusion or exclusion criteria based on patient population or other key variables, as applicable. • **How to use:** • **Targeted patient population:** • **Cautioned out-of-scope settings and use cases:** [45 CFR 170.315(b)(11)(iv)(B)(*3*) (i-ii)] ○ Please describe any tasks, situations, or patient populations where the use of this solution is cautioned or discouraged. ○ Please provide all known risks, inappropriate settings or uses, and any limitations of the solution that users should be aware of.

Warnings

- **Known biases or ethical considerations:** 45 CFR 170.315(b)(11)(iv)(B)(*3*) (i-ii)]
  - Please describe any tasks, situations, or patient populations where the use of this solution is cautioned or discouraged.
  - Please provide all known risks, inappropriate settings or uses, and any limitations of the solution that users should be aware of.
- **Clinical risk level:** See resources in section 5.2 including "Software as a Medical Device": Possible Framework for Risk Categorization and Corresponding Considerations

Trust Ingredients

AI System Facts:

- **Outcome(s) and output(s):** [45 CFR 170.315 (b)(11)(iv)(B)(1)(iv)]
  - Please describe the type and value of the solution output. Specify whether the output is a prediction, classification, recommendation, evaluation, analysis, or another form of result.
- **Model type:**
  - Please determine if this model is predictive or generative (as well as relevant subclass within the broader category), and describe how it interacts with any other systems, such as EHRs, medical devices, or any other integrated platforms.
- **Foundation models used in application, if applicable:**
  - Please list any foundational models (with version information) utilized in the solution, such as ChatGPT 4, Claude 3, or others.
- **Input data source:**
  - Please specify the source of data that is necessary as input into the AI solution.
- **Output/Input data type:** [45 CFR 170.315(b)(11)(iv)(B)(*4*) (i-iii)]
  - Please specify the types of data used by the solution (e.g., EHR, waveform, imaging, etc.).
  - Please indicate whether the data is real-world data or synthetic data.
  - Please outline exclusion and inclusion criteria that influenced the training data set.
  - Please clarify the use of following variables as input features: race, ethnicity, language, sexual orientation, gender identity, sex, date of birth, social determinants of health elements (USCDI v3), and health status assessments (USCDI v3).
  - Please describe the demographic representativeness based on race, ethnicity, language, sexual orientation, gender identity, sex, date of birth, social determinants of health elements (USCDI v3), and health status assessments (USCDI v3).
- **Development data characterization:** [45 CFR 170.315(b)(11)(iv)(B)(*4*) (i-iv)]
  - Please provide details on the data used to train the solution, including Exclusion and inclusion criteria that influenced the training dataset.

- Use of the variables as identified in USCDI v3:
  - Race, Ethnicity, Language, Sexual Orientation, Gender identity, Sex, Date of Birth, Social determinants of health data, Health status assessment data
- Description of the demographic representativeness of the data, including demographic distribution as identified in USCDI v3
  - Race, Ethnicity, Language, Sexual Orientation, Gender identity, Sex, Date of Birth, Social determinants of health data, Health status assessment data
- Description of relevance and alignment of training data to the intended deployment setting.

- **Bias mitigation approaches:** [45 CFR 170.315(b)(11)(iv)(B)(5) (i-ii)]
  - Please describe the approach the developer(s) took to ensure the output is fair.
  - Please describe the approaches to manage, reduce, or eliminate bias.
- **Ongoing Maintenance:** Update Schedules, Monitoring, and Response Approach including fairness performance
- [45 CFR 170.315(b)(11)(iv)(B)(*8*) (i-iv) & 45 CFR 170.315(b)(11)(iv)(B)(9) (i-ii)]
  - Please provide the following details regarding to the ongoing maintenance and monitoring of the AI solution:
    - **Monitoring Validity**: Describe the process and frequency for monitoring the solution's validity over time, including its validity when applied to local data.
    - **Monitoring Fairness:** Outline the process and frequency for assessing the solution's fairness, both in general and when applied to local data.
    - **Update Process**: Explain the process and frequency for updating the solution to ensure continuous improvement.
    - **Risk Correction:** Describe how often the solution's performance is corrected when risks related to validity or fairness are identified.
    - **Monitoring Tools**: List any built-in monitoring tools or functionalities provided by the developer or quality assurance lab.
    - **Anticipated Improvements**: Note any anticipated changes or improvements based on ongoing monitoring and evaluation efforts.
- **Security and compliance environment practices or accreditations, if applicable:**
  - Please provide a list of accreditations or practices adhered to with respect to security and compliance. E.g., *SOC2 Type 2, ISO, NIST, Fed Ramp, etc.*
- **Transparency, Intelligibility, and Accountability mechanisms, if applicable:**
  - Please provide descriptions on the implemented features or mechanisms regarding transparency, intelligibility, and accountability, such as providing saliency maps, confidence indicators as part of the system output, and feedback channels.

Transparency Information:

- **Funding source of the technical implementation:** [45 CFR 170.315 (b)(11)(iv)(B)(1)(ii)]
  - Please provide information regarding the funding source of the technical implementation for the solution's development
- **3rd Party Information, If Applicable:**
  - Please provide name, product, contact information for third parties integrated into the overall solution.
- **Stakeholders consulted during design of intervention (e.g. patients, providers):**

Key Metrics

Usefulness, Usability, and Efficacy		Fairness and Equity		Safety and Reliability	
Goal of metric(s):		**Goal of metric(s):**		**Goal of metric(s):**	
Result:	**Interpretation:**	**Result:**	**Interpretation:**	**Result:**	**Interpretation:**
Test Type:		**Test Type:**		**Test Type:**	
Testing Data Description:		**Testing Data Description:**		**Testing Data Description:**	
Validation Process and Justification:		**Validation Process and Justification:**		**Validation Process and Justification:**	

[45 CFR 170.315(b)(11)(iv)(B)(*6*) (i-iv) & 45 CFR 170.315(b)(11)(iv)(B)(*7*) (i-v)]

For each of the principle areas above (definitions provided in Appendix Section 5.3), describe the **specific metric and goal** of the metrics selected, the **quantitative results** of performance relevant to those principle areas, the **interpretation** of the quantitative results, the **type of test** used, the **description of the data used** to conduct the test, and the **validation process and justification as it relates to the AI solution.** See relevant additional information below to help with scoping of this information.

- **External Validation Process and Measures of Performance**

- Description of the data source, clinical setting, or environment where an intervention's validity and fairness has been assessed other than training data source
- Party that conducted the external testing
- Description of the external validation process
- Description of validation impact metrics

- **Quantitative Measures of Performance**
  - Validity of intervention in test data derived from same source as initial training data
  - Fairness of intervention in test data derived from same source as initial training data
  - Validity of intervention in data external to or from a different source than the initial training data
  - Fairness of the intervention in data external to or from a different source than the initial training data
  - References to evaluation of use of the intervention on outcomes including, bibliographic citations or hyperlink to how well the intervention reduced morbidity, mortality, length of stay, or other outcomes
- **Fairness & Equity**
  - Describe fairness and equity across notable subgroups such as but not limited to national origin, race, color, ethnicity, disability, legal and self-reported sex, age, relevant social determinants of health as applicable
  - In the Validation Process & Justification, please describe what was and was not considered and why.

Test Type Definitions:

- Internal as defined as validation set that comes from the same population as the training set (e.g. cross validation, train-test split)
- External as defined as validation set that comes from a different population
- Local as defined as validation set from data sourced from the same healthcare system or institution
- Prospective as defined as validation set that evaluates performance before it is used in real-world implementation

Resources

- **Evaluation References, If Available:**
- **Clinical Trial, If Available:**
- **Peer Reviewed Publication(s):** Please indicate the status of each (e.g., peer-reviewed, abstract, or under review)
- **Reimbursement status, if applicable:**
- **Patient consent or disclosure required or suggested:**
  - Please indicate if the developer recommends obtaining patient informed consent or providing disclosure for the use of the solution.
- **Stakeholders consulted during design of solution:**
  - Please indicate pertinent stakeholder groups or individuals consulted during the design of the solution. Include specific names where possible.
    - E.g., patient advocacy groups, coalitions, physician groups, individual clinicians, etc.

4. Populated Exemplar Model Card*

Name: BriefCase-Triage for Intracranial Hemorrhage (ICH) **Developer:** Aidoc	**Inquires or to report an issue:** Demetri.g@aidoc.com +1 321-243-1594
Release Stage: FDA Cleared Commercial **Release Date:** May 2022 **Version:** V2 **Global Availability:** CE Marked, UKCA, Canada, Australia, New Zealand, Israel, UAE and South Africa	
Summary: The ICH solution is a radiological computer aided triage and notification software indicated for use in the analysis of non-enhanced head CT images in adults aged 18 and older.	**Uses and Directions:** • **Intended use:** for use in the analysis of non-enhanced head CT images in adults aged 18 and older.

The device is intended to assist hospital networks and appropriately trained medical specialists in workflow triage by flagging and communication of suspected positive findings of Intracranial hemorrhage (ICH) pathologies. The ICH solution uses an artificial intelligence algorithm to analyze images and highlight cases with detected findings on a standalone desktop application in parallel to the ongoing standard of care image interpretation. The user is presented with notification for cases with suspected findings. Notifications include compressed preview images that are meant for informational purposes only and not intended for diagnostic use beyond notification. The results of the ICH solution are intended to be used in conjunction with other patient information and based on their professional judgment, to assist with triage/prioritization of medical images. Notified clinicians are responsible for viewing full images per the standard of care.	• **Primary intended users:** The BriefCase-Triage is intended to be used by appropriately trained medical specialist • **Targeted patient population:** patients 18 and over with non-contrast head CTs. • **How to use:** The user is presented with notification for cases with suspected findings. Notifications include compressed preview images that are meant for informational purposes only and not intended for diagnostic use beyond notification. The results of the ICH solution are intended to be used in conjunction with other patient information and based on their professional judgment, to assist with triage/prioritization of medical images. • **Cautioned out-of-scope settings and use cases:** Only for use in workflow triage, and only for use by appropriately trained medical specialists.

Warnings

- **Known risks and limitations:** The results of Aidoc's solutions are intended to be used in conjunction with other patient information and based on professional judgment, to assist with triage/prioritization of medical images. Notified clinicians are responsible for viewing full images per the standard of care. For false positives, it might take the radiologist additional time to read the scan. For false negatives, it also might take additional time reducing the radiologists' efficiency. Non-availability might do so as well if the radiologists are accustomed to having the Aidoc solution.
- **Known biases or ethical considerations:** None
- **Clinical risk level:** Low

Trust Ingredients

AI System Facts:

- **Outcome(s) and output(s):** The desktop application feed displays incoming suspect cases. Hovering over the feed pops up a compressed, unannotated image that is captioned "not for diagnostic use" and is displayed as a preview function. The radiologist prioritizes cases based on this additional info and reads the case as per standard of care. Output: Binary prediction of suspected ICH cases.
- **Model type:** Custom-built 3D deep convolutional neural network. Output: Binary prediction of suspected ICH cases.
- **Foundation models used in the application, if applicable**: NA
- **Input data source:** Our imaging AI solutions are trained, tuned, and validated using diverse data sets. Specifically, we work with DICOM images, which capture anatomical and physiological data from medical imaging technologies like CT scans and X-rays. These data sets are sourced from a wide range of institutions, including community hospitals, academic centers, teleradiology providers, private clinics, and imaging centers across various global regions (e.g., east and west US, central US, EU, and other parts of the world). The names of our partners are proprietary and confidential information.
- **Output/Input data type:** Input: Digital imaging and communications in medicine (DICOM) images.
- **Development data characterization:** The ICH model has been trained, tuned, and validated on over 10,000 scans. Each image processing algorithm was trained on images of the specific pathology it is intended to analyze. As is customary in the

field of machine learning, deep learning algorithm development consisted of training on manually labeled ("tagged") images. The sites from which images were obtained were from diverse geographies within the USA, as well as diverse settings including urban and rural, academic medical centers and community hospitals, etc. This was done to ensure the data cover the diversity of the US population to support the generalizability of the algorithm.
Additionally, we conduct research and validations with most of our customer sites to evaluate model performance and provide data to the sites for publication and research purposes. The time frame for the data used for the initial and retrained model construction and validation is 2015-2022, with most cases from 2021-2022.

- **Bias mitigation approaches:** At Aidoc, we take bias mitigation seriously, building AI solutions that advance care equity while working within the known constraints of limited accessibility to certain data elements. Our commitment is driven by the understanding that there is a risk of AI solutions underperforming for populations that may need them the most. With this in mind, we approach the development of our imaging AI solutions with diligence and responsibility, rigorously training, validating, and monitoring our models to ensure robust and equitable performance across diverse patient populations.

  Given that some of the aforementioned factors are not accessible to us, we take a two-fold approach: (1) maximizing what can be achieved with the data we have by analyzing key demographic elements such as age, gender, comorbidities, geographical location, patient settings, reason for exam and technological characteristics (e.g., equipment manufacturer, slice thickness, and modality model distribution); and (2) creating proxies that allow us to effectively mitigate risks associated with the absence of certain demographic details. We address potential biases by employing diverse data sets from a wide range of institutions and patient settings, including academic centers, community hospitals, teleradiology providers, private imaging centers and clinics worldwide (specifically East, West and Center of US, EU, Middle East, Oceania, Latin America and far east). We validate and measure the AI across subpopulations to ensure a normal and predictable performance on them.

  When it comes to addressing disability, our training and validation data encompass a broad range of ages, and health statuses. This ensures that our models are well-equipped to handle diverse health conditions, further supporting equitable care for individuals with disabilities.
  Notably, FDA has accepted and validated this approach across Aidoc's FDA-cleared devices.

- As part of our commitment to continuous improvement, Aidoc conducts multiple retrainings throughout the lifecycle of each AI module to ensure increased generalizability across all the aforementioned parameters. While the exact size of our training data sets is proprietary, they may include up to hundreds of thousands of DICOM studies in a product lifecycle.
- **Ongoing Maintenance:** The current ICH model is the second FDA clearance for ICH at Aidoc. Aidoc constantly trains the algorithm on Additional Data and passes the retrained versions to FDAs review. An expanded dataset enhances the model's generalization capabilities and potentially elevates the product's performance (improved or non-inferior to previously cleared device's Time-to-Notification and/or AUC). The inclusion of data from a broader and more diverse array of sources enhances the model's ability to perform effectively across a wider range of clinical and technical settings, particularly if any changes to real-world data were made or if any data was underrepresented in the original dataset.
- **Continuous monitoring:** The Aidoc AI Monitoring team monitors algorithm performance 24/7 for the purpose of mitigating AI drift. The team monitors and considers aspects such as alert correctness, the timeliness of data available, data completeness, slice thickness, series correctness and relevancy, contrast phase, algorithm specificity, and algorithmic positive ratio. We have robust monitoring systems in place to track the AI's performance and identify any emerging issues like AI drift and bias creep. Regular audits and assessments ensure the model continues to deliver fair and accurate

prioritizations. We remain committed to ongoing research and development to refine the AI and address any ethical concerns that may arise.
Aidoc quality management system is ISO 13485 certified, and 21 CFR Part 820 compliant.
Additionally, Aidoc's Governance, Risk, and Compliance (GRC) Program is responsible for enterprise oversight and direction for all Security governance activities; governance for risk related activities includes Risk Identification, Management, Mitigation, and Remediation, and Risk Assessments; and responsibility for the planning, execution, and adherence with Aidoc Security Policies and Procedures, legal, regulatory, and contractual requirements.

- **Security and compliance environment practices or accreditations, if applicable**: The Aidoc aiOS is structured based on international standards and frameworks, including ISO 27001, ISO 27017, ISO 27018, and SOC 2 Type 2. We follow a systematic approach, integrating these frameworks into our Information Security Management System (ISMS) to ensure a holistic and proactive risk management strategy. Our risk management program involves regular risk assessments, leveraging common methodologies, to identify, evaluate, and prioritize risks. Continuous monitoring, periodic audits, and assessments contribute to the dynamic nature of our risk management program, allowing us to adapt to evolving threats and vulnerabilities.
- **Transparency, Intelligibility, and Accountability mechanisms, if applicable:** The company ensures Transparency, Intelligibility, and Accountability in Aidoc's medical devices via its product design and its Quality Management System (QMS), by providing users with:
    - User guide and labeling and additional tutorial materials;
    - Robust user training;
    - Explainability and transparency by design;
    - Reporting and feedback mechanism directly via the device's UI and additional communication in different platforms.

Key Metrics

Usefulness, Usability, and Efficacy		Fairness and Equity		Safety and Reliability	
Goal of metric: Evaluate the software's performance in identifying non-contrast head CT images containing intracranial hemorrhage (ICH) in 220 cases from 5 US-based clinical sites.		**Goal of metric:** Evaluate for differences in performance (sensitivity & specificity) based on available socio-demographic variables of age, location and gender.		**Goal of metric:** The time-to-notification metric observed for the BriefCase software in the five medical centers was compared to the equivalent metric of prior predicate devices.	
Result: Sensitivity was 96.15% (95% CI: 90.44%, 98.94%) and specificity was 94.83% (95% CI: 89.08%, 98.08%).	**Interpretation:** Primary endpoints of sensitivity and specificity with an 80% performance goal were met. Secondary endpoints were BriefCase time-to-notification compared to the predicate device, Positive Predictive Value (PPV), Negative	**Result:** No statistically significant interaction between performance and age, gender, or location.	**Interpretation:** Race distribution for sample was unavailable. Device performance did not meaningfully interact with location, gender or age.	**Result:** Mean Time to Notification (sec) Predicate K203505 = 267.6, 95% CI = 246.0-289.5. Mean Time to Notification (sec) Briefcase = 33.5, 95% CI = 30.9-36.1	**Interpretation:** The time-to-notification results obtained for the subject BriefCase device showed improvement with regard to time savings to the standard of care review.

	Predictive Value (NPV), Positive Likelihood Ratio (PLR), and Negative Likelihood Ratio (NLR).				
Test Type: Aidoc conducted a retrospective, blinded, multicenter study.		**Test Type:** Retrospective, blinded, multicenter study		**Test Type:** Retrospective, blinded, multicenter study	
Testing Data Description: The mean age of patients whose scans were reviewed in the study was 65.1 years, with standard deviation of 18.8 years. Gender distribution was 49.8% male, and 50.2% female. Race distribution within the study data patient population was unavailable. None of the potential covariates demonstrated statistical significance, thus, device performance did not meaningfully interact with location, gender or age.		**Testing Data Description:** The mean age of patients whose scans were reviewed in the study was 65.1 years, with standard deviation of 18.8 years. Gender distribution was 49.8% male, and 50.2% female. Race distribution within the study data patient population was unavailable.		**Testing Data Description:** The BriefCase time-to-notification includes the time to get the DICOM exam, de-identify it, upload it to the cloud, analyze and send a notification on a positive suspect case back to the desktop application. The BriefCase time-to-notification was measured for all True Positive cases (i.e., identified as positive both by the reviewers as well as the BriefCase device).	
Validation Process and Justification: Link to Methods Description		**Validation Process and Justification:** Link to Methods Description		**Validation Process and Justification:** Link to Methods Description AiDoc develops software within a Design Control process that is aligned with FDA 21 CFR Part 820 and ISO 62304 Medical device software – Software life cycle processes.	

Resources

- **Funding source of the technical implementation:** Aidoc's model creation is self-funded.
- **3rd Party Information, If Applicable:** N/A
- **Evaluation References:** Please reference our clinical compendium.
- **Clinical Trial, If Available:** N/A
- **Peer Reviewed Publication(s):** Aidoc's clinical compendium with 100+ Peer-reviewed publications or abstract/conference presentations are available at the following link.
- **Reimbursement status, if applicable:** N/A
- **Patient consent or disclosure required or suggested:** N/A
- **Stakeholders consulted during design of intervention:** Physicians, health system administrators, and patient groups were engaged in the development of the model

**Note: this is not an endorsement of the BriefCase-Triage for Intracranial Hemorrhage (ICH) solution or of Aidoc but is an illustrative example of a populated model card. This card is still under review for alignment with model card instructions and reporting standards and will be updated in the final version of this document. More examples including Generative AI solutions and other use-cases to follow.*

5. Appendix

5.1 References

1. AI Nutrition Facts by Twilio. nutrition-facts.ai. https://nutrition-facts.ai/
2. MD GL. It's Time For "Nutrition Labels" In Artificial Intelligence. *Forbes*. https://www.forbes.com/sites/greglicholai/2023/11/21/its-time-for-nutrition-labels-in-artificial-intelligence/. Published November 21, 2023.
3. The Data Nutrition Project. datanutrition.org. https://datanutrition.org/
4. Sendak MP, Gao M, Brajer N, Balu S. Presenting machine learning model information to clinical end users with model facts labels. *npj Digital Medicine*. 2020;3(1). doi:https://doi.org/10.1038/s41746-020-0253-3
5. Mitchell M, Wu S, Zaldivar A, et al. *Model Cards for Model Reporting*.; 2019. https://arxiv.org/pdf/1810.03993
6. CHAI Draft Assurance Standards Guide
7. CHAI Draft Assurance Standards Checklists: Planning Checkpoint, Checkpoint 1, Checkpoint 2, Checkpoint 3
8. HTI-1 Final Rule
9. FDA 510K Pre-Market Submission

5.2 Resources

- HTI-1 Final Rule
    - HTI-1 Final Rule Overview 2024-01-18 (slides)
    - Link to Regulation
    - ONC HTI-1 Decision-Support Interventions Fact Sheet

ARC 1.4.3 Population Impact Evaluation Tool (page 12) (To assist with population impact assessment)

Assessment criteria/guidance for clinical/device risk level:

See Draft Medical Device Software: Considerations for Device and Risk Characterization (IMDRF, 2024)

See levels described in detail in "Software as a Medical Device": Possible Framework for Risk Categorization and Corresponding Considerations" by IMDRF Software as a Medical Device (SaMD) Working Group (2014).

5.3 Terms Defined

Definitions provided here are taken directly from the CHAI Assurance Standards Guide Draft, which includes source references. They will be further modified to address their scope within the specific model card requirements.

AI solution: A shorthand for the AI model, algorithm , or system and required technical infrastructure (hardware, software, data warehousing, etc.).

AI model: A conceptual or mathematical representation of phenomena captured as a system of events, features, or processes. In computationally-based models used in AI, phenomena are often abstracted for mathematical representation, which means that characteristics that cannot be represented mathematically may not be captured in the model. Often used synonymously with "algorithm," though it may be conceptually distinct, prior to the transformation of inputs to outputs.

Usefulness, Usability, and Efficacy. To be useful, an AI solution must provide a specific benefit to patients and/or healthcare delivery, and it must prove to be not only valid and reliable but usable and effective. The benefit of an AI solution can be measured based on its effectiveness in achieving intended outcomes, as well as its impact on overall health resulting from both intended and potentially unintended uses, as appropriate. An assessment of benefit should consider the balance between positive effects and adverse effects or risks. The usefulness of an AI solution also depends on the cost of its deployment and the capacity of personnel to act as a result of its output or guidance. Relatedly, an effective AI solution can be shown to achieve the intended improvement on health compared to existing standards of care, or it can improve existing workflows and processes; for example, an AI solution intended to increase the efficiency of a workflow can be associated with reduced costs or shorter times to complete tasks. The robustness and reliability of an AI system can be demonstrated by its ability to maintain its level of performance under a variety of circumstances. The solution's testability is more encompassing, demonstrating the extent to which its performance can be verified as meeting best practices for trustworthy AI including safety, equity, transparency, privacy, and security. The usability of an AI solution connotes the quality of the user's experience, including effectiveness, efficiency, and satisfaction with the technology. In this context, an engaged human-centered development and deployment process entails understanding, expressing and prioritizing the needs, preferences and goals of end users and other stakeholders, as well as considering related implications throughout the AI lifecycle. Similarly, an accessible development and deployment process ensures that stakeholder access and engagement is a core feature of each stage of the AI lifecycle and governance. Additionally, an adaptive accountability framework ensures continuous learning and improvement, providing ongoing information on the results of the AI solution.

Fairness and Equity. To be considered fair and equitable, AI solutions may demonstrate (1) parity, meaning that common measures of algorithmic performance are equal across protected subgroups; (2) calibration, meaning that outcomes are independent of protected characteristics (or class) – such as race, gender, or their proxies; and (3) anti-classification, meaning that protected characteristics are not explicitly used to make decisions. Fairness applies beyond diagnostic accuracy to equitable allocation of resources, access to care, and outcomes. Following from that, the ultimate indicator of fairness goes beyond these measures and may include measures of equity. Equity is the absence of systematic disparities in health outcomes between groups with different levels of underlying social advantage or disadvantage – that is, wealth, power, prestige, or other social determinants of health.

Safety and Reliability: A safe AI solution does not endanger human life, health, property, or the environment. In healthcare, this translates to the avoidance, prevention, and amelioration of AI-related adverse outcomes affecting patients, clinicians, and health systems. Harms, or a diminishment of safety may occur due to misuse or model deterioration because of factors like drifts and shifts. An AI system therefore proves reliable to the extent that it can perform as required without failure, incorporating backup plans that ensure continuity, resilience, communication, accountability, and responsive action in the event of any issues.

Transparency, Intelligibility, and Accountability. In this context, transparency is the extent to which information about an AI solution (e.g., capabilities, limitations, and purpose) and its output is available to all relevant stakeholders. Intelligibility is the extent to which the AI system can be understood by relevant stakeholders, often through a representation of the mechanisms underlying an algorithm's operation and through the meaning of its output in the context of its designed functional purposes. The principle of intelligibility addresses the question of whether humans can understand and make sense of the AI solution, encompassing the principles of explainability and interpretability. Explainability is the ability to provide insight into why and how the AI model is generating outputs – the observation of the inner mechanics of the AI/ML method, along with the factors and features that influence the system's decision-making process. Explainability addresses the question of why an AI system made a specific decision. Interpretability, by contrast, is the ability to understand the cause and effect of the AI model's output in human terms, serving as a risk mitigation strategy. It involves making the model structure, parameters, and relationship between inputs and outputs understandable. Observability, then, connotes the ability to observe inputs, outputs, impact, and consequences of model predictions. In applications designed to augment human decisions, a human-machine teaming model should explicitly specify the interface for interaction between the AI solution and the human use. This specification lays out the details needed for the user to operate the model safely, which finally supports the principle of accountability: the responsibility and liability for minimizing harm throughout all stages of the AI lifecycle.

Security and Privacy: The principle of security conveys the extent to which AI systems can maintain confidentiality, integrity, and availability through administrative, technical, and physical safeguards. Privacy is the extent to which AI systems can maintain predictability, manageability, and dissociability through administrative, technical, and physical safeguards that prevent problematic data actions for individuals (including at the group and societal level). In this context, risk is the composite measure of an event's probability of occurring and the magnitude or degree of consequences resulting from the corresponding event. Risk management is a term that signifies coordinated activities to direct and control an organization with regard to risk.

References for Principle Area Definitions and Descriptions:

[1] L. Adams et al., "Artificial Intelligence in Health, Health Care, and Biomedical Science: An AI Code of Conduct Principles and Commitments Discussion Draft," NAM Perspect., Apr. 2024, doi: 10.31478/202404a.

[2] "Blueprint for an AI Bill of Rights | OSTP," The White House. Accessed: Apr. 29, 2024. [Online]. Available: https://www.whitehouse.gov/ostp/ai-bill-of-rights/

[3] National Institute of Standards and Technology, "The NIST Cybersecurity Framework (CSF) 2.0," National Institute of Standards and Technology, Gaithersburg, MD, NIST CSWP 29, Feb. 2024. doi: 10.6028/NIST.CSWP.29.

[4] National Institute of Standards and Technology, "NIST PRIVACY FRAMEWORK: A TOOL FOR IMPROVING PRIVACY THROUGH ENTERPRISE RISK MANAGEMENT, VERSION 1.0," National Institute of Standards and Technology, Gaithersburg, MD, NIST CSWP 01162020, Jan. 2020. doi: 10.6028/NIST.CSWP.01162020.

[5] E. Tabassi, "NIST Artificial Intelligence Risk Management Framework (AI RMF 1.0)," National Institute of Standards and Technology (U.S.), Gaithersburg, MD, NIST AI 100-1, Jan. 2023. doi: 10.6028/NIST.AI.100-1.

[6] “Healthcare Sector Cybersecurity Framework Implementation Guide 1.1”.

[7] W. Raynor, International Dictionary of Artificial Intelligence. Routledge, 2020.

[8] E. B. [D-T.-30 Rep. Johnson, “Text - H.R.6216 - 116th Congress (2019-2020): National Artificial Intelligence Initiative Act of 2020.” Accessed: May 03, 2024. [Online]. Available: https://www.congress.gov/bill/116th-congress/house-bill/6216/text

[9] Z. Obermeyer, B. Powers, C. Vogeli, and S. Mullainathan, “Dissecting racial bias in an algorithm used to manage the health of populations,” Science, vol. 366, no. 6464, pp. 447– 453, Oct. 2019, doi: 10.1126/science.aax2342.

[10] S. Dutchen, “The Importance of Nuance | Harvard Medicine Magazine.” Accessed: Apr. 24, 2024. [Online]. Available: https://magazine.hms.harvard.edu/articles/importancenuance

[11] Coalition for Health AI, “Blueprint for Trustworthy AI: Implementation Guidance and Assurance for Healthcare.” Accessed: Apr. 22, 2024. [Online]. Available: https://www.coalitionforhealthai.org/papers/blueprint-for-trustworthy-ai_V1.0.pdf

[12] B. Vasey et al., “Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI,” BMJ, vol. 377, p. e070904, May 2022, doi: 10.1136/bmj-2022-070904.

[13] A. Downing and E. Perakslis, “Health advertising on Facebook: Privacy and policy considerations,” Patterns, vol. 3, no. 9, Sep. 2022, doi: 10.1016/j.patter.2022.100561.

[14] S. G. Finlayson et al., “The Clinician and Dataset Shift in Artificial Intelligence,” N. Engl. J. Med., vol. 385, no. 3, pp. 283–286, Jul. 2021, doi: 10.1056/NEJMc2104626.

[15] N. K. Corrêa et al., “Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance,” Patterns, vol. 4, no. 10, Oct. 2023, doi: 10.1016/j.patter.2023.100857.

[16] T. L. C. Patient Experts and Community Leaders, “AI Rights for Patients.” Accessed: Apr. 22, 2024. [Online]. Available: https://lightcollective.org/wpcontent/uploads/2024/03/Collective-Digital-Rights-For-Patients_v1.0.pdf 69

[17] A. D. Bedoya et al., “A framework for the oversight and local deployment of safe and high-quality prediction models,” J. Am. Med. Inform. Assoc. JAMIA, vol. 29, no. 9, pp. 1631–1636, May 2022, doi: 10.1093/jamia/ocac078.

[18] “key decision points,” Health AI Partnership. Accessed: May 03, 2024. [Online]. Available: https://healthaipartnership.org/key-decisions-in-adopting-an-ai-solution

[19] A. Ferlitsch, “Making the machine: the machine learning lifecycle,” Google Cloud Blog. Accessed: May 03, 2024. [Online]. Available: https://cloud.google.com/blog/products/aimachine-learning/making-the-machine-the-machine-learning-lifecycle [20] S. Kearney and R. G. Kenny, “SAS in Healthcare: Duke-SAS Health Innovation Lab.” SAS Institute Inc.

[21] “Workbook | AI Toolkit.” Accessed: May 03, 2024. [Online]. Available: https://www.ailawenforcement.org/assess/workbook

[22] K. E. Morse, S. C. Bagley, and N. H. Shah, “Estimate the hidden deployment cost of predictive models to improve patient care,” Nat. Med., vol. 26, no. 1, pp. 18–19, Jan. 2020, doi: 10.1038/s41591-019-0651-8.

[23] K. E. Henry et al., “Human–machine teaming is key to AI adoption: clinicians’ experiences with a deployed machine learning system,” Npj Digit. Med., vol. 5, no. 1, pp. 1–6, Jul. 2022, doi: 10.1038/s41746-022-00597-7.

[24] “ISO/IEC TS 5723:2022(en), Trustworthiness — Vocabulary.” Accessed: Apr. 22, 2024. [Online]. Available: https://www.iso.org/obp/ui/#iso:std:iso-iec:ts:5723:ed-1:v1:en

[25] 14:00-17:00, “ISO 9241-11:2018,” ISO. Accessed: Apr. 22, 2024. [Online]. Available: https://www.iso.org/standard/63500.html

[26] S. Corbett-Davies, J. D. Gaebler, H. Nilforoshan, R. Shroff, and S. Goel, “The Measure and Mismeasure of Fairness,” 2018, doi: 10.48550/ARXIV.1808.00023.

[27] H. E. Wang et al., “A bias evaluation checklist for predictive models and its pilot application for 30-day hospital readmission models,” J. Am. Med. Inform. Assoc. JAMIA, vol. 29, no. 8, pp. 1323–1333, Jul. 2022, doi: 10.1093/jamia/ocac065.

[28] P. Braveman and S. Gruskin, “Defining equity in health,” J. Epidemiol. Community Health, vol. 57, no. 4, pp. 254–258, Apr. 2003, doi: 10.1136/jech.57.4.254.

[29] T. Li, D. Khashabi, T. Khot, A. Sabharwal, and V. Srikumar, “UNQOVERing Stereotyping Biases via Underspecified Questions,” in Findings of the Association for Computational Linguistics: EMNLP 2020, T. Cohn, Y. He, and Y. Liu, Eds., Online: Association for Computational Linguistics, Nov. 2020, pp. 3475–3489. doi: 10.18653/v1/2020.findings-emnlp.311.

[30] “Information bias,” Catalog of Bias. Accessed: Apr. 22, 2024. [Online]. Available: https://catalogofbias.org/biases/information-bias/

[31] T. Panch, H. Mattie, and R. Atun, “Artificial intelligence and algorithmic bias: implications for health systems,” J. Glob. Health, vol. 9, no. 2, p. 010318, Dec. 2019, doi: 10.7189/jogh.09.020318.

[32] R. R. Bond et al., “Automation bias in medicine: The influence of automated diagnoses on interpreter accuracy and uncertainty when reading electrocardiograms,” J. Electrocardiol., vol. 51, no. 6, pp. S6–S11, Nov. 2018, doi: 10.1016/j.jelectrocard.2018.08.007.

[33] To Err Is Human: Building a Safer Health System. Washington, D.C.: National Academies Press, 2000, p. 9728. doi: 10.17226/9728. 70

[34] J. B. Cooper, D. M. Gaba, B. Liang, D. Woods, and L. N. Blum, “The National Patient Safety Foundation agenda for research and development in patient safety,” MedGenMed Medscape Gen. Med., vol. 2, no. 3, p. E38, Jul. 2000.

[35] A. Subbaswamy and S. Saria, “From development to deployment: dataset shift, causality, and shift-stable models in health AI,” Biostat. Oxf. Engl., vol. 21, no. 2, pp. 345–352, Apr. 2020, doi: 10.1093/biostatistics/kxz041.

[36] “Health Data, Technology, and Interoperability: Certification Program Updates, Algorithm Transparency, and Information Sharing,” Federal Register. Accessed: Apr. 22, 2024. [Online]. Available: https://www.federalregister.gov/documents/2024/01/09/2023- 28857/health-data-technology-and-interoperability-certification-program-updatesalgorithm-transparency-and

[37] NIST, “The Language of Trustworthy AI: An In-Depth Glossary of Terms (March 22, 2023) - Google Drive.” Accessed: May 03, 2024. [Online]. Available: https://docs.google.com/spreadsheets/d/e/2PACX1vTRBYglcOtgaMrdF11aFxfEY3EmB31zslYI4q2_7ZZ8z_1lKm7OHtF0t4xIsckuogNZ3hRZAaDQuv_K/pubhtml