Smartening up with Artificial Intelligence (AI) – What's in it for Germany and its Industrial Sector?

Digital/McKinsey

Preface

Artificial intelligence (AI) is finally bringing a multitude of capabilities to machines which were long thought to belong exclusively to the human realm: processing natural language or visual information, recognizing patterns, and decision making. While AI undoubtedly holds great economic potential for the whole world, in this report we explain how and where AI will likely affect the German industrial sector by exploring several questions: Which subindustries are most strongly affected by the automation potential of AI? What are the most promising use cases? What are pragmatic recommendations for managers of industrial players planning to harness the power of AI?

We describe several use cases in which we highlight the impact of AI and aim to quantify it. These use cases were carefully selected based on their economic potential and their ability to demonstrate the benefits of AI in practice. We do not claim that AI – despite its enormous potential – is the silver bullet for every business problem. We realize that AI is very often the enabler for performance improvements whose actual realization requires changing business processes. It is a rapidly evolving field. Thus, the present report needs to be understood as a peek into the future based on the current state of the art. With these caveats we are confident that this report will provide managers in the German industrial sector with valuable guidance on how they can benefit from AI.

Acknowledgements

This study was conducted by McKinsey & Company, Inc. We wish to express our appreciation and gratitude to UnternehmerTUM's[1] artificial intelligence application unit for their support and valuable contributions.

The authors would especially like to thank:

Andreas Liebl, Partner New Venture Creation, UnternehmerTUM
Alexander Waldmann, Visionary Lead AI, UnternehmerTUM

1 UnternehmerTUM, founded in 2002, is one of the leading centers for entrepreneurship and business creation in Europe.

Contents

Executive summary

Self-learning machines are the essence of artificial intelligence (AI). While concepts already date back more than 50 years, only recently have technological advances enabled successful implementation at industrial scale. According to the McKinsey Global Institute (MGI), at least 30% of activities in 62% of German occupations can be automated, which is at a similar level as the US[2]. Freed-up capacity can and needs to be put to new use in value-adding activities to support the health of Germany's economy. AI has proven to be the core enabler of this automation based on advances in such fields as natural language processing or visual object recognition.

Highly developed economies, like Germany, with a high GDP per capita and challenges such as a quickly aging population will increasingly need to rely on automation based on AI to achieve its GDP targets. About one-third of Germany's GDP aspiration for 2030 depends on productivity gains. Automation fueled by AI is one of the most significant sources of productivity. By becoming one of the earliest adopters of AI, Germany could even exceed its 2030 GDP target by 4%[3]. However, if the country adopts AI more slowly – and productivity is not increased by any other means – it could lag behind its 2030 GDP target by up to one-third.

AI is expected to lift performance across all industries and especially in those with a high share of predictable tasks such as Germany's industrial sector. AI-enabled work could raise productivity in Germany by 0.8 to 1.4% annually.

We selected eight use cases covering three essential business areas, (products and services, manufacturing operations, and business processes) to highlight AI's great potential in the industrial sector.

Products and services:
- Highly *autonomous vehicles* are expected to make up 10 to 15% of global car sales in 2030 with expected two-digit annual growth rates by 2040. The efficient, reliable, and integrated data processing that these cars require can only be realized with AI.

Manufacturing operations:
- *Predictive maintenance* enhanced by AI allows for better prediction and avoidance of machine failure by combining data from advanced Internet of Things (IoT) sensors and maintenance logs as well as external sources. Asset productivity increases of up to 20% are possible, and overall maintenance costs may be reduced by up to 10%.

- *Collaborative and context-aware robots* will improve production throughput based on AI-enabled human-machine interaction in labor-intensive settings. Thereby, productivity increases of up to 20% are feasible for certain tasks – even when tasks are not fully automatable.

- *Yield enhancement in manufacturing* powered by AI will result in decreased scrap rates and testing costs by linking thousands of variables across machinery groups and sub-processes. E.g., in the semiconductor industry, the use of AI can lead to a reduction in yield detraction by up to 30%.

2 See MGI "A future that works," January 2017.

3 Assumption: Displaced labor is redeployed into productive uses.

- *Automated quality testing* can be realized using AI. By employing advanced image recognition techniques for visual inspection and fault detection, productivity increases of up to 50% are possible. Specifically, AI-based visual inspection based on image recognition may increase defect detection rates by up to 90% as compared to human inspection.

Business processes:

- *AI-enhanced supply chain management* greatly improves forecasting accuracy while simultaneously increasing granularity and optimizing stock replenishment. Reductions between 20 and 50% in forecasting errors are feasible. Lost sales due to products not being available can be reduced by up to 65% and inventory reductions of 20 to 50% are achievable.

- The application of machine learning to enable *high-performance R&D projects* has large potential. R&D cost reductions of 10 to 15% and time-to-market improvements of up to 10% are expected.

- *Business support function automation* will ensure improvements in both process quality and efficiency. Automation rates of 30% are possible across functions. For the specific example of IT service desks, automation rates of 90% are expected.

Our findings concerning AI – as well as our observations of the most successful players in both the industrial and adjacent sectors – reveal five effective recommendations that address the challenges of AI and help get firms in the industrial sector started on their AI journey:

- Get a grasp of what AI can do, prioritize use cases, and don't lose sight of the economics – without a business case no innovation survives.

- Develop core analytical capabilities internally but also leverage third-party resources – trained people are scarce.

- Store granular data where possible and make flat or unstructured data usable – it is the fuel for creating value.

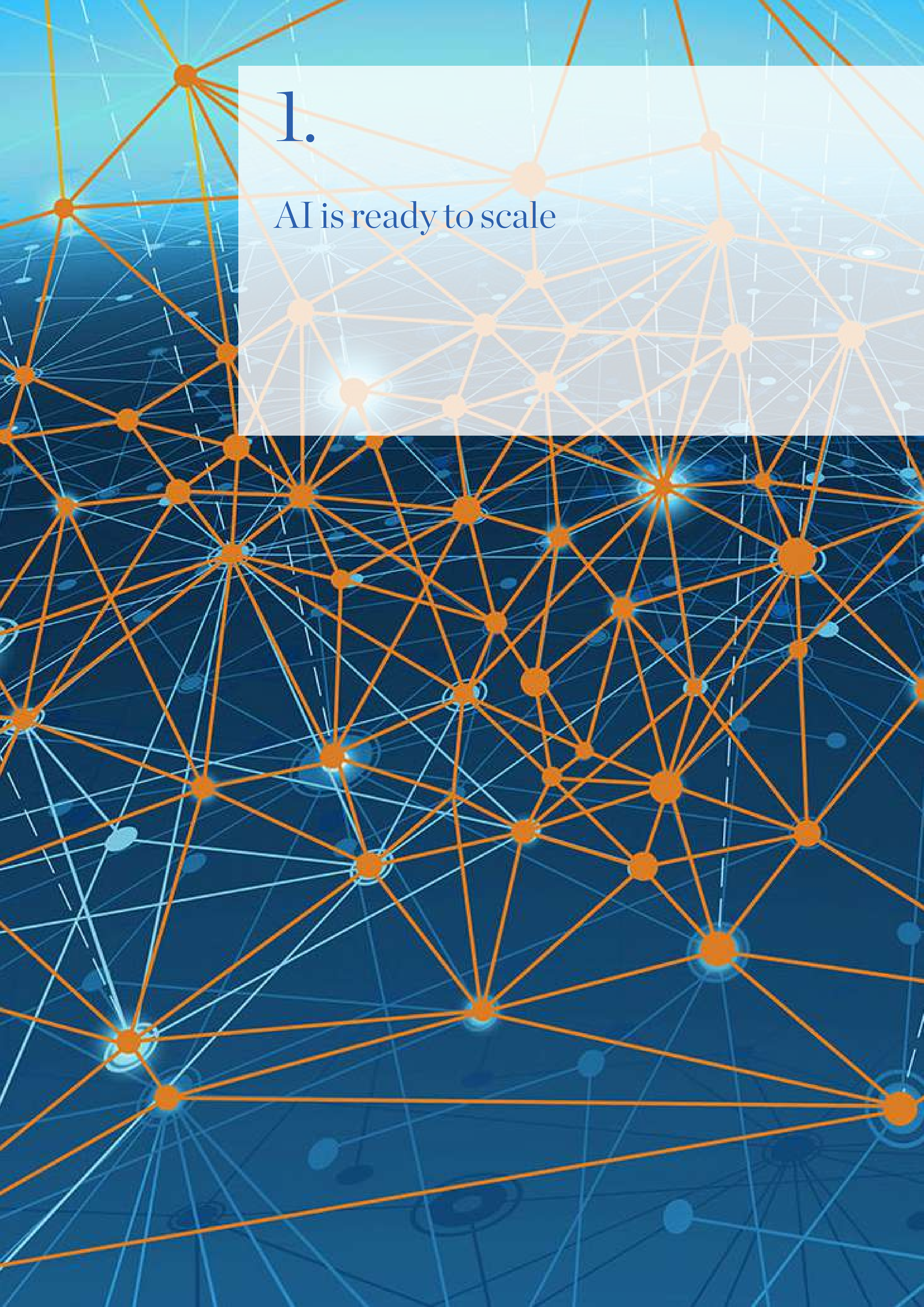
- Leverage domain knowledge to boost the AI engine – specialized know-how is an enabler to capture AI's full potential.

- Make small and fast steps through pilots, testing, and simulations – the AI transformation does not require large up-front investments, but agility is a prerequisite for success.

Beyond deciding where and how to best employ AI, an organizational culture open to the collaboration of humans and machines is crucial for getting the most out of AI. Trust is among the key mindsets and attitudes of successful human-machine collaboration. Initially, cultural resistance may be strong because the relationship between the inner workings of an artificially intelligent machine and the results it produces can be rather obscure. In a sense, it is no longer the algorithm but mainly the data used to train it that leads to a certain result. Humans will need some time to adjust to this shift. Getting started early not only helps produce results quickly but also helps speed up an organization's journey toward embracing the full potential of AI.

1.

AI is ready to scale



The essence of intelligence is learning. Just as humans learn how to communicate, identify visual patterns, or drive a car, machines can similarly be trained to perform such tasks based on powerful learning algorithms. A common method of training machines consists of providing them with labeled data, e.g., photographs of cats combined with the word “cat” as a label. Such machines are then said to possess AI[4] if they can – given their training – ascribe the correct label to a previously unknown data set with sufficient accuracy. Following the previous example, a machine would then be able to correctly identify a cat in an unfamiliar photograph.

Typical applications of AI include autonomous driving, computer vision, decision making, or natural language processing. AI holds the benefit of being adaptable to very heterogeneous contexts just like humans. Well-trained AI is capable of performing certain tasks at the same skill level as humans but with the additional advantages of high scalability and no need for pauses. AI can discover patterns in the data that are too complex for human experts to recognize. In some specific applications such as computer vision, AI has already achieved performance levels surpassing that of humans (e.g., in skin cancer diagnostics).

The idea of AI dates back to the 1950s when AI successes were largely limited to the scientific field. In the last years, established IT giants like Google, IBM, and nVidia – fueled by the abundance of data, algorithmic advances, and the usage of high-performance hardware for parallel processing – have begun bridging the gap between science and business applications. Nowadays, adoption of AI has become increasingly easier due to freely available algorithms and libraries, relatively inexpensive cloud-based computing power, and the proliferation of sensors generating data. Hence, not only established firms but also start-ups play a significant role in bringing AI to life. Start-ups with AI-savvy founders are capable of developing AI-based products in less than three months.

In the industrial sector, AI application is supported by the increasing adoption of devices and sensors connected through the Internet of Things (IoT). Production machines, vehicles, or devices carried by human workers generate enormous amounts of data. AI enables the use of such data for highly value-adding tasks such as predictive maintenance or performance optimization at unprecedented levels of accuracy. Hence, the combination of IoT and AI is expected to kick off the next wave of performance improvements, especially in the industrial sector.

Given its growing accessibility, broadening applications, and specific relevance to the industrial sector, it comes as no surprise that AI is a hot topic for leading researchers, investors, think tanks, and companies. It is hard to open a newspaper without coming across an article on AI. As per a Tracxn[5] analysis, start-ups dealing with AI-related topics have raised around USD 6 billion in funding in 2016 alone.

4 The process described here refers to supervised learning, a type of machine learning. See Text Box 1 on the differentiation between AI and machine learning. Within AI, there is the distinction between strong AI and weak AI. Strong AI or true AI is often defined by using the Turing Test. According to the Turing Test a machine possesses AI if it can provide a human with written responses to a set of questions so that the human cannot tell whether answers were given by a machine or another human being. In this report we follow a broader definition of AI that includes machines capable of learning that would not pass the Turing Test (“weak AI”).

