



ResearchInChina
www.researchinchina.com

China Autonomous Driving Data Closed Loop Research Report, 2022

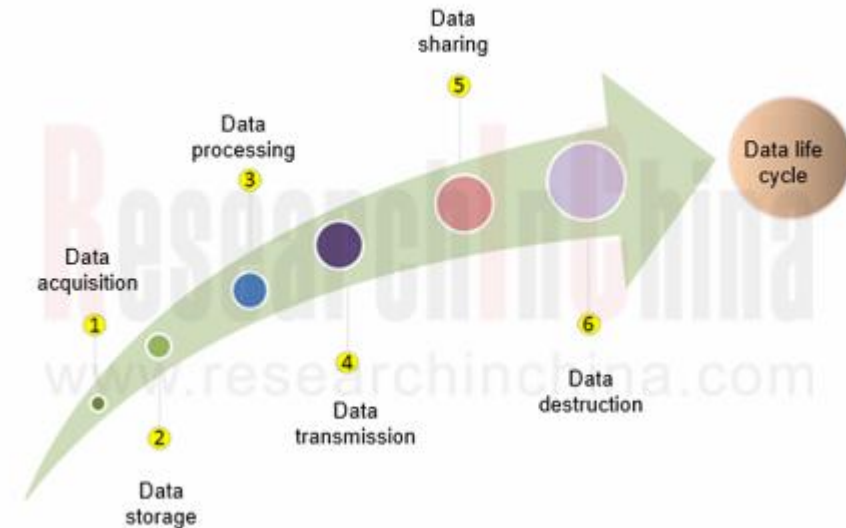
Aug. 2022

1. The development of autonomous driving is gradually driven by data rather than technology

Today, autonomous driving sensor solutions and computing platforms have become increasingly homogeneous, and the technology gap between suppliers is narrowing. In the past two years, the iteration of autonomous driving technology has advanced rapidly, and mass production has accelerated. According to ResearchInChina, a total of 4.79 million passenger cars with L2 assisted driving were insured in China in 2021, a year-on-year increase of 58.0%. From January to June 2022, the penetration rate of L2 assisted driving in the Chinese new passenger car market climbed to 32.4%.

For autonomous driving, data runs through the entire life cycle ranging from R&D, testing, mass production, operation to maintenance. As the number of sensors in intelligent connected vehicles swells, the amount of data generated by ADAS and autonomous vehicles is growing exponentially, from gigabytes to terabytes, petabytes, exabytes, and even zettabytes in the future. The evolution of data-driven vehicles can meet the personalized demand of users, and facilitate the long-term development of automakers.

According to "Safety Guidelines for Processing of Data Collected by Automobiles", the data collected by automobiles refer to the data collected by automotive sensors and control units, as well as additional data generated after aforementioned data are processed, including out-of-vehicle data, cockpit data, operation data, position data, trajectory data, etc..



The "Several Provisions on Management of Automobile Data Security (Draft)" issued by Cyberspace Administration of China in August 2021 details regulations for collection, analysis, storage, transmission, query, application, deletion, etc. of automobile data. It requires that automobile data processing should adhere to the principles of "in-vehicle processing", "data should be not collected by default", "applicable accuracy range ", "desensitization processing" and so on, so as to reduce the disorderly collection and illegal abuse of automobile data. During the development of autonomous driving technology, data collection and processing must be legal and compliant.

Data Collection/Cleaning

The massive unstructured data (images, video, speech) collected by automotive cameras, radar, LiDAR, and ultrasonic radar can be raw and messy. To make them meaningful, they should be cleaned, structured, and organized. At first, the data from multiple sources should be imported into appropriate repositories with their formats being standardized and they should be aggregated according to relevant rules. Then, checks should be made to detect corrupt, duplicated, or missing data points, and the data that might affect the overall quality of the dataset should be discarded. Finally, labels should be used to classify videos captured under different conditions, such as daytime, night, sunny day, rain, etc. This step provides the cleaned structured data that will be used for training and validation.



Data annotation

The structured data that are cleaned after data collection should be labeled. Labeling is the process of assigning encoded values to raw data. Encoded values include, but are not limited to, assigning class labels, drawing bounding boxes, and marking object boundaries. High-quality annotation is needed to teach supervised learning models what objects are and to measure the performance of trained models.

In the field of autonomous driving, data annotation usually covers scenarios where vehicles are changing lanes to overtake, passing through intersections, turning left or right without traffic light control, running red lights and parking on roadsides illegally, pedestrians are jaywalking, etc.

