Background

In Canada, privacy is considered a human right. As the majority of the data we handle at IRCC is personal information, privacy requirements must be top-of-mind when planning, developing and monitoring any initiative involving data-driven technology.

This document is intended to outline the privacy protections in place for this particular model or tool. The requirement statements in this document are based off the Baseline Privacy Requirements for Disruptive Technology that lays out the minimum privacy requirements that must be met for all initiatives involving disruptive technology.

This document **does not replace** the need for a Privacy Impact Assessment (PIA) as it is intended only to analyze privacy compliance at a model level and document steps taken to increase privacy protections. This document may be used to assist in completing a larger initiative or program-level privacy assessment as required. Program areas are responsible for filling out a Privacy Needs Assessment (PNA) and sending it to the ATIP Division. Information about the PNA and the template can be found on Connexion.

Details

Name of Initiative:	
Branch/Division who is responsible for the Model:	
Type of Model (triage, predictive, risk pattern identification, etc):	
Date model is expected to be launched:	
Version 1.0 Model Privacy Assessment is completed:	
Update MPA whenever the Model or Tool is updated	

Model Privacy Assessment

Lighthouse – Study Permit Pilot #2

Version 1.0 Final – May 2021

Model or Tool Update Date:	
Version 2.0 Model Privacy Assessment completed:	

Summary of Initiative

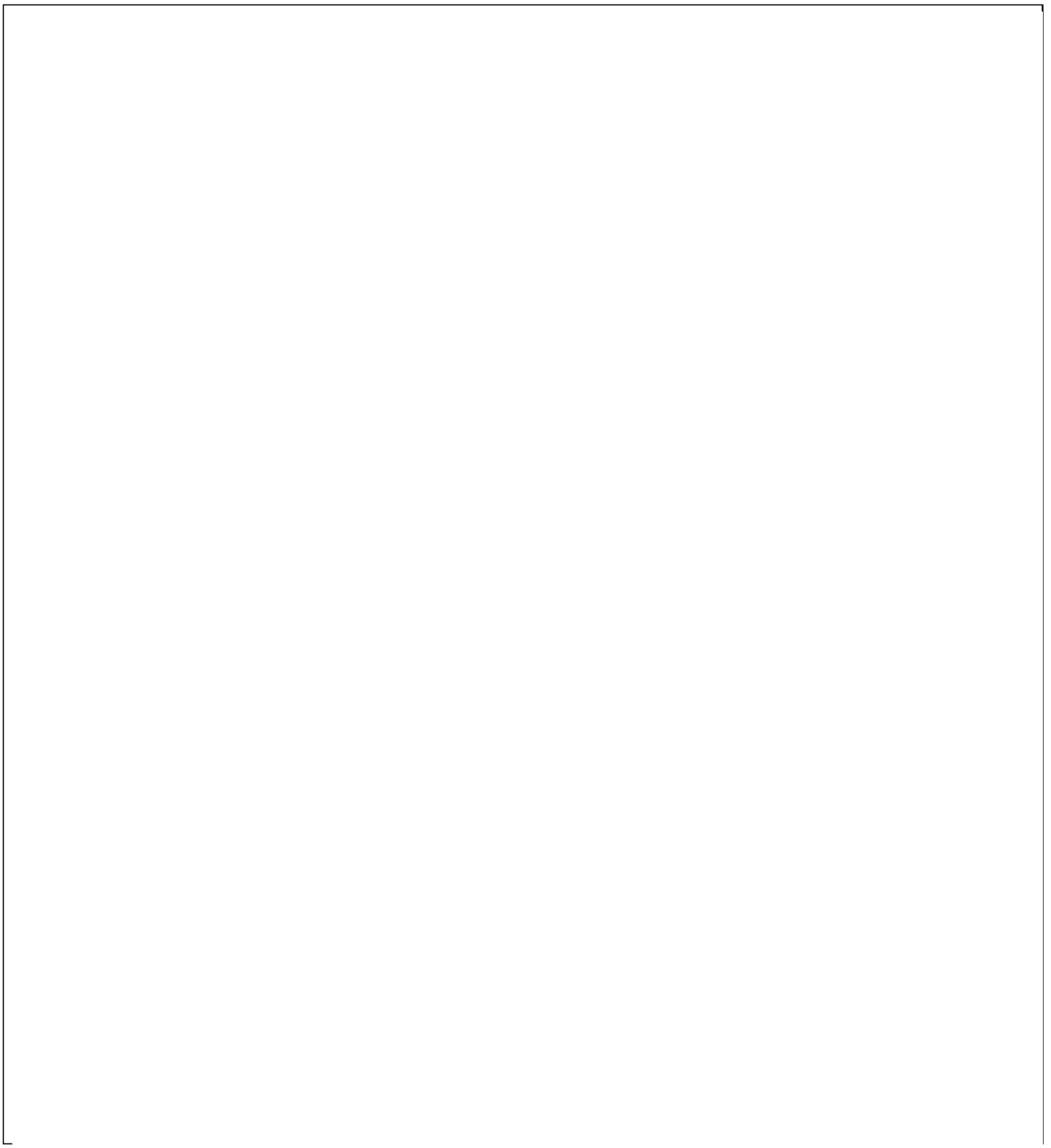
What is the problem this model is trying to solve? Why is this the best solution to that problem? Why is the use of personal information necessary? What does the model do? What is the population the model is being applied to (ex: study permit applicants from a specific country)? In what way does the model support officers in making decisions? Does it suggest decisions for officers? Please include any and all useful information.

