PRIVACY IMPACT ASSESSMENT

Predictive Analytics (PA)

1. Contact Information

A/GIS Deputy Assistant Secretary
Bureau of Administration
Global Information Services

2. System Information

(a) Date of completion of this PIA: March 2022
(b) Name of system: Predictive Analytics
(c) System acronym: PA
(d) Bureau: Consular Affairs (CA)
(e) iMatrix Asset ID Number: 302939
(f) Child systems (if applicable) iMatrix Asset ID Number: N/A
(g) Reason for performing PIA:

☒ New system
☐ Significant modification to an existing system
☐ To update existing PIA for a triennial security reauthorization

(h) Explanation of modification (if applicable):

3. General Information

(a) Does the system have a completed and submitted data types document in Xacta?
☒Yes ☐No - Contact IRM/IA at IASolutionCenter@state.gov for assistance.

(b) Is this system undergoing an Assessment and Authorization (A&A)?
☒Yes ☐No

If yes, has the privacy questionnaire in Xacta been completed?
☒Yes ☐No

(c) Describe the purpose of the system:

The Predictive Analytics (PA) platform is a data science and data analytics computing platform hosted on the Microsoft Azure Commercial Cloud (MAC). The PA platform provides a wide range of tools that allows data scientists and data analysts to perform data analysis to create and share reports, and to develop and deploy machine learning models. PA provides a development environment for data science experiments, reporting, and machine learning /artificial intelligence (AI) development.

The platform contains a suite of supporting capabilities for data engineering and data governance. PA will initially contain a subset of the Consular Consolidated Database (CCD) (Nonimmigrant Visa (NIV) tables) to identify areas of opportunity and methods for improving efficiency in visa-related processes (such as fraud or visa overstay).

(d) Describe the personally identifiable information (PII) that the system collects, uses, maintains, or disseminates:

Noncitizen PII
PA collects the following PII on noncitizens:

- Full Name
- Aliases
- Phone number
- Address
- Email address
- Social security numbers
- Passport number
- National identification numbers
- U.S. Alien registration Number
- Tax identification number
- Date of birth
- Place of birth
- Citizenship
- Nationality
- Medical Information
- Social media indicators
- Race
- Gender
- Arrests and convictions
- Legal information
- Education information
- Financial information
- Employment information
- Family information
- Mother's maiden name
- Business contact information (occupation, work/company name, phone number, and work address).

U.S. Citizen PII: PA may also contain information on U.S. citizens if the applicant enters information in the source system. U.S. citizen Point of Contact (POC) information is a U.S. person who is affiliated with the non-immigrant applicant or provides a means for an immigrant to travel to the U.S.

U.S. citizen POC PII:

- Full Name
- Phone number
- Email address
- Personal Address
- Date of birth
- Affiliation with NIV applicant

Government Business Contact Information: Federal Government Employees (Department of State Adjudicator ID # only).

(e) What are the specific legal authorities and/or agreements that allow the information to be collected?

U.S.C. 1101-1504 (Immigration and Nationality Act of 1952, as amended, Titles I-III, General, Immigration, Nationality and Naturalization)
8 U.S.C. 1701 et seq., Enhanced Border Security and Visa Entry Reform Act
22 U.S.C. 211a-218, 2705, Passports and Consular Reports of Birth Abroad (CRBAs)
Executive Order 11295, August 5, 1966,
31 FR 10603 (Department of State Authority to Issue, Deny, Limit Passports);22 U.S.C. 2651a (Organization of Department of State)
22 U.S.C. 3904 (Functions of service)

(f) Is the information searchable by a personal identifier (e.g., name or Social Security number, etc.)?

☒Yes, provide:
SORN Name and Number: Visa Records, STATE-39
SORN publication date: November 8, 2021

☐No, explain how the information is retrieved without a personal identifier.

(g) Does the existing SORN need to be amended to reflect the inclusion of this new or significantly modified system? ☐Yes ☒No

If yes, please notify the Privacy Office at Privacy@state.gov.

