

Date of Approval: **August 19, 2019**

PIA ID Number: **4313**

SYSTEM DESCRIPTION

Enter the full name and acronym for the system, project, application and/or database.

Chatbot Prototype Application, N/A

Is this a new system?

Yes

What governance board or Executive Steering Committee (ESC) does this system report to? Full name and acronym.

This system reports to the Chatbot Integrated Project Team (Chatbot IPT), which is comprised of resources from IRS OLS (Office of Online Services), IRS IT Enterprise Services, IRS PGLD (Privacy, Governmental Liaison and Disclosure), IRS Cybersecurity, and TE/GE (Tax Exempt and Government Entities) Division.

Current ELC (Enterprise Life Cycle) Milestones:

Vision & Strategy/Milestone 0

Is this a Federal Information Security Management Act (FISMA) reportable system?

No

General Business Purpose

What is the general business purpose of this system? Provide a clear, concise description of the system, application or database, the reason for the system, and the benefits to the IRS to use the information, and how the information will be used.

The chatbot prototype application is an HTML (Hypertext Markup Language)-based system that takes text input from the end-user and provides back an intelligent text response. The application supports the Online Services (OLS) Chatbot Research Project, which will aid the IRS in learning about the effectiveness of incorporating chatbots into their customer service model. This research will allow IRS to build a practical understanding of how taxpayers interact with chatbots and enable OLS to observe how users interface with a new content discovery tool. The information will be used to inform requirements for a potential procurement in the future. The chatbot prototype application is an experimental prototype and will not be used in production long-term.

The chatbot prototype application will operate using regular expressions, a form of natural language processing (NLP). We may expand our chatbot in the future to include machine learning, but all machine learning would be supervised. Our chatbot will leverage two open-source models, listed below, but these will not be updated using user data.

Spell Correction - corrects generic spelling errors included within user questions.

Lemmatization - linguistic model to group together inflected forms of a word so that they can be analyzed as a single item (identified by the word's lemma, or dictionary form).

The scope of the prototype application does not include unsupervised machine learning, deep learning, or a neural network. There are currently no plans to include additional algorithms beyond the aforementioned items. All content provided to users by the chatbot will be non-personalized content that is publicly available on IRS.gov. No PII will be created through the regular expression NLP (Natural Language Processing). The chatbot also won't store IP (Internet Protocol) addresses, but they may be stored in audit/security logs that are separate from the chatbot logs.

PII DETAILS

Does the system use, collect, receive, display, store, maintain, or disseminate IR Code 6103 taxpayer information; or any other type of Sensitive but Unclassified (SBU) information or PII such as information about IRS employees or outside stakeholders?

Yes

Does the system use, collect, receive, display, store, maintain, or disseminate Social Security Numbers (SSN's) or tax identification numbers (i.e. last 4 digits, etc.)?

Yes

What types of tax identification numbers (TIN) apply to this system?

Social Security Number (SSN)

List the approved Treasury uses of the SSN:

Another compelling reason for collecting the SSN

Explain why the authorized use(s) above support the new or continued use of SSNs (or tax identification numbers)

The chatbot prototype application will not solicit or disseminate any PII or SBU. The application will store user entered information for later research analysis. There exists the potential for end-users to enter unsolicited/unrequested PII. Note that end-user entered information is not available to other users of the application.

Describe the planned mitigation strategy and forecasted implementation date to mitigate or eliminate the use of SSN's (or tax identification numbers).

To minimize the risk of unsolicited PII, an information consent notice will be displayed and must be accepted in order for the user to access the chatbot prototype application. This message will state that the chatbot prototype application is part of a research study, that participation is voluntary, that chatbot data will be collected, that no PII should be provided, and that the scope is limited to the selected content. Additionally, the development team will create scripts to identify and redact all chatbot data resembling common PII and SBU formats (For example, in the case of SSNs, multiple variations of 9-digit numbers will be redacted: #####, ###-##-####, etc.). The forecasted implementation date to mitigate the use of unsolicited PII is by November 1, 2019, prior to the chatbot being made available.