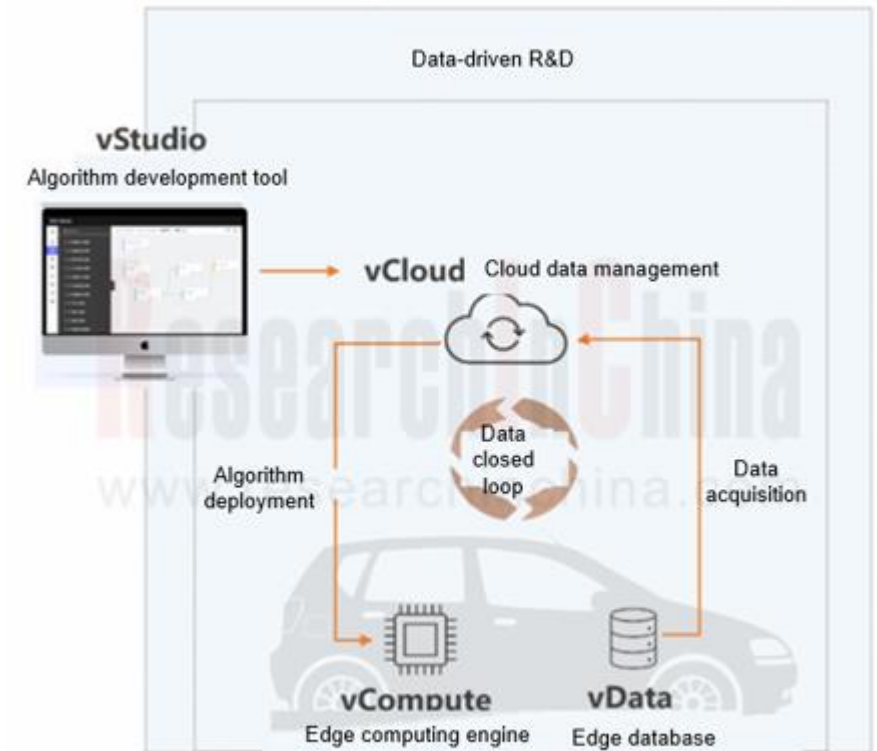
Popular annotation tools are involved with general picture frames, lane line annotation, driver face annotation, 3D point cloud annotation, 2D/3D fusion annotation, panoramic semantic segmentation, etc. Prompted by development of big data and the spike in the number of large datasets, data annotation tools are used more and more widely.

Data transmission

Nowadays, data collection occurs every few milliseconds, requiring high-precision data in thousands of signal dimensions (such as bus signals, the internal state of sensors, software embedment, user behaviors, and environmental perception data, etc.). At the same time, in order to avoid data loss, disorder, hopping and delay, the transmission/storage cost is greatly reduced under the premise of high precision and high quality. The long uplink and downlink (from automotive MCU, DCU, gateways, 4G/5G to the cloud) of IoV data require the data transmission quality of each link node.

In response to new changes in data transmission, some companies have been able to provide efficient data acquisition and vehicle-cloud integrated transmission solutions. For example, EXCEEDDATA's flexible data acquisition platform solution implements 10-millisecond real-time operations based on real-time data in the automotive computing environment to trigger flexible data collection and upload. After being calculated and filtered, the amount of uploaded data is significantly reduced. In addition, 100-300 times lossless compression and storage of the original signals at the vehicle is performed. The cloud management platform saves lossless high-quality signals of the vehicle with a high compression ratio, supports the issuance of data acquisition algorithms, the triggering of multiple acquisition modes, and the one-click download of acquired data uploaded to the business desktop in real time. The data can be flexibly filtered by vehicle, event, time, etc., and the storage and calculation are separated, realizing the closed loop of collection-calculation-upload-processing of vehicle-cloud isomorphic data. In 2021, HiPhiX became China's first production model equipped with EXCEEDDATA's solution.

