

Date of Approval: 10/18/2024
Questionnaire Number: 1046

Basic Information/Executive Summary

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

AI Contract Document Toolbox

Acronym:

AICDT

Business Unit

Office of Chief Procurement Officer

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.

Generative artificial intelligence (AI) tool is designed to assist with procurement document tasks, streamlining the process and improving overall efficiency. This innovative tool utilizes machine learning algorithms to automate the drafting and review of government contracting documents, reducing the time and effort required for these tasks. The benefits of using our generative AI tool for procurement document tasks include: 1. Time Saving: Automating routine tasks frees up personnel to focus on higher-level activities, such as strategic planning and decision-making. 2. Improved Procurement Document Quality: The AI tool ensures consistency and accuracy in procurement documents, reducing errors and improving overall quality. 3. Increased Efficiency: Automated drafting and review processes enable faster turnaround times, enabling procurement teams to respond quickly to changing requirements. What information is collected: Contract file documents produced by IRS personnel and vendor business

submitters are processed by the system. These procurement documents including work statements, market research, vendor proposals, evaluation reports assessing supplier capabilities, and pre-award / post-award procurement documents generally. intended use of the agency of the information 2. Data Sharing: Collected data will only be shared with authorized employees and contractors who have undergone approved background investigations and cybersecurity training. 3. Voluntary Participation: Individuals provide procurement data voluntarily by choosing to participate in the federal acquisition process. How Information is Secured: System is designed in accordance with privacy principles, data protection, and to comply relevant laws and regulations. The system designed to protect sensitive information from unauthorized access, use, or disclosure. System developers followed National Institutes of Standards and Technology (NIST) guidance to included proper security controls for federal IT systems. By leveraging our generative AI tool for procurement document tasks, we aim to improve the efficiency, effectiveness, and transparency of the federal acquisition process while maintaining the highest standards of data protection and security.

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).

System will process documents containing procurement sensitive and personally identifiable information (e.g. vendor phone numbers and emails). Documents are uploaded by IRS employees and staff-like access contractors to the system. Submission of information by external vendors is voluntary. However, failure to provide requested information may affect consideration for current or future government contract opportunities. The information provided will be used to evaluate vendor qualifications and capabilities and may be shared with authorized government personnel and contractors as necessary to support the procurement process. Data can be destroyed when no longer needed (e.g. the timeframes specified after contract completion or final payment per National Archives and Records Administration policies).

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

Address

Email Address

Procurement Sensitive Data

Telephone Numbers

Cite the authority for collecting SBU/PII/FTI (including SSN if relevant).
PII for personnel administration - 5 USC

Product Information (Questions)

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

No

1.3 What type of project is this (system, project, application, database, pilot/proof of concept, power platform/visualization tool)?

Initial proof of concept and experimentation with different AI solutions. Which may later grow into a pilot project.

1.35 Is there a data dictionary for this system?

IRS employees can ask questions and submit documents to the system. The artificial intelligence large language model (LLM) provides a written response that the users can incorporate into government contracting documents. See attachment for a data flow diagram visualizing data flows in and out of the tool.

1.36 Explain in detail how PII and SBU data flow into, through and out of this system.

System will use model information stored in a GPT-Generated Unified Format (GGUF) or similar type of file. GGUF file stores Generative Pre-trained Transformer (GPT) models for inference. GGUF data is open source, public information. Some personally identifiable information from Internet sources is encoded in models. SBU data such as procurement sensitive or confidential commercial information will be uploaded into the system by users. For example, a draft solicitation document may be uploaded into the system and revised using AI. Also, contractor proposals may be uploaded so the AI can draft an evaluation of vendor strengths and weaknesses for review by IRS employees during the source selection process. Sensitive data may be stored in the system depending on the configuration of specific tools. That said, sensitive data will not flow into open access AI foundation models accessible to the general public.

1.4 Is this a new system?

Yes

1.8 If the system is on the As-Built-Architecture, what is the ABA ID of the system? If this PCLIA covers multiple applications shown on the ABA, please indicate the ABA ID for each application covered separated by a comma.

