PBOT-PSU Numina pilot sensor project

Threshold Privacy Analysis

Delivered

Smart City PDX

August 4, 2022

THRESHOLD PRIVACY ANALYSIS REPORT

City of Portland Privacy Toolkit

WHAT IS THE THRESHOLD PRIVACY ANALYSIS?

The Threshold Privacy Analysis ("TPA") is a method to quickly evaluate what are the general privacy risks of a technological solution or a specific use, transfer or collection of data to City bureaus or offices. The TPA is a way to identify factors that contribute to privacy risks and lead to proper strategies for risk mitigation or alternatives that may even remove those identified risks.

The Threshold Privacy Analysis may lead to a more comprehensive Privacy Impact Assessment (PIA) and a Surveillance Assessment depending on the level or risks identified and the impacts on civil liberties or potential harm in communities.

In the interests of transparency about data collection and management, the City of Portland has committed to publishing all Privacy Assessments on an outward facing website for public access. TPAs do not include specific uses of technology or data other than those initially evaluated.

WHEN IS AN THRESHOLD PRIVACY ANALYSIS RECOMMENDED?

A TPA is recommended when:

- A project, technology, data sharing agreement, or other review has been flagged as having some privacy risk due to the collection of private or sensitive data.
- A technology has high financial impact and includes the collection, use or transfer of data by city bureaus or third parties working for or on behalf of the city.

HOW TO COMPLETE THIS DOCUMENT?

City staff complete two documents:

- *The Threshold Privacy Analysis form.* This document identifies all important information related to the project description, data collection, use, safekeeping, and management; as well as a verification of existing privacy policies and measures to protect private information.
- *The Privacy Risk Assessment.* This document breaks the privacy risk into six different areas of evaluation: (1) Individual Privacy Harms; (2) Equity, Disparate Community Impact; (3) Political, Reputation & Image; (4) City Business, Quality & Infrastructure; (5) Legal & Regulatory; and, (6) Financial Impact. Then compares risks to the likelihood of happening to create a single risk measure based on the worst case scenario.

Executive summary

This is a pilot project managed by the Portland Bureau of Transportation in partnership with PSU using Numina sensors on PBOT assets to collect street level data on localized travel behaviors.

Numina's sensors capture Detection Images multiple times per second, and use edge processing to extract object data from the images in real-time. These images are not stored or sent to any other server, and the image is deleted from the sensor as soon as it has been processed. Object data extracted from the processed images are transmitted to the cloud and deleted immediately after transmission.

The worst case scenario analyzed in this assessment resulted in Medium Risk. In a nutshell, Numina sensors do not collect personal identifiable information and the only main risk is due to the lack of transparency on the pilot, sensors installed in the public realm, and what is the purpose of the project.

This main risk can be mitigated by informing local communities properly about the goals of the project and connecting them with existing needs with quantifiable outcomes. Transparency can also be improved if labels are used to inform the purpose of the technology and where to find more information on the sensors located in the public realm.

Another identified risk is having lower sensor performance or identification errors from the system. This technology uses algorithms to identify objects and people moving in the field of view and verification on the ground is encouraged, as well as implementing vendor techniques that may improve performance, as long as impacts on privacy or cost are not outside a reasonable threshold.

For more information about this report contact the Smart City PDX team at smartcitypdx@portlandoregon.gov

Threshold Privacy Analysis

	Portland threshold privacy analysis for a technology, project, data sharing agreement or app solution
Version 0.3	**This information is considered restricted and for internal use only until the client clears it to the public. This notice must be remove when authorized for publication**
Information	**Request information**
Bureau	Portland Bureau of Transportation
Assessment done by (name/email)	Hector Dominguez / hector.dominguez@portlandoregon.gov
Date of Assessment	July 19, 2022
Document status	**Delivered**
Name of the assessment	**PBOT-PSU Numina pilot sensor project**
General description	A pilot project in partnership with PSU to install Numina sensors on PBOT assets to collect street level data on localized travel behaviors.
Evaluation topic	**Assessment**
Purpose of the technology, project, data sharing or application	Numina's sensors capture Detection Images multiple times per second, and use edge processing to extract object data from the images in real-time. These images are not stored or sent to any other server, and the image is deleted from the sensor as soon as it has been processed. Object data extracted from the processed images are transmitted to the cloud and deleted immediately after transmission. The data collected does not include any images or personal information that could be used to identify a particular person, vehicle, etc. The objects identified are: pedestrian, car, bicycle, bus, truck. The resulting dataset includes traffic volume metrics Volume counts for the following categories: people, bikes, cars, trucks, and buses. Volume by mode with timestamp Desire lines and movement patterns accounted for by mode

