

**How anonymization  
can solve autonomous  
driving data privacy  
challenges**

# Solving the challenges of autonomous vehicles

Solving the challenge of developing autonomous vehicles for the road as a mainstream transportation option has proven difficult even for the world's leading automotive and technology companies.

Much of the difficulty lies in dealing with the vast amounts of data collected by autonomous vehicles and determining the best ways to comply with the various government regulations being implemented around the management and sharing of that data.

Yet another challenge is how to protect the privacy of drivers and pedestrians whose identities or personal information may be among the data captured, stored and analyzed as autonomous vehicles travel in public spaces. This paper closely examines these privacy issues and shows how anonymization technology can help companies solve one of the most vexing autonomous driving challenges: how best to protect personal privacy.

## Key drivers in autonomous vehicle development

As automotive manufacturers develop autonomous driving technology, they run vehicles through lengthy and rigorous road tests, during which hundreds of petabytes of data are gathered through onboard and exterior cameras and sensors. Road tests conducted in metropolitan areas can potentially capture personal data such as the faces and license plates — together with time, date and location — of drivers and pedestrians in sight of a test car.

This data must be kept confidential and secure, and governments around the world have implemented strict regulations that address the collection, storage, use and sharing of personal data to safeguard personal data privacy. One of the most prominent sets of regulations is the European Union's General Data Protection Regulation (GDPR), along with more specific guidelines for autonomous driving being developed by the European Data Protection Board. Many other jurisdictions around the world have issued regulations related to data privacy, including Japan (APPI), Canada (PIPEDA) and in the United States, the California Consumer Privacy Act (CCPA). In 2018 Singapore approved guidelines that vehicles for hire must comply with the Personal Data Protection Act (PDPA) when using in-vehicle recording devices to monitor passengers.

To comply with these regulations, companies handling large amounts of personal data must establish measures to protect personal privacy. For example, when processing personally identifiable information (PII), companies must define clear structures and processes to guarantee data privacy at all times. Collecting and analyzing data while complying with privacy regulations poses a significant challenge to auto manufacturers.

Gathering and processing a lot of personal data involves great responsibility. For example, the tremendous amount of data required for autonomous driving development, along with the involvement of other companies in the process, increases the risk of data being mishandled.

One way to meet the challenge of using driving data effectively while meeting privacy regulations is to anonymize personal data with advanced software.

### **Managing massive volumes of data**

Developing autonomous vehicles involves rigorous testing and using the information gathered to improve safety and performance. Information collected includes vehicle data such as system status or information on the vehicle's powertrain, data related to location and navigation, and data from sensors, cameras and radar that is very memory intensive. Sensor data gauges parameters such as driving conditions and different scenarios the vehicle has encountered. Video data holds sensitive personal information, specifically, faces and license plates.

From working with global manufacturers, DXC Technology knows that sensors and cameras in autonomous vehicles generate up to 19 terabytes of data per hour, depending on the level of autonomy. Dealing with such enormous amounts of data — from collecting, storing and organizing it, to applying analytics to gain value from it — requires the orchestration of a robust IT infrastructure with adequate storage and compute power, advanced software and often third-party partner support. The DXC Robotic Drive data development platform, for example, has been designed to solve the engineering challenges of dealing with the hundreds of petabytes involved with a fleet of test vehicles.

Automotive manufacturing companies typically don't develop all aspects of autonomous driving on their own, so data has to be made available to third parties. Even as the data collected is used to improve vehicle performance, the data analysis process has become increasingly reliant on these third parties, which, in turn, often need to share data with subcontractors to cope with all the work.

This role of third parties and subcontractors increases the risk of some of that data being leaked, published, accidentally sent to the wrong location, or stored in a location that is not compliant with local privacy regulations. Furthermore, suppose a breach of privacy regulations results in a lawsuit by a third party or a subcontractor. In that case liability is with the auto manufacturer, because original equipment manufacturers are ultimately responsible for that data.

### **Dealing with data responsibly**

Gathering and processing a lot of personal data involves great responsibility. For example, the tremendous amount of data required for autonomous driving development, along with the involvement of other companies in the process, increases the risk of data being mishandled and resulting in fines or legal action. Still, dealing correctly with personal data is not just about conforming to government regulations such as GDPR or avoiding lawsuits. It is also about company branding, positioning in the market and letting customers know that the enterprise is being responsible.

Many of the regulations related to personal privacy are not crystal clear, so it can be difficult to know whether the collection, processing and analysis of personal data is truly compliant. It is crucial that companies do not simply pay lip service to complying

with regulations, but instead make every effort to ensure that personal data and images are being protected. If an original equipment manufacturer can demonstrate that every technical effort is being made to protect personal data, the company can be confident of compliance with regulations.