(h) Is there a records retention schedule submitted to or approved by the National Archives and Records Administration (NARA) for this system? ☒Yes ☐No
(If uncertain about this question, please contact the Department's Records Officer at records@state.gov.)

If yes provide (Consolidate as much as possible):

Schedule number: B-09-002-39a
Disposition Authority Number: N1-084-93-14, item 1a

Length of time the information is retained in the system: After 1 year, archive a copy of log files, data, and reports onto a disk, tape, CD, or other electronic media (to allow records to be used in future fraud investigations). Verify copy, then destroy/delete on-line reports.
Type of information retained in the system: Online electronic reports

Schedule number: B09-002-39b(1)
Disposition Authority Number: N1-084-93-14, item 1b(1)(a)
Length of time the information is retained in the system: Destroy when 1 years old.
Type of information retained in the system: Locally created paper logs, logbooks, log files, data, and reports; and off-line electronic reports archived on disks, tapes, CDs, or other electronic media. Because the degree of visa fraud varies by post, the Consular Officer shall determine which of the following time periods is best for post to destroy or retire the information. (1) No fraud problem.

Schedule number: B-03-003-39b(2)
Disposition Authority Number: N1-084-93-14, item 1b(1)(b)
Length of time the information is retained in the system: Destroy when 3 years old.
Type of information retained in the system: Locally created paper logs, logbooks, log files, data, and reports; and off-line electronic reports archived on disks, tapes, CDs, or other electronic media. Because the degree of visa fraud varies by post, the Consular Officer shall determine which of the following time periods is best for post to destroy or retire the information. (2) Low degree of fraud.

Schedule number: B-03-003-39b(3)
Disposition Authority Number: N1-084-93-14, item 2a
Length of time the information is retained in the system: Retire to the RSC when 3 years old. Destroy when 10 years old
Type of information retained in the system: Locally created paper logs, logbooks, log files, data, and reports; and off-line electronic reports archived on disks, tapes, CDs, or other electronic media. Because the degree of visa fraud varies by post, the Consular Officer shall determine which of the following time periods is best for post to destroy or retire the information. (3) Medium degree of fraud.

Schedule number: B-03-003-39b(4)
Disposition Authority Number: N1-084-93-14, item 2a
Length of time the information is retained in the system: Destroy when 20 years old
Type of information retained in the system: Locally created paper logs, logbooks, log files, data, and reports; and off-line electronic reports archived on disks, tapes, CDs, or other electronic media. Because the degree of visa fraud varies by post, the Consular Officer shall determine which of the following time periods is best for post to destroy or retire the information. (4) High degree of fraud.

4. Characterization of the Information

(a) What entities below are the original sources of the information in the system? Please check all that apply.

☐ Members of the Public
☐ U.S. Government employees/Contractor employees
☒ Other (people who are not U.S. Citizens or LPRs)

(b) On what other entities above is PII maintained in the system?

☒ Members of the Public
☐ U.S. Government employees/Contractor employees
☐ Other
☐ N/A

(c) If the system contains Social Security Numbers (SSNs), is the collection necessary?
☒Yes ☐No ☐N/A

- If yes, under what authorization?
8 USC 1182 and 26 U.S.C. 6039E (Information Concerning Residence Status)

(d) How is the PII collected?

The information is transmitted electronically from a subset of the CA Consular Consolidated Databases (CCD) (NIV tables) of nonimmigrant applicants requesting and awarded visa services.

(e) Where is the information housed?

☐ Department-owned equipment
☒ FEDRAMP-certified cloud
☐ Other Federal agency equipment or cloud
☐ Other
- If you did not select "Department-owned equipment," please specify.
The PA platform and its information is hosted on the Department of State authorized SE-Azure Cloud platform.

(f) What process is used to determine if the PII is accurate?

Accuracy of the information in the PA platform is the responsibility of the applicant completing the application for submission or entering data into the source system for visa services. The original PII is validated for accuracy via the source system processes in which the applicant requests and apply for the visa service.

(g) Is the information current? If so, what steps or procedures are taken to ensure it remains current?

Information is checked for currency via the source system where the applicant is requesting service. Any updates or changes to applicant visa information are updated in the CCD system automatically, where the PA platform pulls information to conduct assessments and analytics.