Compliance Data Warehouse (ABA ID:210212) Enterprise Data Platform (ABA ID: 211416)

1.9 What OneSDLC State is the system in (Allocation, Readiness, Execution)?

Allocation

1.95 If this system has a parent system, what is the PCLIA Number of the parent system?
Not applicable

2.1 If this system discloses any PII to any third party outside the IRS, does the system have a process in place to account for such disclosures in compliance with IRC 6103(p)(3)(A) or Subsection c of the Privacy Act? Contact Disclosure to determine if an accounting is required. Enter "Yes" or "No". If Exempt, type "Exempt".
Yes

2.2 Please provide the full name of and acronym of the governance board or Executive Steering Committee (ESC) this system reports to.
Data and Analytics Strategic Integration Board (DASIB)

3.1 Does your project/system involve any use of artificial intelligence (AI), including virtual assistant, chat bot, and robotic process automation, as defined in Executive Order 13960?
Yes

3.2 What is the methodology used and what database is training your AI?
System will use open source, publicly available large language model (LLM) trained on a vast amount of data (e.g. 1 trillion words) from the Internet. LLMs are trained using a technique called self-supervised learning. This means the model learns by analyzing patterns in unlabeled text data, without the need for humans to manually label or annotate the training data. The model takes in vast amounts of text data and uses that data to pick up on the statistical relationships between words - learning the nuances of linguistics including grammar, semantics, fluency, and the context in which words appear. Some AI tools within the system are fine-tuned with federal procurement specific information such as the Federal Acquisition Regulations.

3.3 Does this system use cloud computing?
Yes

3.31 Please identify the Cloud Service Provider (CSP), FedRAMP Package ID, and date of FedRAMP authorization.
Treasury Cloud (FedRAMP Package ID: FR1801046750 and Authorization Date 03/02/2020) and Azure Government (FedRAMP Package ID: F1603087869 and Authorization Date 04/29/2020)

3.32 Who has access to the CSP audit data (IRS or 3rd party)?
IRS Cybersecurity, Enterprise Security Audit Trails (ESAT)

3.32 Does the CSP allow auditing?
Yes

3.33 Please indicate the background check level required for the CSP (None, Low, Moderate or High).

Moderate

3.4 Is there a breach/incident plan on file?

Yes

3.5 Does the data physically reside in systems located in the United States and its territories and is all access and support of this system performed from within the United States and its territories?

Yes

3.6 Does this system interact with the public through a web interface?

No

3.7 Describe the business process allowing an individual to access or correct their information.

Vendors address any incorrect information in procurement documents with their IRS Contracting Officer (CO). For example, if AI included incorrect information in a solicitation or contract document - vendors can notify the CO (typically via email).

4.1 Who owns and operates the system (IRS Owned and Operated, IRS Owned and Contractor Operated, Contractor Owned and Operated)?

Some tools / components of the system are federally owned and operated. Other system components are contractor owned and operated.

4.2 If a contractor owns or operates the system, does the contractor use subcontractors?

Yes

4.3 What PII/SBU data does the subcontractor have access to?

Subcontractor employees accessing the system have background investigations and non-disclosure agreements completed. Additionally, team has completed Specialized IT Security (SITS) training. The IRS project manager and Contracting Officer's Representative ensures that employees and contractors had background investigations favorably adjudicated and have job duties that require access. Subcontractor employees have read and write access to data.

4.5 Identify the roles and their access level to the PII data. For contractors, indicate whether their background investigation is complete or not.

Contractor employees accessing the system have background investigations and non-disclosure agreements completed. Additionally, team has completed Specialized IT Security (SITS) training. The IRS project manager and Contracting

Officer's Representative ensures that employees and contractors had background investigations favorably adjudicated and have job duties that require access. Contractor employees have read and write access to data.

4.51 How many records in the system are attributable to IRS Employees? Enter "Under 50,000", "50,000 to 100,000", "More than 100,000" or "Not Applicable".

Not Applicable

4.52 How many records in the system are attributable to contractors? Enter "Under 5,000", "5,000 to 10,000", "More than 10,000" or "Not Applicable".

Not Applicable

4.53 How many records in the system are attributable to members of the public? Enter "Under 5,000", "5,000 to 10,000", "More than 10,000" or "Not Applicable".

More than 10,000

4.54 If records are attributable to a category not mentioned above in 4.51 through 4.53, please identify the category and the number of corresponding records to the nearest 10,000. If none, enter "Not Applicable".

More than 10,000

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

Access to AI tools will be determined by BEARS access requested and system authentication mechanisms.

5.1 Please describe any privacy risks, civil liberties and/or security risks identified for the system that need to be resolved and what is the mitigation plan?