5 Venture capital investment tracking company.

The global market for AI-based services, software, and hardware is expected to grow at an astonishing annual rate of 15 to 25% and reach USD 130 billion by 2025. Machine learning is expected to be the dominant methodology (see Text Box 1). In summary, AI is ready to scale across industries and it is has already begun to do so.

In this publication, we:

- Outline the influence that AI will have on the German economy
- Dive into business applications along eight specific use cases, with a special focus on the industrial sector[6]
- Describe five pragmatic recommendations that CEOs should consider in the upcoming months

Text Box 1: the nomenclature of artificial intelligence

Artificial intelligence is a buzzword these days and, hence, subject to multiple interpretations. For the purpose of establishing a common understanding, we have defined various AI terms as they are used in this report. For additional information see also the appendix.

- *Artificial intelligence (AI)* is intelligence exhibited by machines, with machines mimicking functions typically associated with human cognition. AI functions include all aspects of perception, learning, knowledge representation, reasoning, planning, and decision making. The ability of these functions to adapt to new contexts, i.e., situations that an AI system was not previously trained to deal with, is one aspect that differentiates strong AI from weak AI. In this report, we will not make the distinction between weak and strong AI for the sake of simplicity and due to our focus on the business context.
- *Machine learning (ML)* describes automated learning of implicit properties or underlying rules of data. It is a major component for implementing AI since its output is used as the basis for independent recommendations, decisions, and feedback mechanisms. Machine learning is an approach to creating AI. As most AI systems today are based on ML, both terms are often used interchangeably – particularly in the business context.
- Machine learning uses *training*, i.e., a learning and refinement process, to modify a model of the world. The objective of training is to optimize an algorithm's performance on a specific task so that the machine gains a new capability. Typically,

6 Our particular focus is on aerospace, automotive OEMs and commercial vehicles, automotive suppliers, industrial equipment, and the semiconductor industry.

large amounts of data are involved. The process of making use of this new capability is called *inference.* The trained machine-learning algorithm predicts properties of previously unseen data.

- There are three main types of learning within ML, namely *supervised learning, reinforcement learning,* and *unsupervised learning*. They differ in how feedback is provided. Supervised learning uses labeled data ("correct answer is given") while unsupervised learning uses unlabeled data ("no answer is given"). In reinforcement learning, feedback includes how good the output was but not what the best output would have been. In practice, this often means that an agent continuously attempts to maximize a reward based on its interaction with its environment.

- Since the late 2000s, *deep learning* has been the most successful approach to many areas where machine learning is applied. It can be applied to all three types of learning mentioned above. Neural networks with many layers of nodes and large amounts of data are the basis of deep learning. Each added layer represents knowledge or concepts at a level of abstraction that is higher than that of the previous one. Deep learning works well for many pattern recognition tasks without alterations of the algorithms as long as enough training data is available. Thanks to this, its uses are remarkably broad and range from visual object recognition to the complex board game "Go."

2.

AI will increase productivity and transform the German economy

> "In many industries, we do not talk about artificial intelligence, but instead about augmented intelligence. Because the machines will not completely take over the tasks from humans, but instead replace a part of their activities."
>
> ***Helle Valentin,*** Global Chief Operating Officer, Watson Internet of Things at IBM

In a recent report published by MGI ("A future that works", January 2017), we show that about 1% of occupations can be fully automated in the US. At the same time, at least 30% of activities can be automated in 62% of occupations. These numbers are similar with respect to Germany, where roughly 2% of occupations can be fully automated and also 62% of occupations have at least 30% technically automatable activities. AI has proven to be the core enabler of this automation based on advances in such fields as natural language processing and visual object recognition.

We estimate that AI-enabled work could raise productivity[7] in Germany by as much as 0.8 to 1.4% annually. The impact is particularly important for Germany given that it is in the group of advanced economies with quickly aging populations. Germany will simply not have enough workers to maintain GDP projections per capita without productivity gains through automation, whereas younger economies will have more than enough workers to maintain their GDP targets per capita by increasing their share of the working population. These younger economies are typically in emerging markets, where their aspirations are to grow GDP per capita rapidly. About one-third of Germany's GDP target for 2030 depends on productivity gains. AI can provide the productivity boost required to achieve or even overachieve this target. By becoming one of the earliest adopters of AI, Germany could exceed its 2030 GDP aspiration by 4%[8]. However, if the country adopts AI more slowly, it could lag behind its 2030 GDP target by up to one-third.

In order to understand the degree to which certain sectors can benefit from AI, we have grouped work activities into seven high-level categories[9]. We then determined the relative mix of those activities for each sector. Sectors that rely disproportionately on automatable activity categories (i.e., data processing and predictable physical activities) are the strongest candidates for employing AI, while those that emphasize less automatable activities (i.e., people management and content expertise) have less overall potential for the application of AI (see Exhibit 1).

In the German manufacturing sector specifically, around 55% of all activities currently conducted by humans have the potential to be automated by AI technology. Performing physical activities or operating machinery in a predictable environment (e.g., packaging of

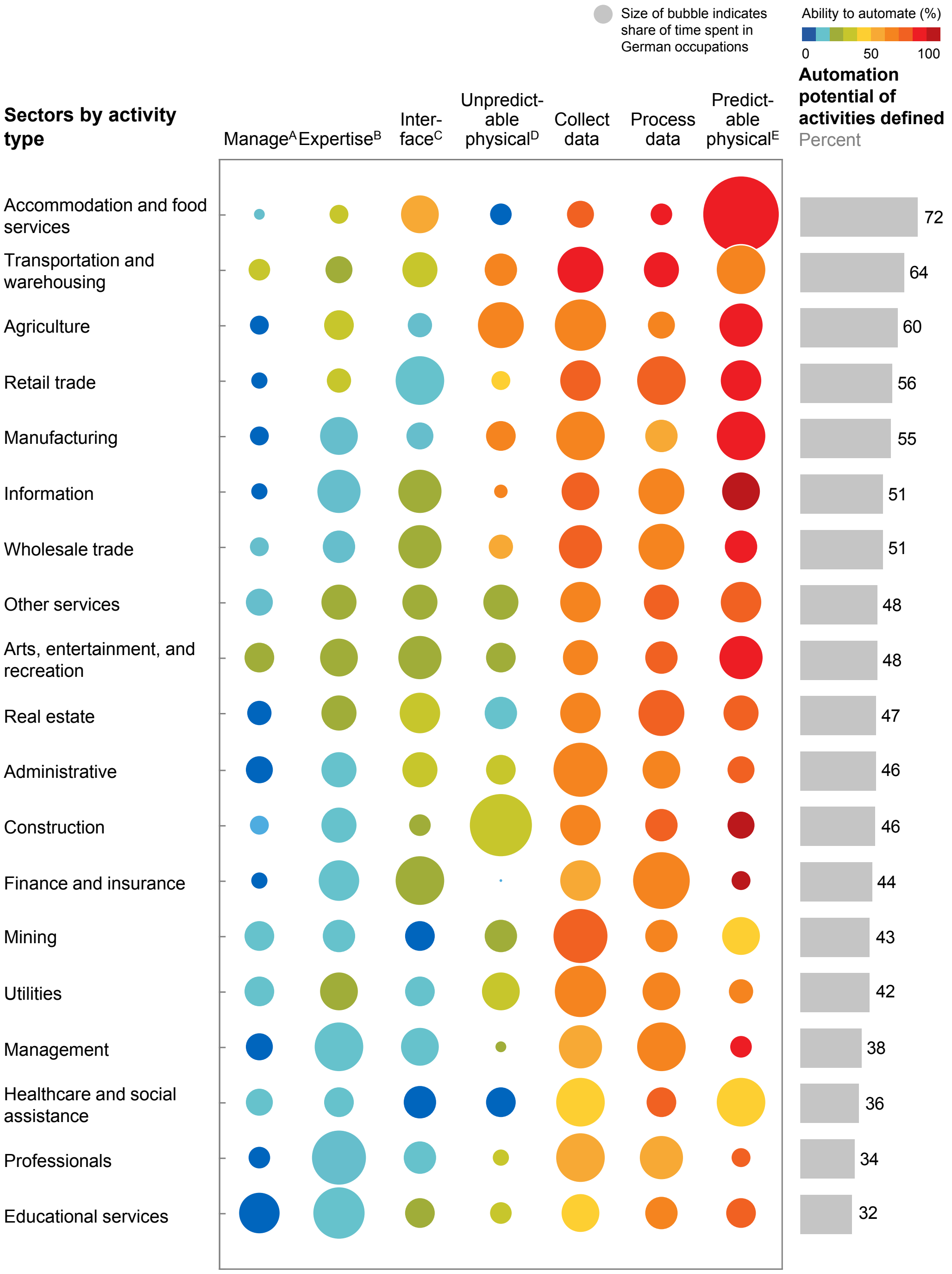
7 Defined as GDP per full-time-equivalent worker.

8 Assumption: Displaced labor is redeployed into productive uses.

9 Manage (managing and developing people), expertise (applying expertise to decision making, planning, and creative tasks), interfacing with stakeholders, unpredictable physical (performing physical activities and operating machinery in unpredictable environments), collect data, process data, and predictable physical (performing physical activities and operating machinery in predictable environments).

Exhibit 1

Technical potential for automation across sectors varies depending on mix of activity types



A Managing and developing people
B Applying expertise to decision making, planning, and creative tasks
C Interfacing with stakeholders
D Performing physical activities and operating machinery in unpredictable environments
E Performing physical activities and operating machinery in predictable environments
SOURCE: MGI analysis

products, welding) represents one-fourth of the overall work time in manufacturing. The automation potential of this activity type is around 90%. All other activity types – except manage, expertise, and interface – have automation potentials well above 50%. In line with US results (see MGI "A future that works", January 2017), the five sectors with the highest automation potential are accommodation and food services, transportation and warehousing, agriculture, retail trade, and manufacturing. The manufacturing sector in Germany has a slightly lower automation potential than that of the US (55 vs. 60%) because of the different composition of manufacturing occupations in each country. In both Germany and the US, the educational sector has the lowest automation potential (less than one-third) because employees working in this sector spend most of their time on creative tasks or activities which require high cognitive capabilities. German enterprises across all sectors need to consciously decide how they will leverage AI to achieve these levels of automation and free up capacity for value-adding growth.

3.

Players in the industrial sector should consider eight use cases of AI to achieve the next level of performance

Given the importance of AI for the German economy and specifically for the industrial sector, several key questions arise: What are the key applications of AI in the industrial sector? To what degree will these applications actually improve performance? How does the technology work in specific contexts and how exactly can it be applied? What will practically change in daily work and production processes? In the following, we will shed light on these questions in the context of eight use cases that demonstrate AI's manifold applications and enormous potential for performance improvement.

Exhibit 2

Impact of use cases across multiple industries

Impact: High, Medium, Low

	Use case	Aerospace	Automotive OEMs/commercial vehicles	Automotive suppliers	Industrial equipment	Semiconductors
Products and services	Autonomous vehicles	High	High	High	Low	Medium
Manufacturing operations	AI-enhanced predictive maintenance	Medium	High	Medium	High	Medium
	Collaborative and context-aware robots	Medium	Medium	Medium	High	Low
	Yield enhancement in manufacturing	Low	Low	Low	Medium	High
	Automated quality testing	Medium	Medium	Low	Medium	Low
Business processes	AI-enhanced supply chain management	Medium	Medium	High	High	Low
	High performance R&D projects	Medium	Medium	Medium	Low	High
	Business support function automation	Low	Medium	Low	Low	Medium

To do this, we first visually highlight the relative impact of each use case across five focal industries in the industrial sector and then describe each use case in detail. The five focal industries are: aerospace[10], automotive OEMs and commercial vehicles[11], automotive suppliers[12], industrial equipment[13], and semiconductors[14]. The impact of a use case on

10 Aerospace includes both commercial and military manufacturers of airplanes, unmanned aerial vehicles (UAVs), and satellites.

11 Includes automotive OEMs, manufacturers of construction equipment or agricultural machinery, and contract manufacturers.

12 Automotive suppliers include suppliers of assembled parts, components, and raw materials.

13 Industrial equipment includes manufacturers of various types of equipment such as power generation, transmission and distribution, storage equipment, industrial automation equipment, building technology, or test and measurement equipment.

