Springer Vieweg

ATZ extra

DATA DRIVEN DEVELOPMENT

Used for Recognition of Emotions and Fatigue

© shutterstock | temp-64GTX | depositphotos | kerenby

Data-driven Development in the Automotive Industry

WRITTEN BY

Marcus Weber
is Group Leader Data Driven Software and Sensors at ITK Engineering in Rülzheim (Germany).

Stefan Held
is Lead Engineer at ITK Engineering in Holzkirchen (Germany).

Data plays a key role in a functioning mobility ecosystem. It is used in the ongoing operation of vehicles, as well as during development. The basis for this is data-driven development methods. ITK Engineering shows how they work using the example of emotion and fatigue detection, which will be mandatory from 2024.

More and more organizations are discovering the data they generate and own can be a tremendous business asset [1]. The automotive industry has certainly come to this realization in recent years. Companies expect that the effort they invest in exploiting this data will pay off by creating added value. However, while businesses develop long-term strategies to build and grow conventional assets, the data economy is more about the short-term pursuit of low-hanging fruit. This is why organizations launch application-focused data initiatives with isolated data sinks rather than building scalable data architectures to use throughout the enterprise. And rather than developing new job profiles, they simply assign additional tasks and responsibilities to existing roles and departments. A great deal of data utilization potential goes untapped because of this failure to integrate initiatives into the overall organization.

These pitfalls can be sidestepped early on with a comprehensive data strategy. Furnishing a framework in which data initiatives can be embedded, a data strategy enables organizations to extract maximum value from data. By taking an agile approach to strategy development, enterprises can make course corrections and mitigate business risk. Companies are also well-advised to factor data into the decision-making process at an early stage. **FIGURE 1** outlines an example of a method for quickly assessing innovative data-based ideas and the data pool's potential. These insights are the foundation of a long-term data strategy.

This method also serves other purposes, for example, to examine the use case for drowsiness and attention warning from several angles – through the lens of technology, from a business perspective, and through users' eyes. Available data can be explored to assess the feasibility of various options such as the use of camera data, the steering angle, or other sensors to detect vital functions. The results obtained with this method provide the insights needed to determine what it will take to put a holistic data strategy into action. This includes technical modifications and data governance measures. The latter are becoming increasingly important in the wake of regulatory requirements such as the EU's Data Act and the AI Act [2, 3].

