Privacy and Data Protection with Pathr.ai

Authors: Jason Sadowski PhD, David Valdez, Zoë Cayetano, Nicole O'Keefe, George Shaw
Published Date: September 10, 2021

Intro

Over the past several years, the EU's General Data Protection Regulation[1] and California's Consumer Privacy Act[2] have brought the technology industry to a new level of privacy compliance. At Pathr.ai, we hold ourselves to the highest standards of personal privacy and anonymity; and through these standards, we are committed to fulfilling and exceeding these privacy regulations. In this paper, we define how our technology protects personal privacy while generating business insights.

The question at the heart of modern data collection is whether an application collects personally identifiable information (PII). We use the definition of PII provided by the U.S. Office of Privacy and Open Government:

> The term personally identifiable information refers to information which can be used to distinguish or trace an individual's identity, such as their name, social security number, biometric records, etc. alone, or when combined with other personal or identifying information which is linked or linkable to a specific individual, such as date and place of birth, mother's maiden name, etc.[3]

In essence, PII can be thought of as data that can identify a unique person (including their likeness). At Pathr.ai, we use anonymous data collection and analysis to remove all identifiable characteristics.

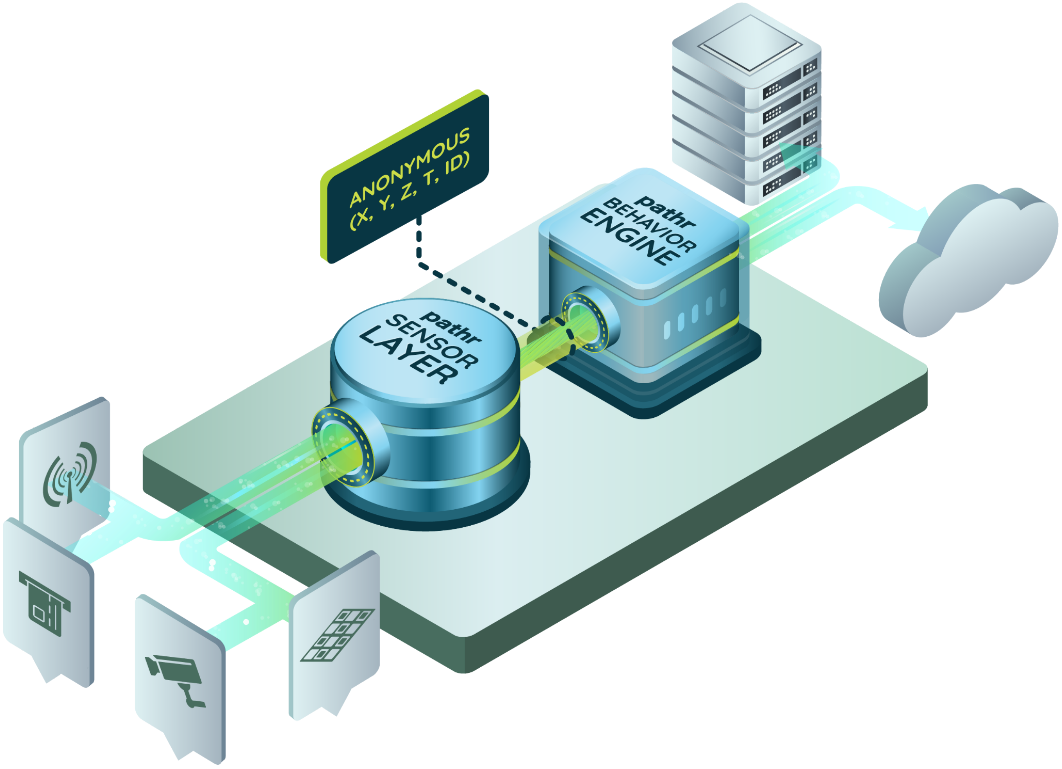
[1] GDPR.eu. *Complete guide to GDPR compliance.* https://gdpr.eu/

[2] State of California Department of Justice. *California Consumer Privacy Act (CCPA).* https://oag.ca.gov/privacy/ccpa

[3] U.S. Office of Privacy and Open Government. *Safeguarding Information.* https://www.osec.doc.gov/opog/privacy/PII_BII.html

How it works

The Pathr.ai stack can be broken up into two sections: 1) the Sensor Layer and 2) the Behavior Engine. Importantly, this stack functions **in real-time** on the edge at the client's premise, ensuring that no PII leaves their site without their permission.