Summary of Initiative
Lighthouse is a prototype risk detection tool developed by IRCC's Advanced Analytics Solution Centre (A2SC). The tool aims to enhance program integrity and Canadian public safety by automatically identifying and summarizing historical risk patterns for IRCC officials. The tool aims to provide neutral, fact-based risk information to IRCC officials, augmenting their capacity to quickly identify and understand organized fraud trends and other risk patterns. It can be used to support frontline decision makers or to understand historical risk patterns. Lighthouse presents opportunities for the department to modernize IRCC's risk assessment activities by placing timely and relevant risk information at the fingertips of IRCC officials in a manner that was previously impossible or prohibitively expensive to do.

Model Privacy Assessment

Lighthouse – Study Permit Pilot #2

Version 1.0 Final – May 2021



Model Privacy Assessment

Lighthouse – Study Permit Pilot #2

Version 1.0 Final – May 2021

High-Level Summary of How the Model Works (as appropriate, include: how the rules are created, how an application runs through a model/how personal information is used in production, an overview of the output/what officers see, etc.)

Model Privacy Assessment

Lighthouse – Study Permit Pilot #2

Version 1.0 Final – May 2021

Requirements

Legal Authority

A program must have the parliamentary authority to collect and use personal information for the specified purposes of the program. This legal authority will be identified in the Privacy Needs Assessment and other required privacy assessments for this initiative.

A program must also be legally allowed to use disruptive technology and/or automation to support the program. Identify which of the legal authorities below grant your program area the **authority to administer your program(s) using electronic means** (disruptive technology):

☒ Part 4.1 of the Immigration Refugees Protection Act

☐ Section 2.2 of the Passport Order

Accountability

A part of ethical and responsible development and deployment of advanced analytics, artificial intelligence and automation initiatives is ensuring that **humans are ultimately responsible for the model's behaviour**. Describe below the processes to guarantee that

model rules are reviewed by humans and to make sure that internal governance and accountability (sign-off) processes are in place.

Model Privacy Assessment

Lighthouse – Study Permit Pilot #2

Version 1.0 Final – May 2021

Source of Data

Personal information must only be collected if it relates directly to an operating program or activity of IRCC and each personal information data element must be necessary to the administration of the program. When possible, personal information should be collected directly from the individual.

For the purposes of disruptive technology initiatives, **only information found in departmental systems of record (ex. GCMS) should be used** unless another data source has been approved by IT Security, and the activities of training models and algorithms should be done outside of those systems of record. Data collected from outside sources should not be used unless demonstrably necessary, and proper information sharing