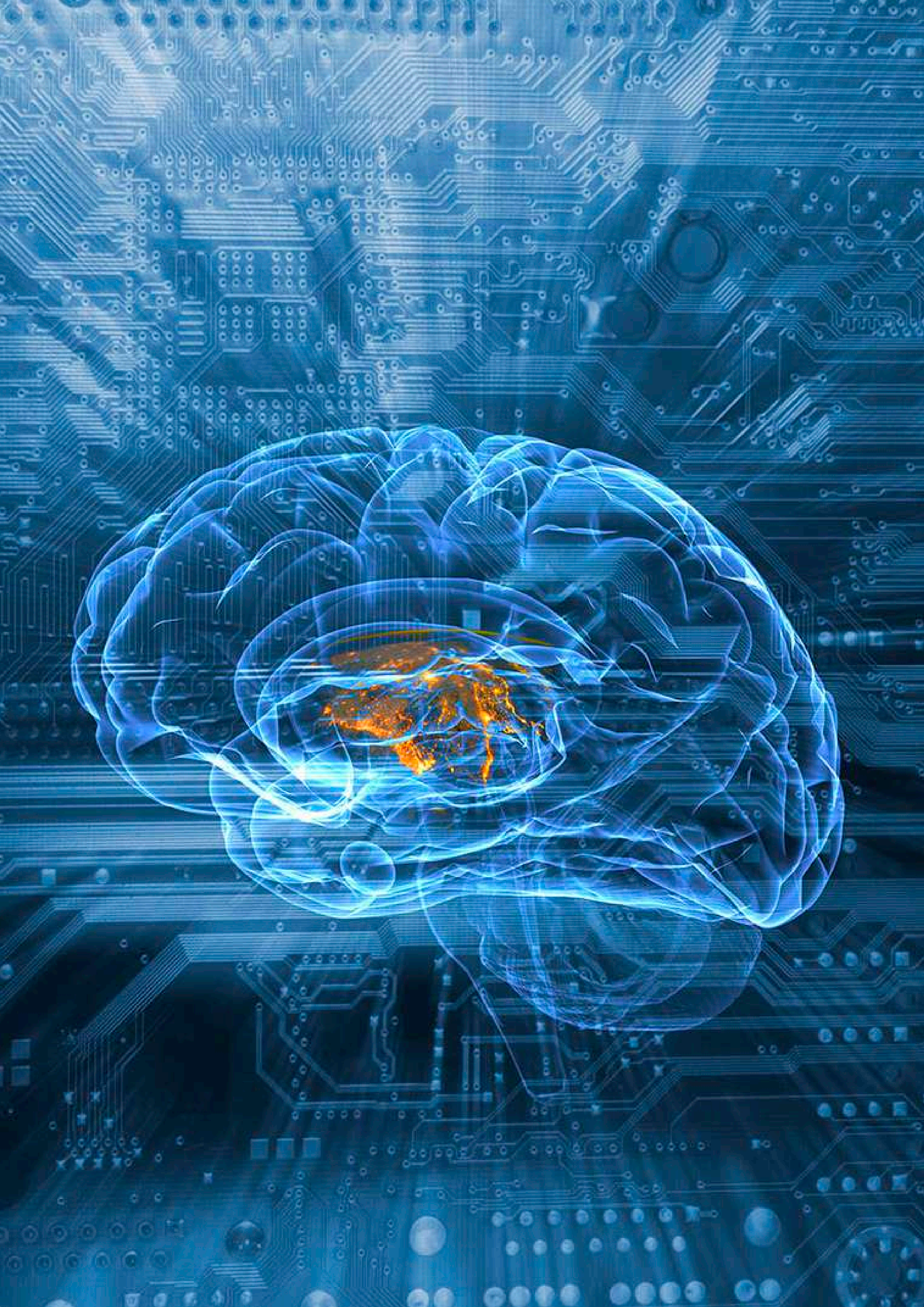
14 The semiconductor industry spans across integrated device manufacturers (IDMs), fabless companies and foundries, capital equipment manufacturers, and suppliers of electronic materials.

each of the five industries differs based on the idiosyncrasies of each industry. Impact levels were estimated as averages to reflect the heterogeneity of some industries. A heat map was generated based on estimates from both functional and industry experts at McKinsey and verified using bottom-up calculations that include various cost and revenue levers (Exhibit 2)[15]. As a tool, the heat map can help players easily identify relevant use cases in their particular industry for starting or continuing their journey to becoming a fully AI-enabled organization.

Looking at the results from a use case perspective shows that the future ubiquity of AI-enabled autonomous vehicles and drones will have a large impact on companies manufacturing these vehicles or supplying parts and components for them. AI-enhanced predictive maintenance is relevant for all of the focal industries because of their heavy reliance on manufacturing machinery. The potential of other use cases – such as yield enhancement in manufacturing – is greatest, however, when applied in the context of specific industries such as semiconductors, where yield is a major driver of economic performance. Still, use cases may be relevant levers across industries given a specific application context.

In the following, we describe all eight use cases in greater detail to elaborate the specific pain points that AI addresses, provide insights into the technology and methods applied, and estimate the impact of AI in various use-case-specific dimensions. To ensure that these descriptions are as vivid and concrete as possible, most of them focus on one specific industry or application. However, use cases generally extend across all industries mentioned and are typically easily transferrable to related applications. Nevertheless, technological adaptations may become necessary when changing the industry context or application context for a specific use case. The general logic, however, remains the same.

15 Among other factors, the impact estimates of use cases incorporate an industry-specific split of the operating revenue across cost types. E.g., the impact estimate for the use case “business support function automation” is medium to low in the five focal industries because the share of G&A on the operating revenue is relatively small.



3.1. Product and service improvement use case

Autonomous vehicles

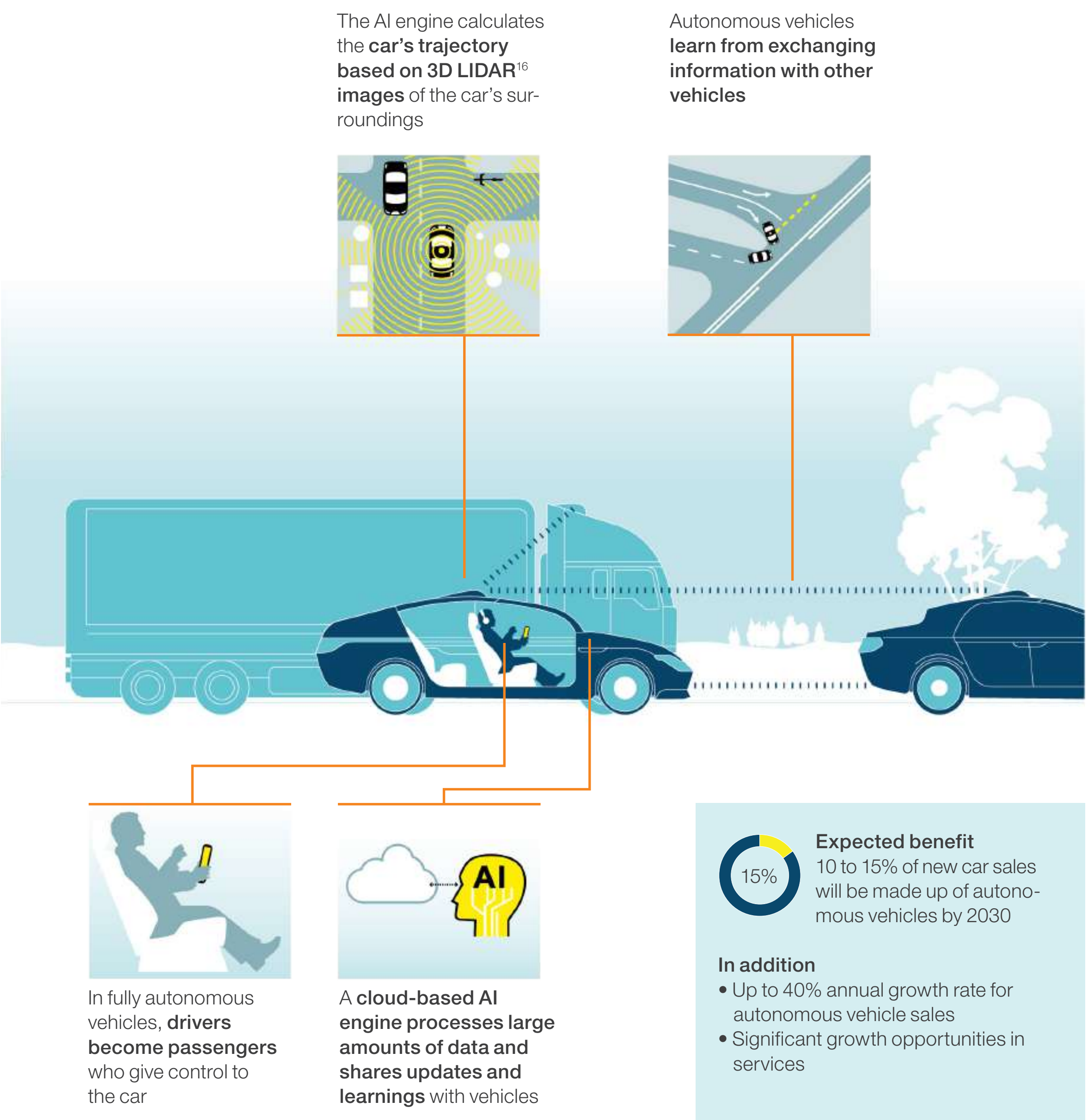
Context

Current and future means of autonomous transport come in various forms such as cars, trucks, unmanned aerial vehicles (drones), or agricultural machinery. The example of autonomous cars is particularly relevant due to their impact on society as a whole. Autonomous driving holds the promise of a smoother, safer, and more comfortable mobility experience. The automotive industry is on a continuous journey from assisted to autonomous driving. Nowadays, the majority of advanced driver assistance systems (ADAS) such as pedestrian recognition are still realized with rule-based programming. Building and maintaining those systems, however, is complex. The number of situations that need to be covered is virtually indefinite, given the large variety and diversity of traffic scenarios. Therefore, defining a full set of rules is not only impractical but rather impossible. In addition, rule-based systems do not offer sufficient performance to efficiently process the entirety of required information from cameras and LIDAR[14] and radar systems for new applications like city driving. To complete the journey toward truly autonomous decisions, the use of modern AI approaches will become a prerequisite.

Approach

Currently, machine-learning methods like neural networks are already starting to complement and, in some cases, replace rule-based systems in ADAS modules. The first hybrid systems have emerged that add self-learning elements to conventional systems and are used, e.g., in Google and Tesla vehicles. In addition, several automotive start-ups aim at extending the usage of AI. Well-known examples are Argo.ai, Drive.ai, nuTonomy, Otto, Preferred Networks, and Zoox. The goal is to build fully integrated, learning-based systems which are enhanced by AI algorithms through four major steps: sensor processing, data interpretation, planning and decision making, as well as execution. The independent training of those systems requires large sets of sensor data and significant computing power. In AI-enabled autonomous cars, the system is trained by humans based on representative scenarios. From there on, autonomous vehicles learn from all situations they encounter to continuously improve performance. Eventually, these vehicles will be able to share their learnings through direct interaction or a centralized platform. Then, the accumulated knowledge of all vehicles on the market can be utilized to improve each individual vehicle. Training data can optimize machine learning algorithms, and the data gathered in the field can be processed largely centrally or offline. For autonomous vehicle operation, the expectation is that AI-based systems and additional learning iterations can be realized with limited but specialized computing power within cars. Advancements in self-driving cars are closely correlated with those of machine-to-machine interactions. As humans start to hand off their decision making to machines, the interaction between machines will become more important. Highly autonomous vehicles are expected to hit the roads in 2025 and become an established part of the mobility landscape by 2030. This timeline depends on technological progress, customer acceptance, and regulatory approval. Highly autonomous vehicles will likely feature a much higher utilization than nonautonomous vehicles as there is no economic reason for autonomous vehicles to be off the road except when refueling or for maintenance. Hence, the erosion of boundary between privately owned and public cars, which was started by car sharing, will progress further in the age of the highly autonomous vehicle.

16 LIDAR – Light Detection and Ranging; a method for measuring distances using laser pulses.



Impact

Starting around 2025, global sales of highly autonomous vehicles will grow significantly until around 2040 following an S-shaped curve. Already in 2030, global sales of highly autonomous vehicles could make up 10 to 15% of new car sales. An annual sales growth rate of up to 40% is expected that flattens out before 2040. Hence, automotive OEMs and suppliers can hardly risk not investing in autonomous vehicles. In addition, they should consider a strategic reevaluation of their business model. While total kilometers driven in a given year are expected to increase continuously until 2040, significantly higher utilization rates of autonomous vehicles as compared to traditional cars may result in stagnation of total car sales. Hence, growth opportunities may lie not only in autonomous vehicles, but also in additional services enhancing the overall mobility experience.