(h) Does the system use information from commercial sources? Is the information publicly available?

No. PA does not use commercial sources of information, nor is the information publicly available.

(i) How was the minimization of PII in the system considered?

The PII items listed in Question 3d are the minimum necessary to perform the actions required by this system. Concerns about collecting and maintaining PII include unauthorized access, disclosure, modification, and/or misuse of the data by users and/or a security breach. These risks were considered during the system design and security configuration. Impact is minimized as collection of PII is limited to only what is required for the system to perform the function of conducting assessments to conduct data analytics of overstay risks.

5. Use of Information

(a) What is/are the intended use(s) for the PII?

The PII allows aggregate analyses of information to detect categories of overstay indicator information. For example, in the case of overstay, a probability model is created to determine the likelihood of overstay for NIV applicants applying for visas, and then the correctness of the predictions are evaluated for various indicators.

(b) Is the use of the PII relevant to the purpose for which the system was designed or for which it is being designed?

Yes, the information collected is used to predict individual as well as group behavior in determining the risk of NIV applicants which is what the system was designed to do.

(c) Does the system analyze the PII stored in it? ☒Yes ☐No

If yes:

(1) What types of methods are used to analyze the PII?
The PA platform is designed to allow a full range of analytics and machine learning processes. The types of analysis that are conducted with the PII Visa data include descriptive analytics (looking at summaries of information, such as frequencies and counts, to determine patterns as well as how variables might be related), and predictive modeling (how individual variables predict outcomes).

Modeled outcomes, such as probability of overstays are associated with mathematical groups of information to predict trends.

(2) Does the analysis result in new information? Yes

(3) Will the new information be placed in the individual's record? ☐Yes ☒No
Information is to predict trends which are not associated with an individual.

(4) With the new information, will the Department be able to make new determinations about the individual that would not have been possible without it?
☐Yes ☒No

(d) If the system will use test data, will it include real PII? ☐Yes ☐No ☒N/A
If yes, please provide additional details.

6. Sharing of PII

(a) With whom will the PII be shared internally and/or externally? Please identify the recipients of the information.

Internal: The data analysis is made available to authorized CA users only via reports on summaries of information such as frequencies to model outcomes such as probability overstays for analysis. No PII put into the system is shared internally via the PA platform.

External: No data is shared externally.

(b) What information will be shared?

Internal: N/A

External: N/A

(c) What is the purpose for sharing the information?

Internal: N/A

External: N/A

(d) The information to be shared is transmitted or disclosed by what methods?

Internal: N/A

External: N/A

(e) What safeguards are in place for each internal or external sharing arrangement?

Internal: N/A.

External: N/A

7. Redress and Notification

(a) Is notice provided to the record subject prior to the collection of his or her information?

The PA platform does not collect information from applicants. PA information is obtained from the CCD system (NIV tables). Respective notices are provided via the source systems collecting information from applicants requesting visa services that are stored in CCD. Information is given voluntarily by the consenting applicants, family members, or the designated agent. Individuals are informed that failure to provide the information necessary to process the application may result in the application being rejected.

(b) Do record subjects have the opportunity to decline to provide the PII or to consent to particular uses of the PII?
☐Yes ☒No

If yes, how do record subjects grant consent?

If no, why are record subjects not allowed to provide consent?
The PA platform does not collect information directly from applicants, but instead pulls information from the CCD (NIV tables) to conduct analysis. Consent is acquired via the source systems in which the information is originally obtained when applicants request consular services.

(c) What procedures allow record subjects to gain access to their information?

Visa customers cannot access information in PA as it is not a public facing system. However, they can acquire access to their records contained in PA by following the procedures in SORN STATE-39, Visa Records.

(d) Are procedures in place to allow a record subject to correct inaccurate or erroneous information?
☐Yes ☒No

If yes, explain the procedures.

If no, explain why not.
Individuals must follow processes of the source system where they applied for the visa services to request correction of information. Individuals can also correct records contained in PA by following the procedures in SORN STATE-39, Visa Records.