	Numina is making a clear effort to prioritize privacy and information protection by using techniques that anonymize or de-identify individuals.
Name of the entity owner of the application and website	Numina https://numina.co
Type of Organization	Private entity
Scope of personal data collected. List all sources of data and information.	Images of individuals or vehicles are collected using a video camera. Edge processing de-identify people and vehicles and reduces image resolution. Numina's sensors capture Detection Images multiple times per second, and use edge processing to extract object data from the images in real-time. These images are not stored or sent to any other server, and the image is deleted from the sensor as soon as it has been processed. Object data extracted from the processed images are transmitted to the cloud and deleted immediately after transmission. Sensors also collect Sample Images for use in training and validation services. Once per hour, at a random time within each 1-hour interval, one of the Detection Images is transmitted to Numina's servers. This sample rate equates to approximately 0.003% of the total images collected and processed by the sensors. These images are used by Numina to evaluate how our models are performing and make adjustments to improve their accuracy.
How personal data is collected	No personal information is collected.
Who can access the data	Numina employees and authorized agents have access to original sample images. Images are de-identified before access. Customer or designated third-parties have access to processed information that includes identified objects via Dashboard with valid login. Also accessible by

	Numina employees for quality control.
Purposes the data is used for	Images are used for different purposes: - Algorithm quality control (accuracy validation) - Algorithm improvement and evaluation - Interpreting object tracks - Object movement data for the API, Dashboard, and reports
Where the data is stored	Data is stored within the Numina sensor and some images may be sent to Numina's servers on AWS for processing, calibration, or other purposes connected to the service, including quality assurance, validation, and specific customer needs.
How data is shared	De-identified Sample Images may be shared in a limited fashion with our customers and partners. For example, Numina may share example De-identified Sample Images to help our customers understand Numina functionality, but Numina will never provide a full set of images collected. For data annotation purposes, Numina may also share de-identified images with authorized contractors. Numina never shares the Original Sample Images and never sells any images collected from sensors.
How long is the data stored?	Original Sample Images are retained for 30 days. Numina employees and authorized agents have access to these images. Images are de-identified before access i.e. no one looks at these images directly. De-identified images may be stored on Numina servers for up to 180 days beyond the original retention period, and, on rare occasion, some permanent retention may occur if the customer gives permission for the inclusion of images in reports and communications. Images and data in calibration mode are not deleted. Calibration data enables to continually improve machine learning algorithms, and is used as a ground truth for benchmarking accuracy of Numina algorithms and to create more accurate models which are periodically pushed out to all sensors. Object data, processed information that includes identified objects, is retained indefinitely. Customer or designated third-parties have access via Dashboard with valid login. Also accessible by Numina employees for quality control.

Effectiveness	Numina Computer Vision System serves three main purposes: I. Object detection II. Object classification III. Object tracking Different sources of errors impact how effective the Numina system is. Object detection is the first gate for accurate data, and multiple correct detections are essential for successful tracking. Errors can include: I. Detection errors. A. Missed detection. False negative error. B. Over-detection. False positive error. C. Location error. II. Classification error. A. Misclassification errors. III. Tracking error. A. Track break error. Numina has implemented data validation strategies to mitigate and minimize these errors. The latest upgrades of Numina's system includes a sensor calibration period where images are collected in a different regime and training and calibration images are kept permanently in the sensor. Calibration Mode is a two-week period of increased data sampling and data retention, designed to improve sensor accuracy in new environments. This mode has been designed to collect the minimum viable amount of data to achieve long-term data accuracy and mobility goals, while preventing the collection of potential PII. Calibration Mode allows Numina to fine-tune sensors to their specific deployment environment and build training datasets to improve, or develop new, detectors for different kinds of objects (e.g. scooters). Performing a calibration period also allows customers to collect accuracy benchmarks for their specific Numina deployment environment. Aggregated data is delivered in the form of: I. object counts II. Path visualization III. Activity heat maps

	IV. Behavior zones https://numina.co/defining-accuracy-for-street-level-mobility-data/ https://numina.co/an-update-to-numinas-privacy-policy-introducing-calibration-mode/

Proportionality, fundamental rights, frequency of the collection, and data protection and privacy issues line unintended data collection or processing.	As the Numina system is not collecting Personal Identifiable Information (PII), the services do not offer individual privacy rights. The frequency of the image collection is designed to feed the machine learning algorithms that provide the service. Numina's sensors encrypt all communication with TLS1.2 using industry standard AES-256 encryption. Only authorized devices can communicate with sensors. This requirement removes pathways for data interception or sensor access by unauthorized third parties. All data is stored in Amazon Web Services (AWS), and access is restricted to current Numina employees and contractors using AWS best practices for identity and access management. Numina applies de-identification processes including lowering the image quality and resolution, and removing PII from people and vehicles.
Privacy safeguards	Numina restrains access to employees and authorized agents by de-identifying images before access. Customers and designated third-parties access aggregated data only through a password protected dashboard or API.
Open source	No
AI/ML claims	Yes
Privacy Policy (link)	https://numina.co/wp-content/uploads/2022/03/Numina-Privacy-Policy-effective-March-15-2022.pdf
Privacy risk	Medium
Surveillance Tech?	Yes
Portland Privacy Principles (P3)	
Data Utility	The Numina system collects just the minimum information and applies privacy protection processes to reduce the risks. The final product delivers information about vital real-time, ground-level intelligence to help urban planners and municipal governments design better streets and public places based on their multimodal use.
Full lifecycle stewardship	The Numina platform de-identify information and constrains the use of images to specific cases and processing for delivering aggregated data. No human agent has access to high resolution images. Algorithms are constraint to the tasks designed for delivering this service.

