

Date of Approval: 12/16/2024
Questionnaire Number: 1705

Basic Information/Executive Summary

What is the name of your project (system, database, pilot, product, survey, social media site, etc.)?

Volunteer Income Tax Assistance (VITA) Generative Artificial Intelligence
(GenAI) Chatbot Proof of Concept (PoC)

Business Unit

Office of Online Services

Preparer

For Official Use Only

Subject Matter Expert

For Official Use Only

Program Manager

For Official Use Only

Designated Executive Representative

For Official Use Only

Executive Sponsor

For Official Use Only

Executive Summary: Provide a clear and concise description of your project and how it will allow the IRS to achieve its mission.

This Generative Artificial Intelligence (GenAI) chatbot Proof of Concept (PoC) will be tested by a select number of Volunteer Income Tax Assistance (VITA) and Tax Counseling for the Elderly (TCE) program volunteers during the 2025 tax filing season while preparing taxes without having to manually search for answers to tax questions across multiple tax publications and instructions. This PoC will allow the Internal Revenue Service (IRS) to achieve its mission by evaluating if using a GenAI chatbot can feasibly improve efficiency and user experience.

Personally Identifiable Information (PII)

Will this project use, collect, receive, display, store, maintain, or disseminate any type of Sensitive but Unclassified (SBU), Personally Identifiable Information (PII), or Federal Tax Information (FTI)?

Yes

Please explain in detail how this project uses sensitive data from inception to destruction (data lifecycle).

The project will receive names and email addresses of 20-30 Volunteer Income Tax Assistance (VITA) volunteers who Stakeholder, Partnership, Education and Communication (SPEC) determines to use. The information will be sent to a designated Online Services (OLS) representative and then forwarded to the Information Technology (IT) developer via IRS email. The names and emails will then be placed on a whitelist to be used for Okta (a commercial identity and access management software service used by IRS) authentication and log in to the chatbot. Internet Protocol (IP) addresses for the volunteer users will not be required; only the names and emails are required for the whitelist activity.

Please select all types of Sensitive but Unclassified data (SBU)/Personally Identifiable Information (PII)/Federal Tax Information (FTI) that this project uses.

Email Address

Internet Protocol Address (IP Address)

Name

Cite the authority for collecting SBU/PII/FTI (including SSN if relevant).

PII for federal tax administration - generally IRC Sections 6001 6011 or 6012

Product Information (Questions)

1 Is this PCLIA a result of the Inflation Reduction Act (IRA)?

No

2 Describe in detail, the Robotic Process Automation (RPA) process; be sure to identify the project title and business unit owner; state what IRS Strategy or initiative it supports; identify the system or process it supports and if PII will be required for the RPA to run; identify activities and workflow controls with the type and capabilities that will be incorporated; lastly indicate how the service benefits from the use of this RPA. (Process, Library, Test Automation, Template.)

The objective of this AI project is to investigate the feasibility and applicability of using Large Learning Model, Retrieval-Augmented Generation AI technology to consolidate multiple sources of designated IRS publicly facing information into a simple question and answer tool. IT accomplishes this purpose by using a select

data set to retrieve answers for questions asked by a small subset of Volunteer Income Tax Assistance personnel during tax preparation. Currently, volunteers must know where to go to find the information for which they are looking. This includes publications, forms, certification training guides, etc. Once they determine which document to use, then they must search for answer within that document. This model uses the information provided to search for an answer and shows the user what document the information came from as well as the page. This saves the user a lot of time and effort. This model has been tested by two separate groups with a 99.5% accuracy rate for answers to questions asked. Humans were very involved in the development and testing of this chatbot to ensure answers were correct, accurate and relevant to the questions asked. The chatbot was built with a survey-type tool whereby each answer is scored by the volunteer for the previously mentioned criteria. The questions, answers, scores, and comments (feedback) from the volunteers is recorded within the chatbot for review and analysis to make further adjustments or fine tuning of the chatbot. The benefits of this chatbot and this specific technology will decrease the time and effort to prepare taxes, and the volunteers will be able to provide service for more taxpayers. During the Subject Matter Expert (SME) testing, it was discovered some volunteers are using free GenAI tools to find information. This information is coming from all sources in the web-world, or wherever that source uses for information gathering. The free Google Gemini tool is one of the examples a tester provided. If this tool is using the information the IRS provides as the only data source it uses to reference and provide an answer, the volunteer will know it is the correct answer, instead of them using a tool of their choice and assuming the answer is correct. This is a proof of concept and will expire at the end of the upcoming tax season (April 2025). The population impacted is VITA volunteers who have already been Certified under the VITA program and who Taxpayer Services>Stakeholder Partnerships, Education and Communication (SPEC) has vetted and asked to test while they are preparing taxes for those eligible under the VITA/Tax Counseling for the Elderly (TCE) Programs. The number will not exceed 30 testers.