Architecture of EXCEEDDATA's Solution

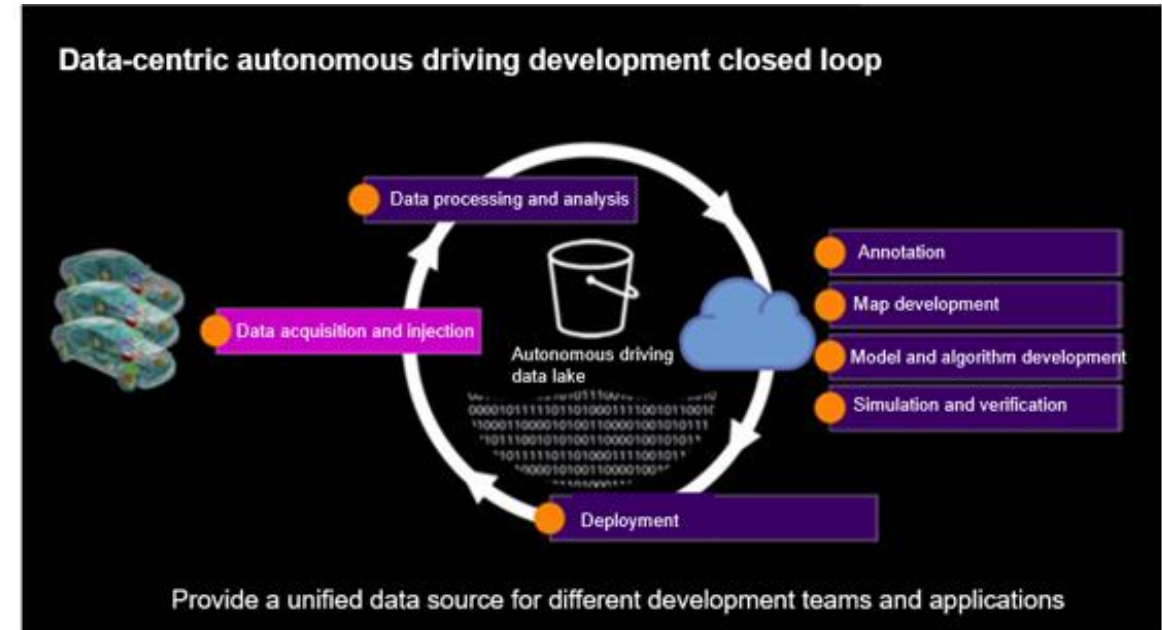


Source: EXCEEDDATA

Data storage

In order to perceive the surrounding environment more clearly, autonomous vehicles carry more sensors and generate massive data. Some high-level autonomous driving systems are even equipped with more than 40 assorted sensors to accurately perceive 360° environment around vehicles. The R&D of autonomous driving systems has to go through multiple links such as data collection, data aggregation, cleaning and marking, model training, simulation, big data analysis, etc.. It involves the aggregation and storage of massive data, the data flow between different systems of different links, and reading and writing of massive data during model training. Data see new challenges from storage bottlenecks.

In this regard, the technology and capabilities of many cloud service providers have become the key to automakers. For example, Amazon Web Services (AWS) offers cloud computing services. AWS is centered on the autonomous driving data lake, helping automakers build an end-to-end autonomous driving data closed loop. Automakers can exploit Amazon Simple Storage Service (Amazon S3) to build an autonomous driving data lake so as to realize data collection, data management and analysis, data annotation, model and algorithm development, simulation verification, map development, DevOps and MLOps, as well as to conduct development, testing and application of autonomous driving easily.