3.2. Manufacturing operations use cases

3.2.1 AI-enhanced predictive maintenance

Context

Predictive maintenance aims at improving asset productivity by using data to anticipate machine breakdowns. A well-established and relatively simple method of recognizing failures early on is condition monitoring. The complexity of forecasting failure is often due to the enormous amount of possible influencing factors. Data sources can be manifold and depend on the scenario. E.g., in engines, gear boxes, or air conditioning, analysis of sound can detect an anomaly in device operation. In switches, machines, and robots, vibrations can be measured and used to detect errors. Since new sensors and IoT devices can be integrated in production processes and operations, the availability of data increases drastically. AI-based algorithms are capable of recognizing errors and differentiating the noise from the important information to predict breakdowns and guide future decisions.

Approach

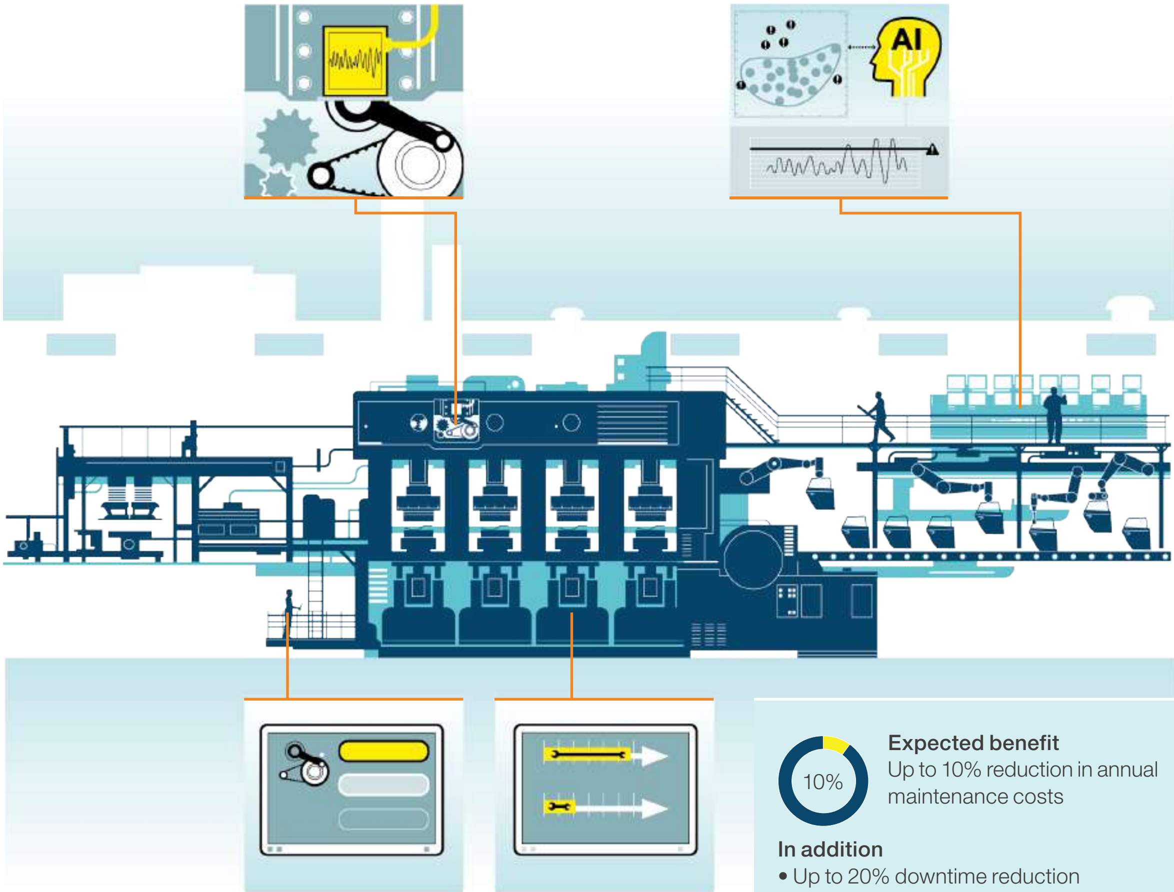
Machine-learning techniques examine the relationship between a data record and the labeled output (e.g., failures) and then create a data-driven model to predict those outcomes. This technique helps recognize patterns from historical events and either predict future failures or prevent them based on learnings from specific breakdown root causes. Companies like Neuron Soundware use artificial auditory cortexes to simulate human sound interpretation, thus automating and improving the detection and identification of potential breakdown causes. KONUX, one of last year's winners of McKinsey's "The Spark"[17] award for digital innovation, uses sensors to detect anomalies. Its cloud-based AI system continuously learns from alerts to further improve the overall performance of the system and give recommendations for optimized maintenance planning and extended asset life cycles. Recent applications of machine learning also combine supervised learning with unsupervised learning and feature[18] learning. This enables an automated classification of machine failure modes and also the identification of relevant features in the data, thereby enhancing expert domain knowledge. Both approaches greatly simplify the deployment of predictive maintenance systems while improving prediction accuracy. In addition to algorithmic advances, the use of a great variety of data sources beyond sensor outputs – such as maintenance logs, quality measurement of machine outputs, and, if applicable, external data sources such as weather data – enables prediction of events that were not possible to model before. Implementing AI-supported predictive maintenance takes, on average, six to eight weeks for pilot cases and several months for a full rollout. It may take longer if sensor development is involved.

17 The Spark is a joint award from Handelsblatt and McKinsey for excellence in innovation and products in the context of digitization in Germany (http://award.handelsblatt.com/the-spark/). In 2017, "The Spark" focuses on AI.

18 Data transformation techniques such as clustering that have the objective of generating a data representation that can be effectively used in machine learning.

Sensors detect e.g., sounds or vibrations and send their data to the AI engine for processing

ML algorithms (e.g., anomaly detection) accurately **predict maintenance needs** of machine parts



A maintenance worker is **automatically given suggestions on the predicted maintenance** and its schedule

Predictive maintenance greatly **reduces machine downtime** caused by maintenance work as compared to other approaches

Expected benefit
Up to 10% reduction in annual maintenance costs

In addition
- Up to 20% downtime reduction
- Up to 25% reduction in inspection costs

Impact

Comparing an AI-based approach to traditional condition monitoring or more classical maintenance strategies like usage-based exchange, a considerable improvement can be expected due to better failure prediction. Depending on the starting point and the level of redundancy, availability can sometimes increase by more than 20%. Inspection costs may be reduced by up to 25% and an overall reduction of up to 10% of annual maintenance costs is possible.

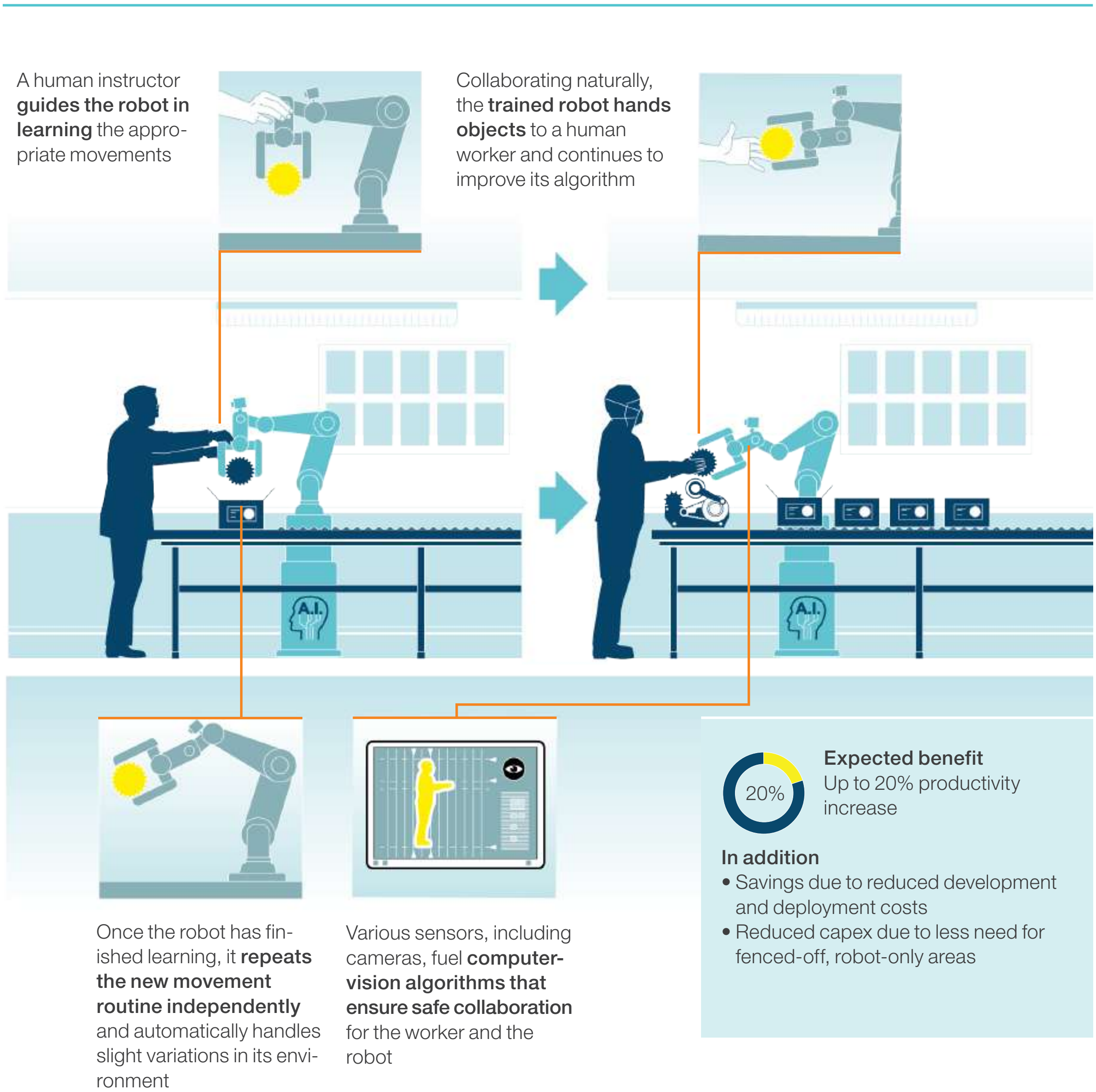
3.2.2 Collaborative and context-aware robots

Context

Today's industrial robots still face the limitation that they mostly cannot react to changes in their environment and that they need to operate in fenced-in, robots-only areas. Significant advancements in AI, however, are enabling a new generation of automation robots: flexible, non-special-purpose robots that need less configuration time and are easier to incorporate into specific environments. This includes a shared environment of robots and humans. The robustness of robots in unprepared environments as well as flexibility in the applicability of robotic systems has increased dramatically. There is a significant upside through further collaboration between the human workforce and context-aware robots, especially in assembly-heavy industries.

Approach

Key drivers behind the developments in collaborative and context-aware robots are advances in computer vision. Enhanced vision is enabled by more powerful computers and new algorithmic models that combine unsupervised and semi-supervised machine-learning approaches (see appendix). Within the field of computer vision, especially object recognition and semantic segmentation, i.e., the capability to recognize the object type (e.g., tool vs. component), are becoming increasingly important. They allow robots to behave as if they were aware of the context they operate in. Fully context-aware robots recognize the properties of the materials and objects they interact with. They are flexible and autonomous systems that are capable of safely interacting with the real world and humans. This is a notable shift since traditional robotic systems typically do not understand how to work or interact with a specific object, cannot react in a flexible manner to changes in an environment, and therefore need to follow predefined steps. E.g., new AI-enhanced logistics robots are taught how to recognize empty shelf space via camera automatically. This leads to a dramatic speed advantage in picking objects over conventional methods. Deep learning is used to correctly identify an object and its position. This greatly enhances flexibility as it enables handling of objects without requiring fixed, predefined positions. AI-enhanced logistics robots are also able to integrate disturbances in their movement routines via an unsupervised learning engine for dynamics. This capability leads to more precise maneuvers and an overall improved robustness of processes. Rethink Robotics is one company designing collaborative robots. Its learning algorithms allow for joint human-robot collaborative work spaces. AI enables the "programming" of a robot by simply showing the desired movements to it. A human "robot instructor" can take the arm of the robot and guide it through the desired movement, including gripping and releasing objects. Robot movement is then the result of the robot's replication and further improvement of the freshly learned movement combined with a computer-vision-based assessment of an object's position in space.



Impact

MGI estimates an automation potential across the German manufacturing sector of 55% of currently performed activities, freeing up capacity to focus on value-added work. Employing context-aware robots in logistics processes related to the delivery of product parts, e.g., is expected to produce efficiency gains of 5 to 10% in picking and a 15 to 20% reduction in travel time. Collaborative robots are particularly relevant with respect to tasks that are not fully automatable. In such settings they hold the potential of increasing productivity by up to 20%. Additional savings stem from reduced development and deployment costs in new tasks (due to the relative simplicity of instructing robots with new movement procedures) as well as simpler factory design made possible by a level of safety that eliminates the need for "robot-only" areas in factories with significant reduction in capex for fences. The implementation of flexible and adaptable robots will change the way manufacturing processes are set up. Manufacturing lines will look more like flexible work aisles where humans and robots work together side by side and an increasing number of tasks are performed by one system.

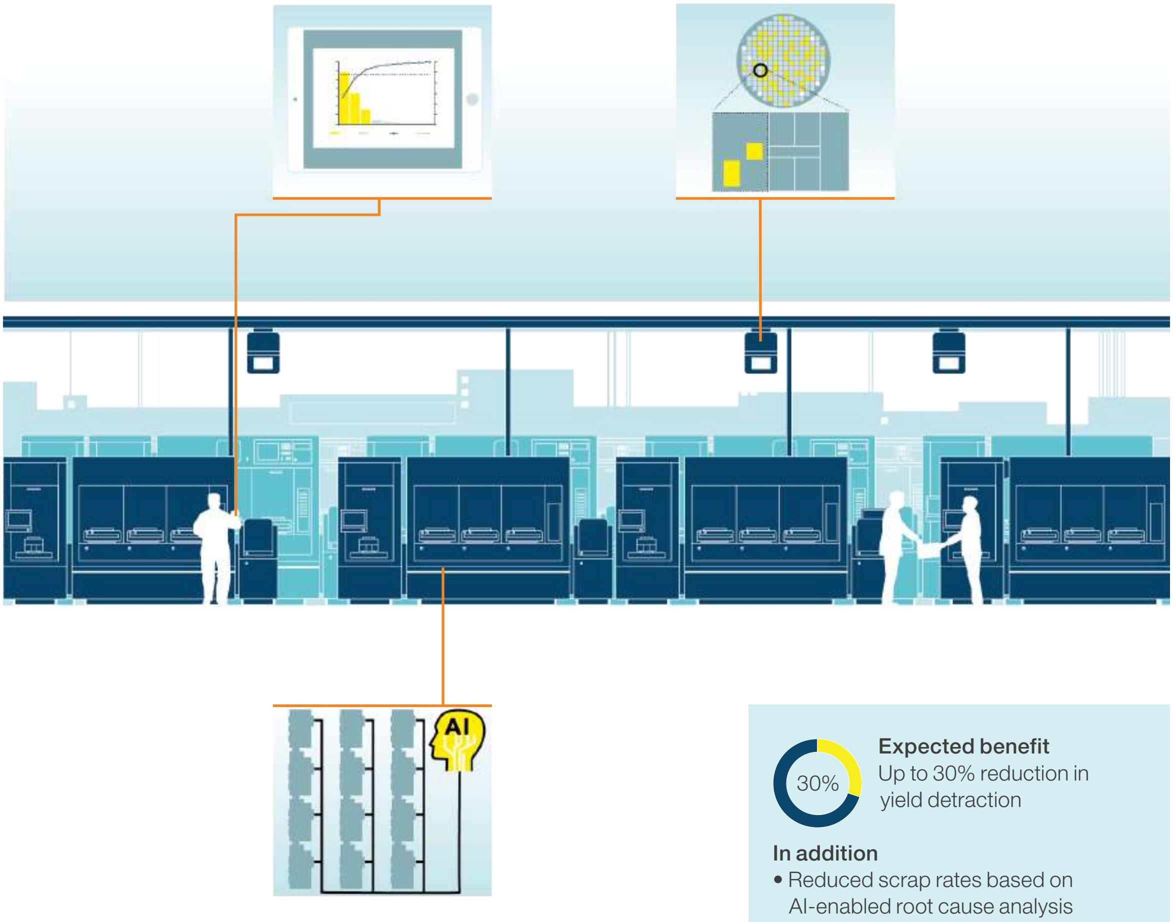
3.2.3 Yield enhancement in manufacturing

Context

Yield losses, i.e., products that have to be disposed of or need to be reworked due to defects, play an important role in complex manufacturing environments. The multistep semiconductor chip production process is a meaningful example because cycle times from the first processing of the wafer to the final chip are typically several weeks to months and include various intermediate quality-testing processes. Testing cost and yield losses within semiconductor production can constitute up to 20 to 30% of the total production cost. Data availability is typically high in semiconductor fabs due to their high degree of automation and advanced production equipment, often with archives that allow insights into detailed production information dating back months to years. However, systematic analysis and linkage of data sources across multiple tool groups is not necessarily performed.

Approach

Linking process control data with quality control and yield data serves as a basis to identify yield losses as well as the root causes of quality loss. Semiconductor manufacturers are starting to use AI engines to identify root causes of yield losses that can be avoided by changes to production processes or chip designs. Enhanced applications are designed to monitor and adjust subprocesses in real time. Established players are often supported by specialized analytics firms. Companies like Qualicent Analytics are improving yield using process or test parameters as predictors. The key variables are determined automatically from thousands of variables with noisy data. The AI engine is built such that it determines the optimized product operating conditions or process conditions to significantly reduce defects in manufacturing. Yield enhancement is also strongly linked to optimized chip design. Companies like Motivo have begun offering both supervised and unsupervised machine-learning systems that predict the locations of yield detractors. Known problematic patterns are algorithmically broken down into key components with the support of AI-based analytics. These components can be searched in existing and potential new designs to identify previously unknown problematic locations in the physical design layout of individual microchips.



Expected benefit
Up to 30% reduction in yield detraction

In addition
- Reduced scrap rates based on AI-enabled root cause analysis
- Reduced cost of testing due to AI optimization

Impact

The money saved by AI-based testing algorithms that lead to better yields spans several dimensions. First, AI-enabled root cause analytics can improve yield by reducing scrap rates. Second, minimizing the amount of equipment and maintenance required, AI can improve overall equipment effectiveness. Third, AI-optimized testing procedures are less expensive (e.g., fewer test wafers required for semiconductors), and, fourth, they can lead to higher throughput by reducing/stabilizing the flow factor. Overall, the use of AI can lead to a reduction in yield detraction by up to 30%.

3.2.4 Automated quality testing

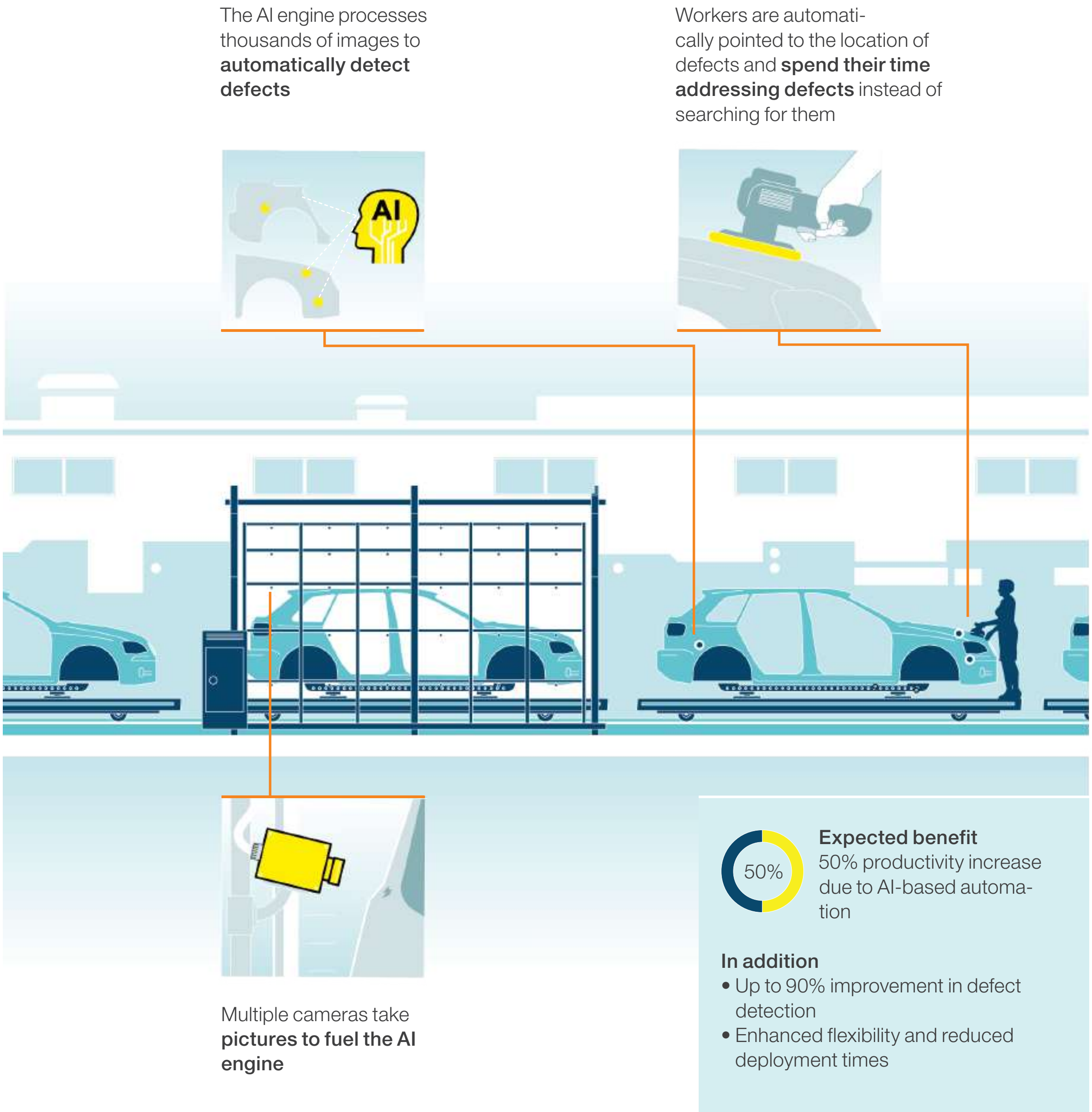
Context

Quality assurance in production is of critical importance for the long-term success of products and services. Quality assurance systems typically require intense upfront investment and extensive testing and calibration. One particular example is a visual quality inspection system for a production line of various products such as printed circuit boards or car bodies. Current automated approaches for visual inspection follow, e.g., a pixel-wise comparison of an ideal reference image of a product with an image of the specific product up for testing. Such a methodology is only feasible, however, if task- and environment-related preconditions are met. Ideally, perfect mounting of the product within the inspection tool is given, lighting conditions are the same across all inspections and types of defects are known in advance. In addition, an operator's trust in the results of the automated inspection process is of high importance. A large number of false positives may reduce this trust, thereby eroding any benefits from automation. Methodologies based on computer vision and machine learning hold the promise of overcoming these challenges.

Approach

State-of-the-art approaches for quality assurance and visual inspection in particular employ various forms of machine learning. In AI-enabled visual quality inspection, reference examples are created by visual imaging of good and defective products from different perspectives that fuel the training of supervised learning algorithms (e.g., based on deep neural networks). Previously unknown defect types can be identified by employing semi-supervised learning (see appendix). This leads to a further enhancement in accuracy. Modern machine-learning algorithms are capable of handling visual inspection and quality assurance of diverse products such as machined parts, solar panels, painted car bodies, or textured metal surfaces. Machine learning abstracts from differences in illumination, imperfect surface orientation, or the presence of irregular background textures and focuses on defects only. Hence, AI enables the detection of defects that would only have become apparent at processing steps much further downstream using a conventional approach. The University of Amsterdam's spin-off Scyfer, e.g., applies deep neural networks to automate the visual inspection of steel surfaces leading to improved performance especially in relation to complicated or rare defects. US-based Nanotronics combines AI with 3D microscopy for defect detection down to the nanometer scale. Machine-learning-based visual inspection costs less and is easier to implement than conventional approaches, so adoption is expected to grow quickly. Given the availability of open-source AI environments and inexpensive hardware in terms of cameras and powerful computers, even small businesses are expected to increasingly rely on AI-based visual inspection. E.g., a Japanese cucumber farmer has used Google's open TensorFlow library to create a computer-vision-based tool for rating cucumber quality.[19]

19 Source: https://cloud.google.com/blog/big-data/2016/08/how-a-japanese-cucumber-farmer-is-using-deep-learning-and-tensorflow



Impact

The recent advancements in AI-based quality assurance promise productivity increases of up to 50%. For AI-powered visual inspection, detection accuracy of defects increases while simultaneously flexibility is enhanced and deployment times decrease. Improvements of up to 90% in defect detection as compared to human inspection are feasible using deep-learning-based systems. This reduces the costs associated with shipping bad products. Efficiency and speed are improved as the need for human input is lowered significantly. Generally, insights from AI-based quality testing can be used for root cause analysis to improve the overall production process.

3.3. Business processes use cases

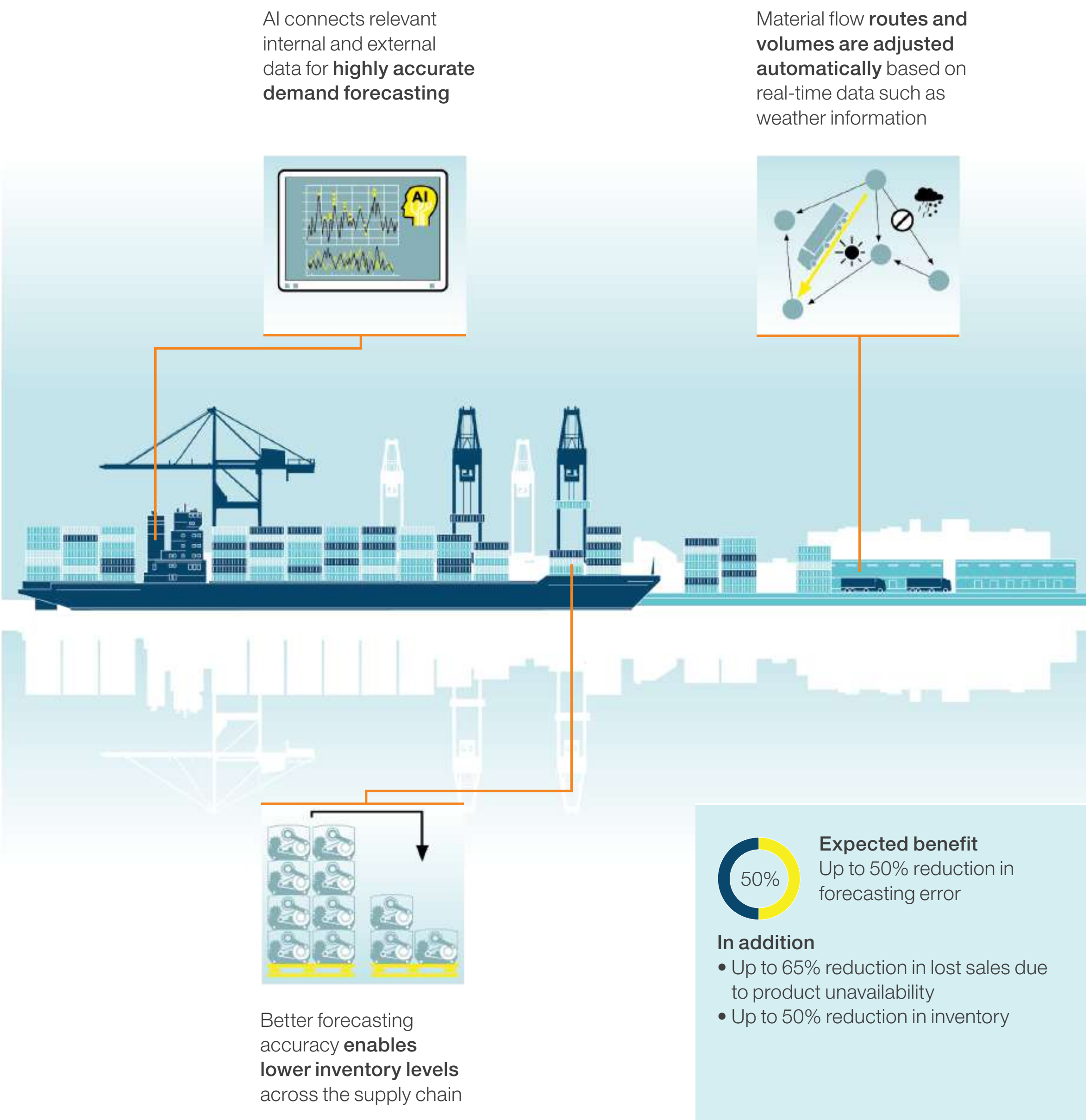
3.3.1 AI-enhanced supply chain management

Context

A well-functioning supply chain is the backbone of virtually every industry. To enable a close match between supply and demand, a highly accurate forecast of demand combined with optimized replenishment strategies is key. This is especially relevant in just-in-time production settings where short cycles and having just the right amount of inventory are critical for achieving a competitive advantage. Currently, a supply chain that works at maximum performance at all times is almost impossible to achieve. This is caused by enormous internal and external complexity. Internal factors include new product introductions, the expansion of distribution networks, or short-term promotions in a B2C context. The presence of long-tail items, extreme seasonality, and changes in customer perception or media coverage are among the external factors. Traditional systems for forecasting and replenishment are overwhelmed by the amount of data that, e.g., come from IoT devices and the sheer number of influencing factors. Hence, just-in-time production often relies on a supply chain that is highly efficient under standard conditions but at the expense of flexibility, i.e., the inability to quickly and effectively react to upstream or downstream changes.

Approach

Supply chain leaders are starting to realize the ability of machine-learning-based methods to increase forecasting accuracy and optimize replenishment. The objective is to reduce the bullwhip effect and increase flexibility. AI-powered supply chain optimization not only focuses on performance in a given setting, but can also flexibly adapt to changes in the product mix or the distribution network due to unforeseen events. In addition, future systems address the entire value chain from the supplier of raw materials to the end customer. Supervised learning approaches, e.g., based on Bayesian networks, not only incorporate historical sales data and the setup of the supply chains, but also rely on near real-time data, e.g., regarding advertisement campaigns, prices, or local weather forecasts. In addition, data granularity can be driven to much higher levels than before by focusing on, e.g., individual stock points. The objective is to achieve a fully automated, self-adjusting decision making system for supply chain management. Demand spikes are predicted accurately, and the routes and volumes of material flows are adjusted automatically. Some companies have started experimenting with their own internal solutions to enable AI-based predictive forecasting and replenishment. Established players like Blue Yonder promote AI techniques to optimize forecasting and replenishment while being capable of simultaneously adjusting pricing.



Impact

AI-based approaches to forecasting are expected to reduce forecasting errors by 30 to 50% in some settings. The benefits of applying AI in supply chain management, however, go far beyond that. Lost sales due to product unavailability can be reduced by up to 65%. Costs related to transport and warehousing and supply chain administration are expected to decrease by 5 to 10% and 25 to 40%, respectively. Due to AI, overall inventory reductions of 20 to 50% are feasible.

3.3.2 High performance R&D projects

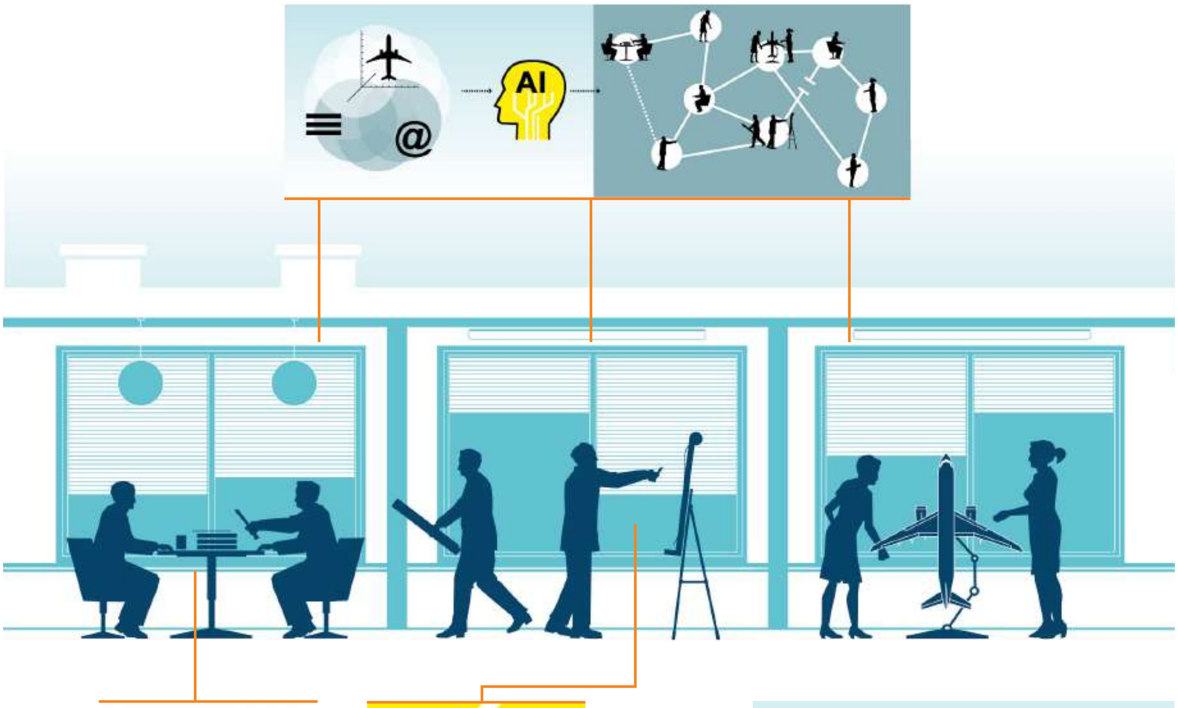
Context

Managing R&D projects at both maximum effectiveness and efficiency is a highly challenging task that requires dealing with technical and market uncertainties while under pressure of ambitious time-to-market targets and budget limitations. This situation is exacerbated by often poorly defined interfaces on the functional level, e.g., between design, development, and production, and also on the level of individual R&D teams. Competencies and proper communication channels remain largely unclear. Given the long-term nature of some projects and the complexity of measuring progress, it is hard to estimate the right point in time to cancel a project in favor of more promising ones from a portfolio perspective. Among the negative consequences for R&D are higher costs and longer times to market, often combined with individual frustration. Outcome variability is high. “Zombie projects” with unclear status and milestones tend to live too long and burn money until they are finally terminated.

Approach

AI-based methodologies hold significant potential for improving R&D project prioritization and raising performance within individual projects, thereby freeing up budget and raising individual projects’ efficiency. R&D projects at Formula 1 racing teams provide the perfect testing ground for the promise of AI-based R&D performance boosts. In F1, the race on the track is accompanied by a race in engineering. Conventionally, a brute force approach with several thousands of R&D projects – some of which are in competition with each other – is used to determine the optimal configuration of racing cars. Failure rates of R&D projects are around 90%. In the context of F1 and also in other industries, McKinsey’s own QuantumBlack has successfully employed AI to streamline the R&D process and identify the most promising R&D projects early on. In an end-to-end approach, QuantumBlack combines data from various sources such as product life cycle management, HR, finance, or team e-mails to fuel machine learning and econometric methods and forecast performance detractors. Interestingly, it is not the sheer amount of available data but the breadth of integrated data sources that is a key success factor in this approach. Team dynamics in terms of the interfaces between functions and also in-between individual project teams often turn out to be a key lever for improving performance and predicting project success or failure. Ideally such interfaces are defined on a per-project basis taking individual project characteristics into account. Projects with an inherently large degree of interconnectedness should have a higher intensity of communication while rather isolated projects should accordingly have a lower communication intensity in a fully optimized setting. The team staffing process and team member familiarity can also play an important role in driving efficiency. AI-based methods allow QuantumBlack to infer the optimal project environment and degree of communication for each project, thereby significantly raising performance. Once key patterns are discovered, early-warning capabilities can be established that enable managers to predict project performance early on and react accordingly.

AI optimizes R&D project communication networks using data from CAD, HR, e-mail, and other sources to improve R&D performance by suggesting suitable levels of interaction



AI-based analysis uncovers needs for **frequent communication between R&D projects ...**

… and also suggests **reduction of exchanges** between more independent R&D projects

15%

Expected benefit
10 to 15% productivity gain in R&D projects

In addition
- 10% time-to-market acceleration; up to 40% possible in some settings
- Decrease in development risk

Impact

QuantumBlack's AI-based approach to increasing R&D project performance results in productivity gains of 10 to 15% depending on the industry. Time to market is typically accelerated around 10% while up to 40% is possible in specific settings. This results directly in additional revenues. Time-to-market advantages are enabled by a focus on the most successful R&D projects and increased project efficiency. In addition, a considerable decrease in overall product development risk is expected.