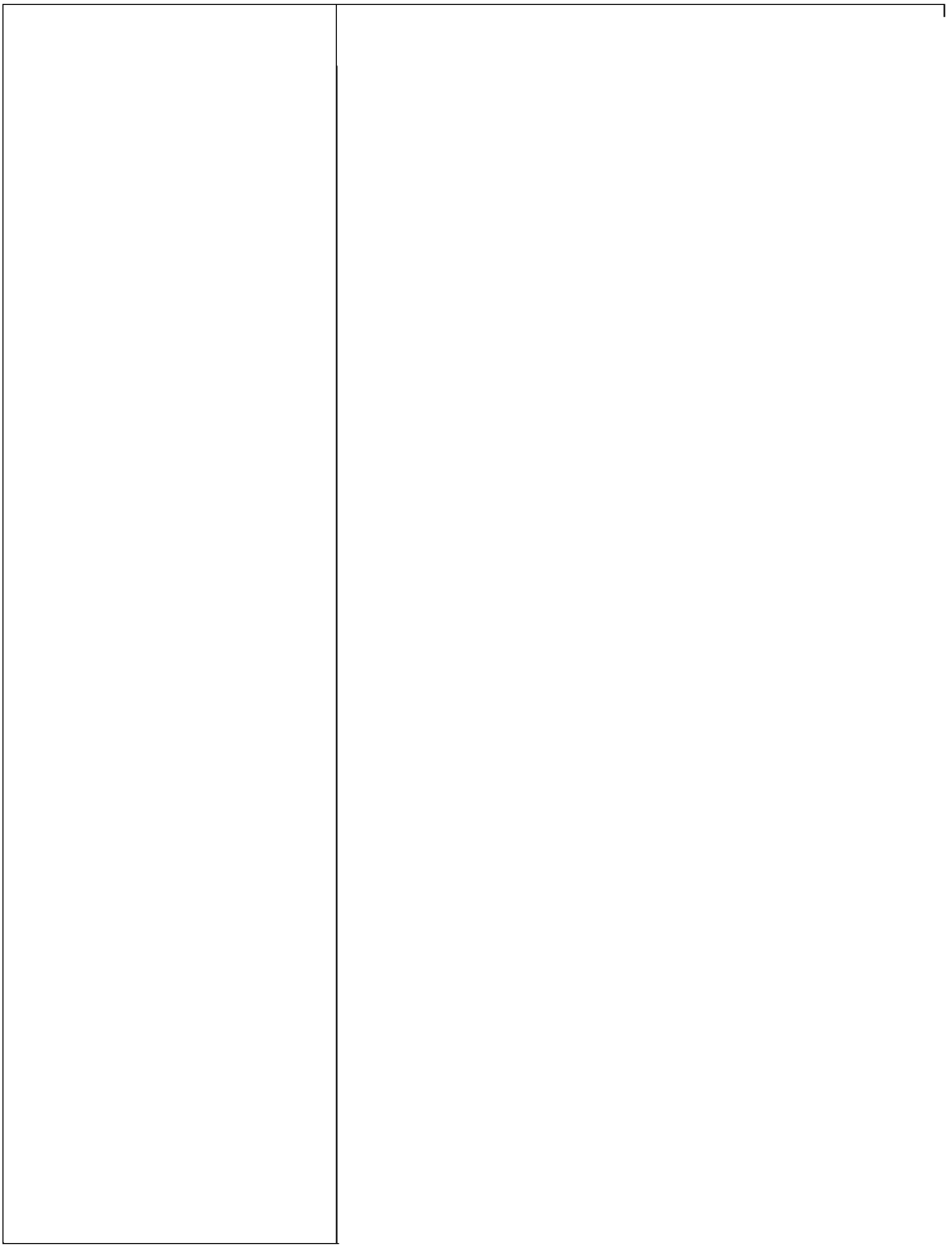
agreements, memoranda of understanding, service level agreements etc. should be in place and followed.

Source of Data – Where the data comes from and reasonableness of using that data	
Source of the data (ex: GCMS)	
If outside data is used, through what means is it collected (ex: MOUs, ISAs, etc.)	
Demonstrate how Necessity, Proportionality, Effectiveness, and Minimal Intrusiveness, were considered and applied in considering what data to use in the model (Oakes Test).	

Model Privacy Assessment

Lighthouse – Study Permit Pilot #2

Version 1.0 Final – May 2021



Model Privacy Assessment

Lighthouse – Study Permit Pilot #2

Version 1.0 Final – May 2021

Notice / Informed of Purpose / Transparency / Explainability

IRCC must notify individuals (clients and the general public) of the purpose for which their information is being collected, commonly referred to as a 'privacy notice.' This notice must be given at or before the time of collection. IRCC must notify past applicants that their information was used to train or build models. Individuals have a right to know exactly how their personal information was processed through a disruptive technology system. Ensuring that plain language explanations are available on demand would allow individuals to see how technology was used to support decision-making.