DATA QUALITY AS PREREQUISITE FOR PRECISE ANALYSES

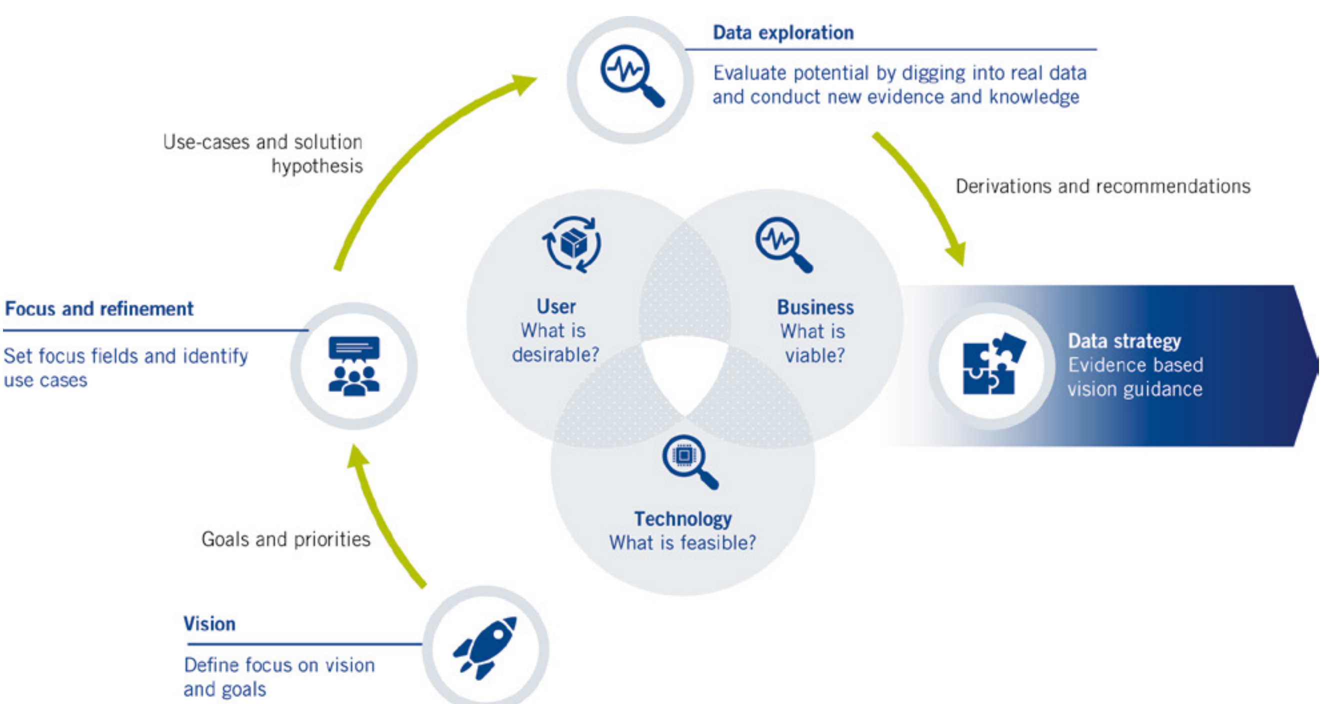
Data quality is crucial to using data successfully and achieving accurate, reliable results. The key criterion for data quality is its suitability for the task at hand. Several factors have to be considered when setting out to measure and improve data quality [4]:

- Data has to meet the development project's specific requirements and serve its objectives. It has to be representative of all use cases. For this use case, the data would have to cover accessories such as glasses and headgear as well as all variables in physical appearance such as height, head shape, skin tone, hair, and so on.
- Evidence needs to be presented to substantiate that data errors are minimal. Data corruption during storage and transmission have to be prevented.
- Accuracy is another critical aspect of data quality that merits consideration, especially when labeling for machine learning (ML) development.
- Completeness is crucial. Requirements for the minimum viable dataset have to be determined to ensure all relevant elements of the input domain are properly represented. Data must correlate with the variables in the operational design domain [5, 6] as described by scenarios. The variables in this use case would include different cameras and their positions, the interior design as it relates to seat position and lighting, and the occupants.
- Traceability is vital to assessing the data's reliability and credibility.
- Good data governance practices are essential to obtaining high-quality data. Robust data governance entails setting data standards, implementing data validation methods, and enforcing quality assurance processes.

SELECTING AND MOVING OF DATA TO THE CLOUD

Another key aspect of data-driven development is the selection and transfer of data from the vehicle to the cloud, where

FIGURE 1 A comprehensive data strategy provides the underpinning for companies seeking to extract value from the data they have available (© ITK Engineering)



it can be processed, analyzed, and exploited to develop various automotive functions. Take, for example, ML systems: They could be trained using legacy datasets, often require transfer learning data to adapt to specific use cases. Drowsiness and attention warning is a case in point: It involves specific positions and specific physical properties of cameras, as well as the vehicles' specific spatial and lighting conditions. Choosing the right data collection methods is an important step. Undirected, continuous collection generates vast amounts of data that demand a lot of labeling work. However, data campaigns can be defined with fixed boundaries to enable directed data collection within preset parameters. This can be done with external triggers to start and stop data collection and with specific analytics such as comparisons with specific scenarios. For fatigue and emotion detection, for example wide variances in key points for face tracking can be used to indicate data collection issues.