3.3.3 Business support function automation

Context

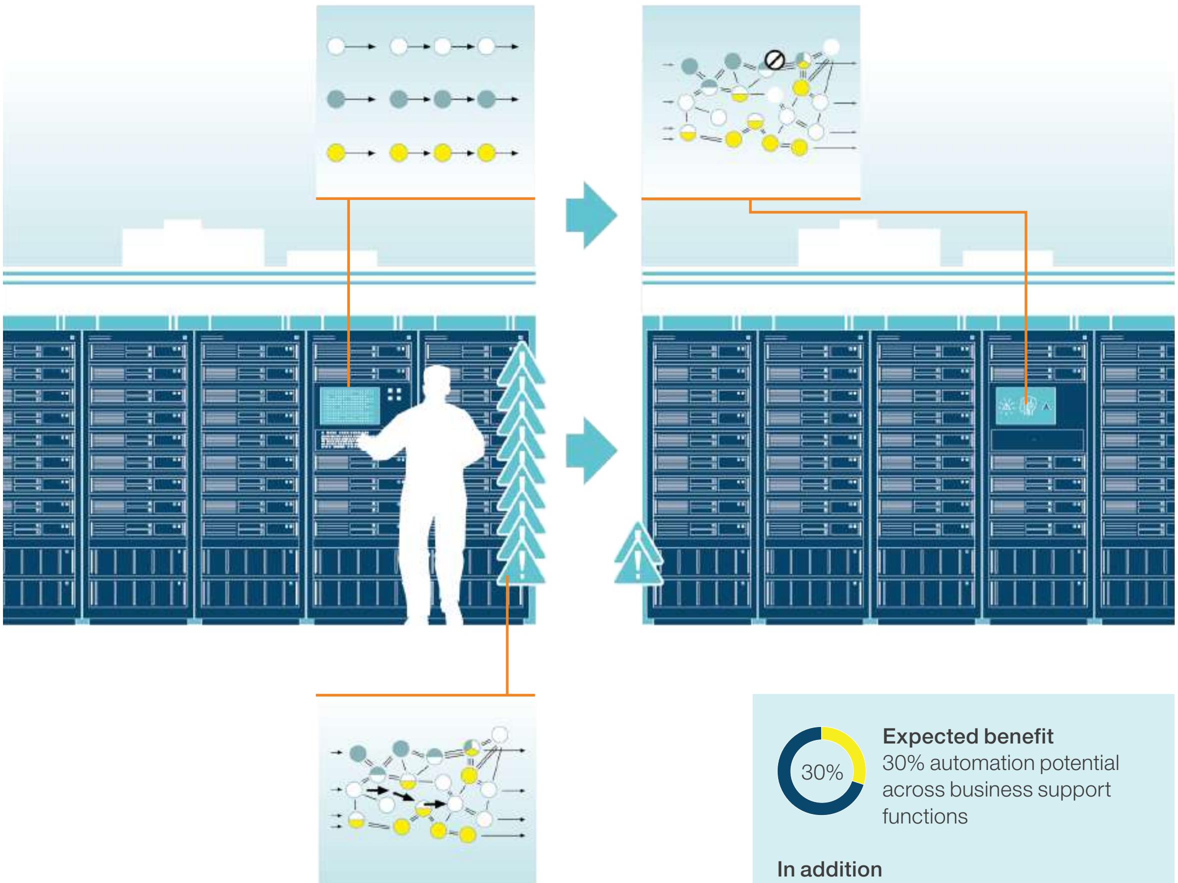
While business support functions such as Finance, HR, and IT are key to ensuring a business' effective operation, they are usually resource intense. Challenges arise due to globalization, high flexibility requirements, and cost pressures. The next wave of digitization enables companies to leverage robotic process automation (RPA) in business support functions to raise efficiency. In RPA, activity sequences are recorded and then repeated automatically. Such an approach fails where task complexity surpasses a certain threshold. Here intelligent process automation (IPA) – the combination of RPA and AI – provides a competitive advantage. IPA has an especially high potential for automating tasks in business support functions that are naturally supported via computer systems such as IT or finance. This is because current AI systems function best in environments where the range of inputs and outputs are clearly defined, such as controlling server operations in IT. In contrast, current AI systems tend to struggle in situations where tasks require higher-level cognitive abilities – e.g., the discussion of a performance evaluation where human empathy is needed.

Approach

The automation of complex tasks based on IPA often combines natural language processing with other applications of machine learning. One example is the automation of an IT service desk using AI. This example is particularly interesting because the resolution of IT tickets requires both an understanding of the ticket's intent and the ability to determine the best course of action. One way of solving this challenge consists of a two-step approach. In the first step, human problem solving strategies are codified. Each piece of codified strategy represents one simple task, such as a specific server configuration change, that may be relevant for solving a variety of IT-related issues. In the second step, these pieces are then fed into a self-learning AI algorithm. Depending on a specific IT ticket's content, the AI engine combines individual pieces of knowledge to create a tailor-made ticket resolution process. The first step makes sure that the AI engine builds upon an organization's unique knowledge that was gradually built up by humans based on the existing IT landscape. The second step employs reinforcement learning, the same technique used by Google's AlphaGo to beat the world's Go champion, to ensure that the AI engine continuously learns from each step in ticket resolution. Feedback enables the system to adapt and propose a more efficient ticket resolution approach in case a similar IT ticket is posted in the future. The automation potential of such a solution rises with the number predefined or additionally created steps that the AI algorithm can draw from.

A human **expert's knowledge is codified** to fuel the AI engine

Based on experience, the **AI engine refines its problem solving capability**



The **AI engine automatically finds the best path to a solution** based on codified knowledge

30%

Expected benefit
30% automation potential across business support functions

In addition

- Up to 90% automation potential in IT service desk
- Overall greater accuracy and consistency, increased scalability and speed, and traceability of results

Impact

Business support functions have a high potential for automation through IPA of approximately 30%. For the specific example on IT service desk automation mentioned above, a degree of automation of around 90% is possible. The self-learning capability of the algorithms enables further efficiency boosts. Automation is accompanied by greater accuracy and consistency, increased scalability and speed, and traceability of results – all at around-the-clock availability. In addition, the effectiveness of decision making and business analysis is expected to improve due to data-driven insights that are not available without the use of AI.

4.

Players in the industrial sector should follow five pragmatic recommendations for enabling AI-based performance improvements

Certainly, no singular, standardized course of action may enable all AI-driven performance improvements within industrial organizations. There are, however, a number of recommended approaches and perspectives that aspiring companies in the industrial space should adopt. These are the result of our findings concerning AI as well as our observations of the most successful players in both the industrial sector and its adjacent industries with similar AI-related challenges.

1. **Get a grasp of what AI can do, prioritize use cases, and don't lose sight of the economics – without a business case no innovation survives.**

> "30% of surveyed executives admit that senior management lacks sufficient understanding of AI which makes building a business case for AI challenging."
>
> *The Economist* (Artificial Intelligence in the Real World), 2016

The speed of innovation in the field of AI is overwhelming. Only a decade ago, good speech recognition algorithms were almost considered science fiction. Today, millions of people rely on them for communicating with the machines around them. E.g., people obtain driving directions by giving voice commands to SIRI, Apple's personal assistant, or interact almost naturally with Echo, Amazon's personal assistant. Even though AI is becoming second nature in certain realms, it is important not to make hasty investments in trendy technology without understanding how it can bring value to one's own business. This is what a business case is for.

The first step in establishing a solid AI business case is separating the hype and buzz around AI from its actual capabilities in a specific, real-world context. This includes a realistic view of AI's capabilities and an honest accounting of its limitations. A prerequisite is certainly a sufficient, high-level grasp of how AI works and how it differs from conventional technological approaches and what it takes to get the AI engine started in one's own business.

Even then, building an actual business case is much easier said than done. Much of the information is imperfect, the returns are often fuzzy and unclear in the early stages, and the doubters line up ready to block new ideas from entering the commercialization process.

What helps is a pragmatic prioritization of potential use cases for AI-enhanced products and processes (e.g., building on the selection discussed in the previous chapter). It should contain two major dimensions. Besides the technical feasibility and complexity of the required AI engine, the overall impact potential – derived from estimates of the financial baseline and optimization potential – should be key prioritization parameters.

2. **Develop core analytical capabilities internally but also leverage third-party resources – trained people are scarce.**

> "When it comes to applying artificial intelligence, a highly skilled combination of algorithms is key. Take cooking as an analogy: even though anyone can buy different spices (i.e., commoditized algorithms), it takes a good cook to create a well-seasoned meal (i.e., combine the right algorithms in the right way)."
>
> *Andreas Kunze,* CEO at KONUX

To really capture the performance-boosting value of AI, companies need to build strong internal capabilities and cooperate with renowned companies or start-ups in the area of AI. Relevant roles that companies need to fill are "quants" who develop the required algorithms and "translators" whose core expertise is in bridging the gap between data scientists and management. At the center of the translator's role lies the ability to help management differentiate hype and buzz around AI from real-world applications and ensure value creation tailored to a specific business. Quants design and develop the AI engine. The current situation for obtaining talent to fill these roles, however, is far from ideal. E.g., out of approximately 100 million North American workers, only 8,900 are trained data scientists.

Given the scarcity of talent and the typical difficulties associated with projects based on new technologies such as AI, it is most practical to combine internal means (development) and external means (recruiting) of obtaining skills with partnering to get the AI engine up and running. During such partnership a process following the phases "build, operate, transfer" has proven to be successful. The tasks performed by quants and translators vary for each phase. In the build phase, translators help prioritize AI-based use cases due to their ability to assess and communicate what is feasible from both a technical and a business perspective. Internal quants and those at the external partner collaborate closely and lay the foundation of an AI application by integrating systems, data, and algorithms. During the operate phase, prioritized use cases are tested to assess the value creation potential of specific AI applications and then they are scaled. Translators ensure that the new solution is accepted throughout the organization. The outcome of the operate phase is a fully working and scaled AI engine that taps the entire value pool. In the transfer phase, all required knowledge to run the scaled system is transferred to internal personnel. Translators are responsible for managing the skill transfer from the external partner to the company owning the AI engine. For a full transfer, companies usually require a number of internal quants to run the AI systems, perform updates, and identify improvement needs. Given the challenges of each phase within the build-operate-transfer process, we suggest complementing internal up-skilling by hiring top-notch analytics talent to oversee the entire process – also on a top-management level.

To address the issue of talent scarcity, it is key to establish an attractive AI-centric environment where unique data and sufficient computing power are easily accessible. Recruiting pure data scientists also provides a good opportunity to adjust AI resources dynamically.

3. Store granular data where possible and make flat or unstructured data usable – it is the fuel for creating value.

> “Data is the new oil. It’s valuable, but if unrefined it cannot really be used. It has to be changed into gas, plastic, chemicals, etc. to create a valuable entity that drives profitable activity; so must data be broken down, analyzed for it to have value.”
>
> *Clive Humby,* UK mathematician, 2006

Data is at the heart of the disruptions occurring across economies and it has been recognized as an increasingly critical corporate asset. Without data it is impossible to get the AI engine started. Because of this, business leaders should know what their data and the information therein is worth, and where they can obtain the data relevant for their company’s future success. There are the well-known examples of Google and Facebook, who obtain most of their revenue through insights they extract from the enormous quantities of data their customers generate on a daily basis by using their services.

One important capability will be making data usable that is not available in a relational format or that cannot be analyzed with traditional methodologies. Examples include pictures and voice transcripts but also data generated by sensors and machines. While the latter is basically structured data, its size and format makes it hard to analyze in the traditional context using relational databases[20], which have been around for 30 years in the business context. Estimates say that from all the data produced in the manufacturing context, 90% of it is flat data without relational structure. Making this data usable requires new approaches that can efficiently handle both data volume and types. Such approaches are most importantly NoSQL[21] technologies with the special purpose of efficiently storing and processing data in its original fidelity, e.g., based on frameworks such as Apache Hadoop[22]. Keep in mind that data needs to be present in a format that can easily be tailored to a specific approach to AI and machine learning. E.g., supervised learning techniques require labeled data in the training mode (see Text Box 1).

Customer sentiment or geo-locational real-time events are examples of differentiating types of data that will help organizations build distinctive AI-enabled services – the competition to gain exclusive access to such data will therefore intensify. On the other hand, certain data may only become valuable if combined with other data sources in a larger ecosystem. Thus,

20 A database where, in simplified terms, data is stored in tables that are linked to each other via keys.

21 A database that, in simplified terms, deviates from the relational model so that data is stored in different structures, e.g., in graphs. Potential advantages of NoSQL databases over conventional relational databases are speed or flexibility.

22 A framework that enables distributed storage and computing of large data sets using mainly commodity hardware. The result is a faster and more efficient processing as well as higher scalability as compared to conventional approaches.

cross-organizational data utilities may become more common to enrich and contextualize data and make it available to participants in the closed data ecosystem.

Given the rapidly increasing data output from sensors, machinery and social networks, organizations face challenges in how to handle such massive streams of data. While some use cases for such data will be very concrete with clear data requirements, other potential uses of data will be fuzzy or not yet fully defined. Some use cases will require significant time series of data (sometimes making up for lack of data quality), while for others, data becomes stale quickly (e.g., when analyzing social media data that is affected by frequent changes in trends). The situation requires a thoughtful approach on which data to store in its original granularity and which to aggregate or preanalyze. With increasing data storage capacities in the cloud as well as more powerful computing facilities close to the sensor, flexibility increases rapidly.

4. Leverage domain knowledge to boost the AI engine – specialized know-how is an enabler to capture AI's full potential.

> "Regardless of how good certain software engineers are or how powerful the algorithms, in order to truly benefit from machine-learning capabilities, companies need to rely on their domain knowledge: an in-depth understanding of their business, process, and industry."
>
> ***Holger Kleck,*** Head of IT Steering and Supporting Processes at AUDI AG

Possessing specialized knowledge on specific domains (e.g., on the parameters for a certain manufacturing process) is one advantage that OEMs will likely not lose to start-ups or service providers, and that they should be well aware of. In relation to applying AI and machine learning to business problems, domain knowledge can help companies in two ways.

First, within their own industry or technological environment, companies are able to best describe the problem to be solved using AI and typically have a deep understanding of the dependencies between systems, technologies, and players.

Second, applications of AI can make domain knowledge an integral part of the system. E.g., domain knowledge can be codified and provide a significant boost to the performance of an AI algorithm before self-learning starts.

In summary, we suggest viewing AI as a tool that can be applied to many problems – however, without a deep understanding of the context, it will fall short of any expected improvement potential.

5. **Make small and fast steps through pilots, testing, and simulations – the AI transformation does not require large up-front investments, but agility is a prerequisite for success.**

> "Cognitive Scale [...] is one of the new service providers adding more intelligence into business processes and applications [...]. Using their '10-10-10 method' they deploy a cognitive cloud in 10 seconds, build a live app in 10 hours, and customize it using their client's data in 10 days."
>
> *Brad Power* in The Harvard Business Review, 2015

If you want to effectively leverage the power of AI within your own company, you need to gain experience as soon as possible. As mentioned earlier, basic programming interfaces can be obtained at a low cost or even for free and if required, computational power at a larger scale can be accessed through cloud-based solutions. Companies can build up initial process know-how with the help of third parties while maintaining ownership of the underlying data and domain knowledge.

Best practice for AI implementation within a company is based on agile management and keeping an open mind on the real power of AI. Neglecting both is a typical pitfall and hinders a successful journey. Small and fast steps will assure the right focus, e.g., through simulation-based pilots that allow companies to quickly test the impact estimates in the business case. Best-practice companies set up cross-functional AI taskforces which are able to prototype a solution in one to three weeks given that data is readily available, test it with the business units, and decide how to proceed. Naturally, the exact pilot timeline depends on the scope of the project, but an agile approach will ensure an efficient use of resources.

To get the most out of AI in the long run, an organizational culture open to the collaboration of humans and machines is also required. Trust is a key enabler here. Due to the interplay of training and inference (see Text Box 1) in AI, the relationship between a machine's inner workings and the results it produces can become rather obscure. Instead of an algorithm's predetermined steps it is – in a sense – the data used to train it that leads to a certain outcome. Humans will need some time to adjust to this paradigm shift. Hence, the creation of an AI-ready culture should be a priority early on.

Outlook

Get started early with the journey towards a fully AI-enabled organization

The application of AI to business problems has never been as easy and promising as today. All pieces of the puzzle – effective algorithms, high-performance computing hardware, and sensors generating data – have fallen into place and AI is ready to scale. Especially the industrial sector with its highly automatable tasks and increasingly connected devices holds great potential for benefiting from AI. We have demonstrated – based on eight practical use cases – that already today AI can lead to performance boosts especially in the industrial sector. Five pragmatic recommendations guide the way towards getting started with AI almost immediately.

Getting “your hands dirty” by testing the first prospective applications of AI technology in your company does not require long preparation or a large up-front investment. Jumping in holds the benefit of producing early results and helping your company make quick progress on its journey toward becoming an organization that embraces the full potential of AI.

What are you waiting for?

Appendix

Nomenclature and terminology of AI

- *Artificial intelligence (AI)* is intelligence exhibited by machines, with machines mimicking functions typically associated with human cognition. AI's functions include all aspects of perception, learning, knowledge representation, reasoning, planning, decision making, and dynamically adapting to new contexts.

  The point of adapting to new contexts is an important one that – among others – differentiates strong AI from weak AI. What is the scope in which AI is able to transfer concepts learned in one context to another context? To what extent can the way knowledge is represented within AI be adapted based on new information or experiences? The first step toward strong AI is to give up the separation between a "learning mode" (see "training" below) and an "operating mode" (see "inference" below) where AI acts but does not incorporate learnings from prior interactions into future ones. Beyond this first step, the level of context changes an AI system is capable of depends mainly on processing power, the data it has access to, and the manually coded boundaries defined by human programmers on the machine-learning algorithm at the core of AI. For many use cases, early implementations will involve little or no context change capabilities because this is simpler and more efficient to program. Those capabilities will be increased over time, but in most cases not within the next three to five years. For the sake of simplicity and due to our focus on the business context, this report does not make the distinction between weak and strong AI.

- *Machine learning (ML)* describes automated learning of implicit properties of data. It is a major component for implementing AI since its output is used as the basis for independent recommendations, decisions, and feedback mechanisms. Many of the current algorithmic advances take place in the area of ML. The transition from ML to (strong) AI is fluid and the terms are often used interchangeably, especially in the business context. In a stricter sense, machine learning is an approach to creating weak and strong AIs. Creating systems that are able to do context changes has, so far, only been possible using machine-learning approaches.

- Depending on the problem statement, various algorithms may be applied, leading to the final model in an *iterative learning and refinement process.* These algorithms are – in contrast to those in traditional computer programs – *adaptive to changes* in their environment (self-learning). Learned data properties are generalized and used, e.g., for pattern recognition or prediction based on previously unseen data. Regarding ML's iterative learning and refinement process, types of learning are divided into *supervised learning, unsupervised learning, and reinforcement learning,* depending on the amount of structured input and provided feedback that are used for learning. In supervised learning, the algorithm is trained using labeled data, i.e., input data is associated with the desired output ("correct answer given"). Unsupervised learning algorithms find structure in unlabeled data ("no answer given"). *Semi-supervised learning* falls in-between supervised and unsupervised learning and uses both labeled and unlabeled data. In reinforcement learning, feedback includes how good the output was but not what the best output would have been. In practice, this often means that an agent continuously attempts to maximize a reward based on its interaction with its environment. For this purpose a mixture of exploration and exploitation is employed.

- The distinction between *training* and *inference* is crucial in the context of machine learning. Before a machine-learning algorithm can be put to work, it needs to be trained using large amounts of data to adjust underlying model parameters and optimize performance. This is usually computationally very intensive and scales with the size of the data set and the complexity of the model. Once training is finished, the machine-learning algorithm has gained a new capability. The process of making use of this new capability is called inference. The trained machine-learning algorithm receives a signal of previously unseen data and then – given its training – infers the most appropriate output. Computational cost of inference scales with inference frequency and model complexity.

- One key property of AI that is often overlooked is *knowledge representation.* While there are methods to represent known information well (e.g., Bayesian Networks that represent information in the form of probabilities), it has proved challenging to automatically learn *scalable* knowledge representations. Ideally, those can be reused in later stages of learning and recombined to learn more complex concepts. This is one of the key strengths of deep learning.

- Since the late 2000s, *deep learning* has been the most successful approach to many areas where machine learning is applied. It is based on learning neural networks[23] and utilizes large amounts of data. Neural networks consist of many layers of nodes with each subsequent layer representing a higher level of abstraction. In most contexts neural networks are trained using supervised learning, even though reinforcement learning has recently gained popularity.

  The concepts used today are closely related to those of the 1980s. However, in comparison to artificial neural networks used back then, modern deep-learning systems often have ten or more layers while traditional systems only had three layers. The increasing computational power of such networks has reduced the need for the preprocessing of input data. Surprisingly, deep learning works well for many pattern recognition tasks without alterations of the algorithms as long as enough training data is available. Thanks to this, its uses are remarkably broad and range from different visual object recognition tasks to the complex board game "Go."

23 An artificial neural network is a computational model that loosely imitates a network of neurons in a human brain. An artificial neuron sends a signal to another neuron in the network if the – usually weighted and mathematically transformed – sum of input signals surpasses a certain threshold. Training the network with data leads to adjustments of the weights.

Important notice

This report has been prepared by a McKinsey team on the basis of public sources as well as proprietary information gathered via in-depth interviews with industry thought leaders and extensive research.

Neither the report nor any part of its contents are intended or suited to constitute the basis of any investment decision regarding any company operating in the markets covered by this report or any similar markets (including, without limitation, the purchase of any securities of any listed company or in connection with the listing of any company).

This report contains certain forward-looking statements (i.e., estimates, forecasts, and projections about industries, trends, players, and consumers). By their nature, forward-looking statements involve uncertainties because they relate to events that may or may not occur in the future. This applies particularly to statements in this report containing information on future developments; earnings capacity; plans and expectations regarding the business, growth, and profitability of companies operating in the markets covered by the manufacturing industry or any related markets; and general economic and regulatory conditions and other factors that affect such companies. Forward-looking statements in this report are based on current estimates and assumptions that are made to the McKinsey team's best knowledge. These forward-looking statements and the overall markets about which they are made are subject to risks, uncertainties, and other factors that could cause actual situations and developments to differ materially from those expressly or implicitly predicted by these forward-looking statements or not occur at all. In addition, McKinsey does not assume any obligation to update any forward-looking statements or to conform these forward-looking statements to actual events or developments.

McKinsey does not assume any responsibility toward any persons for the correctness and completeness of the information contained in this report.

Content contributors

Harald Bauer, Senior Partner, Frankfurt
Peter Breuer, Senior Partner, Cologne
Gérard Richter, Senior Partner, Frankfurt
Jan Wüllenweber, Senior Partner, Cologne
Knut Alicke, Partner, Stuttgart
Matthias Breunig, Partner, Hamburg
Ondrej Burkacky, Partner, Munich
Michael Chui, MGI Partner, San Francisco
Matthias Daub, Partner, Berlin
Kersten Heineke, Partner, Frankfurt
Holger Hürtgen, Partner, Düsseldorf
Philipp Kampshoff, Partner, Houston
Matthias Kässer, Partner, Munich
Richard Kelly, Partner, Stamford
Dominik Wee, Partner, Munich
Simon Williams, Chief Marketing Officer – QuantumBlack, London
Robert Feldmann, Senior Expert, Frankfurt
Malte Hippe, Senior Expert, Hamburg
Ricardo Moya-Quiroga Gomez, Senior Expert, Munich
Jürgen Rachor, Senior Expert, Frankfurt
Matthias Roggendorf, Senior Expert, Berlin
Michael Schmeink, Senior Expert, Düsseldorf
Christian Schröpfer, Senior Expert, Berlin
Jon Swanson, Senior Expert, Miami
Timo Möller, Head of McKinsey Center for Future Mobility, Cologne
Jonathan Tilley, Senior Practice Expert, Southern California
Ulrich Huber, Senior Knowledge Expert, Cologne
Doruk Caner, Engagement Manager, Spark Beyond Partnership, Helsinki
Sven Beiker, Expert, Silicon Valley
Klaus Kunkel, Knowledge Expert EMEA Operations Practice, Düsseldorf
Peter Udo Diehl, Associate, Zurich

Editorial team

Heinz Klein, Engagement Manager, Munich
Jan Paul Stein, Fellow Senior Associate, Munich
Jonas Binding, Engagement Manager, Stuttgart
Matthias Breunig, Partner, Hamburg
Marko Mojsilovic, Fellow Senior Associate, Berlin
Stefan Stahl, Engagement Manager, Munich

Production team

Jörg Hanebrink, Senior Communication Specialist, Düsseldorf
Martin Hattrup-Silberberg, Senior Communication Specialist, Düsseldorf
Heinke Maria Kunze, Senior Media Designer, Berlin
Kristina Leppien, Senior Copy Editor, and Bryce Stewart, Copy Editor, Munich

Use case illustrations: Tobias Wandres

Contacts

Harald Bauer is a Senior Partner at
McKinsey's Frankfurt office.
harald_h_bauer@mckinsey.com

Gérard Richter is a Senior Partner at
McKinsey's Frankfurt office.
gerard_richter@mckinsey.com

Jan Wüllenweber is a Senior Partner at
McKinsey's Cologne office.
jan_wuellenweber@mckinsey.com

Matthias Breunig is a Partner at
McKinsey's Hamburg office.
matthias_breunig@mckinsey.com

Dominik Wee is a Partner at
McKinsey's Munich office.
dominik_wee@mckinsey.com

Heinz Klein is an Engagement Manager at
McKinsey's Munich office.
heinz_klein@mckinsey.com

April 2017

www.mckinsey.com

Digital/McKinsey