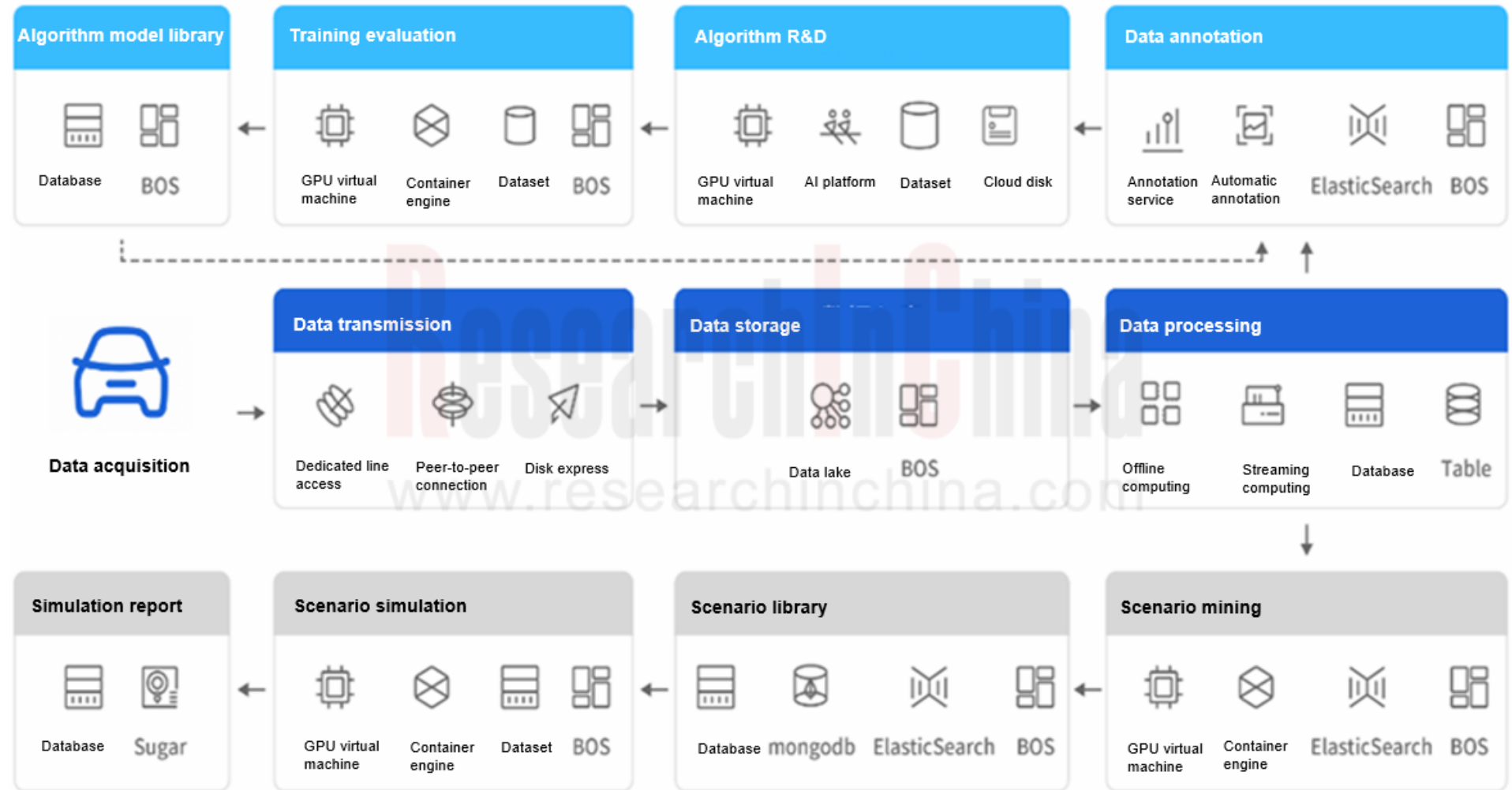


Source: AWS

Baidu Data Closed Loop Solution

For example, Baidu's data closed loop solution provides data retrieval services for multi-source data information of roadsides and vehicles, which are used for massive data search on business platforms, with advantages like multi-dimensional retrieval (vehicle information, mileage, autonomous driving duration, etc.), management of the entire life cycle from data production to destruction, support for panoramic data views, data traceability, data openness and sharing.

Architecture of Baidu's Autonomous Driving Data Closed Loop Solution



2. The efficient development of autonomous driving requires construction of a data closed loop system

The development of autonomous driving is gradually driven by data rather than technology. However, data-driven business models have many difficulties.

Difficult massive data processing: High-level autonomous driving test vehicles collect terabytes of data every day, so development teams need petabytes of storage space. However, less than 5% of the data are available for training as value data. In addition, there are strict security compliance requirements for the data collected by sensors such as automotive cameras, LiDAR, and high-precision sensors, which undoubtedly brings great challenges to the access, storage, desensitization, and processing of massive data.

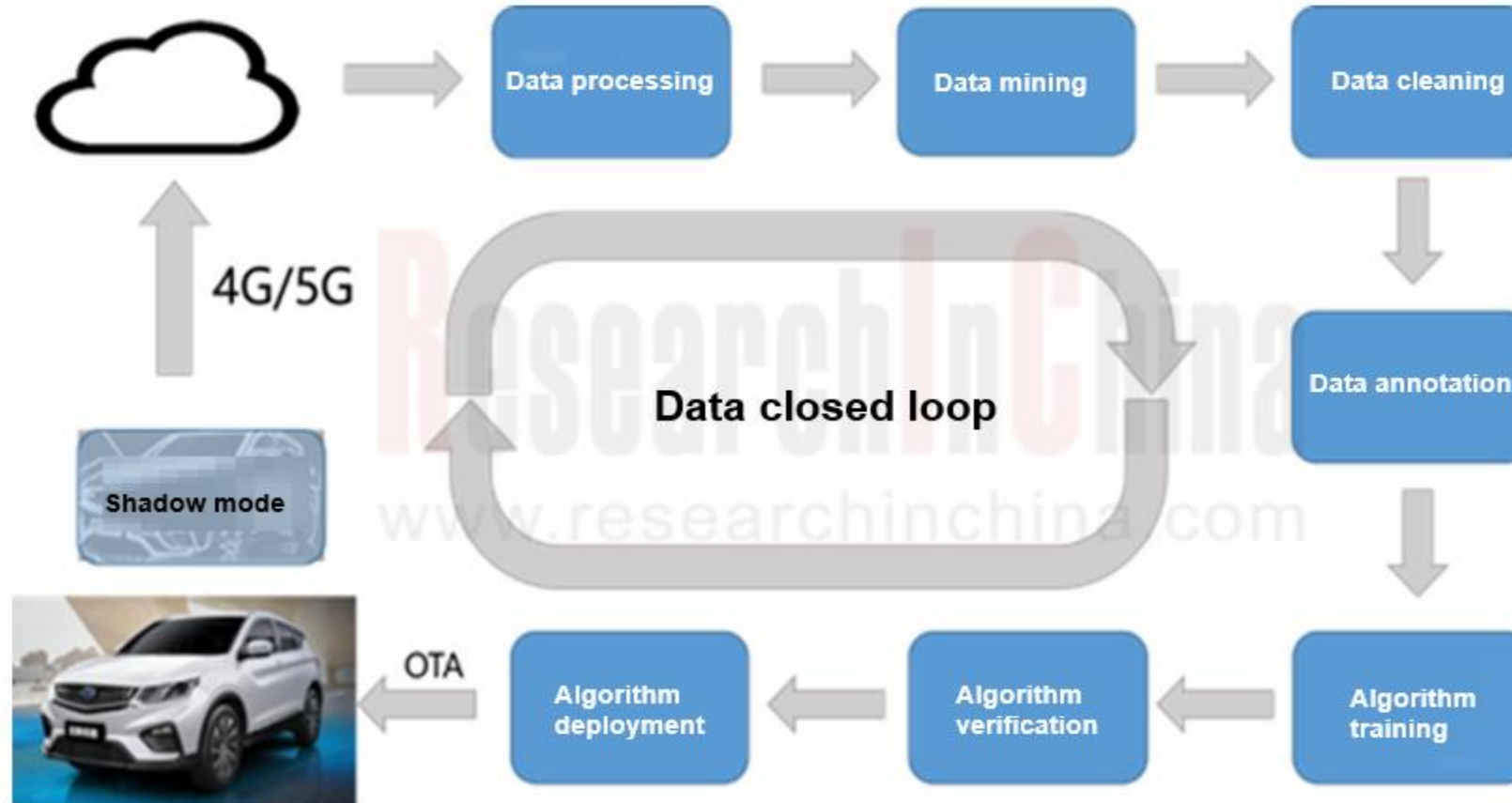
High data annotation cost: Data annotation costs a lot of labor and time. With the development of advanced capabilities of autonomous driving, scenarios are becoming more and more complex, and difficult scenarios will happen. Improving the accuracy of vehicle perception models places higher requirements on the scale and quality of training datasets. In terms of efficiency and cost, traditional manual annotation has been unable to meet the demand of model training for massive datasets.

Low simulation test efficiency: Virtual simulation is an effective means to accelerate the training of autonomous driving algorithms, but simulation scenarios, especially complex and dangerous scenarios, are difficult to construct and embody a low degree of restoration. Plus the insufficient parallel simulation capability, the efficiency of simulation tests is low, and the iteration cycle of algorithms is too long.

Less coverage of HD maps: HD maps mainly rely on self-collection and self-made mapping, and only cover designated roads in the experimental stage. In the future, commercial HD maps will face prominent challenges in coverage, dynamic update, cost and efficiency when spreading to urban streets in major cities across the country.

2. The efficient development of autonomous driving requires construction of a data closed loop system

In order to solve difficulties and problems, the efficient development of autonomous driving requires the construction of an efficient data closed loop system.



Source: Freetech

2. The efficient development of autonomous driving requires construction of a data closed loop system

As far as the closed loop of autonomous driving data is concerned, Corner Cases should be solved in the process of autonomous driving. To this end, there must be enough data samples and convenient automotive verification methods. Shadow mode is one of the best solutions for Corner Cases.

Shadow mode was proposed by Tesla in April 2019 and applied to vehicles so as to compare relevant decisions and trigger data upload. It uses autonomous driving software on the sold vehicles to continuously record data detected by sensors, and selectively sends back autopilot algorithm for machine learning and refinement at the appropriate time.

In 2021, Tesla delivered 936,200 vehicles globally, of which 484,100 ones came from Chinese factory. Tesla delivered 560,000 units in 2022H1. Tesla takes advantage of mass production to continuously optimize its algorithm through shadow mode. Tesla leverages shadow mode to take millions of sold vehicles as test vehicles to perceive the surrounding environment and capture special road conditions, thereby continuously strengthening the capability to predict, avoid, and learn from uncertain events. Thanks to millions of sold vehicles, more Corner Cases and extreme working conditions will be covered. The high-quality data collected by flexible triggering can iterate better algorithms which determines the value of software. In terms of software update subscription services, the energy of data closed loop has just emerged.



The goal of Dojo will be to be able to take in vast amounts of data and train at a video level and do unsupervised massive training of vast amounts of video with the Dojo program or Dojo computer.

3. Data closed loop becomes the core of iterative upgrade of autonomous driving

The premise of continuous iteration of automatic driving systems lies in constant optimization of algorithms which hinges on the efficiency of data closed loop systems. The efficient flow of data in each scenario of autonomous driving development is crucial, and data intelligence will become the key to accelerating mass production of autonomous vehicles.

In December 2021, Haomo.AI officially released MANA (Snow Lake), the first autonomous driving data intelligence system in China, to accelerate evolution of autonomous driving technology from the perspectives of perception, cognition, annotation, simulation and calculation. In the next three years, the assisted driving system of Haomo.AI will land on more than 1 million passenger cars. By virtue of its fully self-developed autonomous driving system, Haomo.AI has achieved remarkable advantages in data accumulation, processing and application. Massive data brings about technological iterative advantages, like obvious cost reduction and efficiency improvement.

Momenta has acquired leading full-process data-driven technology. Algorithmic modules about perception, fusion, prediction and regulation can be efficiently iterated and updated in a data-driven manner. Momenta's Closed Loop Automation (CLA) is a complete toolchain that lets data streams drive automatic iterations of data-driven algorithms. CLA can automatically filter out massive gold data, drive automatic iteration of algorithms, and make autonomous driving flywheel spin faster and faster.

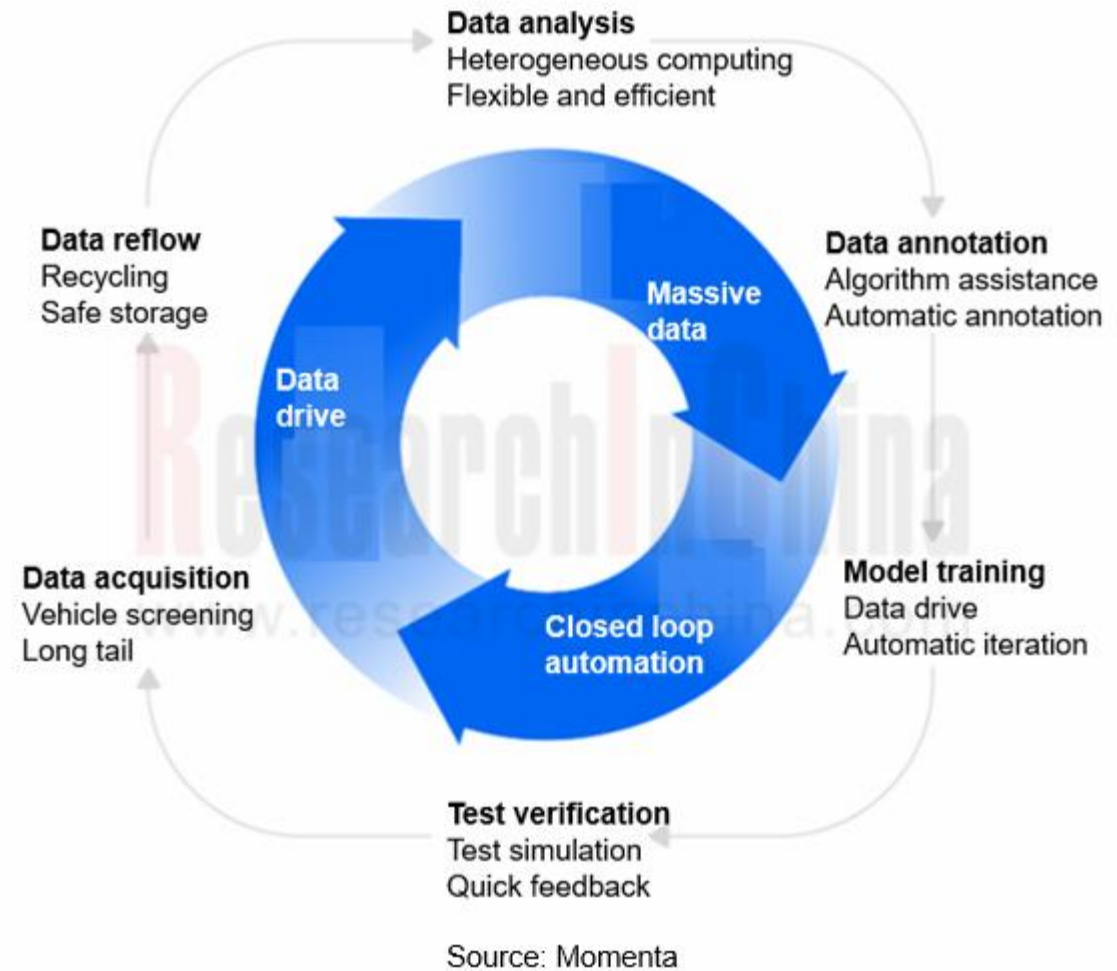


Table of Content (1)