Notice & Transparency – How individuals (clients and the general public) will be notified about the use of this model and how the use of disruptive technology will be explained to applicants.	
Notice at time of collection (link to the privacy notice if applicable)	The privacy notice on the Study Permit forms a well as the corresponding Personal Information Bank (PIB) have been updated to account for the use of analytics.
Transparency to historical applicants and public	The Digital Transparency webpage that is now published accounts for the use of risk screening tools such as Lighthouse by explaining that advanced data analytics systems will be used by IRCC to recognize patterns to help accelerate our work and better inform decision makers.

Explainability of the model (link to the plain language explanation)	The pattern reports that are used in Lighthouse are intended to provide sufficient evidence of any individual pattern. All patterns are "self-contained" in that they are intended to stand on their own merit regardless of the rest of the system. Encrypted data and model information is currently being retained to allow A2SC to recreate a model on-demand and provide the underlying data to explain how Lighthouse patterns were produced and why applications matched against these patterns.

Accuracy

IRCC must take all reasonable steps to ensure that personal information used for an administrative purpose is as **accurate, up-to-date and complete as possible**. This also includes ensuring there are mechanisms to correct inaccurate information.

For initiatives involving disruptive technology, this involves ensuring data is collected from a reliable source, the quality of the data, developing technological mechanisms to make certain that the technology is working (such as feedback loops and blind tests), quality assurance on the outputs, and so on. Additionally, to guarantee the accuracy of the data, program areas must take the necessary steps to minimize unintended bias in the data. Finally, accuracy also involves model maintenance and ensuring the model is trained and re-trained on the most updated, accurate and reliable data.

Accuracy – Ways the model is ensuring accuracy of the data and outputs, and the process to correct inaccuracies. Describe any Quality Assurance (QA) processes that are in place.

Use

Personal information must only be used for the purpose it was initially collected, a use consistent with that purpose or for a purpose for which it is may be disclosed under section 8(2) (see Disclosure, below).

Applying disruptive technology to a dataset involving personal information is a use; this includes all uses whether administrative or not. Personal information must be treated appropriately regardless of the level of automation or support the technology is providing. The use of disruptive technology should be a consistent use of the personal information. To determine what constitutes a consistent use of personal information, the original purpose and the proposed purpose must be so closely related that the individual would expect that the information would be used for the consistent purpose, even if the use is not spelled out.

Model Privacy Assessment

Lighthouse – Study Permit Pilot #2

Version 1.0 Final – May 2021

Use – Mechanisms that are in place to reduce the inappropriate use of the data	
Confirmed that applying this model is a consistent use of the personal information (provide the text from the legal opinion if applicable at time of completing this assessment)	
Data minimization (only use data elements that are absolutely necessary)	
Reducing data granularity (removing precision of some data elements, ex: using only the first three digits of a postal code)	
De-identification (masking/hashing/synthesizing	

data so that it's no longer personally identifiable)	
Need to Know (ensuring access is only granted to those who need to know it)	
Other (Ex: Use of privacy enhancing technologies, anonymizing data for all demonstrations, etc)	

Disclosure

Personal information under IRCC's control must not be disclosed to anyone or any organization for any reason, except for those reasons listed in sub-section 8(2) of the *Privacy Act*.

For initiatives involving disruptive technology, this includes information found in departmental systems of record (ex. GCMS) that other organizations such as CBSA or CSIS can view. Regular information sharing may continue to occur between IRCC and partner organizations, however Memoranda of Understanding (MOUs), Information Sharing Agreements (ISAs) and other formal agreements must be updated and modified through the appropriate channels if there is the desire to disclose outputs on a regular basis.

Disclosure – Mechanisms that are in place to reduce the risk of inappropriate disclosure of the data, outputs and other model-related personal information. List GC partners and MOUs ISAs linked to this project as applicable.

Safeguards

Personal information must be safeguarded appropriately regardless of the kind of technology applied to it. **Appropriate administrative, technical and physical safeguards should be applied to personal information at all stages of a disruptive technology initiative,** and consideration should be given to reducing the likelihood of privacy and security breaches throughout development.