The benefits of transferring data to the cloud are many, particularly in terms of data governance. A central gate in the cloud can provide security. Uploading also enables automated data validation and the creation of specific data pipelines for pre-processing, storage, and visualization in dashboards. Data is instantly available which speeds up workflows and enables real-time analytics and live presentation. There are several prerequisites for moving data from the vehicle to the cloud: Access to the cloud endpoint has to be secured, for example, by authenticating with a bearer token. Data can be buffered in the vehicle and uploaded in batches. This upload can be mobile or stationary depending on the amount of data. Mobile transmission of key points for face tracking [7] is possible because there is not that much data to convey. Camera data is usually cached and transmitted at special download stations. Collected data can be stored in an event store such as Kafka for persistent storage and post-processing.

CENTRAL DATA ACCESS FOR DISTRIBUTED DATABASES

Data-driven development requires an ecosystem that enables developers to intuitively integrate data into their development activities. This requires a low threshold for data access and the assurance that the data's content is reliable and traceable. The reality in projects is that data is usually stored in distributed, heterogeneous silos. Access is usually not standardized; the necessary authorizations are often lacking. In most cases the access problem is down to complex and opaque approval workflows. On a related note, complex formats and poor quality often impede data analysis.

One solution to these problems is a digital twin that forms an abstraction layer between data consumers and data sinks. A knowledge graph represents the actual vehicle and its subsystems. With the benefit of this interface, it is possible to gain read and write access to data in the backend without necessarily knowing the specifics of this data. Data verification and access authorizations can also be mapped to this interface, **FIGURE 2**. In this example, developers can obtain the data needed to train the model via the digital twin. This way, data can be made available to a wide range of users at short notice and without requiring extensive modifications to the underlying data architecture and infrastructure. This breaks down data silos, creates added value, and enables developers to come up with new services. Combining this approach with a large language model is likely to further facilitate data access, thereby driving data democratization [8].

FROM PROTOTYPING TO SERIES DEVELOPMENT

The term coined to describe the automation of ML models throughout the lifecycle – that is, from integration to deployment and monitoring – is machine learning operations, or MLOps for short [9, 10]. When developing an ML system, it is advisable to implement a seamless data processing chain from the source to the point of data usage – that is, an end-to-end (E2E) pipeline. This ought to be done at an early turn to ensure connected systems' compatibility. It is also a good idea to label data and improve its quality while this pipeline is in development. Developing ML systems is an iterative process, the subordinate steps of which are model creation, testing, and refinement. Once the E2E pipeline is ready to go, the next step is to improve the ML system's accuracy and integrate the methods of accountable, explainable AI (XAI) so humans can comprehend ML systems' reasoning, to build trust in these systems,

