Katrin Sjoberg
Senior Editor

Artificial Intelligence and Vehicles

Introduction

Artificial intelligence (AI) will change society in so many different aspects that it is not comprehensible to grasp yet. It will be on the same magnitude as the Industrial Revolution (or most likely much larger). AI carries threats and great opportunities. Given the unknown unknowns of AI, we can spot headlines every day with either alarming news or the opposite. AI plays a major role for enabling autonomous vehicles but also for the development of more sophisticated and efficient advanced driver assistance systems (ADASs) by supporting, for example, rapid object detection and classification. Since the mid-20th century, AI has been a research topic, but not until recently has it gained momentum due to, for example, the performance increase in ubiquitous hardware. It is computationally expensive to execute AI algorithms. There are broadly speaking three "types" of AI discussed today, machine learning (ML), deep learning, and generative AI (GenAI), where GenAI is a specialized case of deep learning, which in turn is a subset of ML. The common denominator for all AI is training data, and data fit for purpose are pivotal for the development of AI algorithms. Garbage in will result in garbage out.

Digital Object Identifier 10.1109/MVT.2024.3355609
Date of current version 19 March 2024

ML algorithms are used for making predictions and decisions when going live. They marry computers science and statistics together. The training of the ML algorithm can be conducted in supervised, unsupervised, and reinforced modes. In supervised training, the algorithm learns from known input and output variables, as opposed to in unsupervised training, where the ML algorithm needs to figure out what output to produce given a defined input. In reinforced training, the ML algorithm is provided feedback with carrots and sticks.

Deep learning is about mimicking how the brain's neural network functions. Between the input and the output in a deep learning algorithm, a complex structure of hidden layers is situated. It must contain at least two layers to be classified as deep learning. To train a deep learning algorithm, an immense amount of data are needed, much more than its less complex cousin, ML. The magic happens in the hidden layers, and those are benighted. Deep learning has gained much attention lately through the application of GenAI, which (as the name suggests) generates new content based on what it has learned from. The breakthrough in deep learning and GenAI lately is, as already mentioned, the increased performance in hardware but also the usage of transfer learning among deep learning algorithms. This implies that an already existing deep learning algorithm that has already been trained on a massive amount of data is transferred to a related field. This transfer can speed up the development of new deep learning algorithms substantially.

Data have for over a decade been perceived as the new oil in the digital economy; however, extracting and refining the new oil has not been as straightforward as anticipated. Data come in unimaginable different formats, and far from all data are of value, or are they? Extracting valuable data is like mining for diamonds: there is a lot of vast before finding a new Kohinoor. Deep learning might be the key to unlock the potential of unstructured data through its inherent ability to search through vast amounts of data and extract valuable information. The downside of deep learning is the energy consumption required due to the training being very computationally expensive when starting from scratch. Here transfer learning among similar deep learning algorithms will be important to cut cost and time.

Data need context and refinement. Still, to date, there is no universal taxonomy on data. Disciplines are using different methods for determining the quality of data, and many data themselves need metadata to be understandable. Data quality is also dictated by context: data sufficient for one type of application

 1556-6072/24/©2024IEEE

might be discarded by another one. Much more harmonization and standardization work on defining data quality across disciplines is necessary. There has been an attempt to establish a data readiness level scheme resembling the technology readiness level ladder, which is today widely used in engineering, but it seems to be more challenging to find a similar scheme for data.

Traditional products such as cars stemming from mechanical engineering are slow in reaping from digital technologies. For sure, modern vehicles contain a lot of data, but where are the diamonds? Legislators believe that vehicles contain not just one diamond but several; they have just not been extracted yet. The legislators are looking for data around the health status of the vehicle to enable more third-party services to emerge that might benefit the customer.

Vehicles are expensive to manufacture, but this is due to the rigorous processes surrounding the development and manufacturing of the vehicle rather than the physical components it is composed of. Vehicles have though been through a huge transformation during the last decades, containing more software and more sensors. AI will be crucial for the development of more advanced ADAS and automated driving system (ADS) features; then data fit for purpose will be needed, which starts with the data preparation.

Data Preparation

The time and the importance of the data collection and data preparation processes should not be underestimated in the development of new ADAS and ADS functionalities using AI. The raw data that have been collected for a specific task must be curated and groomed, involving tasks such as organizing, cleaning, labeling, and feature engineering the data. A high-quality preprocessed training dataset will increase the chances of a successful ML outcome. Deep learning algorithms can to some extent perform excellently without some of the preprocessing steps such as feature engineering, but on the other hand, massive amounts of data are a must. The time it takes to train the ML algorithm is a minor part of the overall work; the preprocessing of the training dataset and the final testing are the two most time-consuming parts.

The steps involving the preprocessing of the training dataset require much manual work to be performed by data scientists even though there are tools available that can be of support. In turn, these tools can also be based on AI.

In summary, the training dataset is the crucial element for succeeding in AI. The data need to be fit for purpose, implying that they should have the necessary quality for the intended functionality, they should be unbiased, and there should be enough data to train on. Deep learning algorithms need a huge amount of data, much more than ML. Spending time in finding the optimal training dataset will pay off in the long run.

Privacy

Privacy concerns have been raised in conjunction with the usage of AI. Protecting the privacy of individuals needs to be considered in all AI development steps from the establishment of the training datasets to the final product. The European Union has agreed upon and published several legislative acts recently to address the holistic perspective of digital technologies: the Data Act, the AI Act, and the Cyber Resilience Act (CRA). Privacy has been addressed since 2016 by the General Data Protection Regulation (GDPR) in Europe, and the new pieces of legislation have the privacy and safety of individuals at their cores. Road vehicles put on the market in Europe must adhere to GDPR already and to the Data Act and AI Act once they become applicable in a couple of years. Of course, the AI Act must only be fulfilled if AI is used in the vehicle. Road vehicles are not in the scope of the CRA; instead, they are covered by the sectoral cybersecurity regulation UNECE R155 in Europe.

Cybersecurity

Securing a product or service entails everything from vetting employees to physical doors to data centers. The ingenuity of adversaries is infinite, and the most secure product is the one that is not yet on the market. Cybersecurity has in the past been an afterthought, but products and services developed must be secure by design.

The transfer of data between sender and receiver, i.e., the digital communications being wired or wireless, is secure due to strong encryption mechanisms. Adversaries exploit software vulnerabilities at the edges of the network before data have been encrypted or transferred.

Open source software speeds up the development of software, driving down cost, but at the same time, vulnerabilities can be exploited on a larger scale by adversaries than if all software was developed from scratch in house. The software bill of materials (SBOM) detailing what software components a product is made up of is necessary to establish for products and services. The SBOM, together with cybersecurity management systems, is a must for complex products such as vehicles. A car contains 100 million lines of software code, and this will triple by 2030. Keeping track of all software inside a vehicle is a tedious but necessary task.

Autonomous Driving

The year 2023 was once again an upheaval year among startup companies dealing with autonomous driving technologies for trucks. TuSimple has, as of December 2023, laid off all its workers in the United States, and they will focus on activities in China. No one wanted to save Embark from bankruptcy, and what was left of the company was sold to Applied Intuition for roughly US$70 million in May last year. Applied Intuition focuses on

the simulation and validation of ADAS and ADS functionality.

Aurora Innovation, developing a driver for mainly automated trucks, has due to competitors' disappearance managed to receive more than US$800 million in new capital. The economic situation for Kodiak Robotics is opaque since it is privately held, but they have, for example, received US$50 million from the U.S. Department of Defense for applying their technology in the army.

The Daimler trucks independent subsidiary Tors Robotics plans for driverless routes with trucks in 2027; having access to Daimler's loyal customer base makes them well positioned to succeed. Two startups that have made AI their call are Canadian-born Waabi and recently founded Stack AV, launched by the Argo AI founders during the last year. Both Waabi and Stack AV focus on automated truck technology. Argo AI was backed by Ford and VW and focused on self-driving technologies for cars but was closed in 2022. Waabi has developed an advanced simulation environment powered through GenAI for testing their autonomous driver. This greatly enhances testing in a realistic, safe, and fast way, to reach the millions of miles required to build confidence in the technology.

Wrap Up

The recent advancements in AI hold great promise for boosting more advanced ADAS features and launching ADS functionality. Maybe it is AI that in the end will bring automated vehicles onto public roads on a large scale. AI will for sure be an integral part in the automotive industry in the future from when a vehicle is born until it is retired.

Safety is at the core of everything in the automotive industry, leading to an inertia in picking up on new technologies. The world surrounding digital technologies is moving with the speed of light compared with the automotive world. A car has an average lifetime expectancy of 12–13 years, and imagine everything that has happened during the last decade on digital technologies such as AI.

Changing the rigid processes surrounding the development and manufacturing of vehicles takes time, not weeks but years, because there is confidence in the process. It will be a slow and steady progress toward integrating AI in different parts of the vehicle ecosystem.

Privacy and security shall be at the core when developing a new AI algorithm. Much time must be spent on curating the training dataset, which in the end will dictate the performance of AI-powered functionalities in road vehicles. ***VT***