Transparency and accountability	Numina offers a variety of documentation about how personal information is de-identified and protected, including their privacy policy and blog posts. However, Numina does not refer to a third party audit of their services. Numina offers an email for public inquiries: inquiries@numina.co https://numina.co/wp-content/uploads/2022/03/Numina-Privacy-Policy-effective-March-15-2022.pdf https://numina.co/blog/
Ethical and non-discriminatory use of data	Numina applies privacy by design principles and refers to 'cities [becoming] more walkable, bikeable, equitable — with data' in their marketing materials.
Data openness	Aggregated and de-identified data is available to customers
Equitable data management	Equitable data management depends on how this pilot is used. The strategic business goals describe the following uses of the Numina pilot: - Safety. Data will help us analyze different modes of travel in targeted areas within PBOT right of way. - Mobility - Inform policies aimed at reducing carbon emissions and promoting transportation justice.
Automated Decision Systems	The Numina system uses machine learning algorithms to identify mobility subjects including people, bikes, cars, and buses. The aggregated data delivered from this service will not be used for any automated decision system.
Optional	
Consent	Subjects captured in images do not have the option of consent. However, images are de-identified and processed to lower their resolution when accessible to stakeholders. Some minimum risks of re-identification may appear based on the timestamp and the observable features of the subject, including people and vehicles.

Privacy Impact Risk Severity Assessment

WORST CASE SCENARIO	Medium

Severity (Evaluate for the worst / highest possible impact)					
	Impact	**Risk**	**Likelihood**	**Comments**	**Risk level**
Individual Privacy Harms	Moderate	1.1 Risks of Re-identification of an individual or a vehicle	Unlikely	Numina has built a system that de-identifies collected images. However, certain metadata and contextual aspects like timestamps, item color, or vehicle types in low transit periods may increase the risk of re-identification. Minimizing the collection of images or creating minimum thresholds for reporting aggregated objects in low traffic periods can minimize the risk of re-identification.	Low

Equity, Disparate Community Impact	Moderate	2.1 Risk of creating collective impacts to specific groups	Unlikely	Connecting demographics of the location of the sensors may provide how specific groups of local residents use public spaces. Location of sensors to community spaces like churches, schools, or community centers may be perceived as invasive. Comment: Work with local residents and organizations when sensors are placed nearby places of gathering in the community.	**Low**
	Moderate	2.2 Risk of biased results or reporting	Unlikely	Deriving conclusions that apply to larger geographic areas from few sensors may create biases. There could exist specific bias connected to the location where the sensors are located. Demographic data may provide insight about the localized mobility patterns and reduce some bias in analysis.	**Low**

Political, Reputation & Image	High	3.1 Risk due to the lack of transparency	Possible	Local residents may perceive the presence of cameras as surveillance. Proper information about the purpose of the devices, who is the entity responsible, and how to find more information is highly recommended. Publishing a report of how effective these sensors are, including how the information is being used and by what entities, can improve transparency, particularly if local residents are involved and understand the benefits of the technology. Data can be released as open data.	**Medium**
	Moderate	3.2 Risk due to impacts in public trust around technology use.	Possible	Public interest technology approach refers to the study and application of technology expertise to advance the public interest in a way that generates public benefits and promotes the public good. Try to have open houses and other public events and use social media to connect with local residents and get local involvement in this project.	**Medium**

City Business, Quality & Infrastructure	Moderate	4.1 Risk of lack quality and device performance	Possible	Many solutions that claim machine learning involvement have not been delivered in the past. Assurance of ground truth matching and verification of object detection in the field are important. A small pilot project will allow quality assurance in a cost effective way and reduce general costs.	**Medium**
	Moderate	4.2 Risk of infrastructure damage	Unlikely	Damage to infrastructure could appear due to faulty electrical damage, installation failure, or vandalism. Make sure compliance with installation practices of devices on public infrastructure is followed. Vandalism is perhaps the hardest to predict. Install devices out of people's reach and verify that devices are in operation.	**Low**
	Moderate	4.3 Risk of cybersecurity attacks or cyber security breach	Unlikely	Work with Information security and assure compliance with City standards. Assure that there is a responsible to keep all access keys under control.	**Low**

Legal & Regulatory	low	5.1 Risk of consumer data privacy compliance and privacy breach.	Unlikely	Given that no personal identifiable information is collected, these risks are low.	**Low**
Financial Impact	low	6.1 Risk due to privacy damage compensation.	Unlikely	Given that no personal identifiable information, financial impacts due to potential lawsuits are practically null.	**Low**
WORST CASE SCENARIO	**Medium**				