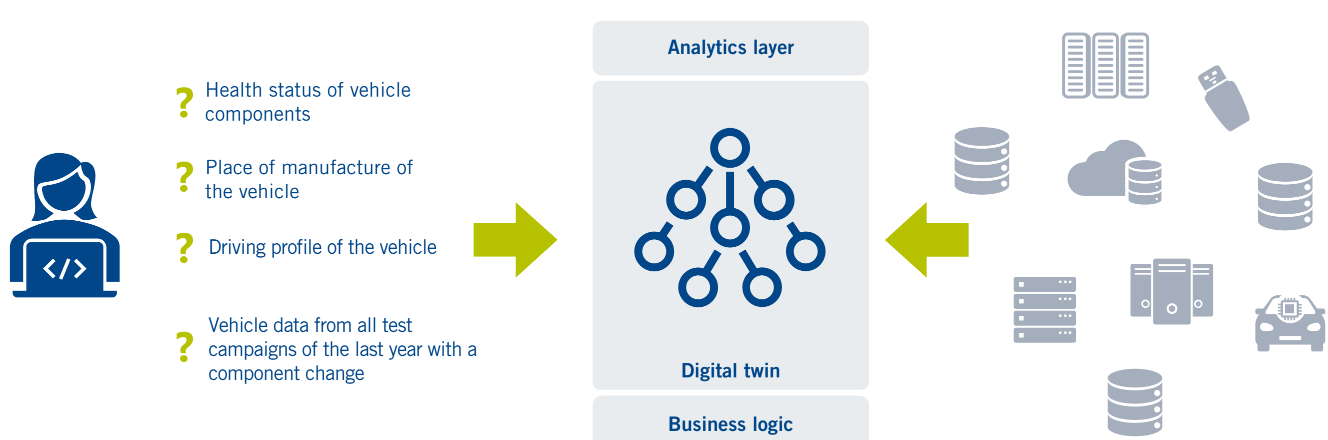


FIGURE 2 A digital twin can enable companies to make the most of their legacy data (© ITK Engineering)

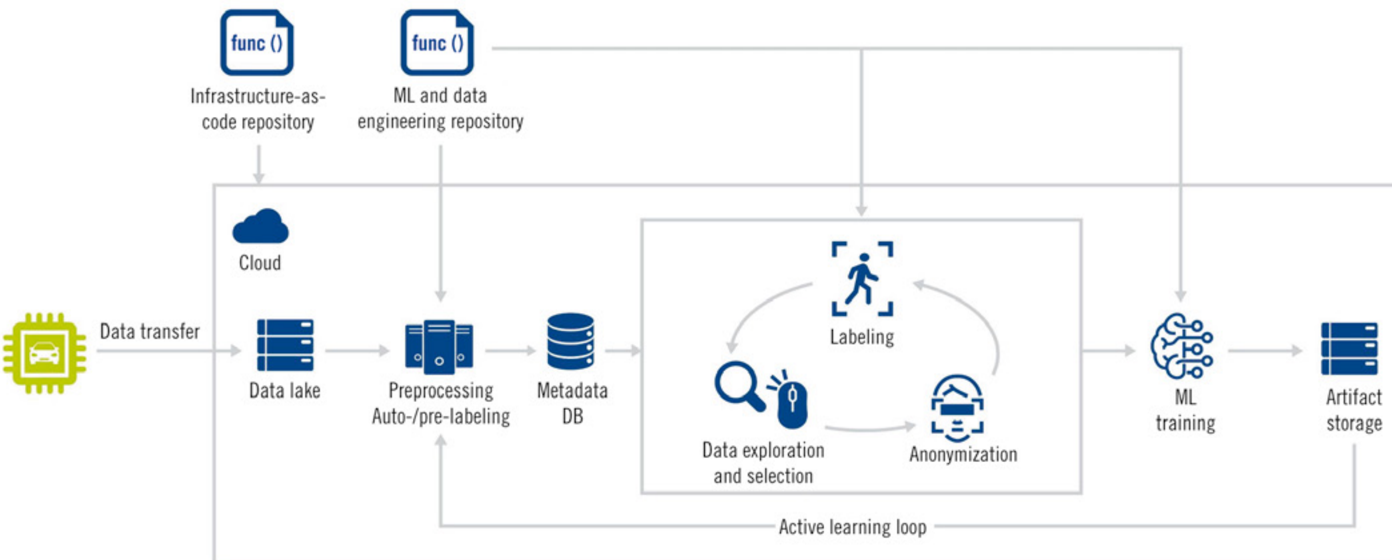


FIGURE 3 A seamless MLOps cycle is the prerequisite for the ML model's successful integration into the working system (© ITK Engineering)

and to pinpoint and correct biases or errors in the decision-making process. Heat maps are an example for visualization of explanations for fatigue detection and emotion recognition.

Scalability is a critical factor in data-driven development. A system can be scaled vertically to increase the capacity of a single resource or it can be scaled horizontally by adding machines or nodes. Scalability can bring down the cost of deploying and maintaining the system over time and make it easier to train more complex systems. To this end, the iterative ML development process is continuously monitored by tracking the ML system's performance metrics. This way, the system will continue to function optimally even if data and requirements change, a phenomenon known as data and concept drift.