Does this system use, collect, receive, display, store, maintain or disseminate other (non-SSN) PII (i.e. names, addresses, etc.)?

Yes

Specify the PII Elements:

Internet Protocol Address (IP Address)

Does this system use, collect, receive, display, store, maintain, or disseminate SBU information that is not PII?

No

Are there other types of SBU/PII used in the system?

No

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

SSN for tax returns and return information is Internal Revenue Code Section 6109

Has the authority been verified with the system owner?

Yes

BUSINESS NEEDS AND ACCURACY

Explain the detailed business needs and uses for the SBU/ PII, and how the SBU / PII is limited only to that which is relevant and necessary to meet the mission requirements of the system. If SSNs (or tax identification numbers) are used, explicitly explain why use of SSNs meets this criteria. Be specific.

There is no business need or use for SBU/PII. The chatbot prototype application will not solicit or disseminate any PII or SBU. The application will store user entered information for later research analysis. There exists the potential for end-users to enter unsolicited/unrequested PII. Note that end-user entered information is not available to other users of the application. To minimize the risk of unsolicited PII:

An information consent notice will be displayed and must be accepted in order for the user to access the chatbot prototype application. This message will state that the chatbot prototype application is part of a research study, that participation is voluntary, that chatbot data will be collected, that no PII should be provided, and that the scope is limited to the selected content. The development team will create scripts to identify and redact all chatbot data resembling common PII and SBU formats (For example, in the case of SSNs, multiple variations of 9-digit numbers will be redacted: #####, ##-##-####, etc.).

How is the SBU/PII verified for accuracy, timeliness and completion?

There is no business need or use for SBU/PII. The chatbot prototype application will not solicit or disseminate any PII or SBU. The application will store user entered information for later research analysis. There exists the potential for end-users to enter unsolicited/unrequested PII.

PRIVACY ACT AND SYSTEM OF RECORDS

The Privacy Act requires Federal agencies that maintain a system of records to publish systems of records notices (SORNs) in the Federal Register for records from which information is retrieved by any personal identifier for an individual who is a US citizen, or an alien lawfully admitted for permanent residence. The Privacy Act also provides for criminal penalties for intentional noncompliance.

Does your application or this PCLIA system pertain to a group of any record from which information is retrieved by any personal identifier for an individual who is a US citizen, or an alien lawfully admitted for permanent residence? An identifier may be a symbol, voiceprint, SEID, or other personal identifier that is used to retrieve information.

Yes

Identify the Privacy Act SORN(s) that cover these records.

IRS 00.001 Correspondence Files and Correspondence Control Files

RESPONSIBLE PARTIES

Identify the individuals for the following system roles:

For Official Use Only

INCOMING PII INTERFACES

Does the system receive SBU/PII from other systems or agencies?

No

DISSEMINATION OF PII

Does this system disseminate SBU/PII?

No

PRIVACY SENSITIVE TECHNOLOGY

Does this system use social media channels?

No

Does this system use privacy-sensitive technologies such as mobile, global position system (GPS), biometrics, RFID, etc.?

No

Does the system use cloud computing?

Yes

Is the cloud service provider (CSP) Federal Risk and Authorization Management Program (FedRAMP) certified?

Yes

Date Certified {MM/DD/YYYY}

11/13/2017 12:00:00 AM

Please identify the ownership of the CSP data.

IRS

Does the CSP allow auditing?

Yes

Who audits the CSP Data?

3rd Party

What is the background check level required for CSP?

Moderate

Is there a breach/incident plan on file?

Yes

Does this system/application interact with the public?

Yes

Was an electronic risk assessment (e-RA) conducted on the system/application?

Not Applicable

Explain why not required.

The system does not require authentication from non-organizational users.

INDIVIDUAL NOTICE AND CONSENT

Was/is notice provided to the individual prior to collection of information?

Yes

How is notice provided? Was the individual notified about the authority to collect the information, whether disclosure is mandatory or voluntary, the purpose for which the information will be used, with whom the information may be shared, and the effects on the individual, if any, if they decide not to provide all or any of the requested information?