### A privacy-compliant anonymization solution

DXC is at the forefront of supporting autonomous driving research and development and is addressing these privacy concerns. Our DXC Robotic Drive solution provides the platform, toolkit and technical expertise for managing and analyzing large amounts of autonomous driving data. Backed by DXC's global presence and ability to scale, the solution serves as an end-to-end platform that provides the IT infrastructure, computing power, accelerators and automation required for dealing with large amount of autonomous driving data.

DXC Robotic Drive uses anonymization technology to address privacy protection challenges, and DXC is also augmenting its capabilities by integrating technology from brighter AI into the DXC Robotic Drive platform. This makes brighter AI available for automatic integration into any existing autonomous driving development environment (**Figure 1**). This solution is one component of the DXC Robotic Drive toolkit, which includes multiple layers to deal with aspects of autonomous vehicle development such as security, applications and analytics. For other solutions that might be relevant for a certain customer use case, DXC is technology agnostic.

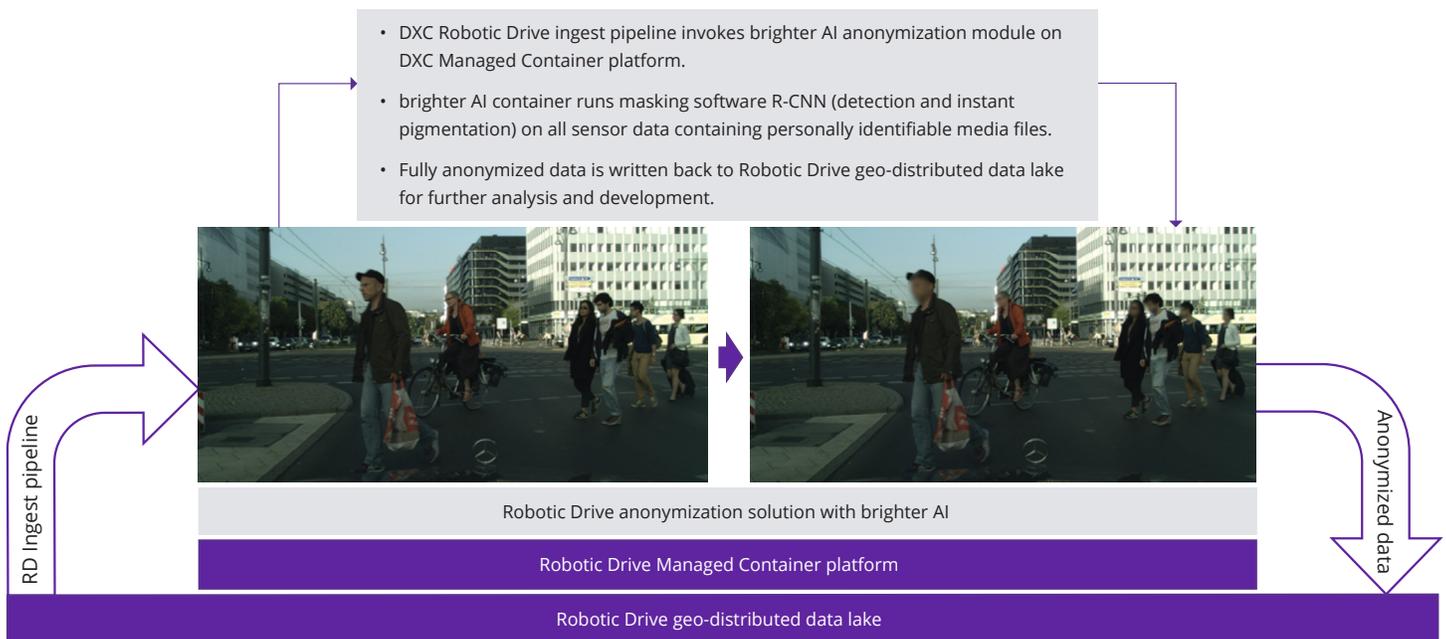
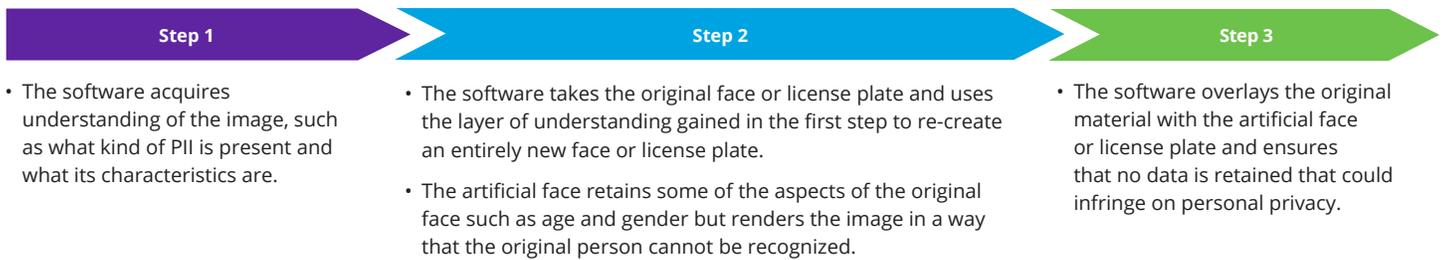