1 Introduction to Autonomous Driving Data Industry Chain

1.1 Overview of Automotive Data and Autonomous Driving Data

1.1.1 Classification of Automotive Data

1.1.2 China's Laws and Regulations for Automotive Data Security

1.1.3 Data Volume and Computing Power Requirements by Autonomous Driving Level

1.1.4 Computing Power of Assisted Driving of Some New Vehicles on the Market

1.1.5 Basic Requirements for Data Storage of Autonomous Vehicles

1.1.6 The Efficient Development of Autonomous Driving Requires Construction of a Data Closed Loop System

1.1.7 Workflow of Conventional Data Closed Loop

1.1.8 Workflow of AI Data Closed Loop

1.2 Data Acquisition

1.2.1 Status Quo

1.2.2 Value

1.2.3 Acquisition Methods of Traditional Structured Data

1.2.4 Unstructured Data

1.2.5 Problems of Data Acquisition in Corner Cases

1.3 Data Annotation

1.3.1 Definition

1.3.2 Industry Chain and Ecology

1.3.3 Autonomous Driving Data Annotation

1.3.4 Types of Autonomous Driving Data Annotation

1.3.5 More Data Required by Model Training

1.3.6 Harder and More Demanding 3D Annotation

1.3.7 L3+ Requires Massive High-quality Data

1.4 Shadow Mode

1.4.1 Definition

1.4.2 Accumulated Mileage of Tesla Autopilot

1.4.3 Application Examples of Shadow Mode of Some Enterprises

1.5 Overview of Autonomous Driving Data Industry Chain

2 Typical Data Acquisition and Annotation Companies

2.1 Testin

2.1.1 Profile

2.1.2 Intelligent Driving Solution

2.1.3 Data Acquisition and Annotation Services

2.1.4 Storage Architecture and Data Visualization

2.1.5 Customers

2.2 MindFlow

2.2.1 Profile

2.2.2 SEED Data Service Platform

2.2.3 Data Annotation Solution

2.2.4 Autonomous Driving Data Middle Platform

2.3 Appen

2.3.1 Profile

2.3.2 Data Annotation Platform

2.3.3 Data Annotation Toolset

2.3.4 Data Quality Control

2.3.5 Data/Platform Service Deployment

Table of Content (2)

- 2.4 Graviti
 - 2.4.1 Profile
 - 2.4.2 Development History
 - 2.4.3 Data Platform
 - 2.4.4 Data Management Advantages
 - 2.4.5 Customers and Partners
- 2.5 Jinglianwen Technology
 - 2.5.1 Profile
 - 2.5.2 Intelligent Driving Data Solution
 - 2.5.3 Process of Data Acquisition and Annotation Solution
- 2.6 Speechocean
 - 2.6.1 Intelligent Driving Data Solution
 - 2.6.2 Autonomous Driving Data Annotation Technology
 - 2.6.3 Performance

3 Data Closed Loop Solution Providers

- 3.1 Kunyi Electronics
 - 3.1.1 Profile
 - 3.1.2 Events
 - 3.1.3 Introduction to Advanced Autonomous Driving Data Acquisition Solution
 - 3.1.4 Advanced Autonomous Driving Data Bypass Acquisition Solution
 - 3.1.5 Advanced Autonomous Driving Data True Value Acquisition Solution
 - 3.1.6 Acquisition Solutions of Different Types of Sensors
 - 3.1.7 Bypass Solutions of Different Types of Sensors
 - 3.1.8 Data Acquisition Solutions and Products
 - 3.1.9 Autonomous Driving Data Equipment

- 3.1.10 Data Acquisition Solution Cloud Platform
- 3.1.11 Data Visualization and Analysis Software
- 3.1.12 Data Synchronization Accuracy in Data Acquisition Solution
- 3.1.13 Data Re-injection Equipment: For Corner Case Reproduction
- 3.1.14 Data Re-injection Equipment: For Clustered Regression Testing
- 3.1.15 Data Re-injection Equipment: For Simulating Closed Loop Testing
- 3.1.16 Overview of Smart Driving Re-injection
- 3.1.17 Smart Driving Re-injection Function - Video Injection
- 3.2 EXCEEDDATA
 - 3.2.1 Profile
 - 3.2.2 Full Life Cycle Business of Smart Vehicles
 - 3.2.3 Vehicle Cloud Computing Product Composition
 - 3.2.4 Vehicle Cloud Computing Solution
 - 3.2.5 Data Closed Loop Solution
 - 3.2.6 Converged Acquisition of Structured and Unstructured Data
 - 3.2.7 Data Acquisition Solution
 - 3.2.8 EXD Mass-produced Intelligent Driving Flexible Data Acquisition Solution
 - 3.2.9 EXD Autonomous Driving Data Acquisition Flexibly Triggers Scenarios
 - 3.2.10 Cross-domain Multi-scenario Solution
 - 3.2.11 Overall Solution Architecture of EXD Data Analysis Platform
 - 3.2.12 Customers (Automakers) and Ecological Cooperation
- 3.3 Baidu
 - 3.3.1 Product Platform and Data Platform of Autonomous Driving
 - 3.3.2 Architecture of Data Closed Loop Solution
 - 3.3.3 Data Closed Loop Solution - Data Acquisition
 - 3.3.4 Data Closed Loop Solution - Data Processing
 - 3.3.5 Data Closed Loop Solution - Data Upload

Table of Content (3)