Sensor Layer

Our proprietary Sensor Layer is designed to use any pre-existing sensors (e.g., cameras, RFID readers, elevator telemetry) installed at a site. Most commonly, these sensors are the existing security cameras used by the client. The Sensor Layer links to the video streams provided by these cameras via HTTPS or RTSP protocol either through each camera directly or through a network video recorder (NVR). It requires only the address of each stream and the proper credentials for video access.

The Sensor Layer organizes the different sensor streams into their appropriate ingestion format, then uses artificial intelligence (AI) to detect individuals. Since most PII concerns are with video streams, we'll use the video stream as a case study.

Once the video stream is ingested, we use AI-based computer vision models to detect people in the frame. These computer vision models use whole body detection and **no** facial recognition algorithms. Once a person has been detected, we convert these detections into a coordinate point on the client's floor plan by using the position and field of view of the camera.

Once a frame has been analyzed and detections have been converted, the frame is deleted from the local memory of the device. No video record is retained without the client's express permission.

Using these points, the Sensor Layer generates tracks of individuals through space and passes these tracks into our Behavior Engine. We represent these tracks as a collection of the following information: location coordinates, timestamps, and a random identifier assigned by person. By design, the tracking information contains absolutely no PII. This tracking information is the only data ever sent through the rest of the Pathr.ai system.

Behavior Engine

The Behavior Engine is the part of the stack that powers our spatial intelligence. It uses the track data generated by the Sensor Layer in tandem with the site's spatial features to create quantifiable analytic insights for our clients. It is at this stage where we can leverage the power of machine learning.

A good example analytic is store traffic. We pass the tracks and the location of the entrance into our Behavior Engine and it then estimates how many people passed through the entrance and into the store. The Behavior Engine looks at the tracks and determines which paths are associated with store entrances. Importantly, these mathematical models contain no identifying characteristics of **who** entered the store, only that an entrance occurred.

Each time the Behavior Engine identifies an entrance, it creates an "entrance event" with the timestamp. The creation of this event further anonymizes the data; no information is retained about **which track** created that event, only that an event was created.

It is only at this stage that analytics and tracks leave the site and are sent to our cloud databases.

Third Parties

Pathr.ai has never and will never sell data to third parties. The point data generated by the processes detailed above is owned by our customers. Our customers have the final say in where their data goes.

In order to maintain our analytic accuracy, Pathr.ai shares a small amount of video data from each installed site with third party annotators. We collect approximately one hour of video data from each appropriate camera (e.g., those on the sales floor). Our annotators then create a validation dataset for our calibration process. We call this video dataset the "canonical dataset".

Canonical Dataset

The only PII dataset that Pathr.ai collects is the "canonical dataset". As mentioned above, in order to generate validation metrics for each site, we require a small portion of video (approximately 1 hour per camera) to be annotated. This video data is always collected at the edge, then transferred to our trusted annotators using industry-standard authentication and encryption methods such as HTTPS or SSH. At the end of the pilot period or once we complete the video annotation and analytic calibration process (whichever comes first), we and our annotators delete all copies of that video file.

Occasionally, the initial canonical dataset is insufficient for client needs. In those cases, if the client has provided advance authorization, we collect additional annotation datasets at our discretion. These additional datasets follow the same protocols as our canonical one.

Each client is provided with a Research and Development opt-in agreement prior to beginning work. If the client opts into our research program, their canonical video may be retained by our computer vision team for model improvements. If the client's data is particularly valuable, we will work with them to collect and retain additional datasets. All of this video will be kept at a central location and encrypted while not in use. We maintain a record of all video retained for these purposes and can easily delete the video if requested by the client or an affected individual.

What This Means for Data Privacy

In summary, Pathr.ai's stack is privacy-sensitive and functions in real-time on the edge at the client's premises, guaranteeing that no PII leaves their site without their permission.

Our commitment to protecting our clients' data is at the forefront of every decision we make. Each part of our technical stack is designed with personal privacy in mind. As data moves through the stack, PII is stripped away until all that is left are anonymous events and anonymous tracks.

We are the stewards of our clients' data and their customers' privacy. Protecting that data is our greatest responsibility and we hold ourselves accountable to the highest standard.

For more information, please email info@pathr.ai.