Figure 1. Integration of brighter AI technology into the DXC Robotic Drive platform

The core software from brighter AI, brighter Redact, uses a process called Deep Natural Anonymization. In autonomous driving, the source material is gathered by recording video footage that contains public content, including people's faces and license plates. Then, applying machine learning algorithms, the software uses a three-step process to anonymize the data and generate a synthetic image (**Figure 2**).



**Figure 2.** brighter AI's Deep Natural Anonymization process for generating synthetic images

The DXC Robotic Drive Containerized Compute Platform helps scale this solution to process the anonymization workloads for millions of images and videos efficiently. DXC Robotic Drive also provides a geo-distributed data lake that stores and manages automotive datasets before (restricted access) and after anonymization (accessible to applications for consumption).

The value of brighter AI's software is that the anonymization process is fully automated and requires no human labor to change PII data. Because the process is held to such high data privacy protection standards, metrics that might need to be provided to a legal department, for example, ensure that no human involvement is required in the anonymization process.

This anonymization technique is much more valuable than simply blurring faces and license plates, because facial features and physical attributes can still be recognized, and that data can be used to train machine learning models. The solution combines technical innovation with effective protection of personal privacy, distinguishing it from other redaction techniques. Importantly, this approach ensures that video recordings are in compliance with the strict data protection guidelines stipulated by GDPR and other regulations.

### **Anonymization in other industries**

DXC's technology platform for handling large amounts of data, including this innovative anonymization technology, can be applied in many use cases beyond autonomous driving. These include healthcare settings (emergency rooms or surgery), retail (tracking customer behavior), research (anonymizing research subjects) and public transportation.

In public transportation, for example, movements of train passengers can be recorded and analyzed to better understand customer behavior and preferences.

Working with huge amounts of data is not just about collecting it, storing it and processing it. Orchestrating the data environment is essential.

This data collection and analysis can lead to more efficient train scheduling to better fit passenger loads and increase convenience for commuters, but it could also pose similar privacy issues.

Anonymization of personal data is just one piece of a complex puzzle that needs to be fitted together and integrated on a large scale. Working with huge amounts of data is not just about collecting it, storing it and processing it. Orchestrating the data environment is essential. In any industry that deals with large amounts of data, organizing the data so that it is available at the right time to the right person with the right quality is essential, and that quality needs to be maintained over time.

### **Conclusion: Winning the race**

Automotive manufacturers across the globe are racing to produce autonomous vehicles that can be driven safely and comply with laws and regulations in all regions. The winners in this race will be companies that can combine technical innovation with regulatory compliance and personal data protection.

Still, ensuring privacy is not just a matter of meeting government regulations. All enterprises, including autonomous vehicle manufacturers, have a social responsibility to their customers and society. Employing responsible and effective privacy measures also can become a key part of company branding and contribute to maintaining a strong, customer-friendly position in the marketplace.

### **How DXC Technology can help**

In the complex world of autonomous driving, successful data collection and analysis requires the ability to orchestrate massive amounts of data. This means having the scalability, infrastructure, software, storage and flexibility to work with numerous providers to turn raw data into usable insights that will improve the safety and performance of autonomous vehicles.

DXC has the technical know-how and state-of-the-art integration capabilities in a highly scalable environment to deliver world-class autonomous driving solutions to automotive manufacturers worldwide. Successfully safeguarding personal data is an important ingredient to meeting strict regulations and integral to doing what's right for consumers. DXC is a reliable, go-to partner when it comes to working out well-planned, end-to-end use cases with large amounts of data in a secure environment. Our ability to integrate leading anonymization technology into the process is a key differentiator and helps make DXC a leading IT services provider to the automotive industry and beyond.

## About the authors

**Philipp Wende** is the innovation program lead for STARTUP AUTOBAHN and a senior consultant at DXC Technology, involved in developing DXC' s go-to-market strategy for the global automotive and mobility sector.

**Sunil Samantaray** is the chief technologist for DXC Robotic Drive engineering, leading a global team of technical architects and engineers developing our state-of-the-art Analytics & AI platform and valuable solution accelerators for autonomous driving.

Learn more at  
[dxc.com/auto](https://dxc.com/auto) and  
[luxoft.com/automotive](https://luxoft.com/automotive)

Get the insights  
that matter.

[dxc.com/optin](https://dxc.com/optin)