3 Is this a new Robotic Process Automation (RPA) project?

No

4 Identify the IRS IT systems, applications, projects, and/or databases this RPA is applied to; include the associated system name.

Accenture Federal Services (AFS) and the IEP Innovation team are using the Google Cloud Platform for this chatbot.

5 Identify why the use of SBU/PII/FTI is required; include any type of Sensitive But Unclassified (SBU), Personally Identifiable Information (PII), or Federal Tax Information (FTI) that this project will create, collect, receive, use, process, maintain, access, inspect, display, store, disclose, disseminate, or dispose of.

The information is required for whitelist and OKTA log in for each tester/user of the chatbot.

6 Is your RPA Attended/Unattended?

Unattended

7.1 Explain the process being replaced/automated.

The manual look-up of documentation pertinent to a tax preparation question from numerous IRS manuals and publications.

8 Indicate what level of complexity the RPA is classified as and if you were required to register with OneSDLC or not or indicate if Information Technology's (ITs) Technical Insertion process was used for approval of this RPA.

Medium level complexity. This is not required to be registered with OneSDLC since this has a definite end date and will not remain in production after the 2025 tax filing season. Research, Applied Analytics & Statistics (RAAS) documents completed.

9 Will connections or interdependencies be established for this RPA?

Yes

9.1 Will the connection be encrypted?

Yes

9.2 Will authentication/credentials be required?

Yes

9.3 Please provide details for the connection/interdependency. Indicate if this occurs on the backend versus through the system/user interface.

The encryption will take place using OKTA when the user logs in to authenticate. The users will be identified on the backend as well.

10 Indicate who has or will have permission to access the data and how users are authenticated.

User will have Read only access. Users will be authenticated using OKTA.

11 Indicate if BEARS entitlements are required for access and if PUMAS (Privileged User Management Access System) control management is applied for granting access to the system(s)? If BEARS/PUMAS are not applied, indicate what access controls are in place.

A whitelist of select volunteers will be used to determine who will be able to log in using a URL provided and OKTA to authenticate.

12 Identify the maintenance tasks or updates performed; state whether the maintenance tasks are inherited from the host (UiPath Platform), or you are using customized maintenance activities.

There are no maintenance tasks related to this PoC. If SPEC determines a document needs to be updated, added, or deleted as a reference for the chatbot, then a developer will have to do this from the backend.

13 Indicate if this product or system shares data outside of the United States or its territories.

No

14 Indicate if this system or Robotic Process Automation (RPA) is trained through the use algorithms; indicate if the algorithm used contains data with a sensitivity classification. (Sensitive but unclassified data might include algorithms, methods, system data, or PII/FTI that could be used to re-identify a person.)

No

Interfaces

Interface Type

IRS Systems, file, or database

Agency Name

IRS email

Incoming/Outgoing

Incoming (Receiving)

Transfer Method

Secure email/Zixmail

Systems of Records Notices (SORNs)

SORN Number & Name

IRS 10.555- Volunteer Records

Describe the IRS use and relevance of this SORN.

Provide information to volunteers who coordinate activities and staffing at taxpayer assistance sites.

Records Retention

What is the Record Schedule System?

General Record Schedule (GRS)

What is the retention series title?

Public customer service operations records GRS 6.5

What is the GRS/RCS Item Number?

GRS 6.5 Item 10

What type of Records is this for?

Electronic

Please provide a brief description of the chosen GRS or RCS item.

Records from operating a customer call center or service center providing services to the public. Services may address a wide variety of topics such as understanding agency mission-specific functions or how to resolve technical difficulties with external-facing systems or programs.

What is the disposition schedule?

Destroy 1 year after resolved, or when no longer needed for business use, whichever is appropriate.

Data Locations

What type of site is this?

Environment

What is the name of the Environment?

Google Cloud Platform

What is the sensitivity of the Environment?

Personally Identifiable Information (PII) including Linkable Data

What is the URL of the item, if applicable?

We do not have the URL to be used by volunteers during the 2025 tax filing season.

Please provide a brief description of the Environment.

VITA chatbot - LLM, RAG Generative AI

What are the incoming connections to this Environment?

Accenture Federal Services (AFS) Google Cloud Environment is where the chatbot resides.

What type of site is this?

Environment

What is the name of the Environment?

Google Cloud Platform

What is the sensitivity of the Environment?

Personally Identifiable Information (PII) including Linkable Data

What is the URL of the item, if applicable?

Do not have the final URL yet; will have by mid-December.

Please provide a brief description of the Environment.

Generative Artificial Intelligence (GenAI) chatbot that uses Retrieval Augmented Generation (RAG); built using Google Gemini within the Google Cloud Project

What are the incoming connections to this Environment?

PDF documents from IRS.gov.

What are the outgoing connections from this Environment?

The data is stored in Integrated Enterprise Portals Google Cloud Project (IEP-GCP), but there are copies of the publicly available documents from IRS.gov.