Template Version Number: 01-2021

U.S. Department of Commerce
U.S. Patent and Trademark Office

Privacy Threshold Analysis
for the
Patent Search AI (PSAI)

U.S. Department of Commerce Privacy Threshold Analysis

USPTO Patent Search AI (PSAI)

Unique Project Identifier: PTOC-00-060-00

Introduction: This Privacy Threshold Analysis (PTA) is a questionnaire to assist with determining if a Privacy Impact Assessment (PIA) is necessary for this IT system. This PTA is primarily based from the Office of Management and Budget (OMB) privacy guidance and the Department of Commerce (DOC) IT security/privacy policy. If questions arise or further guidance is needed in order to complete this PTA, please contact your Bureau Chief Privacy Officer (BCPO).

Description of the information system: *Provide a brief description of the information system.*

The E-Government Act of 2002 defines "information system" by reference to the definition section of Title 44 of the United States Code. The following is a summary of the definition: "Information system" means a discrete set of information resources organized for the collection, processing, maintenance, use, sharing, dissemination, or disposition of information. See: 44. U.S.C. § 3502(8).

Patent End to End (PE2E) is a new web-based system that integrates all phases of the patent application process into a unified set of tools that can be accessed through a single user interface.

Search, one of the primary components of PE2E is the application used by examiners to search prior art documentation to make patentability decisions. Due to exponential growth in patents submissions over the years, Patent Examiners have experienced challenges using traditional search tools to efficiently and effectively find the 'right' references to enable them make appropriate patentability decisions. To address this challenge, USPTO decided to leverage emerging technology and implemented an Artificial Intelligence (AI) solution to augment its current search systems to help the agency's 9,000+ examiners perform search faster, identify more relevant results, deliver better and more thorough output.

Patent Search Artificial Intelligence (PSAI) is the system of choice designed to address the challenges describe above. It combines AI Technologies that are specifically custom-made machine learning (ML) models, cloud-based deployments and user experience development integrated with PE2E Search. The intended purpose of the PSAI system (with AI capabilities) is to augment existing PE2E search user interface in a manner that allows examiners to perform searches faster, identify more relevant search results, deliver better and more thorough output in a high compute and secure cloud environment hosted in Google Cloud Platform (GCP).

A component of the PSAI search experience is the 'Similarity Search' feature that leverages AI Capabilities to provide examiners with similar results based on application data to inform search strategies and identify relevant prior art. Similarity Search will be delivered as a new gadget within the existing PE2E System and will utilize similar patterns and show users patent application and document data accordingly. The gadget operates within an iFrame and will integrate directly with PE2E features within a workspace.

Address the following elements:

a) *Whether it is a general support system, major application, or other type of system*
Patent Search AI is a Minor Application.

b) *System location*

The system lives in two places. The end users of the application use the Gadget which operates within an iFrames and integrates directly with PE2E features within a workspace. The backend of the system resides in the USPTO's private Google Cloud environment (Google Cloud Platform us-east4 region) and deployed across three different availability zones (us-east4-a, us-east4-b and us-east4-c).

c) *Whether it is a standalone system or interconnects with other systems (identifying and describing any other systems to which it interconnects)*

The PSAI system interconnects with:

Patent End to End (PE2E) – PE2E serves all of Patents and is composed of 3 components, PE2E-OC, PE2E-DAV, and PE2E-Search, and together they provide capabilities for users to review their dockets, manage their work, open applications and review contents, perform prior art searches against foreign and domestic patents and create official communications to patent applicants explaining the Office's position on patentability.

USPTO Google Cloud Services (UGCS) - a standard infrastructure platform used to support the USPTO Patent Search AI system hosted in the Google Cloud Platform (GCP) us-east4, Northern Virginia, environment.

Network and Security Infrastructure (NSI) – facilitates the communications, secure access, protective services, and network infrastructure support for all USPTO applications.

d) *The purpose that the system is designed to serve*

The intended purpose of the PSAI system (with AI capabilities) is to augment existing PE2E search user interface in a manner that allows examiners to perform searches faster, identify more relevant search results, deliver better and more thorough output in a high compute and secure cloud environment hosted in Google Cloud Platform (GCP).

e) *The way the system operates to achieve the purpose*

PSAI combines AI Technologies that are specifically custom-made machine learning (ML) models, cloud-based deployments and user experience development integrated with PE2E Search to augment existing PE2E search user interface in a manner that allows examiners to perform searches faster, identify more relevant search results, deliver better and more thorough output in a high compute and secure cloud environment hosted in Google Cloud Platform (GCP).

Outside of platform and systems architecture, PSAI leverages NOAA N Wave (Wide Area Network) WAN for network interconnections, Big Data Reservoir (BDR) Data Lake as the source for unpublished datasets pushed/transformed by Google Cloud Platform (GCP) Cloud composer to PSAI Landing zones and USPTO's CICM/SCDAD services for source code management via GitLab as well as, Continuous Integration Continuous Delivery (CICD) pipelines for collaboration with on-premises configuration components shared with other PE2E Search teams. The System is registered with all applicable enterprise system registries at USPTO, including Dynamic Operational Support Plan (DOSP) and GEARS.

f) A general description of the type of information collected, maintained, used, or disseminated by the system

The information across the system includes user interactions within the application's visual screens, similarity context between patents, application logs, and Cooperative Patent Classification (CPC) suggestion data. Published and Unpublished patent information including claim results and/or applications could be queried by application number under certain circumstances if the application number is used.

g) Identify individuals who have access to information on the system

PII/BII and Unpublished Patent Data in the IT system will be shared within the bureau by Patent examiners and development teams on a case-by-case basis, bulk transfer, and direct access.

Patent examiners and Designated System Admins will only have a 'Read Only' access to the information on the system. There'll be zero modification right to unpublished patent data in the system.

PSAI's infrastructure is separate from PE2E Search's infrastructure but, the applications are conjoined meaning, the user interface of PE2E Search host the PSAI's user-facing components. As such, PSAI is built on top of an existing enterprise system (PE2E), leveraging access control and identity management for new and existing assets in PE2E

Search. PSAI will leverage Okta for enterprise identity management using OIDC to secure access to user-owned assets within GCP.

h) How information in the system is retrieved by the user

The information in the system is retrieved by the user interface (UI) through the PE2E (via the embedded Gadget UI Search application programming interface) hosted in a USPTO secure cloud environment.

BDR extracted and transformed data through application pipeline exports using GCP Cloud Composer can be retrieved by System Admins after they have been loaded in the PSAI landing zones from BDR.

i) How information is transmitted to and from the system

The information is transmitted through private encrypted network traffic between end-user (examiners) machines, PE2E Search/Gadget/Similarity Search User Interfaces, Application APIs (e.g. Search API) and the USPTO secure cloud environment. PSAI application traffic is logically protected using the USPTO PKI/signed TLS (Transport Layer Security).

BDR Patent Applications data is transmitted using Application Pipeline Export processed by GCP Cloud Composer to PSAI landing zones.

Questionnaire:

1. Status of the Information System

1a. What is the status of this information system?

☐ This is a new information system. *Continue to answer questions and complete certification.*

☐ This is an existing information system with changes that create new privacy risks. *Complete chart below, continue to answer questions, and complete certification.*

Changes That Create New Privacy Risks (CTCNPR)					
a. Conversions	☐	d. Significant Merging	☐	g. New Interagency Uses	☐
b. Anonymous to Non-Anonymous	☐	e. New Public Access	☐	h. Internal Flow or Collection	☐
c. Significant System Management Changes	☐	f. Commercial Sources	☐	i. Alteration in Character of Data	☐
j. Other changes that create new privacy risks (specify):					

☐ This is an existing information system in which changes do not create new privacy risks, and there is not a SAOP approved Privacy Impact Assessment. *Continue to answer questions and complete certification.*

☒ This is an existing information system in which changes do not create new privacy risks, and there is a SAOP approved Privacy Impact Assessment. *Skip questions and complete certification.*

1b. Has an IT Compliance in Acquisitions Checklist been completed with the appropriate signatures?

☐ Yes. This is a new information system.

☐ Yes. This is an existing information system for which an amended contract is needed.

☐ No. The IT Compliance in Acquisitions Checklist is not required for the acquisition of equipment for specialized Research and Development or scientific purposes that are not a National Security System.

☒ No. This is not a new information system.

2. Is the IT system or its information used to support any activity which may raise privacy concerns?

NIST Special Publication 800-53 Revision 4, Appendix J, states "Organizations may also engage in activities that do not involve the collection and use of PII, but may nevertheless raise privacy concerns and associated risk. The privacy controls are equally applicable to those activities and can be used to analyze the privacy risk and mitigate such risk when necessary." Examples include, but are not limited to, audio recordings, video surveillance, building entry readers, and electronic purchase transactions.

☐ Yes. *(Check all that apply.)*

Activities			
Audio recordings	☐	Building entry readers	☐
Video surveillance	☐	Electronic purchase transactions	☐
Other (specify):			

☒ No.

3. Does the IT system collect, maintain, or disseminate business identifiable information (BII)?

As per DOC Privacy Policy: "For the purpose of this policy, business identifiable information consists of (a) information that is defined in the Freedom of Information Act (FOIA) as "trade secrets and commercial or financial information obtained from a person [that is] privileged or confidential." (5 U.S.C.552(b)(4)). This information is exempt from automatic release under the (b)(4) FOIA exemption. "Commercial" is not confined to records that reveal basic commercial operations" but includes any records [or information] in which the submitter has a commercial interest" and can include information submitted by a nonprofit entity, or (b) commercial or other information that, although it may not be exempt from release under FOIA, is exempt from disclosure by law (e.g., 13 U.S.C.)."

☒ Yes, the IT system collects, maintains, or disseminates BII.

☐ No, this IT system does not collect any BII.

4. Personally Identifiable Information (PII)

4a. Does the IT system collect, maintain, or disseminate PII?

As per OMB 17-12: "The term PII refers to information that can be used to distinguish or trace an individual's identity either alone or when combined with other information that is linked or linkable to a specific individual."

☒ Yes, the IT system collects, maintains, or disseminates PII about: *(Check all that apply.)*

- ☒ DOC employees
- ☒ Contractors working on behalf of DOC
- ☐ Other Federal Government personnel
- ☒ Members of the public

☐ No, this IT system does not collect any PII.

If the answer is "yes" to question 4a, please respond to the following questions.

4b. Does the IT system collect, maintain, or disseminate Social Security numbers (SSNs), including truncated form?

☐ Yes, the IT system collects, maintains, or disseminates SSNs, including truncated form.

Provide an explanation for the business need requiring the collection of SSNs, including truncated form.
Provide the legal authority which permits the collection of SSNs, including truncated form.

☒ No, the IT system does not collect, maintain, or disseminate SSNs, including truncated form.

4c. Does the IT system collect, maintain, or disseminate PII other than user ID?

☒ Yes, the IT system collects, maintains, or disseminates PII other than user ID.

☐ No, the user ID is the only PII collected, maintained, or disseminated by the IT system.

4d. Will the purpose for which the PII is collected, stored, used, processed, disclosed, or disseminated (context of use) cause the assignment of a higher PII confidentiality impact level?

Examples of context of use include, but are not limited to, law enforcement investigations, administration of benefits, contagious disease treatments, etc.

☐ Yes, the context of use will cause the assignment of a higher PII confidentiality impact level.

☒ No, the context of use will not cause the assignment of a higher PII confidentiality impact level.

If any of the answers to questions 2, 3, 4b, 4c, and/or 4d are "Yes," a Privacy Impact Assessment (PIA) must be completed for the IT system. This PTA and the SAOP approved PIA must be a part of the IT system's Assessment and Authorization Package.

CERTIFICATION

☒ The criteria implied by one or more of the questions above **apply** to the Patent Search AI (PSAI) and as a consequence of this applicability, a PIA will be performed and documented for this IT system.

☐ The criteria implied by the questions above **do not apply** to the Patent Search AI (PSAI) and as a consequence of this non-applicability, a PIA for this IT system is not necessary.

System Owner Name: Jonathan Horner Office: Office of Information Technology for Patents (P/OITP) Phone: (571) 270-7358 Email: Jonathan.Horner@uspto.gov Signature: Users, Horner, Jonathan R. Digitally signed by Users, Horner, Jonathan R. Date: 2023.08.29 08:56:12 -04'00' Date signed: _____	**Chief Information Security Officer** Name: Timothy S. Goodwin Office: Office of the Chief Information Officer (OCIO) Phone: (571) 272-0653 Email: Timothy.Goodwin@uspto.gov Signature: Users, Goodwin, Timothy Digitally signed by Users, Goodwin, Timothy Date: 2023.09.05 17:16:36 -04'00' Date signed: _____
Privacy Act Officer Name: Heaton John Office: Office of General Law (O/GL) Phone: 703-756-1240 Email: Ricou.Heaton@uspto.gov Signature: John Ricou Heaton Digitally signed by John Ricou Heaton Date: 2023.08.02 17:46:44 -04'00' Date signed: _____	**Bureau Chief Privacy Officer and Co-Authorizing Official** Name: Henry J. Holcombe Office: Office of the Chief Information Officer (OCIO) Phone: (571) 272-9400 Email: Jamie.Holcombe@uspto.gov Signature: Users, Holcombe, Henry Digitally signed by Users, Holcombe, Henry Date: 2023.09.06 13:03:38 -04'00' Date signed: _____
Co-Authorizing Official Name: Vaishali Udupa Office: Office of the Commissioner for Patents Phone: (571) 272-8800 Email: Vaishali.Udupa@uspto.gov Signature: Vaishali Udupa Digitally signed by Vaishali Udupa Date: 2023.09.11 10:38:52 -04'00' Date signed: _____	