PCLIA document mitigates risk of lack of public notice regarding IRS use of generative AI. Further, agency personnel have been informed of AI-specific agency policies including Treasury's Policy and Governance over the development and use of Artificial Intelligence and Large Language Learning Model Automations. Human review will be required for documents / language drafted by the system to ensure accuracy and respect for privacy rights / civil liberties.

5.11 Is there a Risk Assessment Form and Tool (RAFT) associated with this system on file with your organization or the IRS Risk Office.

No

5.2 Does this system use or plan to use SBU data in a non-production environment?

No

Interfaces

Interface Type

IRS Systems, file, or database

Agency Name

Contract Writing System

Incoming/Outgoing

Outgoing (Sending)

Agency Agreement

No

Transfer Method

Other

Other Transfer Method

Users will transfer procurement sensitive documents via uploads from laptops.

Systems of Records Notices (SORNs)

SORN Number & Name

Treasury .009 - Treasury Financial Management Systems

Describe the IRS use and relevance of this SORN.

Contract documents

Records Retention

What is the Record Schedule System?

General Record Schedule (GRS)

What is the retention series title?

Financial Management and Reporting Records

What is the GRS/RCS Item Number?

1.1

What type of Records is this for?

Electronic

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

Records documenting procurement of goods and services.

What is the disposition schedule?

Generally, procurement records are retained for 6 years after final action pursuant to 48 CFR 4.805.

Data Locations

What type of site is this?

Environment

What is the name of the Environment?

ARC Select

What is the sensitivity of the Environment?

Sensitive But Unclassified (SBU)

What is the URL of the item, if applicable?

<https://arcselect.com/>

Please provide a brief description of the Environment.

AI powered tool for vendor proposal evaluation and contractor source selection.

What are the incoming connections to this Environment?

Proposals and other procurement documents can be uploaded.

What are the outgoing connections from this Environment?

Procurement documents can be exported.

What type of site is this?

Environment

What is the name of the Environment?

IBM Cloud Pak for Data

What is the sensitivity of the Environment?

Sensitive But Unclassified (SBU)

What is the URL of the item, if applicable?

<https://cpd-cpd4-dev.apps.micp.dmd.enterprise.irs.gov/zen/#/homepage>

Please provide a brief description of the Environment

Server environment, managed by RAAS Business Unit, that can store model files and run Python code.

What are the incoming connections to this Environment?

Files are uploaded into environment.

What are the outgoing connections from this Environment?

Files can be exported to IRS laptops.

What type of site is this?

Environment

What is the name of the Environment?

NIPRGPT

What is the URL of the item, if applicable?

<https://niprgpt.mil/>

Please provide a brief description of the Environment.

Non-classified Internet Protocol (IP) Router Network (NIPRNet)
Generative Pre-Trained Transformer (GPT)

What type of site is this?

Environment

What is the name of the Environment?

Virtual Acquisition Office Ally

What is the sensitivity of the Environment?

Not Applicable

What is the URL of the item, if applicable?

<https://www.gotovao.com/VAO-Ally>

Please provide a brief description of the Environment.

Website with AI tool that answers procurement policy questions.

What are the incoming connections to this Environment?

Users access a chatbot style interface and can ask question about
the Federal Acquisition Regulations or other procurement policies.

What are the outgoing connections from this Environment?

None

What type of site is this?

Environment

What is the name of the Environment?

Enterprise Data Platform

What is the sensitivity of the Environment?

Sensitive But Unclassified (SBU)

What is the URL of the item, if applicable?

<https://irs.gov.sharepoint.com/sites/edp>

Please provide a brief description of the Environment.

Enterprise Data Platform provides trustworthy advanced analytics and artificial intelligence/machine learning models.

What type of site is this?

Environment

What is the name of the Environment?

Federal Acquisition Regulation AI Sidekick

What is the sensitivity of the Environment?

Not Applicable

What is the URL of the item, if applicable?

<https://www.aframesolutions.com/far-ai-sidekick>

Please provide a brief description of the Environment.

Website with AI tool that answers procurement policy questions.

What are the incoming connections to this Environment.

Users access a chatbot style interface and can ask question about the Federal Acquisition Regulations or other procurement policies.

What type of site is this?

Environment

What is the name of the Environment?

IRS End User Laptops

What is the sensitivity of the Environment?

Sensitive But Unclassified (SBU)

Please provide a brief description of the Environment.

IRS laptops end user computing (General Support System (GSS)
32)

What are the incoming connections to this Environment?

IRS laptops can generally upload to agency servers.

What are the outgoing connections from this Environment?

IRS laptops can generally download from agency servers.