Safeguards – Safeguards that are in place in and around the model to protect the data

Retention and Disposal

Personal information used for an administrative purpose (as a part of a decision-making process that affects the individual) must be retained for at least two years, and in accordance with the appropriate Retention and Disposition Schedule. All data (with the exception of training data in the Exploration Zone) must be kept such that in the event of a complaint or legal action, the decision can be replicated. The data in the Exploration Zone that is an exact duplicate of production data and that is used to generate and retrain model rules can be considered transitory.

Retention and Disposition schedules	
Training Data in EZ	
Client Input Data (during production)	
Model Outputs	

Model (the code)
Reports (outputs for officers, legal tables, others as appropriate)

Monitoring Plan

Monitoring for privacy compliance to the above-noted requirements should be built in from the model development phase and a monitoring schedule post-deployment should be followed.

In addition to being certain that the disruptive technology is working properly, here is a list of non-exhaustive monitoring activities to plan for:

- **Collection**: Make sure no data from sources other than departmental systems of record (ex. GCMS) and other IRCC data repositories are included in the disruptive technology, and if there is outside data, put in place the appropriate ISA or MOU and keep these up to date.
- **Notice**: Review privacy notices and transparency and explainability practices for accuracy and for current information. Update when required.
- **Retention and Disposal**: Review the retention and disposal practices and ensure that no information is retained beyond IRCC's prescribed retention period.
- **Accuracy**: Build in regular data quality practices to ensure data is accurate, up to date and as complete as possible, and modify the information when required.

- **Use**: Review data handling practices, ensure mitigation measures against inappropriate use are functioning properly, employ new measures as required, and update practices periodically.
- **Disclosure**: Review disclosure practices so that disclosures are occurring as a part of up to date MOUs and/or ISAs, and ensure other government organizations can only see information they are permitted to see in GCMS under those MOUs/ISAs. Make modifications when required.
- **Safeguards**: Complete the mandatory IT Security Assessment and Authorization process and conduct security checks to confirm that the training data, the technology itself and the outputs are secure.

Monitoring – Monitoring for privacy compliance once the model is in production. Please describe the steps you will follow to develop and establish the required monitoring plan.

Monitoring Notes:

The quality assurance and monitoring plans will monitor the ongoing efficacy of the model and of the data quality. The quality assurance and monitoring activities will contribute to monitoring the Accuracy and Use privacy requirements. At the time of completing this document, a detailed plan to monitor all of the privacy requirements was not in place. However, the measures that are in place should mitigate many potential privacy risks in the future (such as encrypting data, limiting access to only those with a 'need to know' etc.) This model will be in a pilot mode from June 2021 to October 2021, and many assessments will take place during that time. Building in additional privacy controls may occur during the pilot and they will be recorded in future versions of this document.

As of June 2, 2021, a formal plan to monitor the privacy requirements for the Lighthouse model has not been completed.

Gap Analysis and Proposed Recommendations

Requirement	Gap	Proposed Recommendation	Status
Retention and Disposal	No clear retention and disposal schedules have been determined.	Meet with IM to determine a retention and disposal schedule for all Lighthouse data.	Not begun – to engage with IM in the coming weeks

Monitoring	A clear plan to monitor the effectiveness of privacy controls and adherence to privacy requirements.	Currently quality assurance mechanisms are in place to monitor the accuracy and effectiveness of the model. However, Privacy recommends that a more detailed monitoring plan be developed to monitor adherence to privacy requirements or include it in existing documentation.	Not begun

Documents Reviewed to Gather Information

The documents below are linked to their source in GCDOCS. As future readers of this MPA may not have access to those documents, below is a point-in-time capture of the documents as of May 20, 2021. Because they are point-in-time, note that they may be drafts, so please use the hyperlink when possible.

1. Lighthouse Privacy Needs Assessment - http://gcdocs2/otcs/cs.exe?func=ll&objaction=overview&objid=392470772

A2SC - PNA - Lighthouse for SP.d

2. Lighthouse Project Charter - http://gcdocs2/otcs/cs.exe?func=ll&objaction=overview&objid=389516701


A2SC - Project Charter - Lighthouse

3. Lighthouse Study Permit Pilot Legal Opinion - http://gcdocs2/otcs/cs.exe?func=ll&objaction=overview&objid=392337429


Lighthouse Study Permit Pilot Legal Op

Signature

Invalid signature

X

Tracy Perry
ATIP Director
Signed by: Perry, Tracy