SUMMARY

Data-driven development requires a suitable data strategy to ensure developers pursue the right goals, take appropriate measures, and make the most of resources. Then the available data can be tapped to sustainably create added value. This data has to have the requisite level of quality. Developers also need an ecosystem that enables them to intuitively integrate available data. A digital twin brings data use barriers down and creates a virtual model of the vehicle and its subsystems. The right project strategy and, particularly for ML development projects, a supporting MLOps pipeline are also success factors. In conclusion, data-driven development in the automotive industry requires a combination of technical skills, experience, and passion.

REFERENCES

[1] Lempp, M.; Siegfried, P. (ed.): Automotive disruption and the Urban Mobility Revolution. Cham: Springer, 2022, pp. 35 - 37

[2] Europäische Kommission (Hg.): Gestaltung der digitalen Zukunft Europas – Datengesetz. Online: https://digital-strategy.ec.europa.eu/de/policies/data-act, access: June 30, 2023

[3] Future of Life Institute (Hg): The AI Act. Online: https://artificialintelligenceact.eu/, access: June 30, 2023

[4] Safety-Critical Systems Club (Hg.): SCSC Publications. Working Group Documents. Online: https://scsc.uk/scsc-127H?tab=wgdoc, access: June 30, 2023

[5] Pegasus (Hg.): PEGASUS Sicherheitsargumentation. Online: https://www.pegasusprojekt.de/files/tmpl/pdf/PEGASUS%20Sicherheitsargumentation.pdf, access: June 30, 2023

[6] Blank, F.; Hüger, F.; Mock, M.; Stauner, T.: Assurance Methodology for In-vehicle AI. In: ATZ Worldwide 124 (2022), No. 7-8, pp. 54 - 59

[7] ITK Engineering (Hg.): Individual Virtual Environment & Sensor Simulation: Driver/Occupant Monitoring Simulation. Online: https://www.youtube.com/watch?v=Un8nOLCmlUQ, access: June 30, 2023

[8] Patil, D.; Mason, H. (ed.): Data Driven. O.O: O'Reilly Media, 2015

[9] Luber, S.: Was ist MLOps?. Online: https://www.bigdata-insider.de/was-ist-mlops-a-1116284/, access: June 30, 2023

[10] Nayak, Y.: A Gentle Introduction to Mlops. Online: https://towardsdatascience.com/a-gentle-introduction-to-mlops-7d64a3e890ff, access: June 30, 2023

IMPRINT
Special Edition 2023 in cooperation with ITK Engineering GmbH, Bergfeldstraße 2, 83607 Holzkirchen; Springer Fachmedien Wiesbaden GmbH, Postfach 1546, 65173 Wiesbaden, Amtsgericht Wiesbaden, HRB 9754, USt-IdNr. DE81148419

MANAGING DIRECTORS: Stefanie Burgmaier | Andreas Funk | Joachim Krieger
PROJECT MANAGEMENT: Anja Trabusch

COVER PHOTO: © Limitless Visions | stock.adobe.com

ITK Engineering

Stability, reliability and methodological expertise – this is what we have stood for since our founding in 1994. At all times, our customers have benefitted from our dedicated multi-industry know-how, especially in the fields of control systems design and model-based design. Customers can count on us – from conception through to deployment, we cover the entire development process.

Our areas of expertise include:

- Software development
- Hardware development
- Electrical & electronic systems
- System integration
- Software as a product
- Turnkey systems
- Customer specific development
- Technical consulting
- Seminars
- Quality assurance

The satisfaction of each of our partners and mutually respectful cooperation shape our corporate philosophy, in which four values are firmly anchored: Read more about this on the web.

V.1.0.0_e_2021

ITK Engineering GmbH
Headquarters: Ruelzheim
Im Speyerer Tal 6
76761 Ruelzheim, Germany
T: + 49 (0)7272 7703-0
F: + 49 (0)7272 7703-100
info@itk-engineering.com

www.itk-engineering.com
www.itk-career.com

Founded in 1994
Branch offices throughout Germany – ITK companies worldwide.

Follow us on: