Under the hood of AI

Demystifying how leaders scale value of AI solutions by ensuring seamless data access.

At a glance

Technology-based competitiveness is increasingly driven by artificial intelligence (AI). Yet few companies have successfully scaled the value of AI solutions.

Under the hood of AI, powerful engines built on machine learning are becoming table stakes. The most critical building block for success is the data that fuels the engines. Data puts a cap on the scale of AI solutions' business impact.

Coupled with successful AI solutions, we find three cornerstones of data management that together ensure seamless data access:

i) A data strategy that puts focus on the most valuable data assets and defines how to secure control of and access to these;

ii) A data platform which ensures that valuable data from different sources can be accessed efficiently where and when its needed; and

iii) Ways-of-working adapted for data-driven and AI-based business.

The three cornerstones allow AI to be purpose-built from the start, enable learning loops with ever-better data quality, and ultimately scale customer benefits and the business impact of AI.

AUTHORS

Oscar Brodin
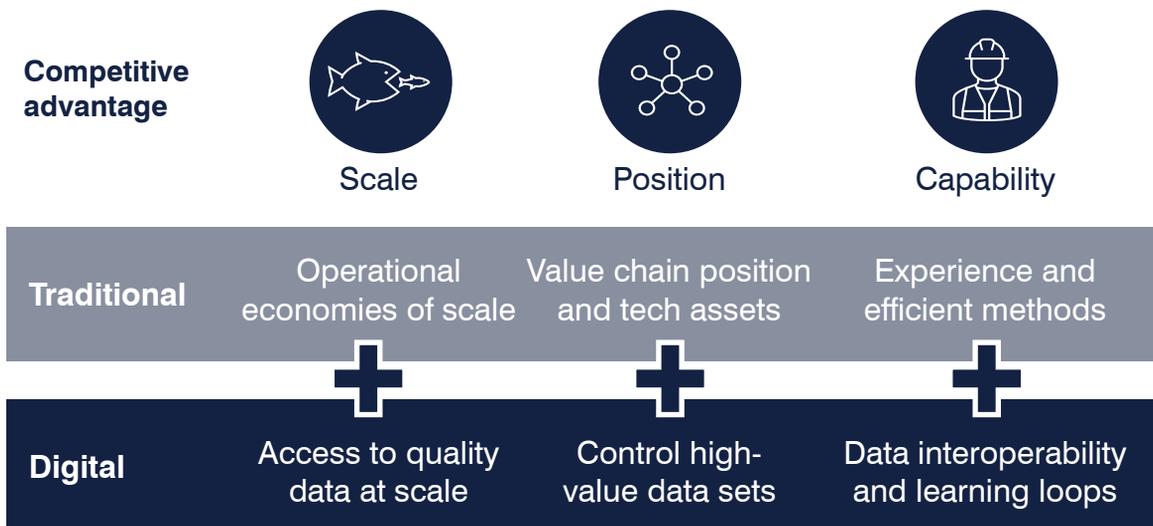Manager

Robin Sparrefors
Director

AI holds the promise of unlocking immense business value by expediting, automating and optimizing tasks previously reserved for the human mind, and by solving complex problems beyond the capability of previous methods. AI is expected to generate trillions of dollars of value in coming years and tens of billions in venture capital funding have been invested into AI companies.[1]

In today's competitive environment, companies cannot afford to lag in essential technology. Every advantage counts. To provide an analogy from the sports world, the 6 different winners of Wimbledon over the last 20 years have only won 52% of the points they played. Having just a slight edge over your competition is what wins out, and increasingly this edge comes from AI.

The growing importance of AI largely stems from shifts in what competitive advantage means when business becomes digital *(see Exhibit 1)*. Using automotive OEMs as an example, competitive advantage from scale has traditionally meant production scale. With the entrance of intelligent systems and self-driving cars, today's automotive leaders cares as much about collecting quality vehicle data at scale. Position has traditionally meant value chain control, i.e. securing the right cost levels with suppliers and controlling differentiating system-level technology assets. In a digital setting, it increasingly means control of high-value data sets secured from multiple sources in automotive ecosystems. Capability has meant developing your people and methods so that the next seven-year cycle produces better cars at higher efficiency. With digital, it also means having learning loops leveraging real-time data from vehicles in the market, and continuously updating the cars over-the-air. While all the traditional interpretations of competitive advantages are still valid today, their relative importance is decreasing.

Exhibit 1: Shifts in what competitive advantage mean

Competitive advantage	Scale	Position	Capability
Traditional	Operational economies of scale	Value chain position and tech assets	Experience and efficient methods
Digital	Access to quality data at scale	Control high-value data sets	Data interoperability and learning loops

Although most large companies have AI initiatives with executive management support, few consider themselves AI-advanced and are satisfied with business outcomes from AI.[2] The benefits from optimizing internal processes using AI, and the revenue from AI-based customer value propositions are not scaling.

There are many reasons why companies struggle with AI, but it predominantly comes down to data.[2] Data is the most critical building block of competitive AI solutions – in many regards more important than the machine learning code. Data quantity and quality are directly correlated with performance of AI solutions. Failing to understand the value of, and secure seamless access to, data puts a cap on the benefits a company could realize from AI. Also large companies with vast amounts of data, holding great potential for generating business value from AI applications, struggle to scale benefits.[3]

**Case: Ericsson creating competitive edge from AI**

Telecom equipment provider Ericsson leverages AI for e.g. self-organizing telecom networks and carrier aggregation, optimizing the network management for their customers. Such solutions could provide up to 25% better 5G coverage and close to 20% better network load distribution.[4] These intelligent and autonomous networks with close to zero touch are constantly improving and optimizing based on the learning it generates from the real-world network data.

As with most AI offerings, one can expect the customer benefit of the service to increase over time and competitors without the same contextual training of their AI will find it increasingly harder to compete.

AI leaders have three key data management cornerstones in place. They approach data with strategic intent, anchored in a data strategy. They are equipped with data platforms that technically enable the development and delivery of AI solutions. And, they adapt ways-of-working to be data-driven. Together, these cornerstones ensure the seamless data access needed to create AI solutions with real business impact.

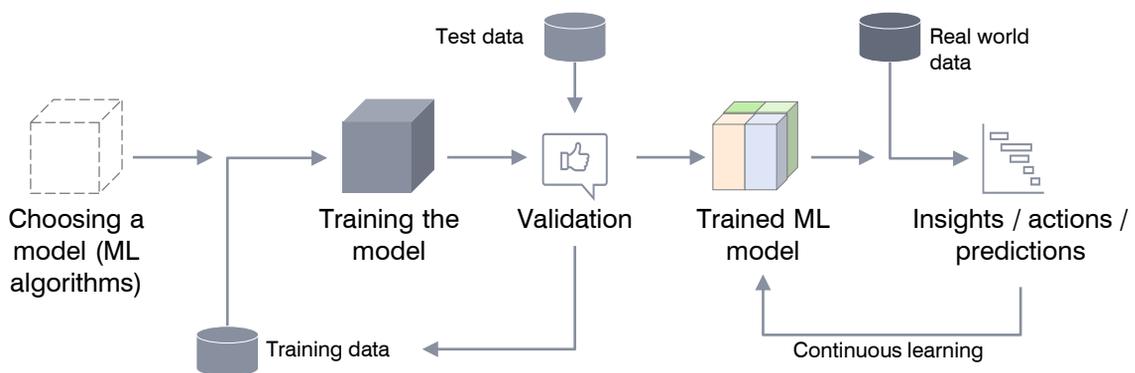Exhibit 2: Data management cornerstones to scale the value of AI

Data strategy

Seamless data access

Data platform

Ways-of working

Scaling the value of AI
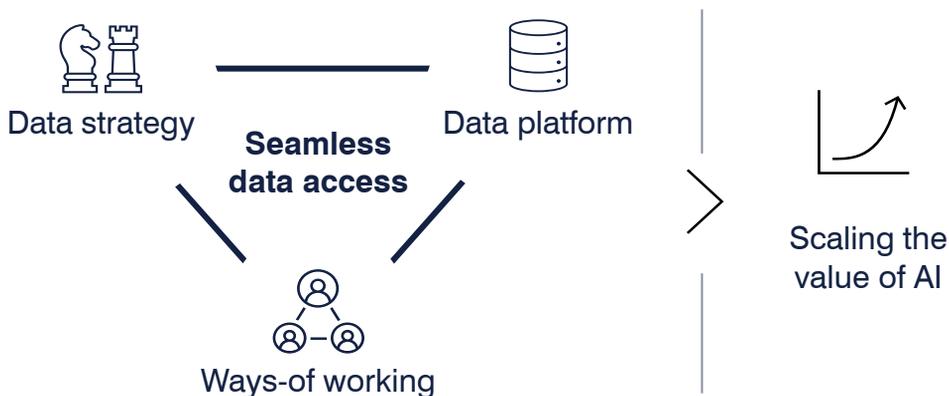
AI learns and improves through data

To understand why data is essential to the success of AI solutions, let's first look at how AI works under the hood. Artificial intelligence is often defined as a system able to carry out tasks normally reserved for humans and considered to some extent intelligent. In most cases this is made possible through machine learning models that are trained with data and continuously learn and improve upon themselves based on data feedback loops.

Exhibit 3: Training a Machine Learning Model

Choosing a model (ML algorithms) → Training the model → Validation → Trained ML model → Insights / actions / predictions

Test data → Training the model

Training data

Real world data
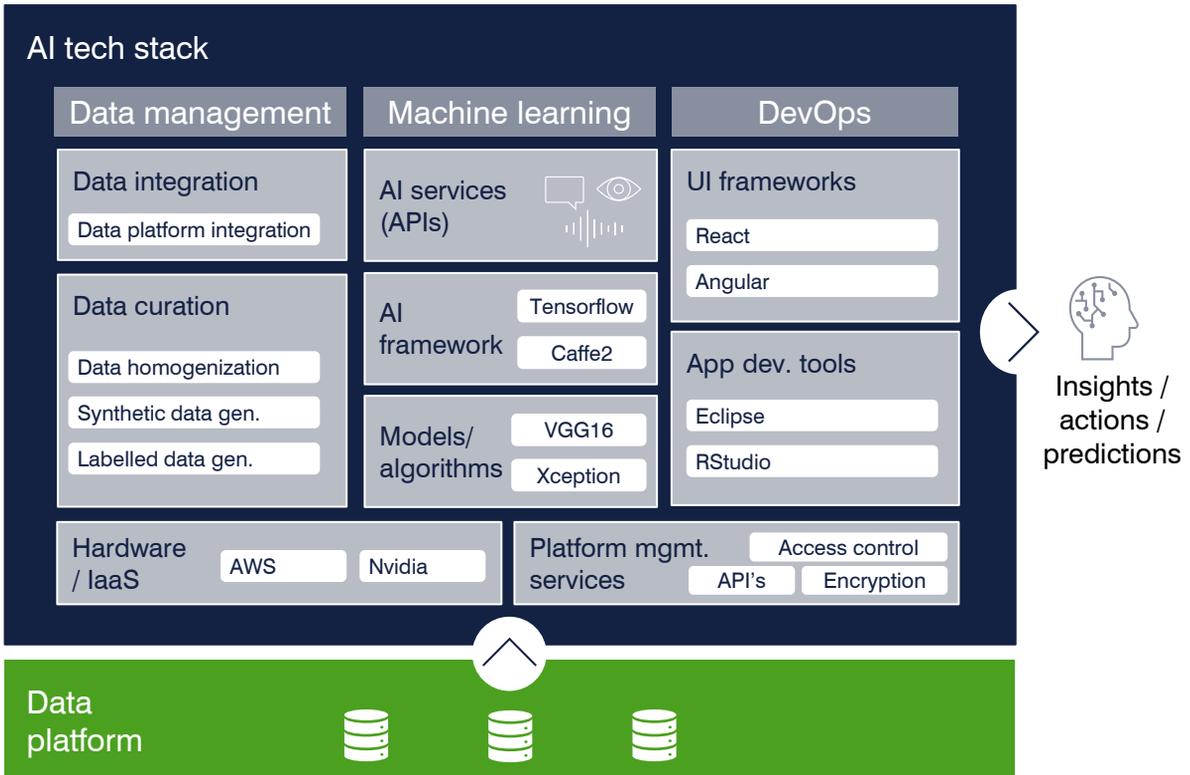
Continuous learning

Developing AI solutions is mostly about data management and DevOps

There is a perception that the primary effort in developing AI solutions is the machine learning element. However, writing machine learning code and finetuning algorithms is only a portion of the effort that goes into a solution or a system, considering the full AI stack (see *Exhibit 3*).

A larger portion of the effort is spent managing data (e.g. integration and curation of data) and on DevOps (e.g. developing, testing, releasing and operating software).[5] This is in part why creating successful AI systems is inherently complex. One needs to manage both the usual technical complexity of a codebase and the machine learning specific complexity from data and modelling.

AI tech stack

Data management	Machine learning	DevOps
Data integration Data platform integration	**AI services (APIs)**	**UI frameworks** React Angular
Data curation Data homogenization Synthetic data gen. Labelled data gen.	**AI framework** — Tensorflow, Caffe2 **Models/ algorithms** — VGG16, Xception	**App dev. tools** Eclipse RStudio
Hardware / IaaS — AWS, Nvidia	**Platform mgmt. services** — Access control, API's, Encryption	

Insights / actions / predictions

Data platform

Note: Activities, functions and providers in different parts of the AI stack are examples for illustrative purposes.

Main differentiation in AI comes from data

But isn't the machine learning code the secret sauce? Likely not. High performing machine learning frameworks and pre-trained models are increasingly becoming open source. The machine learning part of the stack will therefore not be the differentiator between AI solutions for most companies. Differentiation will instead come from the data you use to train and improve the AI. As a result, companies should focus on the three cornerstones of data management to scale the benefits of AI.

Vast data sets are not enough

Although synthetic data generation is becoming more advanced, most successful applications of machine learning techniques rely on historical data for training the models. If this is the case, why do companies across industries that sit on vast data assets, accumulated over years of operations, often struggle to make good use of it with AI? The likely main cause is failing to treat data as a business asset.[6] This leads to poor consistency and quality of data over time and decreases the learning value for AI solutions. Methods such as supervised machine learning rely on annotated or labeled data to learn, which are made more difficult by poor quality data.

In managing data, companies therefore need to identify which data sets constitutes core assets for present as well as future use cases for AI applications. Otherwise, companies might risk becoming constrained by limited access to strategic data. Especially companies with existing and potential for building future data assets at scale have an incentive to treat data as a strategic business resource, as they would with any financial assets.

Change anything, change everything

Well-managed data in terms of quality and accessibility is the foundation to reducing complexity in developing and maintaining AI solutions. From a technical perspective, a key shortcoming of machine learning systems is their natural susceptibility to entanglement or what is often referred to as the "change anything, change everything" principle. Many times, this leads to individual changes, e.g. to the type of input data, impacting the whole system in ways that cannot be isolated or predicted. Hence, leading to unwanted or suboptimal output from the AI and increasingly complex and costly software stacks.

Risks stemming from such unexpected changes to the system can partly be mitigated with technical solutions, such as system monitoring, but this is largely retroactive. The most efficient way to handle this complexity is through proactivity in the capture and curation of data. Your data strategy is therefore key to scaling value from AI.

Further reading:
The winning card in industrial digitalization

Winning in digital ecosystems comes down to using intellectual property to control key positions, build trust, and balance exclusivity and openness

Pinpoint critical data assets and secure seamless access

Proactive data management can be a challenging endeavor and treating all data the same will quickly become costly and resource heavy. The data strategy must provide focus. It must prioritize data based on the business potential of the envisioned use cases, i.e. identify the most valuable data assets that exist today and that will be created in the future.

When critical data sets have been defined, functional requirements should dictate how different stakeholders must manage the assets. Functional requirements can include how data shall be created, collected, refined and used, as well as targets or minimum requirements for quantity and quality of data. More importantly, the data strategy must define how to secure access to critical data sets and how to control these data sets, which does not necessarily equate to ownership of data. This is vital as critical data will in many cases originate from external sources such as customers' use of products and ecosystem partners' sensors. Losing access to critical data sets can significantly reduce the value of AI solutions, e.g. if learning loops with cross-value chain data are disrupted, predictions can quickly lose their accuracy.

Exhibit 5: Data strategy pinpoints and secures access to critical data
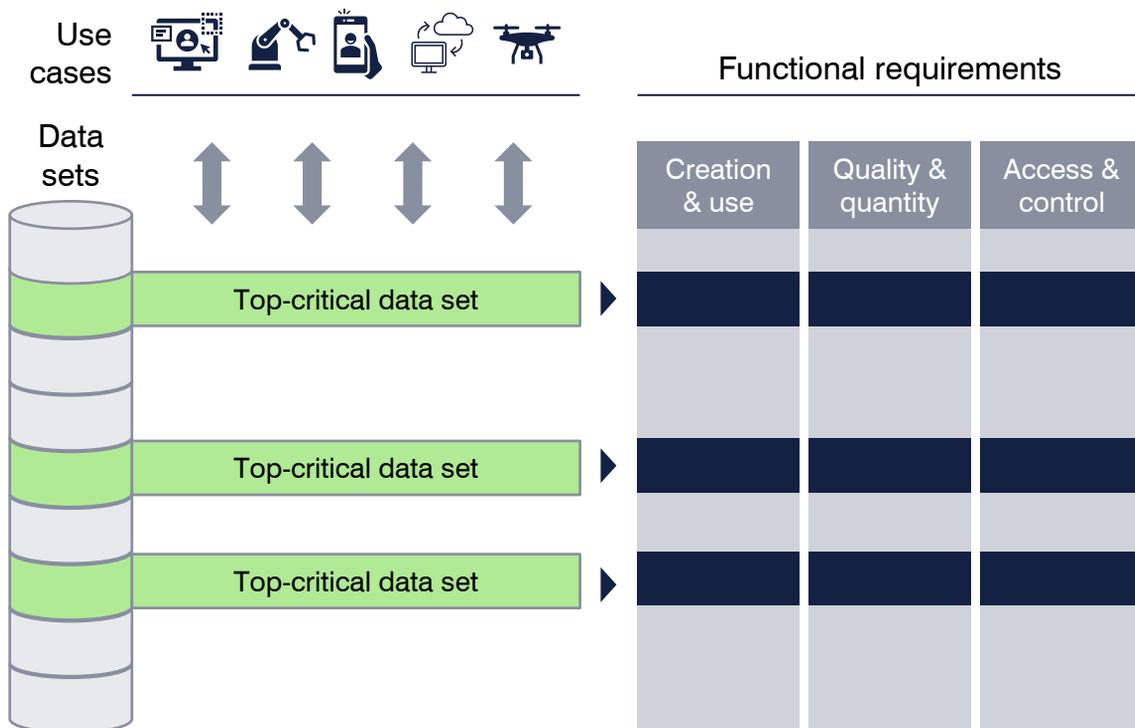
Cornerstone 2: Data platform

Technology-enabled seamless data access

A data strategy that pinpoints critical data assets is a first step to a strong foundation for AI success. A data platform is the next step to practically ensure that AI solutions can seamlessly access data with consistent quality at scale.

The main function of the data platform is to allow data from different sources to be governed, accessed and processed as a single source of truth. Independent of the organizational structure, whether centralized or decentralized management of data, a data platform introduces both control and flexibility in terms of quality and access of data within the company. This is especially important when data assets are derived from different data sources and different parts of the technology stack, e.g. spanning from edge device data to app usage data. A data platform brings it all together in one place for scalable use in AI systems. This way, it also becomes an enabler for cross-functional and agile collaboration in AI development. Making the data platform modularized, as independent as possible from legacy IT systems, can also help bring data closer to the business while providing additional benefits in decoupling from potential IT transformations.

Multiple vendors provide capable data platforms. When selecting a solution, it is key to understand the company's requirements on the platform to make data access seamless to the right people at the right time. Informed design of the data platform can significantly improve the potential for cross-functional collaboration in AI while managing risks in relation to data.

Exhibit 6: Illustrative overview of data platform
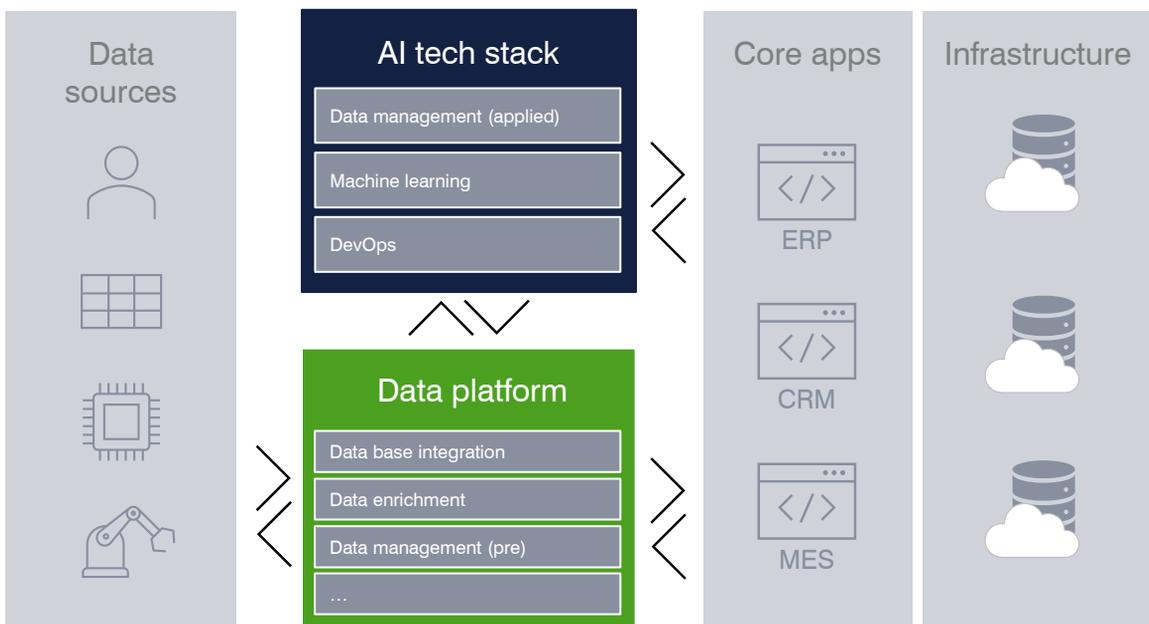
Cornerstone 3: Ways-of-working fit for data

Data managed by everyone

Generating and managing data with AI in mind requires more of companies than hiring people skilled in data science or data engineering. You need to organize for data. The responsibility for keeping data fit for use and accessible should be everybody's. Fixing data shortcomings after the fact is expensive (if even possible) and inefficient data management will limit the performance of AI solutions.

