

Date of Approval: 12/20/2024  
Questionnaire Number: 1717

## Basic Information/Executive Summary

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

Text Analytics - Resume Scanner

Business Unit

Small Business and Self Employed

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 tool pre-screens applicants, ranking unqualified or less qualified applicants lower and well qualified applicants higher. This allows reviewers to focus on analyzing fewer applicants in greater detail. Highly ranked applicants are invited to interviews, and eventually, a select group are given job offers.

## 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).

Resumes are received via USA Staffing. We extract and parse the resume text using R and Python. Then, a keyword search is used to discover specific keywords and phrases present or absent (and their frequency) in the documents. The system does not store PII.

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

Address  
Education Information  
Email Address  
Employment Information  
Geographical Indicators  
Name  
Standard Employee Identifier (SEID)  
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)?

System

1.35 Is there a data dictionary for this system?

No

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

Resumes are received via USA Staffing. We extracted and parsed the resume text using R and Python. We applied natural language processing (NLP) techniques to analyze the data as text documents. We implemented a keyword search to discover specific keywords and phrases present or absent (and their frequency) in the documents. We constructed unsupervised models, namely hierarchical clusters, to partition similar applicants into meaningful clusters. Our team used the NLP data transformation method, Bag-of-Words, to convert text into a digital representation to achieve computational analysis. We then ranked applicants

based on the sum of cluster (keyword-related) quantiles for each applicant. The system does not store PII.

1.4 Is this a new system?

Yes

1.5 Is there a Privacy and Civil Liberties Impact Assessment (PCLIA) for this system?

No

1.8 If the system is on the As-Built-Architecture (ABA), what is the ABA ID number of the system? If this PCLIA covers multiple applications shown on the ABA, please indicate the ABA ID number(s) for each application covered separated by a comma. If the system is not in the ABA, then contact the ABA (<https://ea.web.irs.gov/aba/index.html>) for assistance.

No

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

Execution

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?

Natural language processing. The model uses resumes of applicants who want a variety of jobs in the federal government. Although the requirements of the job and the skills of the applicants may change, fundamentally the model uses job requirements to assess whether applicants have relevant skills.

3.3 Does this system use cloud computing?

No

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

3.63 If individuals do not have the opportunity to give consent to collect their information for a particular use, why not?

Once applicants and new hires are routed from USAJOBS to USA Staffing, USA Staffing provides notice to individual applicants through the Privacy links available in the footer of USA Staffing's applicant and new hire interfaces. As they are applying for positions within the federal government, they are aware that the information collected will be reviewed for the purposes of hiring.

3.64 If the individual was not notified of the following items prior to the collection of information, why not? 1) Authority to collect the information 2) If the collection is mandatory or voluntary 3) The purpose for which their information will be used 4) Who the information will be shared with 5) The effects, if any, if they don't provide the requested information.

Once applicants and new hires are routed from USAJOBS to USA Staffing, USA Staffing provides notice to individual applicants through the Privacy links available in the footer of USA Staffing's applicant and new hire interfaces. As they are applying for positions within the federal government, they are aware that the information collected will be reviewed for the purposes of hiring.

3.65 If information is collected from third-party sources instead of the individual, please explain your decision.

Information is retrieved from USA Staffing. The application process both USAJOBS and USA Staffing for applicants to apply for IRS job openings.

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

Applicants have direct access to USA Staffing and the ability to ensure all data is accurate, relevant, and up to date as well as access to other resources, such as the USA Staffing Help Desk and subscribing agency personnel to provide assistance.

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

IRS owned and operated

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

No

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".

Under 5,000

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

The need for access is determined by the Project Lead and Program Manager based on an individual being added to either resume scanner project team or the review panel.

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?

Privacy Risk Related to Characterization of the Information Privacy Risk: There is a risk of collecting inaccurate or falsified information from the applicants, which may result in unfavorable hiring decisions. Mitigation: The risk of collecting inaccurate information is mitigated using field restrictions that permit applicants to enter data only within certain parameters and using user confirmation, in which USA Staffing asks the user to validate that the information submitted is accurate. Privacy Risk Related to Redress Privacy Risk: There is a risk that individuals will not be able to amend incorrect or incomplete information, impacting the hiring decisions and the utility of USA Staffing. Mitigation: This risk is mitigated by providing applicants with direct access to USA Staffing and the ability to ensure all data is accurate, relevant, and up to date as well as access to other resources, such as the USA Staffing Help Desk and subscribing agency personnel to provide assistance.

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

Other Federal Agencies

### Agency Name

Office of Personnel Management (USAJOBS, USA Staffing)

### Incoming/Outgoing

Incoming (Receiving)

### Transfer Method

Secured channel via HTTPS

# Systems of Records Notices (SORNs)

## **SORN Number & Name**

OPM/GOVT-6 - Personnel Research and Test Validation Records

Describe the IRS use and relevance of this SORN.

This SORN is applicable to information contained in USA Staffing which provides resumes utilized by the scanner.

## **SORN Number & Name**

OPM/GOVT-5 - Recruiting, Examining, and Placement Records

Describe the IRS use and relevance of this SORN.

This SORN is applicable to information in USAJOBS and USA Staffing which supplies resumes utilized by the resume scanning tool.

# Records Retention

What is the Record Schedule System?

General Record Schedule (GRS)

What is the retention series title?

General Records Schedule 2.1: Employee Acquisition Records

What is the GRS/RCS Item Number?

51

What type of Records is this for?

Electronic

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

Resumes

What is the disposition schedule?

Temporary. Destroy 2 years after termination of register.

# Data Locations

What type of site is this?

Shared Drive

What is the name of the Shared Drive?

SBR2

Please provide a brief description of the Shared Drive.

This is a SBSE Research shared drive and access to the project documents is only accessible to the project team and hiring panel.