PA platform data will be pulled from CCD and updated as data in the CCD is updated via the source systems.

(e) By what means are record subjects notified of the procedures to correct their information?

The PA platform does not collect information directly from applicants. PA pulls information from the CCD system (NIV tables), which resides outside the boundary of PA. Individuals can follow procedures in SORN STATE-39 regarding points of contacts to inquire about corrections to their information.

PA platform data will be pulled from CCD and updated as data in the CCD are updated via the source systems.

8. Security Controls

(a) How is all of the information in the system secured?

PA is hosted on the Department of State (Department) authorized SE-Azure platform which has undergone FedRAMP and Department security assessments. The SE-Azure platform incorporates multiple layers of Cloud Service Provider and Department implemented security features including management, operational and technical security controls, access management, auditing, firewalls, physical security, and continuous monitoring.

PA is further secured within SE-Azure using an Azure private network to further limit access to the system. Access to the services provided by PA is limited to authorized Department users, including cleared contractors who have a justified need for the information to perform official duties. Access to PA is protected with additional access controls set at the storage and service level. All system accounts and access are granted in accordance with established Department account management policies.

(b) Explain the different roles that have been created to provide access to the system and the PII (e.g., users, managers, developers, contractors, other).

Access to PA is role-based and the user is granted only the role(s) required to perform officially assigned duties and roles approved by the supervisor. Department of State PA users, system administrators, and database administrators have access to the system to aid in conducting analytics and trends of potential visa applicant stays.

(c) Describe the procedures established to limit system and data access to only those individuals who have an "official" need to access the information in their work capacity.

Separation of duties and least privilege access are employed; users have access to only the data that the supervisor and local Information System Security Officers (ISSOs) approve to perform official duties. Access is role-based, and the user is granted only the role(s) required to perform officially assigned duties.

The Azure cloud leverages Azure Role Base Access (RBAC) to control and require authorization to Azure services, security functions, security protection and the monitoring of permissions. Least privileges are restrictive rights/privileges or access users need for the performance of specified tasks. The Department of State works to ensure that users who must access records containing PII only have access to the minimum amount of PII, along with only those privileges (e.g., read, write, execute) that are necessary to perform their job duties. Users are uniquely identified and authenticated before accessing PII.

(d) How is access to data in the system determined for each role identified above?

Access to PA is role-based and the user is granted only the role(s) required to perform officially assigned duties as approved by the supervisor. Supervisors and the local Information System Security Officers (ISSO) determine the access level needed by a user to ensure it correlates to the user's particular job function, manager's approval, and level of clearance

All system account/access is granted in accordance with established Department account management policies.

(e) What monitoring, recording, auditing safeguards, and other controls are in place to prevent the misuse of the information?

Azure RBAC on PA along with a Departmental entitlement approval process are in place to restrict non-privileged users from disabling, circumventing, or altering implemented security safeguards/countermeasures. All user activities are audited and monitored.

In accordance with Department of State requirements, auditing is enabled to track the following events:

- Multiple log-on failures
- Log-ons after-hours or at unusual times
- Addition, deletion, or modification of user or program access privileges

The purpose of the audit trail is to document unintended modification or unauthorized access to the system and to dynamically audit retrieval access to designated critical data.

(f) Are procedures, controls or responsibilities regarding access to data in the system documented? ☒Yes ☐No

The PA System Security Plan (SSP) contains the procedures, controls, and responsibilities regarding access to data in the system.

(g) Explain the privacy training provided to each role identified in 8(b) that has access to PII other than their own.

All system administrators must take the IA210 System Administrator Cybersecurity Foundation Course which has a privacy component. In accordance with Department of State computer security policies, mandatory security training (PS800 Cyber Security Awareness) is required for all authorized users. Each user must annually complete the Cyber Security Awareness Training, which has a privacy component, to access or use systems. Additionally, all Department of State personnel are required to take the course PA318 Protecting Personally Identifiable Information biennially.

The State Department’s standard “Rules of Behavior” regarding the use of any computer system and the data it contains require that users agree to the rules and that they must protect PII through appropriate safeguards to ensure security, privacy, and integrity.