Organizing for data is not about centralization versus decentralization; central data office teams can be a considerable bottleneck while local data teams tend to create data silos rather than seamless accessibility. It is about taking distributed responsibility for data and working cross-functionally to solve problems and leverage insights from data. Very much with a "startup mindset". This is not to say that a Chief Data Officer does not play a significant role in being accountable for how everyone in the company should work with data. Managing data should be a part of the job description for key data stakeholders across business functions but also be an expectation of most people in any company where data can provide a competitive edge.

Exhibit 7: Operating model implications from ways-of-working fit for data

Operating model	Way-of-working implication (examples)	
Organization	Distributed responsibility for keeping data fit for AI and seamlessly accessible	
Strategy	Readily available functional requirements guiding people on data asset use and access	
Processes & tools	Equipped to cross-functionally create and use data in developing and delivering AI solutions	
Governance	Let AI inform and augment decision making	

Equip people for seamless data access

Cross-functional collaboration is important for the execution of AI initiatives. Many AI use cases span across stakeholders, inside and outside the company. Development should not be siloed, as this will ultimately lead to AI solutions addressing marginal isolated tasks rather than more strategic implementations. People across e.g. data science, R&D, DevOps and commercial teams as well as customers and partners, need to help contextualize data, train models and reduce biases. To do so, they need be equipped with appropriate processes and tools.

In addition to the data platform, solving the technical side of the tool belt, a key part of ways-of-working is ensuring that people can "live and breathe" the data strategy. Ultimately, a data strategy is only as good as the actions it yields. If the functional requirements on data quality, quantity and access are not comprehended and accepted by its users, they will not have the intended effect.

Tools can include triggers and alerts in the user interface when accessing critical data assets. Another example is a data strategy-on-a-page or quick reference guide that ensues different business functions have the functional requirements applicable to them readily available when they use data assets. For a developer in an R&D project it can be requirements on data quality for key data sets, for a procurement lawyer what data to contractually secure and pay for in a collaboration agreement, and for a sales rep how to think about data in customer communication and enable monitoring of data usage.

AI-driven decision making

Governance fit for AI is important as companies move towards agile ways-of-working across functions. The autonomy of agile teams benefits highly from swift decisions based on up-to-date data. Taking it one step further, teams should aim to employ AI-driven decision making to make full use of data at scale and reduce bias.[7]

As an example, an industrial company might develop an AI tool for predictive maintenance, identifying issues to be addressed and their optimal response. Rather than having a manager propose and approve a schedule, the maintenance engineer is now able to leverage the AI output for her planning. Following this logic, governance structures should be specified for tasks where AI can augment decision making. This should make it clear where AI can provide recommendations and which employees that can make the final decision, if not the AI system itself. This has the potential to free up resources but also improve decision making where humans may fall short.

Machine learning can be compared to an increasingly commoditized drivetrain under the hood of AI solutions. In contrast, the data that fuels AI engines provides an increasingly larger share of the competitive advantage. To succeed in AI, you should not stop investing in machine learning, but allocate investments proportionally to the value contribution of everything that is under the hood – and recognize the importance of the data.

- Put a data strategy in place to pinpoint existing and future critical data assets and define how to secure access to and control of these.

- Select the right data platform to technically enable AI solutions to access data where and when it is needed.

- Avoid the pitfall of thinking that a strategy and an IT system will solve it all. Ensure that your ways-of-working are fit for data and AI-based business.

These three cornerstones will create seamless data access that allows AI solutions to evolve with better data, continuously improve customer benefits and scale the business impact of AI.

Sources

[1] PwC (2020), Artificial Intelligence Study

[2] NewVantage Partners (2021), Big Data and AI Executive Survey

[3] Mindtree (2019), AI Readiness Report 2019

[4] Ericsson (2021), AI in Networks, Retrieved March 11 from https://www.ericsson.com/en/ai

[5] Sculley et al. (2015), Hidden Technical Debt in Machine Learning Systems

[6] Informatica (2020), How to Govern Your Data as a Business Asset

[7] Eric Colson (2019), What AI-Driven Decision Making Looks Like, Harvard Business Review