A logo with a blue circle with a eagle and a flag with a white text

Description automatically generated with medium confidence

Ambient Dictation Evaluation Criteria and Process: Participant Guide

|  |
| --- |
| LAST UPDATED: JANUARY 18, 2024 |

DEPARTMENT OF VETERANS AFFAIRS

**2024**

**Contents**

I. [Overview 2](#_Toc156482990)

II. [Expected Solution Details 3](#_Toc156482991)

[Option 1: 3](#_Toc156482992)

[Option 2: 3](#_Toc156482993)

III. [Phase Details and Evaluation Components 4](#_Toc156482994)

[Gate 0 Details: 4](#_Toc156482995)

[Gate 1 Details: 4](#_Toc156482996)

[Gate 2 Details 5](#_Toc156482997)

[Gate 3 Details 6](#_Toc156482998)

[Award Ceremony 7](#_Toc156482999)

IV. [Expected Data Output from Solutions 7](#_Toc156483000)

[Expected Data VA will Present to the Solution 7](#_Toc156483001)

# Overview

The overall structure of the Ambient Dictation evaluation will include three phases, which are also referred to as gates. Each gate (or phase) will be subjected to distinct evaluation criteria.

## Expected Solution Details

**Process for operationalizing products on Veterans Affairs (VA) equipment:** There are two options for deploying solutions to VA for this AI Tech Sprint:

### **Option 1:**

* Local client installed software can capture audio from government furnished equipment (GFE) hardware and encrypt data.
* Solutions that require additional services using external connections must have their connections approved and tested prior to testing; or must use the AI Tech Sprint storage provided to transfer files.
* Output data is returned to the local client interface.
  + This could leverage the AI Tech Sprint Application Programming Interface (API) or be handled entirely through their client.
  + Check the “Expected Data Output from Solutions” section for the data format and requirements.
* A contestant must provide their external internet protocol (IP) address(s) and port numbers required for the user to connect to the user interface and for the data connections to occur **within one week of being approved** to participate in the AI Tech Sprint.

### **Option 2:**

* For Federal Risk and Authorization Management Program (FedRAMP) or VA Enterprise Cloud (VAEC) solutions, the testing hardware (i.e., a laptop or phone GFE) must have access to the application and the solution must be able to use the client audiovisual devices or have any vendor required devices reviewed and installed prior to testing.
* The vendor will writeback files to the S3 bucket API. For FedRAMP or VAEC solutions, the testing hardware must have access to the cloud hosted application.
* The solution will output files, including progress notes in a JSON or HL7 file, to the AI Tech Sprint inbound external AWS S3 storage using credentials provided to the contestant.
  + Check the “Expected Data Output from Solutions” section for the data format and requirements.
* A contestant must provide their external IP address(s) and port numbers required for the user to connect to the user interface and for the data connections to occur **within one week of being approved** to participate in the AI Tech Sprint.

**Note:** VA Personally Identifiable Information (PII) regulations prohibit the use of common names in any documents or data distributed as anonymized. Therefore, names are represented by English words that are not associated with any common names. An example may be “Customer, One.”

# Phase Details and Evaluation Components

## **Gate 1 Details:**

Evaluation goal:

* Pass/Fail for basic functionality. If contestant passes, move onto Gate 2.

Deliverables from contestant:

* Solution provided that can run on or be accessed by VA GFE and generates expected output data.
* Responses to the IT Evaluation Criteria.

Deliverables from VA team:

* Report on performance across components for each contestant, determination of overall Pass/Fail.

Evaluation Components:

* Basic quality checks: These checks assess whether the solution can run in VA environment and generate expected output. Basic checks may also assess:
  + Whether the solution is compatible with minimum VA GFE requirements.
  + Satisfactory Behavior during shutdown, connectivity loss, or other temporary interruption, etc.
  + Software stability.
  + Whether the solution generates the correct data format in its output.
  + Response time, ensuring it is below the threshold (120 seconds).
* Acceptable IT performance/requirements: The contestant answers the IT questionnaire.
  + Responses are reviewed and a Pass/Fail status is determined.

Process:

* Contestant provides a solution to be installed and/or provides external connection requirements for VA GFE with instructions for use.
* VA conducts the Gate 1 evaluation:
  + Basic Quality checks:
    - VA attempts to operationalize the solution with the instructions provided by the contestant.
    - Contestants are given an opportunity to fix any installation issues by attending office hours or setting up an appointment. Please note, only one remediation appointment will be granted to each contestant.
* Acceptable IT performance/requirements: The contestant answers the IT questionnaire.
  + Responses are reviewed and a Pass/Fail status is determined.
  + VA generates a report on performance and issues a Pass/Fail determination.
    - A contestant must pass both checks.
  + VA notifies contestants of the results and if they will move onto Gate 2.