- 3.3.6 Data Closed Loop Solution - Data Storage
- 3.3.7 Data Acquisition and Annotation Solution
- 3.3.8 Autonomous Driving: Cloud Simulation Testing Solution
- 3.3.9 Advantages of Cloud Simulation Testing Solution
- 3.3.10 Cloud Simulation Testing Solution - Artificial Design Simulation
- 3.3.11 Cloud Simulation Testing Solution - Others
- 3.3.12 Autonomous Driving Simulation Toolchain - Mass Production Cooperation

3.4 VNET

- 3.4.1 Profile
- 3.4.2 Development History
- 3.4.3 Resource Distribution of Data center nationwide
- 3.4.4 Autonomous Driving Solution
- 3.4.5 Autonomous Driving Solution: Acquisition (1)
- 3.4.6 Autonomous Driving Solution: Acquisition (2)
- 3.4.7 Autonomous Driving Solution: Annotation
- 3.4.8 Autonomous Driving Solution: Training / Simulation
- 3.4.9 Performance
- 3.4.10 Customers

3.5 Momena

- 3.5.1 Profile
- 3.5.2 Three-stage Construction
- 3.5.3 Data Closed Loop Automation
- 3.5.4 Core Technology
- 3.5.5 Autonomous Driving Solution
- 3.5.6 Data Closed Loop Application Case

3.6 CalmCar

- 3.6.1 Profile
- 3.6.2 Data Acquisition
- 3.6.3 AV Data Recording System
- 3.6.4 Self-developed Toolchain

3.7 Molar Intelligence

- 3.7.1 Profile
- 3.7.2 New Data Annotation Platform
- 3.7.3 Standardized Data Production Process

3.8 SandStone

- 3.8.1 Profile
- 3.8.2 Storage Solution

3.9 Amazon

- 3.9.1 AWS for Automotive
- 3.9.2 SageMaker
- 3.9.3 Features of SageMaker
- 3.9.4 Data Annotation of SageMaker
- 3.9.5 EMR Big Data Cloud Platform

4 Data Closed Loop Layout of Main Tier1/Tier2 Suppliers

4.1 Hong Jing Drive

- 4.1.1 Profile
- 4.1.2 Data Cloud Platform

4.2 Pony.ai

- 4.2.1 Profile

Table of Content (4)

- 4.2.2 Autonomous Driving Infrastructure Platform
- 4.2.3 Features of Self-developed Full-stack Data Closed Loop Toolchain
- 4.3 Freetech
 - 4.3.1 Profile
 - 4.3.2 End-to-end Full-stack Process
 - 4.3.3 Data Closed Loop Solution
- 4.4 NavInfo
 - 4.4.1 Profile
 - 4.4.2 "Chip + Data + Algorithm Closed Loop for HD Map"
- 4.5 Idriverplus
 - 4.5.1 Profile
 - 4.5.2 Development History and Financing
 - 4.5.3 Data-driven Autonomous Driving Mass Production Solution
 - 4.5.4 AVDC Data Closed Loop System
- 4.6 iMotion
 - 4.6.1 Profile
 - 4.6.2 Full-scenario Algorithm Capability
 - 4.6.3 Big Data Closed Loop System
- 4.7 Haomo.AI
 - 4.7.1 MANA
 - 4.7.2 Features of MANA
 - 4.7.3 Data Closed Loop
- 4.8 UISEE
 - 4.8.1 U-Drive Intelligent Driving System
 - 4.8.2 Cloud Operation Management Platform
 - 4.8.3 Shadow Mode
- 4.9 Other Tier1/Tier2 Suppliers
 - 4.9.1 MINIEYE

- 4.9.2 Heading Data
- 4.9.3 Leadgentech
- 4.9.4 HoloMatic Technology
- 4.9.5 Joyson Electronics
- 4.9.6 Neusoft Reach
- 4.9.7 MOGO
- 4.9.8 juefx.com
- 4.9.9 QCraft
- 4.9.10 AutoX

5 Data Closed Loop Layout of Other Companies

- 5.1 Chip Vendors
 - 5.1.1 Data Closed Loop Development Platform of Horizon Robotics
 - 5.1.2 Data Closed Loop Solution of Black Sesame Technologies
 - 5.1.3 NVIDIA's Machine Learning Platform: MAGLEV
- 5.2 Tesla
 - 5.2.1 Data Engine System of Autopilot Model Iteration
 - 5.2.2 Data Engine
 - 5.2.3 Dojo Supercomputer
- 5.3 DeepWay
 - 5.3.1 Data Closed Loop Route
 - 5.3.2 Data and Scenario Library
 - 5.3.3 Algorithm Training and Simulation



Beijing Headquarters

TEL: 010-82601561, 82863481

Mobile: 137 1884 5418

Email: report@researchinchina.com

Website:
www.researchinchina.com

WeChat: [zuosiqiche](#)



Chengdu Branch

TEL: 028-68738514

FAX: 028-86930659