A consent notice based on IRM 10.5.1 Privacy and Information Protection Privacy Policy will be provided and must be accepted before participants are able to use the chatbot. Individuals are notified about the authority to collect information, that participation is voluntary, and that PII and SBU should not be provided under any circumstances. If they choose to not accept the consent notice, they will be informed that they are unable to participate in the OLS Chatbot Research Project and returned to IRS.gov.

Do individuals have the opportunity to decline from providing information and/or from consenting to particular uses of the information?

Yes

Describe the mechanism by which individuals indicate their consent choice(s):

Individuals must click the 'Accept' button provided within the consent notice to indicate their agreement. Individuals will not be able to use chatbot unless they have given their consent.

How does the system or business process ensure 'due process' regarding information access, correction and redress?

N/A; This system is not part of a tax administration process subject to due process requirements.

INFORMATION PROTECTION

Identify the owner and operator of the system (could be IRS owned and Operated; IRS owned, contractor operated; contractor owned and operated)

Contractor Owned and Operated

The following people have access to the system with the specified rights:

IRS Employees

Users: Read Only

Managers: Read Only

System Administrators: Read Only

Developers: Read Only

IRS Contractor Employees

Contractor Users: Read Only

Contractor Managers: Read Only

Contractor System Administrators: Administrator

Contractor Developers: Read Only

How is access to SBU/PII determined and by whom?

The chatbot prototype application will not solicit or disseminate any PII or SBU. All contractor employees accessing the chatbot prototype application will have an active Moderate Minimum Background Investigation (MBI) suitability.

RECORDS SCHEDULE

Are these records covered under a General Records Schedule (GRS, IRS Document 12829), or has the National Archives and Records Administration (NARA) approved a Records Control Schedule (RCS, IRS Document 12990) for the retention and destruction of official agency records stored in this system?

Yes

How long are the records required to be held under the corresponding GRS or RCS, and how are they disposed of? In your response, please provide the GRS or RCS chapter number, the specific item number, and records series title.

Records collected by the chatbot prototype application will be managed under GRS 6.5, Item 010 Public Service Operations Records. Temporary. Destroy 1 year after resolved, or when no longer needed for business use, whichever is appropriate. Any additional records developed from the OLS Chatbot Research Project will be scheduled as coordinate with the IRS Records and Information Management (RIM) Program and IRS Records Officer.

SA&A OR ASCA

Has the system been through SA&A (Security Assessment and Authorization) or ASCA (Annual Security Control Assessment)?

No

Is the system secured in accordance with all applicable federal, treasury, and IRS security policy, procedures, and requirements?

Do not know

Please explain the process that was followed to ensure the system safeguards SBU/PII from unauthorized or inappropriate access. Include a description of the formalized documentation, process, and review that was used to analyze the system's security controls.

The chatbot prototype application will not solicit or disseminate any PII or SBU. The chatbot Prototype Application is not a FISMA system. The MITRE team will review the project with an IRS Cybersecurity representative to confirm.

Describe the system's audit trail.

Chatbot logs saved for 90 days in an ElasticSearch database.

PRIVACY TESTING

Does the system require a System Test Plan?

No

Please explain why:

The System Test Plan is not a required artifact because the Chatbot Prototype Application is not related to any business changes.

SBU DATA USE

Does this system use, or plan to use SBU Data in Testing?

No

NUMBER AND CATEGORY OF PII RECORDS

Identify the number of individual records in the system for each category:

IRS Employees: Not Applicable

Contractors: Not Applicable

Members of the Public: Under 100,000

Other: No

CIVIL LIBERTIES

Does the system maintain any information describing how any individual exercises their rights guaranteed by the First Amendment?

No

Is the system information used to conduct 'data-mining' as defined in the Implementing Recommendations of the 9/11 Commission Act of 2007, Public Law 110-53, Section 804?

No

Will this system have the capability to identify, locate, and monitor individuals or groups of people?

No

Does computer matching occur?

No

ACCOUNTING OF DISCLOSURES

Does the system include or require disclosure of tax or employee information to anyone other than IRS employees in the performance of their duties, or to the person to whom the information pertains or to a 3rd party pursuant to a Power of Attorney, tax or Privacy Act consent?

No