## **Gate 2 Details**

Evaluation goal:

* This round of evaluation generates a numeric score to rank contestants for selection. This gate identifies which contestants will move onto Gate 3.

Deliverables from contestant:

* A solution must have instructions provided that can be used on VA GFE. Please note, at this gate, there will no longer be an opportunity offered to troubleshoot installation, as in Gate 1.
* Responses to the Trustworthy AI Questionnaire.

Deliverables from VA team:

* Report on performance/evaluation criteria.
* Selection into Gate 3.

Evaluation Components:

* Trustworthy AI Scoring Manual
* Quantitative evaluation
  + Accurate detection of medical entities and related values compared to manually reviewed and validated controls.
  + Accurate identification of speakers.
* Clinician Review
  + Clinician review of summary note and linked table of contents for Coherence, Factual Consistency, Comprehensiveness, and Harmfulness.

Process:

* The contestant provides a solution to be used on VA GFE with instructions for use/installation (or directs VA to use the solution provided in Gate 1).
* VA deploys/uses solution and runs data through the software, obtaining file outputs.
  + - The order of evaluation will be randomized by day.
  + VA performs quantitative evaluation on the data outputs in a rolling fashion as outputs become available.
  + VA performs clinician review evaluation on the generated notes as outputs become available.
  + VA combines all evaluation components into a single composite score/report.
  + Based on the scores, VA selects contestants who move onto Gate 3.

## **Gate 3 Details**

Evaluation goal:

* The goal here is to generate a numeric score to rank contestants for placement.

Deliverables from contestant:

* A solution provided that can used on VA GFE.
* A recorded presentation or demonstration of software (maximum of 10 minutes in duration) and supporting materials.
  + This presentation or demonstration will use data from Gate 2.

Deliverables from VA team

* VA will provide a report on performance/evaluation criteria.

Evaluation Components:

* Quantitative evaluation:
  + Accurate detection of words will be assessed to compare a generated transcript to a known source of truth transcript and calculate the word information preserved (WIP) metric.
  + An accurate identification of speakers will also be assessed.
* Clinician Review
  + A clinician reviews summary notes and the linked table of contents for coherence, factual consistency, comprehensiveness, and harmfulness.
* Qualitative evaluation of presentation
  + Purposeful: Clarity and specificity of benefits are tailored to clinicians/end users.
  + Effective and Safe: Presentations show the solution has demonstrated proper handling of unexpected scenarios; maintaining safety and effectiveness.
  + Secure and Private: Techniques for data protection and confidentiality are evaluated.
  + Integration: Current workflow optimization, data and system integration capabilities, and future planned capabilities are assessed.

Process:

* Contestant provides a solution to be used on VA GFE with instructions for use or installation.
  + This does not necessarily have to be an entirely new product but can be an updated product.
* VA deploys or uses the solution and runs data through the software, obtaining its file outputs.
* The contestant provides a recorded presentation demonstrating their product.
* VA performs quantitative evaluation on the data outputs.
* VA performs clinician review evaluation on the generated notes.
* VA performs qualitative evaluation on the recorded presentation.
* VA combines all evaluation components into a single composite score or report.
* VA selects final rankings based on composite score.

## **Award Ceremony**

* Teams from Gate 3 will present a demonstration of their solutions.
* 1st, 2nd, and 3rd place winners will be announced.
* The ceremony will be held in May. The location of the ceremony is pending.

# **Expected Data Output from Solutions**

**The solution should return:**

1. A word-for-word transcription of the encounter. This response should be in plain text format returned in a VA standard compliant JSON or HL7 file.
2. For each word in the transcript, a vector of length *n* should designate the speaker likelihood score in the range [0, 1]. The likelihood score indicates that the word is spoken by 1 of up to 10 speakers. This should be returned as a comma-separated values file (.csv), where the first column is the index in the transcript, the second column is the word, and each subsequent column is the likelihood that it came from the *n*th speaker, counting from the first to show up in the recording.
3. A summary of the encounter in the form of a clinical note. This response should be in a plain text format (i.e., JSON or HL7).

## **Expected Data VA will Present to the Solution**

VA GFE will be exposed to audio of the interaction (recorded audio played in room with VA GFE). No specific data files will otherwise be provided.